

# CẢI THIỆN HIỆU QUẢ MẠNG NEURON HỖ TRỢ CHẨN ĐOÁN BỆNH RỐI LOẠN NHỊP TIM BẰNG PHƯƠNG PHÁP GÁN TRỌNG SỐ TƯƠNG ỨNG MỨC CẤU TRÚC CƠ THỂ

## PERFORMANCE IMPROVEMENT OF DIAGNOSTIC NEURON NETWORK FOR ARRHYTHMIA BY METHOD OF ASSIGNING WEIGHTS TO BODY STRUCTURE LEVELS

Huỳnh Lương Nghĩa, Đinh Văn Quang, Đoàn Thị Bích Ngọc, Nguyễn Thị Thủy

Trường Đại học Điện lực

Ngày nhận bài: 25/09/2019, Ngày chấp nhận đăng: 25/12/2019, Phản biện: TS. Vũ Duy Hải

### Tóm tắt:

Hiện nay, việc ứng dụng mạng neuron nhân tạo hỗ trợ chẩn đoán bệnh lý đang ngày càng phổ biến nên việc cải thiện các chỉ tiêu chất lượng cho các mạng này như độ chính xác chẩn đoán, tốc độ hội tụ của quá trình huấn luyện mạng... vốn phụ thuộc vào việc xác định trọng số cho các thành phần tham gia chẩn đoán đang được quan tâm đặc biệt. Để góp phần giải quyết vấn đề này, bài báo trình bày việc sử dụng phương pháp gán trọng số tương ứng mức cấu trúc cơ thể cho các tham số đầu vào của mạng. Kết quả nhận được cho thấy việc gán trọng số cho các thuộc tính đầu vào tương ứng với mức cấu trúc của cơ thể là có ý nghĩa đáng kể.

### Từ khóa:

Mạng neuron nhân tạo, mức cấu trúc cơ thể, gán trọng số.

### Abstract:

Currently, the application of artificial neuron networks to support pathological diagnosis is increasingly popular, so the improvement of quality indicators for these networks such as diagnostic accuracy, convergence speed of training process network... which depends on the determination of weights of diagnostic components is of particular interest. To contribute to solving this problem, the paper presents the method of assigning weight corresponding to body structure level for network input parameters. The results showed that the weighting of input attributes corresponding to the structure level of the body is significant.

### Keywords:

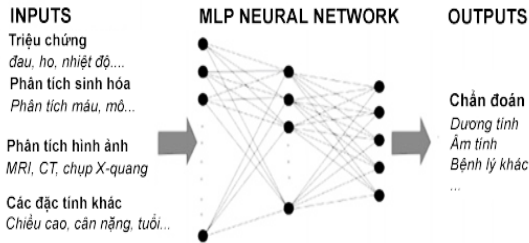
Artificial neuron networks, body structure level, assigning weight.

## 1. GIỚI THIỆU CHUNG

Trong rất nhiều công trình nghiên cứu ứng dụng mạng neuron nhân tạo gần đây điển hình như bài báo [1], đã đưa ra kết luận rằng: “Phân loại bằng cách sử dụng các mô hình mạng neuron nhân tạo cho

thấy tỷ lệ phần trăm thành công rất tốt. Mạng neuron nhân tạo đa lớp truyền thẳng (MLP: multi-layer perceptron networks) được thử nghiệm cho bài toán phân loại bệnh lý dựa trên các thông tin triệu chứng, ảnh chụp y tế và kết quả phân

tích xét nghiêm... là phù hợp để sử dụng trong chẩn đoán y học (hình 1).



**Hình 1. Chẩn đoán y tế sử dụng mạng neuron MLP[4]**

Thông thường đầu vào của mạng này là các triệu chứng - thuộc tính bệnh lý được ghi nhận bằng các phương thức khác nhau và đầu ra là kết quả chẩn đoán được khẳng định bởi thực tế dùng để luyện (dạy) mạng. Nói chung, các mạng này đã cho kết quả có thể chấp nhận được, nhưng khi đầu vào quá lớn có thể dẫn đến giảm độ chính xác và tăng thời gian xử lý (hội tụ chậm hơn). Để giải quyết vấn đề này, nhiều giải pháp đã được đề xuất như chọn lựa cấu trúc tối ưu của mạng neuron nhân tạo, giảm số lượng kích thước thuộc tính đầu vào, chọn thuật toán luyện mạng thích hợp...

Cũng nhằm mục đích này trong bài báo [8] đã bước đầu sử dụng phương pháp gán trọng số cho các thuộc tính đầu vào phù hợp với các mức cấu trúc của cơ thể đối với mạng neuron nhân tạo Hỗ trợ chẩn đoán bệnh ung thư vú. Kết quả nhận được là có ý nghĩa, đặc biệt là đối với những bệnh có liên quan đến sự thay đổi cấu trúc vi mô của cơ thể.

Ý tưởng của phương pháp gán trọng số tương ứng mức cấu trúc cơ thể xuất phát từ việc xem xét thành phần của các thuộc tính đầu vào mạng. Cụ thể các tham

số/thuộc tính đầu vào này - vốn được thu nhận bằng các phương pháp đo lường y sinh khác nhau - sẽ tương ứng với 5 cấp cấu trúc của cơ thể như: cơ thể → cơ quan chức năng → mô → tế bào → phân tử sinh học, như được phân loại trong bảng 1.

**Bảng 1. Mức cấu trúc của cơ thể và các phương pháp chẩn đoán tương ứng**

Số cấp độ	Mức cấu trúc	Phương pháp chẩn đoán
1	Cấp cơ thể	Thu thập triệu chứng ho, đau, sốt... Đo mạch đập, nhiệt độ, huyết áp,...
2	Cấp cơ quan chức năng	Chẩn đoán chức năng (ECG, EEG, EMG,...)
3	Cấp mô	CT, MRI, PET
4	Cấp độ tế bào: các loại tế bào	SPECT, kính hiển vi nano, phương pháp thử phân tích
5	Cấp độ phân tử sinh học	

Một cách trực giác chúng ta thấy độ quan trọng hay trọng số của các tham số/thuộc tính đầu vào này đối với việc chẩn đoán là khác nhau, cụ thể là mức độ cấu trúc càng thấp thì giá trị chẩn đoán tương ứng của chúng càng có ý nghĩa, hay trọng số của chúng càng lớn.

Như vậy, vừa để kiểm tra tính xác đáng và hiệu quả của các phương pháp nêu trên vừa để cải thiện chất lượng của mạng chẩn đoán một số bệnh tim mạch, trong phần tiếp sau sẽ lần lượt xây dựng và khảo sát hiệu quả mạng neuron nhân tạo hỗ trợ chẩn đoán bệnh tim mạch đối với hai trường hợp không gán và có gán trọng số cho các thuộc tính đầu vào tương ứng với các mức cấu trúc của cơ thể.

## 2. XÂY DỰNG MẠNG NEURON NHÂN TẠO HỖ TRỢ CHẨN ĐOÁN BỆNH TIM MẠCH VỚI CÁC ĐẦU VÀO LÀ CÁC THUỘC TÍNH BỆNH LÝ KHÔNG ĐƯỢC GÁN TRỌNG SỐ (HAY CÓ TRỌNG SỐ BẰNG NHAU LÀ 1)

2.1. Để thiết kế mạng neuron trước tiên phải chọn cơ sở dữ liệu phù hợp. Trong trường hợp đang xét ta chọn cơ sở dữ liệu của bệnh rối loạn nhịp tim *Arrhythmia\_dataset* gồm 2 ma trận: ma trận đầu vào mỗi cột tương ứng với các thông số đặc trưng của bản ghi điện tim đồ của một bệnh nhân (*ArrhythmiaInputs*) và ma trận đầu ra mỗi cột tương ứng với loại bệnh chuẩn đoán của một bệnh nhân (*ArrhythmiaTargets*) lấy từ kho cơ sở dữ liệu trực tuyến của Đại học Wisconsin [9]. Được tạo ra vào năm 1997 bởi các tác giả Altay Guvenir, Burak Acar và Haldun Muderrisoglu thuộc Đại học Bilkent ở Thổ Nhĩ Kỳ, cơ sở dữ liệu này chứa các bản ghi điện tâm đồ thu được từ 452 bệnh nhân và 279 thông số đặc trưng (thuộc tính đầu vào) được liên kết với 16 loại bệnh lý tim mạch chuẩn đoán (đầu ra).

2.2. Trước khi dùng để luyện Mạng neuron, cơ sở dữ liệu này cần được xử lý sơ bộ. Do một số bệnh lý chỉ xuất hiện trong một vài mẫu - bản ghi nên chúng được loại trừ giảm số lượng bệnh chuẩn đoán còn lại là 10 (bảng 2).

Ngoài ra dữ liệu được chuẩn hóa bằng cách xác định trị giá phần trăm của thuộc tính so với giá trị tuyệt đối lớn nhất của nó sao cho nằm trong khoảng [0,1] và những thuộc tính không xác định thì bị loại bỏ. Như vậy sau khi xử lý sơ bộ, cơ sở dữ liệu dùng để luyện mạng gồm 2 ma trận: ma trận *ArrhythmiaInputs* có kích thước là  $262 \times 433$  ứng với 262

thuộc tính và 433 mẫu - bản ghi, và ma trận *ArrhythmiaTargets* với kích thước  $10 \times 433$  ứng với 10 loại bệnh được chuẩn đoán và 433 mẫu - bản ghi.

**Bảng 2. Các loại bệnh tim mạch được chuẩn đoán**

Mã lớp	Tên bệnh	Số trường hợp
01	Bình thường	245
02	Thiếu máu cục bộ (bệnh động mạch vành)	44
03	Bệnh nhồi máu cơ tim trước	15
04	Nhồi máu cơ tim kém	15
05	Nhịp tim nhanh xoang	13
06	Nhịp tim chậm xoang	25
07	Nghẽn nhánh trái	9
08	Nghẽn nhánh phải	50
09	Rung tâm nhĩ	5
10	Các loại bệnh khác	22

Các thuộc tính (TT) bao gồm 5 nhóm:

Nhóm 1 từ TT1 → TT4: các đặc trưng chung cơ thể (tuổi, giới tính, chiều cao, cân nặng).

Nhóm 2 từ TT5 → TT15: Các đặc trưng chung tim (độ dài phức hợp QRS, khoảng cách giữa các sóng P và Q, khoảng cách giữa khởi điểm sóng Q và kết thúc sóng T, độ dài sóng T, ..., nhịp đập tim).

Các nhóm còn lại được trình bày trong (bảng 3):

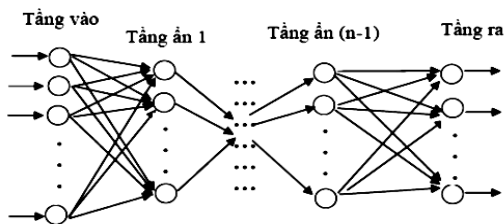
**Bảng 3. Bảng tóm tắt các đạo trình**

Tên đạo trình	Số thuộc tính (TT)
Đạo trình mẫu - cơ bản/Đạo trình song cực chi	
DI	TT16(160) → TT27(169)
DII	TT28(170) → TT39(179)
DIII	TT40(180) → TT51(189)

Tên đạo trình	Số thuộc tính (TT)
Đạo trình tăng cường/Đạo trình đơn cực chi	
aVR	TT52(190) → TT63(199)
aVL	TT 64(200) → TT75(209)
aVF	TT 76(210) → TT87(219)
Đạo trình tim/Đạo trình đơn cực trước tim	
V1	TT88 (220) → TT99(229)
V2	TT 100(230) → TT111(239)
V3	TT112(240) → TT123(249)
V4	TT124(250) → TT135(259)
V5	TT 136(260) → TT147(269)
V6	TT 148(270) → TT159(279)

2.3. Dựa trên cơ sở dữ liệu đã được xử lý tiến hành xây dựng mạng neuron nhân tạo đa lớp (MLP) hỗ trợ chẩn đoán bệnh rối loạn nhịp tim bằng.

Mạng MLP tổng quát là mạng có  $n$  ( $n \geq 2$ ) tầng (thông thường tầng đầu vào không được tính đến): trong đó gồm một tầng đầu ra (tầng thứ  $n$ ) và  $(n-1)$  tầng ẩn.



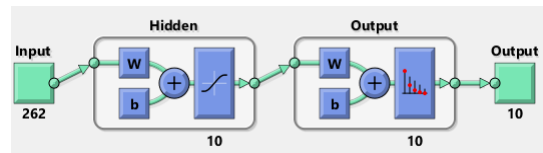
Hình 2. Mạng Neuron MLP

- Đầu vào là các vector  $(x_1, x_2, \dots, x_p)$  trong không gian  $p$  chiều, đầu ra là các vector  $(y_1, y_2, \dots, y_q)$  trong không gian  $q$  chiều. Đối với các bài toán phân loại/chẩn đoán,  $p$  chính là kích thước của mẫu đầu vào,  $q$  chính là số lớp cần phân loại.

- Đầu ra của nơron tầng trước là đầu vào của nơron thuộc tầng liền sau nó. Mỗi nơron thuộc tầng sau liên kết với tất cả các nơron thuộc tầng liền trước nó.

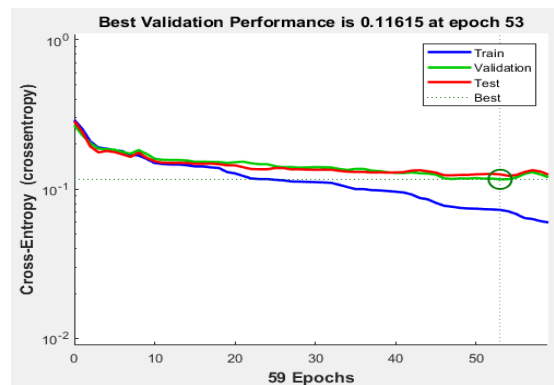
- Hoạt động của mạng MLP như sau: tại tầng đầu vào các nơron nhận tín hiệu vào xử lý (tính tổng trọng số, gửi tới hàm truyền) rồi cho ra kết quả (là kết quả của hàm truyền); kết quả này sẽ được truyền tới các nơron thuộc tầng ẩn thứ nhất; các nơron tại đây tiếp nhận như là tín hiệu đầu vào, xử lý và gửi kết quả đến tầng ẩn thứ 2;...; quá trình tiếp tục cho đến khi các nơron thuộc tầng ra cho kết quả.

Để bảo đảm tính khoa học và giản tiện trong việc xây dựng và khảo sát mạng, ở đây ta sử dụng công cụ Neural Network Toolbox trong phần mềm MATLAB. Như vậy, phù hợp với mục đích và yêu cầu nghiên cứu đã đặt ra, ta chọn Mạng neuron có cấu trúc như sau (Hình 3):



Hình 3. Mạng neuron chẩn đoán hai lớp MLP

Cụ thể, mạng có đầu vào với 262 tế bào nơron ứng với 262 thuộc tính, lớp ẩn sigmoid và lớp đầu ra với 10 tế bào nơron/10 đầu ra ứng với 10 loại bệnh được chẩn đoán.



Hình 4. Kết quả xác thực tốt nhất

2.4. Dữ liệu chẩn đoán đã chuẩn hóa của

bệnh nhân từ ma trận *ArrhythmiaInputs* và ma trận *ArrhythmiaTargets* được áp dụng để huấn luyện mạng MLP đã thiết kế. Dữ liệu này được phân chia một cách phù hợp thành các nhóm dữ liệu đào tạo

(70%), kiểm tra (15%), và xác thực (15%). Mạng được luyện bằng phương pháp lan truyền ngược gradient liên hợp. Kết quả thu được từ quá trình luyện mạng được trình bày trên hình 3 và các bảng:

**Bảng 4. Bảng chỉ số tốt nhất của Mạng neuron chẩn đoán bệnh rối loạn nhịp tim**

		Bệnh được phân loại/chẩn đoán bởi mạng neuron									
		D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
Bệnh thực tế	D1	235	18	2	6	10	16	0	9	0	14
	D2	8	24	0	1	0	0	0	0	3	2
	D3	0	0	13	0	0	0	0	0	0	0
	D4	0	0	0	8	0	0	0	0	1	0
	D5	0	0	0	0	2	0	0	0	0	1
	D6	0	0	0	0	0	7	0	1	0	2
	D7	0	0	0	0	0	0	9	0	0	0
	D8	2	2	0	0	1	2	0	40	1	3
	D9	0	0	0	0	0	0	0	0	0	0
	D10	0	0	0	0	0	0	0	0	0	0

**Bảng 5. So sánh tỷ lệ thành công cho các mô hình phân loại của bệnh rối loạn nhịp tim**

Tỷ lệ thành công [%]	76,3
Tỷ lệ thất bại (%)	23,7

### 3. XÂY DỰNG MẠNG NEURON NHÂN TẠO HỖ TRỢ CHẨN ĐOÁN BỆNH TIM MẠCH VỚI CÁC ĐẦU VÀO LÀ CÁC THUỘC TÍNH BỆNH LÝ ĐƯỢC GÁN TRỌNG SỐ PHÙ HỢP VỚI CÁC MỨC CẤU TRÚC CỦA CƠ THỂ (HAY CÓ TRỌNG SỐ KHÁC NHAU)

**Bảng 6. Kết quả kiểm tra đánh giá hiệu quả mạng neuron chẩn đoán bệnh rối loạn nhịp tim cho trường hợp gán trọng số thuộc tính**

Case number	Gán trọng số cho thuộc tính	Kết quả đào tạo mạng			
		Tỷ lệ thành công [%]	Tỷ lệ thất bại [%]	Kết quả xác thực tốt nhất	Chu kỳ hội tụ
1	Trường hợp không gán	76,3	23,7	0,116	53
2	Gán trọng số đạo trình tăng cường bằng 0	73,8	26,2	0,107	37
3	Gán trọng số đạo trình cơ bản bằng 0	55,3	44,7	0,154	5
4	Gán trọng số thuộc tính DI đạo trình cơ bản bằng 0	73,6	26,4	0,130	55
5	Gán trọng số thuộc tính DII đạo trình cơ bản bằng 0	70,4	29,6	0,115	39
6	Gán trọng số thuộc tính DIII đạo trình cơ bản bằng 0	77,9	22,1	0,105	61
7	Gán trọng số thuộc tính DIII đạo trình cơ bản và đạo trình tăng cường bằng 0	55,3	44,7	0,157	8

3.1. Với mục đích chính là xác định tầm quan trọng của việc gán trọng số bằng cách so sánh kết quả luyện mạng giữa hai trường hợp không và có trọng số, nên ta duy trì cấu trúc của mạng (số lớp, số tế bào thần kinh cho mỗi lớp, độ sai lệch...), nghĩa là vẫn giữ nguyên mạng như trên hình 2.

3.2. Như trên đã phân tích, trong cơ sở dữ liệu *ArrhythmiaInput* có thể chia các thuộc tính thành 5 nhóm có thể coi là tương ứng với các mức cấu trúc của cơ thể. Do vai trò ảnh hưởng tới kết quả chẩn đoán khác nhau, các thuộc tính của các mức cao (1,2) được giữ nguyên hay gán trọng số 1, các mức thấp còn lại sẽ lần lượt được gán trọng số 0 để đánh giá độ quan trọng của chúng.

Kết quả kiểm tra đánh giá hiệu quả Mạng neuron chẩn đoán bệnh rối loạn nhịp tim cho trường hợp gán trọng số thuộc tính được đưa ra trong bảng 6.

#### 4. KẾT QUẢ VÀ THẢO LUẬN

Kết quả đào tạo mạng neuron chẩn đoán bệnh rối loạn nhịp tim trong hai trường hợp không và gán trọng số tương ứng với các mức cấu trúc của cơ thể cho thấy:

- Mô hình mạng MLP chẩn đoán bệnh rối loạn nhịp tim có độ chính xác và độ hội tụ tương đương với kết quả nhận được trong [1], do đó phù hợp cho sử dụng trong chẩn đoán y học
- Đối với mô hình gán trọng số:
  - Từ bảng 6 ta thấy kết quả nhận được ở

trường hợp 2 tốt hơn trường hợp 1 suy ra đạo trình tăng cường là dư thừa nên để tiết kiệm thời gian tính toán ta có thể bỏ qua chúng.

- Trường hợp 3 cho thấy tỉ lệ sai số rất cao tỉ lệ thành công thấp nên việc bỏ thuộc tính đạo trình cơ bản là không hợp lý. Điều này cho phép khẳng định đạo trình cơ bản là thuộc tính quan trọng nhất trong quá trình luyện mạng neuron chẩn đoán bệnh rối loạn nhịp tim.

- Dựa vào kết quả của các trường hợp 4,5,6 ta thấy DI là dạng thuộc tính quan trọng nhất tiếp đến là DII và để tiết kiệm thời gian và thu được kết quả tốt hơn có thể bỏ dạng thuộc tính DIII

- Việc bỏ các thuộc tính đạo trình tăng cường và thuộc tính DIII của đạo trình cơ bản (trường hợp 7) không cho kết quả xác đáng có thể vì dữ liệu đầu vào bị bớt đi nhiều làm thiếu liên kết giữa các thuộc tính với nhau do số lớp ẩn trong mạng không đủ.

**Kết luận:** Việc gán hệ số trọng cho các thuộc tính đầu vào tương ứng với mức cấu trúc của cơ thể là có ý nghĩa đáng kể. Nếu trọng số được chỉ định chính xác theo thứ tự của các cấp cấu trúc, thì hiệu quả sẽ được cải thiện và tốt hơn trường hợp tham chiếu không được gán trọng số hay với trọng số 1 cho tất cả các thuộc tính đầu vào. Kết quả nghiên cứu có thể áp dụng cho việc chẩn đoán bệnh lý với các triệu chứng/thuộc tính đặc trưng bệnh thu nhận được ở các mức cấu trúc cơ thể khác nhau.

#### TÀI LIỆU THAM KHẢO

[1] S. Kajan, D. Pernecký, J. Goga. "Application of neural network in medical diagnostics".

- [2] Deepa Rao, Sujuan Zhao. "Prediction of Breast cancer" (2012).
- [3] <https://gsm672.wikispaces.com/Prediction+of+Breast+cancer>, last visit: 5/14/18
- [4] Q.K. Al-Shayea. "Artificial neural networks in medical diagnosis". International Journal of Computer Science Issues, 2011, 8.2: 150-154.
- [5] S. Kajan. GUI for classification using multilayer perceptron network, Technical Computing Prague, 2009.
- [6] F. Amato, et al. Artificial neural networks in medical diagnosis. Journal of Applied Biomedicine, 2013, 11.2: 47-58. ISSN 1214-0287.
- [7] Kornel Papik, et al. Application of neural networks in medicine — a review. Med Sci Monit, 1998; 4(3): 538-546.
- [8] Huynh Luong Nghia, Dinh Van Quang, Nguyen Thi Thuy. "Pathological diagnosis neuron network with inputs corresponding with structure levels of the body". Journal of Military Science and Technology, 2018, 11.17: 72-78. ISSN 1859 - 1043.
- [9] UCI: Machine Learning Repository  
<http://archive.ics.uci.edu/ml/>

### Giới thiệu tác giả:



Tác giả Huỳnh Lương Nghĩa tốt nghiệp Trường Kỹ sư TLPK Odetsa (Liên Xô) năm 1973; nhận bằng Tiến sĩ ngành điều khiển học năm 1999 tại Minsk - Belarus; được phong Phó giáo sư chuyên ngành tự động hóa năm 2006. Hiện nay tác giả là giảng viên Khoa Điện tử viễn thông - Trường Đại học Điện lực.



Tác giả Đinh Văn Quang là sinh viên năm cuối lớp D10 Thiết bị điện tử y tế - Khoa Điện tử viễn thông - Trường Đại học Điện lực. Tác giả đã tham gia nghiên cứu khoa học theo hướng học thuật "Ứng dụng mạng neuron network trong y tế" từ năm thứ 4; hiện đang thực hiện luận văn tốt nghiệp với đề tài "Ứng dụng trí tuệ nhân tạo xây dựng mạng neuron hỗ trợ chẩn đoán bệnh rối loạn nhịp tim".



Tác giả Đoàn Thị Bích Ngọc sinh viên năm cuối lớp D10 Thiết bị điện tử y tế - Khoa Điện tử viễn thông - Trường Đại học Điện lực. Tác giả tham gia nghiên cứu khoa học theo hướng học thuật "Ứng dụng mạng neuron network trong y tế" từ năm thứ 4; hiện đang thực hiện luận văn tốt nghiệp với đề tài "Ứng dụng trí tuệ nhân tạo xây dựng mạng neuron hỗ trợ chẩn đoán bệnh ung thư vú".

