KJCEM 24. 5. 022~034
September 30, 2023

www.kicem.or.kr (pISSN 2005-6095 / eISSN 2465-9703)
http://dx.doi.org/10.6106/KJCEM.2023.24.5.022

인공신경망 기반의 공공청사 공사비 예산 예측모델 개발 연구

김현진[1] · 김한수[2*]
[1]세종대학교 건축학과 석사과정 · [2]세종대학교 건축학과 교수

A Study on the Development of Construction Budget Estimating Model for Public Office Buildings based on Artificial Neural Network

Kim, Hyeon Jin[1], Kim, Han Soo[2*]
[1]Graduate Student, Department of Architecture, Sejong University
[2]Professor, Department of Architecture, Sejong University

Abstract : Predicting accurately the construction cost budget in the early stages of construction projects is crucial to support the client's decision-making and achieve the objectives of the construction project. This holds true for public construction projects as well. However, the current methods for predicting construction cost budgets in the early stages of public construction projects are not sophisticated enough in terms of accuracy and reliability, indicating a need for improvement. The objective of this study is to develop a construction cost budget prediction model that can be utilized in the early stages of public building projects using an artificial neural network (ANN). In this study, an artificial neural network model was developed using the SPSS Statistics program and the data provided by the Public Procurement Service. The level of construction cost budget prediction was analyzed, and the accuracy of the model was validated through additional testing. The validation results demonstrated that the developed artificial neural network model exhibited an error range for estimates that can be utilized in the early stages of projects, indicating the potential to predict construction cost budgets more accurately by incorporating various project conditions.

Keywords : Construction Cost Prediction Model, Construction Budget Estimating, Public Building Projects, Artificial Neural Network

1. 서론

1.1 연구의 배경 및 목적

건설사업은 사업이 진행됨에 따라 다양한 리스크가 발생하고 수많은 변경이 일어나기 때문에 적정한 공사비를 예측하고, 이를 관리하는 것은 매우 중요하다. 특히, 사업초기단계부터 산정된 공사비 예산은 발주자의 올바른 의사결정을 지원하는 기준이 되며 공사비 부족으로 발생하는 다양한 리스크에 선제적으로 대응하기 위해서 적정한 비용으로 산정되어야 한다(Jo et al., 2020).

이는 공공 건설사업의 경우도 마찬가지이다. 정책적 목적의 달성과 사회적 수요의 충족 등을 목표로 수행되는 공공건설사업은 국비로 수행되며 비용의 준수 및 절감이 성과의 기준이 되기 때문에 비용의 관점에서 민감할 수밖에 없다. 또한 공공 건설사업에서 적정한 공사비 예산이 책정되지 않는다면 사업의 목적달성 뿐만 아니라 사업의 존폐에도 영향을 미칠 수 있다. 따라서 사업초기단계에서 보다 정확한 공사비를 책정하고, 이를 준공까지 적절히 관리하는 것은 매우 중요하다.

공공 건설사업의 총사업비는 공사비, 보상비, 부대경비로 이루어지며 그 중 공사비는 기획단계의 기본구상단계에서 추정되어 타당성 조사, 설계, 입찰 및 계약, 시공단계까지 사업이 진행됨에 따라 계속해서 관리되고 있다. 이와 관련해 정부는 매년 공사비 예산 산정을 위한 기준과 공사비 정보를 공표하고 있다(Jung et al., 2021).

공사비 예산 견적과 관련된 기준 및 지침 등에 따르면, 공사비 예산은 과거 수행했던 사례 데이터를 활용해 공사 유형별 단위면적당 공사비를 도출하고 도출된 단위면적당 공

* **Corresponding author:** Kim, Han Soo, Department of Architecture, Sejong University, Korea
E-mail: hskim@sejong.ac.kr
Received May 29, 2023; **revised** July 21, 2023
accepted August 30, 2023

사비를 신축공사의 연면적에 곱하여 산정하고 있다.

그러나 단위면적당 공사비를 연면적에 곱하여 산출하는 현행 방식은 시설물의 다양한 조건이나 특성을 반영하거나 예측의 신뢰성 확보에서 한계점이 있는 것으로 나타나고 있다(Ji et al., 2008; Kim & Park, 2022). 이러한 문제점으로 인해 건설산업계에서는 공사비 예산의 예측 방식의 개선을 요구하고 있으며, 기술이 발전함에 따라 더욱 높은 정확도의 비용예측 기술 활용이 가능해지고 있다(Heo, 2014).

비용예측에 관련한 연구들은 과거부터 활발하게 진행되어왔지만, 최근 인공신경망이 비용예측에 있어 정확도 높은 결과를 보여주고 있는 것으로 나타나고 있다(Van et al., 2019; Chandgude and Sarode, 2020). 공사비 예산 예측을 위해 인공신경망을 활용하면, 기존의 방식보다 조기에 더욱 빠르고 높은 정확도로 공사비를 예측할 수 있는 장점이 있는 것으로 나타나고 있다(Lee, 2018).

본 연구의 목적은 인공신경망을 활용하여 공공청사 프로젝트 사업초기단계에서 활용 가능한 공사비 예산 예측모델을 개발하는데 있다. 이를 통해 공공청사 프로젝트 공사비 예산 산정시 보다 다양한 프로젝트의 조건(변수)을 반영하여 보다 정교하게 공사비 예산을 예측할 수 있으며 이는 공사비 예산 예측의 정확성과 효율성을 높이는데 기여할 수 있다.

1.2 연구의 범위 및 방법

본 연구에서는 다양한 공공 건설공사의 유형 중 공공청사를 중심으로 인공신경망 기반의 공사비 예산 예측모델 개발을 연구 범위로 하고 있다. 예측모델 개발을 위해 활용된 데이터는 조달청의 공사비정보광장(http://pcae.g2b.go.kr)에서 제공하는 121개 일반청사 신축 프로젝트 공사비 예산데이터를 활용하였으며, 연구의 흐름을 정리하면 다음과 같다.

첫째, 문헌조사를 통해 공공 건설공사 공사비 예산 산정방식, 인공신경망에 대한 개념을 파악하고 관련 연구를 분석하였으며 연구 방법론인 인공신경망에 대해 고찰하였다.

둘째, 모델 구축을 위한 데이터를 확보하기 위해 조달청의 공공청사 프로젝트 공사비 예산데이터를 수집 및 분석하였으며, 연구범위에 적합한 데이터인 일반청사 신축 프로젝트의 공사비 예산 데이터를 선별, 이를 바탕으로 사업초기단계 공사비 예산 예측에 영향을 미치는 요인들을 변수로 선정해 정의하였다.

셋째, SPSS Statistics 프로그램을 활용하여 오차가 가장 낮게 나타나는 인공신경망 모델 구축을 위한 환경을 비교분석하였으며, 최종적으로 오차제곱합이 가장 낮게 나타나는 인공신경망 모델을 설계하고 최적 모델로 선정, 예측결과를 분석하였다.

마지막으로, 인공신경망 모델의 구축 시 사용하지 않은 별도의 10개의 추가 검증데이터를 대입하여 공사비 예측값을 추출하였으며, 이를 바탕으로 예측결과의 오차율을 분석하고 인공신경망 모델의 정확성을 검증하였다.

2. 이론적 고찰

2.1 공공 건설공사비 예산 산정 방식

건설 프로젝트에서 적정한 공사비 예산을 예측하는 것은 발주자의 원활한 의사결정을 지원하고, 시공단계에서 발생할 수 있는 공기 지연, 품질 미달 등의 리스크에 대해 선제적으로 대응할 수 있게 하는 중요한 요소이다(Lee et al., 2010). 이는 공공 건설사업에서도 마찬가지이다. 현재 공공 건설사업에서 총사업비는 공사를 수행하기 위한 공사비, 직접보상비와 간접보상비를 포함하는 보상비, 설계, 감리 및 시설 부대비 등을 포함하는 부대경비로 구성된다. 그 중 공사비는 사업초기단계인 기본구상단계에서부터 예측되어 관리된다.

일반공사의 경우, 기본구상단계에서 추정된 공사비 예산은 타당성 조사 단계를 거쳐 사업이 진행됨에 따라 계속해서 관리되며 대형공사인 경우 기본구상단계 이후 예비타당성조사를 거치며 공사비가 재협의되고, 이후 타당성조사 단계를 거쳐 관리된다(Korea Institute of Procurement, 2021). 이와 관련하여 정부는 총사업비관리지침, 국가재정법, 지방재정법, 건설기술진흥법 등 관련 기준, 지침 및 법령을 매년 공표하고 있으며 조달청, 공공데이터포털 등을 통해 공사비 데이터 또한 계속해서 업데이트하고 있다.

그러나 현재 공공 건설사업 기본구상단계에서 활용되는 방식은 정확도와 신뢰도 측면에서 한계가 있는 것으로 나타나고 있다. 특히, 사업초기단계 공사비 예산 예측 방식에 대한 문제가 다수 지적되고 있다(Lee et al., 2021; Kim & Son, 2006; Park et al., 2010).

공공 건설 신축사업의 사업초기단계 공사비 예산 예측의 경우 단위면적당 단가를 연면적에 곱하여 예측하는 것으로 나타나고 있다. 일반청사의 경우 조달청의 수행데이터(실적데이터)를 바탕으로 기획재정부가 공사 유형별 단위면적당 단가를 계산하고, 공공청사 업무를 담당하는 행정안전부에서 신축사업의 연면적을 계산해 기획재정부가 도출한 단위면적당 단가와 행정안전부가 도출한 연면적을 곱하여 기본구상단계 공사비 예산을 예측하고 있는 것으로 파악되었다(총사업비 관리지침 제90조; 총사업비 관리지침 제91조; 예비타당성조사 수행 총괄지침).

그러나 과거 실적데이터를 통해 도출한 단위면적당 단가에 연면적을 곱하여 공사비 예산을 예측하는 현행 방식은 해당 프로젝트의 다양한 특징(변수)을 고려하여 예측하는

것이 아닌, 유사 프로젝트의 단위면적당 단가만을 해당 프로젝트에 단순 대입하여 산출하는 것이기 때문에 정확성과 신뢰도 측면에서 한계가 있는 것으로 나타나고 있다. 또한, 시설물의 구성 및 조합이 다양하게 나타날 수 있는 학교, 청사 등의 공공 건축물에서는 적용성이 부족하며, 신뢰도 측면에서 한계가 있는 것으로 나타나고 있다(Jung et al., 2021).

즉, 프로젝트의 다양한 조건을 반영하고 있지 못하며, 예측을 위한 기술적 한계가 존재하고 있는 것이다. 이와 같은 한계점으로 인해 사업초기단계에서 제한적인 정보를 활용하여 적정한 공사비를 책정하기 위한 보다 정교한 공사비 예산 예측 기술의 적용 필요성이 높아지고 있다.

2.2 인공신경망

인공지능 기술은 인간의 지능을 모방하여 작업을 수행하고 수집한 정보를 기반으로 반복적으로 개선할 수 있는 기술이다. 즉, 인간이 가진 지적 능력을 컴퓨터를 통해 구현하는 기술을 말하며, 대표적인 인공지능 기술로는 머신러닝(machine learning)이 있다(Hashemi et al., 2020).

머신러닝 기술은 컴퓨터가 데이터로부터 스스로 학습하여 인공지능의 성능을 향상시키는 기술로 신경망, 결정 트리, 선형 회귀 등이 속한다. 그중에서도 인공신경망은 다양한 분야의 예측 관련 연구에서 활발하게 활용되고 있는 것으로 나타나고 있으며 인공신경망의 기본적인 구조는 〈Fig. 1〉에 제시된 바와 같다.

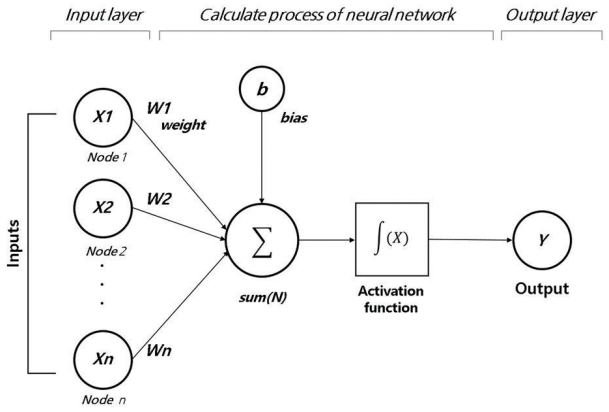

Fig. 1. Structure of artificial neural network

인공신경망은 인간의 신경망 구조와 유사한 구조를 갖는 학습 알고리즘으로, 기능에 따라 입력층(input layer), 은닉층(hidden layer), 출력층(output layer)으로 구성되며 각각의 층은 노드(node)로 구성되어 서로 기능적으로 연결되어 있다. 이때 노드란 인간의 신경망을 구성하는 신경세포와 같은 개념으로, 각 층을 구성하는 요소이다. 노드는 각 층에서 전달되는 신호를 주고받아 정보를 처리하는 기본적인 구성단위라고 할 수 있다.

입력층은 예측값(종속변수)을 도출하기 위한 입력변수(독립변수)의 값을 입력하는 층이며, 입력값이 n개라면 n개의 노드를 가지게 된다. 입력층에 입력된 입력값은 입력 노드를 지나 각각의 입력값에 대한 가중치와 곱해지고, 이 값들이 합산된다. 이때, 활성화 함수 임계값의 변동을 흡수하기 위한 가중치값인 편향(bias)이 함께 합산되는 연산과정을 거쳐 활성화함수에 적용 후 출력층에 전달되며 신경망 뉴런의 연산과정을 정리한 수식은 다음과 같다(Han et al., 2011).

$$N = \sum X_i W_i + b = (X_1 W_1 + X_2 W_2 + ... + X_n W_n) + b$$

이러한 연산과정을 거쳐 출력값이 전달되는 층이 출력층이며, 입력층과 출력층 사이의 층을 은닉층이라고 하고 각각의 층은 가중치를 갖고 연결된다.

인공신경망은 수학적 모형이 필요하지 않고, 예측에 유용한 기술이며 크기가 큰 자료를 사용한다는 특징을 지니고 있다. 또한, 인공신경망은 불완전하거나 사전에 알 수 없었던 입력값을 표현하는 경우, 합리적인 입력을 생성하는 일반화(generalization)의 장점이 있으며, 입력값이 변화할 경우 새로운 환경을 만들어 즉각적으로 프로그램을 갱신하고 학습 및 유지하는 적용성(adaptability)을 지닌다(Lee, 2018). 이러한 장점을 통해 인공신경망은 학교, 청사와 같이 규모 및 유형이 다양한 사업에서도 적용성이 높을 것으로 기대되고 있다.

인공신경망에서 입력층과 출력층 사이에 존재하는 중간층인 은닉층을 여러 개 갖는 더욱 심화된 구조의 신경망을 심층신경망(deep neural network)이라고 하며, 이 심층신경망을 학습하기 위해 고안된 기술 및 알고리즘을 딥러닝(deep learning)이라고 한다(Hong, 2021).

본 연구에서는 간단한 구조로 정확도 높은 결과를 도출하는 딥러닝 알고리즘인 다층 퍼셉트론(multi layer perceptron)을 활용하여 공공청사 공사비 예산 예측을 위해 활용하였으며 다층 퍼셉트론의 구조는 〈Fig. 2〉에 제시된 바와 같다.

다층 퍼셉트론은 하나 이상의 은닉층이 추가된 구조로 분석능력이 향상되었으며, 현재 다양한 분야의 예측을 위해 활용되고 있는 것으로 나타나고 있다(Yoon, 2019).

본 연구에서는 인공신경망 기반의 공공청사 공사비 예산 예측모델 개발을 위해 다층 퍼셉트론을 활용하였으며, 다층 퍼셉트론의 환경의 변화에 따른 결과값을 비교하여 가장 높은 정확도의 환경을 구축, 가장 오차가 낮게 나타나는 인공신경망 모델을 구축하고자 하였다.

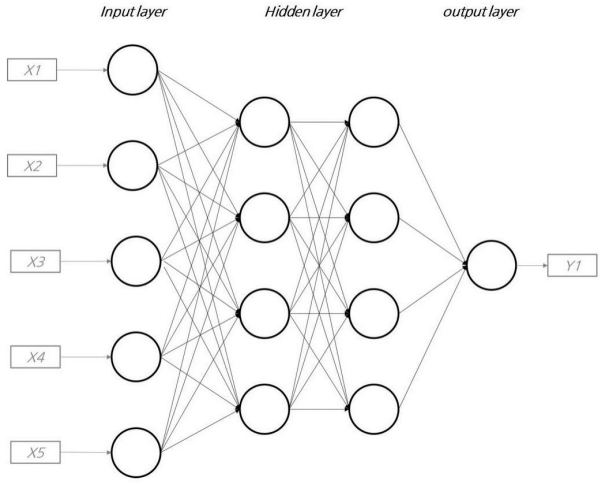

Fig. 2. Structure of multi layer perceptron

2.3 인공신경망을 활용한 비용예측

문헌 조사 결과 인공신경망은 다양한 분야에서 다양한 예측을 위해 활용되고 있는 것으로 나타났다. 건설산업에서도 인공신경망은 다양한 예측을 위해 활용되고 있으며, 특히 비용예측의 관점에서 활발한 연구가 진행되고 있는 것으로 나타났다.

Kim and Son (2006)은 교육시설 신축프로젝트 개념단계에서의 공사비 예측을 위해 오차역전파 학습 알고리즘을 활용한 인공신경망 모델을 구축하고 오차율을 분석, 교육시설 공사비 예측을 위한 인공신경망 모델의 활용가능성을 검증하였다. Lee et al. (2021)은 오차역전파 학습 알고리즘을 활용한 인공신경망을 활용하여 실적데이터를 기반으로 발전시설의 사업초기단계 견적모델을 개발하였으며, 인공신경망과 다양한 기법들을 조합해 기법들의 조합에 따른 오차를 검증하였다. Kulkarni et al. (2017)은 건설사업관리 분야에서 인공신경망의 활용 방안에 대해 분석하고, 인공신경망이 비용예측에 다수 활용되고 있음을 제시하였다. Han et al. (2011)은 오차역전파 기반의 인공신경망을 활용하여 기획단계의 도로건설 공사비 예측모델을 개발하고, 회귀분석과 비교하여 인공신경망 모델의 정확성이 더 높음을 검증하였다. Baccarini (2005)는 건설공사 예비비 예측을 위한 오차역전파 기반의 인공신경망 모델을 구축하고, 이를 회귀분석과 비교하여 정확성을 비교 및 검증, 인공신경망 모델의 정확성이 더 높음을 검증하였다.

위와 같이 건설산업에서는 비용예측의 관점에서 인공신경망의 활용이 다수 관찰되었다. 그러나 건설사업의 관점에서, 특히 공공 건설사업의 공공청사 관점에서 인공신경망을 활용한 공사비 예산 예측에 관한 연구는 미흡한 것으로 나타나고 있다. 또한, 신경망 환경 구축에 있어 학습알고리즘을 오차 역전파 알고리즘으로 한정하여 설계했다는 측면에서 한계가 나타나고 있다.

따라서 국내 공공 건설사업 중 공공청사의 사업초기단계 공사비 예산 예측을 위한 모델 개발을 목표로 하는 본 연구는 주제와 연구범위 측면에서 기존 연구와는 차별성을 지니고 있다. 또한, 기존 선행연구의 한계점을 보완하기 위해 모델 구축시 오차역전파, 척도화된 켤레기울기, 경사하강법 등의 다양한 학습 알고리즘을 실험하고, 특정한 기준을 바탕으로 한 모델 검증을 통해 기존 연구와 연구 방법론 측면에서 차별성을 지니고 있다고 할 수 있다.

3. 데이터 수집

3.1 데이터 수집 개요

데이터 수집이란 조직 내외부에 존재하는 다수의 데이터 소스로부터 필요로 하는 데이터를 수집하는 것을 말하며, 수집한 데이터를 분석하기 위해 데이터를 변환 또는 통합하는 것도 넓은 의미의 데이터 수집이라고 할 수 있다.

본 연구에서는 인공신경망 기반의 공공청사 공사비 예산 예측모델 구축을 위한 관련 데이터 수집을 위해 조달청에서 제공하는 공사비 정보광장 포털을 활용하였다. 데이터는 공사비 정보광장 포털에 게시된 데이터 중 가장 데이터의 수가 많았던 공공청사 데이터를 활용하였다. 공공청사 데이터 전체 201개 중, 사업 규모에 따라 일반청사와 대형청사를 구분하여 139개의 일반청사 데이터를 활용하였고, 그중에서도 발주년도 및 공사유형을 구분하여 2015년도부터 2021년도까지 수행된 일반청사 신축프로젝트의 121개의 공사비 예산데이터를 활용 및 분석하였다.

3.2 데이터 선별 및 정의

본 연구에서 데이터 선별 및 정의란 연구에 활용 가능한, 혹은 필요한 데이터를 식별하여 수집하고, 데이터가 지니는 의미와 단위 등 정보를 정의하는 것을 의미한다.

조달청의 공사비 정보광장 포털에는 기수행된 공공 건설프로젝트의 전체적인 예산정보가 기재되어 있었다. 데이터는 크게 공사개요, 대표공종 공사비 현황, 공종별 공사비, 세부공종별 공사비, 순공사원가 및 직접공사비 기준의 공사비 분석, 표준시장단가 시장시공가격 구성비, 공종별 조사금액 및 계약금액 현황으로 구성되어 있었다.

공사비 정보광장에서 제공하는 데이터들은 전체적인 예산정보를 갖는 데이터로, 연구 범위에 맞추어 사업초기단계 공사비 예산 예측에 영향을 미치는 데이터들을 선별할 필요가 있었다. 이를 기준으로 하여 선별된 데이터의 변수 및 정의는 다음과 같다.

- 용도(use): 건축물의 건축법상의 용도로 업무시설(business facilities), 문화집회시설(cultural and assembly facilities), 근린생활시설(residential neighborhood facilities), 공장(factory), 교육연구시설(educational research facilities), 방송통신시설(Broadcasting and communication facilities), 교정시설(correctional institution) 등이 용도로서 구분되었으며, 두 가지 이상의 용도가 기재되어 있는 데이터는 두 가지 용도를 모두 기입하여 활용하였다.
- 지역(district): 건축물이 위치한 지역의 데이터로 서울(Seoul), 경기(Gyeonggi), 인천(Incheon), 강원(Gangwon), 충남(Chungnam), 충북(Chungbuk), 경남(Gyeongnam), 경북(Gyeongbuk), 전남(Jeonnam), 전북(Jeonbuk), 부산(Busan), 울산(Ulsan), 세종(Sejong), 대전(Daejeon), 대구(Daegu), 제주(Jeju)로 구분되었다.
- 구조(structure): 건축물의 구조 데이터로, 철근콘크리트조(reinforced concrete building), 철골조(steel skeleton building), 철골철근콘크리트조(steel framed reinforced concrete building)로 구분되었다.
- 지하층/지상층(basement floor/floor): 지하층 및 지상층의 층수를 나타내는 데이터로, 공사의 규모를 나타내는 데이터이다. 데이터의 입력을 위해 지하층과 지상층을 구분하여 활용하였다. '층' 단위를 활용하였다.
- 대지면적(lot area): 대지의 수평투영면적을 의미하는 데이터로, '㎡' 단위를 사용하였다.
- 연면적(architectural area): 대피공간을 제외한 건축물의 바닥면적에 층수를 곱한 값의 데이터로, '㎡' 단위를 사용하였다.
- 건축면적(building area): 건축물의 수평투영면적을 나타내는 데이터로, '㎡' 단위를 사용하였다.
- 조경면적(landscaping area): 건축물의 부지 내 조경의 면적을 나타내는 데이터로, '㎡' 단위를 사용하였다.
- 건폐율(building coverage): 대지면적에 대한 건축면적의 비율로, '%' 단위를 사용하였다.
- 용적률(floor area ratio): 대지면적에 대한 연면적의 비율로, '%' 단위를 사용하였다.
- 층고(floor hight): 건축물의 층과 층 사이의 높이를 나타내는 데이터로, 건축법상의 건축물 층의 높이를 말한다. 'm' 단위를 사용하였다.
- 최고높이(hight of building): 해당 건축물의 가장 높은 부분의 높이를 나타내는 데이터로, 'm' 단위를 활용하였다.
- 주차대수(parking capacity): 건축물에서 수용 가능한 주차대수를 나타내는 데이터로, 건축물 규모에 따라 다양한 규모를 갖고 있는 것으로 나타났다. '대(cars)' 단위를 사용하였다
- 공사기간(construction period): 착공일부터 준공일까지 실제 공사가 소요된 공사 기간이 아닌, 도급계약 당시 계약된 공사 기간 데이터로, '일(days)' 단위를 사용하였다.
- 총공사비(total construction cost): 최종 준공정산금이 아닌, 입찰시 입찰평가를 위해 활용된 조달청의 기초금액을 기준으로 하여 관급자재비를 포함한 금액으로 '원(Won)' 단위를 사용하였다. 본 연구의 예측값(종속변수) 데이터로 활용하였다.

본 연구에서는 조달청의 공사비 정보광장에 기재된 일반청사의 예산데이터 중 사업초기단계부터 추정이 가능한 데이터를 구분하여 상기 총 16개 유형의 변수를 선별, 인공신경망 예측모델 구축을 위해 활용하였다.

상기에 제시된 변수 중 총공사비는 예측의 결과인 종속변수에 해당되며 나머지 변수 15개는 독립변수에 해당된다. 이들 독립변수의 설정은 두 단계로 진행되었다. 첫째는 유사연구 분석을 통해 공사비 예산 예측을 위해 중요한 독립변수를 1차적으로 도출한 후, 이를 조달청 공사비정보광장에서 제공하는 정보(독립변수)와 상호 비교하여 공통적인 독립변수를 선정하였다. 결국 조달청 공사비정보광장에서 제공하는 정보(독립변수)가 선택에 있어서 가장 중요하였으며 유사 연구에서 도출된 독립변수가 대조 및 참고되었다.

본 연구에서 사용된 사례 데이터의 수에 비해서 독립변수가 다소 많은 것도 연구 시 검토의 대상이었다. 그럼에도 불구하고 세 가지 이유로 인해 상대적으로 다소 많은 15개 독립변수를 예측모델 개발에 사용하였다. 첫째, 원(原)데이터(조달청 공사비정보광장)에서 제시한 데이터 속성을 임의로 조정하기보다는 있는 그대로 활용하고자 하였고, 둘째, 공사비 예측을 위해 모두 중요한 변수들이라 임의적으로 일부 변수를 제외하는 판단이 어려웠으며, 마지막으로 인공신경망(머신러닝) 기법의 특성상 가중치 분석 등을 통해 무의미한 독립변수들은 '자동으로' 제외하여 예측모델을 구축하기 때문이었다.

4. 인공신경망 모델 구축

인공신경망 모델 구축이란 연구목적 달성을 위한 인공신경망 환경 및 조건을 구축하는 것이다. 본 연구에서는 독립변수(independent variables) 및 종속변수(dependent variables) 설정, 예측모델 환경을 분석하는 인공신경망 환경설정에 대한 분석, 최적의 환경 조건을 도출하여 모델을 설계하는 최소오차 모델 설계, 모델의 예측결과를 분석하여 최적 모델을 도출하는 모델 분석 및 최적모델 선정의 단계

로 수행되었으며, 자세한 내용은 각 절에 기술하였다.

4.1 독립변수 및 종속변수 설정

변수 설정이란 예측모델을 구축하기 위해 독립변수(입력변수)와 종속변수(예측변수)를 설정하는 과정을 의미하며, 이때 독립변수란 종속변수의 값을 결정하는 입력값, 원인 등 조건을, 종속변수란 독립변수에 의해 결정되는 예측을 통해 알고자 하는 결과값(출력값)을 의미한다(Kim et al., 2017).

SPSS Statistics 프로그램은 변수를 숫자, 문자, 날짜, 시간의 유형으로 입력할 수 있으며, 변수의 특성에 따라 변수의 고유한 값이나 범주 수가 제한된 데이터는 '요인'변수로, 그렇지 않은 변수는 '공변량' 변수로 입력하여 분석한다.

본 연구에서는 독립변수 중 변수의 고유한 값이나 범주가 제한적인 변수인 용도, 지역, 구조, 지하층, 지상층의 5개 변수는 '요인(factors)'변수로, 변수의 값이나 범주가 제한적이지 않은 대지면적, 연면적, 건축면적, 조경면적, 건폐율, 용적률, 층고, 최고높이, 주차대수, 공사기간의 10개 변수는 '공변량(covariances)'변수로 입력하여 분석하였고 총 15개 유형의 독립변수와 1개 유형의 종속변수를 설정하였으며, 이를 정리하면 〈Table 1〉에 제시된 바와 같다.

본 연구에서는 총 16가지의 변수 중 용도, 지역, 구조, 지하층, 지상층, 대지면적, 연면적, 건축면적, 조경면적, 건폐율, 용적률, 층고, 최고높이, 주차대수, 공사기간의 15가지 변수를 독립변수로 설정하였다. 예측하고자 하는 종속변수는 총 공사비로 설정하여 분석을 진행하였다.

4.2 인공신경망 환경설정에 대한 분석

인공신경망 환경설정이란 보다 정확한 예측을 위해 인공신경망의 성능에 영향을 미치는 환경을 조율 및 조정하는 것으로, 환경의 조율을 통해 가장 최적의 성능을 발휘하는 인공신경망 모델을 설계하는 것이 중요하다.

인공신경망의 성능에 영향을 미치는 환경으로는 은닉층의 수, 은닉층의 노드 수, 학습 최적화, 활성화 함수 등이 있으며, 각 환경별 특징은 다음과 같다.

다층 퍼셉트론에서의 은닉층은 1개 이상의 층으로 구성되며, 은닉층의 최적화에는 규칙이 명확하지 않아 시행착오적 방법(trial error method)을 활용하는 것이 일반적이다. 문헌조사 결과, 인공신경망의 은닉층 수가 증가할수록 훈련 데이터의 예측 정확도는 일반적으로 증가하지만 은닉층의 개수가 증가함에 따라 검증데이터에 대한 손실이 증가해 검증데이터의 정확도는 오히려 떨어지게 된다(Oh & Lim, 2019). 즉, 은닉층의 수가 많다고 항상 종속변수 예측값의 정확도가 높아지지는 않으며, 예측에 적합한 최적의 은닉층 수를 설계해야 한다.

Table 1. Variables of the model

		Variables	Sub-variables/Unit
Independent variables	factors	use	business facilities
			cultural and assembly facilities
			residential neighborhood facilities
			factory
			Broadcasting and communication facilities
			educational research facilities
			correctional institution
		district	Seoul, Gyeonggi, Incheon, Gangwon, Chungnam, Chungbuk, Gyeongnam, Gyeongbuk, Jeonnam, Jeonbuk, Busan, Ulsan, Sejong, Daejeon, Daegu, Jeju
		structure	reinforced concrete building
			steel skeleton building
			steel framed reinforced concrete building
		basement floor	0 - 2
		floor	2 - 8
	covariances	lot area	㎡
		architectural area	㎡
		building area	㎡
		landscaping area	㎡
		building coverage	%
		floor area ratio	%
		floor hight	m
		hight of building	m
		parking capacity	cars
		construction period	days
dependent variables		total construction cost	won

인공신경망 모델은 하나의 입력층과 출력층, 그리고 몇 개의 은닉층으로 구성된다. 각각의 층은 서로 수평적으로 상호 연결되지만, 하나의 층 안에서 수직적으로는 연결되지 않는다. 각 층에서 최적의 노드 수를 결정하는 규칙은 시행착오적인 방식을 활용해 노드 수를 결정하는 것이 일반적이다(Park et al., 2020).

인공신경망의 예측에서 학습은 가중치 및 편향의 크기를 결정하며 예측의 정확도에 기여하기 때문에 매우 중요하다. 인공신경망에 활용되는 학습 알고리즘으로는 경사하강법, 척도화된 켤레기울기, 오차 역전파 등 매우 다양하며, 학습 알고리즘을 잘못 설계한다면 가중치의 변화가 감소하거나, 예측값의 변화가 크게 일어나 예측값의 오차를 증가시킬 수 있다(Gunaydin & Dogan, 2004). 선행 연구 고찰 결과 나타났던 인공신경망 모델 구축 시 학습 알고리즘을 오차역전파 알고리즘으로 한정지어 모델을 설계했던 한계를 극복하

기 위해 본 연구에서는 다양한 학습 알고리즘에 따른 차이를 분석, 더욱 오차가 낮은 학습 알고리즘을 도출 및 활용하고자 하였다.

인공신경망에서 활성화 함수란 입력층의 입력값으로부터 다음 층으로 전달하는 출력신호를 생성하는 함수이다. 활성화 함수의 선택에 따라 인공신경망 모델의 예측결과가 달라지기 때문에, 모델 설계에서 활성화 함수의 선택이 매우 중요하다(Kim et al., 2021).

본 연구에서는 SPSS Statistics 프로그램을 활용하여 인공신경망 모델을 구축하였다. 인공신경망 모델 구축시 시행착오적 방법을 활용하여 모델의 최적화에 영향을 미치는 은닉층의 수, 은닉층의 노드 수, 학습 환경, 활성화 함수의 변화에 따른 결과를 비교 분석하여 활용하였다.

4.3 최소오차 모델 설계

최소오차 모델 설계란 인공신경망의 성능에 영향을 미치는 환경들을 조합하여 가장 오차가 낮게 나타나는 환경 조건을 도출, 이를 바탕으로 모델을 설계하는 것을 의미한다.

인공신경망의 최적 환경을 도출하는 방법은 시행착오적 방법, 격자 탐색법, 랜덤 탐색법 등의 기법이 있으며 본 연구에서는 시행착오적 방법을 활용하여 인공신경망의 최적 환경을 도출하였다. 본 연구에서 시행착오적 방법을 활용한 이유는 설정 환경의 수가 크지 않은 경우 설정 환경의 시험적인 조합을 통해 목표 학습 성능을 위한 조합을 간편하게 찾을 수 있다는 장점이 있기 때문이다(Lee & Lee, 2021).

본 연구에서 인공신경망 최적 모델 추출을 위한 기준점으로 활용한 것은 다층 퍼셉트론 예측의 결과로서 나타나는 오차제곱합(sum of squares for error)의 값이다. 오차제곱합은 인공신경망의 최적의 가중치값을 탐색하여 모델의 성능을 판단하는 지표인 손실함수(loss function)에서 가장 대표적으로 활용되는 방법으로, 예측값과 실사례 값의 오차를 비교 계산한다. 설계한 인공신경망 모델의 오차제곱합의 값이 작을수록 오차가 적고 정확도가 높다는 것을 의미한다.

오차제곱합이 가장 낮은 모델을 추출하기 위해서는, 인공신경망의 성능에 영향을 미치는 요인들의 변화에 따른 결과값을 비교할 필요가 있었다.

SPSS Statistics 프로그램은 인공신경망 모델 설계 시 프로그램이 자동으로 신경망을 설계하는 자동 신경망 설계 기능과 사용자가 직접 인공신경망의 환경을 설정하여 모델을 설계하는 사용자 정의 신경망 설계 기능이 있다. 사용자 정의 신경망 설계 기능에서는 인공신경망 모델의 성능에 영향을 미치는 은닉층의 수, 은닉층의 노드 수, 활성화 함수 및 모델의 학습 환경인 학습 최적화 알고리즘을 사용자가 직접 설정할 수 있었다.

Table 2. Summary of model information

Layer	environment		information
	variables	number	
Input layer	factors	1	use
		2	district
		3	structure
		4	basement floor
		5	floor
	covariance	1	lot area
		2	architectural area
		3	building area
		4	landscaping area
		5	building coverage
		6	floor area ratio
		7	floor hight
		8	hight of building
		9	parking capacity
		10	construction period
	number of nodes		47
	training optimizer		scaled conjugate gradient
Hidden layer	number of hidden layer		1
	number of nodes in hidden layer		4
	activation function		hyperbolic tangent
Output layer	dependent variable		total construction cost
	error function		sum of squares for error

은닉층의 수는 한 개 또는 두 개, 노드는 자동으로 계산하거나 사용자가 노드의 수를 지정하여 설계할 수 있었다. 활성화 함수는 쌍곡 탄젠트 함수와 시그모이드 함수를 구분하여 설계가 가능했으며, 학습 최적화 알고리즘은 척도화된 켤레기울기, 경사 하강 알고리즘 중에서 선택하여 학습률, 모멘텀, 기울기 람다, 시그마 값 등의 조정 등의 세부 내용을 설계할 수 있었다.

본 연구에서는 사용자 정의 신경망 설계 기능을 활용해 시행착오적 방법으로 신경망 모델의 최적화를 진행함과 동시에 자동신경망 설계 기능을 활용하여 오차제곱합이 낮게 나온 결과의 환경을 분석하고 이를 상호 비교, 비슷한 조건에서 더욱 낮은 오차제곱합을 나타내는 환경을 분석하였다.

인공신경망 모델의 환경별 오차제곱합을 분석한 결과, 다른 환경이 동일한 경우 일반적으로 은닉층의 수는 두 개일 경우보다 한 개일 경우 정확도가 높은 것으로 나타났으며, 은닉층의 노드 수는 2개부터 30개까지 변화를 주어 분석하였을 때 4개와 20개의 노드일 경우 가장 정확도가 높은 것으로 나타났지만, 20개의 경우 학습의 오차제곱합 값은 낮았지만 검증의 오차제곱합 값이 더욱 높게 나타났다.

또한, 활성화함수는 시그모이드 함수보다는 쌍곡 탄젠트 함수일 경우, 학습 최적화 알고리즘은 척도화된 켤레기울기

(scaled conjugate gradient) 알고리즘을 활용하여 설계하였을 경우 오차제곱합 값이 더욱 낮게 나타났으며 설계된 모델의 환경 정보를 요약하면 〈Table 2〉에 제시된 바와 같다.

은닉층의 수, 은닉층의 노드 수, 학습 최적화 알고리즘, 활성화 함수를 조합하여 인공신경망 모델이 최적의 성능을 발휘하는 환경을 분석한 결과, 인공신경망 은닉층이 1개, 은닉층의 노드 수가 4개, 활성화 함수는 쌍곡 탄젠트 함수로 설정하고, 척도화된 켤레기울기 학습 알고리즘으로 설계할 경우 학습 오차제곱합 0.792, 검증 오차제곱합 1.809으로 가장 낮은 오차제곱합 값을 나타내는 결과를 도출할 수 있었다.

4.4 예측 결과 분석

본 연구에서 의미하는 예측결과 분석이란 인공신경망 모델의 예측에서 나타나는 결과의 특징을 분석한 것으로, 예측결과 분석은 케이스 분석, 오차제곱합 결과 분석, 실사례값과 예측값 비교, 잔차 분석, 독립변수 중요도 분석의 과정으로 진행하였다.

본 연구에서는 인공신경망 기반의 공공청사 공사비 예산 예측모델 구축을 위해 조달청에서 제공하는 공사비정보광장 포털에 게시된 공공청사 데이터 중, 신축 일반청사 프로젝트의 121개 예산데이터를 활용하였다. 그 중에서 10개의 데이터를 모델의 예측성능을 보여주기 위한 별도의 검증데이터로 활용하였고, 나머지 111개의 데이터 중 분석에 부적합하여 케이스에서 제외된 2개의 데이터를 뺀 109개의 데이터를 최적 인공신경망 모델을 추출하기 위한 데이터베이스로 활용하였다. 도출한 최적의 인공신경망 환경에서 109개 데이터를 분석한 결과는 다음과 같다.

인공신경망 분석을 위해서는 학습케이스와 검증케이스를 설정해야 한다. 이때 학습 케이스란 인공신경망 모델을 훈련하기 위한 변수들의 데이터 세트이며, 검증 케이스란 과도한 훈련을 방지하기 위해 훈련 중 오차를 추적하는 데 활용되는 독립된 데이터 세트를 의미한다. 본 연구에서는 데이터의 학습과 검증은 학습 70%, 검증 30%로 설정하여 분석을 진행하였고, 프로그램 내에서 케이스를 무작위로 할당해 진행하였으며 결과는 〈Table 3〉에 제시된 바와 같다.

Table 3. Summary of case processing

Case information		Number	Percentage
Case	Training	86	78.9%
	Testing	23	21.1%
Valid case		109	100%
Excluded case		2	-
Total		111	-

인공신경망 분석을 위해 총 111개의 케이스 중 2개 케이스를 제외한 109개의 케이스를 활용하였고 그 중 86개(78.9%)의 케이스를 학습에, 23개(21.1%)의 케이스를 검증에 활용하였다.

본 연구에서는 인공신경망의 성능 판단을 위해 오차제곱합과 상대오차의 값이 가장 낮은 값으로 도출되는 모델을 추출하였다. 오차제곱합(sum of squares error)은 종속변수(예측변수)의 예측 수준에 따라 계산되며, 상대오차(relative error)는 모델 구축을 위한 종속변수의 입력값인 실사례 값에 대한 오차의 비율을 나타낸 값으로 실사례 값에 오차를 나누어 계산한 백분율의 평균값이다. 분석을 통해 도출된 오차제곱합과 상대오차 값이 낮을수록 모델의 정확도가 높음을 의미한다. 본 연구에서 설계한 인공신경망 모델의 학습은 오차의 감소가 더 이상 나타나지 않는 단계에서 중지하였으며 도출된 오차를 요약하면 〈Table 4〉에 제시된 바와 같다.

Table 4. Summary of sum of squared errors

Case	error function	result
Training	sum of squares error	0.792
	relative error	0.019
	stopping rule	1 consecutive step(s) with no decrease in error
Testing	sum of squares error	1.809
	relative error	0.092

오차제곱합 분석 결과, 학습 데이터의 오차제곱합은 0.792로 낮은 수준의 값을 보였고, 검증데이터의 오차제곱합은 1.809로 학습 데이터의 값보다는 높지만, 낮은 수준의 값을 보이는 것으로 나타났다. 상대오차 또한 학습에서 0.019, 검증에서 0.092로 낮은 수준의 값으로 나타났으며, 시행착오적 방법을 통해 도출한 가장 낮은 오차제곱합 값을 나타내는 본 모델을 최적모델로서 선정하였다.

구축된 인공신경망 모델의 예측 정확도 수준을 분석하기 위해 모델 구축을 위해 입력된 종속변수의 실사례값(입력값, observed value)과 모델의 예측을 통해 도출된 종속변수의 예측값(결과값, predicted value)을 비교 분석할 필요가 있었다. 본 연구에서 정의하는 실사례값이란 인공신경망 모델의 학습을 위해 모델 구축시 입력된 종속변수의 입력값을 의미하며, 예측값이란 구축된 인공신경망 모델의 예측을 통해 도출된 입력값에 대한 예측값을 의미한다.

인공신경망 분석을 통해 종속변수의 실사례값과 예측값을 비교한 결과는 〈Fig. 3〉에 제시된 바와 같다.

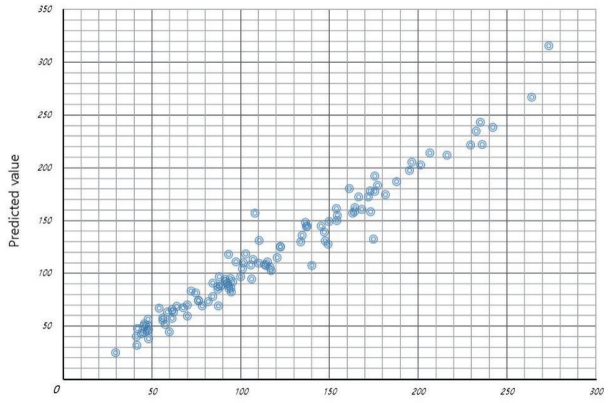

Fig. 3. Graphs comparing predicted and observed values

총공사비의 실사례값과 예측값을 비교한 결과, 실사례값과 예측값이 유사한 값에 위치하는 선형에 가까운 비교 그래프를 도출할 수 있었으며, 이는 본 인공신경망 모델의 공사비 예산에 대한 예측의 오차 수준이 낮다는 것을 의미한다.

그러나 실사례값과 예측값의 차이인 잔차(residual value)를 분석한 결과, 모든 데이터에 대해 일정한 값의 잔차를 갖는 것이 아닌 데이터의 특성에 따라 각기 다른 값의 잔차를 갖는 것으로 나타났으며 이를 그래프로 표현하면 〈Fig. 4〉에 제시된 바와 같다.

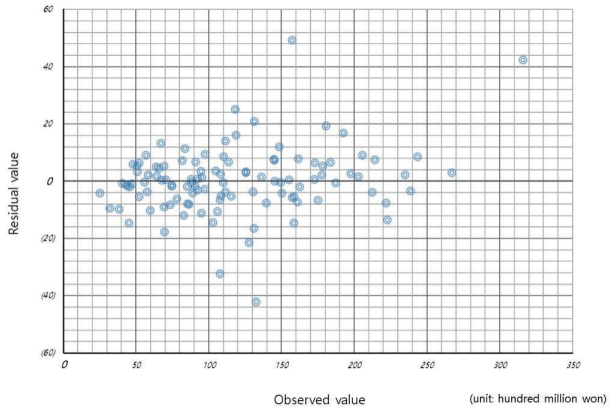

Fig. 4. Graphs comparing residual and observed values

〈Fig. 4〉에서 나타나는 바와 같이 대부분의 데이터가 잔차값이 0이 되는 부분에서 오차가 20억원 이하로 나타나는 범위 안에 속하는 것으로 나타났으며 각각의 데이터는 특정한 규칙을 갖지 않고 각기 다른 잔차 값을 보이는 것으로 나타났다. 각각의 데이터가 나타내는 잔차를 활용해 [|(잔차/실사례값)|×100]의 계산을 통해 도출하는 오차율로 계산한 결과, 가장 높은 오차율을 보이는 데이터는 약 32.45%, 가장 낮은 오차율을 보이는 데이터는 약 0.02%의 오차율을 나타내고 있었다. 실사례값과 예측값을 비교한 오차율의 평균은 약 7.5%의 낮은 수준의 오차율을 나타내고 있었다.

높은 오차를 나타내는 데이터의 특징을 분석하기 위해 높은 오차의 데이터를 추적하여 분석한 결과, 높은 오차의 데이터는 공사비가 특히 높거나, 연면적이 특히 높은 데이터 등 일반적인 데이터들과 차이를 나타내는 데이터 또는 오류 및 손실(missing)이 있는 데이터값이 존재하는 데이터임을 알 수 있었다. 이는 각각의 데이터의 질이 예측모델의 오차에 영향을 미치고 높은 질의 데이터는 낮은 오차를, 낮은 질의 데이터는 높은 오차의 예측을 수행한다는 것을 의미한다. 즉, 높은 정확도의 예측을 위해서는 좋은 질의 데이터를 선별 활용해야 한다는 것이다.

또한, 각각의 독립변수(입력변수)에 따라 인공신경망 모델의 예측에 미치는 중요도가 다른 것으로 나타났다. 중요도란, 독립변수의 중요성, 영향력을 수치화한 척도로 각 독립변수가 종속변수(예측변수)에 미치는 중요도를 분석한 결과를 요약하면 〈Table 5〉에 제시된 바와 같다.

Table 5. Importance of independent variables

Independent variables	Importance	Normalized importance
use	0.059	27.8%
district	0.056	26.4%
structure	0.034	16.3%
basement floor	0.038	18.1%
floor	0.029	13.8%
lot area	0.085	40.3%
architectural area	0.211	100%
building area	0.088	41.9%
landscaping area	0.053	25.2%
building coverage	0.083	39.2%
floor area ratio	0.052	24.8%
floor hight	0.004	2%
hight of building	0.059	28.0%
parking capacity	0.072	34.2%
construction period	0.076	36.1%

각 독립변수의 중요도 분석에서 나타나는 가장 큰 특징은 세 가지로 정리할 수 있다. 첫째, 각각의 독립변수들은 각기 다른 중요도 값을 지니며, 모델의 변화에 따라 중요도 값이 변화한다. 이는 각각의 독립변수들이 모델의 예측값에 기여하는 정도가 각기 다르며, 모델 환경의 변화에 따라 해당 모델의 예측값에 대한 독립변수들의 중요도가 변화함을 의미한다. 둘째, 가장 높은 중요도 값을 나타내는 독립변수는 연면적이었다. 이는 공공청사 공사비 예산 예측을 위해 가장 중요한 영향을 미치는 변수는 연면적이며, 연면적의 값이 인

공신경망 모델의 예측값에 가장 큰 영향을 미친다는 것을 의미한다.

셋째, 중요도가 가장 높게 나타나는 변수들은 연면적, 대지면적, 건축면적, 건폐율, 공사기간 등 프로젝트의 스코프(scope)에 직접적인 영향을 미치는 변수들인 것으로 나타났다. 즉, 공공청사 공사비 예산 예측을 위한 인공신경망의 설계에서 각각의 독립변수들 모두 예측값을 위해 중요하지만, 공사의 규모와 비용 등 프로젝트의 스코프에 영향을 미치는 변수들의 데이터가 예측값에 기여하는 정도가 크며, 이러한 데이터에 오류가 있을 경우 예측값의 정확도에 더욱 큰 악영향을 미칠 수 있음을 시사한다. 즉, 공사비 예산 예측에 대한 각각의 독립변수들이 갖는 중요도의 값은 각기 다르며, 본 인공신경망 모델에서는 독립변수 중 연면적이 가장 높은 중요도를 지니며 모든 독립변수 데이터의 질이 중요하지만, 높은 중요도의 독립변수 일수록 모델에 더욱 큰 영향을 미칠 수 있음을 시사한다.

5. 모델 추가 검증

5.1 모델 추가 검증 방식

인공신경망 모델 검증이란, 설계된 모델이 얼마나 정확도 높은 결과를 도출하는지 검증하는 것이다. 4장에서 도출된 모델의 오차 수준을 나타내는 오차제곱합 값, 실사례값 예측값 분석, 잔차 분석 등의 결과로도 모델이 낮은 수준의 오차를 나타내는 결과를 보였기 때문에 신뢰도가 확보되었다고 할 수 있다. 그럼에도 불구하고 모델의 신뢰성을 재확인하기 위해 추가 검증을 실시하였다. 그 이유는 두 가지로 정리할 수 있다. 첫째, 구축된 모델이 새로운 데이터를 추가하여 예측했을 경우에도 일관되게 낮은 오차를 보이는지 추가적으로 검증하여 모델의 활용 가능성을 분석하기 위함이었다. 둘째, 문헌 고찰을 통해 선행 연구에서 나타난 특정한 기준을 바탕으로 모델을 추가 검증하지 못했던 한계를 보완하기 위함이었다. 4.4절의 모델 검증과 본 절의 모델 검증의 차이점은 4.4절은 예측모델 개발 후 SPSS Statistics 프로그램에서 제시하는 각종 통계 수치를 해석하여 예측모델의 타당성을 검증하는 것이며, 5절은 예측모델 개발에 사용되지 않은 별도의 추가검증 데이터를 통해 검증을 실시한 것이다.

추가 검증은 예측 모델 구축시 제외된 별도의 데이터 10개를 추가 검증데이터로 활용하였다. 10개의 추가검증 데이터는 무작위(random) 방식으로 선별되었으며 무작위로 데이터를 선별한 이유는 작위성을 피하면서 예측모델의 정확도를 객관적으로 검증하기 위한 목적이었다.

추가 검증데이터를 구축된 인공신경망 모델에 대입하여 총공사비 실사례값과 인공신경망 모델 분석을 통해 도출된 예측값을 비교, 오차율을 계산하여 이를 세계적인 Cost Engineering 협회인 AACEI (Association for the Advancement of Cost Engineering International)에서 발표한 Construction cost estimate accuracy와 비교, 유효성을 검증하였다. 오차율을 비교하는 가장 이상적인 방법은 기존 방법의 결과와 비교하는 것이라고 할 수 있다. 그러나 현재 공공 건설공사의 예산 예측 이후 그 정확도를 평가하는 과정이나 그 결과를 조달청 및 발주청에서 기록으로(데이터로) 남겨두고 있지 않아서 상호 비교할 수 없는 어려운 한계가 있었다. 이러한 한계점으로 인해 대안적으로 AACEI 기준을 차용하게 되었다.

Construction cost estimate accuracy는 미국의 Cost Engineering 협회인 AACEI에서 발표한 건설사업 단계별 견적의 허용 오차범위를 나타내며 이를 정리하면 〈Table 6〉에 제시된 바와 같다.

Table 6. Project stage's estimate error tolerance of AACEI

Estimation class	Project definition level	Project stage	error tolerance rate
Class 5	0% ~ 2%	Programming and budgeting stage	-20% ~ -50% +30% ~ +100%
Class 4	1% ~ 15%	Feasibility study stage	-15% ~ -30% +20% ~ +50%
Class 3	10% ~ 40%	Design development stage	-10% ~ -20% +10% ~ +30%
Class 2	30% ~ 75%	Construction documents stage	-5% ~ -15% +5% ~ +20%
Class 1	65% ~ 100%	Bidding stage	-3% ~ -10% +3% ~ +15%

※ Association for the Advancement of Cost Engineering International

AACEI에서는 설계의 완성도에 따라 건설사업의 견적에서 허용되는 오차의 범위를 5단계로 구분하여 기준을 발표하였다. 국내 건설사업 견적의 기준으로 해석하였을 때, Class 5는 사업초기단계의 예산 편성 및 기획단계의 견적이며, Class 4는 타당성 분석 단계에서의 견적, Class 3, Class 2는 각각 기본설계, 실시설계단계에서의 견적, Class 1은 입찰계약단계에서의 견적으로 해석할 수 있다(Jo et al., 2020). 따라서 본 연구는 모델의 검증을 위해서 모델의 예측값이 기본구상단계와 매칭되는 Class 5의 오차범위인 −30~50% 범위 내에 속하는지 비교 분석하여 유효성을 판단하였고, 예측값이 어느 단계의 Class 수준에 속하는지 분석하였다.

5.2 모델 추가 검증 결과

본 연구에서는 기존 모델의 변화를 최소화하기 위해 검증데이터를 한 개씩 추가하여 분석을 진행하였으며, 분석 결과는 〈Table 7〉에 제시된 바와 같다.

추가 검증데이터를 대입하여 예측을 진행한 결과, 도출된 인공신경망 모델의 오차의 범위는 −10.96~5.28%로 Class 2의 오차범위인 −15~20% 내에 속하는 것으로 나타나 기본구상단계 견적을 위해 유효성을 지니는 것 뿐만 아니라 실시설계단계의 견적 오차 허용범위와 유사한 값으로 나타났다. 본 연구는 사업 초기 기본구상단계의 견적과 매칭되는 Class 5의 오차범위인 −50~100%의 범위 내에 속하면 유효성이 있는 것으로 해석할 수 있지만, 본 연구를 통해 추가 검증된 인공신경망 모델은 실시설계단계의 허용 오차율 수준의 정확도로 예측을 수행할 수 있는 것으로 나타났다. 또한, 구축 모델의 오차율은 −10.96~5.28%로 입찰계약 단계로 해석할 수 있는 Class 1의 오차범위인 −10~15%와 매우 근접한 오차율을 나타냈다.

Table 7. Rsults of validation data predict

Case	Actual value	Pradicted value	Error rate (%)	Absolute Value (%)
C 1	15,046,169,037	16,098,214,932	-6.99	6.99
C 2	4,330,294,685	4,404,964,634	-1.72	1.72
C 3	8,165,376,717	8,942,401,964	-9.51	9.51
C 4	15,463,137,801	14,857,694,095	3.91	3.91
C 5	6,754,434,668	6,978,011,794	-3.31	3.31
C 6	20,211,479,485	21,350,497,668	-5.63	5.63
C 7	8,229,921,858	8,997,024,422	-9.32	9.32
C 8	4,341,534,056	3,913,726,267	-9.85	9.85
C 9	9,705,666,535	8,641,882,447	-10.96	10.96
C 10	20,631,397,228	19,541,672,117	5.28	5.28
Average				6.64

즉, 구축된 인공신경망 모델의 공사비 예산 예측은 실시설계 단계 허용 오차율 내에 속하고, 입찰계약 단계의 견적 허용 오차율과 유사한 값을 나타내 기존 기본구상단계에서 활용하는 현행 방식보다 정확한 수준의 견적을 수행할 수 있는 것으로 나타났다.

또한, 오차율을 절대값으로 변환하여 평균을 분석한 결과, 오차율의 평균은 6.64%로 낮은 값을 나타내고 있으며 예측의 정확도 또한 입찰계약 단계에서 수행하는 견적과 유사한 수준으로 본 인공신경망 모델은 사업초기단계 공사비 예산 예측을 위해 유효성을 지니는 것으로 나타났다.

추가 검증데이터(사례)의 오차가 모두 허용 가능한 범위로 확인되었지만 상대적으로 오차범위가 높은 C3, C7, C8, C9 사례에 대한 세부 내용을 추가적으로 검토하였으며 두 가지 특징을 관찰할 수 있었다. 첫째, 높은 오차율을 보인 사례들은 연면적 값이 타 사례들에 비해 낮은 수준이었다. 이는 가장 주요 변수인 연면적 값이 과소 또는 과대인 아웃라이어(outlier)일 경우 예측의 정확도에 영향을 미칠 수 있다는 것을 시사한다. 또한 C3 및 C8 사례는 데이터 값에 일부 누락이 있는 사례로 확인되었다. 이는 입력 사례가 갖는 데이터 상의 오류나 누락이 있을 경우 예측의 오차율이 높아질 수 있음을 시사한다. 즉, 높은 정확도의 예측을 수행하기 위해서는 양질의 입력 데이터가 필요하며, 높은 중요도의 독립변수 값이 예측의 정확도에 영향을 미칠 수 있음을 확인할 수 있었다.

6. 결론

사업초기단계에서 공사비 예산을 보다 정확하게 예측하는 것은 발주자의 올바른 의사결정을 지원하고, 공사비 부족으로 인해 발생하는 다양한 리스크들에 선제적으로 대응할 수 있는 가능성을 높이는 매우 중요한 요소이다. 본 연구에서는 공공 건설공사 중 공공청사의 사업초기단계 공사비 예산 예측을 위한 인공신경망 모델을 구축하였으며 주요 특징 및 시사점은 다음과 같다.

- 인공신경망을 활용하여 사업초기단계에서 공공청사 공사비 예산을 예측할 수 있는 모델의 구축 가능성을 확인할 수 있었다.
- 동일한 모델 안에서도 입력 데이터에 따라 오차가 다르게 나타나며 오차가 높게 나타나는 데이터는 데이터 상에 오류 및 손실이 있는 데이터인 것으로 나타났다. 이는 인공신경망 모델을 활용하여 공사비 예산을 예측할 경우, 입력 데이터의 질이 모델의 정확성에 영향을 미칠 수 있음을 시사한다.
- 인공신경망을 활용한 공공청사 공사비 예산 예측에서는 연면적에 의존하는 기존 방식에 비해 보다 다양한 변수들이 반영된 예측이 가능하고, 활용된 변수들은 각기 다른 중요도를 지니고 있으며, 특히 연면적, 대지면적, 건축면적, 공사기간 등의 입력변수가 높은 중요도를 갖는 것으로 나타났다.
- 공공청사 공사비 예산 예측을 위한 인공신경망 모델은 AACEI의 Construction cost estimate accuracy를 기준으로 할 때 사업초기단계에서 허용되는 오차율에 유효할 뿐만 아니라 실시설계 단계에서 허용되는 오차율과 유사한 오차율을 지닌다. 이는 인공신경망 모델을 활용하여 사업초기단계 공사비 예산에 대한 보다 신뢰성 있는 예측을 수행할 수 있으며, 이는 사업초기단계 사업비관리에 긍정적인 영향을 미칠 수 있다.

본 연구에서는 인공신경망이 보다 정확한 공공청사 공사비 예산 예측을 위해 활용될 수 있는 가능성을 확인하였으

며. 이는 공사비 예산뿐만 아니라 건설사업의 다양한 비용예측에서도 활용될 수 있다는 것을 의미한다. 인공신경망 모델 구축에 있어 더욱 정확도 높은 모델을 구축하기 위해서는 예측 목적에 적합한 데이터를 선정하고 이에 맞는 모델을 구축, 높은 질의 데이터를 활용하는 것이 중요하며 이는 데이터 수집시 예측 목적에 적합하고 높은 질의 데이터를 수집하여 활용해야함을 의미한다.

본 연구에서 개발된 예측모델을 활용함에 있어서 한가지 고려해야 할 사항은 보정에 관한 것이다. 본 연구에서 사용된 공사비 데이터는 조달청 공사비정보광장에서 제시하고 있는 공사비 데이터를 그대로 추출하여 사용한 데이터이다. 따라서 각 공사비 데이터는 해당 공사가 발주되었던 시점의 물가를 반영하고 있는 데이터이다. 따라서 현 시점에서 예측모델을 통해 공사비 예산을 설정하는 경우에는 최근 물가 동향을 반영하여 보정 할 필요가 있다. 특히, 최근처럼 원자재 가격 상승으로 인해 공사비 예산이 급격하게 상승한 시기에는 이러한 보정이 더욱 필요하다. 결국 본 연구에서 개발된 예측모델이 산출한 공사비 예산은 과거 데이터를 기반으로 산출된 참고 가격(reference price)의 성격을 지니며 발주자가 이를 활용해서 현 시점의 공사비 예산을 산정하기 위해서는 별도의 보정이 필요하다고 할 수 있다. 이러한 보정 과정은 본 연구에서 개발된 예측모델에 국한된 것이 아니라 과거 실적 데이터를 활용하는 견적에서는 필수적인 과정이라고 할 수 있다.

본 연구에 이어 후속 연구에서는 사업초기단계 예비비(contingency) 예측을 위한 모델 개발에 관한 연구를 추진할 계획이다.

감사의 글

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2022R1F1A1064006).

References

Baccarini, D. (2005). Estimating Project Cost Contingency – Beyond the 10% Syndrome. Conference Paper of Australian Institute of Project Management National Conference, Nov 9 2005. Victoria: Australian Institute of Project Management.

Chandgude, S.E., and Sarode G.C. (2020). "To Study Cost Prediction Analysis of Construction Project Using ANN Model(Artificial Neural Network) SVM by Matlab." *Journal of Emerging Technologies and Innovative Research (JETIR)*, JEITIR, 7(7), pp. 291-306.

Gunaydin, H.M., and Dogan, S.Z. (2004). "A Neural Network Approach for Early Cost Estimation of Structural Systems of Buildings." *Journal of Elsevier*, Elsevier, 22(7), pp. 595-602.

Han, H.D., Kim, J.H., Yoon, J.H., and Seo, J.W. (2011). "Road Construction Cost Estimation Model in the Planning Phase Using Artificial Neural Network." *Journal of Korean Society of Civil Engineering*, Korean Society of Civil Engineering, 31(6), pp. 829-837.

Hashemi, S.T., Ebadati, O.M., and Kaur, H. (2020). Cost Estimation and Prediction in Construction Projects: a Systematic Review on Machine Learning Techniques. SN Applied Science Research Report, Review Paper.

Heo, J.Y. (2014). "Prediction of Road-Construction Cost Using an Artificial Neural Network Method." Master Thesis, *Journal of Kang-Won University*, Kang-Won University, Kangwondo, Korea.

Hong, S.I. (2021). Recent Research Trends and Revitalization of Machine Learning in Building Operation and Maintenance. KICEM Research Report, 22(2).

Ji, S.H., Park, M.S., Lee, H.S., and Yoon, Y.S. (2008). "Direction for Improving Cost Estimation and Management of Construction Projects: Comparing to Australian System." *Korean Journal of Construction Engineering and Management*, KICEM, 9(2), pp. 170-181.

Jo, Y.H., Choi, H.J., Kim, J.W., and Yun, S.H. (2020). "Conceotual Cost Estimate Method of Public Office Building Structural Frame Work by Regression Analysis." *Journal of the Korea Institute of Building Construction*, KIC, 20(2), pp. 147-153.

Jung, E.H., Yeom, D.J., Kim, J.S., and Kim, Y.S. (2021). "Development of Construction Cost Estimation Model for Elementary, Middle and High School Facilities." *Korean Journal of Construction Engineering and Management*, KICEM, 22(3), pp. 69-78.

Kim, C.Y., and Son, J.H. (2006). "A Study on the Model of Artificial Neural Networkfor Construction Cost Estimation of Educational Facilities at Conceptual Stage." *Korean Journal of Construction Engineering and Management*, KICEM, 7(4), pp. 91-99.

Kim, I.K., and Park, J. (2022). "A Study on Improvement of Construction Cost Estimation Method in the Feasibility Study Stage of Health Facilities in ODA Project." *Journal of Korea Association of International Development and Cooperation*, KAIDEC, 14(1), pp. 99-118.

Kim, J.H., Jun, S.M., Hwang, S.H., Kim, H.K., Heo, J.M., and Kang, M.S. (2021). "Impact of Activation Functions on Flood Forecasting Model Based on Artificial Neural

Networks." *Journal of the Korean Society of Agricultural Engineers*, KSAE, 63(1), pp. 11-25.

Kim, Y.C., Yoo, W.S., and Shin, Y.S. (2017). "Application of Artificial Neural Networks to Prediction of Construction Safety Accidents." *Journal of Korean Society of Hazard Mitigation*, KOSHAM, 17(1), pp. 7-14.

Korea Institute of Procurement. (2021). 「2022 Construction Order Guidebook.」 Korea Institute of Procurement (2021), pp. 1-51.

Kulkarni, P.S., Londhe, S.N., and Deo, S.N. (2017). "Artificial Neural Networks for Construction Management: A Review." *Journal of Soft Computing in Civil Engineering*, SCCE, 1(2), pp. 70-88.

Lee, D.H. (2018). Artificial Neural Network to Improve Learning Effectiveness of Artificial Intelligence. Telecommunications Technology Association (TTA) Research Report.

Lee, H.S., Jung, M.J., Park, M.S., and Son, B.S. (2010). "Cost Estimating for Public Facilities at Early Stage Using Functional Area Cost -Focusing on Army Barracks-" *Korean Journal of Construction Engineering and Management*, KICEM, 6(1), pp. 3-13.

Lee, J.G., Lee, H.S., Park, M.S., and Seo J.O. (2021). "Early-Stage Cost Estimation Model for Power Generation Project with Limited Historical Data." *Journal of Emarald, Emerald*, ISSN 0969-9988.

Lee, J.Y., and Lee, B.J. (2021). "Application of Artificial Neural Network to Predict Aerodynamic Coefficients of the Nose Section of the Missiles." *Journal of The Korean Society for Aeronautical & Space Science*, KSAS, 49(11), pp. 901-907.

Oh, S.B., and Lim, H.I. (2019). "A Study on Hidden Layer for Neural Network Model." *Proceedings of the Korea Information Processing Society Conference*, KIPS, 26(1), pp. 467-769.

Park, H.J., Choi, S.B., and Kang, C.W. (2020). "Determination of Input Nodes for Neural Network Model in Time Series Data." *Journal of the Korean Data Analysis Society*, KDAS, 22(3), pp. 1055-1065.

Park, M.S., Seong, K.H., Lee, H.S., Ji, S.H., and Kim, S.Y. (2010). "Schematic Cost Estimation Method using Case-Based Reasoning: Focusing on Determining Attribute Weight." *Korean Journal of Construction Engineering and Management*, KICEM, 11(4), pp. 22-31.

Van, S.G., Le-Hoai, L., and Dang, C.N. (2019). "Predicting Implementation Cost Contingencies for Residential Construction Projects in Flood-Prone Areas." *Journal of Emerald*, Emerald, 12(10), pp. 1097-1119.

Yoon, J.M. (2019). "Effectiveness Analysis of Credit Card Default Risk with Deep Learning Neural Network." *Journal of Korea Money and Finance Association*, KMFA, 33(1), pp. 151-183.

요약 : 건설사업의 사업초기단계에 산정되는 공사비 예산을 적절히 예측하는 것은 발주자의 올바른 의사결정을 지원하고 건설사업의 목표를 달성하기 위해 매우 중요한 현안이다. 이는 공공 건설사업의 경우에서도 마찬가지이다. 그러나 현재 공공 건설사업의 사업초기단계에서 수행되는 공사비 예산의 예측방식은 정확성 및 신뢰성 관점에서 정교하지 못해 이에 대한 개선의 필요성이 제기되고 있다. 본 연구의 목적은 인공신경망을 활용하여 공공청사 프로젝트 사업초기단계에서 활용할 수 있는 공사비 예산 예측모델을 개발하는데 있다. 본 연구에서는 조달청에서 제공하는 데이터와 SPSS Statistics 프로그램을 활용하여 인공신경망 모델을 구축하였으며, 공사비 예산 예측의 수준을 분석하고 추가 검증을 통해 모델의 정확성을 검증하였다. 검증 결과, 개발된 인공신경망 모델은 사업초기단계에서 활용할 수 있는 견적의 오차범위를 보여주었으며 이를 통해 다양한 프로젝트 조건(변수)을 활용하여 보다 정교하게 공사비 예산을 예측할 수 있는 가능성을 시사하였다.

키워드 : 공사비 예측 모델, 공사비 예산 예측, 공공청사, 인공신경망