

DETECTION OF MALNUTRITION LEVELS IN PEDIATRIC PATIENTS BASED ON AI

Tran Quoc Tuan^{1*}, Tran Quoc Tu², Tran Quynh Huong¹,
Nguyen Thi Thu Hau¹, Bui Nguyen Doan Thu¹, Vo Thi Ngoc Chau²

¹Children's Hospital 2 -14 Ly Tu Trong, Sai Gon Ward, Ho Chi Minh City, Vietnam

²University of Technology, Ho Chi Minh City National University -268 Ly Thuong Kiet, Dien Hong Ward, Ho Chi Minh City

Received: 11/08/2025

Revised: 28/08/2025; Accepted: 25/09/2025

ABSTRACT

Objectives: This study aims to develop a predictive model to support the detection of malnutrition levels in pediatric patients, analyze significant prognostic factors, and evaluate the model's performance on specific disease groups.

Methods: A descriptive cross-sectional study was conducted on 1,563 pediatric inpatients at Children's Hospital 2. A predictive model was developed to predict five levels of malnutrition based on the screening tool currently used at the hospital.

Results: The Logistic Regression model achieved an overall accuracy of 96%. Factors related to underlying diseases (metabolic disorders, malabsorption) and eating status had the greatest impact on the prediction, outperforming simple anthropometric indices. When evaluated on specific disease groups, the model performed well for lower respiratory infections (95% accuracy) but showed reduced performance for acute diarrhea (82% accuracy). The model struggled to classify the "Very Severe" malnutrition group due to a serious data imbalance.

Conclusion: Developing artificial intelligence models to support malnutrition severity classification is an effective and interpretable approach. Underlying health conditions and dietary status emerge as the most significant indicators. To enhance model applicability, further studies should address data imbalance and tailor predictive models to specific patient subgroups.

Keywords: Pediatric malnutrition, logistic regression, predictive model, risk screening, AI.

*Corresponding author

Email: tqquoctuannd2@gmail.com Phone: (+84) 916874795 <https://doi.org/10.52163/yhc.v66iCD16.3344>

PHÁT HIỆN CÁC MỨC SUY DINH DƯỠNG Ở BỆNH NHI DỰA VÀO TRÍ TUỆ NHÂN TẠO

Trần Quốc Tuấn^{1*}, Trần Quốc Tú², Trần Quỳnh Hương¹,
Nguyễn Thị Thu Hậu¹, Bùi Nguyễn Đoan Thư¹, Võ Thị Ngọc Châu²

¹Bệnh viện Nhi Đồng 2 -14 Lý Tự Trọng, P. Sài Gòn, Tp. Hồ Chí Minh, Việt Nam

²Trường Đại Học Bách Khoa, Đại Học Quốc Gia TP. HCM -268 Lý Thường Kiệt, P. Diên Hồng, TP.HCM

Ngày nhận bài: 11/08/2025

Chỉnh sửa ngày: 28/08/2025; Ngày duyệt đăng: 25/09/2025

ABSTRACT

Mục tiêu: Xây dựng mô hình hỗ trợ phát hiện sớm mức độ suy dinh dưỡng ở bệnh nhi. Các mục tiêu cụ thể bao gồm: (1) Đánh giá độ chính xác của mô hình; (2) Phân tích các yếu tố quan trọng nhất trong thang đo; và (3) Đánh giá hiệu quả của mô hình trên một vài nhóm bệnh cụ thể.

Đối tượng và phương pháp nghiên cứu: Nghiên cứu cắt ngang mô tả được thực hiện trên 1.563 bệnh nhi nội trú tại Bệnh viện Nhi Đồng 2. Mô hình trí tuệ nhân tạo được xây dựng để dự đoán 5 mức độ suy dinh dưỡng dựa trên thang đo đang áp dụng tại bệnh viện.

Kết quả: Mô hình Hồi quy Logistic đạt độ chính xác tổng thể 96%. Các yếu tố về bệnh nền (rối loạn chuyển hóa, kém hấp thu) và tình trạng ăn uống có ảnh hưởng lớn nhất đến kết quả dự đoán, vượt trội so với các chỉ số nhân trắc đơn thuần. Khi đánh giá trên các nhóm bệnh lý cụ thể, mô hình hoạt động tốt trên nhóm bệnh viêm hô hấp dưới (độ chính xác 95%) nhưng giảm hiệu quả trên nhóm tiêu chảy cấp (82%). Mô hình gặp khó khăn trong việc phân loại nhóm suy dinh dưỡng “Rất nặng” do dữ liệu bị mất cân bằng nghiêm trọng.

Kết luận: Việc xây dựng mô hình hỗ trợ phân loại mức độ suy dinh dưỡng là một phương pháp hiệu quả, đồng thời đảm bảo khả năng diễn giải cao. Các yếu tố bệnh lý nền và tình trạng ăn uống là các dấu hiệu quan trọng nhất. Các nghiên cứu xử lý mất cân bằng dữ liệu và cá nhân hóa mô hình cho từng nhóm bệnh lý cần được áp dụng để tăng hiệu quả ứng dụng.

Từ khóa: Suy dinh dưỡng trẻ em, hồi quy logistic, mô hình dự đoán, sàng lọc nguy cơ, trí tuệ nhân tạo.

1. ĐẶT VẤN ĐỀ

Suy dinh dưỡng là một vấn đề y tế công cộng nghiêm trọng, đặc biệt phổ biến ở trẻ em, với tỷ lệ cao được ghi nhận tại các cơ sở khám chữa bệnh hiện nay[7]. Suy dinh dưỡng không chỉ ảnh hưởng đến quá trình điều trị và hồi phục mà còn làm giảm chất lượng sống và tăng nguy cơ tử vong. Việc đánh giá nguy cơ suy dinh dưỡng ở trẻ nhập viện là một quy trình phức tạp, đòi hỏi đánh giá toàn diện trên cơ sở tình trạng lâm sàng, phản ứng chuyển hóa và tiến trình phát triển của trẻ. Các công cụ sàng lọc hiện có như SGNA, STRONGkids hoặc “Phiếu sàng lọc và đánh giá dinh dưỡng bệnh nhi nội trú” do Hội Nhi khoa Việt Nam khuyến nghị hiện được xem là hiệu quả, tuy nhiên việc sử dụng đòi hỏi nhiều thời gian và chuyên môn từ nhân viên y tế. Trong thực tế, quá trình đánh giá

dinh dưỡng thường bị bỏ sót hoặc thực hiện không đầy đủ, đặc biệt trong bối cảnh quá tải và thời gian tiếp xúc bệnh nhân hạn chế, nhất là tại các phòng khám ngoại trú. Do đó, phát hiện sớm và can thiệp kịp thời là yếu tố then chốt, đồng thời giúp tiết kiệm thời gian.

Trong bối cảnh như vậy, trí tuệ nhân tạo (AI) có thể đóng vai trò hỗ trợ quan trọng[5,8]. Cụ thể, AI đã được ứng dụng hiệu quả trong việc phát hiện, phân loại và quản lý suy dinh dưỡng. Từ góc độ thực tiễn, một mô hình hỗ trợ chẩn đoán với độ chính xác đủ cao có thể rút ngắn đáng kể thời gian chẩn đoán và hỗ trợ bác sĩ trong điều kiện làm việc áp lực cao. Một số nghiên cứu ứng dụng mô hình dự đoán như hồi quy logistic (logistic regression), rừng ngẫu nhiên đã

*Tác giả liên hệ

Email: tqtuantuannd2@gmail.com Điện thoại: (+84) 916874795 <https://doi.org/10.52163/yhc.v66iCD16.3344>

được sử dụng rộng rãi để đánh giá mức độ suy dinh dưỡng[4]. Gần đây, một số nghiên cứu sử dụng hình ảnh khuôn mặt trẻ em kết hợp với mô hình ResNet-50 để phát hiện suy dinh dưỡng[3]. Những kết quả như vậy cho thấy tiềm năng ứng dụng rộng rãi của AI trong sàng lọc suy dinh dưỡng.

Tuy nhiên, tại Việt Nam, việc ứng dụng AI trong y tế còn khá hạn chế[8] mặc dù đã có nhiều nghiên cứu đề xuất. Một nghiên cứu tại Bệnh viện Nhi Trung ương[2] cho thấy tỷ lệ suy dinh dưỡng thể thấp còi và nhẹ cân lần lượt là 25,6% và 22,7%. Tại tỉnh Nghệ An[1], các tỷ lệ này còn cao hơn, đạt 27,5% và 23,5%, với các yếu tố liên quan bao gồm tiền sử bệnh, tình trạng tiêm chủng và phương pháp cho ăn. Tuy nhiên, cả hai nghiên cứu đều chưa khai thác ứng dụng các mô hình AI.

Về mặt khoa học, nghiên cứu đóng góp vào lĩnh vực AI y tế tại Việt Nam, cung cấp những hiểu biết thực tiễn về tính ứng dụng của AI, đặc biệt là các mô hình học máy, trong chẩn đoán dinh dưỡng, đồng thời mở ra các hướng tích hợp mới trong thực hành lâm sàng.

2. ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Thiết kế nghiên cứu

Nghiên cứu cắt ngang mô tả.

2.2. Đối tượng nghiên cứu

Đối tượng nghiên cứu là các bệnh nhi được điều trị tại các khoa lâm sàng nội trú trong bệnh viện, bao gồm các khoa có liên quan đến các bệnh lý ảnh hưởng đến tình trạng suy dinh dưỡng như nội tổng hợp, hô hấp, tiêu hóa, truyền nhiễm, ngoại tổng hợp cùng nhiều khoa khác.

2.3. Thời gian nghiên cứu:

Bắt đầu từ đầu tháng 2 đến tháng 5 năm 2025.

2.4. Cỡ mẫu nghiên cứu:

Cỡ mẫu được tính toán theo công thức:

$$n = Z_{1-\alpha/2}^2 \frac{p(1-p)}{d^2}$$

+ n: cỡ mẫu cần thiết,

+ d: sai số cận biên (giả định là 0,05).

+ $Z_{1-\alpha/2}$: Z giá trị tương ứng với mức ý nghĩa α

+ P: giả sử tỷ lệ ước tính suy dinh dưỡng ở trẻ em là 0,162.

Tính theo tỷ lệ mất mẫu dự kiến khoảng 10%, quy mô mẫu cần thiết là khoảng 410.

2.5. Phương pháp chọn mẫu

Một kỹ thuật lấy mẫu liên tục đã được áp dụng, theo đó tất cả các bệnh nhi nội trú nhập viện tại Bệnh viện Nhi Đồng 2 được tuyển chọn liên tiếp cho đến khi đạt đủ cỡ mẫu cần thiết.

2.6. Tiêu chuẩn chọn

Bệnh nhi nhập viện và điều trị nội trú tại bệnh viện.

2.7. Tiêu chuẩn loại trừ

Bệnh nhân có tình trạng bất động do chấn thương, gãy xương.

Bệnh nhân đã tham gia nghiên cứu trước đó và được nhập viện trở lại (nhập viện lại không được đánh giá lại).

2.8. Các thông số đánh giá và đơn vị tính

- Thông số nền:

+ Tuổi

+ Giới tính

+ Chiều cao

+ Cân nặng

+ BMI

- Thông số dinh dưỡng được chuẩn hóa theo hướng dẫn của WHO:

+ BMI theo tuổi (BMI-for-age Z-score)

+ Chiều cao theo tuổi (Height-for-age Z-score)

+ Cân nặng theo tuổi (Weight-for-age Z-score)

+ Cân nặng theo chiều cao (Weight-for-height Z-score)

- Thông số đánh giá suy dinh dưỡng theo hướng dẫn của Hội Nhi khoa Việt Nam, hiện đang được áp dụng tại Bệnh viện Nhi Đồng 2:

+ Cân nặng/ Chiều cao hoặc BMI < -1SD (hoặc <90%) hoặc teo cơ, mất cơ, mất lớp mỡ dưới da trên lâm sàng (quan sát)

+ Sụt cân hoặc không lên cân trong 1 tháng qua

+ Lượng ăn giảm <50% nhu cầu bình thường trong tuần qua

+ Có bệnh nền gây kém hấp thu hoặc hạn chế lượng ăn qua đường tiêu hóa (tiêu chảy, nhiễm trùng nặng, co giật, suy tim, suy hô hấp, phẫu thuật tiêu hóa, bệnh lý vùng hầu họng, u tiêu hóa...)

+ Có bệnh gây rối loạn chuyển hóa dinh dưỡng (gan, thận, nội tiết, rối loạn chuyển hóa...)

Các thông số được liệt kê là những chỉ số quan trọng và đã được công nhận rộng rãi, vì một số trong đó

đã được sử dụng trong các nghiên cứu trước đây⁶ mà không áp dụng mô hình dựa trên AI. Ngược lại, nghiên cứu của chúng tôi hướng đến việc xây dựng một mô hình phát hiện suy dinh dưỡng dựa trên AI, trong đó các thông số đánh giá trên đóng vai trò là các đặc trưng đầu vào.

Ngoài ra, mỗi thông số đánh giá suy dinh dưỡng sẽ được chấm 1 điểm. Tổng điểm sau đó sẽ được tổng hợp theo thang điểm nguy cơ từ 0 đến 5, kèm theo các hướng dẫn can thiệp tương ứng, và được phân loại như sau:

+ 0 điểm: Không có nguy cơ suy dinh dưỡng cấp. Theo dõi lại sau 1 tháng.

+ 1-2 điểm: Nguy cơ nhẹ. Hướng dẫn ăn bồi dưỡng, đánh giá lại sau 2 tuần.

+ 3 điểm: Nguy cơ trung bình. Chuyển khám dinh dưỡng, theo dõi lại sau 1 - 2 tuần.

+ 4 điểm: Nguy cơ nặng. Hội chẩn dinh dưỡng và can thiệp chế độ ăn, đánh giá lại mỗi 3 - 5 ngày.

+ 5 điểm: Nguy cơ rất nặng. Hội chẩn dinh dưỡng tích cực, áp dụng chế độ ăn điều trị và theo dõi sát mỗi 3 - 5 ngày.

2.9. Quy trình thu thập dữ liệu

Quy trình thu thập dữ liệu được thực hiện theo các bước sau:

- Bước 1: Thu thập thông tin nền và bệnh sử: Thông tin nền của bệnh nhi, bao gồm tuổi, giới tính, chẩn đoán, tiền sử bệnh lý và các chỉ số lâm sàng, được thu thập từ hồ sơ bệnh án trong thời gian bệnh nhi nhập viện.

- Bước 2: Đánh giá tình trạng dinh dưỡng ban đầu: Tình trạng dinh dưỡng của từng bệnh nhi được đánh giá ngay sau khi nhập viện. Việc đánh giá ban đầu được thực hiện bởi bác sĩ điều trị thông qua thăm khám lâm sàng và đối chiếu với các tiêu chí trong phiếu.

- Bước 3: Kiểm tra và xác nhận dữ liệu: Sau khi hoàn tất, phiếu đánh giá được chuyển đến bác sĩ tham gia nghiên cứu và chuyên viên dinh dưỡng để rà soát. Các bên liên quan trao đổi để thống nhất đánh giá trong trường hợp có chênh lệch hoặc chưa rõ ràng.

- Bước 4: Theo dõi định kỳ trong thời gian điều trị: Tình trạng dinh dưỡng được theo dõi định kỳ sau mỗi 3 hoặc 7 ngày tùy theo diễn tiến bệnh, bằng cách cân đo lại thể trạng và so sánh với đánh giá trước đó.

- Bước 5: Rà soát và phê duyệt cuối cùng: Trước khi đưa vào phân tích, toàn bộ phiếu đánh giá được bác sĩ trưởng khoa Dinh dưỡng kiểm tra và duyệt lại lần

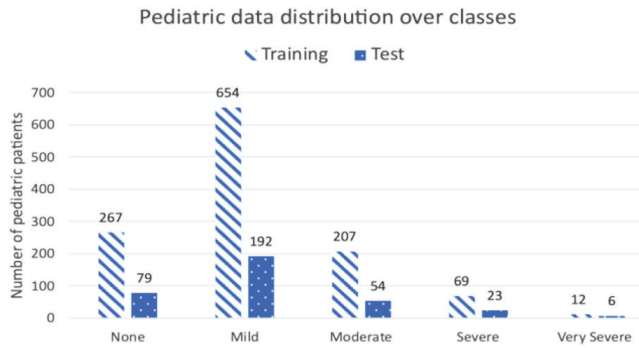
cuối để đảm bảo tính chính xác, đầy đủ và nhất quán của dữ liệu nghiên cứu.

2.10. Phương pháp phân tích dữ liệu

Khác với các phương pháp thống kê, nghiên cứu đã áp dụng phương pháp phân tích hiện đại dựa trên học máy, lập trình Python và môi trường Google Colab. Đây được coi là lựa chọn phù hợp với khoa học dữ liệu và trí tuệ nhân tạo. Toàn bộ quy trình phân tích được triển khai bằng Python, được hỗ trợ bởi các thư viện chuyên dụng như Pandas và NumPy để xử lý dữ liệu, scikit-learn cho các mô hình học máy và Matplotlib/Seaborn để trực quan hóa dữ liệu. Tất cả các quy trình đều được thực hiện bởi một kỹ sư công nghệ thông tin có kinh nghiệm trong nghiên cứu y sinh, do đó đảm bảo tính chính xác, minh bạch và khả năng tái tạo của kết quả.

3. KẾT QUẢ NGHIÊN CỨU

Tập dữ liệu thu được bao gồm 1.563 mẫu, được phân chia thành bốn nhóm tuổi chính: 0 - 13 tuần, 14 tuần - 2 tuổi, 2 - 5 tuổi và 5 - 16 tuổi. Cân nặng trung bình là 17,5 kg và chiều cao trung bình là 96,0 cm, với độ lệch chuẩn lớn, phản ánh sự đa dạng về độ tuổi và thể trạng trong các mẫu nghiên cứu. Chỉ số BMI trung bình là 16,6, cho thấy xu hướng nhẹ cân phổ biến ở nhóm bệnh nhi nội trú. Tất cả các giá trị Z-score trung bình (BMI theo tuổi, cân nặng theo chiều cao, chiều cao theo tuổi, cân nặng theo tuổi) đều có giá trị âm, gợi ý rằng tình trạng dinh dưỡng của nhóm bệnh nhi nội trú thấp hơn so với các chuẩn tham chiếu của Tổ chức Y tế Thế giới (WHO). Sự biến động đáng kể trong các giá trị Z-score làm nổi bật sự đa dạng về tình trạng dinh dưỡng trong mẫu nghiên cứu. Về giới tính, tập dữ liệu bao gồm 964 trẻ nam (61,6%) và 599 trẻ nữ (38,4%). Xét về các bệnh lý nền có liên quan đến hệ thống chấm điểm nguy cơ, 33,85% trẻ có bệnh lý đi kèm liên quan đến kém hấp thu và 22,84% có các bệnh lý liên quan đến rối loạn chuyển hóa. Đây là những yếu tố quan trọng ảnh hưởng đến nguy cơ suy dinh dưỡng. Về hành vi ăn uống và thể trạng, phần lớn đối tượng (63,4%) có biểu hiện giảm ăn trong tuần vừa qua. Tuy nhiên, chỉ một tỷ lệ nhỏ (9,66%) có tình trạng sụt cân trong tháng gần nhất. Phân bố mức độ suy dinh dưỡng cho thấy sự mất cân đối rõ rệt giữa các nhóm: Không suy dinh dưỡng (22,1%), Nhẹ (54,1%), Vừa (16,7%), Nặng (5,9%) và Rất nặng (1,2%). Sự mất cân đối như vậy có thể ảnh hưởng đến hiệu quả xây dựng mô hình đối với các nhóm có mức độ suy dinh dưỡng nặng hơn.



Hình 1. Phân bố các mức độ suy dinh dưỡng trong tập huấn luyện và kiểm tra

Bảng 1. Phân bố dữ liệu trong tập huấn luyện và kiểm tra

Mức độ suy dinh dưỡng	Tập huấn luyện		Tập kiểm tra	
	#	%	#	%
Không (None)	267	22,08	79	22,32
Nhẹ (Mild)	654	54,09	192	54,24
Vừa (Moderate)	207	17,12	54	15,25
Nặng (Severe)	69	5,71	23	6,50
Rất nặng (Very Severe)	12	0,99	6	1,69
Tổng cộng	1,209	100	354	100

Ngoài ra, tập dữ liệu thu được đã được chia thành tập huấn luyện và tập kiểm tra, với phân bố được thể hiện trong Hình 1 và mô tả chi tiết trong Bảng 1. Quá trình chia dữ liệu được thực hiện sao cho mô hình được xây dựng dựa trên dữ liệu trong quá khứ và được kiểm tra trên dữ liệu thu thập ở giai đoạn sau. Hiện tượng mất cân đối dữ liệu cũng xảy ra như đã đề cập trước đó.

Bên cạnh đó, hồi quy logistic được sử dụng để xây dựng một mô hình dự đoán trên tập huấn luyện, với đầy đủ các đặc trưng đầu vào từ các chỉ số đánh giá của chúng tôi. Mô hình đạt được độ chính xác tổng thể rất cao, lên đến 96%. Mặc dù là một mô hình tuyến tính, hồi quy logistic vẫn có khả năng học và khai thác hiệu quả các đặc trưng đầu vào để thực hiện phân loại với độ chính xác cao trong bối cảnh dữ liệu lâm sàng có mức độ phức tạp vừa phải. Mô hình cho thấy khả năng phân biệt rõ ràng giữa các mức độ suy dinh dưỡng khác nhau, đặc biệt là trong các nhóm có kích thước mẫu lớn và phổ biến hơn.

Bảng 2. Hiệu suất của mô hình

Nhãn	Precision	Recall	F1
Không	0.988	1.00	0.994
Nhẹ	1.00	0.995	1.000
Vừa	0.91	0.96	0.94
Nặng	0.71	0.74	0.72
Rất nặng	0.00	0.00	0.00

Trong Bảng 2, nhóm “Không” suy dinh dưỡng đạt độ chính xác và khả năng nhận diện gần như hoàn hảo. Không có trường hợp nào bị phân loại sai trong nhóm này, tức là mô hình đã phân biệt rất rõ giữa trẻ không suy dinh dưỡng với các nhóm còn lại. Tương tự, nhóm “Nhẹ”, chiếm tỷ lệ lớn nhất trong tập dữ liệu, cũng được mô hình phân loại hiệu quả, thể hiện sự ổn định cao trong việc nhận diện suy dinh dưỡng nhẹ. Đối với nhóm suy dinh dưỡng “Vừa”, mô hình cũng cho thấy hiệu suất đáng tin cậy, với khả năng phân loại chính xác mức độ suy dinh dưỡng trung bình. Tuy nhiên, hiệu suất bắt đầu suy giảm ở nhóm “Nặng”, khi điểm F1 giảm xuống còn 0,72. Mặc dù đây vẫn là kết quả khá tốt, nhưng mô hình đã gặp nhiều khó khăn hơn trong việc phân biệt nhóm nặng, có thể do đặc trưng phân biệt không đủ mạnh hoặc có sự chồng lấn giữa nhóm “Nặng” và các mức nhẹ hơn. Đáng chú ý, nhóm “Rất nặng” cho thấy hiệu suất cực kỳ thấp, khi mô hình không phân loại đúng bất kỳ trường hợp nào thuộc nhóm rất nặng. Nguyên nhân chủ yếu là do sự mất cân đối dữ liệu trong tập huấn luyện, với chỉ 12 mẫu thuộc nhóm “Rất nặng”.

Một lý do khác để lựa chọn Hồi quy Logistic là hiệu suất ổn định của mô hình, được thể hiện trong Bảng 3. Hồi quy Logistic vượt trội hơn một số mô hình như AdaBoost và K-Nearest Neighbors, đồng thời cho kết quả tương đương với các mô hình mạnh hơn như Random Forest và một số mô hình học sâu, bao gồm Mạng nơ-ron hồi tiếp (RNN), Bộ nhớ ngắn dài hạn (LSTM) và Mạng nơ-ron tích chập (CNN). So với các mô hình trên, vốn yêu cầu tinh chỉnh nhiều siêu tham số, Hồi quy Logistic có cấu hình đơn giản hơn, từ đó cho thấy tính ứng dụng cao hơn.

Bảng 3. So sánh các mô hình

Model	Accuracy	Precision	Recall	F1
AdaBoost	0.69	0.55	0.69	0.58
K-Nearest Neighbors	0.91	0.92	0.91	0.91
Naïve Bayes	0.88	0.89	0.88	0.88
RNN	0.95	0.95	0.95	0.95

Model	Accuracy	Precision	Recall	F1
LSTM	0.94	0.93	0.94	0.93
CNN	0.95	0.94	0.95	0.94
Random Forest	0.96	0.96	0.96	0.95
Logistic Regression	0.96	0.95	0.96	0.95

Trên hết, để làm rõ mức độ đóng góp của từng đặc trưng đầu vào trong việc dự đoán mức độ suy dinh dưỡng, chúng tôi đã thực hiện phân tích tầm quan trọng (feature importance) của các đặc trưng dựa trên mô hình đã xây dựng. Kết quả cho thấy các yếu tố liên quan trực tiếp đến “tình trạng bệnh lý nền và hành vi ăn uống gần đây” có ảnh hưởng lớn nhất, vượt qua các chỉ số nhân trắc học truyền thống như chiều cao và cân nặng. Như kết quả trong Bảng 4, “Bệnh nền liên quan đến rối loạn chuyển hoá” là yếu tố có ảnh hưởng lớn nhất, tiếp theo là “Bệnh nền liên quan đến tình trạng kém hấp thu” và “Lượng ăn giảm trong tuần vừa qua”. Những kết quả trên phản ánh mối tương quan chặt chẽ giữa các tình trạng bệnh lý nền, đặc biệt là các rối loạn liên quan đến chuyển hóa và hấp thu dưỡng chất, với nguy cơ và mức độ nghiêm trọng của suy dinh dưỡng ở bệnh nhân. Đáng chú ý, yếu tố “Lượng ăn giảm trong tuần vừa qua”, một tiêu chí quan trọng trong nhiều công cụ đánh giá lâm sàng, cũng cho thấy mức độ ảnh hưởng cao, củng cố thêm giá trị lâm sàng của chính yếu tố trong sàng lọc nguy cơ suy dinh dưỡng. Trong số các chỉ số nhân trắc học, chỉ số “Cân nặng/Chiều cao hoặc BMI < -1SD” xếp thứ tư về tầm quan trọng, cao hơn đáng kể so với các chỉ số thường được tính như BMI hoặc các chỉ số Z-score tiêu chuẩn của WHO như Chiều cao theo tuổi và Cân nặng theo tuổi, trong khoảng [0,10 - 0,27]. Sự chênh lệch trên cho thấy rằng đánh giá dinh dưỡng lâm sàng bởi bác sĩ đóng vai trò then chốt trong việc phát hiện suy dinh dưỡng.

Bảng 4. Các đặc trưng có ảnh hưởng lớn nhất

Tên đặc trưng - Tiêu chí đánh giá	Mức độ ảnh hưởng
Bệnh nền liên quan đến rối loạn chuyển hoá	3.07
Bệnh nền liên quan đến tình trạng kém hấp thu	3.00
Lượng ăn giảm trong tuần vừa qua	2.98
Cân nặng/Chiều cao hoặc BMI < -1SD	2.61
Sụt cân trong 1 tháng	2.16
BMI	0.33
Chiều cao theo tuổi (SD)	0.27

Hơn nữa, một số đặc trưng như “Giới tính”, “Chiều cao” và “Cân nặng” có hệ số tầm quan trọng tương đối thấp trong khoảng [0.04, 0.14], cho thấy rằng mặc dù chúng thường được sử dụng trong đánh giá lâm sàng, nhưng mức độ đóng góp vào việc phát hiện suy dinh dưỡng là hạn chế khi được phân tích cùng với các đặc trưng lâm sàng và dinh dưỡng cụ thể hơn. Do đó, phân tích tầm quan trọng của đặc trưng đã cho thấy giá trị của việc kết hợp các yếu tố lâm sàng và triệu chứng bên cạnh các chỉ số nhân trắc học để đánh giá nguy cơ suy dinh dưỡng một cách toàn diện.

Nhằm làm rõ khả năng áp dụng mô hình trong các bối cảnh lâm sàng cụ thể, nghiên cứu đã phân tích hiệu suất mô hình trên hai nhóm bệnh đại diện: “Viêm hô hấp dưới” và “Tiêu chảy cấp”. Đây là những bệnh thường gặp ở trẻ em, ảnh hưởng trực tiếp đến tình trạng dinh dưỡng và thường liên quan đến giảm lượng ăn, tăng tiêu hao năng lượng hoặc rối loạn hấp thu. Ở nhóm “Viêm hô hấp dưới”, mô hình đạt độ chính xác tổng thể 95%, cho thấy khả năng phân loại khá tốt trong bối cảnh bệnh lý hô hấp. Về tầm quan trọng của đặc trưng, ba yếu tố ảnh hưởng nhiều nhất là “Giảm ăn trong tuần vừa qua” (1.92), “Bệnh nền liên quan đến kém hấp thu” (1.87), và “Bệnh nền liên quan đến rối loạn chuyển hoá” (1.77). Đây đều là các yếu tố liên quan đến rối loạn tiêu hoá và chuyển hoá, phù hợp với cơ chế bệnh lý của các bệnh hô hấp kéo dài làm ảnh hưởng đến hấp thu và duy trì dinh dưỡng. Ngược lại, ở nhóm tiêu chảy cấp, hiệu suất mô hình giảm đáng kể, với độ chính xác tổng thể chỉ đạt 82%. Về tầm quan trọng của đặc trưng trong nhóm này, các kết quả vẫn tương tự như nhóm nhiễm trùng đường hô hấp dưới: “Bệnh nền liên quan đến tình trạng kém hấp thu” (1.51), “Giảm ăn trong tuần vừa qua” (1.43), và “Bệnh nền liên quan đến rối loạn chuyển hoá” (1.38) là các yếu tố nổi bật nhất. Tuy nhiên, mức tầm quan trọng của chúng thấp hơn so với nhóm viêm hô hấp dưới..

4. BÀN LUẬN

Các kết quả nghiên cứu cho thấy mô hình Hồi quy Logistic đạt độ chính xác cao (96%) khi áp dụng cho dữ liệu bệnh nhi nội trú có đặc điểm nhân trắc học và lâm sàng đa dạng. Ngoài ra, mô hình thể hiện khả năng nắm bắt hiệu quả mối quan hệ giữa các đặc trưng đầu vào và các mức độ suy dinh dưỡng, với điều kiện dữ liệu được chuẩn hóa và tiền xử lý đúng cách. Mô hình cũng cho thấy hiệu suất tốt trong việc phân loại các mức suy dinh dưỡng phổ biến như “Nhẹ” hoặc không suy dinh dưỡng, phản ánh khả năng sàng lọc tốt. Hiệu suất cũng đạt mức chấp nhận được đối với nhóm “Vừa”. Tuy nhiên, hiệu suất giảm rõ rệt ở các nhóm “Nặng” và “Rất Nặng”, chủ yếu do mất cân bằng dữ liệu vì số lượng mẫu “Rất Nặng” hạn chế. Trong phân tích tầm quan trọng của đặc trưng, các yếu tố liên quan đến bệnh nền và hấp

thu như “Bệnh nền liên quan đến rối loạn chuyển hóa”, “Bệnh nền liên quan đến tình trạng kém hấp thu” và “Giảm ăn” có ảnh hưởng lớn nhất đến quyết định phân loại. Đáng chú ý, yếu tố “Lượng ăn giảm trong tuần vừa qua”, một tiêu chí thường được đưa vào các công cụ đánh giá suy dinh dưỡng lâm sàng, cũng cho thấy mức độ ảnh hưởng cao. Ngược lại, các chỉ số nhân trắc học như BMI, cân nặng và chiều cao lại thể hiện mức ảnh hưởng thấp hơn đáng kể. Những kết quả trên đã nhấn mạnh tầm quan trọng của việc kết hợp dữ liệu nhân trắc học với thông tin lâm sàng để có thể đánh giá tình trạng dinh dưỡng một cách toàn diện hơn. Khi được áp dụng cho các nhóm bệnh cụ thể, mô hình hoạt động tốt đối với nhóm “Viêm hô hấp dưới” và giữ nguyên xu hướng tầm quan trọng của các đặc trưng như đã quan sát trong toàn bộ tập dữ liệu. Tuy nhiên, ở nhóm “Tiêu chảy cấp”, hiệu suất của mô hình giảm, cho thấy những hạn chế tiềm tàng khi áp dụng với các tình trạng bệnh lý cụ thể có tính đặc thù cao.

5. KẾT LUẬN

Việc phát triển mô hình dự đoán mức độ suy dinh dưỡng ở bệnh nhi nội trú là hoàn toàn khả thi và đặc biệt hiệu quả đối với các mức suy dinh dưỡng phổ biến. Mô hình không chỉ đạt hiệu suất cao mà còn góp phần làm nổi bật vai trò quan trọng của các yếu tố như bệnh nền và hành vi ăn uống trong việc dự đoán nguy cơ suy dinh dưỡng. Việc phát triển và ứng dụng mô hình như vậy cho các trường hợp bệnh nhi ngoại trú hoặc tại các cơ sở y tế địa phương cũng là hướng đi nên được xem xét.

Để cải thiện hiệu suất đối với các nhóm thiểu số hoặc nhóm bệnh có tính chất đặc thù, chúng tôi sẽ tích hợp các kỹ thuật xử lý mất cân bằng dữ liệu, làm phong phú tập dữ liệu bằng các thông tin lâm sàng hoặc sinh hóa bổ sung, đồng thời thử nghiệm các mô hình học sâu phù hợp hơn với độ phức tạp của dữ liệu y sinh.

LỜI CẢM ƠN

Nghiên cứu được thực hiện với sự hỗ trợ từ Trường Đại học Bách Khoa, Đại học Quốc gia TP. Hồ Chí Minh và Bệnh viện Nhi Đồng 2, TP. Hồ Chí Minh, Việt Nam. Chúng tôi xin gửi lời cảm ơn chân thành đến các đơn vị đã cung cấp dữ liệu và hỗ trợ quý báu trong suốt quá trình nghiên cứu.

TÀI LIỆU THAM KHẢO

- [1] Vũ Thị Quyên, Phạm Thị Thanh Nga, Nguyễn Thị Việt Hà. (2023). Tình trạng dinh dưỡng và một số yếu tố liên quan đến suy dinh dưỡng ở trẻ từ 6 đến 24 tháng tuổi tại Bệnh viện Sản Nhi Nghệ An. *Vietnam Medical Journal*, 527(1B)
- [2] Souliyakane, A., Nguyễn Thị Yến, Nguyễn Thị Thuý Hồng, Chu Thị Phương Mai, Lê Hữu Thành. (2021). Tình trạng dinh dưỡng ở trẻ 6 tháng đến 5 tuổi tại Phòng khám Dinh dưỡng Bệnh viện Nhi Trung ương. *Vietnam Medical Journal*, 507(2), 2021. (in Vietnamese)
- [3] Aanjankumar, S., Sathyamoorthy, M., Dhanaraj, R.K. et al. (2025). Prediction of malnutrition in kids by integrating ResNet-50-based deep learning technique using facial images. *Sci Rep* 15, 7871, 2025.
- [4] Das, S. and Rahman, R.M. (2011). Application of ordinal logistic regression analysis in determining risk factors of child malnutrition in Bangladesh. *Nutrition Journal*, 10 (124).
- [5] Di Martino, F., Delmastro, F., & Dolciotti, C. (2023). Explainable AI for malnutrition risk prediction from m-Health and clinical data. *Smart Health*, 23, 100429.
- [6] Ezzat, M.A, Albassam, E.M., Aldajani, E.A., Alaskar, R.A., & Devol, E.B. (2022). Implementation of new indicators of pediatric malnutrition and comparison to previous indicators. *International Journal of Pediatric and Adolescent Medicine* 9, 216-224.
- [7] Govender, I., Rangiah, S., Kaswa, R., & Nzaumvila, D. (2021). Malnutrition in children under the age of 5 years in a primary health care setting. *South African Family Practice*, 63(1), 5416.
- [8] Janssen, S. M., Bouzembrak, Y., & Tekinerdogan, B. (2024). Artificial Intelligence in Malnutrition: a systematic literature review. *Advances in Nutrition*, 100264.