**ĐẠI HỌC QUỐC GIA TP.HCM**

**TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN**

**BÙI NGÔ TÔN BÁCH**

**DỰ ĐOÁN GIÁ CỔ PHIẾU BẰNG HỌC MÁY**

**LUẬN VĂN THẠC SĨ**

*TP. Hồ Chí Minh, năm 2023*

**ĐẠI HỌC QUỐC GIA TP.HCM**

**TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN**

**BÙI NGÔ TÔN BÁCH**

**DỰ ĐOÁN GIÁ CỔ PHIẾU BẰNG HỌC MÁY**

**Ngành: Vật lý Kỹ thuật**

**Mã số Ngành: 8520401**

**NGƯỜI HƯỚNG DẪN KHOA HỌC**

**TS. Nguyễn Anh Huy**

*TP. Hồ Chí Minh, năm 2023*

# LỜI CAM ĐOAN

Tôi cam đoan luận văn thạc sĩ ngành Vật lý Kỹ thuật với đề tài *“Dự đoán giá cổ phiếu bằng học máy”* là công trình khoa học do Tôi thực hiện dưới sự hướng dẫn của TS. Nguyễn Anh Huy.

Những kết quả nghiên cứu của luận văn hoàn toàn trung thực và chính xác.

Học viên cao học

(Ký tên, ghi họ tên)

Bùi Ngô Tôn Bách

# LỜI CẢM ƠN

Để có thể hoàn thành được khóa luận, lời đầu tiên tôi xin gửi lời cảm ơn chân thành nhất tới TS. Nguyễn Anh Huy, người đã trực tiếp hướng dẫn trong suốt quá trình hình thành ý tưởng và thực hiện đề tài luận văn này.

Xin chân thành cảm ơn quý Thầy/Cô đã tận tình giảng dạy, truyền đạt kiến thức cho tôi trong gần hai năm qua. Tôi xin chân thành cảm ơn gia đình, bạn bè, đồng nghiệp đã có sự động viên, hỗ trợ và đóng góp ý kiến để tôi có thể hoàn thành công trình nghiên cứu.

Tuy nhiên, trong quá trình thực hiện luận văn không tránh khỏi những sai sót do kiến thức còn hạn chế, tôi rất mong nhận được sự đóng góp từ quý thầy cô và các bạn để tôi hoàn thành luận văn một cách hoàn chỉnh nhất và để trao dồi thêm kiến thức cho những nghiên cứu trong tương lai.

Tôi xin chân thành cảm ơn!!!

TP. Hồ Chí Minh, ngày tháng năm 2023

Bùi Ngô Tôn Bách

**MỤC LỤC**

[LỜI CAM ĐOAN i](#_heading=h.gjdgxs)

[LỜI CẢM ƠN ii](#_heading=h.30j0zll)

[DANH MỤC CÁC HÌNH VẼ VÀ ĐỒ THỊ vi](#_heading=h.1fob9te)

[DANH MỤC CÁC BẢNG SỐ LIỆU viii](#_heading=h.3znysh7)

[DANH MỤC CÁC TỪ VIẾT TẮT x](#_heading=h.2et92p0)

[TRANG THÔNG TIN LUẬN VĂN TIẾNG VIỆT xi](#_heading=h.tyjcwt)

[THESIS INFORMATION xiii](#_heading=h.3dy6vkm)

[MỞ ĐẦU 1](#_heading=h.1t3h5sf)

[CHƯƠNG 1 TỔNG QUAN 3](#_heading=h.26in1rg)

[1.1](#_heading=h.4fsjm0b) CẤU TẠO CỦA NÃO NGƯỜI 3

[1.2. TỔNG QUAN VỀ TÍN HIỆU ĐIỆN NÃO (EEG) 5](#_heading=h.lnxbz9)

[1.2.1 Điện não đồ 5](#_heading=h.35nkun2)

[1.2.2 Phương pháp thu tín hiệu điện não 7](#_heading=h.1ksv4uv)

[1.3 MỤC TIÊU NGHIÊN CỨU 9](#_heading=h.44sinio)

[1.4 GIỚI HẠN CỦA LUẬN VĂN 10](#_heading=h.2jxsxqh)

[CHƯƠNG 2 MẠNG NƠ-RON NHÂN TẠO 11](#_heading=h.3j2qqm3)

[2.1 TỔNG QUAN VỀ MẠNG NƠ-RON NHÂN TẠO 11](#_heading=h.1y810tw)

[2.2 MỘT SỐ LOẠI MẠNG NƠ-RON NHÂN TẠO 13](#_heading=h.2xcytpi)

[2.2.1](#_heading=h.1ci93xb) Mạng nơ-ron truyền thẳng (ANN) 13

[2.2.2 Mạng hồi quy (RNN) 15](#_heading=h.qsh70q)

[2.2.3 Mạng bộ nhớ dài-ngắn hạn (LSTM) 17](#_heading=h.49x2ik5)

[2.2.4 Mạng nơ-ron tích chập 19](#_heading=h.3o7alnk)

[2.3 CÁC PHƯƠNG PHÁP HUẤN LUYỆN MẠNG NƠ-RON 25](#_heading=h.3tbugp1)

[2.3.1 Phương pháp học có giám sát 25](#_heading=h.28h4qwu)

[2.3.2 Phương pháp học không giám sát 26](#_heading=h.1mrcu09)

[2.3.3 Phương pháp học tăng cường 27](#_heading=h.111kx3o)

[2.3](#_heading=h.206ipza) THUẬT TOÁN TỐI ƯU HÓA 28

[2.4.1 Thuật toán Momentum 29](#_heading=h.3cqmetx)

[2.4.2 Thuật toán Adagrad 29](#_heading=h.4bvk7pj)

[2.4.3 Thuật toán Adadelta 31](#_heading=h.2r0uhxc)

[2.4.4 Thuật toán Adaptive Moment Estimation (Adam) 32](#_heading=h.1664s55)

[CHƯƠNG 3 PHƯƠNG PHÁP TRÍCH XUẤT TÍNH NĂNG 33](#_heading=h.3q5sasy)

[3.1 TIỀN XỬ LÝ VÀ TRÍCH XUẤT TÍNH NĂNG 33](#_heading=h.25b2l0r)

[3.1.1 Bộ dữ liệu DEAP [5] 33](#_heading=h.34g0dwd)

[3.1.2 Phân đoạn, tách dải tần số và trích xuất tính năng 35](#_heading=h.43ky6rz)

[3.1.3 Xây dựng mô hình 1D, 2D và 3D 38](#_heading=h.1x0gk37)

[CHƯƠNG 4 THIẾT KẾ MÔ HÌNH MẠNG NƠ-RON NHÂN TẠO 41](#_heading=h.pkwqa1)

[4.1 THIẾT KẾ MẠNG NƠ-RON TRUYỀN THẲNG 41](#_heading=h.39kk8xu)

[4.2 THIẾT KẾ MẠNG BỘ NHỚ DÀI-NGẮN (LSTM) 42](#_heading=h.48pi1tg)

[4.3 THIẾT KẾ MẠNG NƠ-RON TÍCH CHẬP 2 CHIỀU (2D-CNN) VÀ 3 CHIỀU (3D-CNN) 42](#_heading=h.1302m92)

[4.4 THIẾT KẾ MẠNG NƠ-RON TÍCH HỢP (HYBRID NETWORK) 45](#_heading=h.haapch)

[CHƯƠNG 5 KẾT QUẢ VÀ NHẬN XÉT 47](#_heading=h.1gf8i83)

[5.1 KẾT QUẢ KHẢO SÁT MẠNG TRUYỀN THẲNG 47](#_heading=h.40ew0vw)

[5.2 KẾT QUẢ KHẢO SÁT MẠNG LSTM 50](#_heading=h.1tuee74)

[5.3 KẾT QUẢ KHẢO SÁT MẠNG 2D-CNN VÀ 3D-CNN 55](#_heading=h.1ljsd9k)

[5.4 KẾT QUẢ KHẢO SÁT MẠNG TÍCH HỢP (HYBRID NETWORK) 58](#_heading=h.4iylrwe)

[5.5 So sánh với các nghiên cứu trước 62](#_heading=h.j8sehv)

[KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN 65](#_heading=h.1idq7dh)

[TÀI LIỆU THAM KHẢO 67](#_heading=h.1vsw3ci)

# DANH MỤC CÁC HÌNH VẼ VÀ ĐỒ THỊ

[**Hình 1.1:** Cấu tạo bộ não của con người 3](#_heading=h.2uxtw84)

[**Hình 1.2:** Cấu trúc của một nơ-ron 4](#_heading=h.1a346fx)

[**Hình 1.1:** Mũ đo điện cực Emotiv 7](#_heading=h.3u2rp3q)

[**Hình 1.2:** Các thiết bị phục vụ quá trình đo tín hiệu điện não 8](#_heading=h.2981zbj)

[**Hình 2.1:** Cấu trúc của một nút 13](#_heading=h.3whwml4)

[**Hình 2.2:** Cấu trúc của một mạng ANN 14](#_heading=h.2bn6wsx)

[**Hình 2.3:** Cấu trúc của một mạng hồi quy 15](#_heading=h.3as4poj)

[**Hình 2.4:** Mô hình chi tiết của một lớp ẩn 16](#_heading=h.1pxezwc)

[**Hình 2.5:** Cấu trúc của một tế bào trong mạng LSTM 17](#_heading=h.2p2csry)

[**Hình 2.6:** Phương pháp nhận diện của máy tính với một hình ảnh 20](#_heading=h.23ckvvd)

[**Hình 2.7:** Các bước tính toán của phương pháp tích chập 21](#_heading=h.32hioqz)

[**Hình 2.8:** Ma trận chập với độ trượt bằng một 21](#_heading=h.1hmsyys)

[**Hình 2.9:** Ma trận chập với độ trượt bằng hai 22](#_heading=h.41mghml)

[**Hình 2.10:** Ma trận chập bị trượt ra ngoài 22](#_heading=h.2grqrue)

[**Hình 2.11:** Lớp padding phủ xung quanh hỗ trợ ma trận chập bị trượt ra ngoài có thể tính toán được các tính năng 23](#_heading=h.vx1227)

[**Hình 2.12:** Quá trình max pooling 23](#_heading=h.1v1yuxt)

[**Hình 2.13:** Quá trình global average polling 24](#_heading=h.4f1mdlm)

[**Hình 2.14:** Mô hình quyết định của lớp kết nối đầy đủ 25](#_heading=h.19c6y18)

[**Hình 2.15:** Mô hình huấn luyện học có giám sát 26](#_heading=h.nmf14n)

[**Hình** **2.16:** Mô hình kiểm tra phương pháp học có giám sát 26](#_heading=h.37m2jsg)

[**Hình 2.17:** Phương pháp học không giám sát 27](#_heading=h.2lwamvv)

[**Hình 2.18:** Mô tả phương pháp học tăng cường 27](#_heading=h.3l18frh)

[**Hình 2.19:** Mô tả tốc độ hội tụ thuật toán SGD khi không và có sử dụng momentum 29](#_heading=h.1rvwp1q)

[**Hình 3.1:** Tổng quát quá trình lọc dải tần số và tính độ phức tạp tín hiệu 36](#_heading=h.2iq8gzs)

[**Hình 3.2:** Quá trình phân loại các nhãn 37](#_heading=h.xvir7l)

[**Hình 3.3:** Mô hình cảm xúc của Russell [9] 38](#_heading=h.3hv69ve)

[**Hình 3.4:** Vị trí các kênh của thiết bị thu trong tập dữ liệu DEAP 38](#_heading=h.4h042r0)

[**Hình 3.5:** Quá trình xây dựng mô hình 1D cho dữ liệu 39](#_heading=h.2w5ecyt)

[**Hình 3.6:** Ma trận mô phỏng các kênh trong tập dữ liệu DEAP 39](#_heading=h.1baon6m)

[**Hình 3.7:** Tồng quan quá trình xây dựng dữ liệu dạng 2D và 3D 40](#_heading=h.2afmg28)

[**Hình 4.1:** Mô hình tổng quan của mạng nơ-ron truyền thẳng 41](#_heading=h.1opuj5n)

[**Hình 4.2:** Mô hình tổng quan của mạng LSTM 42](#_heading=h.2nusc19)

[**Hình 4.3:** Mô hình tổng quan của mạng 2D-CNN 43](#_heading=h.3mzq4wv)

[**Hình 4.4:** Mô hình tổng quan của mạng 3D-CNN 44](#_heading=h.2250f4o)

[**Hình 4.5:** Mô hình tổng quan của mạng tích hợp 45](#_heading=h.319y80a)

[**Hình 5.1:** Mối quan hệ giữa số nút và độ chính xác phân loại 47](#_heading=h.2fk6b3p)

[**Hình 5.2:** Kết quả về độ chính xác khi khảo sát với số nút thay đổi trong mô hình mạng truyền thẳng hai lớp 48](#_heading=h.upglbi)

[**Hình 5.3:** Kết quả kháo sát độ chính xác với các tốc độ học khác nhau 49](#_heading=h.3ep43zb)

[**Hình 5.4:** Kết quả kháo sát độ chính xác và thời gian huấn luyện với số nút ẩn 50](#_heading=h.4du1wux)

[**Hình 5.5:** Mối quan hệ giữa độ chính xác và thời gian huấn luyện với số epoch 51](#_heading=h.2szc72q)

[**Hình 5.6:** Ma trận phân loại của mô hình LSTM 3 lớp 53](#_heading=h.279ka65)

[**Hình 5.7:** Mối quan hệ sự ảnh hưởng giữa số nút ẩn và độ chính xác phân loại trạng thái cảm xúc 58](#_heading=h.1d96cc0)

# DANH MỤC CÁC BẢNG SỐ LIỆU

[**Bảng 1.1:** Thông số các dải tần số cơ bản [14] 6](#_heading=h.odc9jc)

[**Bảng 1.2:** Thông số các thiết bị đo đạc tín hiệu điện não [15] 7](#_heading=h.38czs75)

[**Bảng 2.1:** Bảng so sánh giữa tế bào nơ-ron sinh học và mạng nơ-ron nhân tạo 11](#_heading=h.1nia2ey)

[**Bảng 2.2:** Thông số các giá trị trong tế bào LSTM 18](#_heading=h.147n2zr)

[**Bảng 3.1:** Mô tả cấu trúc của bộ dữ liệu DEAP đã được tiền xử lý 35](#_heading=h.1jlao46)

[**Bảng 3.2:** So sánh giữa phương pháp thị giác máy tính và phương pháp thực hiện 40](#_heading=h.3vac5uf)

[**Bảng 5.1:** Mối quan hệ giữa tốc độ học với độ chính xác và thời gian huấn luyện trong mạng LSTM 52](#_heading=h.184mhaj)

[**Bảng 5.2:** Kết quả khảo sát giữa số lớp với độ chính xác phân loại 53](#_heading=h.3s49zyc)

[**Bảng 5.3:** Sự ảnh hưởng của độ lớn tập dữ liệu đến độ chính xác phân loại 54](#_heading=h.meukdy)

[**Bảng 5.4:** Kết quả khảo sát với độ dài phân đoạn khác nhau đến độ chính xác phân loại các trạng thái cảm xúc 54](#_heading=h.36ei31r)

[**Bảng 5.5:** Kết quả khảo sát giữa các kích thước ma trận chập với độ chính xác phân loại các trạng thái cảm xúc của mô hình 3D-CNN 55](#_heading=h.45jfvxd)

[**Bảng 5.6:** Kết quả khảo sát giữa các kích thước ma trận chập với độ chính xác phân loại các trạng thái cảm xúc của mô hình 2D-CNN 56](#_heading=h.2koq656)

[**Bảng 5.7:** Kết quảkhảo sát giữa độ trượt của các ma trận chập ảnh hưởng đến độ phân loại chính xác của mô hình 3D-CNN 57](#_heading=h.zu0gcz)

[**Bảng 5.8:** Kết quảkhảo sát giữa độ trượt của các ma trận chập ảnh hưởng đến độ phân loại chính xác của mô hình 2D-CNN 57](#_heading=h.3jtnz0s)

[**Bảng 5.9:** Mối quan hệ giữa độ chính xác phân loại và các hàm hoạt động 58](#_heading=h.1yyy98l)

[**Bảng 5.10:** Kết quả khảo sát chi tiết hơn với các bước nhảy nhỏ hơn giữa số nút ẩn tác động đến độ chính xác phân loại 59](#_heading=h.2ce457m)

[**Bảng 5.11:** Kết quả khảo sát giữa các kích thước ma trận chập với độ chính xác phân loại các trạng thái cảm xúc của mô hình mạng nơ-ton tích chập 60](#_heading=h.3bj1y38)

[**Bảng 5.12:** Kết quảkhảo sát giữa độ trượt của các ma trận chập ảnh hưởng đến độ phân loại chính xác 61](#_heading=h.2pta16n)

[**Bảng 5.13:** Mối quan hệ giữa độ chính xác phân loại và các hàm hoạt động trong mô hình tích chập 62](#_heading=h.3oy7u29)

[**Bảng 5.14:** Bảng so sánh độ chính xác phân loại của các phương pháp đề xuất 63](#_heading=h.338fx5o)

# DANH MỤC CÁC TỪ VIẾT TẮT

|  |  |  |
| --- | --- | --- |
| **VIẾT TẮT** | **TÊN TIẾNG ANH** | **TÊN TIẾNG VIỆT** |
| 1D | One-Dimensional | Một chiều |
| 2D | Two-Dimensional | Hai chiều |
| 3D | Three-Dimensional | Ba chiều |
| 2D-CNN | 2D Convolution Neural Network | Mạng nơ-ron tích chập hai chiều |
| 3D-CNN | 3D Convolution Neural Network | Mạng nơ-ron tích chập ba chiều |
| A/D | Analog/Digital | Tương tự/Số |
| Adam | Adaptive Moment Estimation |  |
| BAM | Bidirectional Associative Memory | Bộ nhớ kết hợp hai chiều |
| CNN | Convolution Neural Network | Mạng nơ-ron tích chập |
| DE | Differential Entropy | Entropy vi sai |
| EEG | Electroencephalography | Điện não đồ |
| GPU | Graphics Processing Unit | Bộ xử lý đồ họa |
| LSTM | Long Short Term Memory | Bộ nhớ dài ngắn hạn |
| RNN | Recurent neural network | Mạng nơ-ron hồi quy |
| RMS | Root mean squared | Giá trị hiệu dụng |
| SGD | Stochastic Gradient Descent |  |
| SOM | Self Organization Map | Bản đồ tự tổ chức |

# TRANG THÔNG TIN LUẬN VĂN TIẾNG VIỆT

Tên đề tài luận văn: DỰ ĐOÁN GIÁ CỔ PHIẾU BẰNG HỌC MÁY.

Ngành: Vật lý Kỹ thuật

Mã số ngành: 8520401

Họ tên học viên cao học: Bùi Ngô Tôn Bách

Khóa đào tạo: 30

Người hướng dẫn khoa học: TS. Nguyễn Anh Huy

Cơ sở đào tạo: Trường Đại học Khoa học Tự nhiên, ĐHQG.HCM

1. **TÓM TẮT NỘI DUNG LUẬN VĂN**:

Nội dung của luận văn gồm 5 chương:

**Chương 1: Tổng quan**

Trình bày một số kiến thức cơ sở về não bộ, lịch sử tín hiệu điện não và các đặc điểm, các phương pháp đo nhận tín hiệu EEG.

**Chương 2: Mô hình LSTM**

Trình bày tóm lược về lịch sử hình thành mạng nơ-ron nhân tạo, các loại cấu trúc thiết kế mạng nơ-ron, các phương pháp huấn luyện cũng như các thuật toán tối ưu.

**Chương 3: Phương pháp xử lý dữ liệu**

Xây dựng các thuật toán trích xuất tính năng, biến dổi các tín hiệu EEG thành các dải tần số khác nhau đồng thời biến đổi các dạng dữ liệu thành 1 chiều, 2 chiều và 3 chiều.

**Chương 4: Thiết kế mô hình LSTM**

Xây dựng các mô hình mạng nơ-ron nhân tạo khác nhau bao gồm mạng truyền thẳng (ANN), Mạng LSTM, mạng 2D-CNN, đồng thời đề xuất mô hình mạng 3D-CNN và mạng tích hơp (3D-CNN + 3D-LSTM).

**Chương 5: Kết quả và nhận xét**

Trình bày điểm mạnh và điểm yếu của đề tài đạt được cũng như so sánh các kết quả đạt được với các nghiên cứu tương tự từ đó đưa ra các nhận xét

**2. NHỮNG KẾT QUẢ MỚI CỦA LUẬN VĂN**:

Luận văn đã đề xuất hai mô hình mới trong phân loại tín hiệu EEG cho bộ liệu DEAP mà các tác giả trước đó chưa thực hiện, đó là mô hình 3D-CNN và mô hình mạng tích hợp (3D-CNN + 3D-LSTM):

* Với mô hình mạng 3D-CNN: ma trận chập có kích thước 4×4×2, độ trượt bằng 1×1×1, hàm kích hoạt ReLU và ClippedReLU. Kết quả cho thấy độ chính xác (76.1%) phân loại khả quan nhất.
* Với mạng tích hợp (3D-CNN + 3D-LSTM): số lượng nút ẩn là 1024, kích thước các ma trận chập 4×4×2, độ trượt bằng 1×1×1, hàm kích hoạt ReLU và ClippedReLU là phù hợp nhất.

**3.** **CÁC ỨNG DỤNG/ KHẢ NĂNG ỨNG DỤNG TRONG THỰC TIỄN HAY NHỮNG VẤN ĐỀ CÒN BỎ NGỎ CẦN TIẾP TỤC NGHIÊN CỨU**

- Thu thập bộ dữ liệu có kích thước lớn để phục vụ cho việc nghiên cứu độc lập.

- Kết hợp nhiều thuật toán khác nhau (học sâu, học máy) để tăng hiệu quả của việc phân loại tín hiệu EEG.

- Ứng dụng của việc phân loại tín hiệu EEG trong việc điều khiển thiết bị.

|  |  |
| --- | --- |
| **TẬP THỂ CÁN BỘ HƯỚNG DẪN**  (Ký tên, họ tên) | **HỌC VIÊN CAO HỌC**  (Ký tên, họ tên) |

**XÁC NHẬN CỦA CƠ SỞ ĐÀO TẠO**

**HIỆU TRƯỞNG**

# THESIS INFORMATION

Thesis title: PREDICTING STOCK PRICES USING MACHINE LEARNING.

Speciality: Engineering Physics

Code: 8520401

Name of Master Student: Bui Ngo Ton Bach

Academic year: 30

Supervisor: PhD. Nguyen Anh Huy

At: VNUHCM - University of Science

1. **SUMMARY:**

The content of the the thesis consists of 5 chapters:

**Chapter 1: Overview**

This chapter presents some basic knowledge about the brain, the history of electroencephalography signals and it’s characteristics. At the same time, this chapter also introduce some methods of measuring EEG signals.

**Chapter 2: Artificial Neural Networks**

This chapter introduces a brief presentation on the history of the architecture of an artificial neural networks, then it‘s presents some types of neural network, training methods as well as the theory of some optimal algorithms.

**Chapter 3: Method of Features Extraction**

Feature extraction algorithms are presented in this chapter, it’s include extracted the EEG signals into different frequency waves, and transforming data structure into one-dimensional, two-dimensional, and three-dimensional.

**Chapter 4: Designing an Artificial Neural Network Model**

This chapter design an ideals about models of artificial neural networks including artificial neural network (ANN), LSTM networks, 2D-CNN networks, also it’s proposes 3D-CNN models and hybrid networks (3D-CNN-3D-LSTM).

**Chapter 5: Results and disscusion**

This chapter present the strengths and weaknesses of the thesis as well as compare the survey results with similar research then giving a conclusion.

**2. NOVELTY OF THESIS:**

The thesis propose two new models in EEG signal classification model for DEAP dataset that previous authors have not yet implemented, we name it as 3D-CNN models and hybrid network models (3DLSTM+3D-CNN):

- An 3D-CNN network model compoments: the kernel with dimensions of 4×4×2, stride equal to 1×1×1, an activation function is ReLU or ClippedReLU. The results showed the classification accuracy (76.1%) is the best performance.

- A hybrid networks (3D-CNN + 3D-LSTM) compoments: the number of hidden nodes is 1024 for the 3D-LSTM, the size of the kernel is 4×4×2, the stride is equal to 1×1×1, activation function is the ReLU and ClippedReLU are the most suitable.

**3. APPLICATIONS/ APPLICABILITY/ PERSPECTIVE**

- Collect large-sized dataset for independent research.

- Combine many different algorithms (Deep learning, Machine learning) to increase the efficiency of EEG signal classification.

- Application of EEG signal classification in device control.

|  |  |
| --- | --- |
| **SUPERVISOR** | **Master STUDENT** |

**CERTIFICATION**

**UNIVERSITY OF SCIENCE**

**PRESIDENT**

# MỞ ĐẦU

Thị trường chứng khoán là một trong những thị trường quan trọng nhất trên thế giới. Nó cung cấp nền tảng để các công ty và cá nhân đầu tư vào cổ phiếu, giúp cho việc tăng trưởng và phát triển kinh doanh của công ty. Thị trường chứng khoán cổ phiếu là một thị trường tiềm năng, với nhiều yếu tố bên trong và bên ngoài tác động đến giá cổ phiếu. Để thành công trong đầu tư chứng khoán, nhà đầu tư cần phải hiểu rõ các yếu tố này và cách chúng tác động đến giá cổ phiếu. Một trong những yếu tố quan trọng nhất là tình hình kinh tế, bao gồm tỷ lệ lãi suất, tỷ lệ thất nghiệp, và tỷ giá. Các sự kiện quốc tế cũng có thể tác động đến thị trường chứng khoán, như chiến tranh thương mại hoặc sự kiện chính trị.

Dự đoán giá cổ phiếu là một khía cạnh quan trọng trong quản lý tài chính và đầu tư. Việc dự đoán giá cổ phiếu có thể giúp nhà đầu tư đưa ra quyết định mua bán cổ phiếu một cách hợp lý và hiệu quả. Hiện nay, thông tin về giá cổ phiếu được cập nhật liên tục và nhanh chóng, điều này đòi hỏi các nhà đầu tư phải có khả năng dự đoán giá cổ phiếu một cách chính xác và đồng thời quản lý rủi ro một cách hiệu quả.

Trong luận văn này, tôi sẽ trình bày một phương pháp dự đoán giá cổ phiếu hiệu quả là sử dụng mô hình LSTM. LSTM là một loại mô hình trí tuệ nhân tạo được sử dụng rộng rãi trong các bài toán dự đoán thời gian liên tục. Nó có khả năng học và ghi nhớ các quan hệ giữa các sự kiện trong quá khứ và hiện tại, giúp cho việc dự đoán trong tương lai trở nên chính xác hơn. Tôi sẽ sử dụng dữ liệu giá cổ phiếu thực tế để tiến hành thực nghiệm và đánh giá hiệu năng của thuật toán học máy. Đồng thời, tôi sẽ xem xét các yếu tố ảnh hưởng đến độ chính xác của dự đoán và cách cải thiện độ chính xác.

Nội dung của luận văn gồm 5 chương:

Chương 1: Tổng quan

Chương 2: Mô hình LSTM

Chương 3: Phương pháp xử lý dữ liệu

Chương 4: Thiết kế mô hình LSTM

Chương 5: Kết quả và nhận xét

# CHƯƠNG 1 TỔNG QUAN

## KHÁI NIỆM CHỨNG KHOÁN

Chứng khoán là một loại tài sản quan trọng trong giao dịch tài chính, xác nhận việc góp vốn cho hoạt động kinh doanh của một công ty hoặc tổ chức. Khi một công ty hoặc tổ chức muốn tăng vốn, họ có thể phát hành chứng khoán để bán cho người mua. Chứng khoán chứa thông tin về sở hữu vốn của công ty hoặc tổ chức và quyền lợi liên quan đến chứng khoán đó.

Chứng khoán có thể được chuyển nhượng trên thị trường chứng khoán, cho phép các nhà đầu tư mua và bán chứng khoán của công ty hoặc tổ chức. Giá chứng khoán thường được xác định bởi sự cân bằng giữa cung và cầu trên thị trường chứng khoán. Những chứng khoán giá tăng thường được xem là có giá trị hơn và có thể tạo ra lợi nhuận cao hơn cho nhà đầu tư. Tuy nhiên, chứng khoán cũng có rủi ro liên quan đến sự biến động giá trên thị trường và hoạt động kinh doanh của công ty hoặc tổ chức.

Có nhiều yếu tố quan trọng ảnh hướng đến giá chứng khoán như:

* Tình hình kinh tế: sự thay đổi tình hình kinh tế như như tỉ lệ thất nghiệp, tăng trưởng GDP và lãi suất.
* Tình hình tài chính của công ty hoặc tổ chức: các thông số như lợi nhuận, dòng tiền và tình hình dư nợ của công ty hoặc tổ chức.
* Hoạt động kinh doanh: các hoạt động kinh doanh mới của công ty hoặc tổ chức như mở rộng vào các thị trường mới hoặc làm việc với các đối tác mới.
* Tình hình thị trường chứng khoán: xu hướng giảm hoặc tăng của thị trường chứng khoán (downtrend hoặc uptrend).
* Tin tức và sự kiện: sự kiện và tin tức liên quan đến công ty hoặc tổ chức như thông báo tài sản, quản lý thay đổi hoặc sự kiện chính trị.
* Đánh giá của chuyên gia: các đánh giá của chuyên gia về công ty hoặc tổ chức như xếp hạng của tổ chức tín dụng hoặc các chuyên gia tài chính.

## 1.2. PHÂN LOẠI CHỨNG KHOÁN

### 1.2.1 Cổ phiếu

Cổ phiếu là loại chứng khoán xác nhận quyền sở hữu và lợi ích hợp pháp đối với thu nhập và tài sản của công ty cổ phần. Số vốn đóng góp để thành lập công ty được chia ra thành nhiều phần nhỏ bằng nhau gọi là cổ phần. Người mua cổ phần gọi là cổ đông. Cổ phiếu có thể được phát hành dưới dạng chứng chỉ vật chất hoặc bút toán ghi sổ, chỉ có công ty cổ phần mới có cổ phiếu. Giá trị ban đầu ghi trên cổ phiếu là mệnh giá của cổ phiếu. Mệnh giá là giá trị danh nghĩa. Số tiền nhận được từ khoản góp vốn gọi là cổ tức. Giá cổ phiếu dao động qua các phiên giao dịch trên thị trường chứng khoán và tách rời so với mệnh giá. Cổ phiếu được chia thành 2 loại

* Cổ phiếu thường: là loại cổ phiếu không có kì hạn, tồn tại cùng với sự tồn tại của công ty phát hành ra nó, không có lãi suất cố định, số lãi được chia vào cuối mỗi kì tất toán. Cổ đông nắm giữ cổ phiếu này có quyền bỏ phiếu, mua thêm cổ phiếu mới, tham gia họp đại hội cổ đông.
* Cổ phiếu ưu đãi: gồm 2 loại là cổ phiếu ưu đãi biểu quyết và cổ phiếu ưu đãi tài chính. Cổ phiếu ưu đãi biểu quyết là cổ phiếu dành cho các cổ đông sáng lập. Cổ đông nắm giữ loại cổ phiếu này phải nắm giữ trong một khoảng thời gian nhất định, không được chuyển nhượng hay trao đổi. Cổ phiếu ưu đãi tài chính tương tự cổ phiếu thường nhưng có một số hạn chế như cổ đông nắm giữ cổ phiếu này không được tham gia bầu cử, ứng cử vào hội đồng quản trị, ban kiểm soát của công ty. Bù lại, họ được hưởng ưu đãi về tài chính theo một mức cổ tức riêng biệt, có tính cố định hằng năm, được ưu tiên chia cổ tức và phân chia tài sản còn lại của công ty sau khi thanh lý, giải thể trước cổ phiếu thường.

### 1.2.2 Trái phiếu

Trái phiếu là loại chứng khoán quy định nghĩa vụ của người phát hành phải trả cho người nắm giữ chứng khoán đó một khoản tiền xác định vào những thời hạn cụ thể và theo những điều kiện nhất định. Đây là những chứng khoán nợ, được phát hành dưới dạng chứng chỉ vật chất hoặc bút toán ghi sổ. Trái phiếu bao gồm các loại:

* Trái phiếu vô danh: là loại trái phiếu không ghi tên trái chủ trên cả chứng chỉ và sổ sách của các tổ chức phát hành. Việc chuyển nhượng loại trái phiếu này rất dễ dàng nên thường được giao dịch trên thị trường chứng khoán.
* Trái phiếu ghi danh: là loại trái phiếu ghi tên, địa chỉ của trái chủ trên chứng chỉ và sổ sách của tổ chức phát hành. Loại trái phiếu này ít được giao dịch trên thị trường.
* Trái phiếu chính phủ: là loại trái phiếu do chính phủ phát hành nhằm bù đắp cho việc thâm hụt ngân sách. Đây là loại chứng khoán mà các nhà đầu tư an toàn rất thích vì nó hầu như không có rủi ro thanh toán.
* Trái phiếu công trình: là loại trái phiếu được phát hành để huy động vốn xây dựng các công trình cơ sở hạ tầng hay công trình phúc lợi công cộng.
* Trái phiếu công ty: là loại trái phiếu do công ty phát hành để vay vốn trung và dài hạn. Khi công ty bán trái phiếu thì công ty là người đi vay của người mua, người mua là chủ nợ của công ty. Công ty phải cam kết trả cả lãi và gốc cho trái chủ như đã nêu trong hợp đồng. Trái phiếu công ty bao gồm trái phiếu có đảm bảo, trái phiếu không có đảm bảo, trái phiếu có thể mua lại, …

### 1.2.3 Chứng chỉ quỹ

Chứng chỉ quỹ là chứng khoán được phát hành bởi công ty quản lý quỹ để huy động vốn từ các nhà đầu tư. Vốn được dung để mua bán, kinh doanh các loại chứng khoán khác để kiếm lời, sau đó chia tiền lời đó cho các nhà đầu tư. Có thể phân loại chứng chỉ quỹ đầu tư thành:

* Quỹ đầu tư chung: là loại quỹ mà tất cả mọi người đều có thể tham gia.
* Quỹ đầu tư riêng: là loại quỹ chỉ giới hạn ở một nhóm người nhất định.
* Quỹ đầu tư dạng mở: là loại quỹ mà nhà đầu tư có quyền bán lại chứng chỉ quỹ cho quỹ phát hành chứng khoán liên tục để huy động vốn và sẵn sàng mua lại số chứng khoán mà quỹ đã phát hành ra.
* Quỹ đầu tư cổ phiếu: là loại quỹ chuyên đầu tư vào cổ phiếu.
* Quỹ đầu tư trái phiếu: là loại quỹ chuyên đầu tư vào trái phiếu.
* Quỹ đầu tư hỗn hợp: là loại quỹ đầu tư vào bất kì loại chứng khoán nào nếu chúng có hiệu quả.

### 1.2.4 Phái sinh

Phái sinh là chứng khoán dưới dạng một hợp đồng mà trong đó giá trị phụ thuộc vào một hay nhiều loại tài sản cơ sở. Cụ thể, hợp đồng xác nhận quyền và nghĩa vụ của các bên đối với việc thanh toán tiền, chuyển giao số lượng tài sản cơ sở vào thời điểm xác định trong tương lai với mức giá nhất định được thoả thuận trước. Tài sản cơ sở của phái sinh cũng được quy định là chứng khoán, các chỉ số chứng khoán hoặc tài sản khác (thực phẩm, nông sản, kim loại, …). Các sản phẩm phái sinh cho phép nhà đầu tư *“đặt cược”* vào sự *“tăng”* hoặc *“giảm”* của tài sản cơ sở trong tương lai. Nếu sự thay đổi diễn ra đúng như dự đoán, nhà đầu tư sẽ có lời và ngược lại. Phái sinh gồm 4 loại chính:

* Hợp đồng kì hạn: là thoả thuận pháp lý giữa hai bên tham gia về việc mua và bán một loại tài sản tại một thời điểm nhất định trong tương lai với mức giá được xác định trước tại ngày thực hiện giao dịch.
* Hợp đồng tương lai: là một dạng hợp đồng kì hạn được chuẩn hoá, niêm yết và giao dịch tại thị trường tập trung.
* Hợp đồng quyền chọn: là thoả thuận pháp lý mà trong đó một bên có quyền yêu cầu thực hiện và bên kia có nghĩa vụ phải mua hoặc bán một lượng tài sản cơ sở theo mức giá đã được xác định trước tại hợp đồng trong một khoảng thời gian hoặc tại một thời điểm nhất định trong tương lai.
* Hợp đồng hoán đổi: là thoả thuận pháp lý mà trong đó có hai bên cam kết hoá đổi dòng tiền của công cụ tài chính của một bên với dòng tiền của công cụ tài chính của bên còn lại trong một khoảng thời gian nhất định.

## 1.3 MỤC TIÊU NGHIÊN CỨU

Trong đề tài này, tôi đã đề xuất những yếu tố sau:

* Nghiên cứu về giá cổ phiếu thuộc VN30.
* Nghiên cứu các phương pháp xử lý dữ liệu và các phương pháp đánh giá hiệu suất.
* Nghiên cứu các mô hình học máy như LSTM, SVM, BAYES, ANN và GAN.
* Xử lý, tính toán, đánh giá các kết quả đạt được của mô hình LSTM với các bộ dữ liệu khác nhau.
* Đề xuất mô hình học máy hiệu quả cho việc dự đoán giá cổ phiếu và hướng cải thiện hiệu suất mô hình LSTM

## 1.4 GIỚI HẠN CỦA LUẬN VĂN

Trong luận văn, tôi chỉ sử dụng bộ dữ liệu của một vài cổ phiếu thuộc VN30, là 30 cổ phiếu được niêm yết tại sàn HoSE, có giá trị vốn hoá thị trường và tính thanh khoản cao nhất (tính đến 12/2022) (bộ dữ liệu đã được thu thập trên các trang về chứng khoán) nhằm phục vụ quá trình khảo sát. Những mô hình đã nghiên cứu hoặc đề xuất chỉ được thiết kế phù hợp với bộ dữ liệu này. Để có thể sử dụng bộ dữ liệu mới hoặc tự thu thập thì tôi cần phải tìm hiểu sâu về các thuật toán học máy cũng như các phương pháp xử lý dữ liệu khác. Ngoài ra, bộ dữ liệu có độ mất mát và thiếu hụt.

# CHƯƠNG 2 MÔ HÌNH LSTM

## 2.1 TỔNG QUAN VỀ HỌC MÁY VÀ HỌC SÂU

Cách mạng công nghiệp 4.0 là một khái niệm để miêu tả sự ứng dụng cao của công nghệ thông tin vào các quá trình sản xuất và kinh doanh của doanh nghiệp. Điều này bao gồm việc sử dụng công nghệ như Internet vạn vật, điều khiển tự động, trí tuệ nhân tạo và tương tác trực tuyến để tăng năng suất và hiệu quả của quá trình sản xuất và kinh doanh. Trong đó, học máy đang trở thành một phần không thể thiếu trong lĩnh vực tài chính.

Học máy (Machine Learning) là một lĩnh vực của trí tuệ nhân tạo (Artificial Intelligence) liên quan đến việc nghiên cứu và xây dựng các kĩ thuật cho phép các hệ thống *“học”* tự động từ dữ liệu để giải quyết những vấn đề cụ thể như các máy có thể *“học”* cách phân loại thư điện tử xem có phải thư rác (spam) hay không và tự động xếp thư vào thư mục tương ứng. Học máy hiện nay được áp dụng rộng rãi như máy truy tìm dữ liệu, chẩn đoán y khoa, phát hiện thẻ tín dụng giả, phân tích thị trường chứng khoán, phân loại các chuỗi DNA, nhận dạng tiếng nói và chữ viết, dịch tự động, chơi trò chơi và cử động robot.

Học máy cũng đóng vai trò quan trọng trong lĩnh vực tài chính như:

* Giúp tự động hoá các quá trình tài chính như xử lý giao dịch, tạo hoá đơn, thanh toán và rút tiền. Điều này giúp giảm thiểu những lỗi do con người làm và tăng năng suất.
* Phân tích dữ liệu và dự đoán xu hướng giá của các mã chứng khoán hoặc tín dụng của khách hàng. Điều này giúp các nhà đầu tư và các công ty tài chính quản lý rủi ro và tìm kiếm cơ hội đầu tư tốt hơn.
* Xác thực và kiểm soát rủi ro cho các giao dịch tài chính. Nó có thể được sử dụng để phát hiện các giao dịch giả mạo hoặc không đúng, giúp ngăn chặn các rủi ro tội phạm.
* Tạo ra sản phẩm và dịch vụ tài chính mới và tiên tiến hơn như tài khoản ngân hàng tự động hoá, thẻ tín dụng điện tử hoặc dịch vụ chuyển tiền qua điện thoại.
* Chăm sóc khách hàng bằng các chatbot hoặc hệ thống tự động trả lời hỗ trợ khách hàng. Điều này giúp cho các công ty tài chính cung cấp dịch vụ khách hàng tốt hơn và giảm chi phí cho công ty.
* Đánh giá và quản lý rủi ro cho các khoản vay hoặc giao dịch tài chính khác và đưa ra quyết định hợp lý.

Học sâu (Deep Learning) là một phần của học máy và cũng được ứng dụng trong nhiều lĩnh vực khác nhau như nhận dạng ảnh, giọng nói, văn bản và các lĩnh vực như tài chính. Với sự phát triển của tốc độ máy tính, đặc biệt là khả năng tính toán trên GPU và sự tăng nhanh của dữ liệu cùng với các thư viện như TensorFlow hay Pytorch, việc *“huấn luyện”* các máy *“học”* trở nên dễ dàng hơn.

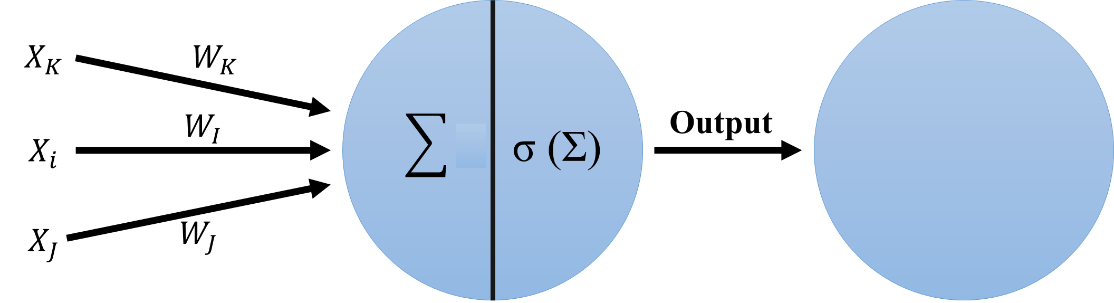
Học sâu sử dụng mạng nơ-ron nhân tạo để học và dự đoán. Mạng nơ-ron nhân tạo giống như một phiên bản mô hình hoá của các quá trình tình cảm và suy nghĩ của con người. Nó bao gồm các tầng được gọi là *“lớp”* và mỗi lớp là một tập hợp các *“nơ-ron”* được kết nối với nhau. Mỗi nơ-ron nhận vào dữ liệu từ các nơ-ron trong lớp trước và sau đó tính toán một giá trị đầu ra. Mạng nơ-ron được huấn luyện bằng cách sử dụng một tập dữ liệu huấn luyện và thay đổi các trọng số của các nơ-ron để tối ưu hoá đầu ra. Khi học sâu, mạng nơ-ron có thể học được các mẫu phức tạp và tìm ra các quan hệ giữa các đặc trưng và nhãn.

Học sâu nói riêng và học máy nói chung cung cấp nhiều loại mô hình và phương pháp học để giải quyết các vấn đề phức tạp khác nhau.

## 2.2 MẠNG NƠ-RON

### Mạng nơ-ron truyền thẳng (ANN)

Mạng nơ-ron nhân tạo được mô hình hóa dựa trên mạng nơ-ron sinh học. Dựa vào đặc điểm đó, các mạng nơ-ron là một mô hình gồm các kết nối của các nút, việc kết nối tương tự như tế bào thần kinh. Với mỗi kết nối sẽ có một giá trị đặc trưng gọi là trọng số. Trọng số là thông số cơ bản cho bộ nhớ dài hạn của mô hình ANN, các trọng số đại diện cho tầm quan trọng của các kết nối giữa các nút nơ-ron. Một mạng nơ-ron *“học”* nghĩa là quá trình điều chỉnh lặp đi lặp lại sau mỗi chu kì huấn luyện của các trọng số sao cho các kết quả thu được từ mô hình là tương đương nhất so với kết quả mong muốn. Mỗi mạng nơ-ron có ba thành phần quan trọng: thông số nút, cấu trúc của mạng và quy tắc huấn luyện [18].



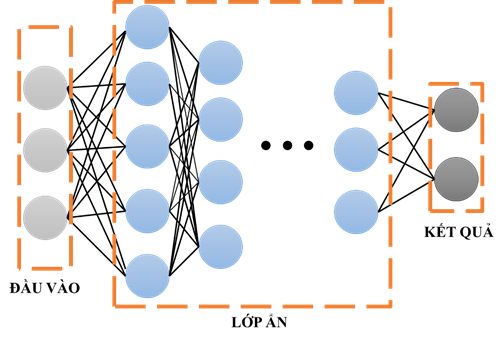
##### **Hình 2.1:** Cấu trúc của một nút

* Thông số nút: bao gồm việc xác định cách xử lý tín hiệu trong các nút, ví dụ như số lượng đầu vào và đầu ra và sự liên kết các nút với nhau, trọng số của từng mỗi nút và các hàm hoạt động. Mỗi nút nhận được nhiều đầu vào từ những nút khác mỗi đầu vào giữa các nút là các kết nối đi kèm với các trọng số liên quan. Khi tổng số đầu vào trọng số vượt quá giá trị ngưỡng của nút, các nơ-ron sẽ được kích hoạt và truyền tín hiệu thông qua các liên kết. Qua đó, các thông tin sẽ được gửi đến các nút lân cận. Quá trình này có thể được biểu diễn theo công thức toán học bên dưới:

|  |  |
| --- | --- |
|  | (2.1) |

trong đó *y* là đầu ra của nút, là hàm hoạt động, *wi* là trọng số của đầu vào *xi* và *T* là giá trị ngưỡng. Chức năng của hàm hoạt động có công thức khác nhau. Tuy nhiên, hàm hoạt động phi tuyến sẽ hữu ích hơn so với những hàm hoạt động tuyến tính trong quá trình huấn luyện [26].

* Cấu trúc mạng: xác định cách các nút được tổ chức và kết nối. Các nút được tổ chức thành các mảng tuyến tính, được gọi là các lớp ẩn. Tuy nhiên, một vài mạng nơ-ron có thể không có lớp ẩn. Thiết kế cấu trúc mạng bao gồm việc xác định số lượng các nút ở mỗi lớp, số lượng các lớp trong mạng và các đường liên kết giữa các nút, trong đó mặc định là các kiến trúc mạng truyền thẳng. Thông thường, những thông số ban đầu được thiết lập một cách ngẫu nhiên và mạng sẽ được tối ưu hóa qua quá trình huấn luyện với các dữ liệu đầu vào khác nhau. Một số cấu trúc mạng truyền thẳng nổi tiếng như perceptron đã được sử dụng rộng rãi hoặc một số mạng phản hồi nổi tiếng bao gồm mạng Hopfield và bản đồ tự tổ chức Kohonen [26].



##### **Hình 2.2:** Cấu trúc của một mạng ANN

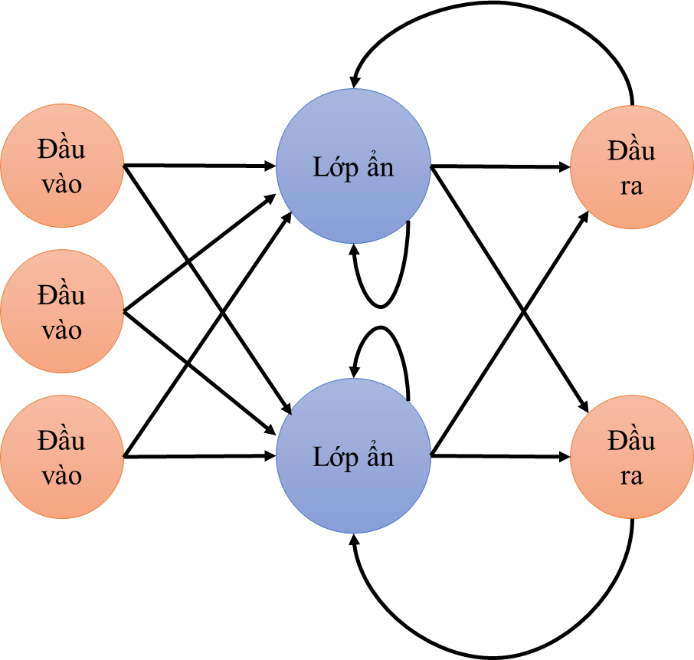
* Quy tắc huấn luyện xác định các trọng số được khởi tạo từ đó điều chỉnh các trọng số đó thông qua các dữ liệu được cho vào huấn luyện. Có nhiều quy tắc huấn luyện khác nhau như học có giám sát, học không giám sát và học tăng cường.

Tuy nhiên, hầu hết các mạng ANN ngày nay đều không phù hợp cho quá trình xử lý dữ liệu dưới dạng tuần tự (như ngôn ngữ, văn phạm, …), đây là một dạng dữ liệu rất phổ biến trong đời sống con người. Vì đặc điểm của các mạng ANN chỉ truyền thẳng mà không hề có phản hồi trở lại của các nút. Do đó, các thông tin đã đi trước đó sẽ bị bỏ qua, điều này sẽ dẫn đến mô hình ANN không thể nhận ra được sự phụ thuộc vào thứ tự và mối liên hệ giữa các tín hiệu chuỗi khi được đưa vào trong mô hình.

### 2.2.2 Mạng nơ-ron hồi quy (RNN)

Tuy nhiên để khắc phục vấn đề giải quyết các dữ liệu tuần tự vào đầu những năm 1980, các nhà nghiên cứu đã tìm ra phương pháp giải quyết bằng cách đưa vào các kết nối phản hồi giữa các lớp và các nút của mạng nơ-ron. Bằng cách này, mối tương quan giữa dữ liệu đầu vào trước đó và các dữ liệu đang được huấn luyện hiện tại đã có thể được huấn luyện cho mạng nơ-ron. Cấu trúc đó được đặt tên là mạng nơ-ron hồi quy.

Tương tự như các thành phần của mạng ANN, tuy nhiên mạng RNN có sự khác biệt với kiến trúc truyền thẳng đó là mô hình được xây dựng theo kiến trúc phản hồi, một vài nút trong mạng sẽ thực hiện công việc như kiến trúc truyền thẳng, một số nút còn lại sẽ có khả năng không chỉ nhận các giá trị từ các nút phía trước mà còn nhận các giá trị trả về từ những nút lân cận hoặc các nút phía sau, cấu trúc của một mạng hồi quy được mô tả như Hình 2.3.



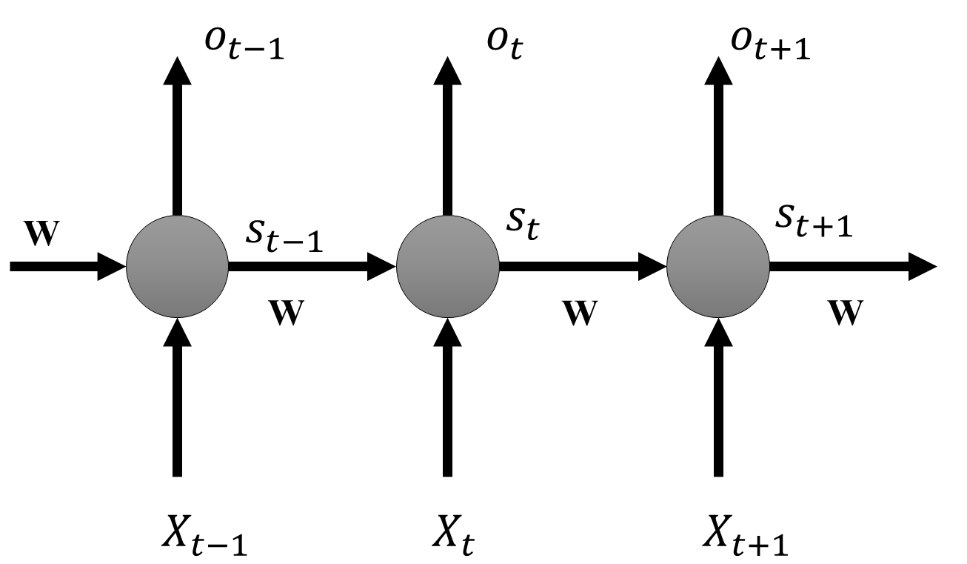
##### **Hình 2.3:** Cấu trúc của một mạng hồi quy

Hình 2.4 mô tả phép triển khai của một RNN. Mạng nơ-ron hồi quy được vẽ dưới dạng nhiều mạng nơ-ron được nối với nhau thành các chuỗi tuần tự. Nguyên tắc hoạt động của mạng nơ-ron có thể được mô tả như sau: nếu có dữ liêụ gồm 5 thành phần, thì mạng nơ-ron sẽ được triển khai sẽ gồm 5 tầng nơ-ron tương ứng với mỗi thành phần. Lúc này, việc tính toán bên trong RNN được thực hiện như sau:

* là đầu vào tại thời điểm t hiện tại
* là bộ nhớ của mạng hồi quy tại thời điểm t. sẽ được tính toán thông qua các trạng thái ẩn trước đó và giá trị đầu vào hiện tại thông qua công thức:

|  |  |
| --- | --- |
|  | (2.2) |

* là giá trị đầu ra tại thời điểm t, một mô hình mạng hồi quy có thể có đưa ra một hoặc nhiều giá trị kết quả, với mỗi kết quả tương đương với từng thời điểm.



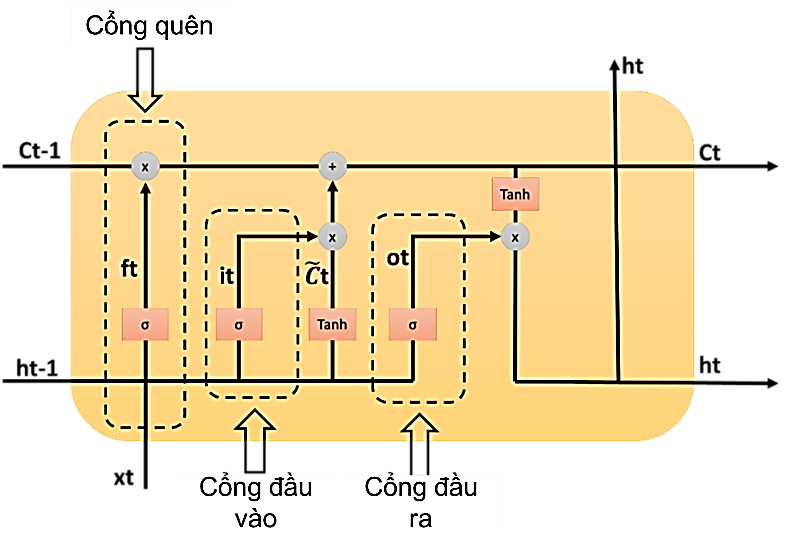
##### **Hình 2.4:** Mô hình chi tiết của một lớp ẩn

Mạng hồi quy mặc dù có khả năng học được các giá trị dữ liệu tuần tự. Tuy nhiên, đối với các bộ dữ liệu có độ dài rất lớn thì trong quá trình huấn luyện, mô hình có thể xảy ra trường hợp mất mát dữ liệu giữa các lớp ẩn từ các dữ liệu phía trước. Điều đó dẫn tới các giá trị tại các nút phía sau cho đến nút cuối cùng sẽ đưa ra các kết quả không chính xác. Về lý thuyết, mô hình RNN hoàn toàn có khả năng xử lý các dữ liệu dài và lớn. Để làm được điều đó mạng cần có khả năng có thể chọn các thông số phù hợp cho việc giải quyết các vấn đề. Nhưng trên thực tế, mô hình RNN dường như không thể thực hiện được điều này [27]. Tuy nhiên năm 1997, mô hình LSTM đã ra đời nhằm khắc phục được các vấn đề mà mạng hồi quy không thể giải quyết.

### 2.2.3 Mạng bộ nhớ dài-ngắn hạn (LSTM)

Mạng LSTM được giới thiệu bởi Hochreiter & Schmidhuber (1997) và đã được cải tiến và phổ biến bởi nhiều nhà nghiên cứu khác nhau bao gồm Felix Gers, Fred Cummins, Santiago Fernandez, Justin Bayer, Daan Wierstra, Julian Togelius, Faustino Gomez, Matteo Gagliolo và Alex Graves [28]. Chúng hoạt động rất hiệu quả trên nhiều vấn đề và hiện đang được sử dụng rộng rãi.

Mạng bộ nhớ dài-ngắn hạn là một mạng nơ-ron được thiết kế để khác phục trường hợp mất mát thông tin của mạng hồi quy khi tín hiệu được đưa qua một lượng lớn các tế bào có trong mạng hồi quy. Vì thế mô hình LSTM là một loại mạng hồi quy có thể xử lý các chuỗi có chiều dài tùy ý và khắc phục được nhược điểm của mô hình RNN. Trong phần sau đây cấu trúc của một tế bào có trong mạng như Hình 2.5:



##### **Hình 2.5:** Cấu trúc của một tế bào trong mạng LSTM

Cấu trúc của một tế bào LSTM được chia làm ba cổng: cổng quên, cổng đầu vàovà cổng đầu ra, trong đó các kí hiệu được mô tả như Bảng 2.2.

###### **Bảng 2.2:** Thông số các giá trị trong tế bào LSTM

|  |  |
| --- | --- |
| **Ký hiệu** | **Thông số** |
|  | Tín hiệu đầu vào |
|  | Tín hiệu đầu ra |
|  | Giá trị lưu trữ cho tế bào tiếp theo |
|  | Giá trị sẽ được lưu trữ |
|  | Cổng đầu vào |
|  | Cổng quên |
|  | Cổng đầu ra |

Các thông tin sẽ được đi qua cổng đầu tiên chính là “*cổng quên*”, chức năng của “*cổng quên”* sẽ xác định những thông tin nào từ tế bào trước nên bị loại bỏ và không cần thiết cho quá trình huấn luyện. Nếu giá trị kết quả là '0', thông tin của tế bào trước đó sẽ bị xóa. Ngược lại, nếu giá trị là '1' thì thông tin đó sẽ được cho phép đi qua cổng quên. Đầu ra của cổng được mô tả tổng quan thông qua phương trình sau:

|  |  |  |
| --- | --- | --- |
|  |  | (2.3) |

Tiếp đó sẽ là “*cổng đầu vào*”, đây là cổng hoạt động với sự hợp tác của hai hàm hoạt động song song với nhau. Hàm *Tanh* sẽ đưa ra các giá trị được chọn, trong khi hàm *Sigmoid* sẽ hoạt động tương tự như cổng quên đã đề cập bên trên. Do đó, trạng thái của giá trị tại vị trí “*cổng đầu vào*” là sự kết hợp các giá trị tính năng của các giá trị trong các tế bào trước và các giá trị được cập nhật thêm của tế bào hiện tại, công thức của “*cổng đầu vào*” có thể được mô tả theo phương trình:

|  |  |  |
| --- | --- | --- |
|  |  | (2.4) |

Cổng cuối cùng là “*cổng đầu ra*”. Tổng quát nhất, “*cổng đầu ra*” sẽ quyết định giá trị ở đầu ra của các tế bào hiện tại, các giá trị này sẽ được đưa đến các tế bào tiếp theo, đồng thời cũng sẽ được hiển thị ra ngoài như là một kết quả cuối cùng, công thức cho “*cổng đầu ra*” được hiển thị trong phương trình 2.5 và 2.6 tương ứng bên dưới:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | |  | | --- | |  | |  | | (2.5)  (2.6) |

Thông qua các cổng này, mạng LSTM có thể khắc phục vấn đề của mạng hồi quy và tránh được xung đột của các trọng số đầu vào và các trọng số đầu ra. Chi tiết hơn, mạng LSTM có thể sử dụng thông tin của tín hiệu đầu vào trước đó để quyết định khi nào tín hiệu sẽ được lưu lại hoặc bị ghi đè thông tin trong bộ nhớ của tế bào và tín hiệu đầu ra sẽ được sử dụng để quyết định khi nào truy cập các giá trị đã lưu trữ nhằm ngăn các đơn vị khác bị xáo trộn bởi các giá trị lưu trữ đó. Tuy nhiên, khi các tín hiệu lỗi đi vào các tế bào (vào các thời điểm khác nhau), *cổng đầu ra* sẽ phải tìm hiểu những lỗi nào cần được để lại trong *cổng quên* và *cổng đầu vào* sẽ phải loại bỏ các lỗi bằng cách thay đổi tỉ lệ độ lớn giá trị.

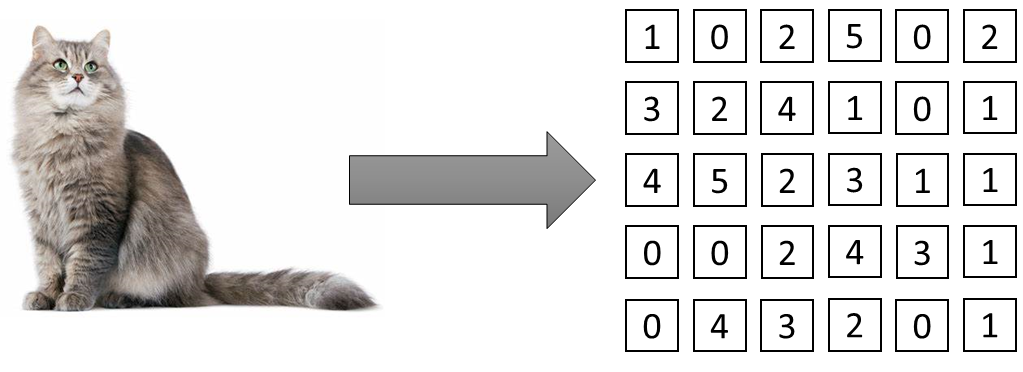
Về cơ bản, các cổng của mạng LSTM sẽ có chức năng cho phép và ngăn chặn quyền truy cập vào các giá trị lỗi thông qua *cổng quên*. Tuy nhiên, việc sử dụng các loại cổng trên không phải lúc nào cũng cần thiết; đôi khi chỉ cần thiết sử dụng *cổng đầu vào* là phù hợp cho quá trình huấn luyện. Tuy nhiên, ngay cả trong các trường hợp sử dụng một cổng, việc sử dụng các *cổng đầu ra* cũng có mang lại rất nhiều lợi ích, ví dụ như các *cổng đầu ra* sẽ ngăn chặn quá trình mạng lưu trữ các thông tin có độ trễ thời gian dài (thường sẽ gây ảnh hưởng đến hiệu suất của mạng).

### 2.2.4 Mạng nơ-ron tích chập CNN

Mạng nơ-ron tích chập là một mạng nơ-ron truyền thẳng. Mô hình này được giới thiệu vào năm 1990 và được thiết kế chuyên dụng để giải quyết các vấn đề liên quan đến hình ảnh [3]. Ngoài ra, mô hình CNN có thể được sử dụng để giải quyết các vấn đề về máy học hoặc khai thác dữ liệu, trong đó các tín hiệu đầu vào có thể được biểu thị bằng một hình ảnh hoặc một tập hợp các hình ảnh. Mô hình này có thể được hình dung như một phiên bản sửa đổi của mô hình mạng nơ-ron đa perceptron [30].

Mạng nơ-ron tích chập có thể được huấn luyện với các dạng tín hiệu thô trực tiếp mà không cần phải trích xuất các tính năng đặc trưng của từng tín hiệu. Do đó, để có thể huấn luyện được một mạng CNN, bộ dữ liệu huấn luyện cần phải lớn vì có hàng triệu tính năng có thể dùng để huấn luyện từ một dữ liệu cụ thể. Chính vì khả năng có thể học được rất nhiều các tính năng, việc tính toán sẽ diễn ra rất phức tạp và thời gian hoàn thành huấn luyện sẽ rất dài. Để khắc phục nhược điểm đó, cần phải có các bộ xử lý đồ họa đặc thù (GPU) để mạng có thể được huấn luyện trong thời gian có thể chấp nhận.

Một mạng nơ-ron tích chập bao gồm các thành phần chính: lớp chập, lớp pooling và lớp fully-connect.



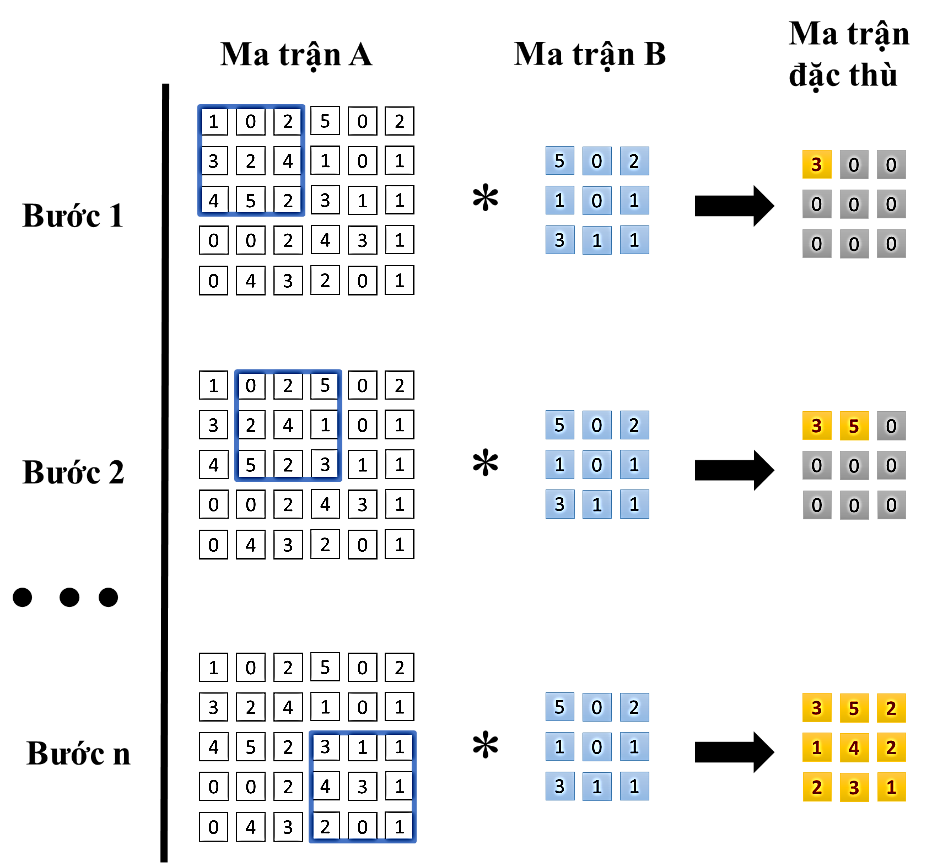
##### **Hình 2.6:** Phương pháp nhận diện của máy tính với một hình ảnh

#### 2.2.4.1 Lớp chập

Trong lớp chập của mạng CNN, các thông số như ma trận chập, độ trượt (stride) và lớp phủ (padding) sẽ cần phải được chú ý đến trong quá trình thiết kế cho một mạng nơ-ron tích chập.

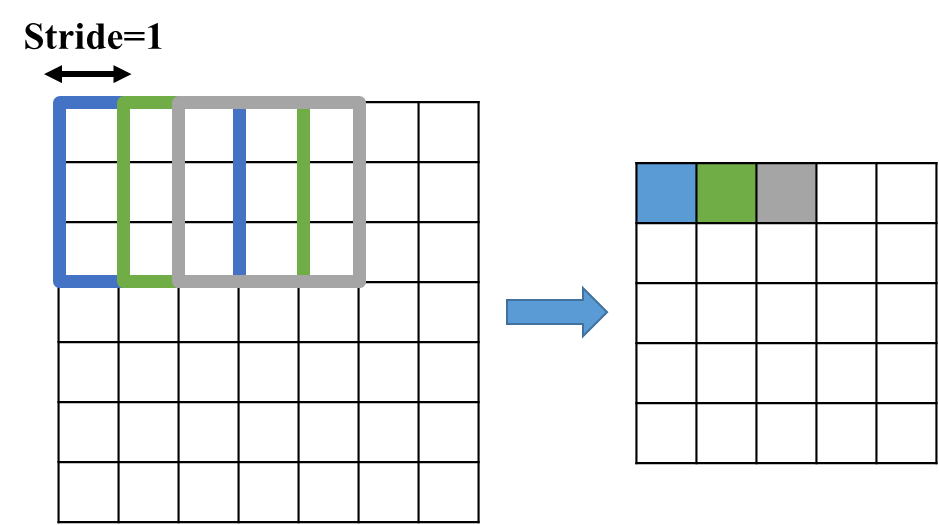
Đầu tiên là ma trận chập, khác với các mạng đã đề cập ở các mô hình mạng truyền thẳng và mạng hồi quy, các trọng số của mạng sẽ tương đương các giá trị bên trong mỗi ma trận chập (Kernel). Nguyên lý hoạt động chủ yếu của các lớp chập sẽ dựa trên các tính toán của tích chập.

Tích chập là một loại phép toán tuyến tính chuyên biệt được sử dụng để trích xuất các tính năng đặc trưng. Để thực hiện tích chập, một ma trận A sẽ được chập bởi một ma trận B để ra một ma trận mới chứa các đặc thù của ma trận A, phương pháp thực hiện được mô tả như Hình 2.7.

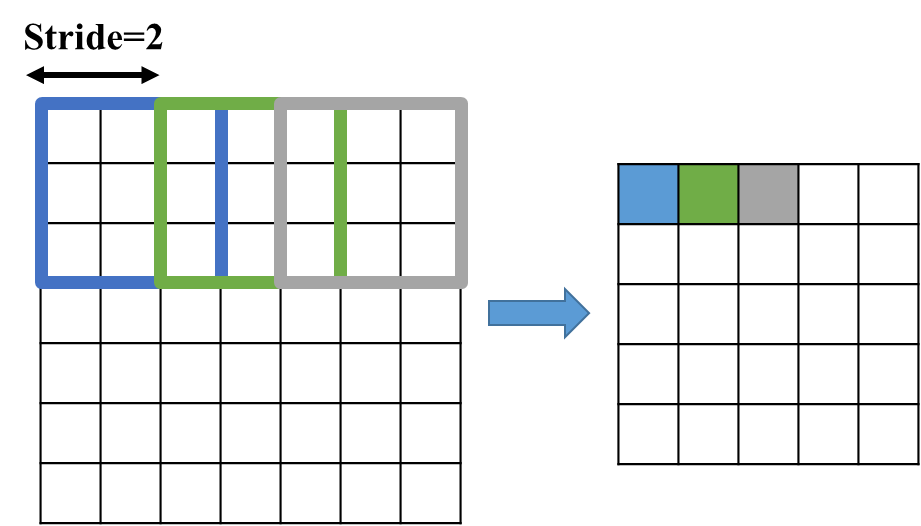


##### **Hình 2.7:** Các bước tính toán của phương pháp tích chập

Vùng mà ma trận B quét tới được gọi là vùng tiếp nhận. Ma trận đó sẽ trượt qua các ma trận A (các nút nơ-ron của mạng), trong quá trình trượt, hai ma trận sẽ nhân các phần tử của ma trận B với các phần tử trong khu vực tiếp nhận của ma trận A. Quá trình trượt bộ lọc được gọi là “*stride*”. Những ma trận kết quả được tạo thành sẽ được gọi là bản đồ tính năng. Trong quá trình chập sẽ có một số lượng lớn bản đồ tính năng được tạo thành bằng phương pháp chập với các ma trận chập khác nhau. Các bản đồ tính năng đấy sẽ đại diện cho một đặc thù như cạnh, đường cong [31].

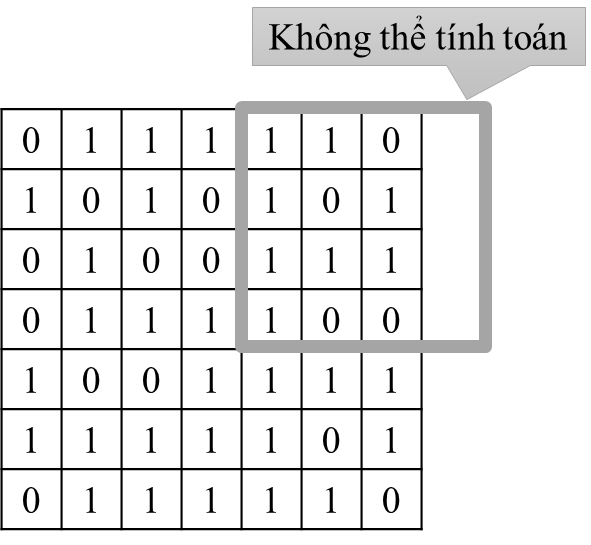


##### **Hình 2.8:** Ma trận chập với độ trượt bằng một



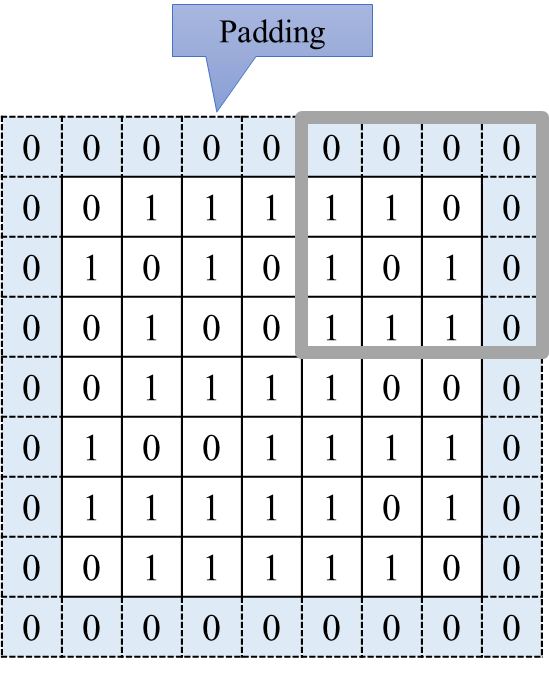
##### **Hình 2.9:** Ma trận chập với độ trượt bằng hai

Thông số tiếp theo đó là độ trượt (Stride), stride là độ lớn mà bộ lọc được phép trượt trên ma trận đầu vào. Thông số này sẽ điều khiển các ma trận chập trượt trên toàn bộ ma trận và kết quả thu của bản đồ tính năng cũng sẽ thay đổi cho từng giá trị trượt. Hình 2.8 và 2.9 mô tả tổng quan quá trình trượt của ma trận chập với độ trượt bằng một và hai. Trong quá trình trượt với các giá trị lớn hoặc các bản đồ tính năng đã được tính toán qua nhiều lớp chập, khi đó các kích thước của các bản đồ tính năng sẽ là rất nhỏ. Khi tiếp tục tính toán trong lớp chập sẽ không thể tránh khỏi trường hợp các ma trận chập sẽ bị trượt ra ngoài (Hình 2.10), khi ấy rìa của ma trận chập sẽ không có giá trị để có thể tính toán được.



##### **Hình 2.10:** Ma trận chập bị trượt ra ngoài

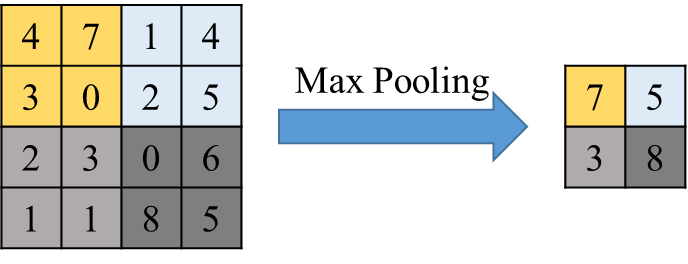
Để khắc phục điều đó, các ma trận bị chập có thể thêm vào các giá trị ‘0’ bao xung quanh chính nó, qua đó quá trình chập sẽ diễn ra mà không gặp bất kì khó khăn nào, người ta gọi đó là phương pháp “Zero padding”. Phương pháp này có tác dụng loại bỏ việc giảm kích thước và duy trì kích thước đầu vào ở đầu ra. Trong các mạng có nhiều lớp, cách tiếp cận này có thể cần thiết để ngăn kích thước của đầu ra bị thu nhỏ quá nhiều [31].



##### **Hình 2.11:** Lớp padding phủ xung quanh hỗ trợ ma trận chập bị trượt ra ngoài có thể tính toán được các tính năng

#### 2.2.4.2 Lớp Pooling

Các lớp pooling thường được sử dụng sau khi các tín hiệu đã đi qua lớp chập. Mục đích của lớp này là đơn giản hóa thông tin xuất ra bởi lớp tích chập. Tóm lại, ý nghĩa của lớp pooling sẽ cho ra một bản đồ tính năng cô đọng hơn từ bản đồ tính năng được tạo ở bước trước. Một trong những phương pháp gộp thông tin đơn giản đó là pooling cực đại (max pooling) và pooling trung bình toàn cục (global average polling).

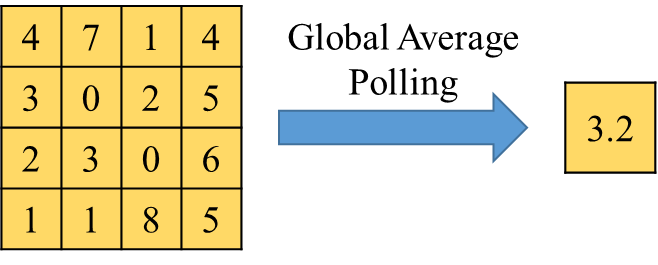


##### **Hình 2.12:** Quá trình max pooling

Phương pháp phổ biến nhất của phương pháp gộp đó là max pooling, trích xuất từ các khu vực của bản đồ tính năng đầu vàovà xuất ra giá trị lớn nhất trong mỗi khu vực đồng thời loại bỏ tất cả các giá trị khác (Hình 2.12). Max polling với ma trận có kích thước 2 × 2 và độ trượt bằng hai thường được sử dụng trong thực tế. Điều này làm giảm kích thước trong mặt phẳng của bản đồ tính năng theo hệ số 2. Không giống như chiều cao và chiều rộng, kích thước chiều sâu của bản đồ đối tượng là không thay đổi [32].

Một phương pháp gộp khác đáng chú ý là phép gộp trung bình toàn cục. Phương pháp này sẽ biến đổi bản đồ tính năng có kích thước chiều cao × chiều rộng được lấy mẫu thành mảng có kích thước 1 × 1 bằng cách chỉ cần lấy giá trị trung bình của tất cả các phần tử trong mỗi bản đồ tính năng, tuy nhiên độ sâu của bản đồ đối tượng là không đổi. Phương pháp này thường chỉ được áp dụng một lần trước khi các bản đồ tính năng được đi vào lớp được kết nối đầy đủ. Việc thực hiện phép gộp trung bình toàn cục sẽ có những ưu điểm như sau:

* Giảm số lượng các tham số có thể được huấn luyện
* Cho phép CNN chấp nhận các tín hiệu đầu vào có kích thước khác nhau.

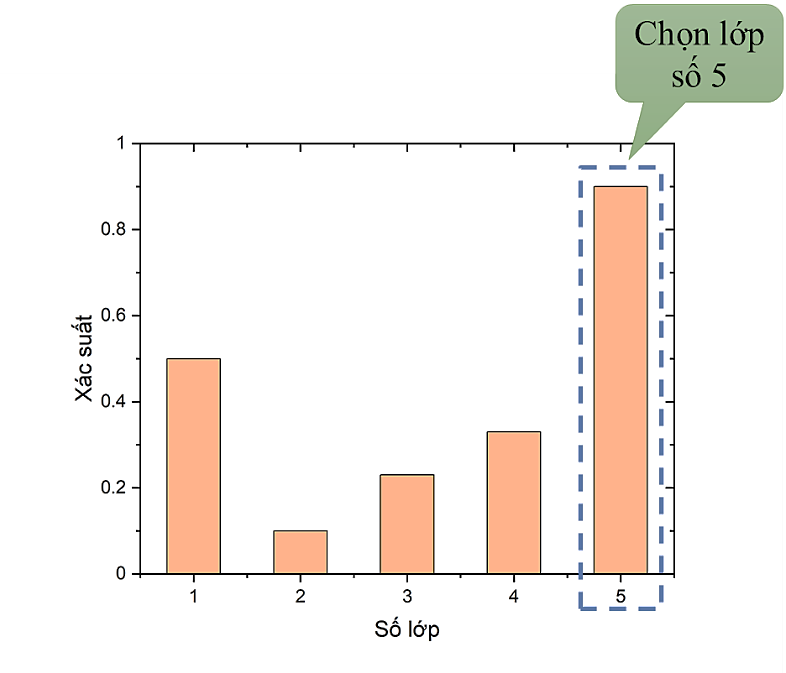


##### **Hình 2.13:** Quá trình global average polling

#### 2.2.4.3 Lớp kết nối đầy đủ (Fully-Connect)

Tại lớp kết nối đầy đủ, các bản đồ tính năng của lớp tích chập hoặc lớp gộp sẽ được làm phẳng. Chi tiết hơn, các bản đồ tính năng sẽ được chuyển thành mảng một chiều (1D) và được kết nối với một hoặc nhiều lớp kết nối đầy đủ, trong đó các tính năng đầu vào được kết nối với các đầu ra thông qua trọng số.

Ví dụ: Một vectơ 1 chiều có 5 giá trị lần lượt là: 0,5 0,1 0,23 0,33 0,9 sẽ tương ứng với 5 lớp được trích xuất ra bởi lớp được kết nối đầy đủ, 0,5 là xác suất của đầu vào được cung cấp cho mô hình thuộc lớp 1; 0,1 là xác suất của nó thuộc lớp 2 và tương tự cho các giá trị tiếp theo. Vì xác suất của đầu vào thuộc lớp 5 là cao nhất (có giá trị bằng 0.9), mô hình CNN sẽ ra quyết định tín hiệu đầu vào tương ứng sẽ thuộc lớp thứ 5 [31]. Để mô hình có thể thực hiện được việc ra quyết định thì cần phải sử dụng hàm *softmax*. Hàm sẽ chuẩn hóa các giá trị thực đầu ra từ lớp kết nối đầy đủ đến xác suất của lớp mục tiêu, trong đó mỗi giá trị nằm trong khoảng từ 0 đến 1 và tất cả các giá trị tổng bằng 1. Từ đó sẽ quyết định giá trị xác suất nào là lớn nhất và gắn nhãn hoặc lớp cho các tín hiệu đầu vào tương ứng



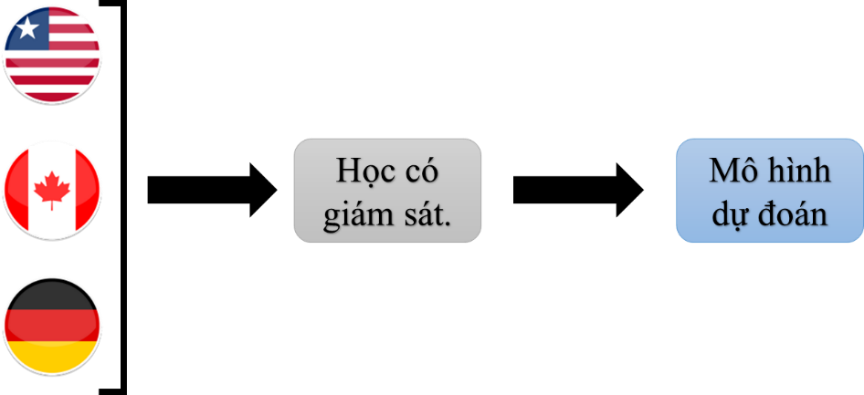
##### **Hình 2.13:** Mô hình quyết định của lớp kết nối đầy đủ

## 2.3 CÁC PHƯƠNG PHÁP HUẤN LUYỆN MẠNG NƠ-RON

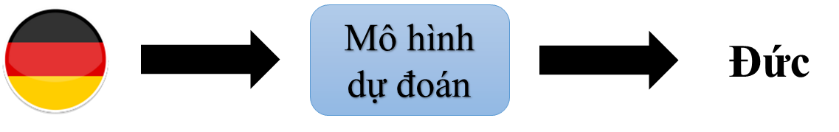
Trong quá trình huấn luyện, các trọng số được điều chỉnh để đạt được các giá trị mong muốn, để có thể thực hiện quá trình đó các mạng nơ-ron cần phải có phải được huấn luyện. Quá trình huấn luyện có thể được phân thành hai loại chính: học giám sát và học không giám sát, học tăng cường.

### 2.3.1 Phương pháp học có giám sát

Học có giám sát, các dữ liệu huấn luyện sẽ bao gồm đầu vào và kết quả tương ứng được cung cấp từ trước, chính vì vậy việc có một tập dữ liệu chính xác là vô cùng quan trọng cho quá trình này. Trọng số sẽ được điều chỉnh từ đó giảm thiểu độ sai số giữa đầu ra của mạng và kết quả tương ứng đã có trước. Một mô hình mạng nơ-ron được huấn luyện một cách lý tưởng phải có sai số giữa kết quả đầu ra và các kết quả mong muốn ở mức thấp nhất. Đối với các mô hình sử dụng học có giám sát, mô hình phải được đào tạo trước, sau đó mô hình sẽ tạo ra các kết quả cho một loạt các đầu vào, các trọng số sẽ được cố định khi sai số là nhỏ nhất hoặc khi quá trình huấn luyện kết thúc. Sau đó, mô hình mạng có thể được đưa vào hoạt động [26].



##### **Hình 2.14:** Mô hình huấn luyện học có giám sát

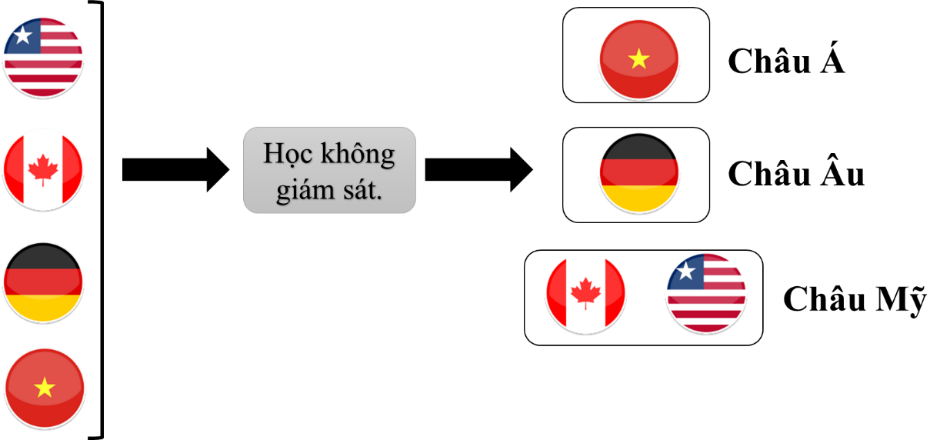


##### **Hình** **2.15:** Mô hình kiểm tra phương pháp học có giám sát

### 2.3.2 Phương pháp học không giám sát

Học không giám sát không sử dụng giá trị đầu ra mục tiêu từ bộ dữ liệu. Từ đó mạng nơ-ron sẽ tự phát triển và xây dựng đại diện của đầu vào có thể được sử dụng để đưa ra quyết định, dự đoán đầu ra trong tương lai. Nói cách khác, mạng nơ-ron sẽ cố gắng để tạo ra các mô hình phù hợp cho bộ dữ liệu đó hoặc tìm ra xu hướng trong dữ liệu đầu vào [33].

Trong quá trình học này, các dữ liệu đầu vào thường được giả định là các mẫu độc lập. Cho một ví dụ, chẳng hạn như hình ảnh của một ảnh sẽ được các đối tượng tại các địa điểm chiếu sáng bởi các góc chiếu khác nhau, dẫn tới tạo ra các hình ảnh khác nhau trong đó có một số hình ảnh tốt nhất cho quá trình muốn thực hiện. Từ đó mô hình sẽ phân cụm chúng lại với nhau [33].



##### **Hình 2.16:** Phương pháp học không giám sát

### 2.3.3 Phương pháp học tăng cường



##### **Hình 2.17:** Mô tả phương pháp học tăng cường

Trong quá trình học tăng cường các mô hình mạng tương tác với các dữ liệu huấn luyện bằng cách sinh các phương pháp giải quyết khác nhau: A1, A2, ... Những hành động này sẽ có ảnh hưởng tuần tự tới bộ dữ liệu, với mỗi hành động, các kết quả nhận được sẽ trả về một phần thưởng (hoặc trừng phạt) tới mô hình, các kết quả nhận được sẽ được gọi là R1, R2, ... tương ứng với A1, A2 (Hình 2.18). Mục tiêu của phương pháp học tăng cường nhắm tới đó là học cách hành động theo cách tối đa hóa phần thưởng (hoặc giảm thiểu hình phạt) trong suốt quá trình huấn luyện. Học tăng cường thường được áp dụng trong các lĩnh vực có liên quan đến lý thuyết quyết định (trong thống kê và khoa học quản lý) và lý thuyết điều khiển (trong kỹ thuật) [34].

## THUẬT TOÁN TỐI ƯU HÓA

Thuật toán tối ưu hóa được áp dụng để đưa giá trị của hàm mục tiêu sao cho gần bằng giá trị cực tiểu cục bộ bằng cách thay đổi các thông số. Quá trình thay đổi các thông số phải làm cho giá trị đạo hàm có chiều hướng thu nhỏ tức là giá trị của đạo hàm của mục tiêu sẽ phải nhỏ dần sao cho bằng 0, đạo hàm đó còn được gọi là Dựa vào độ lớn dữ liệu được sử dụng cho quá trình tính toán, có thể chia thuật toán tối ưu hóa thành ba phương pháp:

* Batch gradient descent: phương pháp này sử dụng toàn bộ dữ liệu để tính toán chỉ trong một lần cập nhật, công thức tính toán của phương pháp được mô tả như sau:

|  |  |  |
| --- | --- | --- |
|  |  | (2.7) |

* Stochastic gradient descent: khác với phương pháp batch gradient descent, phương pháp này sẽ cập nhật các thông số cho từng dữ liệu một, công thức tính toán được mô tả như sau:

|  |  |
| --- | --- |
|  | (2.8) |

* Mini-batch gradient descent: phương pháp này cập nhật các thông số của hàm theo từng bộ dữ liệu nhỏ trong một lần cập nhật và số lần cập nhật có thể có n, công thức tổng quan được mô tả như sau:

|  |  |  |
| --- | --- | --- |
|  |  | (2.9) |

Tuy nhiên các phương pháp trên vẫn còn có một vài nhược điểm khi sử dụng: việc lựa chọn giá trị tốc độ học thích hợp sẽ rất khó khăn; tốc độ học quá nhỏ sẽ khiến quá trình hội tụ diễn ra rất chậm, kéo dài thời gian huấn luyện. Ngược lại, khi độ học quá lớn sẽ làm cản trở sự hội tụ. Do đó trong quá trình huấn luyện, mô hình có thể không học được các tín hiệu mới. Ngoài ra, một nhược điểm khác là quá trình giảm thiểu hàm sai số cho các mạng nơ-ron là quá trình tránh bị mắc kẹt trong vùng cực tiểu cục bộ.

Trong các mục tiếp theo, tôi xin giới thiệu một số thuật toán đã được sử dụng rộng rãi bởi cộng đồng Deep Learning nhằm khắc phục những nhược điểm đã nêu.

### 2.4.1 Thuật toán Momentum

Để cải tiến cho thuật toán Stochastic gradient descent, vốn rất dễ hội tụ ở vùng cực tiểu cục bộ, do độ cập nhật mỗi lần sẽ rất nhỏ (như Hình 2.19). Thuật toán Momentum được đề xuất với mục đích tăng tốc quá trình giảm thiểu hàm mục tiêu bằng cách tăng độ lớn của mỗi lần cập nhật như công thức:

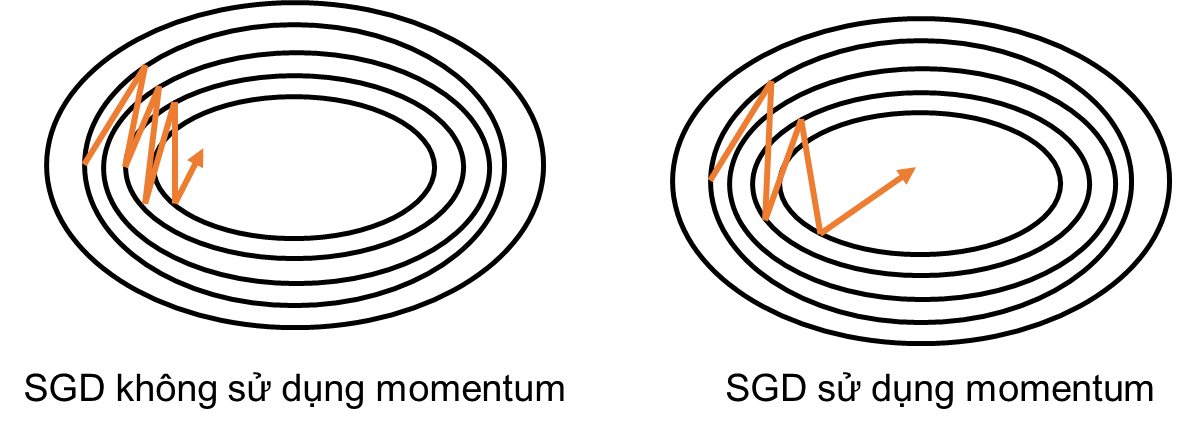
|  |  |
| --- | --- |
|  | (2.10) |

trong đó thường được đặt xấp xỉ bằng 0.9

Dựa vào công thức (2.10) ta có công thức tổng quan cho thuật toán Momentum như sau:

|  |  |  |
| --- | --- | --- |
|  |  | (2.11) |

Thuật toán sẽ có những cập nhật lớn đối với các dữ liệu có sự tương đồng với nhau, tuy nhiên đối với các dữ liệu không tương đồng thì thuật toán sẽ khó có thể huấn luyện được do các cập nhật sẽ rất nhỏ. Do đó, thuật toán Momentum sẽ làm cho mô hình nhanh hội tụ.



##### **Hình 2.19:** Mô tả tốc độ hội tụ thuật toán SGD khi không và có sử dụng momentum

### 2.4.2 Thuật toán Adagrad

Adagrad là thuật toán để tối ưu hóa, thuật toán này tối ưu hóa thông qua việc điều chỉnh tốc độ học, thực hiện thay đổi lớn cho các tham số ít được cập nhật và các thay đổi nhỏ cho các tham số thường xuyên được cập nhật. Đây là thuật toán phù hợp cho việc xử lý các bộ dữ liệu mà độ tương tự thấp hay thưa. Tác giả Dean và các cộng sự đã phát hiện ra rằng thuật toán này có độ ổn định cao và sử dụng nó để đào tạo mạng lưới thần kinh quy mô lớn tại Google. Hơn nữa, tác giả Pennington và các cộng sự đã sử dụng Adagrad để đào tạo hệ thống GloVe, nhờ vào khả năng cập nhật thay đổi lớn cho các từ ít được sử dụng khi so với những từ thường xuyên [5].

Thuật toán Adagrad khác với các thuật toán tối ưu khác, các thông số *θ* sẽ được cập nhật thông qua độ học giống nhau, tuy nhiên Adagrad sẽ sử dụng độ học khác nhau tại mỗi thời điểm huấn luyện để cập nhật *θ*. Quá trình này có thể được làm rõ thông qua công thức:

|  |  |
| --- | --- |
|  | (2.12) |

Trong quá trình cập nhật tại thời điểm t, tỉ lệ học tại thời điểm thứ t cho toàn bộ các thông số sẽ dựa vào các giá trị gradient ở các bước cập nhật các thông số ở trước thời điểm t, từ đó công thức (12) được biến đổi như sau:

|  |  |  |
| --- | --- | --- |
|  |  | (2.13) |

trong đó là ma trận chéo với các giá trị chéo bên trong là tổng của các căn bậc hai của các gradient của thông số tại thời điểm t. Thông số là số dương nhỏ nhất và được đặt bằng để hạn chế trường hợp chia với số 0.

Với là một ma trận chéo, ta có thể vector hóa công thức (2.13) sẽ biến đổi thành dạng như sau:

|  |  |
| --- | --- |
|  | (2.14) |

Một trong những điểm mạnh của thuật toán Adagrad là giúp loại bỏ sự cần thiết phải điều chỉnh độ học theo cách thủ công. Hầu hết các triển khai sử dụng một giá trị mặc định của 0.01.

Tuy nhiên, nhược điểm của thuật toán là sự tích tụ của các gradient bình phương ở trong mẫu số: Vì sau mỗi thời điểm *t*, các giá trị tích lũy tiếp tục sẽ ngày càng lớn dần trong quá trình huấn luyện. Điều này dẫn tới độ học nhỏ dần và cuối cùng trở nên cực kì nhỏ, tại thời điểm đó thuật toán sẽ không còn học được các vấn đề mới.

### 2.4.3 Thuật toán Adadelta

Thuật toán Adadelta là một thuật toán mở rộng của Adagrad nhằm tìm cách giảm độ học khi thông số tăng dần. Khác với Adagrad luôn tích lũy tất cả các gradient trong quá khứ, Adadelta sẽ hạn chế các giá trị gradient tích lũy trong quá khứ ở một mức cố định, ta gọi mức đó là [6].

Thuật toán Adadelta sẽ thay giá trị của công thức (2.13) thành :

|  |  |  |
| --- | --- | --- |
|  |  | (2.15) |

trong đó được gọi là “running average” tại thời điểm *t* và giá trị này phụ thuộc vào các giá trị trung bình gradient trước thời điểm t và gradient hiện tại, công thức tổng quan của tham số này được mô tả như sau:

|  |  |  |
| --- | --- | --- |
|  |  | (2.16) |

trong đó sẽ được gán giá trị là 0.9.

Công thức cập nhật của thuật toán SGD theo tham số như sau:

|  |  |
| --- | --- |
|  | (2.17) |
|  | (2.18) |

Dựa vào công thức (2.15) và (2.18), ta suy ra được:

|  |  |
| --- | --- |
|  | (2.19) |

Vì mẫu số có thể xem là hiệu dụng lỗi bình phương (RMS) của gradient, vì thế mẫu số có thể biến đổi theo dạng viết tắt như sau:

|  |  |  |
| --- | --- | --- |
|  |  | (2.20) |

Cần lưu ý rằng các bản cập nhật phải có các đơn vị giả định tương tự như tham số. Vì thế, trước tiên cần phải xác định một trung bình giảm dần theo cấp số nhân, lần này không phải của gradient bình phương mà là các bản cập nhật tham số bình phương:

|  |  |  |
| --- | --- | --- |
|  |  | (2.21) |

Từ đó lỗi hiệu dụng được cập nhật theo công thức sau:

|  |  |  |
| --- | --- | --- |
|  |  | (2.22) |

Bởi vì giá trị hiệu dụng chưa xác định, giá trị cập nhật hiệu dụng của tham số cho đến khi bước thời gian trước đó. Thay thế tỷ lệ học tập *η* trong quy tắc cập nhật trước đó bằng cuối cùng mang lại quy tắc cập nhật Adadelta:

|  |  |
| --- | --- |
|  | (2.23) |
|  | (2.24) |

### 2.4.4 Thuật toán Adaptive Moment Estimation (Adam)

Thuật toán Adam là phương pháp được sử dụng để tính độ học tương ứng cho mỗi tham số [7]. Ngoài việc lưu trữ trung bình bình phương gradient trong các thời điểm trước , trong thuật toán này được gọi là là , như thuật toán Adadelta, thuật toán còn đồng thời duy trì trung bình độ dốc quá khứ như momentum , trong thuật toán này gọi là là :

|  |  |
| --- | --- |
|  | (2.25) |
|  | (2.26) |

trong đó, và là ước tính của thời điểm đầu tiên và thời điểm thứ hai của các gradient tương ứng. Khi và được khởi tạo thành các vectơ không, thuật toán Adam sẽ có khuynh hướng đưa các giá trị và xấp xỉ bằng với 1. Để tránh điều này, thuật toán phải tính toán các ước tính thời điểm đầu tiên và thứ hai theo công thức sau:

|  |  |
| --- | --- |
|  | (2.27) |
|  | (2.28) |

Sau đó các giá trị này sẽ được sử dụng để cập nhật các thông số như công thức:

|  |  |
| --- | --- |
|  | (2.29) |

trong đó có giá trị là 0.9 và có giá trị là 0.999 và có giá trị là .

**CHƯƠNG 3  
PHƯƠNG PHÁP XỬ LÝ DỮ LIỆU**

## 3.1 TIỀN XỬ LÝ VÀ TRÍCH XUẤT TÍNH NĂNG

Việc trích xuất các tính năng của các tín hiệu điện não là quá trình cực kì quan trọng và cần thiết. Thông qua các tính năng được trích xuất, các tín hiệu điện não mới được mô tả một cách rõ ràng và ngắn gọn hơn so với tín hiệu ban đầu, từ đó giúp giảm thiểu chi phí tính toán cũng như tăng cường độ chính xác của kết quả cuối cùng. Nói cách khác, quá trình trích xuất tính năng sẽ làm tăng tốc độ huấn luyện cho các dữ liệu khi sử dụng các thuật toán học máy có giám sát nhằm phục vụ quá trình phân loại, ở đây là phân loại cảm xúc. Tuy nhiên để có thể thực hiện quá trình này mà hiệu quả đạt được là cao nhất, thì cần phải có một bộ dữ liệu mẫu chuẩn xác và uy tín. Chính vì thế, tôi sẽ chọn bộ dữ liệu nhằm phục vụ cho luận văn của mình là bộ dữ liệu DEAP của Sander Koelstra và các cộng sự.

### 3.1.1 Bộ dữ liệu DEAP [5]

Bộ dữ liệu DEAP là một bộ dữ liệu các tín hiệu điện não đã được công bố và sử dụng rộng rãi trong các nghiên cứu [38, 39, 40], các bản ghi tín hiệu sinh lý và các dữ liệu ngoại vi khác được thu từ những người tham gia khi họ xem một tập hợp các đoạn phim âm nhạc được thiết kế để gợi ra những cảm xúc khác nhau sau đó họ sẽ được cho phép đánh giá các đoạn phim theo mức độ từ 1-9 cho các loại cảm xúc bao gồm: “Arousal, Valence, Ấn tượng, Yêu thích”.

Quá trình đo được thực hiện ở hai môi trường thí nghiệm khác nhau (Geneva và Twente); Hệ thống Biosemi ActiveTwo được sử dụng phục vụ quá trình thu tín hiệu điện não và các tín hiệu sinh lý khác. Trong quá trình thu dữ liệu, những người tình nguyện (độ tuổi trung bình là 26.9) sẽ được xem 40 đoạn phim âm nhạc với mỗi đoạn dài một phút với khoảng cách xấp xỉ 1 mét đối với màn hình, các đoạn phim được trình chiếu thông qua màn hình 17 inch (12801024, 60 Hz); Ngoài ra nhằm giảm nhiễu sinh ra do chuyển động cơ đặc biệt là các chuyển động của mắt, tất cả các đoạn phim đều được thu gọn kích thước và hiển thị ở độ phân giải 800600. Âm lượng sẽ được đặt ở mức tương đối lớn; tuy nhiên, những người tham gia sẽ được hỏi trước khi tiến hành quá trình đo rằng âm lượng có gây khó chịu hay không và sẽ được điều chỉnh khi cần thiết.

Tín hiệu EEG được ghi lại ở tốc độ lấy mẫu 512 Hz thông qua 32 điện cực AgCl hoạt động (được bố trí theo hệ thống 10-20). Ngoài ra còn có mười ba tín hiệu sinh lý ngoại vi khác như: “nhiệt độ cơ thể, huyết áp, ...” cũng được ghi lại. Đối với 22 người tình nguyện đầu tiên, các tác giả có ghi lại đoạn phim quay khuôn mặt của những ngời tham gia bằng thiết bị Sony DCR-HC27E.

Quá trình thu tín hiệu điện não sẽ bắt đầu với việc thu tín hiệu nền trong vòng 2 phútvào lúc này những người tham gia được đề nghị phải trong trạng thái thư giãn hoàn toàn. Sau đó 40 đoạn phim sẽ được trình chiếu tương đương với 40 lần đo khác nhau, trong đó mỗi lần đo sẽ tiến hành các bước như sau:

* + Màn hình hiển thị số sẽ hiện thị trong 2 giây lần thử thứ mấy cho người đo, từ đó thông báo cho những người tham gia về tiến trình của họ.
  + Người tham gia được thu tín hiệu nền trong 5 giây.
  + Người tham gia được thu tín hiệu trong quá trình xem đoạn phim trong 1 phút.
  + Người tham gia tự đánh giá đoạn phim theo các tiêu chí đã đề ra.

Sau 20 lần đo, những người tham gia được cho phép nghỉ ngơi trong một thời gian ngắn. Trong khoảng thời gian đó, các tác giả sẽ kiểm tra chất lượng của các tín hiệu thu được và vị trí điện cực, cuối cùng những người tham gia được yêu cầu tiếp tục phần còn lại của thí nghiệm.

Ngoài bộ dữ liệu gốc đo được, các tác giả của bộ dữ liệu DEAP còn cung cấp cho các nhà nghiên cứu một bộ dữ liệu đã được tiền xử lý (Bảng 3.1); từ đó các nhà nghiên cứu có thể dễ dàng kiểm tra các thuật toán mà họ đã đề ra. Quá trình tiền xử lý bao gồm các bước sau:

* Dữ liệu được lấy mẫu xuống còn 128Hz.
* Tín hiệu được loại bỏ các tín hiệu điện EOG.
* Tín hiệu được đi qua bộ lọc tần số băng tần với tần số từ 4-45 Hz.
* Các kênh EEG đã được điều chỉnh sao cho có cấu trúc thứ tự theo phòng thí nghiệm ở Geneva.
* Dữ liệu được phân đoạn thành các thử nghiệm 60 giây và tín hiệu đo thử trong 3 giây được xóa đi.

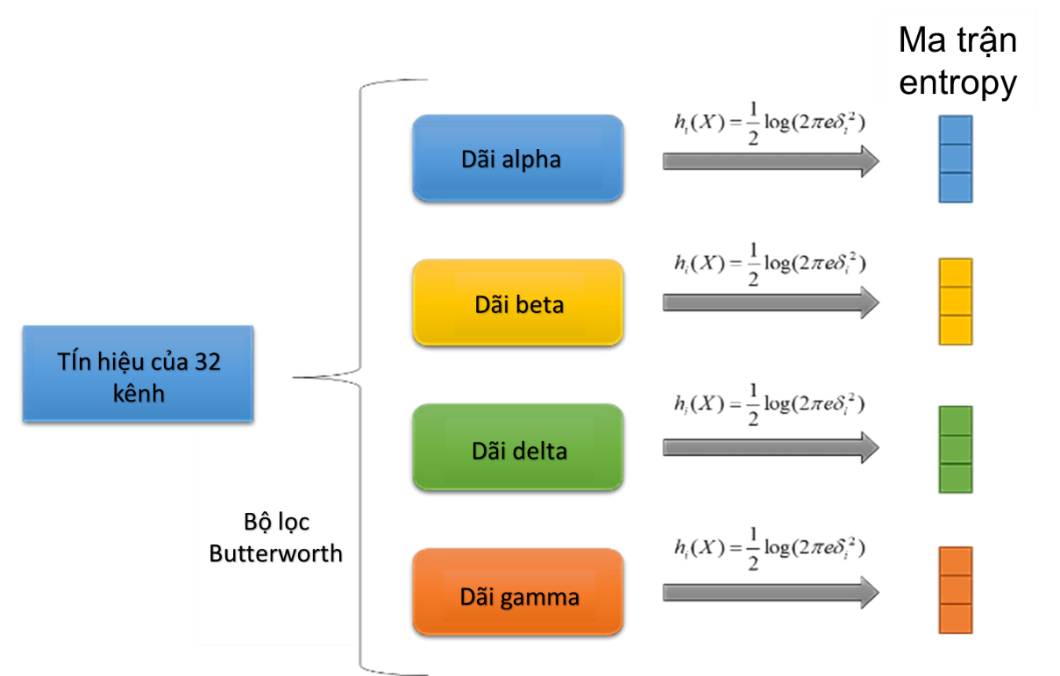
###### **Bảng 3.1:** Mô tả cấu trúc của bộ dữ liệu DEAP đã được tiền xử lý

|  |  |  |
| --- | --- | --- |
| **Tên** | **Cấu trúc** | **Các thành phần** |
| Dữ liệu | 40×40×8064 | Đoạn phim × Số kênh × Dữ liệu |
| Nhãn | 40×4 | Lần thử × Nhãn (Arousal, Valence, Ấn tượng, Yêu thích) |

### 3.1.2 Phân đoạn, tách dải tần số và trích xuất tính năng

Trước khi phân loại dữ liệu thành hai phần là huấn luyện và kiểm tra, các tín hiệu được đi qua bộ lọc Butterworth để tách thành các dải tần số khác nhau bao gồm:   
“Delta, Alpha, Beta và Gamma” như trong hình. Quá trình tách các dải sóng nhằm loại bỏ các thông tin nhiễu không cần thiết có trong tín hiệu, đồng thời trích xuất được thông tin cần thiết [8]. Các dải tần số được lựa chọn đóng vai trò cực kì quan trọng vì chúng có khả năng đại diện cho các trạng thác khác nhau như sau:

* Sóng Delta (0,5–4 Hz): có tần số thấp nhất khi so với các dải tần số khác, sóng delta thường xuất hiện khi con người trong trạng thái ngủ sâuvà khi sóng Delta xuất hiện nhiều chứng tỏ quá trình thiết lập lại cho quá trình cập nhật thêm thông tin mới của não [42].
* Sóng Alpha (8–13 Hz): có thể được phát hiện ở tất cả các bộ phận của thùy sau của não, khi sóng Alpha xuất hiện chiếm đa số sẽ hỗ trợ cơ thể con người đi vào trạng thái bình tĩnh, ngoài ra dải Alpha còn có khả năng thúc đẩy cảm giác thư giãn sâu [42].
* Sóng Beta (14–26 Hz): là dải tần có biên độ thấp và tần số cao, sóng Beta thường được xuất hiện nhiều trong trạng thái con người tỉnh táo, hoặc con người đang tham gia vào quá trình tư duy có ý thức, tư duy logic hoặc sử dụng các chất kích thích như cafein hoặc trà [42].
* Sóng Gamma (30–45 Hz): sóng não có tần số cao nhất và thường xuất hiện khi con người cần xử lý các công việc có mức độ cao. Ngoài ra, dải Gamma có thể được phát hiện khi con người trong điều kiện căng thẳng [42].



##### **Hình 3.1:** Tổng quát quá trình lọc dải tần số và tính độ phức tạp tín hiệu

Sau bước tách dải tần số, các tần số khác nhau có định dạng 40×8064×32 (video × mẫu × kênh) được chuyển sang định dạng 40 × 8064 × 4 × 32 (video × mẫu × dải tần số × kênh). Sau đó, các tín hiệu EEG của mỗi băng tần được chia nhỏ thành nhiều đoạn mỗi phân đoạn có lần lượt là 32, 64 và 128 mẫu. Sau đó, công thức entropy vi sai được áp dụng nhằm kiểm tra sự phức tạp của một biến ngẫu nhiên liên tục [13]. Công thức tính toán của nó được mô tả như sau:

|  |  |  |
| --- | --- | --- |
|  |  | (3.1) |

trong đó *x* là một biến ngẫu nhiên, *f(x)* là hàm mật độ xác suất của *x*. Đối với một chuỗi *x* tuân theo phân bố Gauss *N(μ, δ2)*, entropy vi sai của nó có thể được thể hiện như sau:

|  |  |  |
| --- | --- | --- |
|  |  | (3.2) |

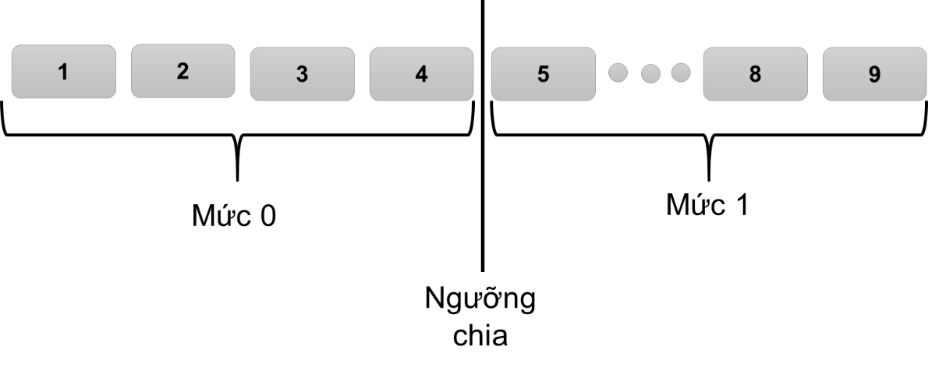
Mặc dù các tín hiệu EEG không tuân theo một phân bố cố định nhất định nào, tuy nhiên các tín hiệu này lại tuân theo phân phối Gaussian trong các dải tần số tương ứng từ 2Hz đến 44Hz [13]. Do đó, định nghĩa của entropy vi sai cho tần số cụ thể là:

|  |  |  |
| --- | --- | --- |
|  |  | (3.3) |

trong đó và đại diện cho entropy vi sai của các tín hiệu EEG tương ứng trong dải tần số thứ và phương sai tín hiệu, tương ứng.

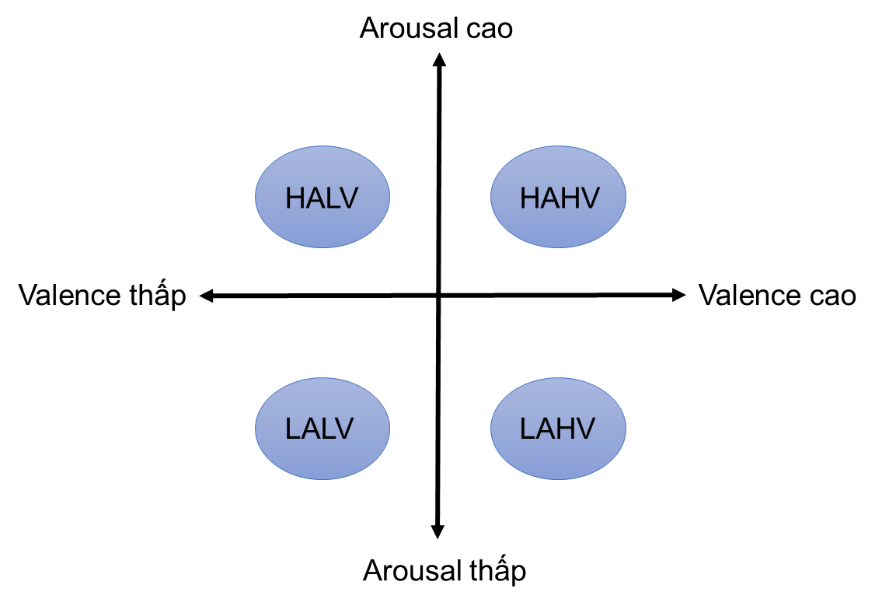
Đối với mỗi phân đoạn trên, dựa vào entropy vi sai có thể tính toán ra được các tính năng cho độ phức tạp hay độ xuất hiện của tín hiệu trong từng thời điểm cụ thể. Dựa vào công thức (3.3), kết quả sau tính toán sẽ là một vec-tơ một chiều, ∈ R32. Cuối cùng, vec-tơ này được chuẩn hóa bằng cách sử dụng chuẩn hóa Z-score.

Nhãn của bộ dữ liệu sẽ được chia ra làm hai phần, các giá trị lớn hơn hoặc bằng 5 sẽ được thay đổi giá trị thành một, tương đương với mức cao và các giá trị nhỏ hơn 5 sẽ được gán giá trị bằng 0 tương đương mức thấp. Quá trình chia nhãn dữ liệu chọn giá trị ngưỡng là 5 bởi vì đây là cách chia phù hợp nhất để mức cao và mức thấp có số lượng dữ liệu là gần như bằng nhau, tránh được trường hợp khi huấn luyện, mô hình bị lệch về một mức cảm xúc trong quá trình phân loại.



##### **Hình 3.2:** Quá trình phân loại các nhãn

Ngoài ra, để tăng độ phức tạp cho quá trình huấn luyện, nhãn của dữ liệu được chia tiếp thành bốn mức: “HAHV (Arousal cao Valence cao), HALV (Arousal cao Valence thấp), LAHV (Arousal thấp Valence cao), LALV (Arousal thấp Valence thấp)

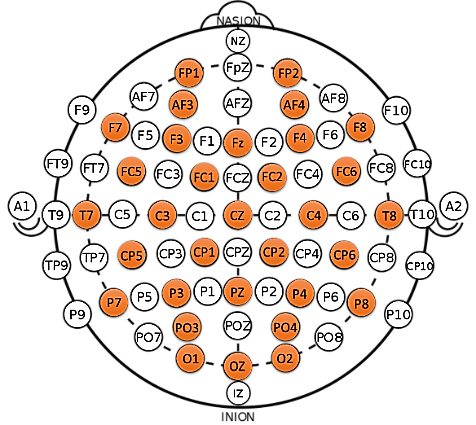


##### **Hình 3.3:** Mô hình cảm xúc của Russell [9]

### 3.1.3 Xây dựng mô hình 1D, 2D và 3D

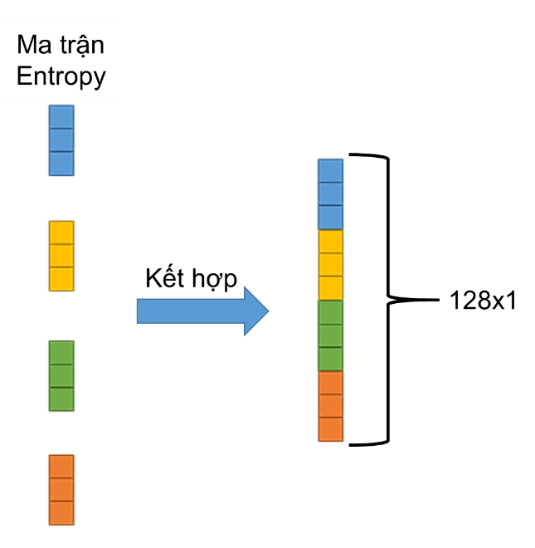
Trong mục này, tôi sẽ xây dựng các cấu trúc 1D, 2D và 3D dựa trên cấu trúc điện cực của bộ dữ liệu DEAP, nói cách khác là hệ thống 10-20. Ngoài ra, phương pháp thị giác máy tính cũng được sử dụng như là hình mẫu cho cấu trúc của các khối EEG ba chiều.

Cấu trúc của tín hiệu sóng não được dựa trên hệ thống 10-20, hệ thống đã được các nhà nghiên cứu quốc tế công nhậnvà có thể mô tả, áp dụng cho vị trí của điện cực và khu vực bên dưới của vỏ não. Số "10" và "20" là khoảng cách thực tế giữa các điện cực liền kề, nói cách khác đó là 10% hoặc 20% tổng khoảng cách phía trước hoặc bên trái của hộp sọ [43], tầm nhìn mặt phẳng của hệ thống được minh họa trong Hình 3.4.



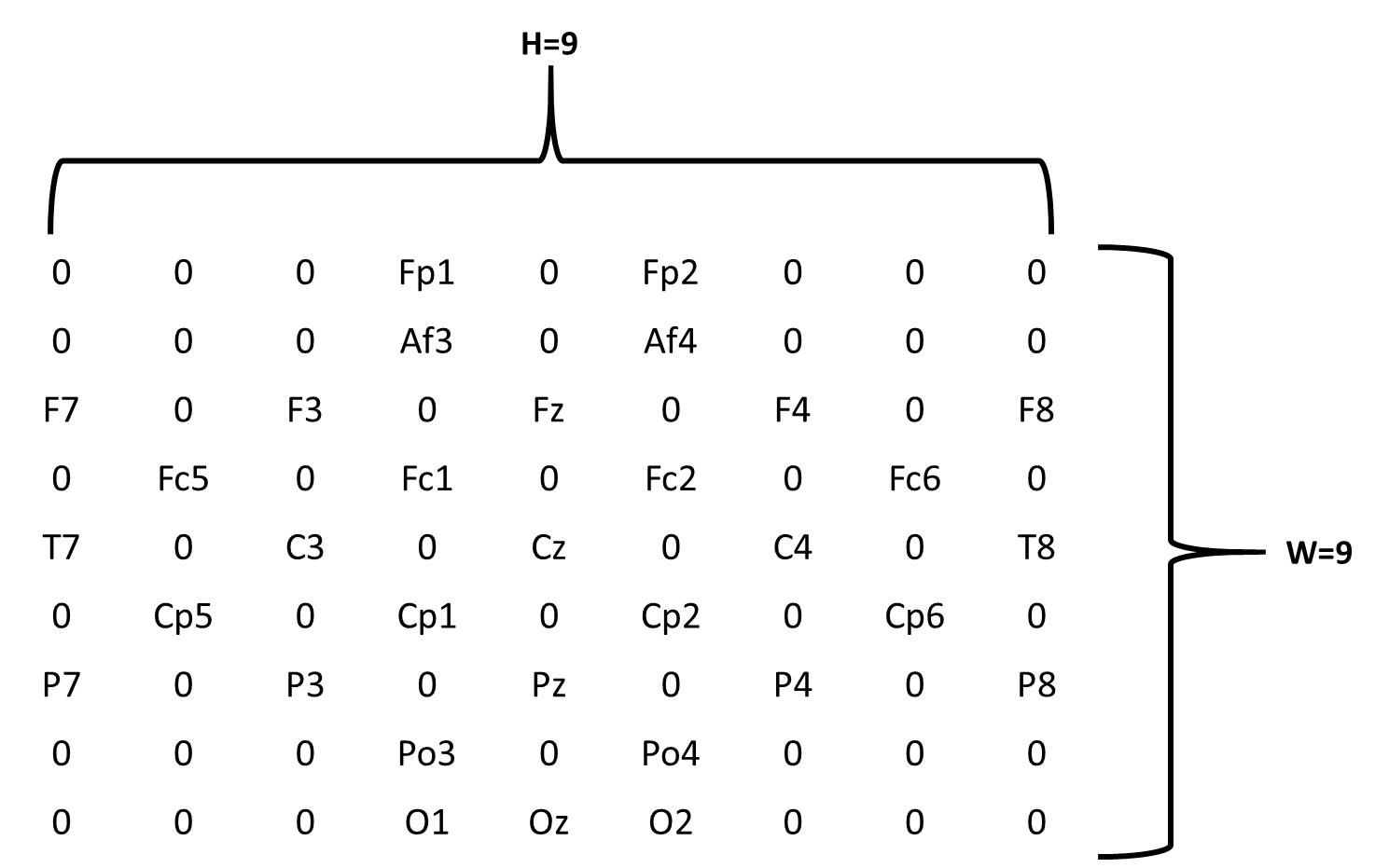
##### **Hình 3.4:** Vị trí các kênh của thiết bị thu trong tập dữ liệu DEAP

Để xây dựng cấu trúc 1D cho bộ dữ liệu, tôi ghép toàn bộ các vec-tơ 1D với nhau từ đó hình thành nên một vec-tơ 1D với độ dài 128 mẫu. Quá trình thực hiện được mô tả như Hình 3.5.



##### **Hình 3.5:** Quá trình xây dựng mô hình 1D cho dữ liệu

Đối với cấu trúc 2D, tôi xây dựng một ma trận 2D với kích thước là H × W để mô phỏng bản đồ điện cực bên trên và nhằm giữ độ chính xác khoảng cách giữa các kênh liền kề, trong đó H và W là số lượng tối đa kích thước tính lần lượt theo chiều x và y của ma trận. Với tập dữ liệu DEAP, giá trị bằng chín là giá trị phù hợp nhất. Ngoài ra, các giá trị 0 sẽ đại diện cho các tín hiệu từ các vị trí không có điện cực.



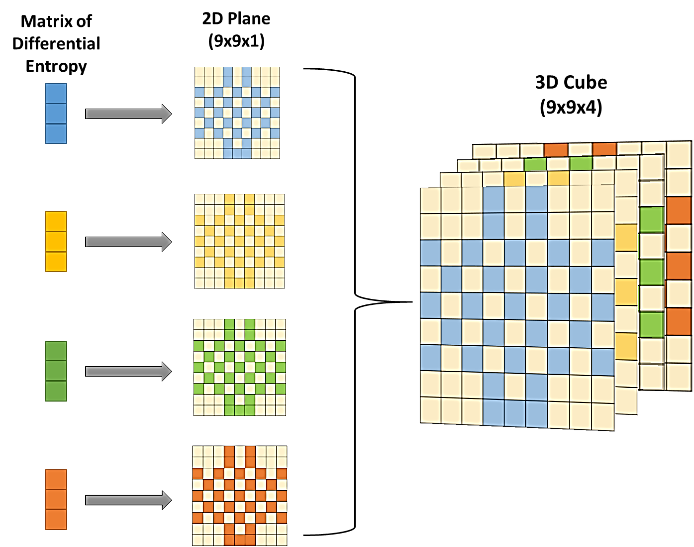
##### **Hình 3.6:** Ma trận mô phỏng các kênh trong tập dữ liệu DEAP

Sau đó, mỗi phân đoạn ma trận của bốn dải tần số được kết hợp với nhau để trở thành một cấu trúc hai và ba chiều. Đối với cấu trúc hai chiều thì tôi biến đổi các dải tần số trở thành một hình ảnh màu với các kênh được liên kết với nhau, còn cấu trúc ba chiều sẽ là hình ảnh trắng đen của các dải tần số với các kênh rời nhau hoàn toàn.

###### **Bảng 3.2:** So sánh giữa phương pháp thị giác máy tính và phương pháp thực hiện

|  |  |
| --- | --- |
| **Thị giác máy tính** | **Phương pháp thực hiện** |
| Ảnh màu | Cấu trúc EEG |
| Kênh màu (Đỏ, Xanh lá, Xanh dương) | Dải tần số (β, α, δ, γ) |
| Cường độ màu sắc | Giá trị của DE |

Phương pháp xây dựng cấu trúc 2D và 3D của tôi được thực hiện được dựa trên phương pháp của thuật toán thị giác máy tính. Trong quá trình phân loại hình ảnh trong phương pháp thị giác máy tính, các hình ảnh được đại diện bởi ba màu chính (đỏ, xanh lá cây và xanh dương). Mỗi màu chính có giá trị từ 0 đến 255 đại diện cho cường độ của màu sắc từ đậm sang nhạt. Với các dữ liệu ở 2D và 3D, mỗi dải sẽ được mô tả là màu chính trong hình ảnh và giá trị tính năng DE sẽ được mô tả là cường độ của màu sắc.

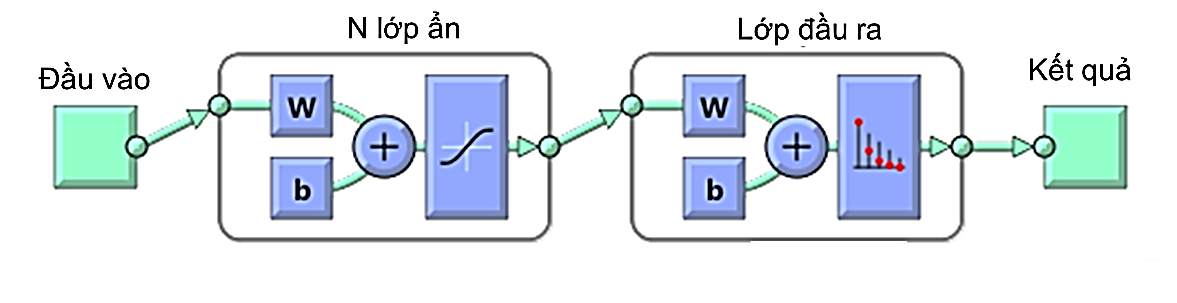


##### **Hình 3.7:** Tổng quan quá trình xây dựng dữ liệu dạng 2D và 3D

# CHƯƠNG 3 MÔ HÌNH ĐỀ XUẤT

## [3.1 THIẾT](about:blank) KẾ MẠNG NƠ-RON TRUYỀN THẲNG

Để khảo sát mạng truyền thẳng, mô hình mạng được thiết kế với một lớp và hai lớp riêng biệt nhau nhằm tìm ra số lớp phù hợp cho quá trình phân loại. Ngoài ra với mỗi lớp ẩn, số nút sẽ được thay đổi từ 10 nút đến 150 nút lần lượt cho từng lớp ẩn, từ đó khảo sát được sự ảnh hưởng của các nút đến quá trình phân loại. Với mỗi nút, hàm hoạt động sử dụng là hàm *Logsi*, hàm có chức năng đưa ra kết quả trong khoảng từ 0 tới 1 khi các giá trị đầu vào sẽ trong khoảng từ âm đến dương vô cực. Dựa vào đó, có thể thấy đây là hàm rất phù hợp với bộ dữ liệu cần được xử lý. Với giá trị khai báo ban đầu của trọng số và mỗi nút sẽ được tạo một cách ngẫu nhiên các giá trị và các giá trị bias đều sẽ được đặt bằng một. Ngoài ra, mô hình sẽ chia bộ dữ liệu làm hai phần theo tỉ lệ 80%/ 20% lần lượt là dữ liệu dùng huấn luyện/ dữ liệu dùng kiểm tra mạng. Luận văn sử dụng phương pháp Stochastic gradient descent để tối ưu hóa mô hình và thuật toán học ngược để huấn luyện mô hình mạng truyền thẳng.

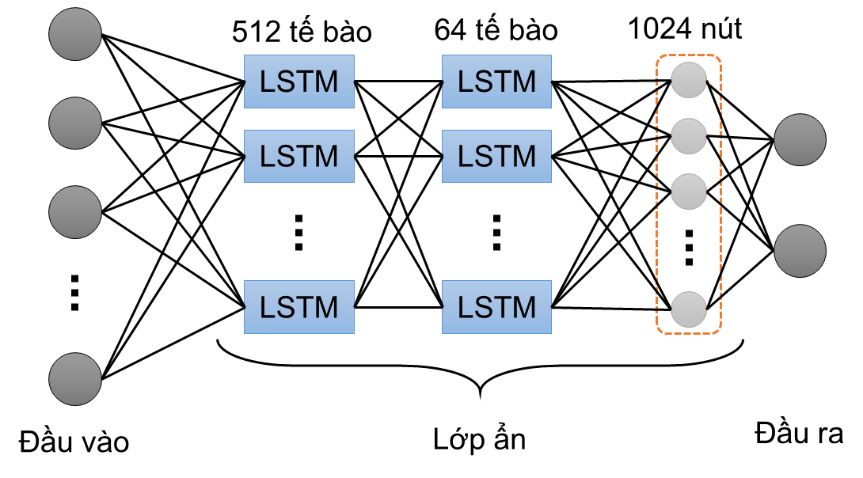


##### **Hình 4.1:** Mô hình tổng quan của mạng nơ-ron truyền thẳng

Tuy nhiên, mạng nơ-ron truyền thẳng chỉ có thể nhận được các vec-tơ một chiều chính vì thế trong mạng này tôi chỉ khảo sát với bộ dữ liệu 1D để tìm ra mô hình phù hợp nhất.

## 3.2 THIẾT KẾ MẠNG BỘ NHỚ DÀI-NGẮN (LSTM)

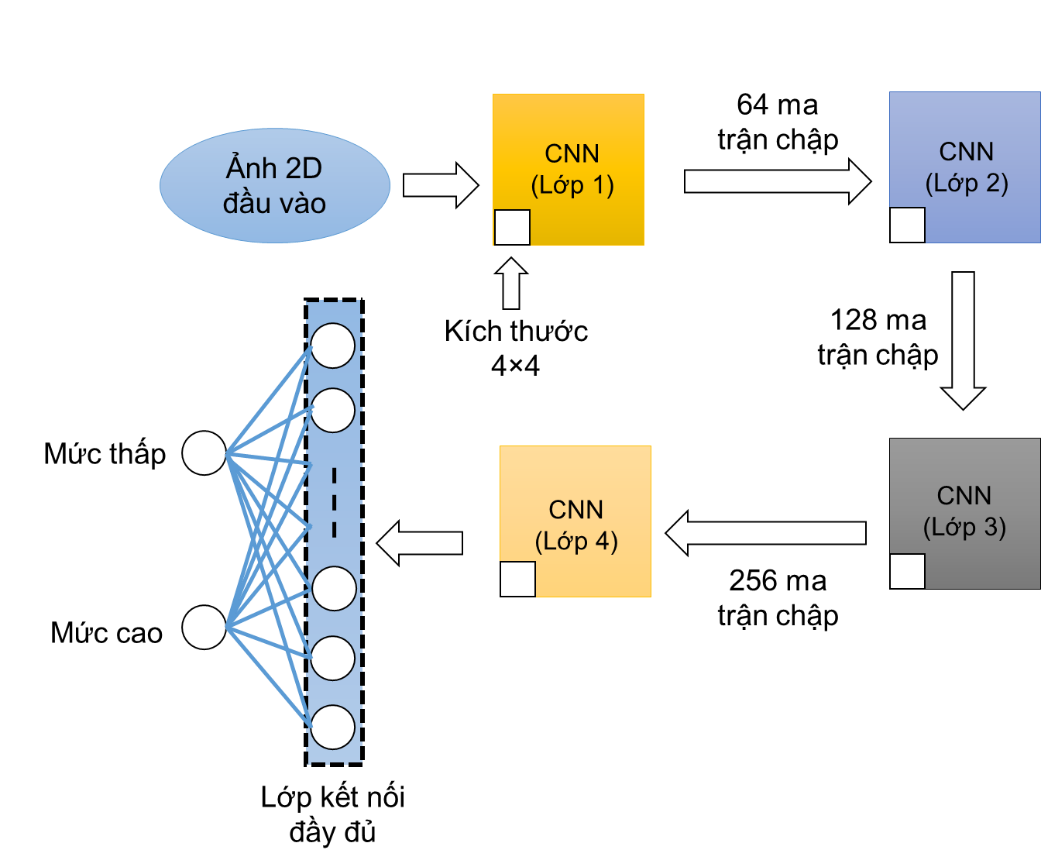
Đối với mạng LSTM, mục tiêu của luận văn muốn hướng tới kiểm tra đối với quá trình sử dụng hình ảnh tuần tự nhau (frame) có đưa ra kết quả tốt hơn so với sử dụng các dữ liệu một cách ngẫu nhiên như mạng truyền thẳng. Chính vì thế, mạng LSTM được khởi tạo với các thông số một cách ngẫu nhiên bên trong của từng tế bào, cấu trúc của mạng sẽ có hai lớp, lớp đầu tiên bao gồm 512 tế bào và lớp thứ hai bao gồm 64 tế bào theo mô hình nghiên cứu [10]. Bên trong mỗi tế bào, hàm hoạt động được sử dụng là hàm *Tanh* và hàm hoạt động cho mỗi cổng sẽ là hàm *Sigmoid*. Sau hai lớp chứa các tế bào trên, các lớp kết nối đầy đủ với 1024 nút được thêm vào để phục vụ cho quá trình bỏ phiếu và lớp cuối cùng sẽ gồm có 2 nút tương ứng cho hai mức cao và thấp của cảm xúc phân loại. Trong mô hình LSTM, tôi sử dụng dữ liệu ở dạng 2D và 3D để làm tín hiệu đầu vào cho quá trình huấn luyện. Nói rõ hơn, luận văn sẽ có hai mô hình mạng LSTM có kiến trúc như nhau nhưng khác nhau ở dạng dữ liêu đầu vào huấn luyện.



##### **Hình 4.2:** Mô hình tổng quan của mạng LSTM

## 3.3 THIẾT KẾ MẠNG NƠ-RON TÍCH CHẬP 2 CHIỀU (2D-CNN) VÀ 3 CHIỀU (3D-CNN)

Kiến trúc mạng nơ-ron tích chập 2 chiều được thiết kế với bốn lớp dựa theo cấu trúc thiết kế của [11]. Cụ thể hơn, trong mỗi lớp tích chập, “*zero-padding*” được thêm vào để ngăn chặn mất thông tin ở rìa của ma trận. Trong ba lớp tích chập đầu tiên, kích thước ma trận chập sẽ là 4×4. Đối với lớp tích chập cuối cùng, kích thước của ma trận sẽ là 1×1, mục tiêu của lớp tích chập cuối là tổng hợp toàn bộ các tính năng đã đi qua sau ba lớp tích chập phía trước. Ở đây sử dụng kích thước ma trận chập là 4×4×2 vì cấu trúc điện cực trong ma trận trong Hình 3.6 được đặt rất xa nhau. Như Hình 3.6, điện cực *Fp1* ở vị trí cột 4 và hàng 1 trong khi *F7* ở cột 1 và hàng 3; tôi có thể lấy giá trị DE trong tích chập đầu tiên mà không cần phải trượt ma trận với kích thước 4×4×2. Đối với các kích thước ma trận nhỏ hơn bên trên sẽ dẫn tới trường hợp kích thước không đủ lớn để bao phủ đến vị trí các điện cực, khi đó các giá trị sau khi chập sẽ tính ra bằng 0. Ngoài ra, nếu kích thước ma trận lớn hơn thì rất có thể quá trình chập sẽ không trích xuất được đầy đủ các tính năng.

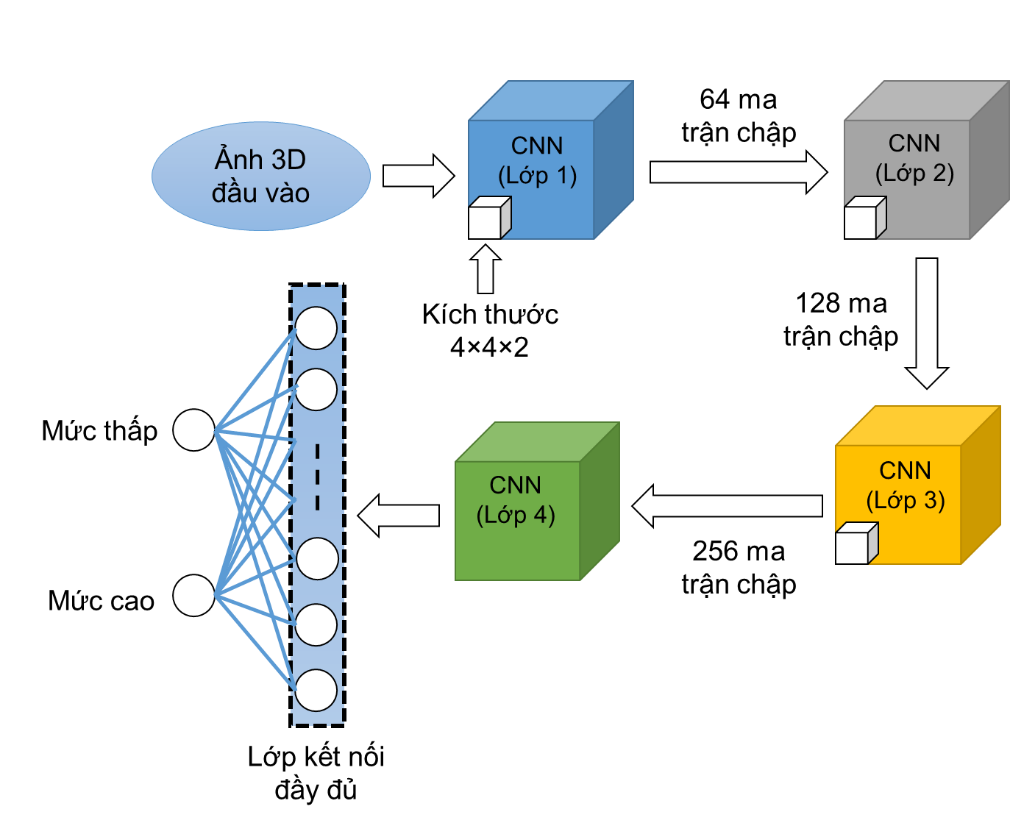


##### **Hình 4.3:** Mô hình tổng quan của mạng 2D-CNN

Độ trượt cho tất cả bốn lớp tích chập được đặt bằng 1×1. Trong mỗi bước của lớp chập, tôi đã sử dụng stride 1×1 để lấy tất cả các tính năng có giá trị DE càng tốt vì điều này có thể giúp mô hình mạng học thêm được nhiều thông tin và cải thiện độ chính xác phân loại của nó.

Luân văn thiết lập ba lớp tích chập đầu lần lượt là 64 ma trận, 128 ma trận và 256 ma trận, đối với lớp tích chập cuối sẽ sử dụng 64 ma trận như ở lớp tích chập đầu. Sau mỗi lớp tích chập, một lớp ReLU được thêm vào để đại diện thay thế cho hàm hoạt động như đối với các mạng nơ-ron truyền thẳng. Lớp ReLU có chức năng biến đổi bất kỳ giá trị âm nào về 0 nhằm mục đích duy trì các giá trị luôn là số dương. Cuối cùng, một lớp được kết nối đầy đủ được thêm vàovà một lớp phân loại (*softmax layer*) được đặt ở cuối để nhận các giá trị đã bỏ phiếu, từ đó đưa ra kết quả cuối cùng.

Trong mô hình 2D-CNN này, lớp tổng hợp (*pooling*) là không cần thiết vì chức năng chính của lớp pooling là giảm kích thước dữ liệu, từ đó giảm độ mất mát thông tin qua các quá trình tích chập. Tuy nhiên, kích thước của dữ liệu mà tôi đã xử lý chỉ có dạng 9×9, được xem là nhỏ khi sử dụng trong mô hình này [11]. Do đó, lớp pooling là không cần thiết.



##### **Hình 4.4:** Mô hình tổng quan của mạng 3D-CNN

Đối với mạng nơ-ron tích chập ba chiều, mạng được thiết kế tương tự với mạng nơ-ron tích chập hai chiều với bốn lớp chập, sử dụng hàm ReLU cho hàm hoạt động, tuy nhiên các giá trị đầu vào sẽ nhận các tín hiệu ba chiều có kích thước là 994. Đối với mỗi lớp chập, các ma trận chập sẽ được thiết kế thành một ma trận khối với kích thước 442 cho ba lớp chập đầu tiên, đối với lớp chập cuối cùng, ma trận 112 để có thể tổng hợp lại tất cả các tính năng đã được trích xuất từ ba lớp chập trước mà vẫn có thể giữ cấu trúc dạng khối cho tính năng.

# CHƯƠNG 4 THỰC NGHIỆM

## 5.1 DỮ LIỆU

Tôi đánh giá mô hình dựa trên dữ liệu của 5 cổ phiếu chọn ngẫu nhiên thuộc nhiều ngành nghề khác nhau trong nhóm VN30 (top 30 cổ phiếu có giá trị vốn hoá thị trường và tính thanh khoản cao nhất) là ngân hàng TMCP Vietcombank (VCB), tập đoàn Vingroup (VIC), tập đoàn Hoà Phát (HPG), CTCP Vinamilk (VNM), CTCP FPT (FPT). Tất cả dữ liệu cổ phiếu được lấy bằng API Investpy từ trang Investing.com, là một trang thông tin tài chính toàn cầu thuộc sở hữu của Fusion Media Ltd. Tôi chọn khoảng thời gian giao dịch trong khoảng 7 năm (từ năm 2015 đến năm 2022), gồm khoảng 1800 mẫu với giá mở phiên (open), giá cao nhất phiên (high), giá thấp nhất phiên (low), giá kết phiên (close) và khối lượng giao dịch (volume). Dữ liệu giao dịch không liên tục vì thị trường nghỉ vào cuối tuần và ngày lễ. Ngoài ra, một số mẫu bị mất mát trong quá trình thu thập và nguồn dữ liệu thiếu hụt.

Cổ phiếu của Vietcombank với giá mở phiên, cao nhất, thấp nhất, kết phiên và khối lượng với khoảng 1800 mẫu.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Symbol** | **Date** | **Open** | **High** | **Low** | **Close** | **Volume** |
| VCB | 2015-02-24 | 16716 | 16927 | 16558 | 16822 | 534023 |
| 2015-02-25 | 16822 | 17929 | 16822 | 17929 | 2901891 |
| 2015-02-26 | 17982 | 18245 | 17454 | 17718 | 2463903 |
| 2015-02-27 | 17771 | 18456 | 17718 | 18245 | 1557643 |
| 2015-03-02 | 18404 | 19511 | 18404 | 19458 | 3939034 |

##### **Bảng 5.7:** Dữ liệu cổ phiếu Vietcombank

Cổ phiếu của Vingroup với giá mở phiên, cao nhất, thấp nhất, kết phiên và khối lượng với khoảng 1800 mẫu.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Symbol** | **Date** | **Open** | **High** | **Low** | **Close** | **Volume** |
| VIC | 2015-02-24 | 27886 | 27886 | 27652 | 27827 | 548730 |
| 2015-02-25 | 27710 | 27944 | 27652 | 27944 | 790900 |
| 2015-02-26 | 27944 | 28061 | 27769 | 27886 | 546700 |
| 2015-02-27 | 27886 | 28295 | 27886 | 28061 | 1348640 |
| 2015-03-02 | 28061 | 28412 | 27944 | 27944 | 1658100 |

##### **Bảng 5.7:** Dữ liệu cổ phiếu Vingroup

Cổ phiếu của Hoà Phát với giá mở phiên, cao nhất, thấp nhất, kết phiên và khối lượng với khoảng 1800 mẫu.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Symbol** | **Date** | **Open** | **High** | **Low** | **Close** | **Volume** |
| HPG | 2015-01-05 | 4566.4 | 4652.6 | 4523.3 | 4566.4 | 2193042 |
| 2015-01-06 | 4480.3 | 4609.5 | 4480.3 | 4609.5 | 2758305 |
| 2015-01-07 | 4609.5 | 4652.6 | 4566.4 | 4566.4 | 2465822 |
| 2015-01-08 | 4566.4 | 4609.5 | 4523.3 | 4566.4 | 1137517 |
| 2015-01-09 | 4566.4 | 4652.6 | 4523.3 | 4652.6 | 4335130 |

##### **Bảng 5.7:** Dữ liệu cổ phiếu Hoà Phát

Cổ phiếu của Vinamilk với giá mở phiên, cao nhất, thấp nhất, kết phiên và khối lượng với khoảng 1800 mẫu.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Symbol** | **Date** | **Open** | **High** | **Low** | **Close** | **Volume** |
| VNM | 2015-02-24 | 66667 | 67014 | 66319 | 66667 | 101218 |
| 2015-02-25 | 66667 | 67361 | 65972 | 66667 | 145282 |
| 2015-02-26 | 66667 | 68403 | 66667 | 68056 | 279547 |
| 2015-02-27 | 68056 | 68750 | 67708 | 68403 | 101736 |
| 2015-03-02 | 69097 | 71528 | 68750 | 71528 | 418162 |

##### **Bảng 5.7:** Dữ liệu cổ phiếu Hoà Phát

Cổ phiếu của FPT với giá mở phiên, cao nhất, thấp nhất, kết phiên và khối lượng với khoảng 1800 mẫu.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Symbol** | **Date** | **Open** | **High** | **Low** | **Close** | **Volume** |
| FPT | 2015-01-05 | 11997.7 | 12047.7 | 11897.7 | 11947.7 | 835215 |
| 2015-01-05 | 11872.7 | 12022.7 | 11822.7 | 12022.7 | 1111534 |
| 2015-01-05 | 12022.7 | 12272.6 | 12022.7 | 12097.7 | 1238274 |
| 2015-01-05 | 12072.7 | 12097.7 | 11947.7 | 11972.7 | 694003 |
| 2015-01-05 | 11972.7 | 12122.7 | 11972.7 | 12072.7 | 977680 |

##### **Bảng 5.7:** Dữ liệu cổ phiếu FPT

## 5.2 XỬ LÝ DỮ LIỆU

### Tiền xử lý dữ liệu

### Phân chia tập dữ liệu

Dữ liệu được chia làm hai phần cho Traning và Testing. Tôi sử dụng 80-85% dữ liệu cổ phiếu (khoảng 1500 mẫu) cho Traning và sử dụng 15-20% dữ liệu cổ phiếu (khoảng 300 mẫu) cho Testing.

## 5.3 HUẤN LUYỆN

Tôi chọn ngẫu nhiên 5 cổ phiếu trong top 30 cổ phiếu lớn nhất của thị trường chứng khoán Việt Nam gồm ngân hàng Vietcombank, tập đoàn Vingroup, tập đoàn Hoà Phát, CTCP Vinamilk, CTCP FPT, với khoảng thời gian giao dịch là 7 năm (từ năm 2015 đến năm 2022), với khoảng 1800 mẫu (khoảng 1500 mẫu cho Training và 300 mẫu cho Testing).

Môi trường thực nghiệm trên Macbook Pro, sử dụng hệ điều hành macOS Ventura 13.1, với chip Apple M1 Pro và RAM 16 GB. Trong quá trình huấn luyện, tôi sử dụng phần mềm PyCharm, Google Colab và ngôn ngữ Python cùng thư viện mã nguồn mở Keras do Google phát triển.

Tôi chọn thực nghiệm từng cổ phiếu bằng mô hình LSTM với 100 epochs. Qua quá trình thực nghiệm, tôi thu được kết quả như sau:

Cổ phiếu Vietcombank

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Symbol** | **RMSE** | **MAE** | **MAPE** | **AR** | **ARTF** |
| VCB | 1982.8 | 1400.2 | 0.0172 | 263.3 | 94.1 |

##### **Bảng 5.7:** Kết quả đánh giá cổ phiếu Vietcombank

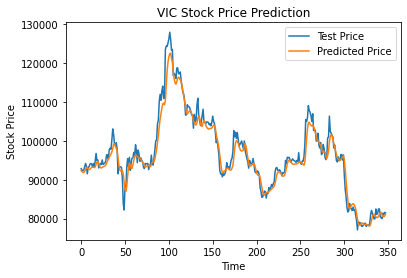
##### 

##### **Hình 5.7:** Biểu đồ giá cổ phiếu Vietcombank

Cổ phiếu Vingroup

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Symbol** | **RMSE** | **MAE** | **MAPE** | **AR** | **ARTF** |
| VIC | 2260.2 | 1557.4 | 0.0158 | 250.2 | 65 |

##### **Bảng 5.7:** Kết quả đánh giá cổ phiếu Vingroup

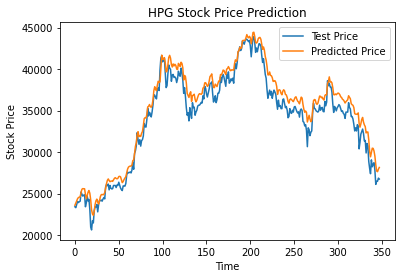


##### **Hình 5.7:** Biểu đồ giá cổ phiếu Vingroup

Cổ phiếu Hoà Phát

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Symbol** | **RMSE** | **MAE** | **MAPE** | **AR** | **ARTF** |
| HPG | 1325.6 | 1171.2 | 0.0352 | 207.7 | 136.7 |

##### **Bảng 5.7:** Kết quả đánh giá cổ phiếu Hoà Phát

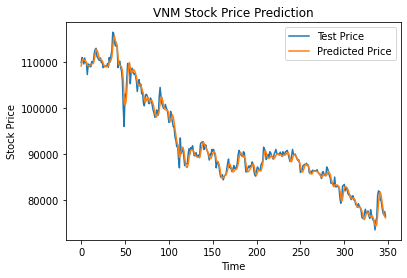


##### **Hình 5.7:** Biểu đồ giá cổ phiếu Hoà Phát

Cổ phiếu Vinamilk

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Symbol** | **RMSE** | **MAE** | **MAPE** | **AR** | **ARTF** |
| VNM | 1121.5 | 772.7 | 0.0083 | 139.8 | -16.2 |

##### **Bảng 5.7:** Kết quả đánh giá cổ phiếu Vinamilk



##### **Hình 5.7:** Biểu đồ giá cổ phiếu Vinamilk

Cổ phiếu FPT

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Symbol** | **RMSE** | **MAE** | **MAPE** | **AR** | **ARTF** |
| FPT | 9326.1 | 8071.8 | 0.1039 | 499.4 | 353.1 |

##### **Bảng 5.7:** Kết quả đánh giá cổ phiếu FPT



##### **Hình 5.7:** Biểu đồ giá cổ phiếu FPT

Tính trung bình, tôi được bảng kết quả như sau:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Mô hình** | **RMSE** | **MAE** | **MAPE** | **AR** | **ARTF** |
| LSTM |  |  |  |  |  |

## 5.4 ĐÁNH GIÁ

Tôi tính RMSE, MAE, MAPE và AR bằng thuật toán dựa trên công thức đã trình bày ở chương 3. Kết quả được trình bày ở trong bảng trên là kết quả khá tốt. Các độ đo RMSE, MAE và MAPE có giá trị thấp nghĩa là dự đoán giá đóng cửa gần đúng với dữ liệu thật và AR chỉ ra lợi nhuận khi giao dịch của những cổ phiếu này. Có thể nhận thấy phương pháp dự đoán của tôi đã đạt kết quả tốt so với những phương pháp khác. So sánh với kết quả báo cáo của một số nghiên cứu với đề tài tương tự, tôi nhận thấy mô hình LSTM của tôi đạt được kết quả tương đối chính xác và đáng khích lệ.

# KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

* **KẾT LUẬN**

Trong luận văn này, tôi đã đề xuất các phương pháp dự đoán giá cổ phiếu bằng học máy. Bằng việc sử dụng mô hình LSTM, tôi đã dự đoán giá cổ phiếu, đánh giá hiệu quả mô hình và đưa ra một vài kết luận như sau:

* Phương pháp và mô hình mà tôi sử dụng có kết quả tương đối tốt.
* Độ chính xác có sự khác nhau giữa các cổ phiếu, điều này là do tính chất từng ngành nghề trong từng giai đoạn, cũng như kết quả kinh doanh của công ty tại thời điểm đó.
* Dữ liệu sử dụng còn nhiều hạn chế về sự liên tục, còn thiếu sót và mất mát nhiều.
* Việc giao dịch dựa trên kết quả dự đoán không được khuyến khích vì thuế và phí khiến độ đo AR bị giảm nhiều.
* **HƯỚNG PHÁT TRIỂN**
* Thử nghiệm trên các mô hình khác nhau để so sánh, đánh giá và tìm ra mô hình tối ưu nhất.
* Trích xuất đặc trưng quan trọng và tối ưu mô hình để kết quả dự đoán chính xác hơn.
* Ứng dụng Internet để cập nhật mô hình, xử lý dữ liệu theo thời gian thực, hỗ trợ việc giao dịch ngay trong phiên.

# TÀI LIỆU THAM KHẢO

|  |  |
| --- | --- |
| [1] | G. Chen, "Automatic EEG seizure detection using dual-tree complex wavelet-Fourier feature," *Expert Systems with Applications,* pp. 2391-2394, 2014 . |
| [2] | W.-L. a. B.-L. L. Zheng, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Transactions on Autonomous Mental Development,* vol. 7, no. 3, pp. 162-175, 2015 . |
| [3] | H. J. a. S. Y. C. Yoon, "EEG-based emotion estimation using Bayesian weighted-log-posterior function and perceptron convergence algorithm" *Computers in biology and medicine ,* vol. 43, no. 12, pp. 2230-2237, 2013. |
| [4] | S. a. A. M. N. Hatamikia, "Recognition of emotional states induced by music videos based on nonlinear feature extraction and som classification" in *21th Iranian Conference on Biomedical Engineering (ICBME). IEEE*, 2014. |
| [5] | S. e. a. Koelstra, "Deap: A database for emotion analysis using physiological signals", *IEEE transactions on affective computing ,* vol. 1, no. 3, pp. 18-31, 2011. |
| [6] | B. A. T. L. a. P. L. Yankner, "The aging brain", *Annu. Rev. Pathol. Mech. Dis.,* vol. 3, pp. 41-66, 2008. |
| [7] | J. Nolte, The human brainq, Mosby/Elsevier, 1993. |
| [8] | A. a. H. N. S. Siegel, Essential neuroscience, Lippincott Williams & Wilkins, 2006. |
| [9] | D. L. a. F. L. D. S. Schomer, Niedermeyer's electroencephalography: basic principles, clinical applications, and related fields, Lippincott Williams & Wilkins, 2012. |
| [10] | M. Teplan, "Fundamentals of EEG measurement," *Measurement science review ,* vol. 2, no. 2, pp. 1-11, 2002. |
| [11] | M. J. Van Putten, Essentials of Neurophysiology, Springer-Verlag Berlin Heidelberg, 2009. |
| [12] | J. R. K. W. B. M. J. Fay S. Tyner, Fundamentals of EEG Technology: Vol. 1: Basic Concepts and Methods, : LWW, 1983. |
| [13] | D. R. P. Joseph D. Bronzino, The Biomedical Engineering Handbook, CRC Press, 2005. |
| [14] | O. e. a. Farooq, "An overview of NFLE" International Journal of Innovative Research in Electrical,," *Electronics, Instrumentation & Control Engg ,* vol. 4, no. 3, pp. 209-211, 2016. |
| [15] | Yannick Roy, et al, NeuroTechX, [Online]. Available: https://neurotechx.com/. |
| [16] | T. W. e. a. Picton, "Guidelines for using human event‐related potentials to study cognition: Recording standards and publication criteria," *Psychophysiology ,* vol. 37, no. 2, pp. 127-152, 2000. |
| [17] | D. Brunet, "Electroencephalography, Guidelines for Clinical Practice and Facility Standards," *College of Physicians and Surgeons of Ontario,* 2000. |
| [18] | Y. H. S.-S. S. Jinming Zou, "Overview of artificial neural networks," *Methods Mol Biol,* vol. 458, no. 15, pp. 15-23, 2008. |
| [19] | W. S. M. &. W. H. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bulletin of Mathematical Biophysics,* vol. 5, pp. 115-133, 1943. |
| [20] | Herausgeber, Brain Theory, Springer-Verlag Berlin Heidelberg, 1984. |
| [21] | F. Rosenblatt, Principles of neurodynamics. perceptrons and the theory of brain mechanisms., Cornell Aeronautical Lab Inc Buffalo NY,, 1961. |
| [22] | N. A. Y. a. M. K. Yadav, An Introduction to Neural Network Methods for Differential Equations, Springer, 2015. |
| [23] | P. D. a. T. S. Wasserman, "What are they and why is everybody so interested in them now?," in *Neural networks.*, IEEE Expert, 1988. |
| [24] | J. Hopfield, " Neural networks and physical systems with emergent collective computational abilities," *Proceedings of the national academy of sciences,* vol. 79, no. 8, pp. 2554-2558, 1982. |
| [25] | J. J. Hopfield, "Neurons with graded response have collective computational properties like those of two-state neurons," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 81, no. 10, pp. 3088-3092, 1984. |
| [26] | M. Minsky, "Overview of Artificial Neural Networks," in *Perceptrons*, MIT Press , 1969, pp. 15-23. |
| [27] | Y. P. S. a. P. F. Bengio, "Learning long-term dependencies with gradient descent is difficult," *IEEE transactions on neural networks,* vol. 5, no. 2, pp. 157-166, 1994. |
| [28] | Y. e. a. Yu, "A review of recurrent neural networks: LSTM cells and network architectures," *Neural computation,* vol. 31, no. 7, pp. 1235-1270, 2019. |
| [29] | K. a. R. N. O'Shea, An introduction to convolutional neural networks., arXiv , 2015. |
| [30] | P. Bambharolia, "Overview of convolutional neural networks," in *Proceedings of the Conference: International Conference on Academic Research in Engineering and Management*, 2017. |
| [31] | P. Bambharolia, "Overview of convolutional neural networks," in *roceedings of the Conference: International Conference on Academic Research in Engineering and Management*, 2017. |
| [32] | R. e. a. Yamashita, "Convolutional neural networks: an overview and application in radiology.," *Insights into imaging,* vol. 9, no. 4, pp. 611-629, 2018. |
| [33] | P. M. S. a. G. D. Dayan, Unsupervised learning, The MIT encyclopedia of the cognitive sciences , 1999. |
| [34] | Z. Ghahramani, "Unsupervised learning," in *Summer School on Machine Learning*, Springer, 2003. |
| [35] | S. Ruder, An overview of gradient descent optimization algorithms, arXiv preprint, 2016. |
| [36] | M. D. Zeiler, Adadelta: an adaptive learning rate method., arXiv preprint, 2012. |
| [37] | D. P. a. J. B. Kingma, "Adam: A methodfor stochastic optimization," in *International Conference onLearning Representations*, 2015. |
| [38] | S. e. a. Tripathi, "Using deep and convolutional neural networks for accurate emotion classification on DEAP dataset.," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, 2017. |
| [39] | Y.-A. e. a. Chen, "The AMG1608 dataset for music emotion recognition," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015. |
| [40] | G. e. a. Placidi, "Classification of Emotional Signals from the DEAP dataset," *International congress on neurotechnology, electronics and informatics,* vol. 2, 2016. |
| [41] | X.-W. D. N. a. B.-L. L. Wang, "EEG-based emotion recognition using frequency domain features and support vector machines," in *International conference on neural information processing*, Berlin, 2011. |
| [42] | W. W. e. a. Ismail, "Human emotion detection via brain waves study by using electroencephalogram (EEG)," *International Journal on Advanced Science, Engineering and Information Technology ,* vol. 6, no. 6, pp. 1005-1011, 2016. |
| [43] | J. A. Russell, Russell, James A. "Measures of emotion, Academic Press, 1989. |
| [44] | Y. e. a. Yang, "Continuous convolutional neural network with 3d input for eeg-based emotion recognition," in *nternational Conference on Neural Information Processing*, 2018. |