

Xây dựng mô hình hỗ trợ báo giá đấu thầu phần kết cấu thép tại Việt Nam

Developing a steel structure bidding support model in Vietnam

> PHAN QUỐC THÁI^{1,3*}, LÊ ANH TUẤN^{2,3}, HỒ THỊ QUỲNH CHÂU⁴, VÕ TIẾN ĐẠT⁵

¹Bộ môn TC&QLXD, Khoa Kỹ thuật xây dựng, Trường ĐH Bách khoa TP.HCM

²Bộ môn VLXD, Khoa Kỹ thuật xây dựng, Trường ĐH Bách khoa TP.HCM

³Đại học Quốc gia TP.HCM

⁴Kiểm toán nhà nước khu vực 13

⁵Công ty TNHH Kirby Đông Nam Á

*Tác giả liên hệ: phanquocthai@hcmut.edu.vn

TÓM TẮT

Sự phát triển mạnh mẽ của các khu công nghiệp (KCN) tại Việt Nam làm gia tăng cạnh tranh trong lĩnh vực xây dựng kết cấu thép nhà xưởng. Nghiên cứu được xây dựng để góp phần hỗ trợ nhà thầu trong việc ước tính chi phí xây dựng kết cấu thép nhà xưởng, phục vụ công tác báo giá dự thầu cũng như tạo cơ sở để so sánh chi phí giữa các phương án xây dựng nhà xưởng. Nghiên cứu đã đưa ra 16 nhân tố ảnh hưởng chính và xây dựng một mô hình ước lượng chi phí ứng dụng mô hình mạng nơ-ron nhân tạo (ANN) trên phần mềm Rapidminer Studio dựa trên nguồn dữ liệu 40 dự án đã thực hiện. Đóng góp của nghiên cứu nhằm tiết kiệm thời gian và tăng tính chính xác cho việc báo giá dự thầu kết cấu thép nhà xưởng và tạo cơ sở tiền đề cho các nghiên cứu sau này.

Từ khóa: Kết cấu thép; nhà xưởng; ước tính chi phí; đấu thầu; mạng nơ-ron nhân tạo (ANN).

ABSTRACT

The strong development of industrial parks in Vietnam has increased competition in the field of steel structure workshop construction. This research is conducted to assist contractors in estimating the cost of constructing steel structure workshops, serving the bidding process as well as providing a basis for comparing costs between different workshop construction options. The study identified 16 main influencing factors and built a cost estimation model applying an artificial neural network (ANN) model on Rapidminer Studio software based on data from 40 completed projects. The contribution of the research is to save time and increase accuracy for bidding on steel structure workshops and create a premise for future studies.

Keywords: Steel structure; factory; cost estimation; bidding; Artificial Neural Network (ANN).

1. GIỚI THIỆU

Trong lĩnh vực xây dựng tại Việt Nam, thép và bê tông cốt thép luôn được xem là những loại vật liệu xây dựng quan trọng. Với việc được ứng dụng rộng rãi trên toàn cầu, kết cấu thép đã du nhập và phát triển mạnh mẽ tại Việt Nam trong hơn ba thập kỷ qua. Sự tham gia của các nhà thầu quốc tế đóng vai trò quan trọng trong quá trình này. Những ưu điểm vượt trội của kết cấu thép đã khẳng định vai trò của nó, trở thành lựa chọn ưu tiên trong nhiều hạng mục công trình như nhà xưởng, nhà nhịp lớn và các công trình công cộng. Sự phổ biến ngày càng tăng của kết cấu thép đòi hỏi phải nâng cao hiệu quả quản lý chi phí, đặc biệt trong bối cảnh giá thép biến động và yêu cầu tối ưu hóa chi phí xây dựng. Điều này đòi hỏi những phương pháp quản lý chi phí hiện đại và hiệu quả.

Để giải quyết vấn đề này, việc ứng dụng các phương pháp hiện đại là vô cùng cần thiết. Trong đó, Mạng nơ-ron nhân tạo (Artificial Neural Network - ANN) nổi lên như một công cụ tiềm năng. Được giới thiệu lần đầu tiên vào năm 1943 bởi Warren McCulloch và Walter Pitts, ANN đã trải qua một chặng đường dài phát triển. Mặc dù những hạn chế về kỹ thuật ban đầu đã cản trở việc nghiên cứu sâu, nhưng những tiến bộ vượt bậc trong lĩnh vực mô phỏng những năm gần đây đã mở ra cánh cửa ứng dụng rộng rãi cho ANN trong nhiều lĩnh vực, bao gồm điện - điện tử, kỹ thuật chế tạo, y học, quản sự, kinh tế, và gần đây nhất là quản lý dự án xây dựng. Với khả năng học hỏi từ dữ liệu, nhận diện các mô hình phức tạp và đưa ra dự báo chính xác, ANN hứa hẹn mang lại những giải pháp hiệu quả trong việc ước tính, kiểm soát và tối ưu chi phí cho các dự án sử dụng kết cấu thép. Những lợi ích cụ thể như giảm thiểu sai số dự toán, rút ngắn thời gian lập kế hoạch, và nâng cao hiệu quả quản lý vật tư sẽ góp phần thúc đẩy sự phát triển bền vững của ngành Xây dựng.

Ngày nay, ứng dụng ANN trong ước tính chi phí đã trở thành chủ đề nghiên cứu phổ biến trên phạm vi quốc tế nói chung và tại Việt Nam nói riêng. Ngay từ những năm đầu thập niên 90, đã có những nghiên cứu tiên phong sử dụng mạng nơ-ron lan truyền ngược để dự đoán biến động chỉ số giá xây dựng dựa trên các yếu tố như giá xây dựng, lãi suất, tỷ giá và số lượng nhà, cho thấy khả năng nhận dạng mẫu tốt mặc dù gặp hạn chế trong dự báo do tính phức tạp của vấn đề [1]. Tiếp nối hướng nghiên cứu này, sau đó đã xuất hiện các mô hình ANN sử dụng cơ sở dữ liệu lớn với nhiều thuộc tính khác nhau nhằm ước tính chi phí xây dựng với độ chính xác cao [2]. Trong một nghiên cứu khác, đã có ứng dụng của mạng nơ-ron chính quy (Regularization Neural Network) kết hợp với phương pháp kiểm chứng chéo để dự đoán chi phí xây dựng. Mô hình xem xét các yếu tố đầu vào như diện tích sàn, số tầng, số lượng và diện tích trung

bình hạng mục, vị trí, loại mái và mức độ hoàn thiện. Kết quả cho thấy độ tin cậy cao và khả năng dự báo ổn định, không bị ảnh hưởng bởi kiến trúc mạng hay tham số huấn luyện [3]. Các nghiên cứu tiếp theo đã ứng dụng ANN để ước tính chi phí kết cấu nhà thép từ dữ liệu của nhiều dự án, cho thấy hiệu quả vượt trội so với các phương pháp hồi quy truyền thống nhờ vào độ chính xác cao và khả năng ước tính được cả chi phí trực tiếp lẫn lợi nhuận, đồng thời đáp ứng linh hoạt với biến động của thị trường [4]. Sự kết hợp giữa ANN và các phương pháp dựa trên tình huống đã mang lại độ chính xác cao với sai số trung bình rất thấp, khẳng định tiềm năng ứng dụng của các phương pháp này [5]. Một số nghiên cứu khác đã phát triển các mô hình kết hợp dựa trên kinh nghiệm và phương pháp phân tích thứ bậc, mặc dù yêu cầu cấu trúc phức tạp, nhưng góp phần làm nổi bật tiềm năng của các mô hình cải tiến [6]-[7]. Ngoài ra, ANN cũng đã được ứng dụng hiệu quả trong dự báo chi phí xây dựng đường cao tốc, cho thấy khả năng nắm bắt các thay đổi của thị trường và hỗ trợ lập kế hoạch ngân sách [8]. Một số nghiên cứu khác phát triển mô hình kết hợp giữa ANN, mạng nơ-ron bậc cao và logic mờ đã đạt được độ chính xác cao ngay từ giai đoạn đầu dự án, vượt trội hơn so với mạng nơ-ron tuyến tính đơn lẻ [9]. Tại Việt Nam, các ứng dụng của phương pháp mô phỏng như động lực học hệ thống trong phân tích và dự báo chi phí xây dựng nhà thép tiền chế cũng đã cung cấp thông tin quý báu về khối lượng công trình và các chi phí liên quan [10].

Từ đó có thể thấy, mặc dù nhiều nghiên cứu về ứng dụng ANN trong dự toán chi phí xây dựng đã đạt được kết quả khả quan, việc chuyển giao các mô hình này vào thực tiễn, đặc biệt trong lĩnh vực nhà thép tiền chế tại Việt Nam, vẫn còn nhiều tiềm năng cần được khai thác. Trong bối cảnh cạnh tranh gay gắt với các hợp đồng giá cố định và yêu cầu tiết kiệm chi phí, việc dự đoán chính xác chi phí thực hiện dự án là yếu tố then chốt, vì ước tính quá cao có thể làm mất cơ hội trúng thầu, trong khi ước tính quá thấp lại dẫn đến thua lỗ cho nhà thầu. Do đó, xây dựng một mô hình hỗ trợ báo giá đấu thầu cho phần kết cấu thép, dựa trên nền tảng ANN và các kỹ thuật tiên tiến, là hết sức cần thiết nhằm rút ngắn thời gian lập dự toán, nâng cao năng lực cạnh tranh và tạo dựng niềm tin với khách hàng thông qua tính minh bạch và chính xác trong dự báo chi phí.

2. TỔNG QUAN VỀ ƯỚC TÍNH CHI PHÍ XÂY DỰNG

Chi phí phần kết cấu thép của dự án nhà xưởng cơ bản bao gồm các khoản: chi phí thiết kế, nguyên vật liệu, gia công, vận chuyển và bảo quản, chi phí thi công cùng các chi phí khác. Theo nghiên cứu [11], chi phí kết cấu thép của dự án nhà xưởng được phân chia thành: khoảng 25% cho nguyên vật liệu; 35% cho gia công, sản xuất (bao gồm vận chuyển và bảo quản); 25% cho thi công; và 15% cho các chi phí khác. Trong các khoản mục trên, chi phí thi công là khó xác định nhất do phụ thuộc vào đơn giá thi công, vốn bị chi phối bởi nhiều yếu tố như quy mô dự án, địa điểm xây dựng, biện pháp thi công, thời điểm và tiến độ thi công. Vì vậy, việc xác định một đơn giá lắp dựng chính xác và tối ưu là thách thức lớn đối với cả chủ đầu tư lẫn nhà thầu.

Văn phòng Trách nhiệm Chính phủ Hoa Kỳ (GAO) đã định nghĩa ước tính chi phí là “tổng hợp các yếu tố chi phí riêng lẻ, được tính toán theo các phương pháp đã được thiết lập và sử dụng dữ liệu hợp lệ để dự báo chi phí tương lai của chương trình, dựa trên những thông tin hiện có” [12]. Việc ước lượng chi phí được thể hiện qua 5 lớp, như trình bày trong Bảng 1 [13]:

Nghiên cứu này áp dụng mức ước tính thuộc lớp 1 và lớp 2 khi mức định nghĩa dự án đạt ít nhất 30%.

Trong thực tiễn đầu tư xây dựng, công tác định giá và quản lý giá vẫn còn tồn tại một số bất cập. Như thiếu căn cứ khoa học để xác

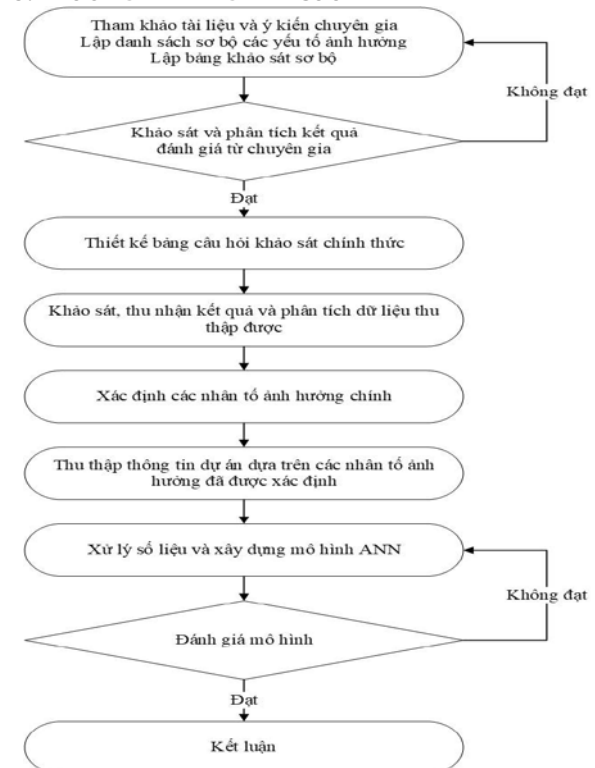
định các chỉ tiêu về giá sản phẩm xây dựng trong các giai đoạn đầu tư, chẳng hạn như tổng mức đầu tư hay tổng dự toán, vốn thường được xác định theo phương pháp ước tính. Đồng thời, nhiều hạng mục công việc chưa có định mức rõ ràng hoặc đang áp dụng các đơn giá lỗi thời, không đáp ứng được yêu cầu của sự phát triển công nghệ và thực tiễn quản lý giá hiện nay. Hơn nữa, quá trình quản lý giá trong đấu thầu xây dựng còn nhiều bất cập, khi giá gói thầu trong kế hoạch đấu thầu khó có thể được xác định rõ ràng so với dự toán hay tổng dự toán đã được duyệt.

Bảng 1: Các lớp trong ước lượng chi phí

Lớp ước tính	Tên	Mục đích	Mức độ định nghĩa dự án
Lớp 5	Theo độ	Sàng lọc hoặc khả thi	0% đến 2%
Lớp 4	Trung gian	Nghiên cứu khái niệm hoặc tính khả thi	1% đến 15%
Lớp 3	Sơ bộ	Ngân sách, ủy quyền hoặc kiểm soát	10% đến 40%
Lớp 2	Thực chất	Kiểm soát hoặc bỏ thầu/đấu thầu	30% đến 70%
Lớp 1	Dứt khoát	Kiểm tra ước tính hoặc bỏ thầu/đấu thầu	50% đến 100%

Tình trạng bỏ thầu giá thấp đang phổ biến, dẫn đến hiện tượng thi công ẩu, cắt giảm vật liệu, qua đó làm giảm chất lượng công trình và ảnh hưởng tiêu cực đến lợi ích lâu dài cũng như sự phát triển của doanh nghiệp. Vì vậy, việc ước tính chi phí xây dựng đóng vai trò vô cùng quan trọng. Nghiên cứu này hướng tới mục tiêu xác định các nhân tố ảnh hưởng đến chi phí xây dựng nhà xưởng, cung cấp cái nhìn tổng quan cho chủ đầu tư và nhà thầu, đồng thời xây dựng mô hình ước lượng chi phí dựa trên mạng nơ-ron nhân tạo (ANN) nhằm tối ưu hóa thời gian báo giá và nâng cao độ chính xác, từ đó đặt nền móng cho các nghiên cứu tiếp theo trong lĩnh vực này.

3. PHƯƠNG PHÁP NGHIÊN CỨU



Hình 1. Sơ đồ quy trình nghiên cứu

Hình 1 minh họa quy trình nghiên cứu với các bước triển khai cụ thể. Trước hết, để xác định các nhân tố ảnh hưởng đến chi phí xây dựng kết cấu thép nhà xưởng, nghiên cứu tiến hành hai giai đoạn chính. Giai đoạn đầu là tổng quan tài liệu trong và ngoài nước, qua đó xây dựng bảng tổng hợp sơ bộ các nhân tố tiềm năng. Bảng này được hiệu chỉnh và hoàn thiện thông qua việc khảo sát ý kiến chuyên gia giàu kinh nghiệm trong lĩnh vực xây dựng nhà xưởng kết cấu thép. Các chuyên gia đã tiến hành đánh giá, loại bỏ và bổ sung các nhân tố nhằm đảm bảo tính phù hợp với bối cảnh thực tiễn tại Việt Nam. Kết quả phân tích giá trị trung bình cho thấy có 18 nhân tố ảnh hưởng chính: Thời điểm thi công, Địa điểm công trình, Công năng sử dụng, Biện pháp thi công, Tiến độ thi công, Kết cấu khung, Khẩu độ nhà (nhịp nhà), Bước nhà, Chiều cao nhà, Cầu trục, Số tầng, Địa hình tính gió, Tải treo (bao gồm tải phụ và tải pin mặt trời), Diện tích xây dựng, Loại mái, Loại vách, Hệ số trượt giá và Hệ số ứng suất (Stress ratio). Trên cơ sở 18 nhân tố này, nghiên cứu tiếp tục khảo sát diện rộng để đánh giá mức độ ảnh hưởng của từng nhân tố. Đối tượng khảo sát là các cá nhân giữ vị trí quản lý và chuyên môn trong lĩnh vực xây dựng, bao gồm Giám đốc dự án, Trưởng/Chuyên viên phòng Dự án, Trưởng phòng Thi công, Chỉ huy trưởng, Ban Chỉ huy công trình, và Trưởng/Chuyên viên phòng Quản lý dự án. Tổng cộng 80 phiếu khảo sát được phát ra, thu về 75 phiếu hợp lệ (tỷ lệ phản hồi 93,75%). Để kiểm tra độ tin cậy của thang đo, nghiên cứu áp dụng phân tích hệ số Cronbach's Alpha. Kết quả cho thấy một số nhân tố có hệ số tương quan biến-tổng thấp, làm giảm độ tin cậy tổng thể của thang đo. Sau khi loại bỏ các nhân tố này, 16 nhân tố quan trọng nhất ảnh hưởng đến chi phí xây dựng kết cấu thép nhà xưởng trong giai đoạn đầu thầu được xác định, gồm: Thời điểm thi công, Địa điểm công trình, Biện pháp thi công, Tiến độ thi công, Kết cấu khung, Khẩu độ nhà (nhịp nhà), Bước nhà, Chiều cao nhà, Số tầng, Địa hình tính gió, Tải treo (bao gồm tải phụ và tải pin mặt trời), Diện tích xây dựng, Loại mái, Loại vách, Hệ số trượt giá và Hệ số ứng suất (Stress ratio).

Kế tiếp, từ 16 nhân tố ảnh hưởng đã xác định, nghiên cứu tiến hành thu thập dữ liệu từ hồ sơ của 40 dự án KCN. Các thông tin liên quan đến từng nhân tố được tổng hợp làm dữ liệu đầu vào cho mô hình. Để đảm bảo tính khách quan trong việc đánh giá hiệu năng, dữ liệu này được chia thành hai tập riêng biệt: tập dữ liệu dùng để huấn luyện (đào tạo) và tập dữ liệu dùng để kiểm tra mô hình. Cụ thể, theo phương pháp 5-fold cross-validation, 80% dữ liệu được sử dụng để đào tạo mô hình và 20% còn lại được dùng để kiểm tra, với tỉ lệ này được lựa chọn nhằm tối đa hóa thông tin huấn luyện trong khi vẫn đảm bảo được độ tin cậy của kết quả kiểm tra. Trên cơ sở dữ liệu này, mô hình mạng nơ-ron nhân tạo (ANN) được xây dựng và huấn luyện bằng phần mềm Rapidminer Studio.

Sau khi hoàn thành quá trình đào tạo, hiệu suất của mô hình được kiểm nghiệm trên tập dữ liệu kiểm tra. Để đánh giá mức độ chính xác và tính hiệu quả của mô hình, nghiên cứu áp dụng các chỉ tiêu đánh giá phổ biến như hệ số tương quan tuyến tính (R), sai số trung bình tuyệt đối (MAE), sai số phần trăm trung bình tuyệt đối (MAPE) và sai số toàn phương trung bình (RMSE).

Đồng thời, nhằm so sánh và khẳng định khả năng dự báo, nghiên cứu cũng triển khai các thuật toán đối chứng gồm hồi quy tuyến tính và máy véc-tơ hỗ trợ (SVM) trên cùng bộ dữ liệu, qua đó cho phép so sánh trực tiếp hiệu quả của các phương pháp dự báo chi phí xây dựng. Qua đó, kết quả thu được sẽ giúp xác định mô hình dự báo có khả năng tổng quát hóa tốt nhất và ứng dụng hiệu quả trong thực tiễn.

4. XÂY DỰNG VÀ THỰC HIỆN MÔ HÌNH

Trong quá trình triển khai nghiên cứu, việc sử dụng Rapidminer góp phần quan trọng trong quá trình xử lý và xây dựng mô hình. Rapidminer là một môi trường mã nguồn mở, được viết bằng ngôn

ngữ lập trình Java, áp dụng mô hình Client/Server và có thể triển khai trên máy chủ on-premise, public cloud hoặc private cloud. Nền tảng này cung cấp đầy đủ các lược đồ và thuật toán trong khai phá dữ liệu, giúp thực hiện các tác vụ từ tải và chuyển đổi dữ liệu (ETL), xử lý và trực quan hóa dữ liệu, đến xây dựng các mô hình dự báo và phân tích thống kê. Đặc biệt, Rapidminer cho phép mở rộng các chức năng bằng cách tích hợp ngôn ngữ R và Python, giúp linh hoạt trong việc ứng dụng các thuật toán học máy.

Trong nghiên cứu này, để dự báo chi phí xây dựng, các thuật toán thuộc nhóm học giám sát được áp dụng, trong đó có ba phương pháp đối chứng: Mạng nơ-ron nhân tạo (ANN), Máy véc-tơ hỗ trợ (SVM) và Hồi quy tuyến tính. Mỗi thuật toán đều có những ưu điểm riêng biệt.

4.1. Thu thập và xử lý số liệu

Dữ liệu: Các thông tin được thu thập từ hồ sơ của 40 dự án xây dựng, dựa trên 16 nhân tố ảnh hưởng đã được xác định (xem Bảng 2 và Bảng 3).

Bảng 2: Các nhân tố và kiểu dữ liệu được chọn để thu thập

STT	Nhân tố ảnh hưởng	Phân loại	Thứ nguyên
1	Thời điểm thi công	2018, 2019, 2020, 2021	Không thứ nguyên
2	Địa điểm công trình	Bắc, Trung, Nam	Không thứ nguyên
3	Biện pháp thi công	Khó, trung bình	Không thứ nguyên
4	Tiến độ thi công		Ngày
5	Kết cấu khung	RF, MS, BC1, BC2, BC3, SS	Không thứ nguyên
6	Khẩu độ nhà (Nhịp nhà)		m
7	Bước nhà		m
8	Chiều cao nhà		m
9	Số tầng	[0÷4]	Không thứ nguyên
10	Địa hình tính gió	B, C, D	Không thứ nguyên
11	Tải treo (bao gồm tải phụ và tải pin mặt trời)	[0.05÷0.47]	kN/m ²
12	Diện tích xây dựng		m ²
13	Loại mái	Tôn mạ màu, tôn thường, panel	Không thứ nguyên
14	Loại vách	Tôn mạ màu, tôn thường, panel	Không thứ nguyên
15	Hệ số trượt giá		%
16	Hệ số ứng suất	[0.93÷1]	Không thứ nguyên
17	Giá trúng thầu		VNĐ

Trong nghiên cứu, các dữ liệu về địa điểm xây dựng, biện pháp thi công, kết cấu khung, loại tường và loại mái được thu thập và phân loại một cách chi tiết. Cụ thể, dữ liệu về địa điểm xây dựng được thu thập từ 20 khu vực khác nhau, bao gồm An Giang, Bình Dương, Bình Phước, Cần Thơ, Đà Nẵng, Đồng Nai, Hà Nội, Hải Phòng, Hậu Giang, Kiên Giang, Lâm Đồng, Long An, Nghệ An, Quảng Nam, Quảng Bình, Quảng Ngãi, Tây Ninh, Thừa Thiên Huế, TP.HCM và Trà Vinh. Về biện pháp thi công, dữ liệu được phân loại theo hai mức độ: khó và trung bình. Ngoài ra, dữ liệu về địa hình tính gió được thu thập dựa trên tiêu chuẩn ASCE 7-05, với các nhóm phân loại là B, C

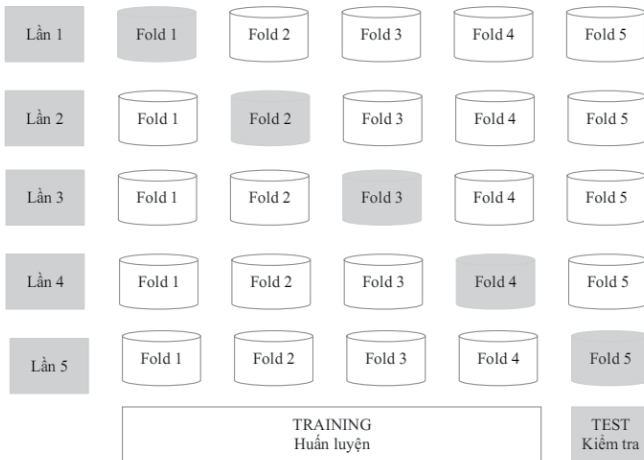
và D. Đối với loại tường và loại mái, mỗi yếu tố được chia thành ba loại: tôn mạ màu, tôn thường và panel.

Lớp nơ-ron đầu vào của mô hình sẽ nhận 40 bộ dữ liệu. Trước khi tiến hành xử lý, các kiểu dữ liệu này cần được chuyển đổi sang dạng số hoặc dạng mã hóa phù hợp. Bảng 3 cung cấp thông tin chi tiết về cách thức mã hóa và dạng dữ liệu tương ứng cho từng nhân tố.

Bảng 3: Mã hóa và gọi tên biến các nhân tố

Ký hiệu	Nhân tố ảnh hưởng	Dạng dữ liệu	Ký hiệu
X1	Thời điểm thi công	Dạng số nguyên	Integer
X2	Địa điểm công trình	Dạng chữ: vị trí (tỉnh)	Polynomial
X3	Biện pháp thi công	Dạng chữ	Polynomial
X4	Tiến độ thi công	Dạng số nguyên (ngày)	Integer
X5	Kết cấu khung	Dạng chữ	Polynomial
X6	Khẩu độ nhà (Nhịp nhà)	Dạng số thực (m)	Real
X7	Bước nhà	Dạng số thực (m)	Real
X8	Chiều cao nhà	Dạng số thực (m)	Real
X9	Số tầng	Dạng số nguyên	Integer
X10	Địa hình tính gió	Dạng chữ: B, C, D	Polynomial
X11	Tải treo (bao gồm tải phụ và tải pin mặt trời)	Dạng số thực (m)	Real
X12	Diện tích xây dựng	Dạng số thực (m)	Real
X13	Loại mái	Dạng chữ	Polynomial
X14	Loại vách	Dạng chữ	Polynomial
X15	Hệ số trượt giá	Dạng số thực	Real
X16	Hệ số ứng suất	Dạng số thực	Real
Y	Giá trúng thầu	Dạng số thực	Real

Chia dữ liệu: Áp dụng phương pháp 5-fold cross-validation, tập dữ liệu được chia thành 5 phần (mỗi phần gồm 8 bộ dữ liệu).



Hình 2. Phân chia tập dữ liệu thành 5-fold

4.2. Thiết lập các thông số

Mạng nơ-ron nhân tạo (ANN) là công cụ đáng tin cậy và được ứng dụng rộng rãi trong nhiều lĩnh vực. Tuy nhiên, độ chính xác của ANN bị ảnh hưởng bởi một số hạn chế từ các yếu tố bên ngoài và bên trong. Các yếu tố bên ngoài bao gồm đặc điểm dữ liệu đầu vào và quy mô nghiên cứu, trong khi các yếu tố bên trong liên quan đến cấu trúc mạng, trọng số ban đầu, số lần lặp, hàm truyền và tốc độ học. Do đó, thách thức lớn nhất trong thiết kế mạng nơ-ron là xác định các thông số tối ưu nhằm tối thiểu hóa sai số và tối đa hóa độ chính xác.

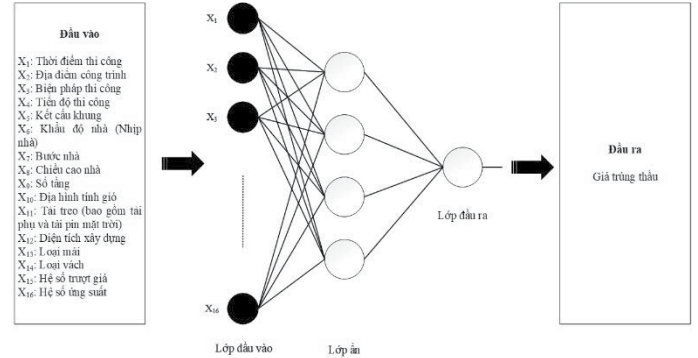
Các thông số cần thiết cho mô hình đã được tính toán và liệt kê ở Bảng 4. Việc thiết lập các thông số này trong phần mềm Rapidminer Studio sẽ dựa theo tài liệu hướng dẫn đi kèm. Quá trình này được thực hiện trước khi đưa các tập dữ liệu vào mô hình.

Bảng 4: Thiết lập các thông số cho mô hình mạng ANN

Thông số	Giá trị trong Rapidminer
Số nút của lớp vào	16
Số lớp ẩn	1
Số nút của lớp ẩn	4
Số vòng lặp	200
Momentum (động lượng)	0.9
Learning rate (tỉ lệ học)	0.02
Hàm truyền	Sigmoid
Sai số cho phép	1.0E-4

4.3. Xây dựng mô hình ANN

Sau khi dữ liệu được xử lý và chuyển đổi sang dạng mã hóa phù hợp, các tham số chính của mô hình ANN, bao gồm số lớp ẩn, số nút của lớp ẩn và hàm truyền, được xác định trước. Dựa trên tập dữ liệu đã mã hóa, mô hình ANN được xây dựng trên phần mềm Rapidminer Studio, trong đó: sử dụng 80% dữ liệu để đào tạo mô hình nhằm tối ưu hóa các trọng số, sau đó, 20% dữ liệu còn lại được dùng để kiểm tra, từ đó đánh giá khả năng tổng quát hóa của mô hình.



Hình 3. Mô hình huấn luyện ANN

4.4. Đánh giá mô hình và so sánh với thuật toán đối chứng

4.4.1. Đánh giá mô hình ANN

Sau khi hoàn thành quá trình huấn luyện, mô hình được đánh giá trên tập dữ liệu kiểm tra bằng các chỉ tiêu dự báo phổ biến hệ số tương quan tuyến tính (R), sai số trung bình tuyệt đối (MAE), sai số phần trăm trung bình tuyệt đối (MAPE) và sai số toàn phương trung bình (RMSE).

4.4.2. So sánh các thuật toán đối chứng

Để khẳng định tính hiệu quả của mô hình ANN, nghiên cứu còn xây dựng và so sánh với hai mô hình đối chứng:

- Máy véc-tơ hỗ trợ (SVM): Sử dụng hàm nhân để ánh xạ dữ liệu vào không gian đặc trưng, tìm kiếm siêu phẳng tối ưu nhằm phân tách và dự báo, đặc biệt hiệu quả với dữ liệu phân bố không đồng đều.

- Hồi quy tuyến tính: Phương pháp thiết lập mối quan hệ tuyến tính giữa các biến số và giá trị dự báo, tạo cơ sở so sánh khách quan với các phương pháp phức tạp hơn.

Cả ba thuật toán được huấn luyện và kiểm tra dựa trên bộ dữ liệu 40 công trình đã được thu thập. Qua đó, so sánh các chỉ tiêu (R, MAE, MAPE, RMSE) giúp khẳng định ưu thế của mô hình ANN trong việc xử lý các mối quan hệ phi tuyến phức tạp giữa các biến số, từ đó đưa ra kết luận về tính khả thi và ứng dụng của từng phương pháp.

5. KẾT QUẢ VÀ ĐÁNH GIÁ MÔ HÌNH

Kết quả đánh giá hiệu năng của các mô hình được trình bày qua các chỉ số thống kê phổ biến, gồm: hệ số tương quan (R), phần trăm sai số trung bình tuyệt đối (MAPE), sai số trung bình tuyệt đối (MAE) và sai số toàn phương trung bình (RMSE).

5.1. Đánh giá mô hình ANN

Bảng 5 dưới đây cho thấy hiệu suất của mô hình ANN đạt được từ quá trình huấn luyện và kiểm tra theo phương pháp 5-fold cross-validation:

Bảng 5: Kết quả hiệu suất mô hình ANN

Mô hình	R	MAPE (%)	MAE (nghìn VNĐ)	RMSE (nghìn VNĐ)
ANN	0.931	26.56%	1,707,889	2,276,932

Giá trị MAE của mô hình được xem là có thể chấp nhận đối với các dự án có tổng mức đầu tư lớn; tuy nhiên, đối với dự án quy mô nhỏ, sai số này có thể trở nên đáng kể. Do đó, MAE không phải là chỉ số duy nhất quyết định hiệu năng của mô hình. Mặt khác, MAPE đạt 26.56% được xem là mức sai số chấp nhận được trong bối cảnh báo giá đầu thầu, trong khi hệ số tương quan (R) đạt 0.931 cho thấy mối liên hệ tuyến tính chặt chẽ giữa giá trị thực tế và giá trị dự báo.

5.2. So sánh hiệu suất giữa các mô hình

Ngoài mô hình ANN, nghiên cứu còn triển khai và so sánh hiệu suất của hai mô hình đối chứng khác: Hồi quy tuyến tính và Máy véc-tơ hỗ trợ (SVM). Kết quả so sánh được trình bày trong Bảng 6 dưới đây:

Bảng 6: So sánh kết quả các mô hình

Mô hình	R	MAPE (%)	MAE (nghìn VNĐ)	RMSE (nghìn VNĐ)
ANN	0.931	26.56%	1,707,889	2,276,932
Hồi quy	0.670	37.70%	1,448,755	1,808,003
SVM	0.256	40.86%	5,283,202	8,558,973

Nhìn chung, mô hình ANN đạt hiệu suất cao nhất với hệ số tương quan (R) là 0.931, vượt trội so với mô hình hồi quy và SVM. Điều này cho thấy mô hình ANN có mức độ tương quan tuyến tính tốt hơn giữa giá trị thực tế và giá trị dự báo, đồng thời phản ánh độ tin cậy cao hơn so với hai mô hình còn lại.

6. KẾT LUẬN

Nghiên cứu đã xây dựng và đánh giá một mô hình dự báo chi phí xây dựng kết cấu thép nhà xưởng dựa trên dữ liệu thực tiễn từ 40 dự án. Nghiên cứu không chỉ nhằm xác định các nhân tố ảnh hưởng chủ chốt mà còn làm rõ mối quan hệ giữa các yếu tố này với chi phí xây dựng, từ đó tạo nên một công cụ hữu ích hỗ trợ quyết định trong quá trình báo giá đầu thầu và lựa chọn phương án thiết kế, thi công. Dữ liệu được thu thập thông qua quá trình tổng hợp tài liệu và khảo sát ý kiến chuyên gia, giúp xác định được 16 nhân tố quan trọng.

Mô hình dự báo được xây dựng trên nền tảng Rapidminer, một môi trường mã nguồn mở mạnh mẽ cho khai phá dữ liệu, cho phép xử lý và trực quan hóa dữ liệu cũng như xây dựng các mô hình dự báo. Qua quá trình huấn luyện và kiểm tra, mô hình ANN cho thấy hiệu năng vượt trội với các chỉ số đánh giá đạt giá trị mong đợi (hệ số tương quan $R = 0.931$, $MAPE = 26,56\%$, cùng với MAE và RMSE ở mức chấp nhận được). Kết quả so sánh với các mô hình đối chứng như hồi quy tuyến tính và máy véc-tơ hỗ trợ (SVM) càng khẳng định tính ưu việt của mô hình ANN trong việc nắm bắt các mối quan hệ phi tuyến phức tạp giữa các biến số.

Với những kết quả từ nghiên cứu cho thấy mô hình dự báo chi phí xây dựng không chỉ cung cấp một cái nhìn tổng thể về các yếu tố ảnh hưởng mà còn giúp đưa ra các ước tính chính xác, từ đó hỗ trợ hiệu quả cho các nhà thầu và chủ đầu tư trong bối cảnh thị trường xây dựng ngày càng cạnh tranh và biến động. Nghiên cứu mở ra hướng đi mới trong việc áp dụng các mô hình dự báo chi phí trong thực tiễn đầu tư xây dựng tại Việt Nam, đồng thời tạo cơ sở cho các nghiên cứu cải tiến nhằm tăng cường độ chính xác và khả năng ứng dụng của các mô hình dự báo trong tương lai.

Lời cảm ơn: Nhóm tác giả xin chân thành cảm ơn Trường Đại học Bách khoa, Đại học Quốc gia TP.HCM và các đơn vị đã hỗ trợ cho nghiên cứu này.

TÀI LIỆU THAM KHẢO

- [1] T. P. Williams, ASCE, "Predicting Changes In Construction Cost Indexes Using Neural Networks," *Journal of Construction Engineering and Management*, vol. 120, no. 2, Jun. 1994.
- [2] T. M. S. Elhag and A. H. Boussabaine, "An Artificial Neural System For Cost Estimation Of Construction Projects," in *Proc. 14th Annual ARCOM Conf.*, 1998.
- [3] H. Adeli and M. Wu, "Regularization Neural Network For Construction Cost Estimation," *Journal of Construction Engineering and Management*, vol. 124, no. 1, Feb. 1998.
- [4] I. Siqueira, "Neural Network-Based Cost Estimating," Project Management Institute, 1999.
- [5] E. A. Lotfy and A. S. Mohamed, "Applying neural networks in case-based reasoning adaptation for cost assessment of steel buildings," *Int. J. Comput. Appl.*, vol. 24, no. 1, pp. 28-38, 2002.
- [6] S.-H. An, G.-H. Kim, and K.-I. Kang, "A case-based reasoning cost estimating model using experience by analytic hierarchy process," *Building and Environment*, vol. 42, pp. 2573-2579, Oct. 2004.
- [7] G.-H. Kim, S.-H. An, and K.-I. Kang, "Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning," *Building and Environment*, vol. 39, no. 10, pp. 1235-1242, Oct. 2004.
- [8] C. G. Wilmot and B. Mei, "Neural Network Modeling of Highway Construction Costs," *Journal of Construction Engineering and Management*, vol. 131, no. 7, Jul. 1, 2005.
- [9] M.-Y. Cheng, H.-C. Tsai, and E. Sudjono, "Engineering Applications of Artificial Intelligence," Elsevier Ltd., 2009.
- [10] V. M. Hồ, "Ứng dụng System Dynamics trong phân tích và dự báo chi phí xây dựng 86 nhà thép tiền chế," Thạc sĩ, Đại học Bách Khoa, ĐHQG-HCM, 2012.
- [11] C. J. Carter, T. M. Murray, and W. A. Thornton, "Cost-effective steel building design," 2000.
- [12] Văn phòng kế toán chính phủ Hoa Kỳ, "Hướng dẫn đánh giá và ước tính chi phí GAO, Thực tiễn tốt nhất để phát triển và quản lý chi phí chương trình vốn," GAO-09-35P, 2009.
- [13] ASTM E2516-11, "Phân loại tiêu chuẩn cho hệ thống phân loại ước tính chi phí."