

ÁP DỤNG MÔ HÌNH DỰ BÁO DEEP LEARNING TRÊN THỊ TRƯỜNG CHỨNG KHOÁN VIỆT NAM: THỰC TRẠNG VÀ GIẢI PHÁP

DEEP LEARNING MODELS IN STOCK PRICE PREDICTION: A STUDY IN VIETNAMESE STOCK MARKET

Đỗ Hồng Nhung^{1,*}, Hoàng Nhật Khánh¹,
Phạm Hoàng Vũ¹, Nguyễn Thị Duy Linh¹,
Bùi Thị Thúy Nga¹, Đặng Chí Bách¹

DOI: <http://doi.org/10.57001/huiv5804.2025.216>

TÓM TẮT

Trong bối cảnh thị trường chứng khoán Việt Nam có nhiều biến động phức tạp, việc dự báo chính xác giá cổ phiếu luôn là một thách thức lớn. Bài báo đã so sánh hiệu quả của các mô hình học sâu (Deep Learning) LSTM và BiLSTM khi áp dụng cho dữ liệu giá cổ phiếu VN-Index trong khoảng thời gian từ 01/01/2018 đến 22/10/2024. Các chỉ số RMSE, MSE và MAE để đánh giá lại mô hình cho thấy các mô hình đơn giản (chỉ sử dụng dữ liệu giá cổ phiếu của Vn-Index) có độ chính xác cao hơn so với các mô hình phức tạp (giá cổ phiếu VN-Index kết hợp thêm dữ liệu về giá vàng và dầu thô). Kết quả này có ý nghĩa thực tiễn, giúp các nhà đầu tư đưa ra các quyết định dựa trên những mô hình dự báo đơn giản nhưng có tính hiệu quả cao.

Từ khóa: Dự báo giá chứng khoán, Deep Learning, LSTM, BiLSTM.

ABSTRACT

Accurately predicting stock prices remains a considerable challenge, particularly within the complex and volatile environment of the Vietnamese stock market. This article investigates the predictive performance of Deep Learning models - namely LSTM and BiLSTM - applied to VN-Index stock data covering the period from January 1, 2018, to October 22, 2024. Using RMSE, MSE, and MAE as evaluation metrics, the analysis reveals that models relying solely on VN-Index stock prices outperform more complex models that incorporate additional variables such as gold and crude oil prices. These results underscore the practical value of employing simpler yet more effective forecasting models, offering investors a robust foundation for data-driven decision-making.

Keywords: Stock price prediction, Deep Learning, LSTM, BiLSTM.

¹Viện Ngân hàng Tài chính, Đại học Kinh tế Quốc dân

*Email: nhungdh@neu.edu.vn

Ngày nhận bài: 06/2/2025

Ngày nhận bài sửa sau phản biện: 20/4/2025

Ngày chấp nhận đăng: 28/6/2025

1. GIỚI THIỆU

Chủ đề về dự báo xu hướng giá chứng khoán luôn là bài toán thu hút sự quan tâm đông đảo của các nhà đầu tư và các nhà nghiên cứu. Tuy nhiên, việc dự báo chính xác xu hướng biến động của giá cổ phiếu thật sự là một thách thức khi giá chứa đựng nhiều thông tin và yếu tố nhiễu, ngoài ra còn ảnh hưởng bởi tâm lý của đám đông [1, 2]. Ngoài ra, giá cổ phiếu cũng bị ảnh hưởng bởi các

chính sách Chính phủ hay sự kiện địa chính trị trên thế giới. Với đặc điểm thị trường “cận biên”, việc dự báo xu hướng giá trên thị trường Việt Nam còn nhiều khó khăn, mang những rủi ro tiềm tàng cho các nhà đầu tư.

Trên thực tế, có nhiều nghiên cứu đã chỉ ra những dự đoán về thị trường chứng khoán bằng phương pháp truyền thống là các mô hình thống kê, sử dụng dữ liệu lịch sử để nhận diện mẫu và đưa ra dự đoán [3]. Tuy nhiên, nhờ

sự phát triển của công nghệ, phương pháp học máy (Machine Learning) đã ra đời và cho kết quả dự báo hiệu quả hơn so với phương pháp truyền thống [4, 5].

Mục tiêu của nghiên cứu là so sánh các mô hình một tính năng và đa tính năng của hai phương pháp học máy LSTM và BiLSTM trong dự đoán xu hướng giá chỉ số VN-Index tại thị trường Việt Nam. Nghiên cứu hướng đến đánh giá độ chính xác và hiệu quả của các mô hình này trong dự báo giá cổ phiếu, từ đó giúp các nhà đầu tư có cái nhìn chính xác hơn về xu hướng thị trường chứng khoán “cận biên” của Việt Nam.

2. CƠ SỞ LÝ LUẬN

2.1. Thị trường chứng khoán và các yếu tố ảnh hưởng

Định nghĩa thị trường chứng khoán tiến hoá theo thời gian, phát triển từ cơ sở thực tiễn giao dịch và các điều kiện lịch sử nhất định. Smith A. bắt đầu với các ý tưởng cơ sở về thị trường, cung - cầu và vai trò của vốn đã đặt nền tảng cho các quan điểm về thị trường khoán sau này [6]. Theo Downes J. và Goodman J. [7], thị trường chứng khoán đóng vai trò là nền tảng để các công ty huy động vốn và để các nhà đầu tư mua và bán cổ phiếu sở hữu. Một số quan điểm khác cho rằng thị trường chứng khoán không hoàn toàn đồng nhất với thị trường vốn mà chỉ giao dịch, mua bán các công cụ tài chính trung và dài hạn như trái phiếu và cổ phiếu. Nhưng tựu chung lại thị trường chứng khoán là nơi các tổ chức và cá nhân tham gia thực hiện các lệnh trao đổi, mua bán chứng khoán và các giấy tờ có giá theo những quy tắc đã được ấn định trước.

Thị trường chứng khoán có đặc thù biến động liên tục, nơi giá cả của các tài sản tài chính như cổ phiếu và trái phiếu thay đổi theo thời gian kèm theo các rủi ro mà các chủ thể tham gia phải đối mặt. Đồng thời, đây cũng là một hệ thống phức tạp, chịu ảnh hưởng từ nhiều các yếu tố mà trong đó tăng trưởng kinh tế, chính sách pháp lý, tâm lý xã hội và các sự kiện bất ngờ là điển hình.

2.2. Machine Learning và Deep Learning

Trong thời đại công nghệ số, học máy (Machine Learning - ML) và học sâu (Deep Learning - DL) đã trở thành lĩnh vực cốt lõi trong trí tuệ nhân tạo (Artificial Intelligence - AI), giúp giải quyết nhiều vấn đề phức tạp. Học máy, một nhánh của AI, phát triển các thuật toán học từ dữ liệu mà không cần lập trình trực tiếp, gồm: học có giám sát (Supervised Learning), học không giám sát (Unsupervised Learning) và học tăng cường (Reinforcement Learning). Học sâu, một nhánh của ML, sử dụng mạng nơ-ron nhân tạo để xử lý bài toán như nhận

diện hình ảnh, xử lý ngôn ngữ tự nhiên và phân tích dữ liệu lớn. Cụ thể, thuật toán Convolutional Neural Network (CNN), đặt nền tảng cho DL trong xử lý hình ảnh nhờ khả năng học đặc trưng qua nhiều lớp nơ-ron [8]. Thuật toán tối ưu, như Gradient Descent, được sử dụng để huấn luyện các mô hình này [9]. ML và DL hiện được ứng dụng rộng rãi, khẳng định vai trò quan trọng của chúng trong việc thúc đẩy sự phát triển công nghệ và giải quyết các bài toán thực tiễn.

2.3. Dự báo giá chứng khoán

Dự báo giá chứng khoán là một lĩnh vực nghiên cứu quan trọng trong tài chính, với các phương pháp tiếp cận đa dạng. Trước khi các kỹ thuật học máy (ML) và học sâu (DL) phát triển, các mô hình thống kê và kinh tế lượng truyền thống như ARIMA hay GARCH được sử dụng phổ biến để phân tích chuỗi thời gian tài chính. Tuy nhiên, các phương pháp này thường giả định rằng mối quan hệ trong dữ liệu quá khứ sẽ tiếp tục tồn tại trong tương lai - điều không phù hợp trong bối cảnh thị trường luôn biến động bởi các yếu tố kinh tế và chính trị [10]. Ngoài ra, chúng gặp khó khăn trong việc xử lý các quan hệ phi tuyến và thường kém hiệu quả trong dự báo dài hạn.

Sự phát triển của ML và DL đã mở ra hướng tiếp cận mới cho phân tích tài chính. Krauss C., Do X. A. và Huck N. [11] đã áp dụng các mô hình như Gradient-Boosted Trees và Random Forests để dự báo thành phần S&P 500. Các nghiên cứu sau đó như Khare K. và cộng sự [12] cho thấy các thuật toán Deep Learning mang lại độ chính xác cao trong dự báo giá cổ phiếu ngắn hạn. Đặc biệt, Mehtab S., Sen J. và Dasgupta S. [13] khẳng định các mô hình hồi quy học sâu, khi hiệu chỉnh phù hợp, có tiềm năng lớn trong việc hỗ trợ nhà đầu tư đưa ra quyết định.

Tại Việt Nam, việc ứng dụng các phương pháp học sâu trong dự báo thị trường chứng khoán mới chỉ bắt đầu thu hút sự quan tâm trong những năm gần đây. Một số nghiên cứu, điển hình như của Oanh Đ.L.K. và Châu N.T.M. [14], đã thử nghiệm các mô hình như LSTM và Temporal Convolutional Networks (TCN) để dự báo chỉ số VN-Index, cho thấy kết quả khả quan với độ chính xác cao, trong đó LSTM đạt hiệu suất vượt trội. Tuy nhiên, so với bối cảnh nghiên cứu quốc tế, số lượng nghiên cứu học thuật ứng dụng Deep Learning trong dự báo thị trường tài chính Việt Nam vẫn còn khiêm tốn.

Đáng chú ý, phần lớn các công trình hiện nay vẫn tập trung vào các mô hình truyền thống như ARIMA, GARCH hoặc các kỹ thuật học máy cơ bản (linear regression, decision tree), trong khi các mô hình học sâu hiện đại như LSTM hay BiLSTM - vốn đã chứng minh hiệu quả vượt trội

trong môi trường tài chính biến động - lại chưa được khai thác đầy đủ. Hơn nữa, chưa có nhiều nghiên cứu thực nghiệm so sánh hiệu năng giữa các mô hình DL khác nhau trên tập dữ liệu tài chính phức tạp và giàu tính nhiễu như ở thị trường Việt Nam, nơi chịu ảnh hưởng lớn từ các yếu tố kinh tế vĩ mô như chính sách tài khóa, tỷ giá hối đoái và tâm lý nhà đầu tư nước ngoài. Khoảng trống này đặt ra nhu cầu cấp thiết cho các nghiên cứu chuyên sâu hơn nhằm đánh giá hệ thống, so sánh hiệu quả các mô hình học sâu trong bối cảnh thị trường nhiều biến động Việt Nam.

3. DỮ LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

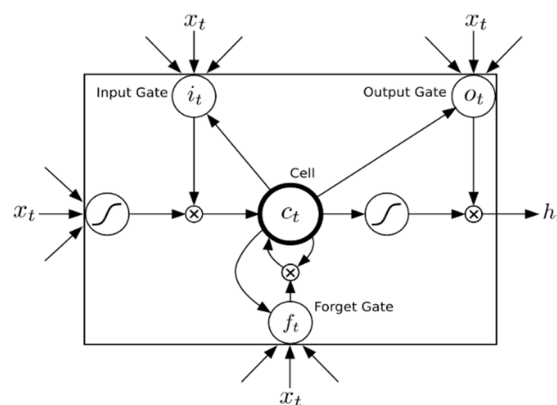
3.1. Dữ liệu

Bộ dữ liệu nghiên cứu sử dụng chỉ số VN-Index, một chỉ số giá có trọng số của các chứng khoán niêm yết trên Sở Giao dịch Chứng khoán TP.HCM (HOSE), kèm theo đó là chỉ số giá dầu thô và vàng được áp dụng trong mô hình đa đặc trưng. Chỉ số VN-Index được nhóm nghiên cứu chọn làm đối tượng dự báo trong nghiên cứu là hoàn toàn có cơ sở cả về lý luận lẫn thực tiễn. VN-Index là chỉ số thị trường chính của Sở Giao dịch Chứng khoán TP. Hồ Chí Minh (HOSE), phản ánh biến động tổng thể của toàn bộ các cổ phiếu niêm yết trên sàn HOSE và được xem là thước đo quan trọng thể hiện xu hướng chung của thị trường chứng khoán Việt Nam. Dù thị trường có mức độ phân hóa cao giữa các cổ phiếu thành phần, song thị trường chứng khoán Việt Nam đang phát triển và chịu ảnh hưởng lớn từ tâm lý đám đông cũng như chính sách điều hành vĩ mô. Ngoài ra, thị trường Việt Nam có đặc điểm là sự chi phối mạnh mẽ của nhà đầu tư cá nhân, dễ phản ứng với các yếu tố địa chính trị và thông tin kinh tế vĩ mô, làm cho chỉ số VN-Index trở thành một công cụ tổng hợp phản ánh kỳ vọng thị trường trong ngắn và trung hạn. Do đó, việc dự báo chỉ số này, thay vì dự báo các mã cổ phiếu riêng lẻ với độ nhiễu cao, sẽ giúp mô hình hóa xu hướng chung hiệu quả hơn và phục vụ tốt cho mục tiêu ra quyết định đầu tư, phân bổ tài sản và quản trị rủi ro thị trường. Bộ dữ liệu này gồm 1674 quan sát, được thu thập từ ngày 01/01/2018 đến ngày 22/10/2024 và lưu trữ trong tệp CSV. Việc phân tích nhờ dữ liệu VN-Index với các chỉ số giá dầu thô và vàng giúp ích đáng kể trong việc ra quyết định của nhà đầu tư bởi phần lớn sự vận hành của thị trường được phản ánh dựa trên chỉ số này.

3.2. Phương pháp nghiên cứu

Nghiên cứu sử dụng mạng bộ nhớ dài - ngắn hạn (LSTM) và mạng bộ nhớ dài - ngắn hạn 2 chiều (BiLSTM) để học sâu và đưa ra các kết quả dự báo. Mô hình LSTM thể hiện sự tiến bộ vượt trội trong cấu trúc của mạng thần

kinh tái phát (RNN), RNN [15], là một nhánh nhỏ trong DL, và được đề xuất để khắc phục vấn đề về biến mất độ dốc (vanishing gradient) - một trong những vấn đề của mạng nơ ron khi học sâu được nhắc tới trước đó trong nghiên cứu của Hochreiter S. và Schmidhuber J. [16]. Nhờ cải tiến này mà LSTM được sử dụng rộng rãi để giải quyết các bài toán có trình tự, đặc biệt là các bài toán liên quan đến chuỗi thời gian và dự báo. Trong mô hình này, dữ liệu đầu vào là chuỗi thời gian gồm các giá trị chỉ số VN-Index theo từng ngày. Tại cổng vào (input gate), các giá trị này được đưa vào mô hình, nơi thông tin mới được xử lý để xác định mức độ ảnh hưởng đến trạng thái hiện tại. Cổng quên (forget gate) sẽ xem xét trạng thái trước đó để xác định phần thông tin nào không còn cần thiết và loại bỏ chúng. Cổng đầu ra (output gate) kết hợp thông tin mới và trạng thái đã cập nhật để tạo ra đầu ra - chính là giá trị dự báo chỉ số VN-Index cho ngày tiếp theo. Quá trình này cho phép mô hình ghi nhớ hoặc quên thông tin một cách có chọn lọc, đảm bảo giữ lại những đặc điểm quan trọng trong chuỗi thời gian. Cấu trúc và cách thức hoạt động của mô hình LSTM được thể hiện trên hình 1.

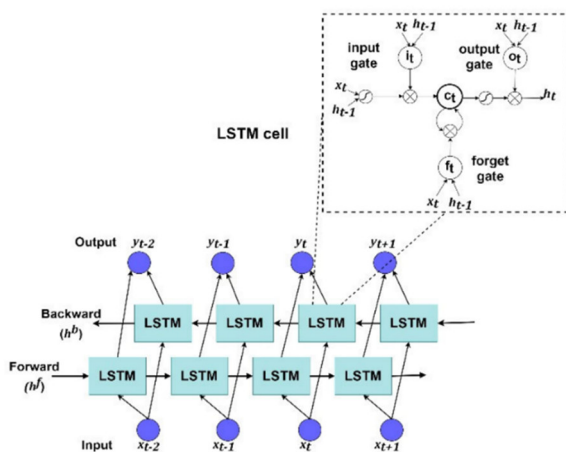


Hình 1. Cách thức hoạt động của các cổng trong mô hình LSTM (Nguồn: [17])

Điểm mới trong nghiên cứu này là không chỉ sử dụng mô hình LSTM truyền thống như nhiều nghiên cứu trước, mà còn áp dụng mô hình BiLSTM, một biến thể mạnh mẽ hơn, để khai thác đồng thời thông tin từ cả quá khứ và tương lai của chuỗi dữ liệu. Việc so sánh hai mô hình LSTM và BiLSTM trong dự báo chỉ số VN-Index, vốn có tính biến động và phụ thuộc thời gian cao, là một hướng tiếp cận ít được khai thác trong bối cảnh thị trường chứng khoán Việt Nam. Nghiên cứu này nhằm làm rõ liệu BiLSTM, với khả năng biểu diễn ngữ cảnh hai chiều, có mang lại cải thiện đáng kể trong dự báo tài chính hay không. Theo Schuster M. và Paliwal K.K. [18], BiLSTM là một biến thể của LSTM, gồm hai lớp độc lập: lớp tiến (forward) xử lý

chuỗi theo chiều thuận và lớp lùi (backward) xử lý theo chiều ngược. Nhờ cấu trúc này, BiLSTM có thể khai thác thông tin từ cả quá khứ và tương lai, giúp mô hình đưa ra biểu diễn toàn diện hơn tại mỗi thời điểm. Điều này đặc biệt hữu ích trong bài toán phân tích chuỗi thời gian tài chính, nơi các mối quan hệ giữa các ngày giao dịch có thể kéo dài cả về trước lẫn sau thời điểm hiện tại. Việc áp dụng BiLSTM để dự báo chỉ số VN-Index trong nghiên cứu này giúp mở rộng phạm vi ứng dụng của mô hình và cung cấp thêm bằng chứng thực nghiệm về hiệu quả của nó trong thị trường chứng khoán Việt Nam. Cấu trúc của mô hình BiLSTM được thể hiện trên hình 2.

Để đánh giá hiệu suất dự báo của mô hình, nghiên cứu áp dụng tập hợp các chỉ số sai số nhằm đo lường các khía cạnh khác nhau của độ chính xác và cung cấp thông tin định lượng về hiệu quả của mô hình trong dự đoán chỉ số VN-Index, nghiên cứu sử dụng các chỉ số Mean Absolute Error (MAE), Mean Squared Error (MSE) và Root Mean Squared Error (RMSE). MAE đo lường sai số trung bình tuyệt đối, giúp diễn giải độ lệch trung bình. MSE nhấn mạnh sai số lớn, phù hợp khi cần hạn chế các độ lệch đáng kể. RMSE, là căn bậc hai của MSE, trả về sai số cùng đơn vị với dữ liệu gốc, giúp diễn giải trực quan hơn. Các giá trị của các chỉ số này càng nhỏ, hiệu quả của mô hình càng được cải thiện.



Hình 2. Các bước vận hành của một mô hình BiLSTM (Nguồn: [19])

4. THỰC TRẠNG ÁP DỤNG PHƯƠNG PHÁP HỌC MÁY TRONG DỰ BÁO GIÁ CỔ PHIẾU TRÊN THỊ TRƯỜNG CHỨNG KHOÁN VIỆT NAM

4.1. Thực trạng thị trường chứng khoán Việt Nam

Thị trường chứng khoán Việt Nam là một trong ba trụ cột của thị trường tài chính với các phân khúc: cổ phiếu, trái phiếu và phái sinh. Giai đoạn 2011 - 2020, huy động vốn qua thị trường này đạt gần 2,9 triệu tỷ đồng,

gấp 10 lần so với 2000 - 2010, đóng góp bình quân 19,5% tổng mức đầu tư toàn xã hội [20]. Trong 4 năm trở lại đây, thị trường trải qua biến động, đặc biệt khi COVID-19 gây nên khủng hoảng y tế, suy thoái kinh tế và những căng thẳng gia tăng về địa chính trị. Chỉ số VN-Index giảm sâu, có thời điểm dưới 660 điểm, nhưng đã có sự phục hồi ấn tượng, đạt gần 1.100 điểm cuối 2020 và hơn 1.500 điểm đầu 2022. Đồng thời thị trường không ngừng mở rộng các ngành hàng giao dịch. Theo thống kê thị trường của Ủy ban Chứng khoán Nhà nước (SSC), đến cuối năm 2022, có tổng 2162 doanh nghiệp niêm yết và đăng ký giao dịch, tăng 5,92% vào năm 2023 và giảm trở lại khoảng 2100 doanh nghiệp vào năm 2024. Những áp lực đến từ tình hình kinh tế trong và ngoài nước, đặc biệt khi các thị trường lớn như Mỹ hay EU đều suy giảm, diễn biến của thị trường Việt Nam bước vào giai đoạn điều chỉnh mạnh [21]. Cuối 2022, chỉ số giảm gần 33% so với mức đỉnh đầu năm, giao dịch bình quân sụt giảm. Sau năm 2022 vẫn tiếp tục xảy ra biến động nhưng xu hướng nhìn chung là hồi phục. Dưới tác động của suy thoái kinh tế, địa chính trị leo thang và lợi nhuận doanh nghiệp giảm, thị trường vận động phân hoá rõ rệt qua VN-Index. Chỉ số này đạt đỉnh trong năm 2023 ở mức 1.245,5 điểm vào tháng 9, thanh khoản thị trường cải thiện đáng kể. Nhưng từ cuối quý 3 đến giữa quý 4/2023, do Ngân hàng Nhà nước hút tiền qua kênh tín phiếu nhằm ổn định tỷ giá, khối ngoại tiếp tục bán ròng, và nền kinh tế phục hồi chậm hơn kỳ vọng [22], VN-Index bước vào xu hướng giảm, lùi về gần mốc 1.000 điểm vào cuối tháng 10. Cuối năm 2023 đầu 2024, thị trường ghi nhận diễn biến khởi sắc. Đáng chú ý, vào phiên giao dịch ngày 25/3/2024, VN-Index tăng lên mức 1.267,9 điểm, trong khi giá trị khớp lệnh trên sàn HOSE đạt 27,2 nghìn tỷ đồng, phản ánh sự sôi động trở lại của dòng tiền trên thị trường.

Dù các hoạt động giao dịch trong nước đã được công khai minh bạch; việc quản lý, giám sát cũng bài bản hoá, thị trường chứng khoán Việt Nam vẫn còn hạn chế tồn đọng. Thứ nhất, quy mô vốn hóa so với GDP thấp, tập trung vào một số mã lớn (bluechips), gây mất cân đối và hạn chế đa dạng hóa danh mục. Thứ hai, vẫn còn vấn đề trong khâu quản lý, khi các vụ thao túng giá và giao dịch nội gián vẫn xảy ra [23], hệ thống giám sát xử lý vi phạm chưa đủ mạnh để ngăn chặn kịp thời các hành vi sai trái. Thứ ba, hệ thống công nghệ còn lạc hậu, tình trạng nghẽn lệnh trên sàn giao dịch gây bất tiện cho nhà đầu tư. Việc ứng dụng công nghệ mới như AI và ML vào hoạt động giao dịch và quản lý vẫn chưa phổ biến.

4.2. Thực trạng dự báo giá cổ phiếu trên TTCK Việt Nam bằng phương pháp học máy

Với sự phát triển của thị trường chứng khoán và nhu cầu ngày càng tăng từ nhà đầu tư, dự báo giá cổ phiếu đã trở thành yếu tố quan trọng trong chiến lược đầu tư. Các phương pháp truyền thống như phân tích kỹ thuật và phân tích cơ bản vẫn được sử dụng rộng rãi. Tuy nhiên, theo Agrawal M., Khan A.U., Shukla P.K. [24], hiệu quả của các phương pháp này không còn nhất quán. Phân tích kỹ thuật, dù phổ biến, ngày càng bị nghi ngờ về tính tin cậy, trong khi phân tích cơ bản cũng không còn phù hợp với bối cảnh thị trường biến động mạnh mẽ như Việt Nam [25].

Trước những hạn chế của các phương pháp truyền thống, các mô hình hiện đại dựa trên AI như ML và DL đang thu hút sự chú ý lớn từ giới đầu tư. Các nghiên cứu cho thấy, ML và DL có khả năng dự đoán hiệu quả hơn các phương pháp truyền thống [26, 27]. Tuy nhiên, thực tế việc ứng dụng các mô hình này tại Việt Nam vẫn còn rất hạn chế. Hầu hết nhà đầu tư cá nhân thiếu công cụ và kiến thức để triển khai, trong khi các tổ chức lớn như công ty chứng khoán và quỹ đầu tư mới chỉ thử nghiệm tích hợp ML vào hệ thống phân tích tài chính.

Các nghiên cứu hiện nay về dự báo giá cổ phiếu tại Việt Nam chủ yếu tập trung vào các chỉ số lớn như VN-Index, VN30 hoặc các cổ phiếu blue-chip có tính thanh khoản cao như VIC, VHM, VNM, HPG và BID. Chưa có nghiên cứu nào được thực hiện ở quy mô toàn bộ thị trường hoặc trên tất cả các sàn HOSE, HNX và UpCOM. Phần lớn các nghiên cứu sử dụng dữ liệu ngắn hạn (3 - 5 năm) do hạn chế về dữ liệu lịch sử và chỉ tập trung vào yếu tố giá cổ phiếu, khối lượng giao dịch. Một số yếu tố kinh tế vĩ mô như lãi suất, tỷ giá hối đoái đã được đưa vào các mô hình, nhưng vẫn còn hạn chế.

Các mô hình dự báo phổ biến bao gồm phân tích kỹ thuật: sử dụng các chỉ báo như trung bình động (MA) và chỉ số sức mạnh tương đối (RSI); phân tích cơ bản: dựa trên các chỉ số như P/E, EPS, doanh thu, lợi nhuận, mô hình thống kê: ARIMA, GARCH và hồi quy tuyến tính; ML: Random Forest, Gradient Boosting Machines (GBM), Support Vector Machines (SVM); DL: Long Short-Term Memory (LSTM) và Convolutional Neural Networks (CNN) [28-31].

Ngoài ra, các nghiên cứu cũng đã thử nghiệm các mô hình lai kết hợp nhiều phương pháp, đạt được hiệu quả dự báo cao hơn [27, 32, 33]. Nhìn chung, mô hình ML và DL cho độ chính xác cao hơn các phương pháp truyền thống, đặc biệt trong dự báo ngắn hạn, với các chỉ số MAE, MSE, và RMSE thường được dùng để đánh giá hiệu quả.

Tuy có nhiều tiềm năng, việc ứng dụng ML và DL vào dự báo giá cổ phiếu tại Việt Nam vẫn đang ở giai đoạn đầu. Các công ty Fintech như FireAnt, SSI và VNDirect đã tích hợp ML vào nền tảng phân tích tài chính, cung cấp dự báo tự động và cảnh báo đầu tư. Tuy nhiên, quá trình này đòi hỏi thời gian để xây dựng đội ngũ chuyên môn, tối ưu hóa hệ thống, và quan trọng nhất là tạo dựng niềm tin từ các nhà đầu tư [34].

Nhìn chung, dự báo giá cổ phiếu tại Việt Nam đang chuyển đổi từ phương pháp truyền thống sang các công cụ hiện đại dựa trên AI và ML. Mặc dù còn nhiều khó khăn về dữ liệu, công nghệ và nguồn nhân lực, đây là xu hướng tất yếu để đáp ứng nhu cầu của một thị trường ngày càng phức tạp và cạnh tranh. Trong tương lai, với sự đầu tư mạnh mẽ từ các tổ chức, ML và DL sẽ góp phần nâng cao hiệu quả dự báo, mở ra cơ hội lớn cho cả nhà đầu tư và thị trường chứng khoán Việt Nam.

5. THẢO LUẬN KẾT QUẢ NGHIÊN CỨU VÀ ĐÁNH GIÁ THỰC TRẠNG DỰ BÁO GIÁ CỔ PHIẾU TRÊN TTCK VIỆT NAM BẰNG PHƯƠNG PHÁP HỌC MÁY

5.1. Thảo luận kết quả nghiên cứu

Nghiên cứu so sánh độ chính xác dự báo giá của 4 mô hình trên thị trường chứng khoán Việt Nam, hình 3 ÷ 6 minh họa sự so sánh giữa giá đóng cửa thực tế và dự báo của các mô hình.



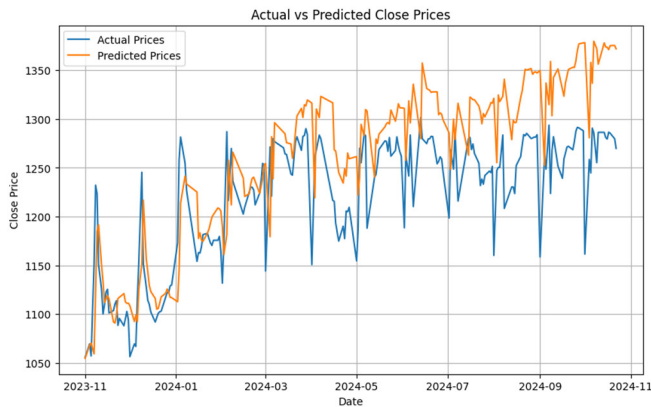
Hình 3. So sánh giá thực tế và giá đóng cửa của mô hình LSTM đơn



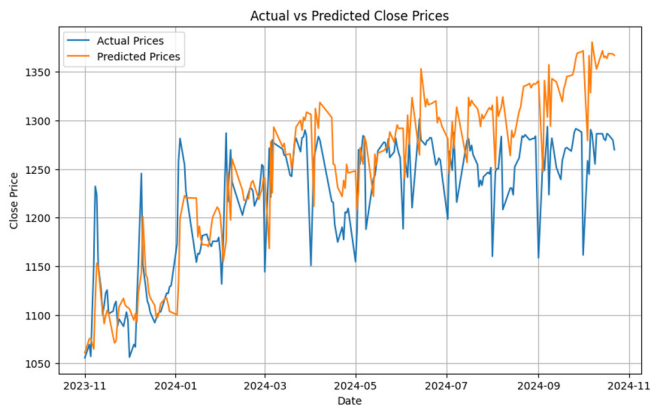
Hình 4. So sánh giá thực tế và giá đóng cửa của mô hình BiLSTM đơn

Dựa trên hình 3 và 4 cho thấy, các mô hình LSTM và BiLSTM đơn bám sát giá thực tế, chỉ sai lệch nhỏ, cho thấy

khả năng nắm bắt tốt xu hướng giá. Trong khi đó, các mô hình đa đặc trưng (bao gồm giá vàng và dầu thô) lại có nhiều dao động, dự đoán thường chệch khỏi giá thực tế như thể hiện trên hình 5 và 6.



Hình 5. So sánh giá thực tế và giá đóng cửa của mô hình LSTM đa đặc trưng (Nguồn: Nghiên cứu của nhóm tác giả)



Hình 6. So sánh giá thực tế và giá đóng cửa của mô hình BiLSTM đa đặc trưng (Nguồn: Nghiên cứu của nhóm tác giả)

Ba chỉ số RMSE, MSE và MAE được dùng để so sánh hiệu suất dự đoán của bốn mô hình, với giá trị nhỏ hơn biểu thị dự đoán tốt hơn [35]. Kết quả được trình bày trong bảng 1.

Bảng 1. Giá trị của RMSE, MSE, MAE trung bình của 4 mô hình

Chỉ số	RMSE trung bình	MSE trung bình	MAE trung bình
Mô hình LSTM đơn	18,2160	352,2103	12,5664
Mô hình BiLSTM đơn	19,8161	418,0885	14,8419
Mô hình LSTM đa đặc trưng	90,3577	9033,1290	65,1513
Mô hình BiLSTM đa đặc trưng	83,7336	7630,4990	60,4518

Nguồn: Nghiên cứu của nhóm tác giả

Theo bảng 1, các mô hình LSTM và BiLSTM đơn đặc trưng có sai số nhỏ hơn, vượt trội về độ chính xác so với các mô hình đa đặc trưng. LSTM cho thấy ưu thế hơn

BiLSTM khi xử lý dữ liệu đơn giản, trong khi BiLSTM tỏ ra hiệu quả hơn với dữ liệu phức tạp. Tuy nhiên, tổng thể, BiLSTM có sai số cao hơn. Điều này cho thấy cách tiếp cận hai chiều (bidirectional) không mang lại nhiều lợi ích trong thị trường biến động như Việt Nam. Ngược lại, LSTM với cách tiếp cận đơn chiều tỏ ra phù hợp hơn.

5.2. Đánh giá thực trạng dự báo giá cổ phiếu trên TTCK Việt Nam bằng phương pháp học máy

Thị trường chứng khoán Việt Nam đã có những bước phát triển mạnh mẽ, trở thành một kênh đầu tư hấp dẫn tại khu vực Đông Á và Đông Nam Á. Tuy nhiên, theo xếp hạng của MSCI và FTSE Russell [36], thị trường Việt Nam vẫn được xếp vào nhóm "cận biên" (frontier market), phản ánh tính non trẻ và sự bất ổn cao. Điều này đặt ra nhiều thách thức cho việc dự báo giá cổ phiếu trong một bối cảnh thị trường đầy biến động.

Hai phương pháp phổ biến nhất hiện nay trong dự báo giá cổ phiếu là phân tích kỹ thuật và phân tích cơ bản. Dù được sử dụng rộng rãi, nhưng các nghiên cứu như của Long N.H., Paientko T. [37] cho thấy hiệu quả của những phương pháp này không còn nhất quán trong bối cảnh thị trường biến động mạnh. Điều này đặt ra nhu cầu cấp thiết về việc áp dụng các mô hình hiện đại dựa trên AI như ML. Tuy nhiên, tại Việt Nam, việc triển khai các mô hình này vẫn còn rất hạn chế. Đa số nhà đầu tư cá nhân thiếu kiến thức và công cụ để áp dụng, trong khi các tổ chức lớn như công ty chứng khoán và quỹ đầu tư mới chỉ bắt đầu thử nghiệm ML trong hệ thống phân tích tài chính.

Sự chuyển đổi từ phương pháp truyền thống sang công cụ hiện đại không chỉ đòi hỏi thời gian mà còn gặp rào cản lớn về nguồn lực và niềm tin từ nhà đầu tư. Để triển khai hiệu quả AI và ML, thị trường cần dữ liệu chất lượng cao và đội ngũ nhân sự có chuyên môn sâu. Tuy nhiên, đây vẫn là thách thức lớn tại Việt Nam, nơi thị trường chịu ảnh hưởng nặng nề bởi tâm lý bầy đàn và sự biến động khó lường.

Số lượng nghiên cứu sử dụng ML và DL để dự báo giá cổ phiếu tại Việt Nam đã tăng lên đáng kể trong những năm gần đây, nhưng số lượng bài được công bố trên các tạp chí quốc tế uy tín vẫn còn hạn chế. Hầu hết các nghiên cứu sử dụng dữ liệu ngắn hạn (3 - 5 năm) do thiếu dữ liệu lịch sử dài hạn. Các yếu tố kinh tế vĩ mô như lãi suất, tỷ giá hối đoái cũng chỉ mới bắt đầu được đưa vào các mô hình dự báo. Tuy vậy, các kết quả thực nghiệm cho thấy ML và DL vượt trội hơn các phương pháp truyền thống, đặc biệt trong dự báo ngắn hạn.

Nhìn chung, dự báo giá cổ phiếu tại Việt Nam đang chuyển đổi từ phương pháp truyền thống sang các công

cụ hiện đại dựa trên AI và ML. Dù còn nhiều khó khăn, đây là xu hướng tất yếu để đáp ứng nhu cầu của một thị trường ngày càng phức tạp và cạnh tranh. Với sự phát triển công nghệ và đầu tư từ các tổ chức, việc dự báo giá cổ phiếu sẽ dần hiệu quả hơn, tạo cơ hội lớn cho cả nhà đầu tư và thị trường chứng khoán Việt Nam.

6. KẾT LUẬN, KHUYẾN NGHỊ VÀ GIẢI PHÁP

Qua việc phân tích, các kết quả được kết luận trong bối cảnh thị trường chứng khoán Việt Nam, các mô hình đơn giản hơn như mô hình đơn tính năng LSTM và mô hình đơn tính năng BiLSTM hoạt động tốt hơn các mô hình đa tính năng tương ứng. Điều này nhấn mạnh rằng các cấu hình đơn giản hơn có thể mang lại độ chính xác tốt hơn trên các thị trường có đặc điểm biến động cao. Nghiên cứu này nhận thấy rằng sự khác biệt về hiệu suất giữa LSTM một tính năng và BiLSTM là rất nhỏ, cho thấy cả hai đều có độ chính xác cao, kết quả này khác với nghiên cứu của Siami-Namini S., Tavakoli N., Namin A.S. [3] và Han C., Fu X. [38] khi chỉ ra BiLSTM vượt trội hơn do khả năng đọc hai chiều. Tuy nhiên, với cách tiếp cận trực tiếp và đơn hướng, LSTM phù hợp hơn với thị trường Việt Nam. Nghiên cứu này nhấn mạnh giá trị của việc kết hợp các yếu tố kinh tế vĩ mô như giá vàng và giá dầu vào mô hình dự đoán cổ phiếu.

Sau quá trình thực hiện nghiên cứu, phân tích và so sánh hiệu quả dự báo xu hướng giá chỉ số Vn-Index giữa các mô hình LSTM và BiLSTM với các cấu hình một tính năng và đa tính năng, nhóm tác giả đã thu được những kết quả quan trọng, góp phần làm rõ tiềm năng ứng dụng của các mô hình học sâu trong lĩnh vực tài chính. Trên cơ sở những phát hiện từ nghiên cứu, nhóm tác giả xin đưa ra một số khuyến nghị và đề xuất nhằm nâng cao hiệu quả dự báo cũng như hỗ trợ các nhà đầu tư, nhà nghiên cứu và tổ chức tài chính trong quá trình ra quyết định.

Đối với các nhà nghiên cứu và các tổ chức tài chính, cần tập trung vào việc tối ưu hoá các mô hình dự báo đơn giản chẳng hạn như LSTM đơn tính năng, thay vì triển khai các mô hình đa tính năng phức tạp do tính phù hợp của mô hình này đối với thị trường Việt Nam. Mặc dù việc bổ sung các yếu tố kinh tế vĩ mô như giá vàng và dầu thô vào mô hình có thể mang lại thông tin hữu ích, nhưng cần phải đánh giá cẩn thận và chính xác sự tác động của các yếu tố này đối với dự báo giá cổ phiếu trên thị trường Việt Nam. Mô hình đơn giản có thể mang lại hiệu quả cao hơn trong điều kiện thị trường Việt Nam với các đặc trưng biến động mạnh, dễ bị tác động bởi tâm lý nhà đầu tư và các yếu tố nội tại đặc thù. Hơn nữa, việc triển khai các mô hình đơn giản không chỉ

giúp giảm thiểu chi phí tính toán và thời gian huấn luyện mà còn dễ dàng hơn trong việc kiểm định, hiệu chỉnh và giải thích.

Đối với nhà đầu tư tại thị trường Việt Nam, việc ưu tiên sử dụng các mô hình dự báo đơn giản như LSTM đơn tính năng là lựa chọn hợp lý, thay vì áp dụng các mô hình học máy phức tạp vốn đòi hỏi nhiều dữ liệu và nguồn lực tính toán. Trong bối cảnh thị trường chứng khoán Việt Nam còn mang tính "cận biên", với đặc điểm biến động nhanh và chịu ảnh hưởng lớn từ tâm lý và tin tức, mô hình LSTM đơn tính năng đã thể hiện khả năng dự báo xu hướng giá VN-Index một cách hiệu quả và ổn định hơn. Việc sử dụng mô hình này không chỉ giúp nhà đầu tư giảm thiểu sai số trong dự báo, mà còn tạo điều kiện dễ dàng hơn trong việc triển khai thực tế, đặc biệt với các nhà đầu tư cá nhân không chuyên sâu về công nghệ.

Đối với các công ty chứng khoán và quỹ đầu tư, việc ứng dụng các mô hình học máy đơn giản như LSTM vào quá trình dự báo và ra quyết định đầu tư có thể mang lại lợi thế cạnh tranh đáng kể, đặc biệt trong việc nắm bắt xu hướng thị trường một cách nhanh chóng và chính xác. Mặc dù LSTM là mô hình đơn giản hơn so với nhiều kiến trúc học sâu khác, nhưng để triển khai hiệu quả trong môi trường thực tiễn, các tổ chức vẫn cần đảm bảo nhiều yếu tố nền tảng. Trước hết, chất lượng dữ liệu đầu vào đóng vai trò then chốt trong việc cải thiện độ chính xác của mô hình, đòi hỏi một hệ thống thu thập, xử lý và lưu trữ dữ liệu chuyên nghiệp. Bên cạnh đó, nguồn nhân lực có trình độ cao về trí tuệ nhân tạo, học máy và phân tích định lượng là yếu tố không thể thiếu để thiết kế, tối ưu và kiểm soát các mô hình dự báo. Cuối cùng, đầu tư vào cơ sở hạ tầng công nghệ là điều kiện thiết yếu giúp đảm bảo mô hình hoạt động ổn định, nhanh chóng và có khả năng mở rộng trong dài hạn. Do đó, để tận dụng tối đa tiềm năng của các mô hình học máy, các tổ chức tài chính cần có chiến lược phát triển đồng bộ cả về công nghệ, con người lẫn dữ liệu.

Nhìn chung, việc áp dụng các mô hình Deep Learning cho dự báo giá cổ phiếu VN-Index có thể mang lại hiệu quả cao hơn trong bối cảnh thị trường chứng khoán Việt Nam đầy biến động. Tuy nhiên, việc lựa chọn mô hình dự đoán phù hợp cũng như chuẩn bị cơ sở hạ tầng công nghệ và nhân sự là vô cùng quan trọng để tối ưu hoá các mô hình này. Mặc dù còn nhiều thách thức, nhưng việc áp dụng công nghệ AI và học máy sẽ góp phần quan trọng vào việc cải thiện quá trình ra quyết định đầu tư và phát triển bền vững thị trường chứng khoán Việt Nam.

LỜI CẢM ƠN

Nghiên cứu này được tài trợ bởi Đại học Kinh tế Quốc dân, Bộ Giáo dục và Đào tạo, Hà Nội, Việt Nam [Số tài trợ: B2024.KHA.08].

TÀI LIỆU THAM KHẢO

- [1]. Chen Y., Hao Y., "A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction," *Expert Systems with Applications*, 80, 340-355, 2017.
- [2]. Chung H., Shin K.S., "Genetic algorithm-optimized multi-channel convolutional neural network for stock market prediction," *Neural Computing and Applications*, 32, 7897-7914, 2020.
- [3]. Siami-Namini S., Tavakoli N., Namin A. S., "The Performance of LSTM and BiLSTM in Forecasting Time Series," in *2019 IEEE International Conference on Big Data, Los Angeles, CA, USA*, 2019.
- [4]. Mehtab S., Sen J., Dasgupta S., "Robust Analysis of Stock Price Time Series Using CNN and LSTM-Based Deep Learning Models," in *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Tamil Nadu, India, 2020.
- [5]. Qian F., Chen X., "Stock Prediction Based on LSTM under Different Stability," in *2019 IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*, Chengdu, China, 2019.
- [6]. Smith A., *An Inquiry into the Nature and Causes of the Wealth of Nations*. University of Chicago Press, USA, 1776.
- [7]. Downes J., Goodman J., *Dictionary of Finance and Investment Terms*. Barron's Educational Series, USA, 1985.
- [8]. Lecun Y., Bottou L., Bengio Y., Haffner P., "Gradient-based learning applied to document recognition," in *Proceedings of the IEEE*, 86 (11), 2278-2324, 1998.
- [9]. Rumelhart D., Hinton G., Williams R., "Learning representations by back-propagating errors," *Nature*, 323, 533-536, 1986.
- [10]. Goodfellow I., Bengio Y., Courville A., *Deep Learning*, Barron's Educational Series. New York, USA, 2015.
- [11]. Krauss C., Do X.A., Huck N., "Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500," *European Journal of Operational Research*, 259 (2), 689-702, 2017.
- [12]. Khare K., Darekar O., Gupta P., Attar V.Z., "Short term stock price prediction using deep learning," in *2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, Bangalore, India, 2017.
- [13]. Mehtab S., Sen J., Dasgupta S., "Robust Analysis of Stock Price Time Series Using CNN and LSTM-Based Deep Learning Models," in *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, India, 2020.
- [14]. Oanh Đ.L.K., Châu N.T.M., "Ảnh hưởng của bất định thế giới đến rủi ro tín dụng của hệ thống ngân hàng: Vai trò của vốn và tập trung," *Tạp chí Kinh tế và Phát triển*, 320(2), 63-71, 2024.
- [15]. Goodfellow I., Bengio Y., Courville A., *Deep Learning*, Barron's Educational Series. New York, USA, 2015.
- [16]. Hochreiter S., Schmidhuber J., "Long Short-Term Memory," *Neural Computation*, 9 (8), 1735-1780, 1997.
- [17]. Graves A., Mohamed A., Hinton G., "Speech recognition with deep recurrent neural networks," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, BC, Canada, 2013.
- [18]. Schuster M., Paliwal K.K., "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing*, 45 (11), 2673-2681, 1997.
- [19]. Yousaf K., Nawaz T., "A Deep Learning-Based Approach for Inappropriate Content Detection and Classification of YouTube Videos," *IEEE Access*, 10, 16283-16298, 2022.
- [20]. Phạm Bảo Anh, "Thực trạng và giải pháp phát triển thị trường chứng khoán Việt Nam," *Tạp chí tài chính Online*, 2022. <https://tapchitaichinh.vn/thuc-trang-va-giai-phap-phat-trien-thi-truong-chung-khoan-viet-nam.html>.
- [21]. Thái Duy, "Thị trường chứng khoán 2022 biến động mạnh, vốn ngoại là điểm sáng," *Thời báo tài chính Việt Nam*, 2022. <https://thoibaotaichinhvietnam.vn/thi-truong-chung-khoan-2022-bien-dong-manh-von-ngoai-la-diem-sang-119454.html>.
- [22]. Hiếu Phương, "Tổng quan thị trường chứng khoán Việt Nam năm 2023 và triển vọng tăng trưởng," *Kinh tế và dự báo*, 2024. <https://kinhtevadubao.vn/tong-quan-thi-truong-chung-khoan-viet-nam-nam-2023-va-trien-vong-tang-truong-28808.html>.
- [23]. Tạ Thanh Bình, "Thị trường chứng khoán Việt Nam năm 2023 và triển vọng 2024," *Tạp chí kinh tế Việt Nam*, 2024. <https://vneconomy.vn/thi-truong-chung-khoan-viet-nam-nam-2023-va-trien-vong-2024.htm>.
- [24]. Agrawal M., Khan A.U., Shukla P.K., "Stock Price Prediction using Technical Indicators: A Predictive Model using Optimal Deep Learning," *International Journal of Recent Technology and Engineering (IJRTE)*, 8 (2), 2297-2305, 2019.
- [25]. Cao K., You H., "Fundamental Analysis via Machine Learning," *Financial Analysts Journal*, 80 (2), 74-98, 2024.
- [26]. Pham H.V., Trinh T.D., Tieu K.M., Pham H.U., Pham T.B., "Stock-Price Forecasting Based on XGBoost and LSTM," *Computer Systems Science and Engineering*, 40 (1), 237-246, 2022.
- [27]. Tran P., Pham T.K.A., Phan H.T., Nguyen V.C., "Applying machine learning algorithms to predict the stock price trend in the stock market - The case of Vietnam," *Humanities and Social Sciences Communications*, 11. 393, 2024.
- [28]. Zhou H., Maneesoonthorn W. O., Chen X. B., "The Predictive Ability of Quarterly Financial Statements," *International Journal of Financial Studies*, 9 (3), 50, 2021.

[29]. Abdul-Rahim R., Khalid A., Karim Z. A., Rashid M., "Exploring the Driving Forces of Stock-Cryptocurrency Comovements during COVID-19 Pandemic: An Analysis Using Wavelet Coherence and Seemingly Unrelated Regression," *Mathematics*, 10, 2116, 2022.

[30]. Tarawneh A., Abdul-Rahman A., Mohd Amin S. I., Ghazali M. F., "A Systematic Review of Fintech and Banking Profitability," *International Journal of Financial Studies*, 12 (1), 3, 2024.

[31]. Kelany O., Aly S., Ismail M. A., "Deep Learning Model for Financial Time Series Prediction," in *2020 14th International Conference on Innovations in Information Technology (IIT)*, Al Ain, UAE, 2020.

[32]. Trần Trung Kiên, Bành Trí Thành, Nguyễn Hoàng Tú Anh, "Dự đoán giá cổ phiếu trên thị trường chứng khoán Việt Nam bằng phương pháp lai GA-SVR," *Các công trình nghiên cứu, phát triển và ứng dụng công nghệ thông tin và truyền thông*, 1 (27), 12-22, 2012.

[33]. Trần Đăng Tuyên, "Đánh giá hiệu suất mô hình phức hợp LSTM-GRU: nghiên cứu điển hình về dự báo chỉ số đo lường xu hướng biến động giá cổ phiếu trên sàn giao dịch chứng khoán Hồ Chí Minh," *Tạp chí Khoa học Đại học Cần Thơ*, 60 (1), 235-249, 2024.

[34]. Vân Phong, "Doanh nghiệp chứng khoán tăng tốc đầu tư ứng dụng công nghệ," *Kinh tế Sài Gòn*, 2024. <https://thesaigontimes.vn/doanh-nghiep-chung-khoan-tang-toc-dau-tu-ung-dung-cong-nghe/>.

[35]. Jierula A., Wang S., OH T.M., Wang P., "Study on Accuracy Metrics for Evaluating the Predictions of Damage Locations in Deep Piles Using Artificial Neural Networks with Acoustic Emission Data," *Applied Sciences*, 11 (5) 2314, 2021.

[36]. Vietnam News, *Vietnam eyes emerging stock market status by 2025*, 2025. <https://vietnamnews.vn/economy/1595221/viet-nam-eyes-emerging-stock-market-status-by-2025.html>

[37]. Nguyen H.L., Paientko T., "Forecasting the VN30 Index: Insights into Vietnam's Stock Market Trends," *Strategies in Accounting and Management*, 4 (5), 2024.

[38]. Han C., Fu X., "Challenge and Opportunity: Deep Learning-Based Stock Price Prediction by Using Bi-Directional LSTM Model," *Frontiers in Business, Economics and Management (FBEM)*, 8 (2), 51-54, 2023.

AUTHORS INFORMATION

**Do Hong Nhung, Hoang Nhat Khanh, Pham Hoang Vu,
Nguyen Thi Duy Linh, Bui Thi Thuy Nga, Dang Chi Bach**

School of Banking and Finance, National Economics University, Vietnam