

# HỆ HỖ TRỢ QUYẾT ĐỊNH TRONG CHẨN ĐOÁN BỆNH: TIẾP CẬN TỪ HỆ MỜ PHỨC

Lương Thị Hồng Lan<sup>1,2</sup>, Trần Thị Ngân<sup>1,2,3</sup>, Hoàng Thị Minh Châu<sup>4</sup>, Lê Bích Liên<sup>5</sup>, Roãn Thị Ngân<sup>6</sup>

<sup>1</sup> Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

<sup>2</sup> Trường Đại học Thủy lợi

<sup>3</sup> Viện Công nghệ thông tin, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

<sup>4</sup> Trường Đại học Kinh tế kỹ thuật công nghiệp

<sup>5</sup> Trường Đại học Sư phạm, Đại học Thái Nguyên

<sup>6</sup> Trường Đại học Tài nguyên và môi trường

lanlhbk@gmail.com, ngantt@tlu.edu.vn, hgchau85@gmail.com, lebichlien@dhsptn.edu.vn, rtngan@hunre.edu.vn

**TÓM TẮT:** Bài toán hỗ trợ chẩn đoán bệnh có ý nghĩa hết sức quan trọng với các bác sĩ. Nó giúp cho quá trình chẩn đoán của bác sĩ được nhanh và chính xác hơn. Việc xây dựng các hệ hỗ trợ chẩn đoán bệnh đã thu hút sự quan tâm của nhiều nhà khoa học. Trong đó, cách tiếp cận từ hệ mờ cũng đã được sử dụng nhiều trong các nghiên cứu trước đây. Trong bài báo này, chúng tôi trình bày mô hình xây dựng hệ hỗ trợ quyết định trong chẩn đoán bệnh theo hướng tiếp cận suy diễn mờ phức. Chúng tôi cũng tiến hành cài đặt mô hình đề xuất trên bộ dữ liệu về bệnh gan. Các kết quả thực nghiệm đã chỉ ra rằng hướng tiếp cận từ hệ mờ phức nâng cao đáng kể độ chính xác trong chẩn đoán bệnh.

**Từ khóa:** hệ mờ, hệ hỗ trợ quyết định, chẩn đoán bệnh, suy diễn mờ, bệnh gan.

## I. GIỚI THIỆU

Chẩn đoán y tế đóng vai trò quan trọng trong quy trình điều trị của bệnh nhân và cũng là trung tâm trong y học lâm sàng. Chẩn đoán chính xác bệnh là yêu cầu quyết định trong việc đưa ra các phương pháp và lộ trình điều trị đúng. Chẩn đoán y tế là dự báo khả năng mắc bệnh của bệnh nhân dựa vào những thông tin triệu chứng mà bệnh nhân gặp phải. Chẩn đoán được ví như “xương sống” của ngành y tế, do đó việc đảm bảo và nâng cao chất lượng chẩn đoán là vấn đề đang được quan tâm hàng đầu tại Việt Nam cũng như thế giới.

Ngày nay, công nghệ thông tin nổi lên như một điểm sáng trong hợp tác đa ngành. Nó như một công cụ đắc lực hỗ trợ nhằm cải thiện hiệu năng, chất lượng trong nhiều lĩnh vực khác nhau trong cuộc sống. Một trong những lĩnh vực gặt hái được nhiều thành tựu to lớn với việc áp dụng công nghệ thông tin là ngành y tế. Trong một vài năm gần đây, sự kết hợp đa ngành này đã giúp giảm bớt tình trạng quá tải, cải thiện dịch vụ y tế, chất lượng khám và điều trị trong các bệnh viện, v.v.. Nhiều phương pháp học máy khác nhau đã được các nhà nghiên cứu cải tiến và phát triển áp dụng trong chẩn đoán y tế như phương pháp phân lớp [9, 38], phương pháp phân cụm [34, 8] luật mờ [23, 22], mạng nơron [4, 18], hệ hỗ trợ quyết định [31, 37], mô hình lai [33, 36].

Vào năm 1975, hệ MYCIN ra đời đã thành công trong việc áp dụng khoa học trí tuệ nhân tạo vào lĩnh vực y học cụ thể là chẩn đoán và điều trị bệnh nhiễm trùng máu. Vào cuối năm 80, Viện công nghệ thông tin bước đầu xây dựng hệ hỗ trợ chẩn đoán khám chữa bệnh nội khoa, chăm cứu và chẩn trị đông y [2], đã có một số hướng tiếp cận trong việc chẩn đoán lâm sàng đối với việc khám chữa bệnh [1].

Zadel [35] đã phát triển lý thuyết lập luận mờ nhằm đưa ra một cách biểu diễn và lập luận với những thông tin ngôn ngữ không chính xác. Trong tiếp cận của Zadel, lập luận xấp xỉ dựa trên cơ sở logic mờ giá trị ngôn ngữ. Trong hệ hỗ trợ ra quyết định đó là xây dựng hệ cơ sở tri thức và xây dựng lập luận tương ứng với cơ sở tri thức. Đối với cơ sở tri thức, ta dựa trên các luật If.. then, còn đối với cơ chế lập luận suy diễn ta sẽ dựa trên việc lập luận ngôn ngữ tự nhiên từ giá trị của biến ngôn ngữ đã được xây dựng trên nền tảng mờ.

Từ những kết quả nghiên cứu về hệ mờ đã có, nhóm chúng tôi nghiên cứu áp dụng mô hình hệ hỗ trợ ra quyết định dựa trên tập mờ phức trong hỗ trợ chẩn đoán bệnh xơ gan cho kết quả tốt.

Phần tiếp theo của bài báo được bố cục như sau: các lý thuyết nền tảng về hệ suy diễn mờ và các nghiên cứu liên quan đến hệ hỗ trợ ra quyết định trong chẩn đoán bệnh sẽ được trình bày trong phần II. Phần III trình bày cụ thể về tập mờ phức và hệ suy diễn dựa trên tập mờ phức. Các kết quả đánh giá thực nghiệm để so sánh hiệu năng của mô hình áp dụng và các mô hình đã có khác được trình bày trong phần IV. Cuối cùng là một số kết luận trao đổi được đưa ra ở phần V.

## II. CÁC HỆ HỖ TRỢ RA QUYẾT ĐỊNH TRONG CHẨN ĐOÁN BỆNH VỚI TIẾP CẬN TỪ HỆ MỜ

### A. Hệ suy diễn mờ

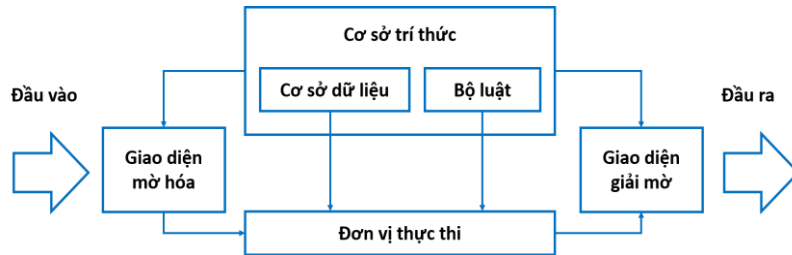
#### 1. Kiến trúc cơ bản của hệ suy diễn mờ

Suy diễn là cơ chế liên kết các tri thức đã có để suy dẫn ra các tri thức mới. Cơ chế suy diễn phụ thuộc rất nhiều vào phương thức biểu diễn tri thức và không có một phương pháp suy diễn duy nhất cho mọi loại tri thức. Hệ suy diễn

mờ là một cơ chế suy diễn thường xuyên được áp dụng khi xây dựng các hệ chuyên gia. Hệ suy diễn mờ tỏ ra hiệu quả trong trường hợp tri thức không đầy đủ, bất định hoặc không chính xác.

Hệ suy diễn mờ (mô tả trong hình 1) có cấu trúc cơ bản như sau:

- Giao diện mờ hóa: chuyển đổi các lớp đầu vào các biên độ phù hợp với các giá trị ngôn ngữ.
- Cơ sở tri thức bao gồm 2 phần:
  - Cơ sở dữ liệu: định nghĩa các hàm thuộc của các tập mờ được sử dụng trong các luật mờ
  - Bộ luật: gồm các luật mờ IF – THEN
- Đơn vị thực thi: thực hiện các hoạt động suy diễn trong các luật
- Giao diện giải mờ: chuyển đổi các giá trị kết quả mờ của hệ suy diễn ra các lớp đầu ra



**Hình 1.** Sơ đồ tổng quan của hệ suy diễn mờ

Các bước suy diễn mờ:

- Mờ hóa các biến đầu vào: ta cần mờ hóa những giá trị rõ để tham gia vào quá trình suy diễn
- Áp dụng các toán tử mờ (AND hoặc OR) cho các giả thiết của từng luật.
- Áp dụng phép kéo theo để tính toán giá trị các giả thiết đến kết luận của từng luật.
- Áp dụng toán tử gộp để kết hợp các kết quả trong từng luật thành một kết quả duy nhất cho cả hệ.
- Giải mờ kết quả tìm được cho ta một số rõ.

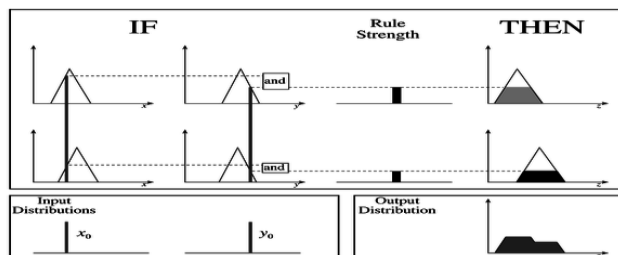
Có ba kiểu hệ suy diễn mờ, đó là Hệ suy diễn Mamdani, Hệ suy diễn Sugeno (hay còn gọi là Takagi – Sugeno), Hệ suy diễn Tsukamoto.

2. Hệ suy diễn Mamdani

Theo hình 2, hệ suy diễn mờ Mamdani có hai biến đầu vào  $x, y$  và một đầu ra  $z$ . Mỗi đầu vào có hai hàm thuộc, được ký hiệu lần lượt là  $\{A_1, A_2\}, \{B_1, B_2\}$  và đầu ra ký hiệu là  $\{C_1, C_2\}$ . Luật thứ  $k$  sẽ có dạng như sau:

$$k: \text{If } x \text{ is } A_i^k \text{ and } y \text{ is } B_j^k \text{ then } z \text{ is } C_l^k$$

với  $k = 1, \dots, R; i = 1, \dots, N; j = 1, \dots, M$  và  $l = 1, \dots, L$  trong đó  $N, M, L$  là số lượng hàm thuộc của hai biến đầu vào và biến đầu ra. Trong hệ suy diễn này, phương pháp giải mờ thường được sử dụng là lấy cực đại và tính toán điểm trọng tâm.



**Hình 2.** Hệ thống suy diễn Mamdani với 2 đầu vào và 2 luật

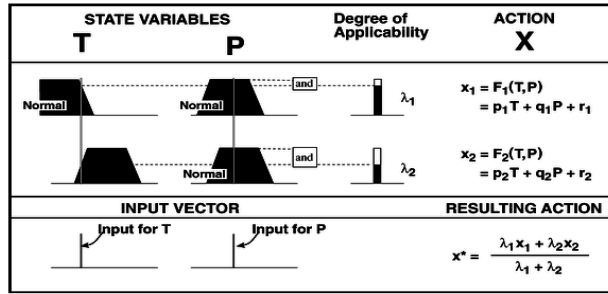
3. Hệ suy diễn Sugeno

Trong hệ suy diễn Sugeno (hình 3), các luật được hình thành như sau:

$$k: \text{If } x \text{ is } A_i^k \text{ and } y \text{ is } B_j^k \text{ then } z^k = f(x, y)$$

Cũng giống như Mamdani,  $k = 1, \dots, R, i = 1, \dots, N$  và  $j = 1, \dots, M$  trong đó  $N$  và  $M$  là số lượng hàm thuộc cho biến đầu vào. Phương pháp giải mờ này sử dụng toán tử độ mạnh trung bình.

Do hệ suy diễn Sugeno có hiệu quả tính toán cao hơn so với hệ suy diễn Mamdani, mô hình Sugeno thường được sử dụng cho các kỹ thuật thích ứng trong việc xây dựng các mô hình mờ. Những kỹ thuật thích ứng có thể được sử dụng để tùy chỉnh các hàm thành viên để đạt được mô hình hiệu quả nhất cho từng loại dữ liệu. Lợi thế của suy diễn Sugeno là tính toán hiệu quả, làm việc tốt với các kỹ thuật tuyến tính, tối ưu hóa, và rất thích hợp để phân tích toán học. Một trong những vấn đề lớn với hệ suy diễn mờ Sugeno là không có phương pháp trực quan tốt để xác định các hệ số,  $p, q,$  và  $r$ . Ngoài ra, hệ suy diễn Sugeno chỉ có đầu ra rõ.



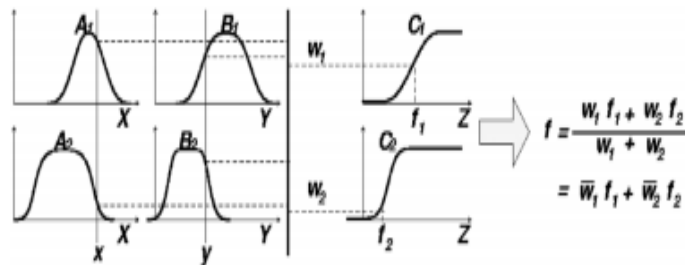
**Hình 3.** Hệ suy diễn Sugeno với 2 đầu vào và 2 luật ( $p_n, q_n, r_n$  là các hằng số)

4. Hệ suy diễn Tsukamoto

Trong hệ suy diễn Tsukamoto (hình 4), luật if – then được biểu diễn như sau:

k: If x is Aki and y is Bkj then z is Ckl

Các luật trong hệ suy diễn này được đại diện bởi một tập mờ với hàm thuộc monotonical. Kết quả suy diễn của từng luật được xác định là một giá trị rõ, giá trị này được tính ra khi tính độ mạnh của luật. Kết quả cuối cùng được tính bằng cách lấy giá trị trung bình đầu ra của từng luật.



**Hình 4.** Hệ suy diễn Tsukamoto

Vì mỗi luật suy diễn ra một kết quả rõ, hệ suy diễn Tsukamoto lấy giá trị cuối cùng là trung bình nên quá trình giải mờ diễn ra nhanh chóng. Tuy nhiên, hệ suy diễn mờ Tsukamoto không được sử dụng thường xuyên vì nó không được hiệu quả như hai hệ suy diễn mờ Mamdani hoặc Sugeno.

**B. Tổng quan các hệ hỗ trợ chẩn đoán bệnh với tiếp cận từ hệ mờ**

Thực tế tất cả các nghiên cứu ứng dụng học mờ trong chẩn đoán y tế đều nhằm mục đích cải thiện kết quả chẩn đoán bệnh bằng cách sử dụng các phương pháp học mờ khác nhau. Các hướng nghiên cứu của hệ học mờ ứng dụng trong chẩn đoán y tế bao gồm hai hướng chính là các hệ thống học được xây dựng từ việc kết hợp hệ suy diễn với các phương pháp mờ như các phương pháp dựa trên luật mờ, hệ chuyên gia mờ, cây quyết định mờ, suy diễn mờ ...; và các hệ thống học mờ được xây dựng dựa trên mô hình mạng nơron. Ở bài báo này chúng tôi chủ yếu tập trung vào ứng dụng hệ suy diễn mờ trong chẩn đoán y tế, có thể nói hệ suy diễn mờ là một trong những hệ hỗ trợ ra quyết định có thể được sử dụng trong chăm sóc sức khỏe để dự đoán và chẩn đoán bệnh.

Cụ thể như Adeli và Neshat [3] đã thực hiện một nghiên cứu với mục đích là thiết kế một hệ mờ chuyên gia để chẩn đoán bệnh tim. Hệ thống được thiết kế dựa trên cơ sở dữ liệu y tế của các trung tâm VA, Long Beach và Cleveland ở Mỹ. Mười ba đầu vào của hệ thống, được thực hiện trong gói MATLAB, liên quan đến đau ngực, huyết áp, cholesterol, đường huyết, nhịp tim tối đa, điện tâm đồ (ECG), bài kiểm tra tập thể dục, đỉnh cũ, quét tali, tuổi và giới tính. Đầu ra duy nhất là bệnh tim xảy ra. Bằng việc so sánh kết quả của hệ thống của họ với từng cơ sở dữ liệu, người ta nhận thấy rằng hệ thống của họ có độ chính xác 94%.

Lee và Wang [20] đã cung cấp một hệ chuyên gia mờ cho một ứng dụng hỗ trợ ra quyết định bệnh tiểu đường. Một bản thể học mờ năm lớp, bao gồm lớp kiến thức mờ, lớp quan hệ nhóm mờ, lớp miền nhóm mờ, mờ lớp quan hệ cá nhân và lớp miền cá nhân mờ được phát triển để mô hình hóa kiến thức không chắc chắn. Opeyemi và Justice [26] đã phát triển một hệ thần kinh mờ so sánh để dự đoán các cơn đau tim. Hệ mờ thần kinh của họ được thiết kế với tám đầu vào và một đầu ra. Đầu vào là nhịp tim, bài kiểm tra tập thể dục, huyết áp, tuổi, cholesterol, loại đau ngực, đường huyết và giới tính. Đầu ra là mức rủi ro đối với bệnh nhân, được phân thành bốn cấp độ rất thấp, thấp, cao và rất cao. Do hệ thống của họ là rất thân thiện với người dùng, nên bệnh nhân có thể sử dụng nó. Sikchi và cộng sự [31] đã chỉ ra xu hướng phát triển mạnh mẽ của hệ chuyên gia mờ trong lĩnh vực chăm sóc sức khỏe.

Kunhimangalam và cộng sự [19] cải thiện kết quả chẩn đoán bệnh thần kinh ngoại bằng việc ứng dụng và phát triển FDSS. Hệ thống dựa vào các triệu chứng lâm sàng trong hồ sơ y tế điện tử của bệnh nhân, và áp dụng kỹ thuật suy luận mờ Mamdani để hỗ trợ đưa ra những chẩn đoán chính xác về bệnh thần kinh ngoại. Devi và Anto [11] đã đề xuất

một hệ chuyên gia mờ tiến hóa để chẩn đoán bệnh động mạch vành dựa trên bộ dữ liệu của phòng khám Cleveland cho bệnh tim. Cây quyết định được sử dụng để chọn các thuộc tính quan trọng nhất và đầu ra được chuyển đổi thành các quy tắc rõ "if-then". Các bộ quy tắc rõ được chuyển thành các quy tắc mờ. Một thuật toán di truyền (GA) sau đó được sử dụng để điều chỉnh các hàm thành viên mờ, và do đó thu được độ chính xác tốt hơn.

Frolova và cộng sự [13] đã sử dụng hệ suy diễn mờ để đánh giá việc sử dụng phân loại quốc tế về các mã bệnh sửa đổi lần thứ 10 từ khoa cấp cứu cho xác định suy tim cấp tính. Dragovic và cộng sự [12] thiết kế hệ suy diễn với sự hỗ trợ của các phụ thuộc mờ để ước tính khả năng bị bệnh viêm phúc mạc của bệnh nhân. Ngoài ra, Gayathri và Sumathi [14] lại hỗ trợ trong việc xác định khối u ung thư bằng suy luận mờ Mamdani. Từ đó, họ tính toán dự báo nguy cơ ung thư vú trên các thuộc tính bệnh chọn lựa.

Chakraborty và Mukherjee [8] đã sử dụng hệ suy diễn mờ để phát hiện bệnh Parkinson. Parkinson được chẩn đoán lâm sàng nhưng xác nhận trong phòng thí nghiệm khó để thu được và cần điều tra rất tinh vi với gánh nặng kinh tế cao. Thakur và cộng sự [32] đã sử dụng hệ suy diễn mờ MAMDANI để chẩn đoán bệnh Thalassemia bằng cách sử dụng mô tả đồ họa MATLAB 8.4 DOE. Một mô hình mờ được thiết kế cho bệnh Thalassemia chỉ ra rằng bệnh này có thể được chẩn đoán trong một số luật đầu vào mờ. Saikia và Dutta [28] áp dụng mờ hình thang và hệ suy diễn Mamdani để cải thiện kết quả chẩn đoán bệnh sốt xuất huyết. Hệ thống sử dụng suy luận mờ để dự báo ngăn chặn sự phát triển của bệnh sốt xuất huyết, từ đó đưa ra hướng điều trị thích hợp. Satyajit, Debashree và Bapi [10] đề xuất một hệ hỗ trợ chẩn đoán bệnh thông qua web nhằm mục đích giúp đỡ cho những người dân sống ở những vùng nông thôn bằng cách thông qua việc kết hợp cơ sở tri thức đã có và hệ suy diễn mờ trực cảm. Sardesai và cộng sự [29] lại nghiên cứu giảm thiểu sự phức tạp của bệnh phụ khoa bằng logic mờ. Nghiên cứu đã sử dụng hệ suy diễn mờ Mamdani để hỗ trợ trong quá trình chẩn đoán bệnh phụ khoa.

Năm 2017, một hệ thống lai dựa trên việc kết hợp hệ suy diễn mờ và các kỹ thuật phân cụm nhằm mục đích phân loại căn bệnh ung thư vú được đề xuất bởi Mehrbakhsh và cộng sự [25]. Nghiên cứu đã được thực nghiệm với độ chính xác khá cao (trên 94%) và được sử dụng như một hệ hỗ trợ quyết định lâm sàng cho các học viên y tế trong việc thực hành chăm sóc sức khỏe. Romero và cộng sự [27] lại phát triển một FDSS mới để cải thiện chẩn đoán bệnh đau cơ trong một tập hợp các bệnh mờ liên quan đến vấn đề đau cơ.

Gần đây, Somayeh Nazaria và cộng sự [24] đã trình bày một cách tiếp cận sáng tạo dựa trên quy trình phân cấp phân tích mờ (AHP) và hệ suy diễn mờ để đo lường khả năng phát triển bệnh tim ở bệnh nhân. Họ thực hiện AHP vì có một cấu trúc phân cấp giữa các tiêu chuẩn và tiêu chuẩn phụ tác động đến việc phát triển bệnh tim. Sau đó, hệ suy diễn mờ được áp dụng để phát triển hệ thống chuyên gia.

### III. HỆ SUY DIỄN MỜ PHỨC

#### A. Tập mờ phức

Khái niệm về tập mờ phức (Complex Fuzzy Set - CFS) và logic mờ phức (Complex Fuzzy Logic - CFL) [31, 32] đã được đề xuất bởi Ramot và các cộng sự như là một phần mở rộng của lý thuyết tập mờ và logic mờ. Một tập mờ phức được đặc trưng bởi một hàm thuộc giá trị phức  $\mu_S x$  mà phạm vi giá trị của nó là đường tròn đơn vị trong không gian phức, và được biểu diễn có dạng:

$$\mu_S x = r_S x \cdot e^{j\omega_S x}, \quad j = \sqrt{-1}$$

Trong đó:  $r_S x$  là biên độ và  $\omega_S x$  pha, và cả 2 đều là các hàm có giá trị thực với điều kiện  $r_S x \in [0, 1]$ . Pha không phải là 1 hàm mờ có thể nhận bất kỳ giá trị thực nào nên nó không ảnh hưởng tới mức độ thuộc. Pha bổ sung thêm một số thông tin mở rộng liên quan tới chu kỳ không gian và thời gian trong tập mờ đã được xác định bởi biên độ. Tuy nhiên khái niệm về tập mờ phức này khác với các khái niệm do Buckley [5, 6] và Zhang [15, 16] đưa ra. Các tập mờ phức giữ lại các đặc điểm về sự không chắc chắn dưới dạng biên độ, trong khi thêm vào thành phần pha để chỉ ra các thuộc tính dạng sóng.

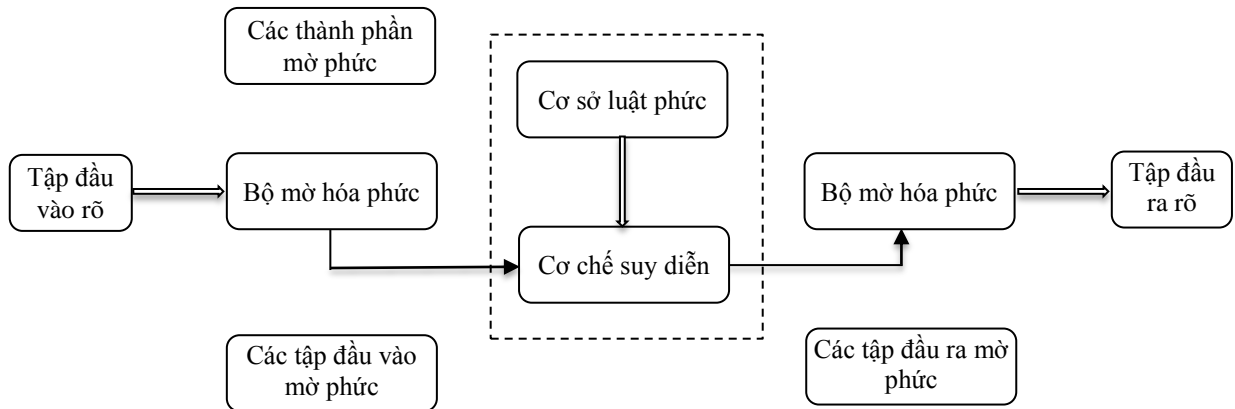
Trong định nghĩa được cung cấp bởi Ramot, giá trị tuyệt đối, hoặc phần biên độ của mức độ thành viên, hoạt động giống như trong tập mờ truyền thống. Giá trị của nó được ánh xạ vào khoảng  $[0, 1]$ . Mặt khác, thành phần pha của biểu thức không phải là hàm mờ; nó là một hàm có giá trị thực sự có thể nhận bất kỳ giá trị thực nào. Vai trò của pha là thêm thông tin liên quan đến không gian hay thời gian trong tập mờ cụ thể được định nghĩa bởi thành phần biên độ. Ví dụ, thông tin mờ liên quan đến hoạt động của mặt trời cùng với thông tin rõ ràng liên quan đến ngày đo hoạt động của mặt trời. Một ví dụ khác mà tập hợp mờ phức tạp có một sự hấp dẫn trực quan xuất phát từ thị trường chứng khoán. Một cách trực giác, sự tuần hoàn của thị trường chứng khoán cùng với ước tính dựa trên mũ của các giá trị hiện tại của cổ phiếu có thể được thể hiện bằng một cấp độ phức tạp của các thành viên như là một trong những đề xuất của Ramot. Biên độ truyền tải thông tin chứa trong tập mờ như "cổ phiếu mạnh" trong khi giai đoạn chuyển tải thông tin rõ nét về giai đoạn hiện tại của chu kỳ thị trường chứng khoán giá định.

**B. Hệ suy diễn dựa trên tập mờ phức**

Hệ suy diễn mờ phức chủ yếu gồm 4 thành phần: Các luật mờ phức, một bộ mờ hóa phức, một cơ chế suy diễn phức và một bộ giải mờ phức.

Một hệ thống CFIS là sự tổng quát hóa của mô hình Mendel’s theo nghĩa sử dụng các tập mờ và logic mờ phức như là tổng quát hóa của tập mờ và logic mờ. Một hệ thống CFIS có khả năng nắm bắt được tính không chắc chắn trong các dữ liệu có tính chất định kỳ/chu kỳ. CFIS/CFLS tương tự như FIS/FLS truyền thống, hệ thống CFIS/CFLS là một ánh xạ phi tuyến của vector dữ liệu đầu vào theo một tỷ lệ đầu ra. Một hệ thống CFIS/CFLS cơ bản được đặc trưng bởi cơ sở luật mờ phức chứa một tập các luật. Những hàm mờ phức được biểu diễn dưới dạng các câu if-then. Nghĩa là, một hệ thống CFIS/CFLS được khái quát hóa bằng cách thay thế các tập mờ và các hàm mờ trong FIS/FLS truyền thống bằng các tương đương phức của chúng.

Hệ suy diễn dựa trên tập mờ phức có cấu trúc chung như hình 3.1 sau:



**Hình 5.** Cấu trúc của hệ CFIS

Đầu ra của một hệ thống có thể được xác định thông qua 3 giai đoạn, như biểu diễn trong Hình 5. Giai đoạn đầu tiên là mờ hóa phức, được dùng để ánh xạ dữ liệu đầu vào rõ thành các tập dữ liệu đầu vào mờ. Các tập mờ này có thể có dạng phức hoặc không tùy thuộc vào từng ứng dụng.

Ở giai đoạn thứ 2, bước suy diễn mờ, sử dụng một cơ sở luật mờ phức để ánh xạ các tập dữ liệu đầu vào mờ thành các tập dữ liệu đầu ra mờ. Mỗi một luật được kết hợp với các tập dữ liệu đầu vào mờ có liên quan (đặc biệt là các tập dữ liệu xuất hiện ngay từ đầu). Tiếp theo, bằng cách thực hiện tích hợp Vector, các dữ liệu đầu ra mờ phức của các luật riêng biệt sẽ được kết hợp để tạo ra một tập đầu ra mờ phức đơn.

Giải mờ là giai đoạn cuối cùng của việc thực hiện ánh xạ bởi CFIS/CFLS. Trong giai đoạn này, giải mờ của tập dữ liệu đầu ra phức sẽ đưa ra một dữ liệu đầu ra rõ. Một cách tiếp cận khác để giải mờ của đầu ra mờ phức là bỏ qua tất cả các thành phần pha mà chỉ xem xét thành phần biên độ của tập đầu ra. Bất kỳ một kỹ thuật giải mờ nào khác được sử dụng trong FIS/FLS truyền thống có thể được sử dụng cho mục đích này.

Trong bài báo này, chúng tôi khái quát hóa phương pháp Wang-Mendel [40] bằng cách xây dựng các tập mờ phức và logic mờ phức để tạo thành CFIS/CFLS. Quá trình này bao gồm 4 bước:

*Bước 1: Xây dựng các vùng mờ phức.*

Trong bước này chúng tôi sẽ xây dựng các vùng mờ phức bằng cách chia thành từng khoảng miền và mỗi miền sẽ chứa lớp giá trị mờ phức của hàm thuộc đối với đầu vào hoặc đầu ra.

*Bước 2. Sinh các luật mờ phức.*

Trong bước này, chúng ta sẽ tạo ra các luật mờ phức từ các dạng dữ liệu số có tính chất định kỳ/ chu kỳ để xác định độ thuộc giá trị phức từ mỗi cặp dữ liệu có tính chu kỳ và tạo ra luật đầu vào- đầu ra. Giá trị thuộc lớp phức lớn nhất đối với từng dữ liệu đầu vào và đầu ra được lấy từ mỗi bộ dữ liệu có tính chu kỳ riêng biệt.

*Bước 3. Giản lược cơ sở luật.*

Từ các luật mờ phức, chúng ta có thể tạo ra được một cơ sở luật có kích thước bằng với tập dữ liệu định kỳ ban đầu vì mỗi điểm dữ liệu có tính chu kỳ riêng biệt tạo ra một luật đơn lẻ. Khi đó thật khó quản lý về kích thước của cơ sở luật phức. Để khắc phục khó khăn này, chúng tôi sẽ phải giản lược kích thước của cơ sở luật phức này và loại bỏ các xung đột, mỗi một luật phức được gán một mức độ giá trị phức dựa trên tích hợp tối đa của các tập dữ liệu đầu vào và các tập dữ liệu đầu ra riêng biệt.

*Bước 4. Ánh xạ đầu ra thông qua giải mờ phức.*

Trong giai đoạn cuối cùng này sẽ đưa ra ánh xạ giữa đầu vào và đầu ra bằng cách giải mờ phức của các dữ liệu đầu vào.

### C. Xây dựng hệ hỗ trợ quyết định chẩn đoán bệnh gan dựa trên hệ suy diễn mờ phức

Trong mô hình bài toán xây dựng hệ hỗ trợ quyết định dựa trên hệ suy diễn mờ phức có đầu vào là  $n$  bệnh nhân với các triệu chứng bệnh của mỗi bệnh nhân, ta cần nội suy ra yếu tố đầu ra tức chẩn đoán bệnh nhân có bệnh viêm gan hay không?

Input:  $P = (P_1, \dots, P_n)$  là danh sách các bệnh nhân,  $S = (S_1, \dots, S_m)$  là triệu chứng

Output: Nhận bệnh viêm gan (có bệnh/không có bệnh)

Việc chẩn đoán bệnh có hay không được thực hiện dựa trên hệ suy diễn mờ phức như sau:

*Bước 1:* Mô tả triệu chứng bệnh  $S = (S_1, \dots, S_m)$  đầu vào thành các tập mờ phức

Ở đây chúng tôi sử dụng hàm mờ phức Gaussian [21]

$$cGaussian(x, m, \sigma) = Re(cGaussian(x, m, \sigma)) + j Im(cGaussian(x, m, \sigma))$$

với

$$Re(cGaussian(x, m, \sigma)) = \exp\left[-0.5\left(\frac{x-m}{\sigma}\right)^2\right]$$

$$Im(cGaussian(x, m, \sigma)) = -\exp\left[-0.5\left(\frac{x-m}{\sigma}\right)^2\right]x\left(\frac{x-m}{\sigma^2}\right)$$

với  $\{m, \sigma\}$  là trung bình và dải của hàm Gaussian,  $x$  là một thành phần thuộc không gian nền  $U$

*Bước 2:* Xây dựng hệ luật mờ phức

- Khởi tạo bộ luật mờ phức ban đầu
- Xây dựng bộ luật mờ tối ưu bằng cách loại bỏ các luật dư thừa, luật yếu

*Bước 3:* Thực hiện quá trình suy diễn dựa trên bộ luật mờ tối ưu đã xây dựng ở bước 2

*Bước 4:* Giải mờ phức để tính toán kết quả cho giá trị đầu ra có bệnh/ không có bệnh

Mô hình hỗ trợ chẩn đoán bệnh gan gồm 2 phần chính: phần thứ nhất dành cho chuyên gia để thực hiện quá trình tạo ra hệ thống cơ sở luật phức ban đầu, phần thứ hai được dùng để suy luận ứng với giá trị các triệu chứng bệnh đầu vào. Quá trình suy luận chẩn đoán bệnh được thực hiện thông qua hệ suy diễn mờ phức để từ đó có thể có được giá trị đầu ra.

## IV. PHÂN TÍCH THỰC NGHIỆM

### A. Tập dữ liệu thực nghiệm

Trong thực nghiệm với bộ dữ liệu bệnh gan LiverDisease của các bệnh nhân được bác sĩ chỉ định các xét nghiệm cận lâm sàng với các xét nghiệm công thức máu và sinh hóa máu để chẩn đoán bệnh viêm gan. Từ đó có thể đánh giá bệnh viêm gan một cách rộng rãi, do đó rất có ý nghĩa trong việc phát hiện và theo dõi diễn tiến bệnh. Tập dữ liệu này gồm 4156 bệnh nhân đến khám và điều trị bệnh do rối loạn men gan tại Bệnh viện Gang thép Thái Nguyên và Bệnh viện Đa khoa Trung ương Thái Nguyên. Mỗi hồ sơ bệnh chứa các thông tin liên quan đến bệnh nhân đến khám và điều trị bệnh do rối loạn men gan. Trong số 4156 hồ sơ bệnh nhân, có 1202 hồ sơ bệnh nhân được chẩn đoán là mắc bệnh viêm gan.

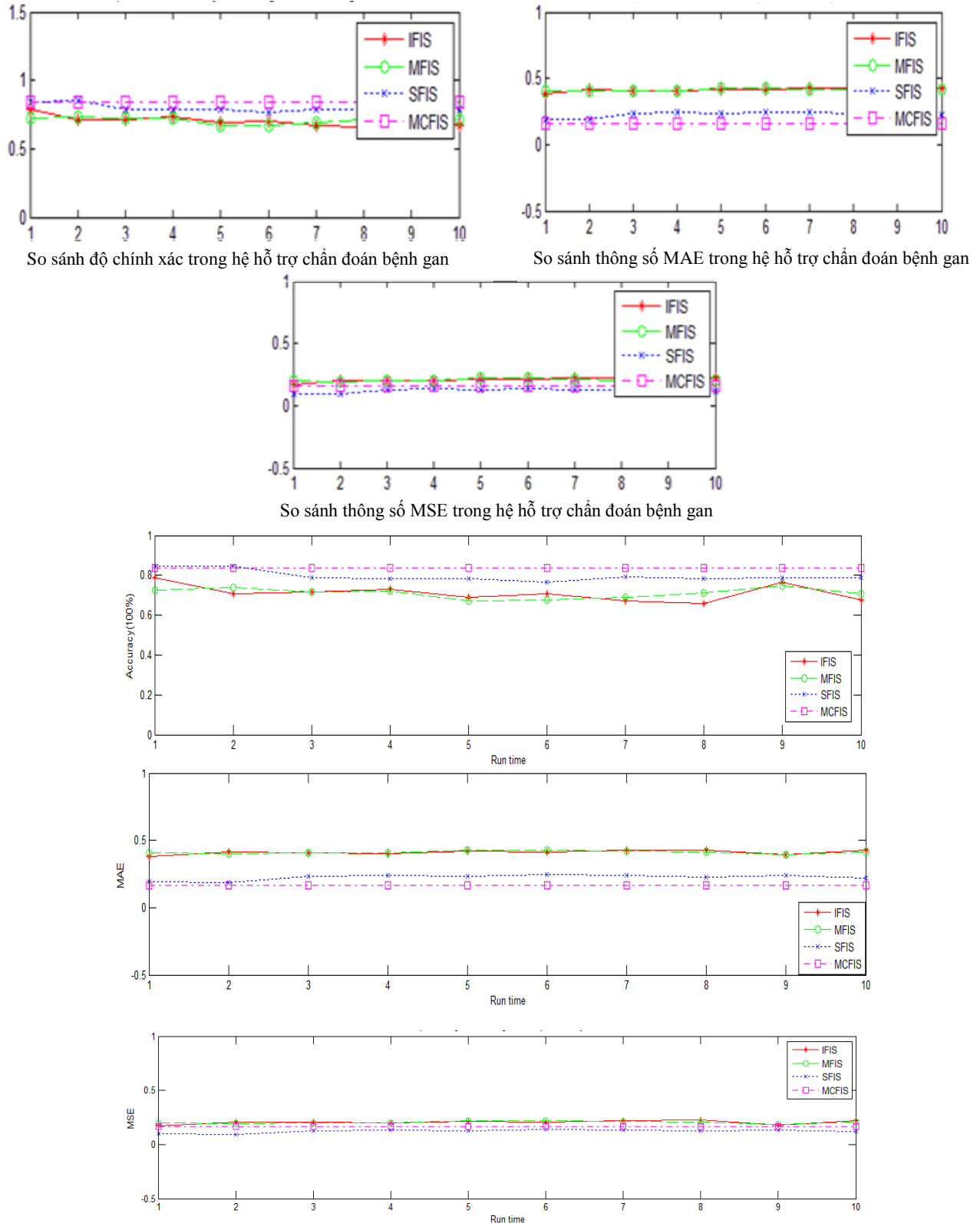
Thông tin về dữ liệu bệnh gan được mô tả trong Bảng 1 được trích từ hồ sơ bệnh án liên quan đến kết quả xét nghiệm (sinh hóa máu và công thức máu) và chẩn đoán bệnh từ các bác sĩ (các thông tin khác được bảo vệ).

**Bảng 1.** Các thuộc tính dữ liệu đầu vào trong tập dữ liệu bệnh gan

Ký hiệu	Thuộc tính	Mô tả
1	Tuổi (Age)	Tuổi tính đến ngày xét nghiệm
2	Giới tính (Gender)	0: nam; 1: nữ
3	AST	Chỉ số men AST
4	ALT	Chỉ số men ALT
5	Tiểu cầu (PLT)	Số tiểu cầu ( $10^9/l$ )
6	GGT	Chỉ số GGT
7	Albumin	Chỉ số Albumin
8	TB	Chỉ số Total Bilirubin
9	DB	Chỉ số Direct Bilirubin
10	ULN	Giới hạn bình thường trên của phòng xét nghiệm nơi gửi mẫu máu để đo ngưỡng AST huyết thanh.

**B. Thực nghiệm**

Ngôn ngữ để cài đặt thực nghiệm là Matlab 2014. Các độ đo dùng để đánh giá và so sánh hiệu năng của các thuật toán được cài đặt trong bài báo này gồm Accuracy (Acc) [17], MSE [17], MAE [17]. Các phương pháp để so sánh chẩn đoán bệnh gan gồm có: suy diễn mờ Mamdani (MFIS), Sugeno (SFIS), hệ suy diễn dựa trên tập mờ trực cảm (IFIS) và hệ suy diễn mờ phức (MCFIS). Kết quả chạy thực nghiệm được thể hiện rõ trong bảng 2 và hình 4.1. Thực nghiệm cho thấy độ chính xác của hệ suy diễn dựa trên tập mờ phức cao hơn các hệ suy diễn so sánh tuy nhiên thời gian chạy vẫn còn khá cao.



**Hình 6** Biểu đồ chạy thực nghiệm trên bộ dữ liệu gan

**Bảng 2.** Kết quả chạy thực nghiệm trên bộ dữ liệu gan

Hệ suy diễn	ACC	MAE	MSE
MFIS	0.7106	0.41135	0.20579
SFIS	0.79711	0.22729	0.12287
IFIS	0.71133	0.41125	0.2057
<b>MCFS</b>	<b>0.83686</b>	<b>0.16314</b>	0.16314

## V. KẾT LUẬN

Trong bài báo này chúng tôi sử dụng hệ suy diễn dựa trên tập mờ phức để hỗ trợ trong chẩn đoán bệnh viêm gan. Bài báo có một số đóng góp chính như sau: (i) đã vận dụng mô hình hệ suy diễn dựa trên tập mờ phức trong bài toán chẩn đoán bệnh; (ii) đã cài đặt thực nghiệm hệ suy diễn mờ phức trên bộ dữ liệu thực tế thu thập tại bệnh viện Gang thép Thái Nguyên và Bệnh viện Đa khoa Thái Nguyên nhằm hiện thực các bước xây dựng hệ hỗ trợ chẩn đoán bệnh; (iii) Kết quả thực nghiệm dựa trên 3 độ đo Accuracy, MSE, MAE đã chỉ ra phương pháp đề xuất của nhóm tốt hơn so với một số hệ suy diễn khác. Nghiên cứu này sẽ tạo tiền đề cho hướng phát triển với tập mờ phức, cho việc giải quyết một số bài toán hỗ trợ chẩn đoán bệnh khác trong tương lai.

## VI. LỜI CẢM ƠN

- Nghiên cứu này được thực hiện dưới sự tài trợ của đề tài sau tiến sĩ, mã số: GUST.STS.ĐT2017- TT02 từ Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam.

- Ngoài ra, nhóm tác giả xin chân thành cảm ơn sự hỗ trợ và hợp tác từ đơn vị phối hợp, Viện Công nghệ thông tin, Viện Hàn lâm Khoa học và Công nghệ Việt Nam.

## TÀI LIỆU THAM KHẢO

- [1]. Đỗ Văn Thành, Một cách tiếp cận ra quyết định trong chẩn đoán lâm sàng, Tạp chí tin học và điều khiển học, Viện Công nghệ Thông tin, 16(1), (2000), 52-58
- [2]. Nguyễn Thanh Thủy, Hệ trợ giúp và kiểm tra đơn thuốc chữa bệnh tăng huyết áp ES – TENSION, Tạp chí tin học và điều khiển học, Viện Công nghệ Thông tin, 12(3), (1996), 10-18
- [3]. Adeli, A., Neshat, M. (2010). A fuzzy expert system for heart disease diagnosis. Lecture Notes in Engineering and Computer Science. Hong Kong. March.
- [4]. Asogbon, M. G., Samuel, O. W., Omisore, M. O., & Awonusi, O. (2016). Enhanced neuro-fuzzy system based on genetic algorithm for medical diagnosis. *J Med Diagn Meth*, 5(205), 2
- [5]. Buckley, J. J. (1992). Fuzzy complex analysis II: integration. *Fuzzy Sets and Systems*, 49(2), 171-179
- [6]. Buckley, J. J., & Qu, Y. (1991). Fuzzy complex analysis I: differentiation. *Fuzzy Sets and Systems*, 41(3), 269-284
- [7]. Camlica, Z., Tizhoosh, H. R., & Khalvati, F. (2015, December). Medical image classification via SVM using LBP features from saliency-based folded data. In *Machine Learning and Applications (ICMLA), 2015 IEEE 14th International Conference on* (pp. 128-132). IEEE
- [8]. Chakraborty A., Chakraborty A., Mukherjee B. (2016) Detection of Parkinson's Disease Using Fuzzy Inference System. In: Berretti S., Thampi S., Srivastava P. (eds). Intelligent Systems Technologies and Applications. Advances in Intelligent Systems and Computing, vol 384. Springer, Cham
- [9]. Chaudhri, I. N., Dastmalchi, S. S., Rogers, R. D., Sethumadhavan, V., & Ansari, S. A. (2015). *U.S. Patent No. 9,043,901*. Washington, DC: U.S. Patent and Trademark Office
- [10]. Das, S., Guha, D., & Dutta, B. (2016). Medical diagnosis with the aid of using fuzzy logic and intuitionistic fuzzy logic. *Applied Intelligence*, 45(3), 850-867
- [11]. Devi, Y. N., Anto, S. (2014). An Evolutionary-Fuzzy Expert System for the Diagnosis of Coronary Artery Disease. *International Journal of Bio-Science and Bio-Technology*, 3(4), 1478-1484
- [12]. Dragović, I., Turajlić, N., Pilčević, D., Petrović, B., & Radojević, D. (2015). A Boolean consistent fuzzy inference system for diagnosing diseases and its application for determining peritonitis likelihood. *Computational and Mathematical Methods in Medicine*, 2015
- [13]. Frolova, N., Bakal, J. A., McAlister, F. A., Rowe, B. H., Quan, H., Kaul, P., Ezekowitz, J. A. (2015). Assessing the use of international classification of diseases-10th revision codes from the emergency department for the identification of acute heart failure. *JACC: Heart Failure*, 3(5), 386-91.
- [14]. Gayathri, B. M., & Sumathi, C. P. (2015, December). Mamdani fuzzy inference system for breast cancer risk detection. In *Computational Intelligence and Computing Research (ICCIC), 2015 IEEE International Conference on* (pp. 1-6). IEEE



- [15]. Guangquan, Z. (1992). Fuzzy Distance and Limit of Fuzzy Numbers [J]. *Fuzzy Systems and Mathematics*, 1, 003
- [16]. Guang-Quan, Z. (1992). Fuzzy limit theory of fuzzy complex numbers. *Fuzzy Sets and Systems*, 46(2), 227-235
- [17]. Hyndman R. J., & Koehler A. B. (2006), "Another look at measures of forecast accuracy", *International journal of forecasting*, 22(4), pp. 679-688.
- [18]. Ho, D. T., & Russell, J. A. (2018). Mercury and motor neuron disease: Hooked on a hypothesis. *Muscle & nerve*
- [19]. Kunhimangalam, R., Ovallath, S., & Joseph, P. K. (2014). A clinical decision support system with an integrated EMR for diagnosis of peripheral neuropathy. *Journal of medical systems*, 38(4), 38.
- [20]. Lee, C. S., Wang, M. H. (2011). A fuzzy expert system for diabetes decision support application. *IEEE Transactions on Systems, Man, and Cybernetics*, 41(1), 139-153
- [21]. Li, C., & Chiang, T. W. (2011). Function approximation with complex neuro-fuzzy system using complex fuzzy sets—a new approach. *New generation computing*, 29(3), 261
- [22]. Mendez, J. A., Leon, A., Marrero, A., Gonzalez-Cava, J. M., Reboso, J. A., Estevez, J. I., & Gomez-Gonzalez, J. F. (2018). Improving the anesthetic process by a fuzzy rule based medical decision system. *Artificial intelligence in medicine*
- [23]. Mitchell, R., & Chen, R. (2015). Behavior rule specification-based intrusion detection for safety critical medical cyber physical systems. *IEEE Transactions on Dependable and Secure Computing*, 12(1), 16-30
- [24]. Nazari, S., Fallah, M., Kazemipoor, H., & Salehipour, A. (2018). A fuzzy inference-fuzzy analytic hierarchy process-based clinical decision support system for diagnosis of heart diseases. *Expert Systems with Applications*, 95, 261-271.
- [25]. Nilashi, M., Ibrahim, O., Ahmadi, H., & Shahmoradi, L. (2017). A knowledge-based system for breast cancer classification using fuzzy logic method. *Telematics and Informatics*, 34(4), 133-144.
- [26]. Opeyemi, O., Justice, E. O. (2012). Development of neuro-fuzzy system for early prediction of heart attack. *International Journal of Information Technology and Computer Science*, 4(9), 22-8.
- [27]. Romero-Córdoba, R., Olivás, J. A., Romero, F. P., Alonso-Gonzalez, F., & Serrano-Guerrero, J. (2017). An application of fuzzy prototypes to the diagnosis and treatment of fuzzy diseases. *International Journal of Intelligent Systems*, 32(2), 194-210
- [28]. Saikia, D., & Dutta, J. C. (2016, January). Early diagnosis of dengue disease using fuzzy inference system. In *Microelectronics, Computing and Communications (MicroCom), 2016 International Conference on* (pp. 1-6). IEEE
- [29]. Sardesai, A., Kharat, V., Sambarey, P., & Deshpande, A. (2016). Fuzzy Logic-Based Formalisms for Gynecology Disease Diagnosis. *Journal of Intelligent Systems*, 25(2), 283-295
- [30]. Sikchi, S. S., Sikchi, S., Ali, M. S. (2013). Fuzzy expert systems (FES) for medical diagnosis. *International Journal of Computer Applications*, 63(11), 7-16
- [31]. Sun, C., Wang, G., Liu, C., Jin, S., Hu, Y., & Wu, L. (2017). Research on the Application of Reading Recommendation System in High Medical College Library Based on the Local Data. *DEStech Transactions on Social Science, Education and Human Science*, (ssme).
- [32]. Thakur, S. A. P. N. A., Raw, S. N., Sharma, R. A. V. I. N. D. R. A. (2016). Design of a Fuzzy Model for Thalassaemia Disease Diagnosis: Using Mamdani Type Fuzzy Inference System. *International Journal of Pharmacy and Pharmaceutical Sciences*, 8(4), 356-361.
- [33]. Thong, N. T., Son, L.H. (2015). HIFCF: An effective hybrid model between picture fuzzy clustering and intuitionistic fuzzy recommender systems for medical diagnosis. *Expert Systems With Applications*, 42(7), 3682-3701
- [34]. Yelipe, U., Porika, S., & Golla, M. (2017). An efficient approach for imputation and classification of medical data values using class-based clustering of medical records. *Computers & Electrical Engineering*
- [35]. Zadeh, L. A. (1972). A fuzzy-set-theoretic interpretation of linguistic hedges.
- [36]. Zaidi, N. L. B., Santen, S. A., Purkiss, J. A., Teener, C. A., & Gay, S. E. (2016). A Hybrid Interview Model for Medical School Interviews: Combining Traditional and Multisampling Formats. *Academic Medicine*, 91(11), 1526-1529
- [37]. Zhang, Y., Chen, M., Huang, D., Wu, D., & Li, Y. (2017). iDoctor: Personalized and professionalized medical recommendations based on hybrid matrix factorization. *Future Generation Computer Systems*, 66, 30-35
- [38]. Zhu, H., Liu, X., Lu, R., & Li, H. (2017). Efficient and privacy-preserving online medical prediagnosis framework using nonlinear SVM. *IEEE journal of biomedical and health informatics*, 21(3), 838-850.

## DECISION MAKING SUPPORT SYSTEM IN DISEASE DIAGNOSIS: AN APPROACH USING COMPLEX FUZZY INFERENCE SYSTEM

Luong Thi Hong Lan, Tran Thi Ngan, Hoang Thi Minh Chau, Le Bich Lien, Roan Thi Ngan

**ABSTRACT:** Disease diagnosis is an important stage in treatment process. It helps the doctor can diagnosed early and accurately. The construction of disease diagnosis system have attracted many scientists. In particular, the approach from fuzzy set has also been used in many previous reseaches. In this paper, we propose a novel model of decision making support system in disease diagnosis using complex fuzzy inference system. We also present the experiments of proposed model on real liver dataset. The experimental results show that our model has a higher performance than other available ones in disease diagnosis.