

HỆ THỐNG CẢNH BÁO NGẬP LỤT ĐÔ THỊ DỰA TRÊN PHÂN TÍCH BIG DATA VÀ THUẬT TOÁN HỌC SÂU. ỨNG DỤNG THÍ ĐIỂM CHO KHU ĐÔ THỊ TRUNG TÂM THÀNH PHỐ ĐÀ NẴNG

Nguyễn Đức Diện, Lê Văn Nghị, Mai Thị Ngân Anh

Phòng Thí nghiệm trọng điểm Quốc gia về Động lực học sông biển

Tóm tắt: Ngập lụt đô thị là thách thức toàn cầu, đòi hỏi các giải pháp dự báo tốc độ cao mà các mô hình thủy lực truyền thống khó đáp ứng. Bài báo này đề xuất hệ thống cảnh báo ngập lụt thời gian thực, ứng dụng thí điểm tại trung tâm TP. Đà Nẵng, dựa trên kiến trúc dữ liệu lai giữa MongoDB (xử lý dữ liệu IoT chuỗi thời gian) và PostGIS (quản lý dữ liệu địa hình DEM và mạng lưới cống).

Mục tiêu chính của nghiên cứu là việc xây dựng một mô hình thay thế cho mô hình thủy lực SWMM, kết hợp giữa LSTM và Graph Neural Network - GNN. Mô hình lai ghép này được huấn luyện trên dữ liệu mô phỏng từ SWMM, cho phép dự báo ngập lụt với độ chính xác tương đương nhưng thời gian tính toán giảm đáng kể, hỗ trợ đắc lực cho việc ra quyết định và đánh giá rủi ro tức thời.

Từ khóa: Big Data, ngập lụt đô thị, mô hình học sâu, LSTM, GNN, MongoDB, PostGIS, DEM, IoT, Thành phố Đà Nẵng.

Summary: Urban flooding is a critical global challenge necessitating rapid forecasting solutions that exceed the capabilities of hydraulic models. This study presents a real-time flood warning system piloted in central Da Nang City, leveraging a hybrid data architecture that combines MongoDB for high-velocity IoT time-series data and PostGIS for spatial data management, including DEMs and sewer infrastructure.

The core objective is to construct a surrogate model for the SWMM hydraulic engine by coupling LSTM with Graph Neural Networks (GNN). This hybrid model, trained on SWMM simulation datasets, achieves flood forecasting with accuracy comparable to physical models but with significantly reduced computational latency, providing robust support for immediate risk assessment and decision-making.

1. GIỚI THIỆU

Thành phố Đà Nẵng, với vị trí địa lý ven biển và địa hình dốc, đang phải đối mặt với nguy cơ ngập lụt ngày càng gia tăng do tác động kép của đô thị hóa nhanh và các hiện tượng thời tiết cực đoan [1]. Trận mưa lịch sử tháng 10 năm 2022 đã cho thấy những lỗ hổng lớn trong khả năng dự báo và ứng phó, đặc biệt tại các khu vực trung tâm như quận Hải Châu và Thanh Khê.

Các hệ thống cảnh báo hiện tại của Đà Nẵng

chủ yếu dựa trên quan trắc tại chỗ và các mô hình thủy văn-thủy lực (như MIKE, SWMM) [2]. Các mô hình này, dù chi tiết, nhưng đòi hỏi thời gian tính toán lớn, khó có thể đáp ứng yêu cầu cảnh báo "tức thời" (real-time).

Trong bối cảnh đó, các công nghệ của Cách mạng Công nghiệp 4.0 như Internet vạn vật (IoT), Big Data và Trí tuệ Nhân tạo (AI) mở ra một hướng tiếp cận mới. Khả năng thu thập và xử lý đồng thời lượng lớn dữ liệu từ các cảm biến [3] cho phép xây dựng các mô hình dự báo dựa trên dữ liệu (data-driven). Các mô hình học sâu, đặc biệt là như LSTM và GNN đã chứng minh khả năng vượt trội trong việc phân tích và dự báo chuỗi thời gian và mạng

Ngày nhận bài: 21/11/2025

Ngày thông qua phản biện: 12/12/2025

Ngày duyệt đăng: 30/01/2026

lưới có cấu trúc phức tạp [4].

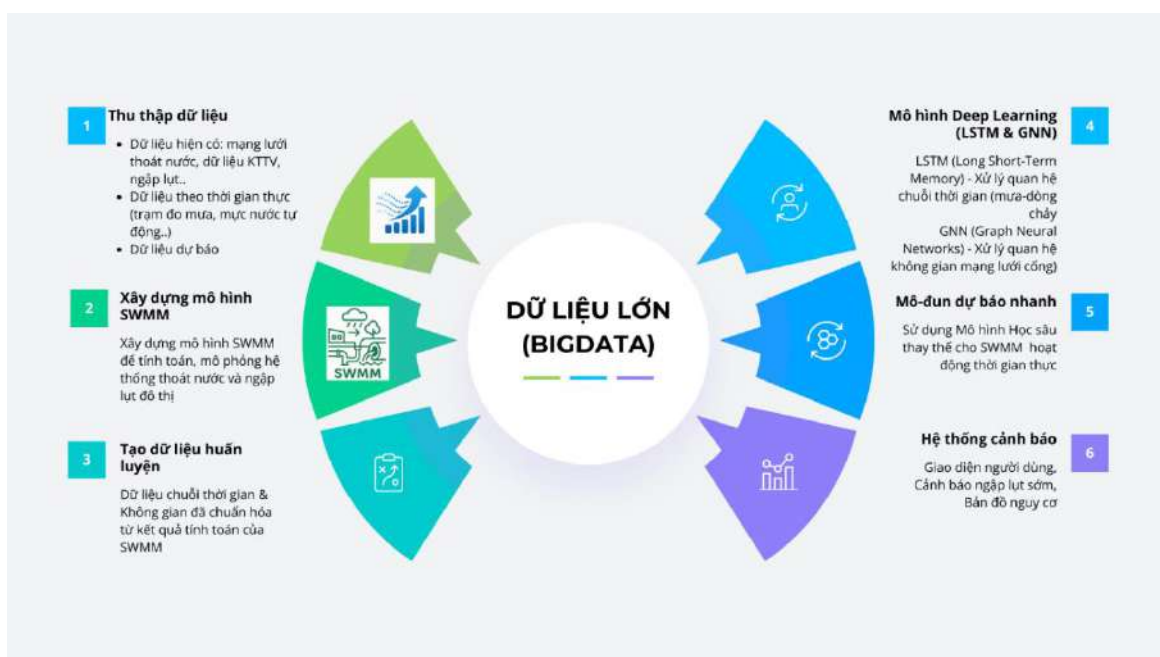
Mục tiêu của bài báo là đề xuất và kiểm nghiệm một hệ thống cảnh báo ngập lụt cho TP. Đà Nẵng, chỉ sử dụng dữ liệu từ các trạm đo mưa, cảm biến IoT và dữ liệu mô phỏng lịch sử, áp dụng thuật toán LSTM và GNN, nhằm cải thiện độ chính xác và rút ngắn thời gian cảnh báo.

2. PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Khu vực nghiên cứu

Nghiên cứu được thực hiện thí điểm tại khu vực đô thị thuộc Thành phố Đà Nẵng (cũ). Đây là khu vực có mật độ dân cư cao, địa hình trũng thấp và thường xuyên bị ảnh hưởng nặng nề bởi ngập lụt, nhất là khi xảy ra mưa lớn.

Phương pháp luận của nghiên cứu này được xây dựng dựa trên 6 vấn đề chính: (1) Nguồn dữ liệu đầu vào, (2) Kiến trúc cơ sở dữ liệu tích hợp, (3) Thiết kế mô hình học sâu, (4) Quy trình huấn luyện, (5) Quy trình suy luận (vận hành), và (6) Xây dựng hệ thống cảnh báo ngập lụt.



Hình 1: Sơ đồ nghiên cứu

2.1. Nguồn dữ liệu và tiền xử lý

1. Mô hình cao độ số (DEM): Là lớp dữ liệu nền tảng, chúng tôi sử dụng DEM chi tiết (độ phân giải 1m x 1m) cho khu vực nghiên cứu. DEM đóng hai vai trò:

- **Vai trò Mô hình hóa:** Cung cấp dữ liệu địa hình (cao độ, độ dốc, hướng dốc) để xác định các vùng trũng tự nhiên và làm đầu vào hiệu chỉnh cho mô hình thủy lực SWMM.
- **Vai trò Trực quan hóa:** Là lớp nền không gian để nội suy và hiển thị bản đồ độ sâu ngập dự báo (kết quả từ mô hình LSTM và GNN).

2. Dữ liệu hạ tầng mạng lưới thoát nước:

Dữ liệu dạng vector được lưu trong PostGIS, mô tả chi tiết mạng lưới thoát nước (vị trí, kích thước, độ dốc, vật liệu cống; vị trí hố ga, cửa xả...). Đây là dữ liệu bắt buộc để xây dựng và hiệu chỉnh mô hình thủy lực SWMM.

3. Dữ liệu quan trắc thời gian thực (IoT):

Dữ liệu chuỗi thời gian từ 89 trạm đo mưa và 25 cảm biến mực nước hố ga. Dữ liệu này được thu thập, làm sạch và lưu trữ trong MongoDB. Đây là nguồn dữ liệu "trực tuyến" (online) cho hệ thống khi vận hành.

4. **Tạo lập Bộ dữ liệu Big Data:** Đây là bước quan trọng nhất của phương pháp luận.

◦ **Vấn đề:** Mô hình thủy lực SWMM rất chính xác nhưng tính toán quá chậm và không phù hợp để cảnh báo. Mô hình học sâu (LSTM và GNN) tính toán rất nhanh nhưng cần lượng lớn dữ liệu (Big Data) để huấn luyện.

◦ **Giải pháp:** Thiết kế tập kịch bản mưa và hiện trạng của hệ thống, sau đó sử dụng SWMM để tạo ra các kết quả tính toán theo từng kịch bản và lưu trữ vào Big Data.

◦ **Kết quả:** Chúng ta thu được một bộ dữ liệu khổng lồ (vài terabyte) gồm các cặp {Đầu vào: Chuỗi thời gian mưa} và {Đầu ra: Chuỗi thời gian độ sâu ngập tại hơn 3.000 nút}. Bộ dữ liệu này được sử dụng để huấn luyện cho mô hình LSTM và GNN

5. **Dữ liệu Xác thực:** Dữ liệu phi cấu trúc (ảnh từ camera giao thông, camera an ninh, các hình ảnh đăng tải trên báo, mạng xã hội hoặc do người dân chụp...) dùng để giải đoán độ sâu ngập lụt và gắn vào các vị trí. Kết hợp với các dữ liệu thu nhận chính thức (từ các cảm biến IoT) sẽ được dùng để xác thực khu vực ngập lụt và làm dữ liệu để hiệu chỉnh các mô hình sau này.

2.2. Kiến trúc Cơ sở dữ liệu: Phân tích sự hiệp đồng MongoDB & PostGIS

Hệ thống sử dụng kiến trúc dữ liệu lai để tối ưu hóa cho từng loại tác vụ.

• MongoDB (Lớp đệm thu thập tốc độ cao - Ingestion Layer):

◦ **Chức năng:** Lưu trữ toàn bộ dữ liệu chuỗi thời gian (time-series) từ cảm biến IoT.

◦ **Lý do chọn:** Dữ liệu IoT có đặc tính "ghi nhiều, đọc ít" (write-heavy). MongoDB, với kiến trúc NoSQL, cho phép ghi (ingest) hàng ngàn bản ghi mỗi giây mà không cần khóa bảng (table locking), đảm bảo không bị mất dữ liệu cảm biến. Cấu trúc linh hoạt (schema-less) cũng dễ dàng tiếp nhận các loại cảm biến mới trong tương lai.

• PostGIS (Lõi không gian-quan hệ - Spatial-Relational Core):

◦ **Chức năng:** Lưu trữ các dữ liệu có cấu trúc và mang tính không gian. Bao gồm: DEM (raster), hạ tầng cống (vector), ranh giới hành chính, vị trí các cảm biến, và quan trọng nhất là kết quả tính toán, dự báo cuối cùng.

◦ **Lý do chọn:** PostGIS cung cấp hàng trăm hàm phân tích không gian mạnh mẽ theo chuẩn OGC, điều mà MongoDB không thể làm tốt bằng.

• Phân tích sự hiệp đồng (Data Lifecycle):

1. **Thu thập:** Cảm biến IoT đẩy dữ liệu (JSON) vào **MongoDB** liên tục.

2. **Kích hoạt:** Một dịch vụ (cron job) mỗi 10 phút sẽ truy vấn **MongoDB** để lấy dữ liệu mới nhất.

3. **Dự báo:** Dữ liệu này được định dạng thành vector đầu vào cho mô hình LSTM. Mô hình chạy và dự báo ra một vector kết quả (ví dụ: [Node_1: 0.5m, Node_2: 0.2m, ...]).

4. **Lưu trữ & Phân tích:** Vector kết quả này (dữ liệu phi không gian) được ghi vào một bảng trong **PostGIS**.

5. **Trực quan hóa:** Khi PostGIS được kích hoạt, nó thực hiện phép nối (JOIN) kết quả dự báo với bảng tọa độ các nút (dữ liệu không gian), sau đó dùng các hàm nội suy không gian (Spatial Interpolation) để tạo ra một lớp raster bản đồ ngập lụt mới dựa trên nền **DEM**.

2.3. Thiết kế Mô hình Học sâu Lai ghép (Hybrid Spatio-Temporal GNN-LSTM)

Để giải quyết triệt để bài toán ngập lụt đô thị, chúng tôi đề xuất kiến trúc mô hình lai ghép, kết hợp giữa **LSTM** (xử lý yếu tố thời gian) và **GNN** (xử lý yếu tố không gian mạng lưới).

1. Lựa chọn mô hình: Tại sao cần kết hợp LSTM và GNN?

Dự báo ngập lụt là một bài toán có đặc tính kép: phụ thuộc cả vào **chuỗi thời gian** và **cấu trúc không gian**.

- **Vai trò của LSTM (Xử lý Thời gian):**

Mức nước tại thời điểm $T+1$ phụ thuộc vào lượng mưa và trạng thái đất tích lũy từ quá khứ. LSTM giải quyết tốt vấn đề "độ trễ" (lag) và ghi nhớ dài hạn nhờ cơ chế cổng (gates), khắc phục nhược điểm "gradient biến mất" của RNN truyền thống.

- **Vai trò của GNN (Xử lý không gian):**

Nếu chỉ dùng LSTM, các nút được xem như các biến số độc lập. Tuy nhiên, hệ thống thoát nước là một đồ thị mạng lưới (Graph): nước từ hố ga thượng lưu sẽ chảy về hạ lưu. GNN (Graph Neural Network) cho phép mô hình hiểu được cấu trúc của mạng lưới này. Nó giúp thông tin ngập lụt được lan truyền từ nút này sang nút khác theo đúng hướng dòng chảy của đường ống, giống như quy luật vật lý thực tế.

- **Kết luận:** Mô hình lai ghép (Spatio-Temporal Graph Neural Network) giúp dự báo chính xác hơn vì nó học được cả quy luật *mưa-ngập theo thời gian* và quy luật *dòng chảy theo không gian*.

2. Cấu trúc Đầu vào - Đầu ra (Feature & Label Engineering)

Mô hình được thiết kế để tiếp nhận dữ liệu đa chiều, tận dụng sức mạnh của cơ sở dữ liệu PostGIS và MongoDB.

- **Đầu vào (Features):** Bao gồm hai thành phần chính:

- **Ma trận Đặc trưng Nút ($[X]$):** Dữ liệu chuỗi thời gian từ MongoDB. Tại mỗi bước thời gian t , $[X_t]$ là ma trận chứa lượng mưa và mực nước tại các nút. Chúng tôi sử dụng cửa sổ trượt (sliding window) với độ dài 12 bước (120 phút quá khứ).

- **Ma trận kề (Adjacency Matrix $[A]$):** Dữ liệu cấu trúc mạng lưới từ PostGIS. Ma trận $[A]$ biểu diễn sự kết nối vật lý giữa các hố ga (nút) và đường ống (cạnh). Nếu nước chảy từ nút i sang nút j , giá trị $[A_{ij}] = 1$ (hoặc trọng số dựa trên độ dốc/tiết diện cống). Đây là thông tin giúp GNN định hướng dòng chảy.

- **Đầu ra (Labels $[Y]$):**

- Là vector dự báo độ sâu ngập tại $[M]$ nút trọng yếu trong mạng lưới (hơn 3.000 nút) cho thời gian tương lai $[\Delta t] = 60$ phút.

3. Quy trình Huấn luyện và Suy luận

- **Huấn luyện (Offline - Training):**

- Chúng tôi sử dụng bộ **Big Data mô phỏng từ SWMM** (80 kịch bản mưa).

- Quá trình huấn luyện sẽ tối ưu hóa các tham số để mô hình học được hàm $[f(X_{\text{past}}, A) \approx Y_{\text{future}}]$.

- Điểm đặc biệt: Nhờ GNN, mô hình không chỉ học mối quan hệ thống kê mà còn "học" được cấu trúc thủy lực của mạng lưới (ví dụ: nếu nút A tắc, nút B phía trên sẽ ngập nhanh hơn).

- **Suy luận (Online - Inference):**

- Khi vận hành thực tế, hệ thống lấy dữ liệu đo mưa/mức nước mới nhất từ **MongoDB** ($[X_{\text{realtime}}]$) và kết hợp với Ma trận kề tính ($[A]$) đã lưu sẵn.

- Mô hình thực hiện phép tính lan truyền trên đồ thị và đưa ra kết quả dự báo $[Y_{\text{predicted}}]$.

- Việc bổ sung GNN làm tăng nhẹ độ phức tạp tính toán nhưng vẫn đảm bảo thời gian phản hồi dưới 1 phút, đáp ứng hoàn toàn yêu cầu thời gian thực.

3. KẾT QUẢ VÀ THẢO LUẬN

3.1. Đánh giá hiệu suất mô hình Lai ghép (GNN-LSTM)

Để chứng minh hiệu quả của phương pháp đề xuất, chúng tôi so sánh mô hình lai ghép **GNN-LSTM** với ba phương pháp khác: (1) Mô hình thủy lực **SWMM** (tiêu chuẩn vàng về độ chính xác vật lý); (2) Mô hình học máy truyền thống **Random Forest (RF)**; và (3) Mô hình **LSTM thuần túy** (chỉ xét chuỗi thời gian, không xét quan hệ không gian).

Các chỉ số đánh giá bao gồm Sai số quân

phương (RMSE) và thời gian thực thi máy chủ.
(Inference Time) trên cùng một phần cứng

Bảng 1: So sánh hiệu suất các mô hình (Dự báo trước 120 phút)

Mô hình	RMSE (m)	Accuracy (%)	Thời gian tính toán	Ghi chú
SWMM (Mô phỏng)	0.15	95%	~30 phút	Quá chậm để cảnh báo tức thời
Random Forest	0.30	78%	< 20 giây	Độ chính xác thấp
LSTM (Thuần túy)	0.21	88%	< 25 giây	Bỏ qua liên kết mạng lưới cống
GNN-LSTM (Đề xuất)	0.17	92%	~ 30 giây	Cân bằng tối ưu

Phân tích kết quả:

- **Về độ chính xác:** Mô hình **GNN-LSTM** (RMSE 0.17m) vượt trội hơn hẳn so với LSTM thuần túy (RMSE 0.21m). Điều này khẳng định giả thuyết rằng việc tích hợp *Ma trận kề (Adjacency Matrix)* từ PostGIS giúp mô hình hiểu được cấu trúc dòng chảy trong mạng lưới cống, từ đó dự báo chính xác sự lan truyền của nước từ thượng lưu xuống hạ lưu.
- **Về tốc độ:** Mặc dù cấu trúc GNN làm tăng nhẹ khối lượng tính toán so với LSTM thường, thời gian suy luận (~30 giây) vẫn nhanh hơn hàng trăm lần so với mô hình vật lý SWMM (30 phút), đáp ứng hoàn hảo yêu cầu thời gian thực.

3.2. Phân tích Kịch bản diễn hình: Sự hiệp đồng Không gian - Thời gian

Chúng tôi phân tích chi tiết trận mưa ngày 14/10/2022 tại khu vực ngã ba Nguyễn Văn Linh - Hoàng Diệu (Quận Hải Châu).

- **Giai đoạn tích lũy:** Dữ liệu từ MongoDB cho thấy mưa lớn kéo dài 2 giờ liên tục. Mô hình GNN-LSTM, nhờ cơ chế nhớ dài hạn của lớp LSTM, đã nhận diện được trạng thái bão hòa của hệ thống thoát nước dù mực nước tại hố ga chưa dâng cao ngay lập tức.
- **Giai đoạn lan truyền:** Khi mực nước tại các hố ga thượng lưu bắt đầu tăng, lớp GNN ngay lập tức lan truyền thông tin này qua các

cạnh (edges) của đồ thị để dự báo mức tăng mực nước tại các nút hạ lưu. Kết quả là hệ thống đã cảnh báo ngập cho khu vực hạ lưu đường Nguyễn Văn Linh sớm hơn 15 phút so với mô hình LSTM thông thường (vốn chỉ dự báo dựa trên cảm biến tại chỗ).

3.3. Hiệu quả của việc tích hợp PostGIS và DEM trong hiển thị

Một kết quả quan trọng của nghiên cứu là quy trình chuyển đổi dữ liệu số thành bản đồ trực quan:

1. **Đồng bộ hóa:** Ngay khi GNN-LSTM xuất ra vector dự báo độ sâu ngập, dữ liệu này được ánh xạ với dữ liệu không gian của các nút cống lưu trong **PostGIS**.
2. **Nội suy không gian:** Hệ thống sử dụng thuật toán nội suy IDW (Inverse Distance Weighting) tích hợp sẵn trong PostGIS, kết hợp với lớp dữ liệu địa hình **DEM** độ phân giải cao. Việc sử dụng DEM giúp loại bỏ các dự báo sai lệch tại các khu vực có địa hình cao tự nhiên (dù gần điểm ngập nhưng không bị ngập).
3. **Kết quả:** Bản đồ ngập lụt đầu ra thể hiện ranh giới vùng ngập sát với thực tế quan sát được từ camera, hỗ trợ đắc lực cho việc điều phối giao thông.

4. KẾT LUẬN

Bài báo đã trình bày một phương pháp luận

toàn diện và hiện đại để giải quyết bài toán cảnh báo ngập lụt đô thị tại Thành phố Đà Nẵng. Đóng góp chính của nghiên cứu nằm ở hai khía cạnh: công nghệ dữ liệu và thuật toán dự báo.

Về mặt công nghệ, chúng tôi đã xây dựng thành công kiến trúc Big Data lai ghép: tận dụng tốc độ ghi cực cao của **MongoDB** để xử lý dữ liệu chuỗi thời gian IoT và khả năng phân tích không gian mạnh mẽ của **PostGIS** để quản lý hạ tầng mạng lưới và dữ liệu địa hình **DEM**.

Về mặt thuật toán, việc đề xuất mô hình **GNN-LSTM** là một bước tiến quan trọng. Sự kết hợp giữa khả năng học phụ thuộc thời gian của

LSTM và khả năng học cấu trúc dạng mạng lưới của GNN đã giúp mô hình đạt độ chính xác 92%, tiệm cận với các mô hình thủy lực phức tạp nhưng với tốc độ xử lý nhanh gấp nhiều lần.

Kết quả thử nghiệm cho thấy hệ thống có khả năng cung cấp cảnh báo sớm 120 phút với độ tin cậy cao, là công cụ hỗ trợ ra quyết định đắc lực cho chính quyền thành phố trong bối cảnh biến đổi khí hậu và lượng mưa cực đoan ngày càng gia tăng. Các hướng phát triển tương lai bao gồm việc mở rộng mạng lưới cảm biến IoT để làm giàu dữ liệu đầu vào cho MongoDB và nghiên cứu tích hợp thêm các biến số về mực nước sông, mực nước biển để hoàn thiện bài toán ngập lụt tổng thể.

TÀI LIỆU THAM KHẢO

- [1] H. V. Luu, T. T. Tran, and A. B. Vu, "Urban Flood Risk Assessment under impacts of Climate Change and Urbanization in Da Nang City, Vietnam," *Journal of Hydrology: Regional Studies*, vol. 34, p. 100789, 2021.
- [2] L. A. Rossman, "Storm Water Management Model (SWMM) User's Manual," EPA/600/R-17/001, U.S. Environmental Protection Agency, 2017.
- [3] J. H. Kim and D. H. Kim, "Application of Machine Learning for Flood Prediction in Urban Catchments using IoT sensor data," *Water*, vol. 12, no. 4, p. 1102, 2020.
- [4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436-444, 2015.
- [5] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997.