

Phát triển ứng dụng di động dựa trên tri thức chuyên gia hỗ trợ sàng lọc nguy cơ mang gen bệnh Thalassemia

Phạm Quang Anh¹, Lê Duy Anh¹, Đặng Tiên Dũng², Nguyễn Phương Linh¹,
Nguyễn Thị Trang³, Trần Danh Cường³, Trần Thị Vân Anh^{1*}

¹Trường Đại học Y Dược Thái Bình, 373 Lý Bôn, phường Trần Lãm, tỉnh Hưng Yên, Việt Nam

²Trường Đại học Công nghệ, Đại học Quốc gia Hà Nội, 144 Xuân Thủy, phường Cầu Giấy, Hà Nội, Việt Nam

³Trường Đại học Y Hà Nội, 1 Tôn Thất Tùng, phường Phương Liệt, Hà Nội, Việt Nam

Ngày nhận bài 28/5/2024; ngày chuyển phản biện 1/6/2024; ngày nhận phản biện 23/6/2024; ngày chấp nhận đăng 26/6/2024

Tóm tắt:

Mục tiêu của nghiên cứu là xây dựng cơ sở dữ liệu tri thức chuyên gia về kết quả xét nghiệm và đánh giá độ chính xác, độ nhạy, độ đặc hiệu của ứng dụng di động hỗ trợ sàng lọc nguy cơ mang gen Thalassemia. Nghiên cứu thu thập các tri thức chuyên gia về sàng lọc Thalassemia trên các tài liệu, báo cáo uy tín. Hồi cứu và mô tả cắt ngang trên hồ sơ bệnh án gồm kết quả xét nghiệm tổng phân tích tế bào máu ngoại vi, kết quả chẩn đoán gen Thalassemia, tiền sử liên quan đến Thalassemia của 233 thai phụ và chồng đến sàng lọc bệnh Thalassemia tại Bệnh viện Phụ sản Trung ương từ tháng 8/2023 đến tháng 6/2024. Kết quả thu được cho thấy, tuổi trung bình của các thai phụ tham gia nghiên cứu là 31 tuổi, chỉ có 8,61% đối tượng nghiên cứu có tiền sử sinh con mắc Thalassemia. Thể bệnh hay gặp nhất là thể mang gen alpha-Thalassemia (54,08%). Đã có 34 luật được xây dựng cho cơ sở dữ liệu hệ tri thức chuyên gia sàng lọc nguy cơ mang gen bệnh Thalassemia. Kết quả của ứng dụng di động từ hệ tri thức chuyên gia đạt độ chính xác 91,30%, độ đặc hiệu 72,73% và độ nhạy lên đến 99,22%. Phần mềm được ứng dụng dưới dạng app điện thoại di động, giao diện thân thiện, dễ sử dụng.

Từ khóa: hệ chuyên gia, Thalassemia, trí tuệ nhân tạo, ứng dụng di động.

Chỉ số phân loại: 1.2, 3.1, 3.3

Mobile application development based on expert knowledge supporting screening Thalassemia carriers

Quang Anh Pham¹, Duy Anh Le¹, Tien Dung Dang², Phuong Linh Nguyen¹,
Thi Trang Nguyen³, Danh Cuong Tran³, Thi Van Anh Tran^{1*}

¹Thai Binh University of Medicine and Pharmacy, 373 Ly Bon Street, Tran Lam Ward, Hung Yen Province, Vietnam

²University of Engineering and Technology, Vietnam National University - Hanoi, 144 Xuan Thuy Street, Cau Giay Ward, Hanoi, Vietnam

³Hanoi Medical University, 1 Ton That Tung Street, Phuong Liet Ward, Hanoi, Vietnam

Received 28 May 2024; revised 23 June 2024; accepted 26 June 2024

Abstract:

The goal of the research is to build a knowledge base for an expert system on testing results and evaluating the mobile application's accuracy, sensitivity, and specificity, supporting screening of Thalassemia carriers. This research collected the knowledge base about Thalassemia in reputable documents and reports. A retrospective, cross-sectional study was carried out based on medical records, including complete blood count results, Thalassemia genetic diagnosis results, and Thalassemia-related medical history of 233 pregnant women and husbands who came to screen for Thalassemia at the National Hospital of Obstetrics and Gynaecology from August 2023 to June 2024. The results have shown that the average age of the pregnant women who joined the survey was 31 years old, only 8.61% of the study subjects had a history of giving birth to a child with Thalassemia. The most common type was alpha Thalassemia carrier (54.08%). 34 rules were developed for the knowledge base of the expert system for screening the risk of carrying the Thalassemia gene. The mobile application results from the expert knowledge system achieved an accuracy of 91.30%, a specificity of 72.73%, and a sensitivity of up to 99.22%. The software is applied in the form of a mobile phone app with a friendly and easy-to-use interface.

Keywords: artificial intelligence, expert system, mobile application, Thalassemia.

Classification numbers: 1.2, 3.1, 3.3

*Tác giả liên hệ: Email: tva7377@gmail.com

1. Đặt vấn đề

Thalassemia là một bệnh rối loạn di truyền ảnh hưởng đến quá trình tổng hợp huyết sắc tố, gây thiếu máu mạn tính và kéo theo nhiều biến chứng nghiêm trọng ở tim, gan, xương, tuyến nội tiết, tác động nặng nề đến sức khỏe giống nòi cũng như chất lượng cuộc sống của người bệnh và cộng đồng. Theo thống kê của Tổ chức Y tế Thế giới (WHO) năm 2018, khoảng 5,2% dân số mang gen Thalassemia, 1,1% các cặp vợ chồng có nguy cơ sinh con mắc rối loạn huyết sắc tố và 2,7/1000 trường hợp thụ thai bị ảnh hưởng [1]. Tại Việt Nam, tỷ lệ người mang gen bệnh dao động 7-10%, với khoảng 20.000 bệnh nhân Thalassemia thể nặng và hơn 2.000 trẻ sơ sinh mắc bệnh mỗi năm [2].

Biện pháp phòng ngừa Thalassemia chủ yếu là ngăn chặn việc sinh ra trẻ mắc bệnh, vì vậy cần tiến hành sàng lọc để phát hiện sớm các thai phụ và chồng có nguy cơ mang gen. Vào tháng 4/2020, Bộ Y tế đã ban hành hướng dẫn sàng lọc Thalassemia trước sinh trong quý I của thai kỳ, quy định tất cả thai phụ bắt buộc phải làm xét nghiệm tổng phân tích tế bào máu ngoại vi, kết hợp khai thác tiền sử bệnh bản thân và gia đình, cùng kết quả xét nghiệm máu của chồng. Mục tiêu là nhận diện sớm người mang gen để có biện pháp dự phòng và can thiệp kịp thời [3]. Tuy nhiên, việc sàng lọc trước sinh bệnh này cho tất cả thai phụ tại Việt Nam còn gặp nhiều khó khăn, đặc biệt tại các tuyến dưới hoặc các vùng dân tộc thiểu số bởi sự hiểu biết về bệnh này còn hạn chế không chỉ ở các thai phụ, gia đình thai phụ mà còn ở cả các nhân viên y tế. Ngoài ra, hiện nay, trí tuệ nhân tạo đang phát triển mạnh mẽ và được ứng dụng ngày càng nhiều trong các lĩnh vực của đời sống, trong đó có y tế. Chính vì vậy, việc xây dựng một phần mềm trí tuệ nhân tạo với kiến thức tổ hợp từ hệ tri thức chuyên gia (Expert System) có thể phân tích đồng thời nhiều dữ liệu và thông tin về công thức máu hay tiền sử gia đình, có thể đáp ứng nhu cầu hỗ trợ các nhân viên y tế các tuyến trong việc sàng lọc Thalassemia. Chỉ cần thao tác đơn giản là nhập số liệu thông tin về thai phụ và kết quả công thức máu, phần mềm có thể tự động đưa ra kết luận nguy cơ mắc bệnh cao hay thấp và phân trầm khả năng dự báo nguy cơ.

Xuất phát từ những đặc điểm trên, nhóm nghiên cứu đã tiến hành thực hiện đề tài “Nghiên cứu xây dựng ứng dụng di động trên nền tảng trí tuệ nhân tạo hỗ trợ sàng lọc nguy cơ mang gen bệnh Thalassemia” với mục tiêu: Xây dựng cơ

sở dữ liệu tri thức chuyên gia về kết quả xét nghiệm và đánh giá độ chính xác, độ nhạy, độ đặc hiệu của ứng dụng di động hỗ trợ sàng lọc nguy cơ mang gen Thalassemia.

2. Đối tượng và phương pháp nghiên cứu

2.1. Đối tượng

Đối tượng nghiên cứu cho xây dựng cơ sở dữ liệu tri thức chuyên gia: Các báo cáo, tài liệu, bài báo và hướng dẫn uy tín về Thalassemia trong giai đoạn 2007-2023.

Đối tượng nghiên cứu cho đánh giá ứng dụng di động: Kết quả xét nghiệm của các cặp vợ chồng thai phụ đến khám sàng lọc trước sinh bệnh Thalassemia tại Bệnh viện Phụ sản Trung ương từ tháng 8/2023 đến tháng 6/2024.

Tiêu chuẩn lựa chọn: Đối với xây dựng cơ sở dữ liệu tri thức chuyên gia: Tài liệu chính thống, được công bố, được phép sử dụng; khuyến cáo của cơ quan y tế liên quan chủ đề Thalassemia.

Đối với việc đánh giá ứng dụng di động: Thai phụ/chồng thai phụ có kết quả xét nghiệm máu, có/không có chẩn đoán gen Thalassemia, có/không có tiền sử sinh con/gia đình mắc Thalassemia. Ưu tiên trường hợp đã có chẩn đoán gen chính xác.

Tiêu chuẩn loại trừ: Đối với xây dựng cơ sở dữ liệu tri thức chuyên gia: Nghiên cứu trùng lặp, không phù hợp tiêu chí, không đo lường dữ liệu.

Đối với việc đánh giá ứng dụng di động: Thiếu kết quả xét nghiệm máu, gen, mắc bệnh tâm thần và các bệnh lý làm ảnh hưởng chỉ số huyết học.

2.2. Phương pháp nghiên cứu

Thiết kế nghiên cứu: Mô tả cắt ngang, hồi cứu dữ liệu.

Mẫu nghiên cứu: Cỡ mẫu và cách chọn mẫu

$$TP + FN = \frac{Z_{\alpha}^2 \times p_{se} \times (1 - p_{se})}{\omega^2}$$

$$TN + FP = \frac{Z_{\alpha}^2 \times p_{sp} \times (1 - p_{sp})}{\omega^2}$$

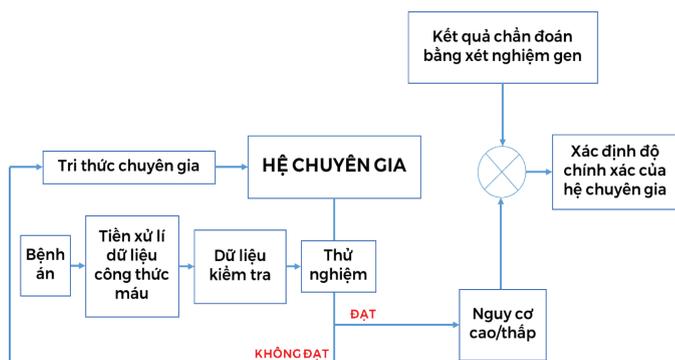
$$N_1 = \frac{TP + FN}{p} = 140$$

$$N_2 = \frac{TN + FP}{1 - p} = 21$$

trong đó: TP, FN, TN, FP lần lượt là số dương tính thật, âm tính giả, âm tính thật và dương tính giả. N_1 là cỡ mẫu nghiên cứu để đánh giá độ nhạy; N_2 là cỡ mẫu nghiên cứu để đánh giá độ đặc hiệu; α là sai lầm loại I. Với khoảng tin cậy là 95%, ta có $\alpha = 0,05$. Như vậy, Z_{α} là 1,96, với ω là chi

số dao động của độ nhạy và độ đặc hiệu ($\omega=0,1$), p_{se} là xác suất dương tính thật (hay độ nhạy) tối thiểu. Ta có $p_{se}=0,95$, với p_{sp} là xác suất âm tính thật (hay độ đặc hiệu) tối thiểu. Ta có $p_{sp}=0,95$, với p là tỷ lệ mang gen bệnh Thalassemia, đến khám tại Bệnh viện Phụ sản Trung ương ước tính có khoảng 13% [4].

Quy trình nghiên cứu được thể hiện ở hình 1.



Hình 1. Quy trình nghiên cứu.

Bước 1: Thu thập và làm sạch số liệu để xây dựng cơ sở dữ liệu nhằm thử nghiệm ứng dụng: Lấy đầy đủ các giá trị đầu vào của phần mềm bao gồm: các chỉ số tổng phân tích tế bào máu ngoại vi: MCV, MCH, HGB, RBC; các chỉ số điện di huyết sắc tố: HbA2, HbF; tiền sử bệnh nhân: tiền sử từng sinh con mắc Thalassemia, tiền sử gia đình có người mắc Thalassemia, tiền sử phù thai (bảng 1).

Bước 2: Tìm kiếm, thu thập, rà soát, loại trừ và soạn thảo các tài liệu để xây dựng tập luật tri thức chuyên gia về sàng lọc người mang gen Thalassemia: Một số tài liệu chính được dùng để xây dựng tập luật có thể kể đến như: Hướng dẫn chẩn đoán và điều trị một số bệnh lý huyết học ban hành kèm theo Quyết định số 1832/QĐ-BYT ngày 1/7/2022 của Bộ trưởng Bộ Y tế [5]; Quyết định số 921/QĐ-BYT ngày 18/9/2014 về việc ban hành tài liệu chuyên môn “Hướng dẫn chẩn đoán và điều trị bệnh Hemophilia và Thalassemia” [6]; hướng dẫn phòng ngừa và chẩn đoán các bệnh lý huyết sắc tố dành cho chuyên gia y tế và các nhà nghiên cứu của J. Old (2016) [7]; hướng dẫn sàng lọc người mang gen bệnh Thalassemia và các bệnh lý huyết sắc tố tại Canada của S. Langlois và cs (2008) [8]; nhận xét vai trò của xét nghiệm Hemoglobin A2 (HbA2) trong chẩn đoán bệnh Thalassemia và các bệnh lý huyết sắc tố liên quan của A. Mosca và cs (2009) [9]; nghiên cứu về đặc điểm di truyền, quản lý lâm sàng và điều trị bệnh β -Thalassemia của A. Cao và cs (2010) [10]; các chương trình phòng ngừa bệnh Thalassemia của A. Cao và cs (2013) [11].

Bảng 1. Ví dụ về tập luật.

Thứ tự	Luật	Kết luận
1	IF HGB \geq 110	THEN Tổng phân tích tế bào máu ngoại vi bình thường
	AND MCV \geq 80	
	AND MCV \leq 100	
	AND MCH \geq 28	
	AND MCH \leq 32	
2	IF HGB \geq 110	THEN Hồng cầu nhỏ
	AND MCV $<$ 80	
	AND MCH \geq 28	
	AND MCH \leq 32	
	AND MCH \leq 32	

Trong ngôn ngữ lập trình, “IF” là lệnh nếu, “AND” là lệnh và, “THEN” là kết quả. Ví dụ, một lệnh IF a AND b THEN c có nghĩa là nếu thỏa mãn điều kiện a và b thì kết quả c sẽ xảy ra. Các chỉ số HGB, MCV, MCH lần lượt là nồng độ hemoglobin trong máu, thể tích trung bình hồng cầu và nồng độ hemoglobin trung bình hồng cầu. Các dấu “ \geq ”, “ \leq ”, “ $<$ ” là các toán tử thể hiện lần lượt các giá trị sau “lớn hơn hoặc bằng”, “nhỏ hơn hoặc bằng”, “nhỏ hơn”. Cột “Đề xuất” sẽ được dùng để ghi các khuyến cáo cho các nhân viên y tế tùy thuộc vào kết luận.

Bước 3: Thử nghiệm tập luật.

Bước 4: Lựa chọn giải pháp kỹ thuật công nghệ, giải pháp phân quyền sử dụng và giải pháp dữ liệu, giải pháp sao lưu phục hồi.

Bước 5: Thiết kế mô hình triển khai, mô hình xây dựng ứng dụng, mô hình kiến trúc ứng dụng.

Bước 6: Thiết kế giao diện.

Bước 7: Đánh giá độ nhạy, độ đặc hiệu và độ chính xác của phần mềm.

Phân tích và xử lý số liệu: Các kết quả đưa ra từ phần mềm dự báo sau đó được phân tích bằng phương pháp thống kê y học theo chương trình SPSS 16.0 và Stata.

Đạo đức trong nghiên cứu: Nghiên cứu được sự cho phép của Trường Đại học Y Dược Thái Bình và được thông qua Hội đồng Y đức của Bệnh viện Phụ sản Trung ương số 1042/CN-PSTW ngày 29/12/2020.

3. Kết quả

3.1. Mô tả đặc điểm bộ dữ liệu input và output

Nghiên cứu sàng lọc trước sinh bệnh Thalassemia tại Bệnh viện Phụ sản Trung ương từ tháng 8/2023 đến tháng 6/2024 thu thập 233 trường hợp, bao gồm 151 thai phụ và 82 chồng thai phụ. Tuổi trung bình của các đối tượng trong nghiên cứu là 31 tuổi. Chỉ có 8,61% thai phụ có tiền sử sinh con mắc Thalassemia. Thể bệnh hay gặp nhất là thể mang gen alpha-Thalassemia (54,08%), không gặp thể mang phối hợp gen alpha và beta Thalassemia (0%) (bảng 2).

Bảng 2. Thông tin chung của đối tượng nghiên cứu.

Đặc điểm	Số lượng (n, %)	
Tuổi trung bình	31	
Giới (n, %)	Nam	82 (35,19%)
	Nữ	151 (64,81%)
Tiền sử (n, %)	Sinh con mắc Thalassemia	13 (8,61%)
	Sinh con không mắc Thalassemia	138 (91,39%)
Thể bệnh Thalassemia (n, %)	Không mang gen Thalassemia	76 (32,62%)
	Mang gen alpha - Thalassemia	126 (54,08%)
	Mang gen beta - Thalassemia	31 (13,30%)
	Mang phối hợp gen alpha và beta Thalassemia	0 (0%)

Phần mềm hệ tri thức chuyên gia được thiết kế dưới dạng một cơ sở tri thức sử dụng các tập luật IF-THEN và một bộ máy suy luận, có khả năng dự đoán kết quả dựa trên dữ liệu đầu vào. Cơ sở tri thức của phần mềm được xây dựng từ các hướng dẫn và khuyến cáo mới nhất của các tổ chức uy tín trong và ngoài nước như WHO, Liên đoàn Thalassemia quốc tế (TIF), Bộ Y tế và được thẩm định bởi đội ngũ chuyên gia giàu kinh nghiệm. Do đó, cơ sở tri thức và các luật của phần mềm là yếu tố quan trọng nhất quyết định hiệu quả của quá trình sàng lọc. Trên thế giới, phương pháp phổ biến để sàng lọc người mang gen Thalassemia là sử dụng hai chỉ số MCV và MCH. Tình trạng hồng cầu nhỏ nhược sắc với $MCV < 80$ fl và $MCH < 27$ pg là 2 chỉ số thông dụng nhất cho mục đích này. Tuy nhiên, Thalassemia là bệnh lý gây sự giảm hoặc mất khả năng tổng hợp chuỗi globin, tạo ra các hemoglobin bất thường, ảnh hưởng đến tất cả các chỉ số huyết học chứ không chỉ riêng MCV và MCH. Do đó, tập luật của phần mềm trong nghiên cứu này được xây dựng dựa trên sự kết hợp của nhiều chỉ số khác nhau, bao gồm: HGB, MCV, MCH, ferritin và các tiền sử liên quan đến Thalassemia. Các loại tập luật nhóm nghiên cứu đã xây dựng được trình bày trong bảng 3.

Bảng 3. Cơ sở dữ liệu tri thức chuyên gia.

Luật	Số lượng luật
Luật sử dụng các chỉ số tổng phân tích tế bào máu ngoại vi	12
Luật sử dụng kết quả ferritin huyết thanh	2
Luật kết hợp công thức máu, ferritin huyết thanh	8
Luật kết hợp công thức máu, ferritin huyết thanh và điện di huyết sắc tố	11
Luật liên quan đến tiền sử bệnh nhân	1

Các luật trong cơ sở dữ liệu tri thức chuyên gia được xây dựng dưới dạng IF-THEN và có tổng cộng 34 luật đã được xây dựng.

3.2. Độ chính xác và khả năng dự báo nguy cơ của ứng dụng di động từ hệ tri thức chuyên gia trong sàng lọc trước sinh bệnh Thalassemia

Kết luận của hệ chuyên gia về công thức máu và tình trạng sắt của đối tượng nghiên cứu được trình bày trong bảng 4. Trong đó, tình trạng hồng cầu nhỏ nhược sắc (36,48%) và tình trạng không thiếu sắt (71,67%) chiếm tỷ lệ cao nhất. Có 5 trường hợp thiếu sắt đều có tình trạng hồng cầu nhỏ nhược sắc, kết quả cho thấy 1 trường hợp có mang gen Thalassemia.

Bảng 4. Kết luận của hệ tri thức chuyên gia.

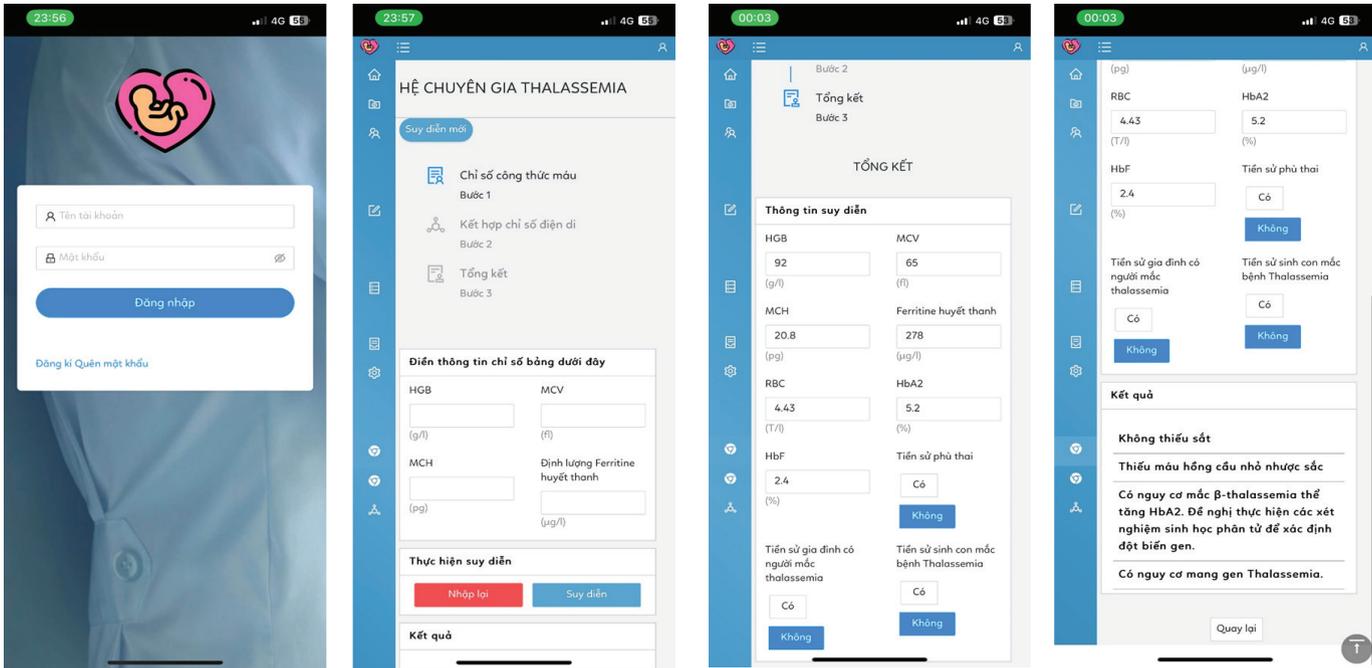
Kết luận tổng phân tích tế bào máu ngoại vi	Số lượng (n, %)
Tổng phân tích tế bào máu ngoại vi bình thường	53 (22,75%)
Hồng cầu nhược sắc	8 (3,43%)
Hồng cầu nhỏ nhược sắc	85 (36,48%)
Thiếu máu hồng cầu bình thường	6 (2,58%)
Thiếu máu hồng cầu nhỏ nhược sắc	73 (31,33%)
Hồng cầu ưu sắc	8 (3,43%)
Kết luận về tình trạng sắt	
Thiếu sắt	5 (2,15%)
Không thiếu sắt	167 (71,67%)
Không xác định được	61 (26,18%)

Kết luận của hệ tri thức chuyên gia sau đó được so sánh với kết quả chẩn đoán chính xác từ các chuyên gia để đánh giá hiệu quả của phần mềm. Các giá trị thống kê được trình bày ở bảng 5, ứng dụng di động từ hệ tri thức chuyên gia cho thấy hiệu quả 72,73% với độ đặc hiệu và độ chính xác lên đến 91,30%.

Bảng 5. Đánh giá kết quả của ứng dụng di động từ hệ tri thức chuyên gia trong chẩn đoán khả năng mang gen Thalassemia.

Chỉ số hiệu lực của biện pháp	Tỷ lệ (%)
Độ nhạy	99,22
Độ đặc hiệu	72,73
Độ chính xác	91,30
Giá trị dự báo dương tính	89,51
Giá trị dự báo âm tính	97,56

Giao diện của ứng dụng được xây dựng dựa trên nền tảng ReactJS nên có thể hiển thị trên mọi loại hệ điều hành và tự động căn chỉnh vừa với kích cỡ của màn hình. Khi người dùng chạy ứng dụng, một giao diện đăng nhập sẽ hiển thị yêu cầu người dùng nhập tài khoản để truy cập. Sau khi đăng nhập thành công, người dùng sẽ được đưa tới giao diện hệ chuyên gia Thalassemia và màn hình đang hiển thị nơi để nhập các dữ liệu cần thiết về tổng phân tích tế bào máu ngoại vi, điện di huyết sắc tố và tiền sử. Sau đó, khi người dùng nhấn “suy diễn”, phần mềm sẽ trả lại kết quả dự báo nguy cơ mang gen Thalassemia cùng các kiến nghị cho nhân viên y tế (hình 2).



Hình 2. Giao diện và thử nghiệm của ứng dụng di động hỗ trợ sàng lọc nguy cơ mang gen bệnh Thalassemia.

4. Bàn luận

Thalassemia là một trong những rối loạn di truyền phổ biến nhất của huyết sắc tố, gây ảnh hưởng nghiêm trọng đến chất lượng cuộc sống và tuổi thọ của người bệnh và gia đình của họ nói riêng cũng như toàn dân số nói chung. Vì vậy, các biện pháp sàng lọc phát hiện người mang gen gây bệnh được coi là một công cụ hiệu quả nhằm phát hiện sớm và ngăn ngừa sự lưu truyền của bệnh trong quần thể dân số. Trọng tâm của các biện pháp này là sàng lọc trước sinh, tức là sàng lọc trên các cặp vợ chồng thai phụ đến khám thai với mục tiêu phát hiện những người mang gen gây bệnh trong các cặp vợ chồng và có nguy cơ sinh ra những đứa trẻ bị Thalassemia. Từ đó, có thể đưa ra chẩn đoán sớm thai phụ mắc các thể nặng ở tuổi thai nhỏ và tư vấn di truyền cho gia đình trong quá trình mang thai hiện tại cũng như các lần mang thai tiếp theo.

Nghiên cứu sàng lọc trước sinh bệnh Thalassemia tại Bệnh viện Phụ sản Trung ương từ tháng 8/2023 đến tháng 6/2024 thu thập 233 trường hợp, bao gồm 151 thai phụ và 82 chồng thai phụ (bảng 2). Kết quả xét nghiệm trên 233 trường hợp cho thấy, tỷ lệ mang gen alpha-Thalassemia chiếm cao nhất 54,08%, kế đến là beta-Thalassemia với 13,30%. Các nghiên cứu trước đây cũng ghi nhận tỷ lệ mang gen alpha-Thalassemia cao hơn so với các thể bệnh khác, dao động 16-18% [12, 13]. Tỷ lệ trường hợp mang gen Thalassemia trong nghiên cứu của chúng tôi cao hơn nhiều so với các nghiên cứu này, nguyên nhân là do chúng tôi ưu tiên lựa chọn những trường hợp đã có kết quả chẩn đoán gen chính xác, chủ yếu là các trường hợp có kết quả sàng lọc nguy cơ cao trong việc đánh giá thử nghiệm phần mềm, nhằm mục đích tránh bỏ sót những trường hợp mắc Thalassemia. Về tuổi thai phụ (bảng 2), tuổi trung bình là 31.

Nghiên cứu của chúng tôi sử dụng phần mềm cơ sở tri thức chuyên gia, được xây dựng dựa trên cơ sở tri thức từ các chuyên gia về bệnh lý Thalassemia, để ứng dụng sàng lọc nguy cơ mang gen Thalassemia của các cặp vợ chồng.

Hệ tri thức chuyên gia được xây dựng dựa trên một tập luật IF-THEN, có khả năng dự đoán kết quả đầu ra dựa trên dữ liệu đầu vào. Cơ sở tri thức của phần mềm được xây dựng từ các hướng dẫn và khuyến cáo mới nhất của các tổ chức uy tín trong và ngoài nước và được thẩm định bởi đội ngũ chuyên gia giàu kinh nghiệm. Do đó, cơ sở tri thức và các luật của phần mềm là yếu tố quan trọng nhất quyết định hiệu quả của quá trình sàng lọc. Trên thế giới, phương pháp phổ biến để sàng lọc người mang gen Thalassemia là sử dụng hai chỉ số MCV và MCH. Tình trạng hồng cầu nhỏ nhược sắc với $MCV < 80$ fl và $MCH < 28$ pg là hai chỉ số thông dụng nhất cho mục đích này. Tuy nhiên, Thalassemia là bệnh lý gây sự giảm hoặc mất khả năng tổng hợp chuỗi globin, tạo ra các hemoglobin bất thường, ảnh hưởng đến tất cả các chỉ số huyết học chứ không chỉ riêng MCV và MCH. Do đó, tập luật của phần mềm trong nghiên cứu này được xây dựng dựa trên sự kết hợp của nhiều chỉ số khác nhau, bao gồm: HGB, MCV, MCH, ferritin và các tiền sử liên quan đến Thalassemia. Nghiên cứu của G.S. Paterakis và cs (1989) [14] về hệ tri thức chuyên gia đã được tiến hành nhằm phát hiện bệnh Thalassemia thể trait và thiếu máu, thiếu sắt cũng đánh giá trên nhiều chỉ số gồm có MCV, RDW, HGB, RBC, MPV và PLT. Trong nghiên cứu này, chúng tôi đã xây dựng được tập luật cho mô hình cơ sở tri thức chuyên gia hỗ trợ sàng lọc trước sinh Thalassemia bao gồm 34 luật, trong đó có 12 luật sử dụng các chỉ số công thức máu; 2 luật sử dụng kết quả ferritin huyết thanh; 8 luật kết hợp công thức máu và ferritin huyết thanh; 11 luật kết hợp công thức máu, ferritin huyết thanh, điện di huyết sắc tố và 1 luật liên quan đến tiền sử bệnh nhân (bảng 3).

Phần mềm hệ tri thức chuyên gia có khả năng kết luận tình trạng huyết học và tình trạng sắt của đối tượng. Thiếu máu trong Thalassemia là hồng cầu nhỏ, nhược sắc do thiếu hụt một loại chuỗi polypeptid của hemoglobin và dư thừa chuỗi còn lại, dẫn đến giảm tổng hợp hemoglobin, sinh hồng cầu không hiệu lực, tan máu, tăng sinh hồng cầu non. Đối với trường hợp chỉ xét nghiệm tổng phân tích tế bào máu ngoại vi, không làm xét nghiệm đánh giá tình trạng sắt thì phần mềm không đánh giá nguy cơ mang gen mà chỉ kết luận về tình trạng công thức máu. Nghiên cứu trên 233 trường hợp cho thấy, phần lớn có thay đổi công thức máu, phổ biến nhất là “Hồng cầu nhỏ nhược sắc” (36,48%), thấp nhất là “Thiếu máu hồng cầu bình thường” (2,58%), 22,75% trường hợp có công thức máu bình thường (bảng 4). Về tình trạng sắt, không thiếu sắt chiếm 71,67%, thiếu sắt chiếm 2,15%, không đánh giá được do thiếu dữ liệu ferritin huyết thanh là 26,18% (bảng 4). Kết luận của phần mềm được xây dựng dựa trên hướng dẫn, khuyến cáo nên đáng tin cậy, độ chính xác cao.

Kết quả ứng dụng phần mềm tri thức chuyên gia trong sàng lọc bệnh Thalassemia được trình bày trong bảng 5, cho thấy, nhóm nghiên cứu đã xây dựng được một phần mềm hệ tri thức chuyên gia có hiệu quả cao với độ chính xác là 91,30%, độ nhạy 99,22%, độ đặc hiệu 72,73%, giá trị dự báo dương tính và âm tính lần lượt là 89,51% và 97,56% (bảng 5). Giá trị dự báo dương tính của phần mềm cao hơn so với phương pháp hiện tại đang được sử dụng tại Bệnh viện Phụ sản Trung ương (50,5%). Các nghiên cứu trên thế giới về phần mềm trí tuệ nhân tạo hỗ trợ tầm soát Thalassemia cũng đánh giá trên nhiều chỉ số và đưa ra kết quả tốt.

Tuy nhiên, hệ tri thức chuyên gia cũng có những điểm hạn chế. Thứ nhất, trên thực tế lâm sàng có rất nhiều tình huống có thể xảy ra với các chỉ số công thức máu, ferritin và điện di huyết sắc tố mà các luật không thể bao phủ hết được. Câu hỏi đặt ra là đối với các trường hợp không xuất hiện trong tập luật, đây thực sự là những bất thường hiếm khi xảy ra hay do lỗi của máy xét nghiệm hoặc lỗi in ấn? Đặc biệt, với Thalassemia là căn bệnh phức tạp có rất nhiều thể bệnh. Thứ hai, việc xây dựng cũng như điều chỉnh tập luật sẽ tương đối khó khăn do cần có sự phân tích, thảo luận và thống nhất giữa các chuyên gia. Do đó, việc nâng cấp và cải thiện các tập luật trong sàng lọc bệnh Thalassemia đôi khi không thể thực hiện một cách liên tục, nhanh chóng và sự bất đồng quan điểm giữa các chuyên gia là không thể tránh khỏi [15]. Phần mềm hệ tri thức chuyên gia trong nghiên cứu này hiện tại vẫn đang dừng lại ở sàng lọc người mang gen hoặc không mang gen Thalassemia, mang đột biến alpha hay beta Thalassemia và vẫn chưa thể phân ra được các thể bệnh phức tạp hơn. Vì vậy, chúng tôi dự định trong tương lai các công thức có thể được đưa vào các luật để làm tăng khả năng phân biệt các thể bệnh của phần mềm.

Ngoài việc xây dựng phần mềm, nhóm nghiên cứu còn xây dựng một giao diện thân thiện với người dùng, dễ sử dụng, tiện lợi khi có thể sử dụng trên điện thoại thông minh (hình 2). Ứng dụng di động này đã giúp việc ứng dụng phần mềm hệ tri thức chuyên gia của chúng tôi trở nên dễ dàng, hiệu quả hơn.

5. Kết luận

Việc xây dựng ứng dụng di động trên nền tảng trí tuệ nhân tạo hỗ trợ sàng lọc nguy cơ mang gen bệnh Thalassemia với độ chính xác 91,30%, độ nhạy 99,22%, độ đặc hiệu 72,73%, mở ra hướng mới cho việc ứng dụng trí tuệ nhân tạo trong y tế, đặc biệt hỗ trợ tầm soát, duy trì giống nòi và hạn chế các bất thường bẩm sinh, để phục vụ chiến lược chăm sóc sức khỏe cộng đồng của Đảng và Nhà nước.

TÀI LIỆU THAM KHẢO

- [1] B. Modell, M. Darlison (2008), “Global epidemiology of hemoglobin disorders and derived service indicators”, *World Health Organization*, **86(6)**, pp.480-487, DOI: 10.2471/blt.06.036673.
- [2] Thalassemia International Federation (2013), *Annual Report*, <https://thalassaemia.org.cy/downloads/annual-reports/>, accessed 15 May 2023.
- [3] Ministry of Health (2020), *Guide to Technical Expertise in Prenatal and Neonatal Screening, Diagnosis and Treatment*, 23pp (in Vietnamese).
- [4] D.T.H. Thien (2019), *Research on Screening for Thalassemia Disease in Pregnant Women Coming for Examination and Treatment at The National Obstetrics Hospital*, Doctoral Thesis in Medicine, Hanoi Medical University (in Vietnamese).
- [5] Ministry of Health (2022), *Guidelines for The Diagnosis and Treatment of Certain Hematological Disorders*, 528pp (in Vietnamese).
- [6] Ministry of Health (2014), *Guidelines for The Diagnosis and Treatment of Hemophilia and Thalassemia*, 27pp (in Vietnamese).
- [7] J. Old (2016), *Prevention and Diagnosis of Hemoglobinopathies: A Short Guide for Health Professionals and Laboratory Scientists*, TIF Publications, 82pp.
- [8] S. Langlois, J.C. Ford, D. Chitayat, et al. (2008), “Carrier screening for Thalassemia and hemoglobinopathies in Canada”, *Journal of Obstetrics and Gynaecology Canada*, **30(10)**, pp.950-971, DOI: 10.1016/S1701-2163(16)32975-9.
- [9] A. Mosca, R. Paleari, G. Ivaldi, et al. (2009), “The role of hemoglobin A2 testing in the diagnosis of Thalassemia and related hemoglobinopathies”, *Journal of Clinical Pathology*, **62(1)**, pp.13-17, DOI: 10.1136/jcp.2008.056945.
- [10] A. Cao, R. Galanello (2010), “Beta - Thalassemia”, *Genetics in Medicine*, **12(2)**, pp.61-76, DOI: 10.1097/GIM.0b013e3181cd68ed.
- [11] A. Cao, Y.W. Kan (2013), “The prevention of Thalassemia”, *Cold Spring Harbor Perspectives in Medicine*, **3(2)**, DOI: 10.1101/cshperspect.a011775.
- [12] T.H. Luc, T.T.H. Bui, V.V. Hac, et al. (2023), “Hematological characteristics and prevention of the general of Thalassemia of people from 15-20 years old in Cao Bang province”, *Vietnam Medical Journal*, **528(1)**, pp.86-90, DOI: 10.51298/vmj.v528i1.5975 (in Vietnamese).
- [13] T.V.V. Nguyen, T.T.H. Nguyen, Q.M. Nguyen, et al. (2023), “Hematological characteristics and prevention of the general distribution of Thalassemia in children in Ha Quang district, Cao Bang province”, *Vietnam Medical Journal*, **528(2)**, pp.109-113, DOI: 10.51298/vmj.v528i2.6092 (in Vietnamese).
- [14] G.S. Paterakis, G. Terzoglou, E. Vasiliou (1989), “The performance characteristics of an expert system for the “on-line” assessment of Thalassemia trait and iron deficiency-micro hema screen”, *Blood Cells*, **15(3)**, pp.541-560.
- [15] C.M. Ezhilarasu, Z. Skaf, I.K. Jennions (2019), “The application of reasoning to aerospace Integrated Vehicle Health Management (IVHM): Challenges and opportunities”, *Progress in Aerospace Sciences*, **105**, pp.60-73, DOI: 10.1016/j.paerosci.2019.01.001.