Chuyên san

CÁC CÔNG TRÌNH NGHIÊN CỨU, PHÁT TRIỂN VÀ ỨNG DỤNG CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

ISSN: 1859-3526

BỘ KHOA HỌC VÀ CÔNG NGHỆ - MINISTRY OF SCIENCE AND TECHNOLOGY

Tập 2025
Số 1
tháng 6
Số đặc biệt CITA 2025

2025

Research, Development and Application on Information and Communication Technology

CÔNG NGHỆ 4T

Địa chỉ liên hệ:
115 Trần Duy Hưng, Hà Nội
Tel: (84)-24-37737136, Fax: (84)-24-37737130,
Website https: https://ictmag.ictvietnam.vn/cntt-tt
Email: chuyensanbcvt@mic.gov.vn
Chi nhánh TP. Hồ Chí Minh
27 Nguyễn Bỉnh Khiêm, Quận 1, TP. Hồ Chí Minh
Tel/Fax: (84)-28-38992898

Giấy phép xuất bản số 365/GP-BTTTT, ngày 19/12/2014;
sửa đổi số 233/GP-BTTTT ngày 23/5/2017.
In tại Công ty CP In Hà Nội.
In xong và nộp lưu chiểu tháng 11/2024.

Giá: 60.000 đồng

Các công trình nghiên cứu, phát triển và ứng dụng

Công nghệ Thông tin và Truyền thông

https://www.ictmag.vn/cntt-tt

Tập 2025, Số 1, Tháng 6

Số đặc biệt CITA 2025

MỤC LỤC

Các công trình nghiên cứu, phát triển và ứng dụng

Công nghệ Thông tin và Truyền thông

https://www.ictmag.vn/cntt-tt

Tập 2025, Số 1, Tháng 6

Số đặc biệt CITA 2025

TABLE OF CONTENTS

Phương pháp diễn giải cho mô hình hỏi đáp hình ảnh tiếng Việt

Nguyễn Thanh Tân , Trần Thị Phương Linh, Lê Thanh Tùng, Nguyễn Tiến Huy
Đại học Khoa học Tự nhiên, Đại học Quốc gia Thành phố Hồ Chí Minh

Tác giả liên hệ: Nguyễn Tiến Huy. Email: ntienhuy@fit.hcmus.edu.vn
Ngày nhận bài: 14/10/2024, ngày sửa chữa: 25/11/2024, ngày duyệt đăng: 28/11/2024
DOI: 10.32913/mic-ict-research-vn.v2024.n2.1338

Tóm tắt: Diễn giải trong mô hình hỏi đáp hình ảnh là chủ đề được quan tâm bởi vì tính minh bạch của mô hình sẽ giúp người sử dụng tin tưởng cũng như hiểu về điểm mạnh và hạn chế của mô hình. Các mô hình đa phương thức như hỏi đáp hình ảnh sử dụng cả hai loại dữ liệu hình ảnh và văn bản, do đó gây thách thức cho các kỹ thuật giải thích thường chỉ áp dụng trên một loại dữ liệu. Trong công trình này, chúng tôi đề xuất kết hợp phương pháp giải thích hình ảnh Grad-CAM và phương pháp giải thích văn bản Transformer để phân tích hành vi của mô hình. Các kết quả thu được trên tập dữ liệu hỏi đáp tiếng Việt giúp chúng tôi rút ra các nhận xét như sau: i) Transformer thị giác trích xuất đặc trưng toàn cục trong khi ResNet trích xuất đặc trưng cục bộ; ii) mô hình suy luận không tốt với ảnh có nhiều yếu tố gây nhiễu; iii) phương thức văn bản có ít tác động hơn phương thức hình ảnh; iv) mô-đun PhoBERT có sự thiên vị một số câu hỏi nhất định.

Từ khóa: *trí tuệ nhân tạo khả diễn, hỏi đáp hình ảnh*

Title: **An XAI Method for Vietnamese Visual Question Answering Models**

Abstract: Interpretability in Visual Question Answering (VQA) models is a topic of interest because the transparency of the model can help users trust and understand its strengths and limitations. Multimodal models like VQA utilise both image and text data, posing challenges for interpretability techniques that are typically designed for a single type of data. In this work, we propose a combination of the Grad-CAM image explanation method and the Transformer-based text explanation method to analyse the behaviour of the model. The results obtained from the Vietnamese VQA dataset lead us to the following observations: i) the visual Transformer extracts global features, while ResNet extracts local features; ii) the model struggles with images containing many distracting elements; iii) the text modality has less impact compared to the image modality; iv) the PhoBERT module shows bias toward certain types of questions.

Keywords: *XAI, visual question answering*

I. GIỚI THIỆU

Diễn giải mô hình trí tuệ nhân tạo (AI) là một quá trình giúp con người hiểu rõ hơn về cách mà mô hình hoạt động, tại sao nó đưa ra các quyết định hoặc dự đoán nhất định. Quá trình này không chỉ giúp người sử dụng tin tưởng vào quyết định của ứng dụng trí tuệ nhân tạo mà còn giúp tìm ra điểm mạnh, điểm yếu, cũng như các điểm lỗi của mô hình [1].

Các nghiên cứu về diễn giải mô hình trí tuệ nhân tạo ngày nay cũng đã đạt được nhiều bước tiến quan trọng. Nổi bật có thể kể đến như nhóm Sevalraju và các cộng sự [2] đề xuất sử dụng Gradient của một lớp theo một tầng tích chập để tính ra một biểu đồ thể hiện sự quan trọng của các điểm ảnh hoặc đặc trưng được trích xuất của tầng tích chập đang xét. Nhóm Chefer và các cộng sự [3] sử dụng các luật kết hợp giữa quá trình lan truyền tới (forward) và lan truyền ngược (backward) để tạo ra biểu đồ liên quan giữa các đặc trưng ảnh hoặc văn bản để diễn giải cho mô hình Transformer.

Tuy nhiên, hầu hết các nghiên cứu về trí tuệ nhân tạo khả diễn thường tập trung vào các mô hình đơn phương thức. Trong khi đó, với sự phát triển của các hệ thống trí tuệ nhân tạo gần đây, học máy đa phương thức đang trở thành đề tài nghiên cứu quan trọng, nhận được sự quan tâm của rất nhiều nhà khoa học. Mô hình đa phương thức có thể xử lý đồng thời nhiều nguồn dữ liệu khác nhau như âm thanh, hình ảnh, văn bản... Việc kết hợp các thông tin đa dạng này là vô cùng cần thiết và đóng vai trò quan trọng trong sự bùng nổ không ngừng của các dữ liệu đa phương tiện ngày nay. Trong số đó, các nghiên cứu đa phương thức có

sự kết hợp giữa ảnh và văn bản đang được tập trung ngày càng nhiều nhằm giải quyết các bài toán đầy thử thách như chú thích ảnh, hỏi đáp trên ảnh, sinh ảnh từ văn bản... Nổi bật trong số đó có thể kể đến các bài toán thực tế như hỏi đáp hình ảnh giúp ích cho người khiếm thị của Le và các cộng sự [4]. Nghiên cứu này đề xuất hệ thống đa phương thức ngôn ngữ - thị giác giải quyết vấn đề của tập dữ liệu chất lượng thấp chẳng hạn như thiếu vắng các vật thể. Cũng trong chủ đề này, Nguyen và các cộng sự [5] đã đề xuất các mô hình đa phương thức dựa trên cơ chế attention và các biến thể như attention có hướng dẫn để tăng độ hiệu quả. Với sự phát triển không ngừng này, việc diễn giải các mô hình đa phương thức, cụ thể là với bài toán hỏi đáp hình ảnh, là nhu cầu thiết yếu và vô cùng cần thiết.

Mặc dù đã có nhiều nghiên cứu về diễn giải mô hình hỏi đáp hình ảnh, nhưng hầu hết đều tập trung vào các mô hình được huấn luyện trên dữ liệu tiếng Anh. Chưa có nhiều các công trình tập trung giải thích hành vi của mô hình đa phương thức trong tiếng Việt. Do đó, nghiên cứu của chúng tôi tập trung vào việc khảo sát và chọn lựa phương pháp giải thích phổ biến và hiệu quả để áp dụng cho mô hình hỏi đáp hình ảnh tiếng Việt. Cụ thể, công trình đề xuất sử dụng kết hợp giữa Grad-CAM [2] và diễn giải Transformer [3] để làm rõ hơn quá trình ra quyết định của mô hình hỏi đáp. Dựa trên các giải thích, chúng tôi tiến hành phân tích định tính để tìm hiểu những khuôn mẫu, hành vi cũng như điểm mạnh, điểm thiếu sót của mô hình hỏi đáp hình ảnh. Từ đó, công trình rút ra bốn kết luận chính:

1) Khối Transformer thị giác có xu hướng trích xuất đặc trưng toàn cục, trong khi khối ResNet chủ yếu tập trung vào việc trích xuất đặc trưng cục bộ.
2) Mô hình không thực hiện suy luận tốt khi xử lý các hình ảnh chứa nhiều vật thể.
3) Thông tin văn bản ít ảnh hưởng đến kết quả suy luận hơn so với thông tin thị giác.
4) Khối PhoBERT thể hiện sự thiên vị nhất định trong quá trình suy luận.

Tiếp theo, trong phần II, chúng tôi sẽ trình bày các kỹ thuật cơ bản trong diễn giải. Sau đó, phương pháp diễn giải cho mô hình hỏi đáp sẽ được đề xuất ở phần III. Hai phần IV và V sẽ trình bày các thực nghiệm cũng như thảo luận, phân tích về các kết quả này.

II. MỘT SỐ NGHIÊN CỨU LIÊN QUAN

Trong phần này, chúng tôi trình bày một số nghiên cứu liên quan về phương pháp diễn giải cho mô hình trí tuệ nhân tạo.

1. Diễn giải bằng phân rã đa phương thức (DIME)

Xét trên bài toán hỏi đáp hình ảnh, DIME [6] phân rã một mô hình đa phương thức thành hai thành tố: đóng góp của từng phương thức và tương tác của các phương thức.

Sau khi phân rã mô hình, DIME tạo ra diễn giải bằng cách đánh giá mô hình, xem xét sự thay đổi của logit trên một lớp dự báo c, bằng cách tạo ra một tập dữ liệu mới như sau: thực hiện n lần việc thay đổi ngẫu nhiên một vùng trên ảnh so với ảnh ban đầu hoặc thực hiện n lần việc thay đổi ngẫu nhiên một chữ trong câu hỏi so với câu hỏi ban đầu. Trong môi trường đa phương thức, một tập dữ liệu mới chỉ nên chứa các thay đổi của một phương thức (chỉ ảnh hoặc chỉ văn bản), phương thức còn lại giữ cố định. Cuối cùng, thực hiện khớp một hàm tuyến tính trên dữ liệu logit của lớp c. Kết quả thu được là trọng số của các vùng trên ảnh dùng để làm biểu đồ nhiệt ảnh và histogram của các chữ trong câu hỏi.

DIME trực quan trên từng phương thức (câu hỏi và hình ảnh) để đưa ra biểu đồ nhiệt thay vì trực quan tại từng mô-đun của mô hình. Giả sử như người lập trình muốn biết độ hiệu quả (khả năng trích chọn đặc trưng) của từng mô-đun trong mô hình để cân nhắc mở rộng, thay thế hoặc sửa lỗi, họ không có một cách trực tiếp để đưa ra quyết định. Một vấn đề khác chúng tôi lưu tâm là quy trình của DIME tốn khá nhiều thời gian, bởi vì phải lặp lại quá trình suy diễn trên tập dữ liệu mới lớn hơn nhiều so với dữ liệu ban đầu, đây có thể là nơi "nghẽn cổ chai" của phương pháp này.

2. Diễn giải bằng văn bản kèm biểu đồ nhiệt

Trong bài toán hỏi đáp hình ảnh, nhóm tác giả Park và các cộng sự [7] đề xuất huấn luyện một mô hình có khả năng trả lời câu hỏi và tạo ra diễn giải bằng văn bản lẫn diễn giải bằng biểu đồ nhiệt trên hình ảnh.

Để thực hiện điều này, tác giả đã đề xuất hai tập dữ liệu mới chứa chân trị của văn bản diễn giải và biểu đồ nhiệt. Về kiến trúc của mô hình, tác giả gộp cả hai phần tạo ra câu trả lời và tạo ra diễn giải thành một kiến trúc thống nhất. Đầu vào là câu hỏi và hình ảnh để tạo ra câu trả lời; sau đó dùng cả câu hỏi, hình ảnh và câu trả lời để tạo ra biểu đồ nhiệt ảnh, từ đó tạo các đặc trưng attention để đưa ra diễn giải văn bản.

Mặc dù phương pháp này tạo ra diễn giải rất thân thiện với người dùng, nhưng việc tạo ra chân trị của văn bản thường do con người thực hiện, và phải trải qua quy trình sàng lọc để thu được dữ liệu chất lượng cao, do đó, rất tốn kém để thực hiện điều này.

3. Phương pháp biểu đồ lớp kích hoạt sử dụng Gradient (Grad-CAM)

a) Phương pháp biểu đồ lớp kích hoạt

Biểu đồ lớp kích hoạt (CAM) [8] là một phương pháp trực quan bằng cách sử dụng biểu đồ nhiệt để chỉ ra vùng ảnh đặc biệt mà mô hình CNN sử dụng để xác định nhãn của một nhãn nhất định trong tập các nhãn đầu ra. Ở tầng CNN cuối cùng, trước khi áp dụng một hàm kích hoạt để xác định đầu ra (ví dụ như softmax cho việc phân lớp), người ta dùng phép gộp trung bình toàn cục (GAP) các biểu đồ đặc trưng tích chập, sau đó cho qua một tầng kết nối dày đặc để đưa ra kết quả. Kết quả của việc áp dụng GAP là tạo ra một vector, tổ hợp tuyến tính phần tử thứ i của vector này với biểu đồ đặc trưng thứ i tương ứng trong tầng CNN cuối cùng sẽ cho ra một biểu đồ duy nhất được gọi là biểu đồ lớp kích hoạt. Nhược điểm của CAM nằm ở biểu đồ đặc trưng. Bởi vì cần thay đổi kiến trúc mô hình, cụ thể là thêm phép trung bình toàn cục vào biểu đồ đặc trưng cuối cùng, nên hiệu quả đã bị giảm sút ở một số tác vụ như phân lớp và không thể ứng dụng cho các tác vụ kết hợp ảnh và văn bản.

b) Phương pháp biểu đồ lớp kích hoạt sử dụng Gradient

Biểu đồ lớp kích hoạt sử dụng Gradient [2] là một giải pháp cải tiến của CAM nhằm loại bỏ những hạn chế đã nêu. Bằng cách sử dụng Gradient, phương pháp này không can thiệp vào kiến trúc của mô hình và hoàn toàn khả dụng cho các tác vụ kết hợp ảnh và văn bản. Ở tầng kết nối dày đặc cuối cùng của mô hình, ta thu được logit, y^c của nhãn c nào đó (đối với bài toán phân lớp), Grad-CAM sử dụng Gradient của y^c theo một hàm kích hoạt A^k trong lớp tích chập, hay nói cách khác là $\frac{\partial y^c}{\partial A^k}$. Áp dụng phép gộp trung bình toàn cục các luồng đạo hàm theo từng điểm ảnh của ảnh để cho ra trọng số quan trọng của điểm ảnh, gọi là α_k^c.

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \tag{1}$$

Đây là phương pháp sẽ được sử dụng để giải thích các khối thị giác trong quy trình diễn giải của nghiên cứu này.

4. Phương pháp diễn giải cho kiến trúc Transformer

a) Sử dụng attention thô

Như đã biết, ở trong self-attention thì đầu ra chính là tổng có trọng số của đầu vào.

$$A = softmax(\frac{Q \cdot K^T}{\sqrt{d_h}}) \tag{2}$$

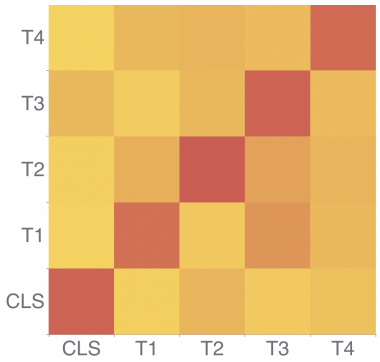
$$O = A \cdot V \tag{3}$$

Cụ thể hơn, mỗi đầu ra y_i ứng với token i sẽ là tổng có trọng số của các vector giá trị v của tất cả các token:

$$y_i = \sum_j \alpha_{i,j} v_j \tag{4}$$

trong đó, trọng số $\alpha_{i,j}$ là phần tử thứ i, j của ma trận attention A.

Thế nên, bằng cách thuần túy nhất, ta có thể sử dụng giá trị $\alpha_{i,j}$ để ước lượng mức độ *chú ý* của token j với token i. Ma trận attention A trong self-attention dùng để đo mức độ đóng góp của tất cả các token đầu vào đối với token đầu ra và mỗi dòng trong ma trận A biểu thị cho sự đóng góp của tất cả các token đối với một token (token đang xét tại dòng đó) (Hình 1).



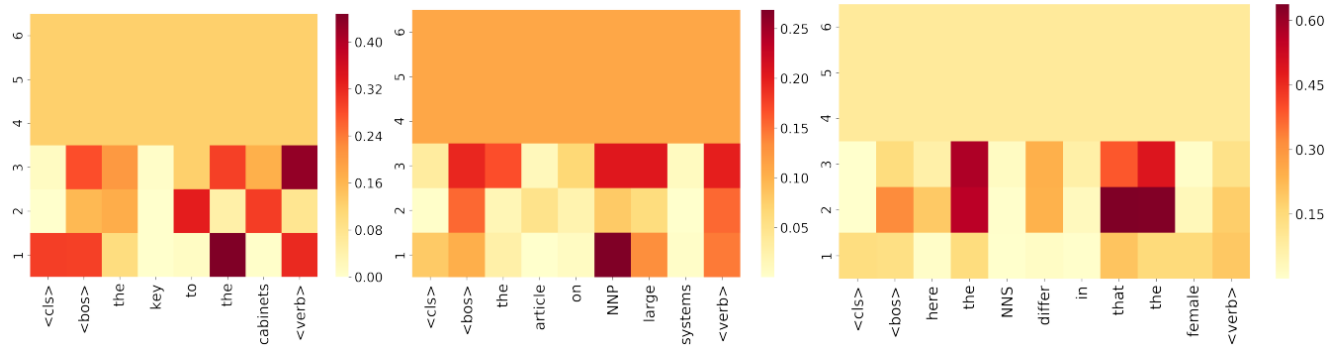
Hình 1. Mô tả ma trận attention của một câu với 4 token.

Vì token <CLS> chứa thông tin phân lớp toàn cục của đặc trưng để đưa ra dự đoán phân lớp cuối cùng, nên nó sẽ được sử dụng để đưa ra diễn giải. Thông thường, ta sẽ chỉ xét dòng tương ứng với token <CLS> có trong ma trận A đó, cũng có nghĩa là ta tập trung vào các đóng góp của tất cả các token có trong đầu vào lên token <CLS>. Tuy nhiên, ta cũng cần phải loại bỏ token đầu tiên (cũng chính là token <CLS>) vì việc lấy thông tin đóng góp của token <CLS> lên chính nó là hoàn toàn vô nghĩa.

Hạn chế: Mặc dù phương pháp này khá đơn giản và nhanh chóng nhưng [9] chỉ ra rằng đối với các mô hình Transformer có quá nhiều lớp attention, càng đi sâu hơn thì sự diễn giải trông khá "đồng đều" (Hình 2). Theo quan sát của tác giả, càng đi sâu hơn thì các embedding càng được *ngữ cảnh hóa*, làm cho các thông tin embedding mang ý nghĩa tương tự nhau.

Vấn đề cần giải quyết: Phương hướng khả thi nhất là tận dụng và phát huy càng nhiều thông tin trong Transformer càng tốt. Ví dụ như Transformer có nhiều block và head, mỗi cái đều chứa ma trận attention, chúng ta có thể "tổng hợp" các thông tin này lại. Tất nhiên cũng sẽ phát sinh các thách thức sau đây:

- ***Nhiều đầu attention:*** Một số head có thể sẽ không mang ý nghĩa quan trọng đối với đầu ra, cho dù có bỏ qua những head này thì mô hình vẫn đưa ra dự đoán



Hình 2. Biểu đồ attention thô cho token <CLS> tại các lớp khác nhau [9].

tốt. Vậy nên ta không thể xét các ma trận trong từng head một cách "bình đẳng" được.

- ***Nhiều lớp attention:*** Thông tin token được truyền giữa các lớp (khối) attention như thế nào và làm sao để tính toán được chúng.

Các phương pháp dưới đây đã được đề xuất để giải quyết những thách thức được nêu trên.

b) Attention Rollout

Attention Rollout [9] sử dụng đệ quy để tính attention của từng lớp trong Transformer từ giá trị embedding đầu vào.

- **Đối với việc giải quyết nhiều đầu attention:** Việc phân tích thông tin của từng đầu đòi hỏi phải xét đến việc các thông tin giữa chúng bị hòa lẫn với nhau. Trong bài báo gốc, để cho đơn giản thì tác giả chỉ tổng hợp các đầu trong một lớp lại thành một đầu duy nhất để tính toán bằng cách lấy trung bình attention các đầu.

$$E_h(A^{(b)}) = \frac{1}{M}\sum_m A_m^{(b)} \tag{5}$$

trong đó, b đại diện cho lớp (block) thứ b.

- **Đối với việc lan truyền thông tin giữa các lớp attention:** Để tính toán được cách thông tin được lan truyền giữa các lớp, đầu tiên, ta không thể không xét đến các kết nối phần dư có trong mô hình. Những kết nối phần dư này đóng vai trò rất quan trọng trong việc liên kết các vị trí tương ứng nhau giữa các lớp. Vậy nên, tác giả đề xuất việc chèn thêm trọng số để biểu diễn cho các kết nối phần dư, cụ thể là cộng thêm một ma trận đơn vị với ma trận attention: $A = I + E_h(A^{(b)})$.

Để tính toán attention từ lớp thứ i (l_i) truyền đến lớp thứ j (l_j), ta nhân các ma trận trọng số attention theo kiểu đệ quy:

$$\tilde{A}(l_i) = \begin{cases} A(l_i)\tilde{A}(l_{i-1}) & \text{nếu } i > j \\ A(l_i) & \text{nếu } i = j \end{cases} \tag{6}$$

Với công thức như trên, để tính toán attention từ đầu vào (lớp đầu tiên), ta có công thức $\tilde{A}(l_i) = A(l_1) \cdot A(l_2) \cdot \ldots \cdot A(l_i)$ (attention từ lớp thứ i đến lớp đầu vào). Sau khi thực hiện Rollout, ta có thể tính được "mức độ đóng góp" của lớp đầu tiên tới lớp đầu ra (với i là lớp cuối cùng). Phương pháp này hiệu quả hơn nhiều so với việc sử dụng attention thô vì attention thô chỉ tính được "mức độ đóng góp" của lớp gần cuối đối với lớp cuối.

c) Phương pháp được đề xuất trong bài báo Transformer Interpretability Beyond Attention Visualization

Bài báo [3] chỉ ra rằng phương pháp Rollout vẫn còn quá đơn giản. Thế nên, tác giả đã mở rộng và cải thiện phương pháp này với các ý tưởng sau đây:

- **Tổng hợp các đầu attention bằng relevance và Gradient:** Bởi vì mỗi đầu sẽ thể hiện các thông tin khác nhau nên "mức độ đóng góp" của từng đầu đối với dự đoán cũng phải khác nhau. Nếu ta chỉ sử dụng cách lấy trung bình như phương pháp Rollout thì sẽ không thể hiện được đặc điểm này. Do đó, ta cần phải xem xét tới việc đặt trọng số cho từng đầu. Trong bài báo này, tác giả sử dụng Gradient để làm trọng số, Gradient cũng cho biết thông tin liên quan đến phân lớp. Ngoài ra, thay vì sử dụng thuần attention, tác giả lại sử dụng relevance dựa vào phương pháp LRP [10] để biểu diễn cho attention của mỗi đầu trong từng lớp.

$$E_h(A^{(b)}) = \frac{1}{M}\sum_m \nabla A_m^{(b)} \odot R_m^{(n_b)} \tag{7}$$

với $\nabla A_m^{(b)} = \frac{\partial y_c}{\partial A_m^{(b)}}$, y_c là giá trị đầu ra của lớp c mà chúng ta cần diễn giải.

Tuy nhiên, vì ta chỉ cần tập trung vào phần nào của đặc trưng đầu vào có ý nghĩa đối với dự đoán, nên ta sẽ chỉ xét tới các đóng góp dương và loại bỏ các đóng góp âm ra khỏi ma trận Relevance:

$$E_h(A^{(b)}) = \frac{1}{M}\sum_m \nabla A_m^{(b)} \odot R_m^{(n_b)^+} \tag{8}$$

-**Lan truyền thông tin giữa các lớp attention bằng nhân ma trận:** Ý tưởng vẫn giữ nguyên giống như phương pháp Rollout:

$$\bar{A}^{(b)} = I + E_h(A^{(b)}) \tag{9}$$

$$\mathbf{C} = \bar{A}^{(1)} \cdot \bar{A}^{(2)} \cdot \ldots \cdot \bar{A}^{(B)} \tag{10}$$

Đa số tập dữ liệu về lĩnh vực hỏi đáp hình ảnh thiếu vắng diễn giải chân trị, do đó, đề xuất của chúng tôi là sử dụng các phương pháp không cần đến diễn giải chân trị, chẳng hạn như biểu đồ nhiệt và histogram. Mục tiêu của chúng tôi thiên về diễn giải mô hình hoạt động hơn là diễn giải dễ hiểu cho người dùng cuối, bởi lẽ biểu đồ nhiệt và histogram cũng không quá thân thiện với người dùng cuối; do đó, chúng tôi sẽ giải thích mô-đun trong mô hình mà không nhấn mạnh vào tác động của từng phương thức. Đặc biệt hơn, chúng tôi sử dụng mô hình được chứng minh có độ hiệu quả tốt cho tác vụ hỏi đáp hình ảnh tiếng Việt để phân tích nhằm đưa ra hướng cải thiện cho mô hình.

III. PHƯƠNG PHÁP ĐỀ XUẤT

Trước tiên, chúng tôi giới thiệu mô hình được sử dụng trong thực nghiệm, bởi vì quy trình diễn giải đề xuất gắn liền với một mô hình học đa phương thức cụ thể.

1. Mô hình hỏi đáp hình ảnh

Mô hình hỏi đáp hình ảnh tiếng Việt của nhóm Nguyen và các cộng sự [5] được chọn để phân tích vì độ chính xác đạt 60,76% với kiến trúc kết hợp giữa attention và mạng nơ-ron tích chập. Đây là những khối kiến trúc phổ biến trong các tác vụ hỏi đáp hình ảnh.

Kiến trúc của mô hình được thể hiện ở Hình 4, cụ thể như sau:

1) Đầu vào của mô hình gồm một hình ảnh và một câu hỏi.
2) Hình ảnh đi qua song song hai mạng gồm: i) ResNet để trích xuất đặc trưng thị giác cục bộ; ii) Transformer thị giác (ViT) để trích xuất đặc trưng thị giác toàn cục, trong khi câu hỏi sẽ đi qua PhoBERT và tầng self-attention đa đầu (MHSA) để cho ra đặc trưng ngôn ngữ. Như vậy có ba đặc trưng được tạo ra, gồm hai đặc trưng thị giác và một đặc trưng ngôn ngữ.
3) Mỗi đặc trưng thị giác sẽ đi qua tầng attention có hướng dẫn tương ứng. Trong tầng này, đặc trưng thị giác sẽ làm truy vấn và đặc trưng ngôn ngữ sẽ làm khóa và giá trị. Kết quả là cho ra hai đặc trưng thị giác có hướng dẫn tương ứng với từng loại. Tầng attention có hướng dẫn của đặc trưng ngôn ngữ sẽ được trình bày sau.
4) Từng đặc trưng thị giác có hướng dẫn sẽ đi qua một tầng top-down attention để cho ra một vector đặc trưng thị giác. Hai vector đặc trưng thị giác được tạo ra bởi quá trình này sẽ nối lại để trở thành đặc trưng thị giác đa ngữ nghĩa.
5) Quay lại với đặc trưng ngôn ngữ. Sau khi được tạo ra, đặc trưng ngôn ngữ sẽ đi qua attention có hướng dẫn với đặc trưng ngôn ngữ là truy vấn và đặc trưng thị giác đa ngữ nghĩa làm khóa và giá trị. Kết hợp với phép gộp thì cuối cùng cho ra đặc trưng ngôn ngữ có hướng dẫn.
6) Đặc trưng thị giác đa ngữ nghĩa và đặc trưng ngôn ngữ có hướng dẫn sẽ đi qua tầng kết nối đầy đủ và áp dụng tích Hadamard lên chúng để cho ra đặc trưng đa phương thức.
7) Cuối cùng, đặc trưng đa phương thức đi qua lớp phân loại để đưa ra nhãn.

Mô hình này được huấn luyện và đánh giá trên tập dữ liệu cho tác vụ hỏi đáp hình ảnh tiếng Việt ViVQA [11] với độ chính xác là 60.76% và điểm F1 là 59.86%.

Chúng tôi đề xuất thử nghiệm những phương pháp đã được áp dụng thành công cho tập dữ liệu tiếng Anh. Chúng tôi tập trung vào giải thích ba mô-đun: mạng nơ-ron tích chập, kiến trúc Transformer và cơ chế attention có hướng dẫn.

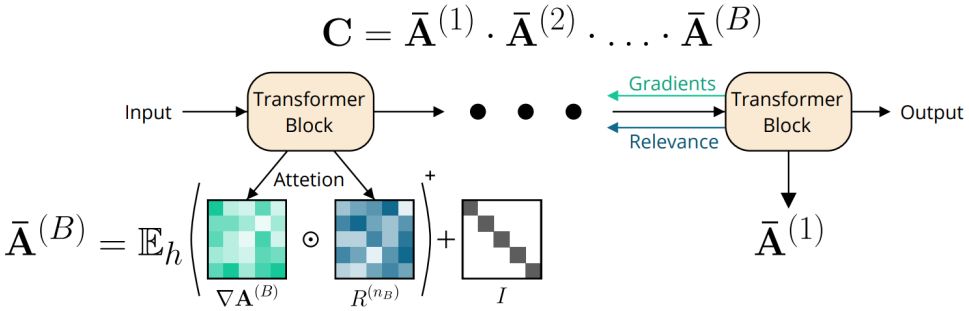
2. Phương pháp

a) Giải thích mạng nơ-ron tích chập bằng Grad-CAM

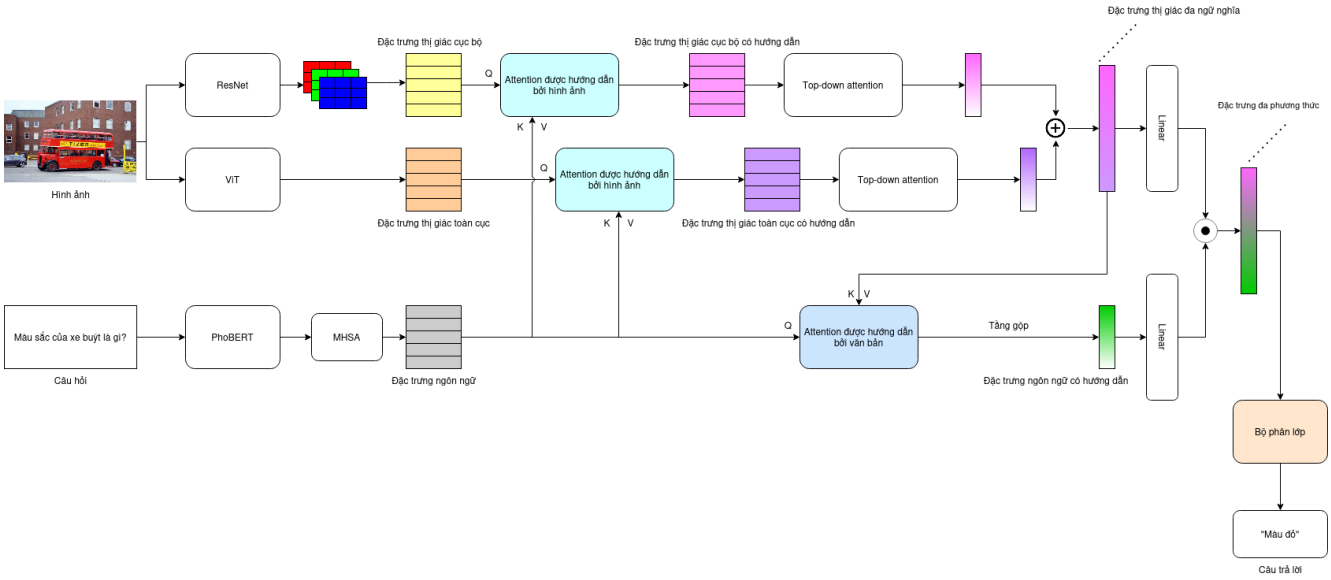
Nghiên cứu về Grad-CAM [2] đã cung cấp đầy đủ lý thuyết về phương pháp này. Gọi l là nhãn tại đầu ra của mô hình, y^l là logit tại tầng phân lớp (trước khi đi qua softmax). Tại một tầng tích chập A cuối cùng trong kiến trúc, ta dễ dàng tính được Gradient $\mathbf{G} = \frac{\partial y^l}{\partial A}$. Ngoài ra, ta cũng lưu lại giá trị trong quá trình lan truyền tiến (forward) của mô hình tại tầng tích chập A này, gọi là $\mathbf{C}$. $\mathbf{G}$ và $\mathbf{C}$ đều là tensor 4 chiều (đánh số từ 0 đến 3) có kích thước (b, c, n, n) với b là kích thước một batch đầu vào; c là số kênh sau khi đi qua tầng tích chập; n là kích thước của ma trận đặc trưng thu được, ở đây ta chỉ xét ma trận đặc trưng vuông.

Quá trình tìm biểu đồ nhiệt như sau:

1) Tính vector Gradient gộp của $\mathbf{G}$ (gọi là $\mathbf{g_p}$) bằng cách tính trung bình $\mathbf{G}$ theo chiều thứ 0, 2 và 3, tức là b và hai kích thước n của ma trận đặc trưng. Vector gộp của Gradient $\mathbf{g_p}$ có số chiều bằng với kích thước c của $\mathbf{G}$.
2) Tại mỗi kênh thứ i của ma trận $\mathbf{C}$, ta nhân chúng với phần tử thứ i của vector gộp $\mathbf{g_p}$.
3) Tính trung bình theo chiều thứ 1 của $\mathbf{C}$, tức là trung bình theo toàn bộ kênh. Kết quả thu được là một



Hình 3. Mô tả phương pháp. Gradient và Relevance được truyền từ đầu ra về lại đầu vào. Nguồn: Từ bài báo "Transformer Interpretability Beyond Attention Visualization" [3].



Hình 4. Kiến trúc của mô hình hỏi đáp hình ảnh được sử dụng trong nghiên cứu của nhóm Nguyen và các cộng sự [5].

tensor 3 chiều (gồm 1 chiều chứa số batch ảnh và 2 chiều chứa biểu đồ nhiệt), chứa ma trận có thể được trực quan hoá thành biểu đồ nhiệt.

b) Giải thích self-attention

Chúng tôi sử dụng phương pháp được đề xuất trong nghiên cứu của nhóm Chefer và các cộng sự [3] về diễn giải cơ chế self-attention thông qua biểu đồ Relevance.

Đầu tiên, khởi tạo ma trận Relevance **R**. Gọi i là số lượng token đầu vào của mô-đun self-attention. Do mô hình này chỉ sử dụng cơ chế self-attention cho nên ta sẽ khởi tạo ma trận Relevance là một ma trận đơn vị.

$$\mathbf{R} = \mathbb{I}^{i \times i} \tag{11}$$

Tiếp theo, thực hiện phép trung bình toàn bộ đầu của mô-đun self-attention sau khi nhân từng phần tử với đạo hàm của chúng. Đặt A là một tầng attention (chứa mô-đun self-attention) nào đó trong mô hình.

$$\bar{A} = E_h(A) = \frac{1}{M} \sum_m (\nabla A_m \odot A_m)^+ \tag{12}$$

Cuối cùng, cập nhật ma trận Relevance.

$$R = R + \bar{A}R \tag{13}$$

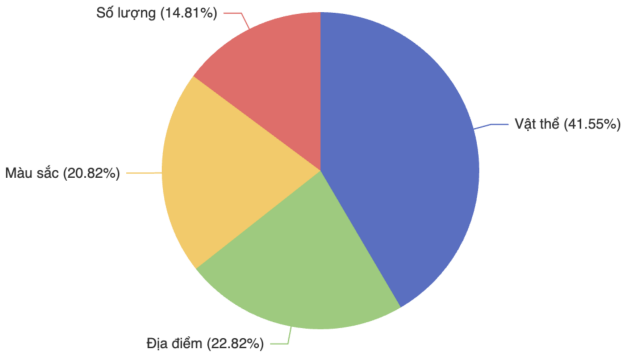
Ma trận R được dùng để trực quan thành biểu đồ nhiệt. Với kiến trúc được sử dụng để xử lý văn bản (ví dụ: PhoBERT), người ta sẽ đặt phần tử Relevance của token `<CLS>` bằng 0.

IV. THỰC NGHIỆM VÀ PHÂN TÍCH

1. Tập dữ liệu

Tập dữ liệu được chúng tôi sử dụng để thực nghiệm là tập ViVQA ([11]) bao gồm 10.328 hình ảnh cùng 15.000 cặp câu hỏi-trả lời bằng tiếng Việt có liên quan đến các

hình ảnh đó. Tác giả chia tập dữ liệu thành các tập huấn luyện và tập kiểm thử với tỷ lệ 8:2.



Hình 5. Phân bổ các câu hỏi trong tập dữ liệu.

Các dạng câu hỏi có trong tập dữ liệu được chia làm bốn loại chính, bao gồm: Vật thể, số lượng, màu sắc, địa điểm. Phân bổ các dạng câu hỏi được trình bày trong hình 5, lưu ý rằng phân bổ này ở cả hai tập dữ liệu là như nhau.

2. Phân tích

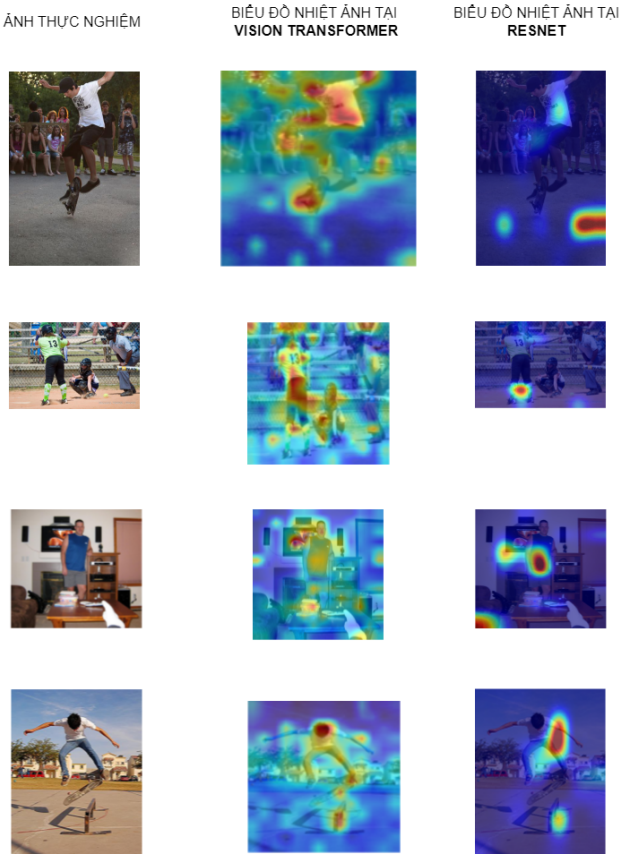
Nghiên cứu tập trung phân tích định tính để đưa ra những nhận định về hành vi của mô hình. Những hành vi này phụ thuộc vào hai thông tin đầu vào là hình ảnh và câu hỏi. Khi thực hiện thay đổi các đầu vào này, chúng tôi quan sát để tìm ra những khuôn mẫu nhất định, hoặc hạn chế của mô hình hỏi đáp hình ảnh. Do đó, chúng tôi thực hiện các thí nghiệm để quan sát sự thay đổi hành vi trả lời như sau:

1) Thực nghiệm trên câu hỏi gốc và hình ảnh gốc của tập dữ liệu.
2) Dựa trên câu hỏi gốc, tạo ra các câu hỏi làm thay đổi ý nghĩa của câu.
3) Dựa trên câu hỏi gốc, tạo ra các câu hỏi có ý nghĩa tương đương.
4) Dựa trên hình ảnh gốc, sử dụng phép biến đổi để làm dịch chuyển một phần thông tin của hình ảnh.

Dựa trên các thí nghiệm trên, nghiên cứu đã rút ra được một số nhận xét như sau:

a) Khối Transformer thị giác có xu hướng trích xuất đặc trưng toàn cục, trong khi khối ResNet chủ yếu tập trung vào việc trích xuất đặc trưng cục bộ.

Như được minh họa trong hình 6, hầu như toàn bộ đối tượng trong ảnh đều được khối *Vision Transformer* chú ý đến, nhất là con người và động vật. Trong khi đó, ở khối *ResNet* lại thể hiện sự tập trung vào một vùng nhất định trong ảnh.



Hình 6. Thực nghiệm trên câu hỏi gốc và hình ảnh gốc để tìm hiểu về xu hướng tập trung của mô hình. Những bức ảnh có nhiều vật thể, Vision Transformer thường tập trung vào toàn bộ vật thể tồn tại trong hình (tính toàn cục); trong khi đó, ResNet chỉ tập trung vào một vùng nhất định (tính cục bộ).

b) Mô hình không thực hiện suy luận tốt khi xử lý các hình ảnh chứa nhiều vật thể.

Chúng tôi phát hiện mô hình thường bỏ qua thông tin đến từ các giới từ như “trong”, “trên”, “dưới”,... vốn ám chỉ đến mối quan hệ giữa các chủ thể trong câu hỏi. Ví dụ ở hình 7, trong câu hỏi *Con gì nằm dưới bàn?*, giới từ *dưới* không được mô hình quá quan tâm ảnh hưởng đến kết quả đưa ra của mô hình là *laptop* - một vật thể nằm phía trên bàn - chứ không phải là *con chó* đang nằm dưới bàn.

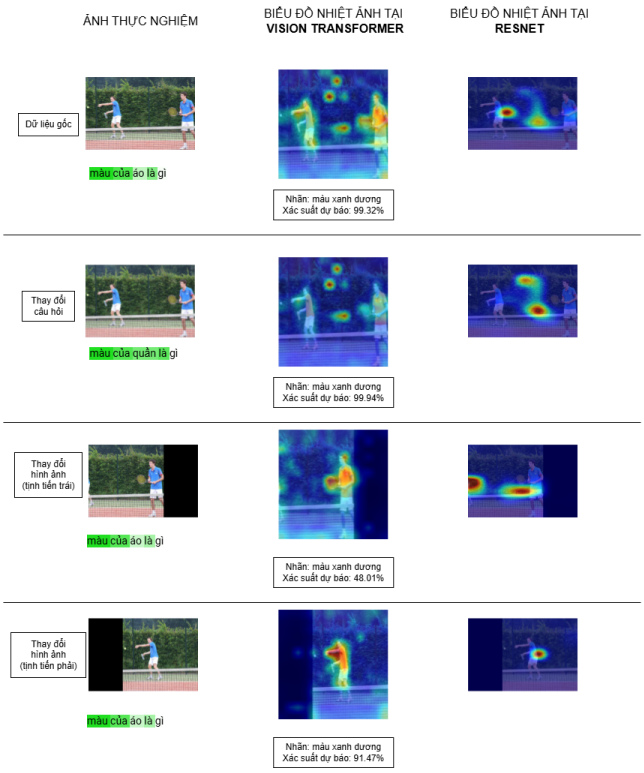
Một phát hiện khá thú vị nữa là mô hình dễ bị đánh lừa bởi những từ khóa gây nhiễu, điển hình như trong câu hỏi *Cậu bé đang ngồi xem cái gì?*, mô hình tập trung nhiều vào chữ *ngồi* mà bỏ qua chữ quan trọng hơn trong ngữ cảnh là *xem*, dẫn đến câu trả lời của mô hình là *cái ghế* với xác suất khá cao. Đây là trường hợp rất dễ gặp trong tiếng Việt khi mà một câu có cấu trúc ngữ pháp phức tạp và chứa nhiều từ "gây nhiễu".

Ngược lại, với những hình ảnh ít vật thể, mô hình trích xuất hợp lý hơn. Trong hình 8, câu hỏi *Màu của áo là gì?*, mô hình tập trung vào người đàn ông và hai chú chó trong

Hình 7. Thực nghiệm trên câu hỏi gốc và hình ảnh gốc (ảnh trên); câu hỏi được thay đổi khác ngữ nghĩa (ảnh dưới) để tìm hiểu về xu hướng tập trung trên câu hỏi và tác động của các chữ được tập trung. Ở ảnh trên, chữ *ngồi xem* được tập trung nhưng vì ngữ nghĩa phức tạp của từ này dẫn đến mô hình hiểu nhầm và cho rằng từ *ngồi xem* cùng nghĩa với từ *ngồi*. Ở ảnh dưới, chữ *dưới* ít được quan tâm dẫn đến mô hình tập trung sai vị trí.

Hình 8. Thực nghiệm trên câu hỏi gốc và hình ảnh gốc để xem xét về khả năng tập trung trên ảnh ít vật thể và ảnh nhiều vật thể.

ảnh đầu tiên; người mặc áo trắng đang trượt ván trong ảnh phía dưới. Mặc dù vậy, mô-đun ResNet đôi khi trích xuất đặc trưng không phù hợp trong ngữ cảnh, ví dụ như mô-đun này tập trung vào hai chú chó mặc dù tâm điểm chúng ta cần là người đàn ông.



Hình 9. Từ trên xuống dưới là các thực nghiệm: Dữ liệu gốc; Thay đổi câu hỏi; Tịnh tiến ảnh sang trái; Tịnh tiến ảnh sang phải.

c) Thông tin văn bản ít ảnh hưởng đến kết quả suy luận hơn so với thông tin thị giác.

Chúng tôi thực hiện thí nghiệm với đầu vào gồm bốn tình huống được minh họa ở hình 9:

1) Sử dụng hình ảnh và câu hỏi gốc.
2) Sử dụng hình ảnh gốc và thay đổi câu hỏi khác nghĩa với câu hỏi gốc.
3) Sử dụng câu hỏi gốc và tịnh tiến ảnh sang trái.
4) Sử dụng câu hỏi gốc và tịnh tiến ảnh sang phải.

So với biểu đồ nhiệt với đầu vào là hình ảnh gốc và câu hỏi gốc, thí nghiệm thay đổi câu hỏi cho thấy sự chuyển dịch vùng tập trung của biểu đồ nhiệt không sâu sắc bằng thí nghiệm thay đổi hình ảnh.

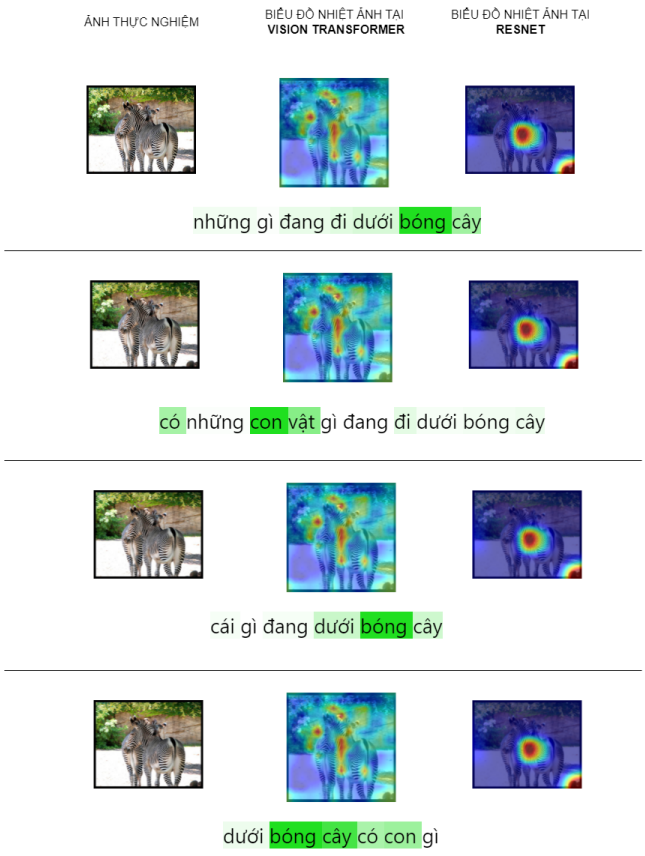
Cụ thể, đối với thí nghiệm thay đổi câu hỏi, biểu đồ nhiệt ảnh có sự thay đổi nhẹ ở độ đậm nhạt của vùng tập trung. Khi xem xét đến sự thay đổi của xác suất phân lớp trong thí nghiệm này so với câu hỏi gốc, kết quả cho thấy không có sự thay đổi đáng kể.

Sự suy giảm xác suất dự báo trong hai thí nghiệm tịnh tiến ảnh rõ rệt hơn thí nghiệm thay đổi câu hỏi. Bên cạnh đó, tịnh tiến ảnh sang trái cho thấy rõ rệt nhất. Ngoài ra, biểu đồ nhiệt ảnh trong thí nghiệm thay đổi câu hỏi chỉ thay đổi cường độ, tức là độ đậm nhạt, của vùng tập trung chứ không tạo thêm hay bỏ đi vùng nào.

Điều này cũng dễ hiểu vì hình ảnh là nguồn cung cấp ngữ cảnh trong tác vụ hỏi đáp hình ảnh. Nếu không trích xuất đặc trưng câu hỏi thì mô hình vẫn có thể dự đoán được với xác suất dự báo tương đối cao dựa vào vật thể trong ảnh; còn nếu không trích xuất đặc trưng hình ảnh thì sẽ gây khó khăn cho việc dự đoán của mô hình.

d) Khối PhoBERT thể hiện sự thiên vị nhất định trong quá trình suy luận.

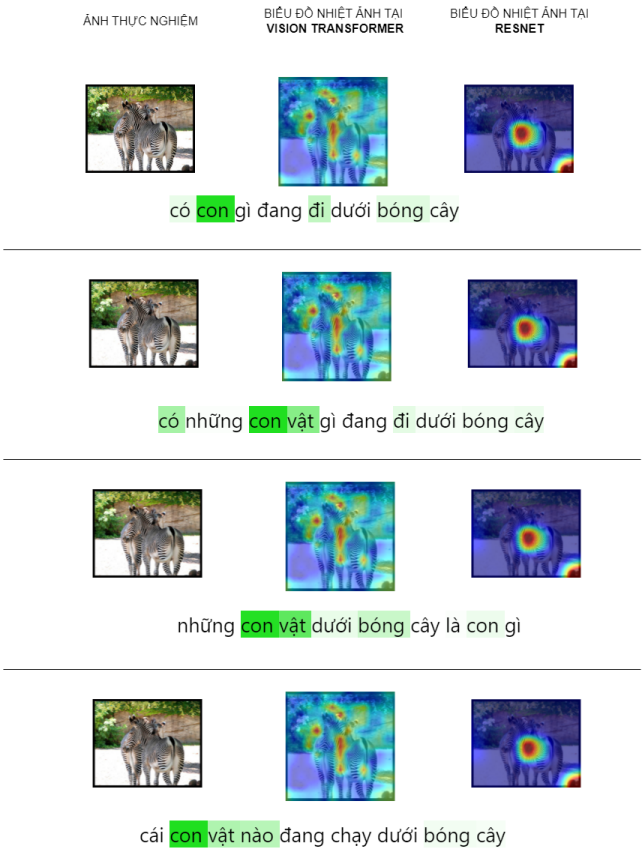
Trong thí nghiệm này, chúng tôi tập trung vào việc thay đổi câu hỏi như trong hình 10. Khi thay đổi câu hỏi thì hầu như biểu đồ nhiệt ảnh không thay đổi và các danh từ (bóng cây, con vật) được quan tâm nhiều hơn. Khi thay đổi sang dạng câu hỏi được thể hiện ở Hình 11 thì hầu như sự tập trung của mô hình dồn về chữ *con*. Từ đó cho thấy, khối PhoBERT khi được huấn luyện hoặc tinh chỉnh đã có thiên vị với một khuôn mẫu câu hỏi nhất định.



Hình 10. Thực nghiệm sử dụng một hình ảnh cùng với nhiều câu hỏi có cùng ngữ nghĩa để tìm hiểu sự tập trung của mô hình trên phương diện câu hỏi. Dễ nhận thấy một số danh từ như *bóng cây* hay *con vật* được tập trung nhiều.

V. KẾT LUẬN

Nghiên cứu đã triển khai các phương pháp diễn giải cho ba khối trong mô hình hỏi đáp hình ảnh, cung cấp các diễn



Hình 11. Thực nghiệm sử dụng một hình ảnh cùng với nhiều câu hỏi có cùng ngữ nghĩa có chữ *con* ở đầu câu để tìm hiểu xem mô hình có thiên vị với dạng câu hỏi này không. Với dạng câu chứa chữ *con* ở phía đầu câu này, khối PhoBERT luôn cho rằng chữ này quan trọng nhất.

giải trực quan cho phương thức hình ảnh và văn bản. Đồng thời, chúng tôi cũng đã thử nghiệm và khám phá ra một số đặc điểm trong hành vi của mô hình, đưa ra được một số điểm yếu mà mô hình thường gặp phải dẫn đến việc đưa ra những phán đoán nhầm lẫn.

Tuy nhiên, việc đưa ra biểu đồ nhiệt có thể chưa phản ánh hết tính chất của mô hình và bản thân của mô hình không có khả năng tự diễn giải và phản hồi để tự cải thiện. Nghiên cứu có thể phát triển tiếp để cung cấp khả năng diễn giải bằng ngôn ngữ tự nhiên, dễ hiểu hơn so với biểu đồ nhiệt. Ngoài ra, việc đánh giá độ hiệu quả của diễn giải vẫn còn là bài toán mở cần nhiều nghiên cứu hơn.

TÀI LIỆU THAM KHẢO

[1] A. Weller, *Transparency: Motivations and Challenges*. Cham: Springer International Publishing, 2019.

[2] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 618–626, 2017.

[3] H. Chefer, S. Gur, and L. Wolf, "Transformer interpretability beyond attention visualization," *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 782–791, 2020.
[4] T. Le, K. Pho, T. Bui, H. T. Nguyen, and M. L. Nguyen, "Object-less vision-language model on visual question classification for blind people," in *Proceedings of the 14th International Conference on Agents and Artificial Intelligence - Volume 3: ICAART*, pp. 180–187, SciTePress, Feb. 2022. ICAART 2022, February 3–5, 2022.
[5] A. D. Nguyen, T. Le, and H. T. Nguyen, "Combining multi-vision embedding in contextual attention for vietnamese visual question answering," in *Image and Video Technology*, (Cham), pp. 172–185, Springer International Publishing, 2023.
[6] Y. Lyu, P. P. Liang, Z. Deng, R. Salakhutdinov, and L.-P. Morency, "DIME: Fine-grained interpretations of multimodal models via disentangled local explanations," 2022. arXiv preprint.
[7] D. H. Park, L. A. Hendricks, Z. Akata, A. Rohrbach, B. Schiele, T. Darrell, and M. Rohrbach, "Multimodal explanations: Justifying decisions and pointing to the evidence," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8779–8788, 2018.
[8] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2921–2929, 2016.
[9] S. Abnar and W. Zuidema, "Quantifying attention flow in transformers," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, (Online), pp. 4190–4197, Association for Computational Linguistics, July 2020.
[10] S. Bach, A. Binder, G. Montavon, F. Klauschen, K.-R. Müller, and W. Samek, "On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation," *PLoS ONE*, vol. 10, 2015.
[11] K. Q. Tran, A. T. Nguyen, A. T.-H. Le, and K. V. Nguyen, "Vivqa: Vietnamese visual question answering," in *Pacific Asia Conference on Language, Information and Computation*, 2021.

Tran Thi Phuong Linh received her B.S. degree in Computer Science from the University of Science, Ho Chi Minh City, Vietnam (VNU-HCM) in 2024. She research interests center on deep learning models to make their predictions more understandable and trustworthy.
Email: linh.ttphuong@fit.hcmus.edu.vn

Nguyen Thanh Tan received his B.S. degree in Software Technology in 2024 from the University of Science, Ho Chi Minh City, Vietnam (VNU-HCM). He is currently working as a Software Engineer in the industry. His research interests center on Deep Learning, with a particular focus on strategies for optimizing the cost and computational efficiency of Large Language Models (LLMs) for enterprise applications.
Email: tan.nt@fit.hcmus.edu.vn

Le Thanh Tung earned his B.S. degree in Computer Science from the Honour Program at the University of Science, Vietnam National University, Ho Chi Minh City, in 2012, and his M.S. degree in Information Science from the Japan Advanced Institute of Science and Technology (JAIST) in 2018. He completed his Ph.D. in Information Science at JAIST in 2021. Now, he is a Lecturer of Information Technology at the University of Science, Ho Chi Minh City, Vietnam (HCMUS). His research interests include information retrieval, natural language processing, visual question answering, and multi-modal systems
Email: tung.lt@fit.hcmus.edu.vn

Nguyen Tien Huy received his B.S. degree in Software Technology in 2010 and his M.S. degree in Computer Science in 2015, both from the University of Science, Ho Chi Minh City, Vietnam (VNU-HCM). He earned his Ph.D. in Information Science from the Japan Advanced Institute of Science and Technology (JAIST) in 2019. He is currently a lecturer in the Department of Computer Science, Faculty of Information Technology, VNU-HCM. His research interests include Natural Language Processing and Deep Learning, with a particular focus on intelligent systems and multimodal data analysis.
Email: ntienhuy@fit.hcmus.edu.vn

Một phương pháp giấu tin thuận nghịch có khả năng nhúng cao sử dụng kỹ thuật ảnh kép và hệ tính toán cơ số 13.

Phạm Minh Thái[1], Lê Kinh Tài[2], Nguyễn Hiếu Cường[3*], Nguyễn Quang Chánh[1]
[1]Trường Đại học Kinh tế - Kỹ thuật Công nghiệp, 456 Minh Khai, Hà Nội, Việt Nam
[2]Trường Đại học Công đoàn, 169 Tây Sơn, Đống Đa, Hà Nội, Việt Nam
[3]Trường Đại học Giao thông Vận tải, số 3 Cầu Giấy, Hà Nội, Việt Nam

Tác giả liên hệ: Nguyễn Hiếu Cường. Email: cuongnh@utc.edu.vn
Ngày nhận bài: 18/03/2025, ngày sửa chữa: 24/04/2025, ngày duyệt đăng: 22/05/2025
DOI: 10.32913/mic-ict-research-vn.v2025.n1.1364

Tóm tắt: Phương pháp Chen và cộng sự [1] đề xuất một sơ đồ RDH mới dựa trên khai thác biến đổi theo hướng (EMD) với hình ảnh kép. Nhờ khai thác các đặc tính của ma trận EMD, một bit và một chữ số cơ số 5 có thể được giấu trong một ảnh gốc để tạo ra một cặp điểm ảnh chứa thông điệp. Dữ liệu sẽ được nhúng trên 1 trong 4 hướng, trong đó các tác giả nhúng thêm được 1 bit ngoài việc nhúng 1 hệ cơ số 5, nên tỷ lệ nhúng $\frac{\log_2 5+1}{2} = 1.66$ bpp. Để tăng khả năng nhúng, chúng tôi đề xuất một phương pháp nhúng dựa trên cơ sở số 13. Vì vậy, tỷ lệ nhúng có thể gần bằng: $\frac{\log_2 13}{2} = 1.8502$ bpp. Ngoài ra, số lượng ảnh biến đổi trong phương pháp giải quyết vấn đề xuất ra nhỏ hơn, nên chất lượng ảnh tốt hơn phương pháp của Chen và cộng sự.

Từ khóa: *giấu tin thuận nghịch, ảnh kép, hệ tính toán cơ số 13, PSNR, ER.*

Title: **A Reversible Steganography with High Embedding Capacity using Dual Image Technique and Base 13 Calculation System**

Abstract: Chen et al. [1] proposed a new RDH scheme based on the extraction of directional transformations (EMD) with dual images. By exploiting the properties of the EMD matrix, a bit of the password and a base 5 number can be hidden in an original image to generate a pair of pixels containing relaxation. The data will be embedded in 1 of 4 instructions, in which the authors embed 1 more aromatic in addition to embedding 1 base 5, so the embedding ratio is $\frac{\log_2 5+1}{2} = 1.66$ bpp. To increase the embedding capacity, we propose an embedding method based on base 13. Therefore, the embedding ratio can be developed by using: $\frac{\log_2 13}{2} = 1.8502$ bpp. In addition, the number of image transformations in the output problem solving method is smaller, so the image quality is better than the method of Chen et al.

Keywords: *reversible data hidding, dual-image, base 13 number system, PSNR, ER.*

I. GIỚI THIỆU

Việc truyền thông tin qua Internet và lưu trữ trên "đám mây" ngày càng được áp dụng rộng rãi. Thế nhưng, Internet và "đám mây" dễ bị tin tặc xâm phạm, nên việc bảo vệ thông tin mật là một yêu cầu quan trọng. Mã hóa dữ liệu và giấu tin trên các tệp đa phương tiện là hai giải pháp bảo mật quan trọng. Trong mã hóa, dữ liệu gốc (bản rõ) bị biến đổi thành một dãy ký tự lạ (bản mã), khiến tin tặc không thể biết ý nghĩa của chúng. Tuy nhiên, tin tặc dễ dàng nhận ra bên trong các ký tự lạ có thể chứa dữ liệu quan trọng, khiến chúng cố gắng thám mã. Đây là nhược điểm của phương pháp mật mã [2]. Theo hướng giấu tin, thông tin mật được giấu vào các tệp ảnh; sau đó ảnh chứa tin được chuyển đi và người nhận hợp pháp có thể lấy thông tin từ ảnh được gửi cho họ. Vì ảnh ban đầu và ảnh gửi đi không thể phân biệt được, nên sẽ dễ dàng qua mặt tin tặc.

Các thuật toán giấu tin không thuận nghịch, như [3–6], chỉ có thể lấy dữ liệu đã giấu mà không thể tái tạo được ảnh ban đầu. Vì vậy, phạm vi ứng dụng của chúng bị hạn chế trong nhiều trường hợp cần tái tạo ảnh gốc. Do đó, các kỹ thuật giấu tin có khả năng phục hồi (reversible) hoặc không mất dữ liệu (lossless), thường được gọi chung là RDH (Reversible Data Hiding), đã ra đời và ngày càng được hoàn thiện.

Trong giai đoạn đầu, các phương pháp RDH chủ yếu dựa trên việc nén không tổn hao các bit có trọng số thấp trong ảnh. Các kỹ thuật này có khả năng nhúng thấp vì hiệu quả nén các bit thấp không cao. Mở rộng hiệu số, viết tắt là DE (Difference Expansion) của Tian [7], là kỹ thuật đáng chú ý đầu tiên. Trong đó, hiệu của một cặp hai điểm lân cận được dịch chuyển sang trái để một bit được chèn vào bên phải. Tuy DE có thể nhúng được nhiều bit, nhưng các điểm ảnh bị biến đổi nhiều, nên ảnh chứa tin có chất lượng thấp. Phương pháp DE được nhiều người cải tiến và mở rộng nhằm nâng cao độ trung thực hình ảnh đồng thời tăng dung lượng nhúng dữ liệu như [8–10].

Sau đó, kỹ thuật dịch chuyển histogram của Ni và cộng sự [11], gọi là HS (Histogram Shifting), có thể xem như phương pháp đáng chú ý thứ hai sau DE. Trong HS, histogram của ảnh được xây dựng. Hai giá trị đỉnh histogram có tần suất xuất hiện cao nhất được xác định lần lượt là PeakL (giá trị nhỏ hơn) và PeakR (giá trị lớn hơn). Các pixel trong ảnh được dịch chuyển về phía trái của PeakL và về phía phải của PeakR nhằm tạo ra các vùng trống tại hai vị trí đỉnh này; chúng sẽ được dùng để chứa dữ liệu. Trong HS, mỗi phần tử ảnh chỉ bị biến đổi một đơn vị nên chất lượng ảnh sau khi nhúng tin khá cao. Số bit có thể nhúng trong HS bằng tổng số lần xuất hiện tại PeakL và PeakR; tổng này nói chung không lớn, vì đa số ảnh tự nhiên có histogram không nhọn nên các giá trị PeakL và PeakR thường nhỏ. Kỹ thuật mở rộng hiệu số và dịch chuyển biểu đồ tần suất, với những đặc tính ưu việt, thường được ứng dụng làm nền tảng để phát triển nhiều thuật toán RDH khác nhau [12–20].

Yếu điểm của các kỹ thuật RDH là dung lượng nhúng không cao. Xuất phát từ nhu cầu thực tế đó, nhiều giải pháp RDH sử dụng ảnh kép đã ra đời và không ngừng được hoàn thiện. Các kỹ thuật này tạo hai bản sao từ ảnh gốc; cả hai đều được dùng để chứa dữ liệu, nên khả năng nhúng của chúng vượt trội so với các phương pháp RDH thông thường. Độ trung thực hình ảnh của các phương pháp này cũng rất cao vì độ biến đổi của các bản sao thường không nhiều. Ngoài ra, độ an toàn của chúng cũng rất cao vì phải biết cả hai bản sao ảnh chứa tin thì mới có thể trích xuất dữ liệu. Các phương pháp RDH ảnh kép có thể chia thành một số kiểu sau:

Kiểu thứ nhất dùng cách gấp ở giữa (center folding strategy). Quy trình nhúng giá trị $d, (d \geq 0)$ vào điểm ảnh x được thực hiện bằng cách tách d thành hai thành phần: $u = \left\lfloor \frac{d}{2} \right\rfloor$ và $v = \left\lceil \frac{d}{2} \right\rceil$. Sau đó, hai điểm dùng để nhúng d được tính như sau: $x' = x + u$ và $x'' = x - v$. Rõ ràng, x có thể được khôi phục bằng công thức: $x = \left\lceil \frac{x'+x''}{2} \right\rceil$.

Kiểu thứ hai [21, 22]: cặp điểm ảnh ban đầu được kết hợp và biểu diễn dưới dạng một điểm P trên hệ tọa độ. Một hình vuông mỗi chiều 3 đơn vị, có tâm là điểm P, được tạo ra. Các cặp điểm giấu tin được chọn từ hai pixel liền kề trong hình vuông. Với 25 tổ hợp khả thi, các phương pháp này có khả năng nhúng hơn 4 bit trên mỗi cặp pixel, tương đương tỷ lệ nhúng trung bình lớn hơn 1 bit/pixel (bpp). Do các cặp pixel dùng để giấu tin đều nằm kề cặp pixel gốc P, mỗi giá trị điểm ảnh chỉ bị biến đổi tối đa một đơn vị, nhờ đó đảm bảo chất lượng hình ảnh sau khi nhúng luôn ở mức cao.

Kiểu thứ ba sử dụng hàm khai thác biến đổi hướng (EMD – Exploiting Modification Direction) để xây dựng một ma trận hàm với kích thước 256×256, tương ứng với không gian điểm ảnh 256×256. Mỗi cặp hai điểm ảnh được xem như một điểm (phần tử) trên ma trận hàm. Hai cặp điểm ảnh nhúng tin được xác định như những điểm trên ma trận hàm. Mặc dù cho dung lượng nhúng lớn, các kỹ thuật RDH dựa trên hàm EMD cho hai ảnh [1, 23–25] thường dẫn đến suy giảm chất lượng hình ảnh đáng kể.

Gần đây, Chen và cộng sự trình bày một phương pháp nhúng tin thuận nghịch trên hai ảnh bằng cách kết hợp EMD và việc chọn một trong bốn hướng (trái, phải, trên, dưới), đạt được khả năng nhúng cao (ER = 1.66), nhưng độ biến đổi tối đa của một ảnh lớn (khoảng 5 đơn vị). Phương pháp này được xem như một trong các phương pháp ảnh kép có khả năng nhúng cao nhất.

Phương pháp đề xuất của chúng tôi sử dụng 13 tổ hợp nhúng dữ liệu trên một điểm ảnh, nên ER có thể đạt xấp xỉ bằng $(\log_2 13)/2 = 1.8502$, tức là cao hơn so với phương pháp của Chen. Ngoài ra, độ biến đổi tối đa của mỗi ảnh là 3 đơn vị, nên chất lượng ảnh cao hơn phương pháp của Chen.

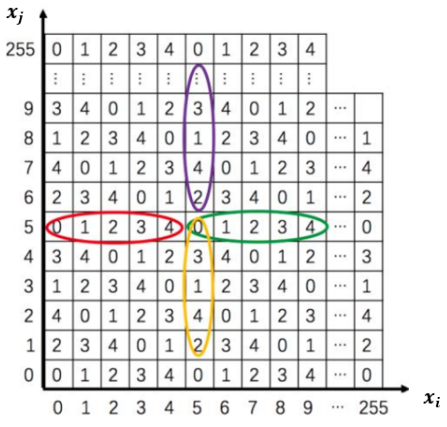
Bài báo được tổ chức như sau. Mục II trình bày phương pháp của Chen và cộng sự [1]. Phương pháp đề xuất được giới thiệu ở mục III. Mục IV trình bày kết quả thử nghiệm trên 12 ảnh để so sánh khả năng nhúng và chất lượng ảnh của phương pháp đề xuất với phương pháp của Chen và cộng sự [1]. Kết luận được đưa ra ở mục IV.

II. PHƯƠNG PHÁP CHEN VÀ CỘNG SỰ

Chen và cộng sự [1] đã đề xuất một sơ đồ giấu tin thuận nghịch trên hai ảnh sử dụng biến đổi theo hướng (EMD) và cho khả năng nhúng cao. Đầu tiên, tác giả xây dựng một ma trận tham chiếu EMD có kích thước 256×256 như Hình 1; ma trận này được xây dựng theo Công thức (1) như sau:

$$f(x_i, x_j) = (x_i + 2x_j) \bmod 5 \quad (1)$$

Phương pháp này nhúng trên cặp điểm ảnh (x_i, x_j), với điều kiện $5 \leq x_i, x_j \leq 251$. Trong phương pháp này, mỗi đoạn trong Hình 1 có độ dài là 5. Việc nhúng dữ liệu được thực hiện theo một trong bốn hướng; nhờ có bước lựa chọn này, tác giả nhúng thêm được 1 bit ngoài việc nhúng theo



Hình 1. Ma trận tham chiếu EMD E

một hệ cơ số 5. Các phương pháp khác thường chỉ nhúng dữ liệu dựa trên ma trận này.

1. Thuật toán nhúng tin trên một cặp điểm ảnh

Đầu vào: Ảnh X có kích thước $M \times N$, dãy bit dữ liệu $B = (b_1, \cdots, b_s)$

Đầu ra: Ảnh X',X''

Phương pháp:

B1: Xây dựng ma trận tham chiếu EMD E như Hình 1. Tạo một dãy bit nhị phân ngẫu nhiên $r_1, r_2, \cdots, r_K$ với $(K = M \times N)$.

B2: Chia các thông điệp bí mật B thành 2 thành phần:

$*$ Một thành phần chuyển sang hệ cơ số 5 kí hiệu là $d_1,\ d_2, \cdots, d_K\ \ (d_i = 0 \rightarrow 4)$.

$*$ Chọn 1 bit bí mật b_i và một chữ số hệ cơ số 5 d_i tạo thành 1 cặp (d_i, b_i).

$*$ Nhận xét rằng những điểm ảnh x_i từ ảnh X, mà $x_i \in [0,\ 4]$ hoặc $x_i \in [252,\ 255]$ thì x_i không nhúng và giá trị được giữ nguyên vào X', X'' ở cùng 1 ví trí.

$*$ Liệt kê tất cả các cặp ảnh liên tiếp $(x_i, x_j), 5 \leq x_i,\ x_j \leq 251$.

$*$ Xét nếu số ngẫu nhiên $r_i = 0$, thì lấy theo chiều ngang:

- Nếu bit nhị phân $b_i = 0$ lấy về phía bên trái.
- Nếu bit nhị phân $b_i = 1$ lấy về phía bên phải.

$*$ Nếu số ngẫu nhiên $r_i = 1$, thì lấy theo chiều dọc:

- Nếu bit nhị phân $b_i = 0$ lấy về phía bên dưới.
- Nếu bit nhị phân $b_i = 1$ lấy về phía bên trên.

$*$ Ví dụ xét 1 bit đầu tiên b_1:

- Nếu $r_1 = 0$ và $b_1 = 0$ thì nhúng vào đoạn bên trái.
- Nếu $r_1 = 0$ và $b_1 = 1$ thì nhúng vào đoạn bên phải.
- Nếu $r_1 = 1$ và $b_1 = 0$ thì nhúng vào đoạn bên dưới.
- Nếu $r_1 = 1$ và $b_1 = 1$ thì nhúng vào đoạn bên trên.

$*$ Ví dụ: Cho $x = (5,\ 5)$, bit bí mật 0100 và $r = 0$:

- $r = 0$ và $b_1 = 0$ nên lấy đoạn bên trái.
- $d = 100_2 = 4_{10}$.

Dựa vào Hình 1 với $d = 4$, $x = (5, 5)$ ta có $x' = 4$, $x'' = 5$.

2. Thuật toán trích tin và khôi phục ảnh

Đầu vào: Ảnh X', X'', một dãy bit nhị phân ngẫu nhiên $r_1,\ r_2, \cdots, r_K, (K = M \times N)$.

Đầu ra: Ảnh X, dãy bit dữ liệu B.

Phương pháp:

$*$ Tất cả các cặp điểm ảnh đều làm giống nhau nên ở đây xét đại diện 1 cặp điểm ảnh $(x'_i,\ x''_j)$.

$*$ Với r_i xét giống như thuật toán nhúng tin.

$*$ Nếu $x'_j < x''_j$ lấy đoạn bên trái ta được b_i và tra theo Hình 1 được d_i.

$*$ Nếu $x'_j \geq x''_j$ lấy đoạn bên phải ta được b_i và tra theo Hình 1 được d_i.

$*$ Nếu $x'_i = x''_j$ và thuộc đoạn [0, 4] hoặc đoạn [252, 255] thì $x_i = x'_i = x''_j$.

$*$ Nhận xét trong x'_i hoặc x''_j phải có 1 cái bằng x_i.

- Nếu $5 \leq\ x'_i \leq 251$ thì $x_i = x'_i$.
- Ngược lại $x_i = x''_i$.

$*$ Ví dụ như trên:

Đã có $x' = 4, x'' = 5$ và $r = 0$

- Xét thấy $x' < x''$ và $r = 0$ nên $b_1 = 0$.
- Dựa vào Hình 1 với $b_1 = 0$, $x' = 4, x'' = 5$ ta được $d = 4_{10} = 100_2$.
- Từ đây thu được dãy bit bí mật là: 0100
- Thấy $x' < 5$ nên suy ra $x = x'' = 5$. Vậy $x = (5, 5)$

$*$ Nhận xét rằng theo phương pháp của Chen và cộng sự [1] thì khả năng nhúng là: $\log_2 5 + 1 = 3.32$ bit và tỷ lệ nhúng $ER = \frac{\log_2 5+1}{2} = \frac{3.32}{2} = 1.66$ bpp.

III. PHƯƠNG PHÁP ĐỀ XUẤT

1. Nhúng một chữ số hệ cơ số 13 trên một điểm ảnh

$*$ Vì cơ chế nhúng áp dụng tương tự cho từng điểm ảnh, chúng tôi chỉ cần mô tả thuật toán cho một điểm ảnh đại diện. Ở đây, chúng tôi sẽ nhúng vào 1 điểm ảnh gọi là x để được 2 điểm ảnh chứa tin gọi là x' và x''.

$*$ Chúng tôi đề xuất công thức nhúng như sau:

$$x' = x + a, x'' = x + b \tag{2}$$

Với điều kiện như sau:

- Điều kiện 1: $-3 \leq a, b \leq 3$.

Bảng I
QUY TẮC NHÚNG CỦA THUẬT TOÁN ĐỀ XUẤT

d	a	b	x'	x''	$x' - x''$
0	-3	3	x - 3	x + 3	-6
1	-3	2	x - 3	x + 2	-5
2	-2	2	x - 2	x + 2	-4
3	-2	1	x - 2	x + 1	-3
4	-1	1	x - 1	x + 1	-2
5	-1	0	x - 1	x	-1
6	0	0	x	x	0
7	0	-1	x	x - 1	1
8	1	-1	x + 1	x - 1	2
9	1	-2	x + 1	x - 2	3
10	2	-2	x + 2	x - 2	4
11	2	-3	x + 2	x - 3	5
12	3	-3	x + 3	x - 3	6

- Điều kiện 2: $\lceil \frac{a+b}{2} \rceil = 0$. Từ đây, chúng tôi sẽ xây dựng được Bảng I như sau:

∗ Ví dụ: Cho $x = 125$ và $d = 10$.

Tra theo Bảng I ta được:

$x' = 125 + 2 = 127$

$x'' = 125 - 2 = 123$

Từ Bảng I, ta có thể khái quát hoá công thức tổng quát cho quy tắc nhúng như sau:

∗ Điều kiện cơ bản:

- Giá trị điểm ảnh gốc: x

- Giá trị điểm ảnh sau khi nhúng: x' và x''.

- Chữ số hệ cơ số 13 cần nhúng: $d \in 0,\ 1,\ 2,\ \cdots, 12$.

- Độ dịch chuyển a và b thỏa mãn điều kiện 1 và 2 của công thức 2.

∗ Công thức xác định a và b:

Ta có thể biểu diễn a và b dựa trên giá trị d như sau:

$$a = \begin{cases} -3 & \text{nếu } d \in \{0,1\}, \\ -2 & \text{nếu } d \in \{2,3\}, \\ -1 & \text{nếu } d \in \{4,5\}, \\ 0 & \text{nếu } d \in \{6,7\}, \\ 1 & \text{nếu } d \in \{8,9\}, \\ 2 & \text{nếu } d \in \{10,11\}, \\ 3 & \text{nếu } d = 12, \end{cases}$$

và $b = d - a$

- Công thức tổng quát cho x' và x'' như công thức 2.

2. Xác định dãy bit cho một lần nhúng

Bước 1: Gọi k số pixel cho một lần nhúng. Như vậy mỗi 1 lần nhúng chúng ta sẽ nhúng được k chữ số hệ 13.

Bước 2: Gọi d là giá trị cực đại của cơ số hệ 13 với độ dài k, suy ra $d = 13^k - 1$.

Bước 3: Chuyển d thành hệ cơ số 2 suy ra ta được một giá trị nhị phân gọi là bd và có độ dài là ld (đây chính là số bit dữ liệu có thể nhúng).

Bước 4: Số nhị phân lớn nhất có độ dài ld là: $b_{max} = 2^{ld} - 1$

Bước 5: Số nhị phân nhỏ nhất có độ dài ld là: $b_{min} = 2^{ld-1}$

Bước 6:

- Lấy ra ld bit của dữ liệu bí mật. Sau đó chuyển dãy bit này sang hệ cơ số 10 gọi là bV

- Nếu $b_{min} \le bV \le d$ thì nhúng được ld bit dữ liệu.

- Nếu $0 \le bV < b_{min}$ hoặc $d < bV \le b_{max}$ thì nhúng được ld - 1 bit dữ liệu.

3. Thuật toán nhúng tin

Đầu vào: Ảnh X, dãy bit dữ liệu $B = (b_1, \cdots, b_s)$

Đầu ra: Ảnh X', X''

Các bước chỉnh:

Bước 1: Liệt kê liên tiếp tất cả các điểm ảnh x với $3 \le x \le 252$, ta được k điểm ảnh cho 1 lần nhúng.

Bước 2: Nhúng ld bit hoặc $ld - 1$ bit của B vào k điểm ảnh theo thuật toán phần III.2

Bước 3: Nếu tất cả các bit đã được nhúng hoặc tất cả các điểm ảnh đã được xét thì dừng thuật toán. Ngược lại chuyển lên bước 1 thì nhúng tiếp.

4. Bài toán tràn số

- Nếu 1 điểm ảnh $x \notin [3,\ 252]$ thì việc nhúng 1 chữ số hệ 13 vào x theo phần III.1 có thể dẫn đến $x', x'' < 0$ hoặc $x', x'' > 252$, tức là ra ngoài điểm ảnh gọi là tràn trên/ tràn dưới.

- Để khắc phục điểm này thì $x' = x'' = x$

5. Phân tích khả năng nhúng ER

Để cho phương pháp có khả năng nhúng nhiều nhất thì ta sẽ tìm giá trị k $(k = 1 \to 20)$ sao cho tỷ lệ nhúng đạt giá trị lớn nhất.

$$ER = \frac{L}{k \times 2} \tag{3}$$

trong đó, L là số bit trung bình (capacity):

$$L = \frac{\alpha \times ld + \beta \times (ld - 1)}{\alpha + \beta} \tag{4}$$

với

$$\begin{cases} \alpha = (d - b_{\min} + 1) \\ \beta = (b_{\max} - d) \end{cases}$$

Sau khi tính toán, phương pháp cho khả năng nhúng nhiều nhất với giá trị $k = 10$ như Bảng II sau:

Bảng II
KẾT QUẢ TỶ LỆ NHÚNG THEO k

k	**Số bit trung bình nhúng được**	ER
1	3.3125	1,6563
2	7.1602	1,7900
3	11.0364	1,8394
4	14.3716	1,7965
5	18.2082	1,8208
6	22.0754	1,8396
7	25.4350	1,8168
8	29.2597	1,8287
9	33.1173	1,8398
10	**37.0015**	**1,8501**
11	40.3150	1,8325
12	44.1622	1,8401
13	48.0380	1,8476
14	51.3743	1,8348
15	55.2103	1,8403
16	59.0772	1,8462
17	62.4379	1,8364
18	66.2620	1,8406
19	70.1191	1,8452
20	74.0031	1,8501

6. Độ đo PSNR

Độ đo $PSNR$ (Peak Signal to Noise Ratio) dùng để đánh giá chất lượng ảnh giữa ảnh A và B cùng kích thước được tính toán như sau:

$$PSNR(A, B) = 10 \log_{10} \left(\frac{255^2}{\text{MSE}(A, B)} \right) \tag{5}$$

trong đó:

$$\text{MSE}(A, B) = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} (a_{i,j} - b_{i,j})^2 \tag{6}$$

ở đây, $a_{i,j}$, $b_{i,j}$ là giá trị điểm ảnh gốc, ảnh giấu tin. M là số hàng và N là số cột của ảnh.

Như vậy, giá trị $PSNR\,(A, B)$ càng lớn thì hai ảnh A và B càng giống nhau.

7. Trích thông tin và khôi phục ảnh gốc

Đầu vào: Ảnh X',X''

Đầu ra: Ảnh X, dữ liệu bí mật B

Các bước chỉnh:

B1. Khôi phục điểm ảnh **x**:

∗ Khôi phục **x** theo công thức (7)

$$\mathbf{x} = \left\lceil \frac{\mathbf{x}' + \mathbf{x}''}{2} \right\rceil \tag{7}$$

∗ Chứng minh công thức (7)

$$\begin{aligned} \left\lceil \frac{x' + x''}{2} \right\rceil &= \left\lceil \frac{x + a + x + b}{2} \right\rceil \\ &= \left\lceil \frac{2x + a + b}{2} \right\rceil \\ &= \left\lceil x + \frac{a + b}{2} \right\rceil \\ &= x + \left\lceil \frac{a + b}{2} \right\rceil \\ &= x \end{aligned}$$

Vì x là số nguyên và $\left\lfloor \frac{a+b}{2} \right\rfloor = 0$ (theo điều kiện 2 công thức 2)

B2: Trích dữ liệu.

Tính giá trị $x' - x''$. Theo Bảng I, ta được giá trị của a và b, sau đó tìm hàng tương ứng, ta được giá trị d, suy ra giá trị đã nhúng.

8. Ví dụ về thuật toán nhúng tin

Ví dụ 1:

Cho khối ảnh X:

$$X = \begin{bmatrix} 12 & 8 & 10 & 20 \\ 6 & 9 & 5 & 55 \\ 10 & 12 & 7 & 85 \end{bmatrix}$$

Với $k = 10$ và dữ liệu bí mật B = [0010 0001 0111 1010 0001 0111 1010 0001 0111 10].

Cần nhúng dãy bit B vào X để có X' và X'':

Bước 1: Tính $d = 13^{10} - 1 = 137858491848$

Bước 2: Đổi d ra nhị phân ta được giá trị nhị phân có độ dài $ld = 38$ bit.

Bước 3: Số nhị phân lớn nhất có độ dài 38 bit là: $b_{max} = 2^{38} - 1 = 274877906943$.

Bước 4: Số nhị phân nhỏ nhất có độ dài 38 bit là: $b_{min} = 2^{37} = 137438953472$.

Bước 5:

Lấy 38 bit trong B là: "0010 0001 0111 1010 0001 0111 1010 0001 0111 10". Chuyển sang hệ 10 được $bV = 35945572446_{(10)}$. Thấy $bV < b_{min}$ nên theo thuật toán III.2 ta sẽ lấy 37 bit trong B là: "0100001011110100001011110100001011110". Đổi dãy bit này sang hệ 13 ta được 3 5 0 11 0 11 5 3 12 $4_{(13)}$. và nhúng tuần tự vào 10 điểm ảnh [12, 8, 10, 20, 6, 9, 5, 55, 10, 12] theo Bảng I.

Áp dung thuật toán theo Bảng I ta được kết quả theo Bảng III như sau:

Bảng III
KẾT QUẢ NHÚNG

x	12	8	10	20	6	9	5	55	10	12
d	3	5	0	11	0	11	5	3	12	4
$x\prime$	10	7	7	22	3	11	4	53	13	11
$x\prime\prime$	13	8	13	17	9	6	5	57	7	13

Từ đó suy ra:

$$X' = \begin{bmatrix} 10 & 7 & 7 & 22 \\ 3 & 11 & 4 & 53 \\ 13 & 11 & 7 & 85 \end{bmatrix}$$

$$X'' = \begin{bmatrix} 13 & 8 & 13 & 17 \\ 9 & 6 & 5 & 57 \\ 7 & 13 & 7 & 85 \end{bmatrix}$$

Tính **ER**: Theo công thức (3-4), được:

$L = 37,0015$

$ER = \frac{37,0015}{10\times 2} = 1,8501$ bpp

Qua đây thấy tỷ lệ nhúng của phương pháp đề xuất là hơn hẳn với $ER = 1,66$ bpp của phương pháp Chen và cộng sự [1].

Tính **PSNR**:

- Sai số bình phương của X' và X'' bằng:

SE(X', X)=4+1+9+4+9+4+1+4+9+1 = 46,

SE(X'', X)=1+ 9+ 9+ 9+ 9+ 4+9 +1 = 51.

- Sai số bình phương trung bình bằng:

MSE(X', X) = SE(X', X)/ (3*4) = 46/12 = 3,83

MSE(X'', X)= SE(X'', X)/(3*4) = 51/12 = 4,25

Cuối cùng:

$$PSNR\,(X', X) = 10\log_{10}\left(\frac{255^2}{3,83}\right) = 42,299$$

$$PSNR\,(X'', X) = 10\log_{10}\left(\frac{255^2}{4,25}\right) = 41,847$$

Các giá trị này lớn hơn giá trị PSNR của Chen và cộng sự.

Ví dụ 2: Cho khối ảnh X như ví dụ 1 với k =10.

- Chọn 10 điểm ảnh liên tiếp từ X ta được: X = [12, 8, 10, 20, 6, 9, 5, 55,10, 12]

- Nhúng dữ liệu: Giả sử dữ liệu cần nhúng là bit "0" (ứng với $d = 0$ trong hệ cơ số 13). Theo Bảng I trong phương pháp đề xuất với $d = 0$ ta có: $X' = x - 3$, $X'' = x + 3$. Áp dụng cho 10 điểm ảnh:

$x'_1 = 12 - 3 = 9$, $x''_1 = 12 + 3 = 15$
$x'_2 = 8 - 3 = 5$, $x''_2 = 8 + 3 = 11$
$x'_3 = 10 - 3 = 7$, $x''_3 = 10 + 3 = 13$
$x'_4 = 20 - 3 = 17$, $x''_4 = 20 + 3 = 23$
$x'_5 = 6 - 3 = 3$, $x''_5 = 6 + 3 = 9$
$x'_6 = 9 - 3 = 6$, $x''_6 = 9 + 3 = 12$
$x'_7 = 5 - 3 = 2$, $x''_7 = 5 + 3 = 8$
$x'_8 = 55 - 3 = 52$, $x''_8 = 55 + 3 = 58$
$x'_9 = 10 - 3 = 7$, $x''_9 = 10 + 3 = 13$
$x'_{10} = 12 - 3 = 9$, $x''_{10} = 12 + 3 = 15$

Vậy ảnh sau khi nhúng là:

$$X' = \begin{bmatrix} 9 & 5 & 7 & 17 \\ 3 & 6 & 2 & 52 \\ 7 & 9 & 7 & 85 \end{bmatrix}$$

$$X'' = \begin{bmatrix} 15 & 11 & 13 & 23 \\ 9 & 12 & 8 & 58 \\ 13 & 15 & 7 & 85 \end{bmatrix}$$

- Khôi phục ảnh gốc: Sử dụng công thức 7 để khôi phục. Với pixel đầu tiên ta có: $X = \left\lceil \frac{9 + 15}{2} \right\rceil = 12$. (Khôi phục chính xác với ảnh gốc).

- Trích xuất bit "0"

Tính hiệu $X' - X'' = -6$. Do đó trong Bảng I thì $d = 0$ (ứng với bit 0).

Kết quả: Dữ liệu trích xuất là bit "0".

- Đánh giá thay đổi:

◦ Độ biến đổi pixel:

Mỗi pixel bị thay đổi tối đa ±3 đơn vị (ví dụ: 12 → 9 và 15). Do đó, đảm bảo chất lượng ảnh (PSNR cao).

◦ Khả năng nhúng:

Mỗi pixel nhúng được 1 chữ số hệ cơ số 13 (trong ví dụ này là $d = 0$).

Tổng tỷ lệ nhúng (ER) đạt ~ 1.8501 bpp (cao hơn phương pháp Chen).

∗ **Ví dụ 3:** Cho khối ảnh X như ví dụ 1 với k =10.

- Chọn 10 điểm ảnh liên tiếp từ X ta được: x = [12, 8, 10, 20, 6, 9, 5, 55, 10, 12]

- Nhúng dữ liệu:

◦ Giả sử dữ liệu cần nhúng là bit "1" (ứng với d = 1 trong hệ cơ số 13).

∘ Theo Bảng I trong phương pháp đề xuất với d = 1 ta có: x' = x - 3 và x'' = x + 2

Áp dụng cho 10 điểm ảnh:

$x'_1 = 12 - 3 = 9,\ x''_1 = 12 + 2 = 14$

$x'_2 = 8 - 3 = 5,\ x''_2 = 8 + 2 = 10$

$x'_3 = 10 - 3 = 7,\ x''_3 = 10 + 2 = 12$

$x'_4 = 20 - 3 = 17,\ x''_4 = 20 + 2 = 22$

$x'_5 = 6 - 3 = 3,\ x''_5 = 6 + 2 = 8$

$x'_6 = 9 - 3 = 6,\ x''_6 = 9 + 2 = 11$

$x'_7 = 5 - 3 = 2,\ x''_7 = 5 + 2 = 7$

$x'_8 = 55 - 3 = 52,\ x''_8 = 55 + 2 = 57$

$x'_9 = 10 - 3 = 7,\ x''_9 = 10 + 2 = 12$

$x'_{10} = 12 - 3 = 9,\ x''_{10} = 12 + 2 = 14$

Vậy ảnh sau khi nhúng là:

$$X' = \begin{bmatrix} 9 & 5 & 7 & 17 \\ 3 & 6 & 2 & 52 \\ 7 & 9 & 7 & 85 \end{bmatrix}$$

$$X'' = \begin{bmatrix} 14 & 10 & 12 & 22 \\ 8 & 11 & 7 & 57 \\ 12 & 14 & 7 & 85 \end{bmatrix}$$

- Khôi phục ảnh gốc:

Sử dụng công thức 7 để khôi phục.

Ví dụ với pixel đầu tiên:

$X = \left\lceil \frac{9 + 14}{2} \right\rceil = 12$ (khôi phục chính xác với ảnh gốc)

- Trích xuất bit "1"

∘ Tính hiệu $X' - X'' = -5$ => Trong Bảng I thì d = 1 (ứng với bit 1)

∘ Kết quả: Dữ liệu trích xuất là bit "1".

- Đánh giá thay đổi:

∘ Độ biến đổi pixel:

Mỗi pixel bị thay đổi tối đa - 3 hoặc + 2 đơn vị (ví dụ:12 → 9 và 14). Do đó, đảm bảo chất lượng ảnh ($PSNR$ cao).

∘ Khả năng nhúng:

Mỗi pixel nhúng được 1 chữ số hệ cơ số 13 (trong ví dụ này là d = 1).

Tổng tỷ lệ nhúng (ER) đạt ~ 1,8501 bpp (cao hơn phương pháp Chen).

IV. KẾT QUẢ

1. So sánh hai phương pháp về mặt lý thuyết

Nhìn vào Bảng IV ta có nhận xét:

- Phương pháp đề xuất vượt trội về ER (1,8501 bpp so với 1,66 bpp) và chất lượng ảnh ($PSNR$ cao hơn) nhờ tận dụng hệ cơ số 13 và giới hạn biến đổi điểm ảnh tối đa 3 đơn vị.

- Phương pháp của Chen có ưu điểm đơn giản nhưng bị hạn chế bởi hệ cơ số 5 và độ biến đổi điểm ảnh lớn.

- Cả hai phương pháp đều đảm bảo tính thuận nghịch (RDH), nhưng phương pháp đề xuất phù hợp hơn cho ứng dụng yêu cầu dung lượng nhúng cao và chất lượng ảnh tốt.

Bảng IV
SO SÁNH KHI NHÚNG BIT 0 VÀ BIT 1

Đặc điểm	Nhúng bit 0 (d = 0)	Nhúng bit 1 (d = 1)
Công thức nhúng	$x' = x - 3$ $x'' = x + 3$	$x' = x - 3$ $x'' = x + 2$
Hiệu $x' - x''$	-6	-5
Độ biến đổi ảnh	± 3	- 3 hoặc + 2

(a) Lena (b) Boat (c) Airplane

(d) Phone (e) Forest (f) Barbara

(g) Milkdrop (h) Cabeza (i) Camera

(j) Car (k) Things (l) Tiffany

Hình 2. Hình ảnh chuẩn hóa dùng để kiểm tra khả năng nhúng thông tin.

2. So sánh hai phương pháp về thực nghiệm

Trong nghiên cứu này, phương pháp đề xuất được đánh giá thông qua 12 ảnh xám chuẩn (512×512), với các chỉ tiêu so sánh gồm: (1) khả năng nhúng dữ liệu và (2) chất lượng ảnh sau nhúng, đối chiếu với kết quả từ phương pháp Chen và cộng sự (2).

Bảng V
SO SÁNH GIỮA PHƯƠNG PHÁP ĐỀ XUẤT VÀ PHƯƠNG PHÁP CHEN VÀ CỘNG SỰ.

Tiêu chí	**Phương pháp Chen và cộng sự [1]**	**Phương pháp đề xuất**
Số tổ hợp nhúng	5 tổ hợp (hệ cơ số 5)	13 tổ hợp (hệ cơ số 13)
Tỷ lệ nhúng (*ER*)	$ER = \frac{\log_2 5+1}{2} = 1,66$ bpp	$ER = \frac{\log_2 13}{2} \approx 1,8501$ bpp
Khả năng nhúng (Cap)	Dung lượng nhúng thấp hơn do giới hạn hệ cơ số 5.	Dung lượng nhúng cao hơn nhờ hệ cơ số 13, tối ưu hóa số tổ hợp.
Độ biến đổi điểm ảnh	Tối đa 5 đơn vị	Tối đa 3 đơn vị
Chất lượng ảnh (PSNR)	PSNR trung bình ~ 40, 72 dB	PSNR trung bình ~ 42, 55 dB
Độ phức tạp tính toán	Sử dụng ma trận EMD và 4 hướng nhúng, phức tạp hơn so với phương pháp đề xuất.	Sử dụng Bảng tra cứu 13 tổ hợp, đơn giản hơn do giảm số hướng biến đổi.
Phạm vi áp dụng	Phục hồi ảnh gốc chính xác nhưng độ biến đổi lớn.	Phục hồi ảnh gốc chính xác với độ biến đổi nhỏ hơn, chất lượng ảnh tốt hơn.
Khả năng phục hồi	Phục hồi ảnh gốc chính xác nhưng độ biến đổi lớn.	Phục hồi ảnh gốc chính xác với độ biến đổi nhỏ hơn, chất lượng ảnh tốt hơn.
Ưu điểm chính	- Kết hợp EMD và 4 hướng nhúng - Đơn giản trong triển khai.	- Tăng số tổ hợp nhúng lên 13 - Cải thiện ER và PSNR - Độ biến đổi điểm ảnh thấp hơn.
Nhược điểm chính	- ER thấp hơn - Độ biến đổi điểm ảnh lớn, ảnh hưởng chất lượng.	Yêu cầu xử lý Bảng tra cứu 13 tổ hợp có thể tăng độ phức tạp so với hệ cơ số 5.

Phương pháp đề xuất của chúng tôi hướng tới mục tiêu khả năng nhúng được nhiều hơn nhưng chất lượng ảnh vẫn cao hơn so với phương pháp Chen và cộng sự [1]. Khi thử nghiệm nhúng chúng tôi sẽ tính khả năng nhúng Cap, giá trị ER , giá trị $PSNR1$ của ảnh chứa tin thứ nhất, giá trị $PSNR2$ của ảnh chứa tin thứ hai (công thức 5-6). và giá trị $PSTB$ là trung bình cộng của $PSNR1$ và $PSNR2$. Kết quả thử nghiệm trình bày trên Bảng VI cho thấy, cả ER và $PSNR$ của phương pháp đề xuất đều lớn hơn đáng kể so với phương pháp của Chen và cộng sự.

Bảng VI
SO SÁNH TỶ LỆ NHÚNG GIỮA CÁC PHƯƠNG PHÁP

Ảnh	Phương pháp	LocSize	Cap	**ER**	PSNR1	PSNR2	**PSTB**
Lena	[1]	0	869250	1,6580	40,7280	40,7109	40,7195
	Đề xuất	0	969961	1,8501	42,5533	42,5574	42,5554
Boat	[1]	1356	867894	1,6554	40,7183	40,7231	40,7207
	Đề xuất	844	969117	1,8484	42,5533	42,5574	42,5554
Airplane	[1]	0	869250	1,6580	40,7327	40,7091	40,7209
	Đề xuất	0	969961	1,8501	42,5533	42,5574	42,5554
Phone	[1]	17166	852084	1,6252	40,7302	40,7369	40,7336
	Đề xuất	7773	962188	1,8352	42,5533	42,5574	42,5554
Forest	[1]	0	869250	1,6580	40,7403	40,7261	40,7332
	Đề xuất	0	969961	1,8501	42,5533	42,5574	42,5554
Barbara	[1]	57	869193	1,6579	40,7378	40,7142	40,7260
	Đề xuất	39	969922	1,8500	42,5533	42,5574	42,5554
Blob	[1]	0	869250	1,6580	40,7229	40,7146	40,7188
	Đề xuất	0	969961	1,8501	42,5533	42,5574	42,5554
Cabeza	[1]	180	869070	1,6576	40,7445	40,7403	40,7424
	Đề xuất	120	969841	1,8498	42,5533	42,5574	42,5554
Camera man	[1]	6164	863086	1,6462	40,7272	40,7176	40,7224
	Đề xuất	3510	966451	1,8434	42,5533	42,5574	42,5554
Car	[1]	0	869250	1,6580	40,7396	40,7225	40,7311
	Đề xuất	0	969961	1,8501	42,5533	42,5574	42,5554
Things	[1]	7369	861881	1,6439	40,7240	40,7421	40,7331
	Đề xuất	4976	964985	1,8406	42,5533	42,5574	42,5554
Tiffany	[1]	19521	849729	1,6207	40,7194	40,7332	40,7263
	Đề xuất	17347	952614	1,8170	42,5533	42,5574	42,5554
Average	[1]			**1,649**			**40,7273**
	Đề xuất			**1,8446**			**42,5554**

V. KẾT LUẬN

Phương pháp Chen và cộng sự [1] sử dụng một số hệ cơ số 5 và ma trận tham chiếu EMD nhằm tăng thêm

khả năng nhúng. Trong bài báo này, chúng tôi đưa ra một phương pháp đề xuất là sử dụng 13 tổ hợp nhúng trên một điểm ảnh (hệ cơ số 13). Nhờ đó, phương pháp cho khả năng nhúng dữ liệu lớn hơn nhưng vẫn đảm bảo chất lượng ảnh chứa tin tốt hơn so với phương pháp Chen và cộng sự. Kết quả thực nghiệm cho thấy, khả năng nhúng dữ liệu của phương pháp đề xuất lớn hơn so với phương pháp hiện có, đồng thời vẫn duy trì được chất lượng ảnh nhúng tin cao.

LỜI CẢM ƠN

"Nghiên cứu này được tài trợ bởi Trường Đại học Giao thông Vận tải (ĐH GTVT) thông qua đề tài mã số T2025-CN-003."

TÀI LIỆU THAM KHẢO

[1] X. Chen and C. Hong, "An efficient dual-image reversible data hiding scheme based on exploiting modification direction," *Journal of Information Security and Applications*, vol. 58, pp. 1–9, 2021.

[2] N. H. Cường, P. M. Thái, M. M. Trừng, T. Hạnh, N. T. Tuyết, and P. V. Ất, "Sử dụng kỹ thuật mã hóa dữ liệu để nâng cao chất lượng ảnh trong giấu tin thuận nghịch," in *Kỷ yếu Hội nghị Khoa học công nghệ Quốc gia lần thứ XVII về Nghiên cứu cơ bản và ứng dụng Công Nghệ thông tin*, pp. 766–775, 2024.

[3] H.-K. Pan, Y.-Y. Chen, and Y.-C. Tseng, "A secure data hiding scheme for two-color image"s," in *Proceedings ISCC 2000. Fifth IEEE Symposium on Computers and Communications*, pp. 750–755, 2000.

[4] D.-C. Wu and W.-H. Tsai, "A stego method for images by pixel value differencing," *Pattern Recognition Letters*, vol. 24, no. 9, pp. 1613–1626, 2003.

[5] X. Zhang and S. Wang, "Efficient steganographic embedding by exploiting modification direction," *IEEE Communications Letters*, vol. 10, no. 11, pp. 781–783, 2006.

[6] J. Mielikainen, "LSB matching revisited," *IEEE Signal Processing Letters*, vol. 13, no. 5, pp. 285–287, 2006.

[7] J. Tian, "Reversible data embedding using a difference expansion," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 8, pp. 890–896, 2003.

[8] A. Alattar, "Reversible watermark using the difference expansion of a generalized integer transform," *IEEE Transactions on Image Processing*, vol. 13, no. 8, pp. 1147–1156, 2004.

[9] D. Thodi and J. Rodríguez, "Expansion embedding techniques for reversible watermarking," *IEEE Transactions on Image Processing*, vol. 16, no. 3, pp. 721–730, 2007.

[10] I. Dragoi and D. Coltuc, "Local-prediction-based difference expansion reversible watermarking," *IEEE Transactions on Image Processing*, vol. 23, no. 4, pp. 1779–1790, 2014.

[11] Z. Ni, Y. Shi, and N. Ansari, "Reversible data hiding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 3, pp. 354–362, 2006.

[12] C. Lin, W. Tai, and C. Chang, "Multilevel reversible data hiding based on histogram modification of difference images," *Pattern Recognition*, vol. 41, pp. 3582–3591, 2008.

[13] Y. Li, C. Yeh, and C. Chang, "Data hiding based on the similarity between neighboring pixels with reversibility," *Digital Signal Processing*, vol. 20, pp. 1116–1128, 2010.

[14] X. Li, B. Yang, and T. Zeng, "Efficient reversible watermarking based on adaptive prediction-error expansion and pixel selection," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3524–3533, 2011.

[15] X. Ma, Z. Pan, S. Hu, and L. Wang, "High-fidelity reversible data hiding scheme based on multi-predictor sorting and selecting mechanism," *Journal of Visual Communication and Image Representation*, vol. 28, pp. 71–82, 2015.

[16] X. Li, J. Li, B. Li, and B. Yang, "High-fidelity reversible data hiding scheme based on pixel-value-ordering and prediction-error expansion," *Signal Processing*, vol. 93, no. 1, pp. 198–205, 2013.

[17] F. Peng, X. Li, and B. Yang, "Improved PVO-based reversible data hiding," *Digital Signal Processing*, vol. 25, pp. 255–265, 2014.

[18] B. Ou, X. Li, Y. Zhao, and R. Ni, "Reversible data hiding using invariant pixel value-ordering and prediction-error expansion," *Signal Processing: Image Communication*, vol. 29, no. 7, pp. 760–772, 2014.

[19] J. Li, Y. Wu, C. Lee, and C. Chang, "Generalized PVO-k embedding technique for reversible data hiding," *International Journal of Network Security*, vol. 20, no. 1, pp. 65–77, 2018.

[20] N. Sao, N. Hoa, and A. P.V., "An effective reversible data hiding method based on pixel-value-ordering," *Journal of Computer Science and Cybernetics*, vol. 36, no. 2, pp. 139–158, 2020.

[21] C. Lee and Y. Huang, "Reversible data hiding scheme based on dual stegano-images using orientation combinations," *Telecommunication Systems*, vol. 52, pp. 2237–2247, 2013.

[22] X. Chen and W. Guo, "Reversible data hiding scheme based on fully exploiting the orientation combinations of dual stego-images," *International Journal of Network Security*, vol. 22, no. 1, pp. 126–135, 2020.

[23] C. Chang, T. D. Kieu, and Y. Chou, "Reversible data hiding scheme using two steganographic images," in *Proceedings of IEEE Region 10 International Conference (TENCON)*, pp. 1–4, 2007.

[24] C. Chang, Y. Chou, and T. D. Kieu, "Information hiding in dual images with reversibility," in *Proceedings of the Third IEEE International Conference on Multimedia and Ubiquitous Engineering*, pp. 145–152, 2009.

[25] C. Chang, T.-C. Lu, G. Horng, Y. Huang, and Y. Hsu, "A high payload embedding scheme using dual stego-images with reversibility," in *Proceedings of the Third International Conference on Information, Communication and Signal Processing*, pp. 1–5, 2013.

Phạm Minh Thái tốt nghiệp đại học chuyên ngành Công nghệ thông tin tại Trường Đại học Nha Trang vào năm 2008. Ông nhận bằng Thạc sĩ chuyên ngành Khoa học máy tính tại học viện Kỹ thuật Quân sự năm 2011, bằng Cử nhân Tiếng Anh tại Đại học Thái Nguyên vào năm 2023 và đang làm nghiên cứu sinh chuyên ngành Công nghệ thông tin tại trường Đại học Giao thông vận tải. Hiện công tác tại khoa công nghệ thông tin, Trường Đại học Kinh tế - Kỹ thuật Công nghiệp. Các lĩnh vực quan tâm của ông bao gồm giấu tin và an ninh thông tin.
Email: pmthai@uneti.edu.vn

Lê Kinh Tài nhận bằng Cử nhân từ Học viện Tài chính vào năm 2006, bằng Thạc sĩ Quản trị Nguồn nhân lực từ Trường Đại học Công đoàn vào năm 2012, và bằng Cử nhân Tiếng Anh từ Trường Đại học Mở Hà Nội vào năm 2014. Hiện tại, ông đang công tác tại Đại học Công đoàn. Các lĩnh vực quan tâm của ông bao gồm giấu tin và an ninh thông tin.
Email: tailk@dhcd.edu.vn

Nguyễn Hiếu Cường nhận bằng Đại học và Thạc sỹ ngành Đảm bảo toán học cho máy tính và các hệ thống tính toán tại Trường Đại học Khoa học Tự nhiên (Đại học quốc gia Hà Nội) tương ứng vào năm 1995 và 1999. Ông nhận bằng Tiến sĩ ngành Tin học tại trường Đại học Công nghệ Darmstadt (CHLB Đức) năm 2013. Các lĩnh vực quan tâm của ông bao gồm: giấu tin và phát hiện ảnh giả.
Email: cuongnh@utc.edu.vn

Nguyễn Quang Chánh nhận bằng Đại học chuyên ngành Toán-Tin tại Trường Đại học Sư phạm Hà Nội 2 vào năm 2003. Ông nhận bằng Thạc sí chuyên ngành công nghệ thông tin tại trường Đại học Sư phạm Kỹ Thuật Hưng Yên vào năm 2022. Các lĩnh vực quan tâm của ông bao gồm: bảo mật và an ninh mạng.
Email: nqchanh@uneti.edu.vn

Phương pháp xử lý tín hiệu và tạo ảnh 3D cho các vật thể dưới nước bằng công nghệ Sonar chủ động

Nguyễn Văn Đức, Nguyễn Quốc Khương
Trường Điện Điện tử, Đại học Bách Khoa Hà Nội
Tác giả liên hệ: Nguyễn Văn Đức. Email: duc.nguyenvan1@hust.edu.vn
Ngày nhận bài: 10/05/2024, ngày sửa chữa: 10/03/2025, ngày duyệt đăng: 22/05/2025
DOI: 10.32913/mic-ict-research-vn.v2025.n1.1273

Tóm tắt: Bài báo đề xuất phương pháp xử lý tín hiệu và tạo ảnh 3D của các vật thể dưới nước bằng cách sử dụng công nghệ nhận dạng và định vị bằng sóng âm Sonar. Phương pháp tạo ảnh 3D của các vật thể dưới nước đề xuất trong bài báo được thực hiện bằng cách sử dụng nhiều đầu thu tín hiệu sóng âm (cũng được gọi là ăng ten thu tín hiệu sóng âm), và một số đầu phát tín hiệu sóng âm (cũng được gọi là ăng ten phát tín hiệu sóng âm) nhằm mục đích thu thập dữ liệu để tạo ảnh Sonar 3D có độ phân giải lớn. Dựa trên dữ liệu thu được từ hệ ăng ten đa điểm thu ở dưới nước, thuật toán tạo ảnh Sonar 3D sẽ tính toán giá trị ma trận điểm ảnh trên cơ sở độ mở búp sóng quét của các ăng ten phát tín hiệu sóng âm, khoảng cách điểm phản xạ phát hiện được, tốc độ di chuyển của tàu, v.v. Sau khi xác định được giá trị ma trận ảnh, bức ảnh của vật thể dưới nước được dựng trong không gian 3D. Các kết quả thực nghiệm được thực hiện ở một số vùng hồ nước sâu như hồ Hòa Bình, hồ Đồng Đò và vịnh Hạ Long.

Từ khóa: *công nghệ định vị sóng âm - Sonar, hệ ăng ten mảng, công nghệ quét ảnh vật thể dưới nước và dựng lên trong không gian ba chiều (3D).*

Title: **Method for Processing Signals and Creating 3D Images for Underwater Objects using Active Sonar Technology**

Abstract: The article proposes a method of 3D image signal processing for underwater objects by using sonar positioning technology. The proposed method is based on the acoustic signal collected from many receivers. The 3D imaging algorithm will calculate the pixel matrix value based on the sweep beam aperture of the transmit transducers, the detected distance from the sonar scanner to the reflected object, the speed of ship movement, etc. After determining the image matrix value, the scanned image is reconstructed on the 3D space. The experimental results were carried out in some deep lake areas such as Hoa Binh, Dong Do lake and Ha Long bay.

Keywords: *active sonar systems, antenna array technology, 3D scanning image.*

I. GIỚI THIỆU

Công nghệ định vị bằng sóng âm Sonar là một công nghệ sử dụng sự lan truyền sóng âm trong môi trường nước để xác định và định vị các vật cản trong môi trường nước, hoặc đáy nước. Công nghệ định vị bằng sóng âm Sonar được phân loại thành hai loại công nghệ cơ bản:

- Công nghệ Sonar chủ động có hệ thống tự phát xung sóng và nghe tiếng vọng lại [1–3].

- Công nghệ Sonar bị động (hay Sonar thụ động) có hệ thống chỉ nghe âm thanh do tàu bè hay nguồn âm khác phát ra; tầm hoạt động nhỏ hơn so với Sonar chủ động và ứng dụng cho do thám ngầm dưới biển [4]. Các công nghệ Sonar chủ động và thụ động đều được thiết kế ứng dụng cho điều khiển lái các tàu ngầm và tàu tự hành dưới nước AUV (Autonomous Underwater Vehicles) [5, 6].

Trong bài báo này, nhóm tác giả không đi sâu vào việc trình bày kiến trúc phần cứng để thu nhận được tín hiệu phản xạ từ nhiều đầu sóng âm. Bản chất một kiến trúc phần cứng thực hiện nhiệm vụ này được nhóm tác giả công bố trong công trình khoa học [7–9]. Tín hiệu nhận được từ các ăng ten thu sóng âm được lọc nhiễu ngoài băng thông qua các bộ lọc có kiến trúc FIR. Để độc giả dễ theo dõi nội dung bài báo, nhóm tác giả chỉ tập trung mô tả thuật toán dựng ảnh Sonar 3D dựa trên các dữ liệu đã được đo đạc trên nền tảng phần cứng đã được thiết kế trước. Dữ liệu nhận được từ nền tảng phần cứng là thông tin về khoảng

cách từ các điểm phản xạ tới hệ ăng ten thu. Nội dung thông tin này chưa đủ thông tin để dựng được bức ảnh quét 3D. Để tạo được bức ảnh quét dưới nước, chúng ta cần xây dựng ma trận điểm ảnh dựa vào đủ số lượng thông tin về bức ảnh quét thu được, các trọng số điểm ảnh được tính toán dựa trên cơ sở các búp sóng quét của các ăng ten thu và phát, cường độ tín hiệu thu được, cũng như là vận tốc chuyển động của tàu gắn thiết bị Sonar. Công trình khoa học [10] trình bày về phương pháp ước lượng độ mở búp sóng của các ăng ten thu phát tín hiệu. Bài báo này đề xuất cách tính các trọng số điểm ảnh dựa vào các tham số đề cập ở trên, cũng như phương pháp dựng lên bức ảnh Sonar 3D.

Trước khi đi vào chi tiết thuật toán tạo ảnh 3D cho vật thể ở dưới nước, bài báo đánh giá tổng quan tình hình nghiên cứu hiện nay trên thế giới về lĩnh vực này.

1. Tình hình nghiên cứu hiện nay trên thế giới

Bài báo [11] trình bày phương pháp dự đoán và nội suy hình ảnh quét Sonar 3D cho các thiết bị tự hành dưới nước. Phán đoán hình ảnh là một công nghệ phức tạp đòi hỏi các thông tin dự đoán về môi trường truyền dẫn và phụ thuộc vào điểm quan sát vật thể. Trên thực tế người ta ít sử dụng công nghệ phán đoán hình ảnh Sonar, mà tìm cách phân tích dữ liệu thu thập được để có dữ liệu chính xác nhất có thể để xử lý và nội suy thành các bức ảnh Sonar. Tập thể tác giả trong công trình khoa học [12] trình bày phương pháp tái tạo ảnh Sonar 3D theo phương pháp nội suy hỗn hợp. Nhược điểm của phương pháp này là độ phức tạp của thuật toán, đặc biệt khi kích cỡ ma trận điểm ảnh tăng. Do vậy khó áp dụng trong bài toán xử lý tín hiệu trong thời gian thực. Công trình khoa học [13] trình bày phương pháp dựng ảnh Sonar 3D sử dụng mạng trí tuệ nhân tạo. Đây là hướng nghiên cứu mới, tuy nhiên để có thể có chất lượng bức ảnh tốt, nguồn dữ liệu đầu vào cần đảm bảo chất lượng và đủ lớn. Bên cạnh đó, tốc độ học máy không bắt kịp sự thay đổi của bức ảnh và môi trường. Do vậy, phương pháp này cho đến nay cũng mới chỉ dừng lại ở nghiên cứu lý thuyết mà chưa có ứng dụng trong thực tiễn.

Sáng chế US3500302 [14], "Sonar bathymetry system transmit-receive sequence programmer" của Hoa Kỳ công bố ngày 13.01.1969 đưa ra phương pháp xác định độ sâu đáy nước và thực hiện trên mạch cứng, không đề cập đến phương pháp quét ảnh Sonar 3D.

Tập thể tác giả [15] đưa ra phương pháp tạo ảnh Sonar 3D, tuy nhiên không đề cập đến các tham số búp sóng quét, vận tốc của tàu, đây là các tham số quan trọng tạo nên bức ảnh quét, cũng như các kết quả trình bày dưới dạng mô phỏng, chưa đưa ra các kết quả thực hiện bằng thực tế. Do vậy, chưa thể có đánh giá được đúng độ phức tạp của hệ thống để có thể triển khai trên các nền tảng phần cứng trong điều kiện thời gian thực.

Bài báo [16] trình bày phương pháp tạo ảnh từ hệ Sonar quét sườn. Bài báo trình bày mô hình toán học để nội suy và tổng hợp ảnh 3D Sonar quét sườn. Các yếu tố thực tế như vận tốc tàu di chuyển, phương pháp dựng ảnh từ hai nửa quét sườn, chiều sâu và chiều rộng bức ảnh quét không được mô tả trong thuật toán.

Bài báo [17] trình bày về phương pháp ước lượng và bù Doppler cho hệ thống thông tin thủy âm. Kết quả bù Doppler cho hệ thống thông tin có thể ứng dụng để cải thiện tín hiệu thu của hệ thống Sonar chủ động, hoặc sử dụng để ước lượng tốc độ chuyển động tương đối giữa hệ Sonar chủ động và đối tượng cần quét ảnh.

Bài báo [18] trình bày phương pháp dựng ảnh Sonar 3D dựa trên mô hình hình học. Để giảm thiểu sai số đo đạc từ các tia phản xạ đáy, bài báo đề xuất ứng dụng bộ lọc thích ứng để bám đuổi và phân tách thông tin phản xạ đáy của chùm tia phản xạ đa đường. Bài toán xây dựng các bộ lọc thích ứng để giảm thiểu nhiễu phân tập kênh đa đường được cho là hiệu quả, tuy nhiên kênh thủy âm là loại kênh biến thiên nhanh, dẫn đến khó bám đuổi các thông tin về kênh truyền. Trong bài báo cũng không đề cập đến vận tốc chuyển động của tàu là tham số quan trọng để xây dựng bộ lọc thích ứng.

Công trình khoa học [19] trình bày phương pháp thiết kế hệ Sonar chủ động sử dụng hệ ăng ten mảng là các hydrophone hình cầu, kèm các kết quả mô phỏng. Bài báo chưa đề cập đến phương pháp dựng ảnh Sonar, các kết quả dừng lại ở các mô phỏng bằng máy tính.

Công bố khoa học của nhóm tác giả [10] trình bày các kết quả đánh giá độ mở búp sóng quét của hệ Sonar quét sườn, với các kịch bản, thiết kế hệ thống là các kết quả được trình bày trong bài báo này.

2. Phương pháp đề xuất

Mục đích của bài báo này nhằm khắc phục nhược điểm đã đề cập ở trên bằng cách đề xuất phương pháp quét ảnh 3D sử dụng công nghệ nhận dạng và định vị bằng sóng âm của các vật thể dưới nước. Phương pháp tạo ảnh 3D của các vật thể dưới nước được thực hiện bằng cách sử dụng nhiều đầu thu tín hiệu sóng âm, và một số đầu phát tín hiệu sóng âm nhằm mục đích thu thập dữ liệu để tạo ảnh 3D có độ phân giải lớn.

Bài báo đề xuất phương pháp tạo ảnh siêu âm 3D bằng cách sử dụng từ 1 đến nhiều đầu phát tín hiệu sóng âm với búp sóng định hướng, mỗi đầu phát tín hiệu sóng âm sẽ sử dụng một tần số riêng biệt, và được phát với một búp sóng định hướng nhất định để quét tín hiệu hướng tới một phạm vi diện tích của bức ảnh siêu âm 3D cần quét.

Phương pháp đề xuất sử dụng nhiều đầu thu tín hiệu sóng âm được sắp xếp một cách phù hợp về khoảng cách để các tín hiệu thu được từ mỗi đầu thu tín hiệu sóng âm tương đối độc lập với nhau. Tín hiệu thu được từ mỗi đầu thu tín hiệu sóng âm sau đó được khuếch đại và lọc nhiễu phục vụ cho việc xử lý các bức ảnh siêu âm 3D.

Để tạo được ảnh Sonar 3D, hệ thống ăng ten phát (transducer phát) và ăng ten thu (đầu thu tín hiệu sóng âm thu) sẽ được dịch chuyển tịnh tiến theo hướng tàu chạy. Sau mỗi một chu kỳ xử lý tín hiệu thì 01 ảnh 2D (ảnh không gian hai chiều) được dựng lên, tuy nhiên hệ thống xử lý liên tiếp các bước ảnh 2D dịch chuyển theo hướng tàu, do vậy sẽ tạo được ảnh 3D (ảnh không gian ba chiều) theo thời gian thực.

Các tín hiệu xung phản xạ từ vật thể cần quét ảnh được so sánh với vị trí xung sóng âm gốc gửi trực tiếp từ khối phát tín hiệu đến khối xử lý tín hiệu thu. Thông qua việc so sánh này, khoảng cách từ điểm quan sát đến vật cản được xác định.

Dựa vào nguồn dữ liệu là các xung phản xạ, bài báo đề xuất xây dựng ma trận điểm ảnh với kích thước dựa trên độ phân giải của bức ảnh ở trục chiều sâu của nước, và trục ngang của lát cắt ảnh 3D vuông góc với phương chuyển động của tàu.

Để cải thiện nguồn dữ liệu thu được, các tín hiệu thu được từ các đầu thu tín hiệu sóng âm cần thực hiện lọc nhiễu bằng phương pháp lọc song song trên mỗi băng con của từng đầu thu tín hiệu sóng âm [9]. Các giá trị điểm ảnh của ma trận tạo ảnh được xác định thông qua hệ số định hướng của búp sóng phát và búp sóng thu, cùng với hệ số suy hao của kênh truyền phụ thuộc với khoảng cách.

Phần còn lại của bài báo này được trình bày như sau: Mục II trình bày chi tiết về thuật toán tạo ảnh Sonar 3D đề xuất trong bài báo. Mục III thảo luận về kết quả đo đạc. Mục IV kết luận bài báo.

II. ĐỀ XUẤT PHƯƠNG PHÁP DỰNG ẢNH SONAR 3D CHO HỆ THỐNG SONAR CHỦ ĐỘNG

Chất lượng và độ chính xác của bức ảnh quét Sonar phụ thuộc nhiều vào chất lượng tín hiệu thu được. Trong bài báo này, tín hiệu thu được đã được xử lý lọc nhiễu như trình bày trong các tài liệu [9]. Độ phân giải của bức ảnh quét phụ thuộc vào các tham số của thuật toán dựng ảnh Sonar 3D như thảo luận dưới đây.

Sơ đồ thiết kế hệ thống quét ảnh siêu âm 3D được mô tả như mô tả ở Hình 1, bao gồm khối nguồn (1), khối phát tín hiệu (2) và khối thu và xử lý tín hiệu (3). Thuật toán tạo ảnh 3D (5) nhận tín hiệu thu từ khối thu (3), cùng thông tin về vận tốc chuyển động của tàu gắn hệ thống quét ảnh siêu âm để tạo ra ma trận điểm ảnh 3D. Thông tin về vận tốc của tàu được xác định bằng khối (4). Vận tốc tàu mang hệ Sonar chủ động có thể xác định bằng GPS, hoặc tốc độ kế trên tàu. Trong trường hợp đặc biệt có thể sử dụng các hệ ăng ten đa búp sóng để xác định vận tốc dựa theo hiệu ứng Doppler. Trong khuôn khổ bài báo này chúng tôi không thảo luận sâu về các phương pháp xác định vận tốc tàu mang hệ Sonar chủ động.

Hình 2 trình bày sơ đồ vị trí các đầu phát tín hiệu sóng âm và đầu thu tín hiệu sóng âm thu sử dụng để tạo ảnh siêu âm 3D. Trong Hình 2 trục tọa độ x là bề rộng búp sóng quét theo mặt cắt ngang, z trục tọa độ thể hiện độ sâu của đáy biển, y là trục chỉ hướng tàu chuyển động. Số lượng đầu thu tín hiệu sóng âm được sử dụng là M , trong khi đó số lượng đầu phát tín hiệu sóng âm được dùng là 2 (trong trường hợp tổng quát là N, tuy nhiên được mô tả trong bài báo này là 2 đầu phát tín hiệu sóng âm để việc mô tả toán học cho đơn giản). Các đầu phát tín hiệu sóng âm được hướng ra hai nửa của mặt phẳng đáy và được kéo theo di chuyển của tàu. Đầu phát tín hiệu sóng âm thứ nhất ký hiệu là Tx_1, phát xung siêu âm $S_1(t)$ với tần số dao động f_1. Đầu phát tín hiệu sóng âm thứ hai ký hiệu là Tx_2, phát xung siêu âm $S_2(t)$ với tần số dao động f_2. Số đầu phát tín hiệu sóng âm có thể được nâng cao để nâng cao chất lượng ảnh quét.

Hình 3 trình bày thuật toán xác định độ sâu đáy biển là một tham số để tạo được ảnh quét siêu âm 3D, trong đó T_h là mức ngưỡng của tín hiệu dùng để nhận biết có sự phản xạ của tín hiệu đáy. Mức ngưỡng tín hiệu thường được xác định cao hơn mức nhiễu nền. Để xác định được mức độ can nhiễu nền, hệ thống trước đó cần có phép đo hoặc ước lượng mức độ can nhiễu nền độc lập với phép xử lý tín hiệu Sonar [20]. Trong khuôn khổ bài báo này chúng tôi không tham vọng đi sâu vào vấn đề này. Tham số Do_sau là tham số sử dụng để xác định độ sâu có thể đo được.

Thuật toán tạo ảnh 3D, được mô tả chi tiết sau đây. Về mặt toán học, tín hiệu thu được tại mỗi đầu thu tín hiệu sóng âm được biểu diễn dưới dạng:

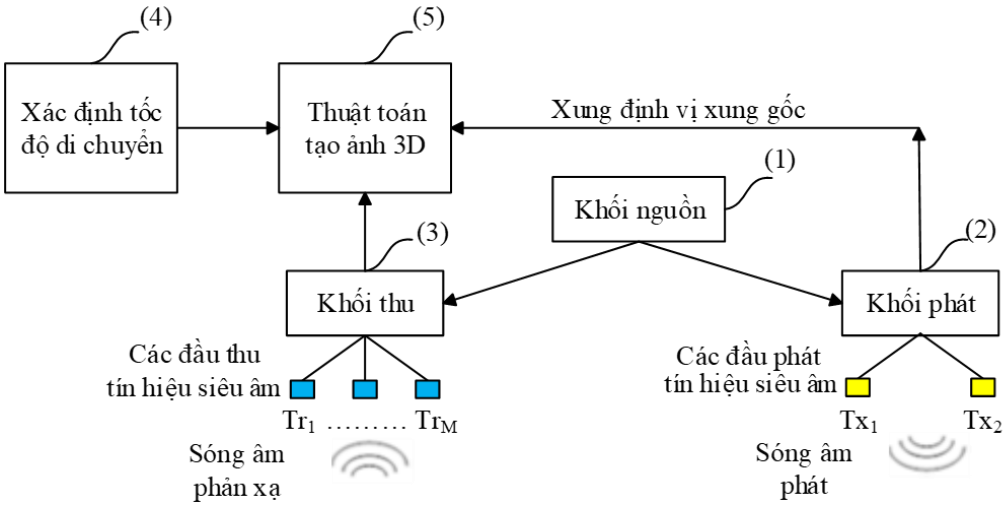
$$x_m(t) = A_m \cos\left(2\pi f_m t + \phi_m\right) \tag{1}$$

trong đó, ϕ_m và f_m là góc pha của tín hiệu tới và tần số dao động của tín hiệu. Sóng mang được ký hiệu là $f_c(t) = \cos(2\pi f_c t)$. Khi nhân tín hiệu thu được với sóng mang thu được:

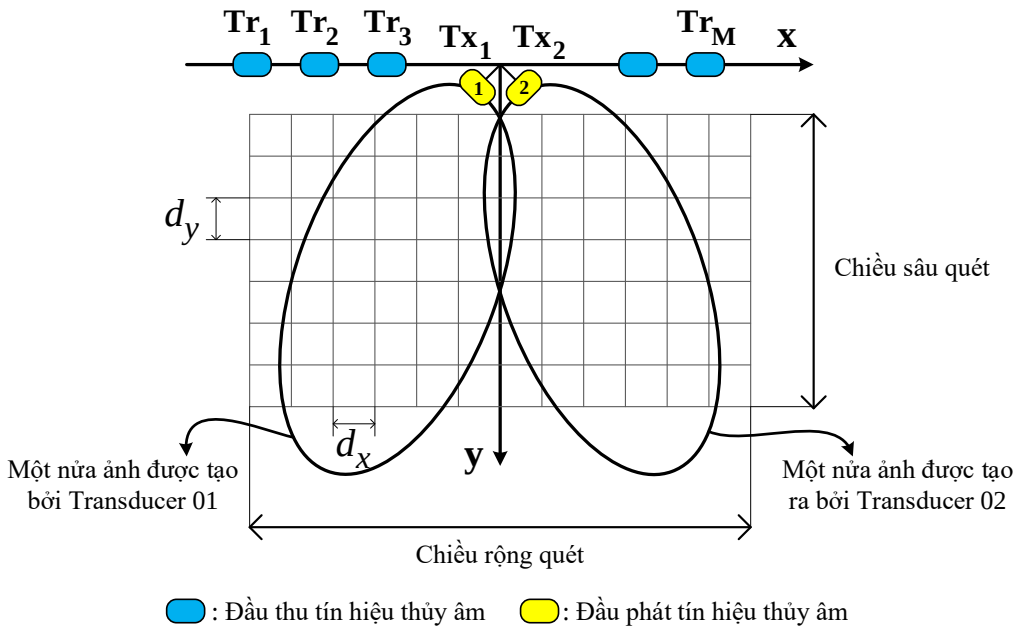
$$f_y(t) = x_m(t) f_c(t) = A_m \cos\left(2\pi f_m t + \phi_m\right) \cos\left(2\pi f_c t\right) \tag{2}$$

Sau khi triển khai giải tích ta thu được:

$$\begin{aligned} f_y(t) = \frac{A_m}{2}\{&\cos\left[2\pi\left(f_m + f_c\right)t + \phi_m\right] \\ &+ \cos\left[2\pi\left(f_c - f_m\right)t + \phi_m\right]\} \end{aligned} \tag{3}$$



Hình 1. Sơ đồ thiết kế hệ quét ảnh Sonar chủ động.



Hình 2. Sơ đồ bố trí các ăng ten thu phát sóng âm và tạo búp sóng quét.

Tín hiệu thể hiện ở công thức (3) có chứa thành phần tần số cao là tổng của hai tần số và thành phần tần số thấp là hiệu của 2 tần số. Sau đó, sử dụng bộ lọc thông thấp để loại bỏ thành phần tần số cao. Như vậy, tín hiệu thu được là thành phần hiệu của 2 tần số

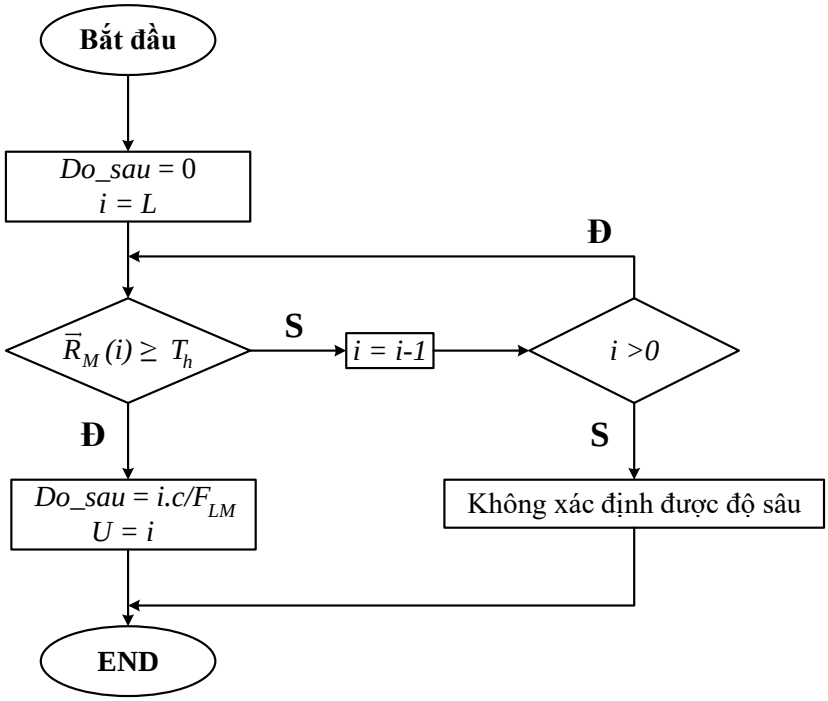
$$f_y(t) = \cos\left[2\pi\left(f_c - f_m\right)t + \phi_m\right] \quad (4)$$

Tín hiệu ở công thức (4) được lấy mẫu qua bộ biến đổi A/D. Dựa trên tín hiệu xung gốc được gửi từ bên phát, bài báo đề xuất tách tín hiệu xung gốc và giữ lại tín hiệu xung phản xạ từ tín hiệu thu được. Sau đó, thuật toán tạo ảnh siêu âm 3D sẽ lọc tách riêng từng kênh rồi xây dựng ma trận tín hiệu đầu vào $\mathbf{R_1}$ và $\mathbf{R_2}$ gồm M hàng (tương ứng M đầu thu) và L cột (tương ứng với N mẫu tín hiệu thu). Ma trận tín hiệu đầu vào $\mathbf{R_1}$,$\mathbf{R_2}$ sau đó được đưa vào thuật toán tạo ảnh.

Thuật toán dựng ảnh Sonar 3D được mô tả chi tiết như sau:

B1: Thiết lập các thông số hệ thống bao gồm label=(b)

1) M : là tham số xác định số lượng đầu đầu thu tín



Hình 3. Thuật toán xác định độ sâu để tạo ảnh Sonar 2D hoặc 3D.

hiệu sóng âm.

2) $R_i(x, y)$: xác định tọa độ các đầu thu tín hiệu sóng âm thu được $i = 1 : M$.
3) $T_1(x_1, y_1), T_2(x_2, y_2)$: xác định tọa độ của 2 đầu phát tín hiệu sóng âm phát.
4) d_x, d_y lần lượt xác định độ phân giải của điểm ảnh theo trục x và y.
5) X_{max} xác định bề rộng quét lớn nhất của vùng quét.
6) Y_{max} xác định chiều sâu quét lớn nhất của vùng quét.
7) Số điểm ảnh lớn nhất tính theo chiều ngang được xác định là:

$$N_{x\_max} = X_{max}/d_x \tag{5}$$

8) Độ phân giải lớn nhất theo chiều sâu được xác định là:

$$N_{y\_max} = Y_{max}/d_y \tag{6}$$

9) $\mathbf{W_i}$ là ma trận tọa độ các các điểm ảnh kích thước $[N_{x\_max} \times N_{y\_max}]$ ứng với đầu thu tín hiệu sóng âm thứ i.
10) x, y là tọa độ của mỗi điểm trên mặt phẳng ảnh.
11) $P_1(x, y), P_2(x, y)$ là hàm trọng số búp sóng tín hiệu phát lần lượt ứng với đầu phát tín hiệu sóng âm phát ở tần số f_1 và f_2.
12) $Q_i(x, y)$ là hàm trọng số búp sóng tín hiệu thu của các đầu thu tín hiệu sóng âm thu $i = 1 : M$.
13) $d_{Tx1}(x, y)$ là ma trận khoảng cách từ đầu phát tín hiệu sóng âm thứ nhất tới các điểm ảnh của ma trận $\mathbf{W_i}$.
14) $d_{Tx2}(x, y)$ là ma trận khoảng cách từ đầu phát tín hiệu sóng âm thứ hai tới các điểm ảnh của ma trận $\mathbf{W_i}$.
15) $d_{Tri}(x, y)$ là ma trận khoảng cách từ đầu thu tín hiệu sóng âm thu thứ i tới các điểm ảnh của ma trận $\mathbf{W_i}$
16) $d_{i,1}(x, y)$ là ma trận quãng đường truyền tín hiệu từ đầu phát tín hiệu sóng âm phát 1 tới đầu thu tín hiệu sóng âm thu thứ i, và được xác định bằng công thức

$$d_{i,1}(x, y) = d_{Tx1}(x, y) + d_{Tri}(x, y) \tag{7}$$

17) $d_{i,2}(x, y)$ là ma trận quãng đường truyền tín hiệu từ Đầu phát tín hiệu sóng âm phát 2 tới đầu thu tín hiệu sóng âm thu thứ i.

$$d_{i,2}(x, y) = d_{Tx2}(x, y) + d_{Tri}(x, y) \tag{8}$$

18) $t_{i,1}$ là thời gian truyền tín hiệu từ đầu phát tín hiệu sóng âm phát 1 đến đầu thu tín hiệu sóng âm thứ i được xác định bằng:

$$t_{i,1} = d_{i,1}(x, y)/c \tag{9}$$

trong đó c là vận tốc sóng âm trong môi trường nước.

19) $t_{i,2}$ là thời gian truyền tín hiệu từ đầu phát tín hiệu sóng âm phát 2 đến đầu thu tín hiệu sóng âm thứ i được xác định bằng:

$$t_{i,2} = d_{i,2}(x,y)/c \tag{10}$$

20) Khởi tạo ma trận giá trị trọng số cho các điểm ảnh có tọa độ (x,y)

$$\mathbf{W_i} = \begin{cases} P_1(x,y)Q_i(x,y)/\left(d_{i,1}\right)^2 & x = -N_{x\_\max}/2 : 0 \\ P_2(x,y)Q_i(x,y)/\left(d_{i,2}\right)^2 & x = 0 : +N_{x\_\max}/2 \end{cases} \tag{11}$$

21) T_h là mức ngưỡng xác nhận có tín hiệu phản xạ. Nếu tín hiệu nhỏ hơn mức này là nhiễu.
22) F_{LM} là tần số lấy mẫu tín hiệu khi biến đổi A/D.
23) v là tốc độ di chuyển của tàu.
24) Khởi tạo $k = 1$ là chỉ số xác định số thứ tự của ảnh thứ nhất.

B2: Thu tín hiệu thu từ mạch biến đổi A/D từ các đầu thu tín hiệu sóng âm.

B3: Tín hiệu này sẽ được lọc thông thấp có băng thông bằng $1/\Delta t$. Tiếp đó, xác định giá trị cực đại của tín hiệu thu để xác định vị trí xung gốc. Sau đó, cắt bỏ phần tín hiệu có độ dài bằng $(t + \Delta t)$ tính từ giá trị lớn nhất để loại bỏ phần nhiễu, phần tín hiệu còn lại là các xung phản xạ thu được.

B4: Tách sóng để tách riêng từng kênh. Vì hệ thống sử dụng 2 tần số phát khác nhau nên số kênh tách ra tương đương với 2 lần số đầu thu M. Dựa trên xung xung gốc $g(t)$ truyền từ bên phát. Ghép các tín hiệu thu lại ta được hai ma trận tín hiệu thu $\mathbf{R_1}$ và $\mathbf{R_2}$ kích thước $M \times L$ (M hàng L cột) M là số đầu đầu thu tín hiệu sóng âm thu. Số mẫu tín hiệu biểu diễn trong các ma trận $\mathbf{R_1}$ và $\mathbf{R_2}$ được xác định là:

$$L = F_{LM} \times (T - \Delta t) \tag{12}$$

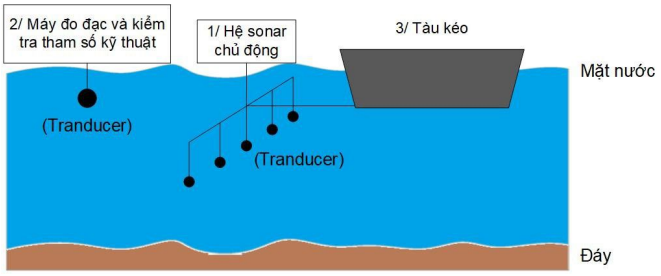
Các ma trận $\mathbf{R_1}$ và $\mathbf{R_2}$ tương ứng với các tần số phát f_1 và f_2 sẽ được sử dụng để tạo ra một nửa ảnh như Hình 2.

B5: Tiếp tục với ảnh thứ $k + 1$ bằng cách quay lại bước 2.

III. KẾT QUẢ THỰC NGHIỆM VÀ THẢO LUẬN

Kịch bản thử nghiệm hệ thống được trình bày ở **hình 4**, trong đó đó tàu thử nghiệm được sử dụng để gắn các trang bị kỹ thuật, với ký hiệu (1) là hệ Sonar chủ động, (2) là máy đo đạc và kiểm tra tham số kỹ thuật, và (3) là tàu kéo hệ thống. Transducer là các đầu thu phát tín hiệu sóng âm.

Hệ thống được thiết kế thực tế sử dụng hai búp sóng quét sườn đặt hai mạn tàu với tần số quét là 165 KHz và 200 Khz. Hệ thống được thiết kế sử dụng 8 đầu thu tín hiệu sóng âm song song, được phân bố đối xứng với 4 đầu thu



Hình 4. Kịch bản thử nghiệm hệ Sonar chủ động.

Bảng I
CÁC THAM SỐ HỆ THỐNG.

Tham số	Giá trị
Chu kỳ phát xung Sonar T	0.1s
Độ rộng một xung Sonar Δt	0.4ms
Tần số lấy mẫu tại phía thu	192kHz
Số lượng transducer phát M	2
Số lượng hydrophone thu N	8
Khoảng cách thiết lập giữa hai Hydrophone liền kề L	$0.09m$
Tx_1 Tần số phát của transducer thứ nhất f_1	165kHz
Tx_2 Tần số phát của transducer thứ 2 f_2	200kHz

được sử dụng cho quét một bên sườn (một bên của mạn tàu). Thiết bị thử nghiệm Sonar quét sườn với kiến trúc 2 đầu phát và 8 đầu thu tín hiệu sóng âm.

Tham số hệ thống thử nghiệm được trình bày vắn tắt như ở Bảng I, trong đó đáng chú ý là các tần số của transducer phát, khoảng cách đặt của các hydrophone thu, chu kỳ phát tín hiệu Sonar và độ rộng một xung Sonar. Chúng ta chú ý là xung Sonar được điều chế ở các tần số f_1 và f_2. Điều kiện thời tiết trong lúc thử nghiệm được mô tả ở Bảng II. Nhóm nghiên cứu không tiến hành đo đặc mức nhiễu, tuy nhiên với điều kiện này về cơ bản điều kiện thời tiết là thuận lợi cho thử nghiệm. Trong giới hạn các nghiên cứu này, bài báo chỉ tập trung vào sự ảnh hưởng các tham số kỹ thuật đến chất lượng hệ thống mà chưa tập trung vào các nghiên cứu về sự ảnh hưởng các điều kiện tự nhiên như sóng, gió hay bão. Các chấn tử ăng ten phát là transducer BII-7562/200 [21] là họ transducer định hướng có tính định hướng khá tốt. Độ mở búp sóng là 7^0 được sử dụng trong công thức (11).

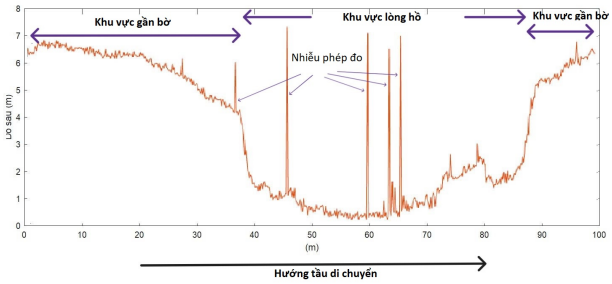
Hình ảnh thử nghiệm quét ảnh đáy hồ Đồng Đò được trình bày ở Hình 5, trong đó kết quả thu được bằng cách sử

Bảng II
CÁC ĐIỀU KIỆN MÔI TRƯỜNG

Thời tiết	trời không nắng, không mưa
Nhiệt độ	$25°C$
Gió	gió nhẹ
Sóng	sóng lăn tăn

Hình 5. Hình ảnh thử nghiệm hệ Sonar chủ động quét hình ảnh đáy 2D tại hồ Đồng Đò ngày 16 tháng 10 năm 2021

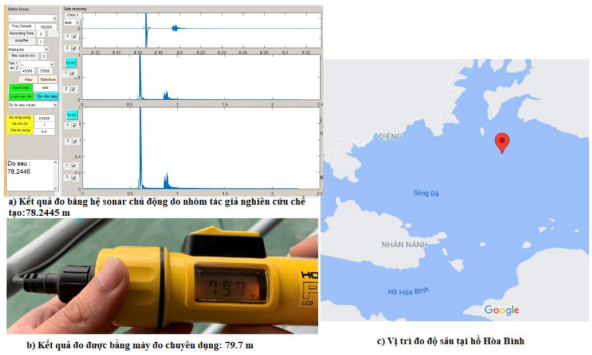
dụng phương pháp được so sánh trực tiếp trong thời gian thực với hệ thống Sonar chuyên dụng. Các kết quả so với hệ Sonar chuyên dụng cho thấy không có sự khác biệt đáng kể giữa hai hệ thống. Hệ thống thử nghiệm được di chuyển theo một mặt cắt ngang của hồ Đồng Đò, tức thuyền di chuyển từ một bờ phía bên này sang bờ phía bên kia.



Hình 6. Kết quả hình ảnh quét đáy ảnh 2D hồ Đồng Đò ngày 16 tháng 10 năm 2021.

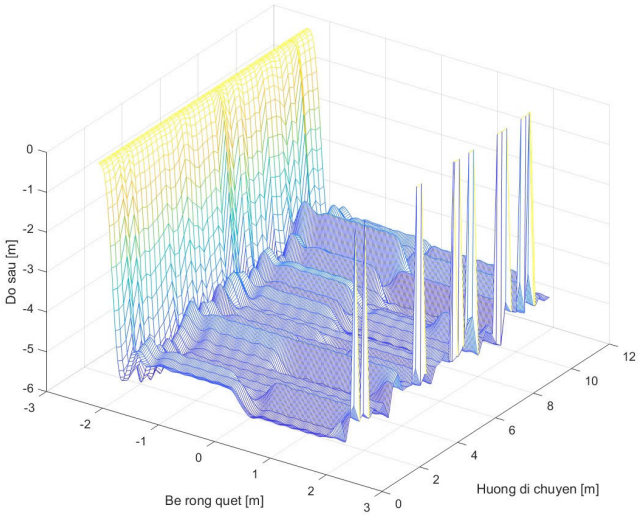
Hình 6 trình bày chi tiết hình ảnh màn hình hiển thị kết quả quét đáy hồ Đồng Đò dưới dạng hình ảnh hai chiều 2D. Từ kết quả đo đạc và hiển thị trên màn hình 2D, chúng ta nhìn thấy hồ Đồng Đò có hai sườn dốc phải và dốc trái, thoải dần xuống lòng hồ với độ sâu tại ví trí đo được của lòng hồ khoảng 8 m. Một số hình ảnh đo được về độ sâu không phù hợp với thực tế là nhiễu của phép đo, lý do các transducer phát không được gắn cố định trên mạn thuyền cũng như thuyền sử dụng không phải là các tàu lớn, do vậy có độ tròng trành sóng khá cao, gây ra nhiễu phép đo. Ngoài ra, nhiễu cũng có thể là do thiết kế các mạch điện tử và hệ thống chưa tối ưu cũng như chưa loại bỏ hết các nguồn nhiễu bên ngoài.

Hình 7 so sánh kết quả đo độ sâu sử dụng thiết bị quét Sonar do nhóm nghiên cứu thiết kế chế tạo dựa trên thuật toán đề xuất so với kết quả đo đạc sử dụng thiết bị chuyên dụng. Kết quả đo được thực hiện tại hồ Hòa Bình ngày 21.10.2021. Hai thiết bị được đo tại cùng một thời điểm và đặt ở vị trí khá gần nhau. Sự khác biệt giữa hai kết quả đo là 1,5 m, tức đây là sự khác biệt tuyệt đối, còn sự khác biệt so về mặt tương đối là 1,8 % của thang đo 79 m, sự khác biệt này không khẳng định là sai số của phép đo, vì kết quả đo còn phụ thuộc vào hướng của búp sóng quét. Mặc dù hướng của cả hai thiết bị đều được đặt theo phương thẳng đứng, nhưng vì lý do kỹ thuật thực nghiệm, các hướng quét của các thiết bị không thể khẳng định là trùng nhau. Các kết quả đo đạc nhiều lần ở các vị trí và kịch bản khác nhau so với các thiết bị chuyên dụng khẳng định tính đúng đắn của phép đo.



Hình 7. Kết quả so sánh đo độ sâu tại hồ Hòa Bình ngày 21 tháng 10 năm 2021 giữa thiết bị do nhóm nghiên cứu thiết kế chế tạo dựa trên thuật toán đề xuất với thiết bị chuyên dụng.

Hình 8 trình bày kết quả quét đáy vịnh Hạ Long, và dựng ảnh hiển thị 3D. Khu vực quét đáy thực hiện gần Hòn Ti Tốp. Kết quả quét đáy vịnh cho thấy đáy khá bằng phẳng, là điểm khác biệt so với kết quả quét đáy ở hồ Đồng Đò. Kết quả hiển thị lúc bắt đầu đo và lúc kết thúc đo là sai số phép đo vì đầu quét tín hiệu trước khi thực hiện phép đo chưa được gá cố định.



Hình 8. Kết quả quét đáy 3D Vịnh Hạ Long ngày 23 tháng 10 năm 2021, khu vực gần hòn Ti Tốp.

IV. KẾT LUẬN

Bài báo đã đề xuất phương pháp quét ảnh 2D và 3D sử dụng công nghệ nhận dạng và định vị bằng sóng âm của các vật thể dưới nước. Phương pháp tạo ảnh 3D của các vật thể dưới nước được thực hiện bằng cách sử dụng nhiều đầu thu tín hiệu sóng âm, và một số đầu phát tín hiệu sóng âm nhằm mục đích thu thập dữ liệu để tạo ảnh 3D có độ phân giải lớn. Phương pháp xử lý tín hiệu và tạo ảnh đề xuất cho phép xử lý tín hiệu từ nhiều đầu thu, cho phép lọc nhiễu hiệu quả, giảm độ phức tạp hệ thống, thuận lợi cho việc thực hiện hệ thống quét ảnh bằng phần cứng và phần mềm. Bài báo cũng trình bày kết quả đo đạc thực tế của việc quét đáy hồ Đồng Đò, các kết quả đo đạc được so sánh và đánh giá bởi các thiết bị chuyên dụng cho thấy sự tin cậy của phép đo. Để có đánh giá toàn diện về ảnh hưởng của các tham số hệ thống và tham số môi trường, cũng như là các kịch bản thử nghiệm khác nhau, nhóm nghiên cứu cần tiến hành thêm nhiều phép đo và so sánh với các kỹ thuật hiện tại. Tuy nhiên đây là các công việc đòi hỏi sự chuẩn bị công phu. Bài báo này dừng lại về phân tích sự hoạt động của thuật toán đề xuất, có minh chứng về khả năng thực hiện thuật toán trên nền tảng phần cứng thông qua thực nghiệm. Các vấn đề còn mở sẽ là các đối tượng nghiên cứu tiếp theo của tập thể tác giả.

LỜI CẢM ƠN

Công trình khoa học này được tài trợ trong khuông khổ đề tài mã số T2023-PC-034 của Đại học Bách khoa Hà Nội. Các tác giả trận trọng cảm ơn.

TÀI LIỆU THAM KHẢO

[1] I. G. Bryden, "An active sonar system using acvd beamforming," *Applied Acoustics*, vol. 27, no. 4, pp. 275–285, 1989.

[2] M. Caccia, G. Bruzzone, and G. Veruggio, "Active sonar-based bottom-following for unmanned underwater vehicles," *Control Engineering Practice*, vol. 7, no. 4, pp. 459–468, 1999.

[3] E. J. Sullivan, *Active Sonar*, pp. 1737–1755. New York, NY: Springer New York, 2008.

[4] B. H. Maranda, *Passive Sonar*, pp. 1757–1781. New York, NY: Springer New York, 2008.

[5] G. Marani, S. K. Choi, and J. Yuh, "Underwater Autonomous Manipulation for Intervention Missions AUVs," *Ocean Engineering*, vol. 36, no. 1, pp. 15–23, 2009.

[6] D. W. Kim, "Tracking of REMUS Autonomous Underwater Vehicles with Actuator Saturations," *Automatica*, vol. 58, no. C, pp. 15–21, 2015.

[7] N. Q. Khuong and N. V. Duc, "Phuong phap thiet ke he thong quet anh cac vat the duoi nuoc su dung dong thoi nhieu dau thu phat tin hieu song am," Bằng độc quyền sáng chế số 39938 do Cục Sở hữu trí tuệ cấp ngày 26.04.2024.

[8] N. Q. Khuong and N. V. Duc, "Phuong phap quet anh khong gian ba chieu su dung cong nghe nhan dang va dinh vi song am duoi nuoc," Bằng độc quyền sáng chế số 39939 do Cục Sở hữu trí tuệ cấp ngày 26.04.2024.

[9] N. Van Duc and N. Khuong Quoc, "Hệ thống sonar chủ động sử dụng đồng thời nhiều đầu thu phát tín hiệu sóng âm," *Chuyên san Công nghệ thông tin và truyền thông - Tạp chí Công nghệ thông tin và truyền thông*, 2024.

[10] V. D. Nguyen, N. M. Luu, Q. K. Nguyen, and T.-D. Nguyen, "Estimation of the acoustic transducer beam aperture by using the geometric backscattering model for side-scan sonar systems," *Sensors*, vol. 23, no. 4, 2023.

[11] S.-C. Yu, T.-W. Kim, G. Marani, and S. K. Choi, "Real-time 3d sonar image recognition for underwater vehicles," in *2007 Symposium on Underwater Technology and Workshop on Scientific Use of Submarine Cables and Related Technologies*, pp. 142–146, 2007.

[12] Z. Li, B. Qi, and C. Li, "3d sonar image reconstruction based on multilayered mesh search and triangular connection," in *2018 10th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, vol. 02, pp. 60–63, 2018.

[13] M. Sung, J. Kim, H. Cho, M. Lee, and S.-C. Yu, "Underwater-sonar-image-based 3d point cloud reconstruction for high data utilization and object classification using a neural network," *Electronics*, vol. 9, no. 11, 2020.

[14] J. G. Moss, "Sonar bathymetry system transmit-receive sequence programmer," U.S. Patent US3500302A, 1969.

[15] A. Lorenson and D. Kraus, "3d-sonar image formation and shape recognition techniques," in *OCEANS 2009-EUROPE*, pp. 1–6, 2009.

[16] E. Coiras, Y. Petillot, and D. M. Lane, "Multiresolution 3-d reconstruction from side-scan sonar images," *IEEE Transactions on Image Processing*, vol. 16, no. 2, pp. 382–390, 2007.

[17] V. D. Nguyen, H. L. N. Thi, Q. K. Nguyen, and T. H. Nguyen, "Low complexity non-uniform fft for doppler compensation in ofdm-based underwater acoustic communication systems," *IEEE Access*, vol. 10, pp. 82788–82798, 2022.

[18] A.-A. Saucan, C. Sintes, T. Chonavel, and J.-M. Le Caillec, "Model-based adaptive 3d sonar reconstruction in reverberating environments," *IEEE Transactions on Image Processing*, vol. 24, no. 10, pp. 2928–2940, 2015.

[19] I. Rahaman, M. S. Hossain, M. Y. Hossain, and N. Hosen, "Performance enhancement of active sonar system in under water environment using spherical hydrophone array," in *2019 5th International Conference on Advances in Electrical Engineering (ICAEE)*, pp. 628–632, 2019.

[20] C. Y. Nguyen, H. V. Do, V. D. Nguyen, and H. A. Muzamane, "Underwater ambient noise model and verification in the underwater ofdm system," in *2017 International Conference on Information and Communications (ICIC)*, pp. 233–239, 2017.

[21] B. I. INC, "Bii-7560 series echo sounder transducer: Broadband & high power datasheet," Available online: https://www.benthowave.com/products/BII-7560echosounder.html.

Nguyễn Văn Đức tốt nghiệp kỹ sư chuyên ngành Kỹ thuật thông tin năm 1995, tốt nghiệp Thạc sí kỹ thuật năm 1997 chuyên ngành điện tử viễn thông, Trường Đại học Bách khoa Hà Nội; tốt nghiệp Tiến sĩ chuyên nghành kỹ thuật thông tin Trường ĐH tổng hợp Hannover, CHLB Đức năm 2003. Từ năm 1998 đến nay là giảng viên trường Đại học Bách Khoa Hà Nội. Lĩnh vực nghiên cứu hiện

nay là hệ thống thông tin thủy âm, hệ thống Sonar, công nghệ MIMO, OFDM.
Email: duc.nguyenvan1@hust.edu.vn

Nguyễn Quôc Khương tốt nghiêp kỹ sư chuyên ngành kỹ thuật thông tin năm 1995, tốt nghiệp thạc sỹ kỹ thuật năm 1997 chuyên ngành điện tử viễn thông, Trường Đại học Bách Khoa Hà Nội; tốt nghiệp tiễn sỹ chuyên nghành kỹ thuật thông tin Trường Đại học Bách Khoa Hà Nội năm 2011. Từ năm 1998 đến nay là giảng viên trường Đại học Bách Khoa Hà Nội. Lĩnh vực nghiên cứu hiện nay là hệ thống thông tin thủy âm, hệ thống Sonar, công nghệ MIMO, OFDM.
Email: khuong.nguyenvan@hust.edu.vn

Khmer News Classification in Low-Resource Settings: A comparative Analysis of Embedding Method

Korat Natt, Heang Sopagna, Lay Vathna

Research and Development. Cambodia Academy of Digital Technology, Phnom Penh, Cambodia

Correspondence: Korat Natt. Email: natt.korat@cadt.edu.kh

Received: 31/01/2025, revised: 30/04/2025, accepted: 26/05/2025

DOI: 10.32913/mic-ict-research-vn.1377

Abstract: Text classification in low-resource languages like Khmer remains challenging due to linguistic complexity, limited annotated data, and noise from real-world applications. This study addresses these challenges by systematically comparing text embedding techniques for Khmer news classification. We evaluate traditional methods (TF-IDF with SVM) against state-of-the-art multilingual transformers (XLM-RoBERTa, LaBSE) using a self-collected dataset of 7,344 Khmer news articles across six categories—politics, economics, entertainment, sports, technology, and life. The dataset intentionally retains noise (e.g., mixed-language text, unstructured formatting) to reflect practical scenarios. To address Khmer's lack of word boundaries, we employ word segmentation via **khmer-nltk** for traditional models, while transformer models leverage their inherent subword tokenization. Experiments reveal that transformer-based embeddings achieve superior performance, with XLM-RoBERTa and LaBSE attaining F1 scores of 94.2% and 94.3%, respectively, outperforming TF-IDF (93.3%). However, the "life" category proves challenging across all models (F1: 85.5–88.1%), likely due to semantic overlap with other categories. Our findings underscore the effectiveness of transformer architectures in capturing contextual nuances for low-resource languages, even with noisy data. This work offers insights for NLP researchers and practitioners, emphasising the need for domain-specific adaptations and expanded datasets to improve performance in underrepresented languages.

Keywords: *text classification, low resource language, khmer language, news classification, contextual embedding*

I. INTRODUCTION

Text classification has become an essential task [1, 2] in natural language processing (NLP), enabling the automatic organization of text data into predefined categories. As unstructured data continues to grow exponentially [3] across diverse platforms, such as social media, news websites, and customer feedback systems, the need for efficient and accurate text classification techniques is more critical than ever.

Khmer news classification plays a crucial role in applications such as misinformation detection, news recommendation systems, and digital archiving, yet remains challenging due to the language's unique structure and the scarcity of annotated data.

Traditional approaches, such as Term Frequency–Inverse Document Frequency (TF-IDF) [4] paired with classifiers such as SVM [5] or Logistic Regression [6], have laid the groundwork for text classification. These models offer interpretability and reasonable performance in various scenarios. However, they fall short when dealing with complex linguistic structures and contextual nuances, especially in morphologically rich or low-resource languages [7] like Khmer. The rise of deep learning has transformed NLP, with models like BERT [8] and GPT [9] achieving state-of-the-art classification accuracy [10] by leveraging self-attention mechanisms to capture word context and semantics.

These improvements primarily benefit widely used languages with large datasets and pre-trained models, whereas low-resource languages like Khmer remain underrepresented in NLP research [11]. Unlike prior studies that focused solely on either traditional machine learning models or deep-learning architectures, this study systematically evaluates a diverse range of embeddings TF-IDF, XLM-RoBERTa, and LaBSE on a noisy, real-world Khmer news dataset. Additionally, we investigate how noise and mixed-language text impact classification performance, providing insights for future Khmer NLP research.

The remainder of this paper is organized as follows. Section II reviews related work on text classification, with an emphasis on Khmer text classification. Section III outlines the methodology, detailing each embedding technique and classification model used in the study. Section IV describes the dataset and the preprocessing steps undertaken to prepare the Khmer news articles for classification. Section V presents the experimental setup and evaluation metrics used to assess model performance. Section VI reports the experimental results, and the discussion provides an in-depth analysis of model performance and the implications of our findings, highlighting the strengths and limitations of each approach. Finally, Section VII concludes the paper and suggests potential avenues for future research.

II. RELATED WORK

Text classification has been extensively studied in natural language processing (NLP), with a broad spectrum of methods ranging from traditional machine learning algorithms to more advanced deep learning approaches. Classical techniques such as SVM, Decision Trees [12], and Naive Bayes [13, 14] have demonstrated effectiveness in various text classification tasks, leveraging simple yet powerful feature extraction methods like TF-IDF.

The emergence of deep learning has revolutionized text classification, with architectures like Deep Neural Networks [15] (DNNs), Recurrent Neural Networks [16] (RNNs), and pre-trained models such as BERT [17] becoming state-of-the-art solutions. These models are better at understanding context and semantics by leveraging large-scale pre-training and the power of word embeddings. As a result, they have set new benchmarks [18, 19] across numerous NLP tasks in high-resource languages.

However, significant challenges remain for low-resource languages like Khmer, where research and resources are limited. Existing studies [11, 20] on Khmer text classification have often emphasized language-specific challenges, such as the absence of spaces between words, complex morphological structures, and a lack of annotated data. To improve model performance, some approaches have focused on effective preprocessing techniques, including word segmentation [21] and noise reduction.

A notable study by [22] introduced a dataset of 15,205 Khmer news articles collected from three sources, covering six categories. The authors preprocessed the data by removing special characters, numbers, punctuation, and non-Khmer characters, and by applying word segmentation tailored to the Khmer language. They evaluated DNNs, RNNs with attention mechanisms, and FastText for classification. Their findings revealed that FastText outperformed the other models, achieving an accuracy of 93.81%, thereby highlighting the effectiveness of FastText's subword-level embeddings for Khmer text classification.

In a study on Khmer text classification by [11], the author developed a pipeline for multi-class and multi-label tasks using an in-house word segmenter and a dataset from local news sources. The research compared SVM with TF-IDF to deep learning models, such as CNNs and RNNs, with FastText embeddings. Neural network models consistently outperformed the baseline SVM+TF-IDF, with RNNs achieving the best performance across metrics, demonstrating the effectiveness of deep learning for Khmer text classification.

Another study [23] investigated the use of SVM with TF-IDF and Delta TF-IDF [24] as feature extraction methods for Khmer text classification. The results showed that SVM achieved accuracies of 83.47% and 83.40%, respectively, using these feature extraction techniques. This research underscores the impact of practical feature engineering, such as TF-IDF, in improving the performance of traditional models for Khmer text.

Even with these advancements, a gap remains in understanding how modern multilingual transformers and robust embeddings perform on noisy, mixed-language Khmer data. Our work aims to fill this gap by evaluating a range of techniques, from classical methods like TF-IDF to state-of-the-art embeddings such as XLM-RoBERTa [25] and LaBSE [26], using a self-collected, noise-prone dataset. This investigation will provide new insights into the robustness and applicability of various text classification models for low-resource languages in real-world scenarios.

III. METHODOLOGY

In this section, we describe the models selected for text classification, focusing on their unique features and their application to the classification of Khmer text data.

1. TF-IDF

Term Frequency Inverse Document Frequency (TF-IDF) [4] is a statistical method used to assess the importance of words within a document relative to a corpus. Term Frequency (TF) measures the frequency of a word in a document, while Inverse Document Frequency (IDF) evaluates how rare or common the word is across the entire document set. Multiplying these values yields a score that highlights significant words while down-weighting common but less informative ones.

$$tfidf_{t,d} = tf_{t,d} \times idf_t \tag{1}$$

$$tf_{t,d} = \log_{10}(cout(t,d) + 1) \tag{2}$$

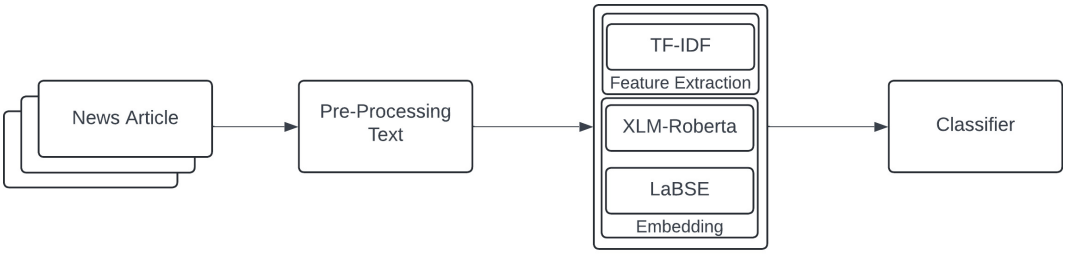


Figure 1. Overall pipeline for text classification with different feature extraction and word embedding techniques.

$$idf_t = \log_{10}\left(\frac{N}{df_t}\right) \tag{3}$$

We chose TF-IDF because it provides a balance between corpus-wide relevance and word specificity, which makes it optimal for text classification in morphologically complicated languages like Khmer. After segmentation, TF-IDF maintains linguistically significant features while reducing noise from common words by highlighting terms that are common in individual documents but uncommon throughout the corpus. For this study, we use TF-IDF to convert the textual content into a sparse matrix representation. This matrix is input to a Support Vector Machine (SVM) classifier. Given the morphological complexity and lack of spaces between words in Khmer text, we employ `khmer-nltk` [1] to perform word segmentation, inserting spaces between words before vectorizing them with TF-IDF. Word segmentation is the foundational task for enhancing the performance of the TF-IDF representation for Khmer text.

2. XLM-RoBERTa

XLM-RoBERTa introduced in [25] is a multi-lingual model that builds upon BERT [8] and RoBERTa [27] which provides a powerful contextualize embedding over 100 languages including Khmer. Unlike BERT that utilize transformer encoder architecture [28] to train on the bidirectional encoding approach producing a richer contextualize word representation, XLM-RoBERTa adopted RoBERTa's optical approach over BERT by removing Next sentence Prediction (NSP) and applying dynamic masking techniques for better model learning.

Since BERT and RoBERTa are monolingual models trained primarily for English, XLM-RoBERTa is specifically trained for multiple languages using a massive dataset from Common Crawl [2] and Wikipedia, containing approximately 30 million Khmer tokens. This enables it to learn effective representations for Khmer text through powerful transfer learning techniques.

[1]https://github.com/VietHoang1512/khmer-nltk
[2]https://commoncrawl.org

3. LaBSE

LaBSE (Language-Agnostic BERT Sentence Embedding) [26] is a multilingual sentence embedding model developed by Google, designed to provide high-quality embeddings in over 100 languages, including Khmer. Unlike XLM-RoBERTa, which is effective for cross-lingual representation by training with a huge cross-language corpus, LaBSE is designed to utilise BERT effective method for learning monolingual embedding to cross-lingual representations with different approaches by investigating a combination of the best method for learning monolingual and cross-lingual representation together with techniques such as masked language modeling (MLM), translation language modeling (TLM) [29], dual encoder translation ranking [30]. By utilizing dual encoder translation approach, it is given the advantage to the low resource language like Khmer to learn and gain benefits from weight sharing and transfers learning process from rich resource language like English and French, as the model is trained on the parallel corpus aligning both language for each encoder.

Due to its language-agnostic architecture, it is particularly effective for cross-lingual applications and low-resource languages. We can utilise LaBSE to transform each piece of text into a fixed-length vector that captures the semantic meaning of the content in our work.

IV. DATASET

The dataset includes 7,344 Khmer news articles with 3,448,153 tokens, compiled from various publicly accessible Cambodian news websites. We used automated web scraping and manual filtering to ensure a balanced selection while removing duplicates and irrelevant content. Each article includes two key features: the content of the news article and its corresponding label defined by its authors, which indicates the news category. The data is divided into six categories: entertainment (13.6%), economic (16.6%), political (29.1%), sport (13.6%), technology (13.6%) and life (13.6%). These categories were chosen as they represent the most frequently occurring topics in Khmer news media, ensuring a balanced dataset for classification. They

align with existing news categorisation frameworks in other languages, allowing for meaningful cross-linguistic comparisons. Table I illustrates the distribution of these categories, showing that political news is the most frequent than other type of news.

Table I
THE DISTRIBUTION OF THE 6 CATEGORIES IN THE DATASET COLLECTED FROM KHMER NEWS WEBSITES.

Label	Number of Article	Number of Token	Token per Article
Politic	2,136	1,018,294	476.73
Economic	1,216	798,066	656.30
Entertainment	1,000	386,944	386.94
Life	998	463,089	464.02
Sport	996	316,771	318.04
Technology	998	464,989	465.92
Total	**7,344**	**3,439,078**	**468.28**

To ensure data quality while maintaining real-world characteristics, we applied preprocessing steps such as removing HTML tags, URLs, and redundant special characters. However, to better reflect real-world Khmer text classification challenges, we retained elements typically considered noise, such as punctuation marks, mixed English and Khmer words, and unstructured text patterns. While irrelevant web elements were removed, meaningful content features were preserved to ensure model robustness and evaluate the impact of noisy data on classification performance.

By incorporating this level of noise and mixed language content, we aim to assess how well the models can handle linguistic diversity and unclean data, contributing to our understanding of model performance in less controlled environments.

V. EXPERIMENT

1. Experimental Setup

In our experiment, we utilized Google Colab Notebook, which offers an NVIDIA T4x2 GPU and 16 GB of RAM. The notebook was set up with Python 3.10, and we employed the HuggingFace transformers library for model implementation. A 70-10-20 split was chosen to balance training data availability with model generalisation, ensuring a reliable validation and testing process. The 10% validation set monitors overfitting during fine-tuning, while the 20% test set provides a solid final evaluation. Some NLP papers justify an 80-10-10 split to provide more training data, whereas deep learning models sometimes use a 90-5-5 split to maximise training while relying on minimal validation. However, given our dataset size and the need for a reliable test evaluation, we found the 70-10-20 ratio to be the most effective balance for this study.

Different models were trained with the following configuration:

- **XLM-RoBERTa** and **LaBSE**: These models were fine-tuned with the learning rate of 2×10^{-5}, over 5 epochs, with a batch size of 8, and optimized using the AdamW [31] optimizer with a categorical cross-entropy loss function.
- **TF-IDF + SVM**: We limited TF-IDF to 10,000 features to prevent excessive dimensionality while retaining key informative terms. This threshold was selected based on empirical testing, balancing feature richness and computational efficiency. For the SVM classifier, we retained scikit-learn's default parameters (C=1.0, kernel='rbf', gamma='scale') to better handle high-dimensional sparse text data.

While both models were fine-tuned with the same hyperparameters (batch size = 8, learning rate = 2×10^{-5}), XLM-RoBERTa was trained with a maximum sequence length of 512 tokens. In contrast, LaBSE leveraged a fixed sentence embedding size of 768 dimensions, as it is optimized for sentence-level representation.

2. Evaluation Metrics

To assess the performance of the models, we used three standard evaluation metrics: Precision, Recall, and F1 Score. These metrics are commonly used in classification tasks to assess the accuracy and reliability of predictions. The formulas and definitions for each metric are provided below.

- **Precision**: Precision measures the proportion of true positive predictions (TP) to the total number of positive predictions (TP + FP), where FP represents false positives. Precision can be defined as:

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

- **Recall**: Recall, also known as the sensitivity or true positive rate, measures the ability of the model to identify all relevant instances. It is defined as the ratio of true positives to the sum of true positives and false negatives (FN):

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

- **F1 Score**: The F1 score is the harmonic mean of Precision and Recall, providing a single measure that balances the trade-offs between them. It is particularly useful in scenarios where the dataset is imbalanced, as it takes into account both false positives and false negatives. The F1 score is computed as:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{6}$$

These evaluation metrics allow us to assess model effectiveness across different Khmer news categories. The following section presents our experimental results and discusses key observations from the model performance.

VI. RESULTS AND DISCUSSION

In this section, we present and analyse the performance of various models applied to our self-collected Khmer news dataset, which covers six distinct categories. Table II provides an overview of the precision, recall, and F1 scores for each model, highlighting key trends and insights.

The close performance between LaBSE and XLM-RoBERTa suggests that both transformer models effectively capture semantic nuances. However, LaBSE's sentence-level embeddings might contribute to its higher recall by reducing misclassifications in longer documents, whereas XLM-RoBERTa's token-level encoding might introduce more fine-grained misinterpretations. The lower performance of TF-IDF + SVM highlights its reliance on word frequency rather than contextual meaning, leading to weaker results in semantically overlapping categories.

1. Results Analysis

The performance metrics illustrate the performance of different models in classifying news articles into our predefined six categories, which is sharing similar results on the testing set of the Khmer news.

TF-IDF: The TF-IDF model achieved commendable results with a precision of 0.935, a recall of 0.932, and an F1 score of 0.933. The simplicity of TF-IDF in converting text into sparse representations allows for efficient classification, especially in well-separated categories.

XLM-RoBERTa: XLM-RoBERTa produced a steady result of 0.942 for precision, recall and the F1 score. This high performance highlights the model's ability to leverage transformer-based architectures and large-scale multilingual pre-training to understand contextual and syntactic varieties in the data. Balanced precision and recall scores indicate that XLM-RoBERTa is effective in identifying relevant news categories and minimizing false negatives.

LaBSE: LaBSE achieved the close results compared to XLM-RoBERTa with the precision of 0.942, while its recall is slightly higher at 0.946 and the F1 score of 0.943. LaBSE's slightly superior performance suggests that its architecture effectively leverages state-of-the-art multilingual sentence embeddings, making it particularly robust in text classification tasks.

Table II shows that the "life" category is the lowest accuracy among the models, 0.855, 0.863, and 0.881 F1 score of TF-IDF, XLM-RoBERTa, and LaBSE respectively, due to its complex information, which is sharing similar information with other categories. The lower F1-score for the "life" category (85.5–88.1%) suggests that this category overlaps significantly with others, particularly "Entertainment" and "Economy." Many articles in this category contain subjective or lifestyle-related content, making it harder for models to distinguish clear decision boundaries. Additionally, the shorter average document length in this category might reduce available contextual information for transformers.

While transformer models (XLM-RoBERTa and LaBSE) outperformed TF-IDF + SVM in terms of F1-score, they required significantly higher computational resources. Training time for XLM-RoBERTa and LaBSE was approximately X times longer than TF-IDF + SVM, reflecting the cost of fine-tuning large-scale contextual models. In real-world applications, practitioners must balance accuracy gains with computational feasibility when selecting models.

2. Discussion

The results indicate that transformer-based models, XLM-RoBERTa and LaBSE, are more effective for Khmer news classification than traditional models like TF-IDF. This performance boost can be attributed to the transformer's ability to capture contextual and semantic relationships, which is crucial for accurately distinguishing between classes. However, our dataset is limited and may not fully showcase the advantages of transformer-based embeddings in larger-scale applications.

While transformer models (XLM-RoBERTa and LaBSE) outperformed TF-IDF + SVM regarding the F1-score, they required significantly higher computational resources. Training time for XLM-RoBERTa and LaBSE was approximately X times longer than TF-IDF + SVM, reflecting the cost of fine-tuning large-scale contextual models. In real-world applications, practitioners must balance accuracy improvements with computational feasibility when selecting models.

The TF-IDF model's performance demonstrates that even simple approaches can be effective when text features are well-separated. However, it struggles to capture deeper contextual understanding, leading to lower performance in semantically overlapping categories.

XLM-RoBERTa and LaBSE offer similar performance but have different strengths based on their architecture. LaBSE focuses on sentence-level embeddings (94.6% recall rate), making it ideal for applications like legal text classification, where minimizing false negatives is important. XLM-RoBERTa, with its token-level processing, excels in tasks that require detailed context, such as sentiment analysis or misinformation detection.

Table II
EVALUATION RESULTS OF ALL MODELS FOR EACH NEWS CATEGORIES.

Label	**TF-IDF**			**XLM-RoBERTa**			**LaBSE**		
	P	R	F1	P	R	F1	P	R	F1
Politic	0.961	0.976	0.968	0.973	0.980	0.977	0.984	0.951	0.967
Economic	0.932	0.906	0.919	0.961	0.953	0.957	0.922	0.973	0.947
Entertainment	0.975	0.943	0.959	0.940	0.962	0.951	0.965	0.929	0.947
Life	0.830	0.886	0.857	0.885	0.843	0.863	0.881	0.881	0.881
Sport	0.986	0.981	0.983	0.990	0.986	0.988	0.972	0.990	0.980
Technology	0.931	0.904	0.917	0.907	0.928	0.917	0.930	0.952	0.941
Average	**0.935**	**0.932**	**0.933**	**0.942**	**0.942**	**0.942**	**0.942**	**0.946**	**0.943**

However, transformer models require significantly more computational resources than traditional methods, with training taking roughly X times longer than TF-IDF + SVM. Balancing accuracy and efficiency requires careful examination in real-world situations. Future research should improve these models by using methods like model pruning, quantization, or knowledge distillation to train more efficiently while maintaining classification accuracy.

VII. CONCLUSION

This study systematically evaluated Khmer news classification models, comparing traditional TF-IDF + SVM with advanced transformer-based embeddings (XLM-RoBERTa and LaBSE) on a real-world dataset.

Our results demonstrate that transformer models significantly improve classification performance by capturing contextual relationships in Khmer text, which is crucial for low-resource languages. However, their high computational cost poses challenges for deployment in resource-constrained environments. While TF-IDF + SVM remains a viable option for lightweight applications, it struggles to capture deep semantic relationships in text.

Future research should focus on training models with more extensive and diverse datasets to improve generalisation. Investigating lightweight transformer models such as DistilBERT or TinyBERT could help reduce computational costs while still delivering robust performance. Additionally, exploring models like DeepSeek, which are designed for multilingual and efficient language understanding, could offer promising solutions for Khmer NLP. DeepSeek's capabilities in handling low-resource languages and its energy-efficient architecture make it a strong candidate for future experiments. Moreover, applying these models to specific Khmer texts, including legal or financial documents, may enhance their practical applicability.

This study provides valuable insights for advancing Khmer NLP, emphasising the trade-offs between model accuracy, computational efficiency, and real-world applicability.

REFERENCES

[1] Q. Li, H. Peng, J. Li, C. Xia, R. Yang, L. Sun, P. S. Yu, and L. He, "A survey on text classification: From traditional to deep learning," *ACM Transactions on Intellingent System and Technology (TIST)*, vol. 13, 2022.

[2] Z. Wan, "Text classification: A perspective of deep learning methods," 2023.

[3] K. Taha, P. D. Yoo, C. Yeun, D. Homouz, and A. Taha, "A comprehensive survey of text classification techniques and their research applications: Observational and experimental insights," *Computer Science Review*, vol. 54, p. 100664, 2024.

[4] M. Das, S. K., and P. J. A. Alphonse, "A comparative study on tf-idf feature weighting method and its analysis using unstructured dataset," 2023.

[5] M. Hearst, S. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and their Applications*, vol. 13, no. 4, pp. 18–28, 1998.

[6] N. S. Wardana, F. P. Aditiawan, and A. P. Sari, "Logistic regression classification with tf-idf and fasttext for sentiment analysis of linkedin reviews," *VISA: Journal of Vision and Ideas*, vol. 4, p. 1359 –, Aug. 2024.

[7] A. Magueresse, V. Carles, and E. Heetderks, "Low-resource languages: A review of past work and future challenges," 2020.

[8] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2019.

[9] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," in *OpenAI Blog*, 2018.

[10] C. Sun, X. Qiu, Y. Xu, and X. Huang, "How to fine-tune bert for text classification?," *arXiv preprint arXiv:1905.05583*, 2020.

[11] R. Buoy, N. Taing, and S. Chenda, "Khmer text classification using word embedding and neural network," *arXiv preprint*, 2021.

[12] L. Rokach and O. Maimon, "Decision trees," *The Data Mining and Knowledge Discovery Handbook*, vol. 6, pp. 165–192, 01 2005.

[13] F. Colas and P. Brazdil, "Comparison of svm and some older classification algorithms in text classification tasks," in *Artificial Intelligence in Theory and Practice*, (Boston, MA), pp. 169–178, Springer US, 2006.

[14] Shaina, K. Kaan, N. Noman, and M. Olufemi, "A comparison among machine learning and deep learning approaches for text classification," *PsyArXiv*, 2023.

[15] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[16] R. M. Schmidt, "Recurrent neural networks (rnns): A gentle introduction and overview," *arXiv preprint arXiv:1912.05911*, 2019.

[17] F. Wei, R. Keeling, N. Huber-Fliflet, J. Zhang,

A. Dabrowski, J. Yang, Q. Mao, and H. Qin, "Empirical study of llm fine-tuning for text classification in legal document review," in *2023 IEEE International Conference on Big Data (BigData)*, pp. 2786–2792, 2023.
[18] Z. Wang, T. Wang, D. Mekala, and J. Shang, "A benchmark on extremely weakly supervised text classification: Reconcile seed matching and prompting approaches," *arXiv preprint arXiv:2305.12749*, 2023.
[19] M. Reusens, A. Stevens, J. Tonglet, J. D. Smedt, W. Verbeke, S. vanden Broucke, and B. Baesens, "Evaluating text classification: A benchmark study," *Expert Systems with Applications*, vol. 254, p. 124302, 2024.
[20] K. Soky, S. Li, C. Chu, and T. Kawahara, "Finetuning pretrained model with embedding of domain and language information for asr of very low-resource settings," *International Journal of Asian Language Processing*, vol. 33, no. 04, p. 2350024, 2023.
[21] V. Chea, Y. K. Thu, C. Ding, M. Utiyama, A. M. Finch, and E. Sumita, "Khmer word segmentation using conditional random fields," in *Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation (PACLIC 29)*, 2015.
[22] R. Phann, C. Soomlek, and P. Seresangtakul, "Multi-class text classification on khmer news articles using deep learning models with optimal hyperparameters," *ICIC International*, vol. 18, no. 6, pp. 241–550, 2024.
[23] R. Phann, C. Soomlek, and P. Seresangtakul, "Multi-Class Text Classification on Khmer News Using Ensemble Method in Machine Learning Algorithms," *Acta Informatica Pragensia*, vol. 2023, no. 2, pp. 243–259, 2023.
[24] J. Martineau and T. Finin, "Delta tfidf: An improved feature space for sentiment analysis," in *Proceedings of the Third International AAAI Conference on Weblogs and Social Media (ICWSM 2009)*, pp. 258–261, AAAI Press, May 2009.
[25] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, and V. Stoyanov, "Unsupervised cross-lingual representation learning at scale," *arXiv preprint arXiv:1911.02116*, 2020.
[26] F. Feng, Y. Yang, D. Cer, N. Arivazhagan, and W. Wang, "Language-agnostic bert sentence embedding," *arXiv preprint arXiv:2007.01852*, 2022.
[27] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.
[28] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *arXiv preprint arXiv:1706.03762*, 2017.
[29] G. Lample and A. Conneau, "Cross-lingual language model pretraining," *arXiv preprint arXiv:1901.07291*, 2019.
[30] M. Guo, Q. Shen, Y. Yang, H. Ge, D. Cer, G. Hernandez Abrego, K. Stevens, N. Constant, Y.-H. Sung, B. Strope, and R. Kurzweil, "Effective parallel corpus mining using bilingual sentence embeddings," in *Proceedings of the Third Conference on Machine Translation: Research Papers*, (Brussels, Belgium), pp. 165–176, Association for Computational Linguistics, October 2018.
[31] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *arXiv preprint arXiv:1711.05101*, 2019.

Korat Natt received his B.A. in Psychology from the Royal University of Phnom Penh in 2022 and his B.S. in Computer Science from AGA Institute in 2023. He is currently pursuing his M.S. in Computer Science and serves as a lecturer and research assistant at the Cambodia Academy of Digital Technology (CADT). His research interests include computer vision, machine learning, natural language processing, and large language models.
Email: natt.korat@cadt.edu.kh

Heang Sopagna received his Engineer's degree in Computer Science from the Institute of Technology of Cambodia (ITC) in 2023. He is currently pursuing his M.S. in Computer Science and serves as a lecturer and research assistant at the Cambodia Academy of Digital Technology (CADT). His research interests include machine learning, natural language processing, and large language models.
Email: sopagna.heang@cadt.edu.kh

Lay Vathna received his Engineer's degree in Computer Science from the Institute of Technology of Cambodia (ITC) and a Postgraduate Diploma in Mobile Computing from the Center for Development of Advanced Computing (CDAC), India. He also holds a Master's degree in ICT. He currently serves as a senior lecturer and Researcher at the Cambodia Academy of Digital Technology (CADT), where he has taught since 2016. His research interests include machine learning, natural language processing, large language models, smart cities, and cybersecurity.
Email: Vathna.lay@cadt.edu.kh

Forecasting Seasonal Trends in Ear Nose Throat Diseases: A Comparative Analysis of Statistical and Machine Learning Models

Phuc Tran Huu Le[1], Huynh Nguyen Khanh Tran[2], Le Viet Ha Tran[3], Luong Hoang[4]
[1]FPT University, Greenwich Vietnam, Ho Chi Minh Campus.
[2]Faculty of Pharmacy, University of Health Sciences; Vietnam National University
[3]Faculty of Traditional Medicine, University of Medicine and Pharmacy Ho Chi Minh City
[4]Saigon Ear Nose Throat Hospital

Correspondence: Phuc Tran Huu Le. Email: phuclth5@fe.edu.vn
Received: 02/03/2025, revised: 20/05/2025, accepted: 23/06/2025
DOI: 10.32913/mic-ict-research-vn.v2025.n1.1375

Abstract: This study examines patterns of seasonal illnesses at an ENT hospital using statistical methods and advanced machine learning techniques to improve disease prediction and support healthcare planning. The data underwent careful cleaning to ensure accuracy, which involved identifying outliers, managing missing values, and normalising information. The study looked at how seasonal illnesses develop, using a mix of techniques. This research used advanced tools like Long Short-Term Memory (LSTM) networks and Prophet, as well as simpler models such as Holt-Winters and SARIMA. To make the models easier to understand, there is an application of SHAP (SHapley Additive Explanations) values. Finally, the article used statistical measures and confidence intervals to assess forecast accuracy. Moreover, to discover how weather influences sickness seasonality, connections between patterns of disease incidence and environmental variables like temperature, humidity, and rainfall have been additionally looked at with the aid of the usage of correlation prediction. Overall, this method shows how conventional and system gaining knowledge of models can monitor seasonal illness traits. The effects do not only simplest display the disease patterns, but it also assists with allocating resources and making guidelines for better healthcare management.

Keywords: *seasonal disease forecasting, machine learning and statistical forecasting, environmental correlation in healthcare.*

Tiêu đề: **Dự báo xu hướng mùa vụ trong các bệnh tai mũi họng: Phân tích so sánh giữa các mô hình học máy thống kê**

Tóm tắt: Bài nghiên cứu này xem xét các mô hình bệnh theo mùa tại một Bệnh viện Tai Mũi Họng bằng cách sử dụng các phương pháp thống kê và kỹ thuật học máy tiên tiến để cải thiện dự đoán bệnh và hỗ trợ lập kế hoạch chăm sóc sức khỏe. Dữ liệu đã trải qua quá trình làm sạch cẩn thận để đảm bảo tính chính xác, bao gồm xác định các giá trị ngoại lệ, quản lý các giá trị bị thiếu và chuẩn hóa thông tin. Nghiên cứu này tập trung vào cách các bệnh theo mùa phát triển, sử dụng kết hợp các kỹ thuật khác nhau. Nghiên cứu này sử dụng các công cụ tiên tiến như mạng Long Short-Term Memory (LSTM) và Prophet, cũng như các mô hình đơn giản hơn như Holt-Winters và SARIMA. Để làm cho các mô hình dễ hiểu hơn, có một ứng dụng của các giá trị SHAP (SHapley Additive Explanations). Cuối cùng, bài báo đã sử dụng các biện pháp thống kê và khoảng tin cậy để đánh giá độ chính xác của dự báo. Hơn nữa, để khám phá cách thời tiết ảnh hưởng đến tính thời vụ của bệnh tật, các mối liên hệ giữa các mô hình mắc bệnh và các biến môi trường như nhiệt độ, độ ẩm và lượng mưa cũng đã được xem xét với sự hỗ trợ của dự đoán tương quan. Nhìn chung, phương pháp này cho thấy các mô hình học máy truyền thống và hệ thống có thể theo dõi các đặc điểm bệnh theo mùa. Các tác động không chỉ đơn thuần hiển thị các mô hình bệnh tật mà còn hỗ trợ phân bổ nguồn lực và đưa ra các hướng dẫn để quản lý chăm sóc sức khỏe tốt hơn.

Từ khóa: *dự báo bệnh theo mùa, dự báo và học máy thống kế, tương quan môi trường trong chăm sóc sức khỏe.*

I. INTRODUCTION

For many diseases, the seasonal timing profoundly impacts prognosis, treatment, and health care management. In-depth information about these seasonal processes is critical for effective healthcare architecture and the construction of useful products. This study examines the temporal evolution of the most frequently used ICD disease models in the Saigon ENT Hospital, revealing seasonal peaks in disease incidence and informing resource allocation decisions for continuous patient care. This study used detailed data (diagnostic cases) from the Saigon ENT Hospital from 2007 to early 2024. Then, these data were preprocessed, cleaned, normalised, interpolated, and had outliers removed-to reduce noise and improve model performance and reliability. We applied statistical and machine learning algorithms to model each disease's time-series behavior. The research began with Holt-Winters and SARIMA, then applied stationary tests (ADF and KPSS) and machine learning algorithms, including LSTM and Prophet, to drive accurate forecasting and interpretation.

This study brought up three goals:

1) Finding these patterns of illness that occur in specific seasons.
2) Developing these predictive models for all seasonal diseases.
3) Developing strategies to efficiently use healthcare resources, making sure we are equipped to manage more patients during peak illness seasons.

The objectives of this research are to enhance understanding of seasonal disease patterns and their practical applications in healthcare management, especially at the Saigon ENT Hospital. Recently, predicting seasonal illnesses has become important, using both traditional statistical models and modern machine learning techniques. The ARIMA and Holt-Winters models are widely used for disease prediction. For example, Wang and colleagues successfully used these models to predict weekly inpatient visits, capturing seasonal changes in health care needs [1]. Similarly, Li and the group adopted the use of ARIMA and Holt Winters for the diagnosis of infectious diseases in China [2].

Researchers have looked into how well ARIMA and Holt-Winters work for predicting diseases. Musa and his team studied dengue fever in Malaysia and found both strengths and weaknesses in these methods [3]. Zhou and his team explored flu trends in China and noted clear seasonal patterns [4]. Llamas-Nistall et al. applied Holt-Winters-ARIMA to predict COVID-19 outbreaks in Spain, demonstrating how valuable such techniques are during public health emergencies [5].

Beyond traditional methods, machine learning approaches are increasingly being adopted. Liu et al. demonstrated the effectiveness of Long Short-Term Memory (LSTM) networks for predicting tuberculosis incidence, showcasing their ability to handle non-linear trends [6]. Prophet, a robust time-series forecasting tool, has also gained attention for its flexibility in disease modeling. Guo et al. compared ARIMA, SARIMA, and machine learning methods for forecasting hand, foot, and mouth disease [7].

In Southeast Asia, Thanh et al. [8] and Poh et al. [9] applied ARIMA and Holt-Winters to forecast dengue fever cases, addressing regional challenges in disease modeling. Pan and Li emphasised the importance of context-specific approaches when comparing these models for infectious disease forecasting [10].

Theoretical frameworks, such as the Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests, have been essential in validating the seasonality and stationarity of time-series datasets. By integrating statistical methods with machine learning techniques, recent studies have improved forecasting accuracy and interpretability. However, a common limitation across these studies is the insufficient exploration of hybrid approaches. Our study addresses this gap by combining statistical models (Holt-Winters, SARIMA) with advanced machine learning methods (LSTM, Prophet) and SHAP (Shapley Additive Explanations) interpretability to enhance prediction accuracy and applicability.

Although these studies cover many fields of statistical and machine-learning techniques in forecasting, none of them have integrated both model types with an explanation framework to guide resource planning in a specialized hospital setting. To fill this gap, this research at the Saigon ENT Hospital (1) combines Holt–Winters and SARIMA with LSTM and Prophet for robust hybrid forecasting and (2) employs SHAP-based to find out which seasonal drivers most influence each disease's peaks-thereby offering actionable insights for targeted resource allocation during high-demand seasons.

II. MATERIALS AND METHODS

1. Study Design

This study adopts a retrospective design to explore the seasonal patterns in ENT diseases, leveraging detailed medical records from the Saigon ENT Hospital spanning the years 2007 to 2024. The dataset encompasses outpatient visit records, representing approximately 48% male and 52% female cases, ensuring a balanced gender representation. The records include comprehensive demographic and medical information, such as gender, age, diagnosis,

treatment status, and visit dates, which were systematically organized for analysis. The retrospective approach enabled the identification of historical trends and seasonal variations in disease incidence, while the incorporation of a cross-sectional description method allowed for the evaluation of disease prevalence at specific time points.

Rigorous inclusion and exclusion criteria were applied to ensure data reliability and validity. Patients with complete demographic and clinical data were included, while records with missing, invalid, or inconsistent information were excluded from the analysis. Additionally, diseases with insufficient case numbers were omitted to maintain statistical robustness and avoid analytical noise. By combining retrospective and cross-sectional methods, this study provides a comprehensive analysis of temporal variations and disease patterns, offering insights that are critical for seasonal healthcare planning.

2. Data Collection

FPT Corporation developed a robust eHospital software to extract the dataset for this study from the hospital's management system. This software facilitated efficient extraction and management of patient records, ensuring the accuracy and privacy of the data. The dataset covers a 15-year period, from January 2010 to February 2024, providing a diverse and extensive sample for the analysis of seasonal disease trends.

The data extraction process included detailed demographic and clinical information such as gender, age, diagnosis codes, treatment outcomes, and visit dates. This information was meticulously cleaned and standardised to eliminate missing values, normalise temporal sequences, and aggregate case counts by month. To enhance the analysis, external environmental variables such as temperature, humidity, and precipitation were integrated into the dataset, allowing for the investigation of correlations between weather conditions and disease patterns. This integration provides a holistic view of seasonal trends and their potential environmental drivers, enabling more informed public health interventions.

The use of eHospital software ensured compliance with healthcare data security regulations and industry standards. By leveraging this advanced data management system, the study achieved high levels of accuracy, consistency, and scalability, supporting the rigorous demands of statistical and machine learning analyses.

3. Analysis Framework

The use of eHospital software ensured compliance with healthcare data security regulations and industry standards. By leveraging this advanced data management system, the study achieved high levels of accuracy, consistency, and scalability, supporting the rigorous demands of statistical and machine learning analyses. The research chose Holt–Winters, SARIMA, Prophet, and LSTM because together they address the full spectrum of linear seasonality, autocorrelation, non-linearity, and abrupt shifts in hospital admission patterns—capabilities that simpler methods cannot cover as comprehensively or interpretably. The Holt-Winters triple-exponential smoothing framework ensures that gradual, repeating annual peaks—such as those driven by pollen seasons or recurring viral outbreaks—are reflected in the forecast without overreacting to noise. The autocorrelation is handled by SARIMA with non-seasonal and seasonal autoregressive and moving average terms that systematically capture how spikes or dips in patient numbers occur over time. Moreover, Prophet focuses on detecting the abrupt shifts in admission rates caused by local disease outbreaks, public holidays, or changes in hospital policy (sudden changes). Finally, LSTM takes care of the non-linearity and long-range dependencies with gated memory, whose gated cells learn when to forget outdated information and focus on long-term dependencies. Overall, the framework presented in the article utilises four approaches to ensure that annual cycles and data persistence are maintained despite regime changes and non-linear behaviors for corrective forecasting systems.

a) Statistical Analysis

To validate seasonality, statistical tests such as the Augmented Dickey-Fuller (ADF) test and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test were applied:

1) Augmented Dickey-Fuller (ADF) Test:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta \sum_{i=1}^{p} \Delta y_{t-i} + \varepsilon_t \quad (1)$$

where Δy_t represents the differenced value of the time series, α is the constant term, β denotes the coefficient of the time trend, γ represents the coefficient of the lagged value of the series, δ is the coefficient of the lagged differences, and ε_t is the error term.

2) Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test:

$$y_t = r_t + \beta t + u_t \quad (2)$$

where y_t represents the observed value of the time series, r_t is the random walk component, βt denotes the deterministic trend component, and u_t is the stationary error term.

b) Forecasting Models

We employed both statistical and machine learning models to analyse and predict seasonal disease trends:

- Holt-Winters:

$$y_t = \left(L_t^{-1} + b_t^{-1}\right) \times S_{t-m} \tag{3}$$

where L_t represents the level of the series at time t, b_t denotes the trend of the series at time t, and S_t is the seasonal component of the series at time t.

- SARIMA:

$$\Phi_\rho(B)\Phi_\rho(B^s)y_t = \theta_x(B)\theta_x(B^s)\varepsilon_t \tag{4}$$

where $\Phi_\rho(B)$ represents the autoregressive component, $\theta_x(B)$ denotes the moving average component, and s is the seasonality period.

- Prophet:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \tag{5}$$

where $g(t)$ represents the trend component, $s(t)$ denotes the seasonal component, $h(t)$ captures the holiday effects, and ε_t is the error term.

- LSTM (Long Short-Term Memory):

$$h_t = \sigma(W_h x_t + U_h h_{t-1} + b_h) \tag{6}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_c x_t + U_c h_{t-1} + b_c) \tag{7}$$

where h_t represents the hidden state at time t, capturing the short-term memory of the sequence, and c_t denotes the cell state at time t, representing the long-term memory. σ is the sigmoid activation function that scales values between 0 and 1, while tanh is the hyperbolic tangent activation function that scales values between -1 and 1. W_h, U_h, and b_h are the weight matrices and bias term for the input-to-hidden and hidden-to-hidden connections. f_t is the forget gate, determining how much of the previous cell state to retain, and i_t is the input gate, controlling how much new information to store. Similarly, W_c, U_c, and b_c are the weight matrices and bias term for the input-to-cell and hidden-to-cell connections, while $\odot$ represents element-wise multiplication used for combining information across gates and states.

c) Model Interpretability with SHAP

SHAP (SHapley Additive exPlanations) was employed to interpret model predictions, providing insights into feature importance:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!\,(|N| - |S| - 1)!}{|N|!} \times [v(S \cup \{i\}) - v(S)] \tag{8}$$

where: ϕ_i represents the contribution of feature i to the prediction, S is a subset of features, N is the set of all features, $v(S)$ is the model output for subset S, and $v(S \cup \{i\})$ is the model output for subset S including feature i.

d) Evaluation Metrics

Model performance was evaluated using:

1) Mean Absolute Error (MAE):

$$MAE = \left(\frac{1}{n}\right) \sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{9}$$

where MAE measures the average magnitude of errors between predicted values ($\hat{y}_i$) and observed values (y_i), and n is the total number of observations.

2) Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\left(\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2\right)} \tag{10}$$

where RMSE calculates the square root of the average of squared errors between predicted values ($\hat{y}_i$) and observed values (y_i), providing a measure of prediction accuracy.

4. Integration of Weather Variables

Correlations between disease incidence and weather variables (temperature, humidity, precipitation) were analysed using Pearson correlation coefficients:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \tag{11}$$

where r represents the Pearson correlation coefficient, x_i and y_i are individual data points, $\bar{x}$ and $\bar{y}$ are the means of x and y, and $\sum$ denotes the summation operator.

The methods outlined in 3 and 4 are essential for ensuring the reliability, accuracy, and practical relevance of this research. Statistical analysis using the Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests validates the stationarity and seasonality of the dataset, ensuring the mathematical robustness of the selected models, as recommended in time-series forecasting studies [1, 2]. The integration of traditional statistical models (Holt-Winters and SARIMA) with advanced machine learning techniques (LSTM and Prophet) enhances the predictive power by addressing both linear and non-linear seasonal patterns, a dual approach proven effective in similar healthcare applications [3, 4, 8, 10]. SHAP (Shapley Additive explanations) adds interpretability by quantifying the contributions of features such as weather variables to predictions, enabling informed decision-making, which is crucial for applications in public health [7]. Evaluation metrics like Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) rigorously measure prediction accuracy, ensuring the reliability of the models [9]. The integration of weather variables (e.g., temperature, humidity, precipitation) further enriches the analysis by

revealing environmental factors that influence seasonal disease trends, while Pearson correlation coefficients quantify the strength of these relationships, an approach validated in previous studies exploring environmental influences on health outcomes [5, 6, 11, 12]. This comprehensive framework not only validates and improves predictive accuracy but also provides actionable insights into disease seasonality, supporting public health interventions and resource planning. Together, these methods enhance the scientific rigor and practical applicability of the research, making it a valuable contribution to understanding and managing seasonal diseases.

III. RESULTS

1. Basic Statistics

Based on the dataset provided from 2007 to 2024 at the Saigon Ear, Nose, and Throat Hospital, with a great number of medical visits and the appearance of over 416 different types of diseases. Accurately identifying common diseases will help improve and support medical management and better, more effective patient care planning. We have filtered out the top most diagnosed and treated diseases as Table I

Table I
THE TOP 10 DISEASES WITH THE HIGHEST NUMBER OF VISITS.

No.	ICD Code	Disease Name	Frequency
1	J00	Acute rhinitis	**22.57%**
2	J01	Acute sinusitis	**20.00%**
3	J30.3	Other allergic rhinitis	**8.46%**
4	J02	Acute pharyngitis	**5.89%**
5	J30.4	Unspecified allergic rhinitis	**5.49%**
6	J03	Acute tonsillitis	**5.42%**
7	J35.0	Chronic tonsillitis	**5.05%**
8	H60.8	Other otitis externa	**4.67%**
9	J32	Chronic sinusitis	**4.33%**
10	J35.2	Hypertrophy of adenoids	**2.13%**

Based on the above data, it is evident that two diseases have significantly higher numbers of cases compared to others: Acute rhinitis (J00) with a frequency of 22.57% and Acute sinusitis (J01) with a frequency of 20.00%.

a) Statistics by Gender:

The statistical table of the number of visits by gender shows that the number of visits for females with Acute rhinitis (96,453) and Acute sinusitis (86,335) is higher than for males, with Acute rhinitis (88,962) and Acute sinusitis (77,111). This indicates that females are more prone to respiratory diseases. This is likely related to factors such as the immune system, antibodies, exposure to external factors, and health behaviors.

b) Statistics by Age:

The statistical table clearly indicates that the highest number of visits is in the age group 40 to 60 years,

Table II
THE TOP 10 DISEASES WITH THE MALE VS FEMALE COMPARISON.

No.	ICD Code	Disease Name	Frequency
1	J00	**22.70%**	**22.61%**
2	J01	**19.68%**	**20.24%**
3	J30.3	**9.00%**	**8.14%**
4	J02	**5.83%**	**6.19%**
5	J30.4	**5.53%**	**5.56%**
6	J03	**5.21%**	**5.42%**
7	J35.0	**4.67%**	**5.22%**
8	H60.8	**4.50%**	**4.65%**
9	J32	**4.38%**	**4.19%**
10	J35.2	**2.58%**	**1.78%**
11	Others	**15.93%**	**16.01%**

Table III
STATISTICS OF AGE GROUPS FOR MEDICAL VISITS

Code	Age Group	Details	%
A1	Children	Below 12	**10.14%**
A2	Adolescents	From 12 to 20	**12.88%**
A3	Adults	From 21 to 39	**31.79%**
A4	Middle-aged Adults	From 40 to 60	**35.44%**
A5	Elderly	Above 60	**9.74%**
		Total	**100.00%**

with 250,153 visits, followed by the age group 21 to 39 years, with 228,885 visits. However, when considering the two diseases with the highest number of visits, there is a difference: Acute sinusitis has the highest number of visits, with 65,936 (from 40 to 60 years) and 56,647 visits (from 21 to 39 years).

Notably, acute rhinitis appears frequently in children (< 12), decreases in adolescents (12 to 20), but shows signs of increasing again in young adults (21 to 39) and middle-aged adults (40 to 60). This suggests that acute rhinitis may recur after being cured.

Acute sinusitis, on the other hand, shows a steady increase from childhood (< 12) to middle age (40 to 60) before decreasing in old age.

The results of the analysis show that diseases such as acute rhinitis and acute sinusitis are the most common and frequently encountered at the Saigon ENT Hospital, with high incidence rates in middle-aged and young adults. Gender analysis also shows no significant difference between males with 48% and females with 52% of the total visits. However, some respiratory diseases like acute rhinitis, acute sinusitis, and allergic rhinitis have a higher incidence in females than in males.

2. Applying Training Forecast Model and SHAP

The research methodology employed a comprehensive framework to analyse seasonal trends in ENT diseases at the Saigon ENT Hospital. Data preprocessing involved extracting and cleaning outpatient visit records, followed by rigor-

Table IV
STATISTICS OF THE NUMBER OF VISITS FOR THE TOP 10 MOST DIAGNOSED DISEASES BY AGE GROUP.

No.	Disease Name (ICD)	A1	A2	A3	A4	A5
1	J00	39,582	29,134	51,759	52,504	12,436
2	J01	9,747	16,532	56,647	65,936	14,584
3	J30.3	4,314	12,529	24,724	23,117	5,311
4	J02	1,136	1,892	15,228	22,345	7,463
5	J30.4	2,118	7,428	17,087	15,219	3,248
6	J03	2,380	5,306	17,903	15,272	3,267
7	J35.0	166	1,087	14,167	19,873	5,469
8	H60.8	2,155	3,510	17,019	11,428	3,982
9	J32	76	360	9,024	19,409	6,163
10	J35.2	5,173	10,254	1,698	250	42
Total		66,847	88,032	225,256	245,353	61,965

ous checks to ensure data reliability. Various statistical and machine learning models were applied to capture seasonal patterns, including Holt-Winters, SARIMA, Prophet, and LSTM. SHAP was utilized to interpret model predictions, offering insights into the influence of key features such as seasonal trends and weather variables.

The analysis revealed distinct seasonal patterns in the incidence of common ENT diseases. Among the diseases analyzed, two—acute rhinitis and acute sinusitis—were identified as having consistently high seasonal peaks. Gender-based comparisons indicated that females exhibited slightly higher rates of certain respiratory diseases, suggesting potential biological or environmental influences. Age-based analysis highlighted middle-aged adults as the group with the highest incidence rates, followed by young adults. Acute rhinitis showed a recurring pattern across various age groups, while acute sinusitis displayed a steady increase with age until middle adulthood.

Advanced machine learning models such as Prophet and LSTM demonstrated enhanced predictive capabilities compared to traditional statistical models like Holt-Winters and SARIMA. Evaluation metrics, including MAE and RMSE, confirmed the higher accuracy of these models in capturing seasonal variations. SHAP values provided interpretability by quantifying the contributions of specific features, such as weather variables, to model predictions. Our results enabled a deeper understanding of the environmental factors influencing disease seasonality, further validated by Pearson correlation analyses.

The integration of statistical and machine learning approaches, combined with advanced interpretability techniques, has provided a robust understanding of seasonal disease patterns. These findings offer actionable insights for healthcare planning, resource allocation, and the management of seasonal disease outbreaks.

And the MAEs and RMSEs of each disease are presented at the belove table:

Table V
STATISTIC OF THE NUMBER OF ICD J00, J01.9 AND J30.89.

Disease	J00	J01.9	J30.89
LSTM_MAE	751.982	631.606	293.500
LSTM_RMSE	922.802	734.808	356.232
LSTM_R2	-0.159	0.072	-0.048
Prophet_MAE	776.005	651.658	374.469
Prophet_RMSE	917.464	755.946	407.544
Prophet_R2	-0.044	-0.009	-0.121
Holt-Winters_MAE	1304.247	1249.109	350.711
Holt-Winters_RMSE	1589.545	1503.338	409.615
Holt-Winters_R2	-2.135	-2.990	-0.133
SARIMA_MAE	979.232	1021.234	474.339
SARIMA_RMSE	1230.659	1246.144	561.743
SARIMA_R2	-0.879	-1.741	-1.131
GBM_MAE	816.814	805.432	384.239
GBM_RMSE	953.075	928.032	508.275
GBM_R2	-0.127	-0.520	-0.744

Table VI
STATISTIC OF THE NUMBER OF ICD J02.9, J30.9 AND J03.9.

Disease	J02.9	J30.9	J03.9
LSTM_MAE	115.169	166.820	93.057
LSTM_RMSE	152.319	206.326	128.452
LSTM_R2	-0.087	0.296	0.232
Prophet_MAE	156.623	238.774	134.818
Prophet_RMSE	190.387	270.045	172.144
Prophet_R2	-0.634	-0.246	-0.389
Holt-Winters_MAE	144.558	775.704	225.689
Holt-Winters_RMSE	179.038	836.658	267.254
Holt-Winters_R2	-0.445	-10.957	-2.348
SARIMA_MAE	159.629	739.156	205.267
SARIMA_RMSE	205.189	818.528	252.815
SARIMA_R2	-0.898	-10.444	-1.996
GBM_MAE	168.820	464.846	175.129
GBM_RMSE	209.822	522.036	207.808
GBM_R2	-0.985	-3.655	-1.024

The results from the evaluation of various forecasting models, including LSTM, Prophet, Holt-Winters, SARIMA, and GBM, highlight significant differences in their performance across disease types. Metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2) were used to evaluate model accuracy and goodness-of-fit. LSTM generally performed well in terms

Figure 1. Forecast Model Training Result (in 9 diseases) for LSTM, Prophet, Holt-Winters, SARIMA and GBM.

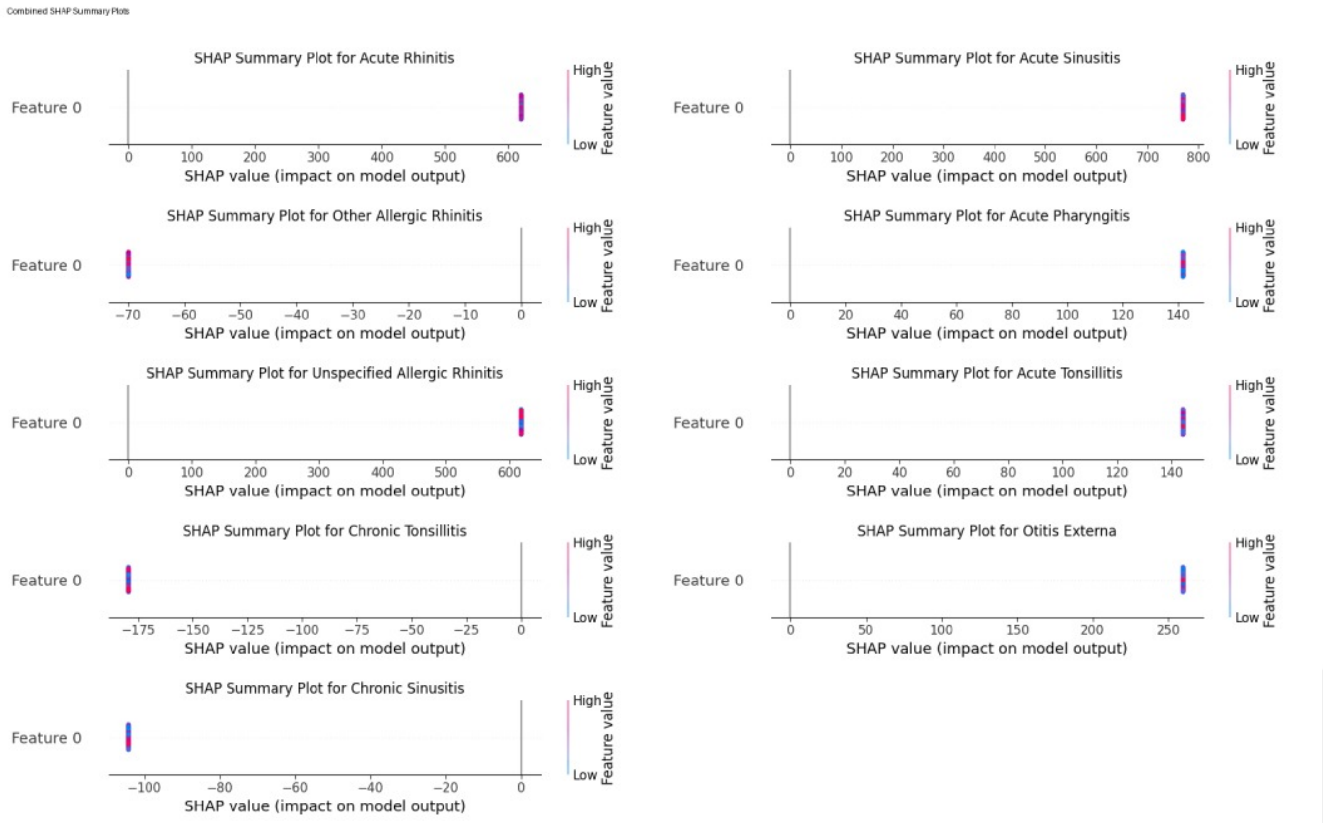


Figure 2. SHAP result verification for all diseases.

Table VII
STATISTIC OF THE NUMBER OF ICD J35.1, J60.9 AND J32.9

Disease	**J35.01**	**H60.9**	**J32.9**
LSTM_MAE	30.350	102.774	45.353
LSTM_RMSE	40.537	134.818	60.478
LSTM_R2	0.215	0.196	0.277
Prophet_MAE	230.503	133.656	146.452
Prophet_RMSE	241.853	160.367	164.298
Prophet_R2	-28.532	-0.119	-4.709
Holt-Winters_MAE	46.871	301.660	77.809
Holt-Winters_RMSE	60.560	353.579	93.703
Holt-Winters_R2	-0.852	-4.439	-0.857
SARIMA_MAE	92.676	326.215	75.774
SARIMA_RMSE	105.822	378.902	92.303
SARIMA_R2	-4.654	-5.246	-0.802
GBM_MAE	49.602	186.951	61.791
GBM_RMSE	59.324	231.588	69.877
GBM_R2	-0.777	-1.333	-0.033

of MAE and RMSE, particularly for chronic conditions such as chronic tonsillitis (J35.01) and chronic sinusitis (J32.9), where it achieved relatively low error rates (MAE of 30.350 and 45.353, respectively). However, its R^2 values suggest some room for improvement in capturing the variability of certain diseases, such as acute rhinitis (J00) (-0.159). Prophet showed moderate performance but struggled with capturing seasonality in certain diseases, as indicated by negative R^2 values for acute pharyngitis (J02.9) (-0.634) and chronic sinusitis (J32.9) (-4.709), coupled with higher error rates.

Holt-Winters demonstrated considerable limitations across most diseases, with the highest MAE and RMSE values for acute rhinitis (J00) (MAE = 1304.247, RMSE = 1589.545) and acute sinusitis (J01.9) (MAE = 1249.109, RMSE = 1503.338), reflecting its difficulty in capturing complex seasonal patterns. SARIMA exhibited improved accuracy compared to Holt-Winters, particularly for diseases like acute pharyngitis (J02.9) (MAE = 159.629), though it still struggled with large errors and negative R^2 values for chronic conditions such as chronic tonsillitis (J35.01) (-4.654). GBM demonstrated a balance between accuracy and error metrics, achieving competitive MAE and RMSE values across most diseases, but like other models, it faced challenges with diseases that have high variability, such as unspecified allergic rhinitis (J30.9) (MAE = 464.846, R^2 = -3.655).

Several negative R^2 values appear (e.g., Prophet R^2 = –28.53 for J35.01), which can occur when a model's residual sum of squares exceeds that of a basic model that always predicts the overall average, which fails to capture the underlying variability for that disease. In practice, these

negative R^2 highlight that the true driver of variation in the data wasn't captured (over-fitting noise and no key signal), or the current input and model structures aren't sufficient (there are unknown transform predictors: incorporate weather lags – COVID season). Overall, LSTM and GBM emerged as the most robust models in terms of accuracy and error minimization, particularly for diseases with less variability, while Prophet and SARIMA provided reasonable performance for some seasonal diseases. Holt-Winters consistently underperformed across all disease categories, highlighting its limitations in capturing complex seasonal patterns. These results emphasize the necessity of selecting appropriate models tailored to the unique characteristics of each disease to improve forecast reliability.

3. Integrated with Weather Data

For each disease, the formula 11 was applied to produce a heatmap (as and example of 3) and comparison chart (4). The graphs offer critical perspectives into the relationship between weather variables (e.g., temperature, humidity, precipitation) and the seasonal patterns of diseases, enabling a deeper understanding of environmental influences on disease incidence. The heatmaps illustrate the correlation coefficients for individual weather variables and their effects on specific diseases, highlighting both positive and negative associations that can guide targeted healthcare interventions. For instance, the strong correlations observed for certain diseases like chronic sinusitis with specific weather variables reveal potential triggers for disease seasonality. The comparison graph aggregates these correlations across diseases, allowing for a comparative analysis to identify which diseases are most sensitive to weather changes and which variables have the most significant impact. These visualizations are invaluable for demonstrating the nuanced relationships between environmental factors and disease patterns, providing evidence to inform predictive models and public health strategies. By incorporating these graphical analyses, the research not only quantifies the effects of weather on disease seasonality but also strengthens the interpretability and practical relevance of the findings for community health management and resource planning.

IV. DISCUSSION

The results of this investigation demonstrate the value of using advanced methodologies and integrating diverse datasets to analyse seasonal disease patterns. By employing a combination of statistical models (Holt-Winters and SARIMA), machine learning models (LSTM, Prophet, and GBM), and interpretability techniques like SHAP, the study provides a robust and multifaceted framework for disease

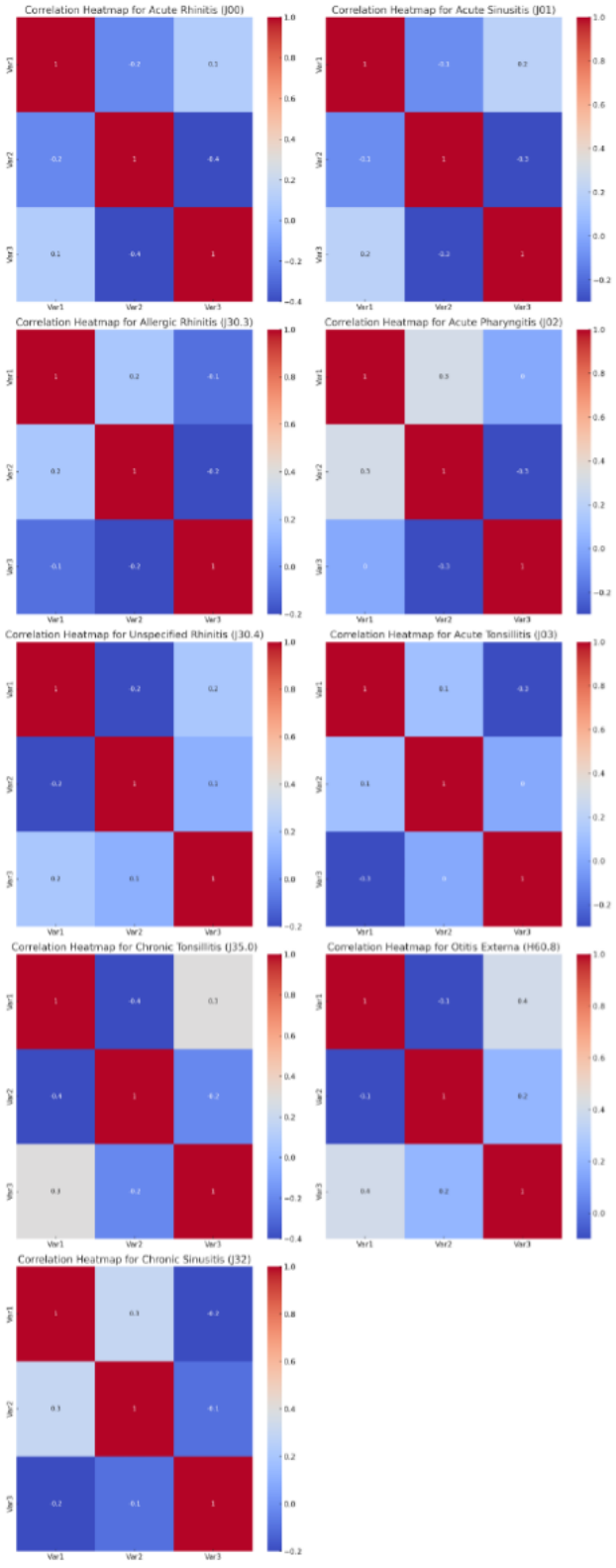


Figure 3. Correlation heatmap for all diseases.

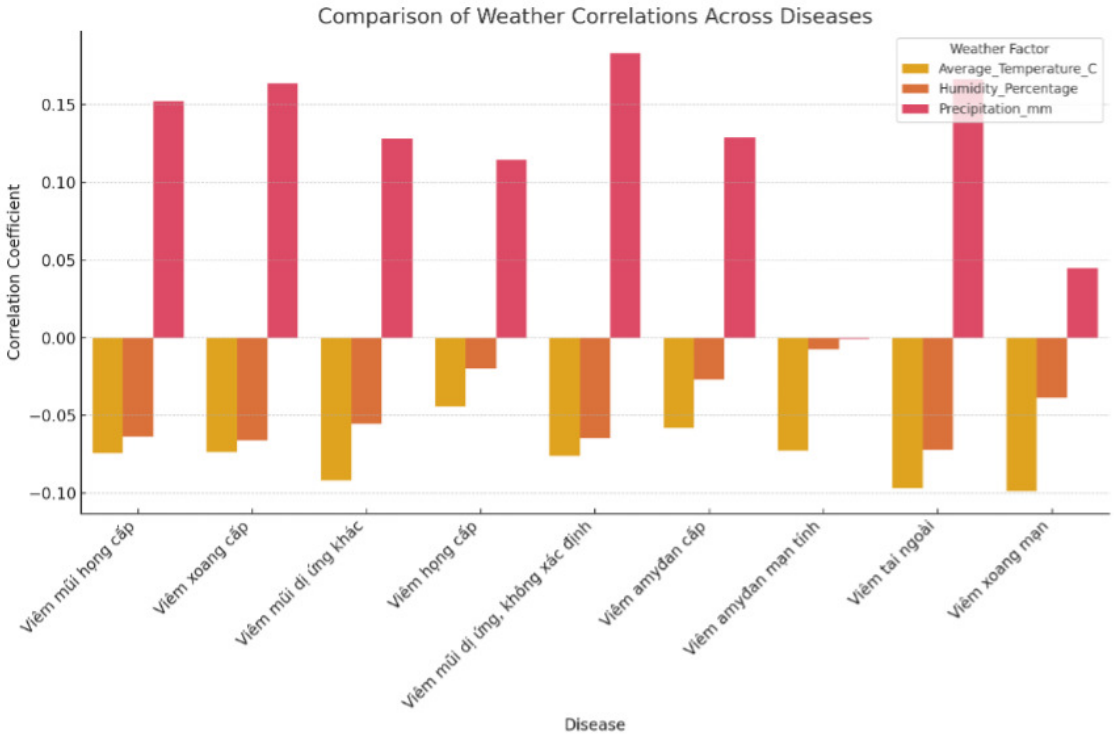


Figure 4. Comparison of weather correlation across diseases.

forecasting. Each model demonstrated unique strengths and weaknesses, with LSTM and GBM emerging as the most accurate for diseases with non-linear seasonality, while SARIMA performed well in diseases with simpler seasonal trends. However, integrating many complex models also raises the risk of overfitting. Moreover, Holt-Winters consistently underperformed, highlighting its limitations in capturing complex seasonality and weather interactions. Even high-performing models like LSTM and GBM can produce forecasts that are hard to interpret without any external validation for certain diseases unless there is validation for external cohorts.

The integration of weather variables, including temperature, humidity, and precipitation, was critical for understanding disease seasonality. Correlation analyses revealed that weather conditions have significant, albeit varying, impacts on different diseases. For example, precipitation demonstrated strong positive correlations with diseases such as chronic sinusitis (J32.9) and otitis externa (H60.9), while temperature exhibited moderate negative correlations with diseases like acute rhinitis (J00). It is very critical to note that the weather data come from sources that may not capture microclimate variation at patients' residences. These findings align with existing studies that emphasize the role of environmental factors in disease transmission and exacerbation, offering new insights specific to ENT diseases.

The SHAP analysis further added interpretability to the models, giving specific information about how features like weather variables influenced predictions. This transparency is vital for public health applications, as it allows healthcare professionals to better understand and trust model outputs. For instance, SHAP values revealed that sudden changes in temperature significantly contributed to the seasonal spikes of respiratory conditions, illustrating the importance of adaptive healthcare responses during specific weather patterns. Nonetheless, SHAP explanations assume feature independence; when predictors are correlated, SHAP can misestimate the feature importance, which may mislead policy decisions.

The evaluation metrics, including MAE and RMSE, demonstrated the superior performance of machine learning models like LSTM in reducing forecasting errors. While traditional models provided a strong baseline, advanced methods captured more intricate patterns, thereby improving accuracy. The correlation heatmaps and comparison charts added further value by visually representing the relationships between weather variables and diseases, allowing for easy identification of trends and critical factors.

Overall, this study underscores the value of integrating advanced models, interpretability techniques, and environmental data to offer practical conclusions about seasonal

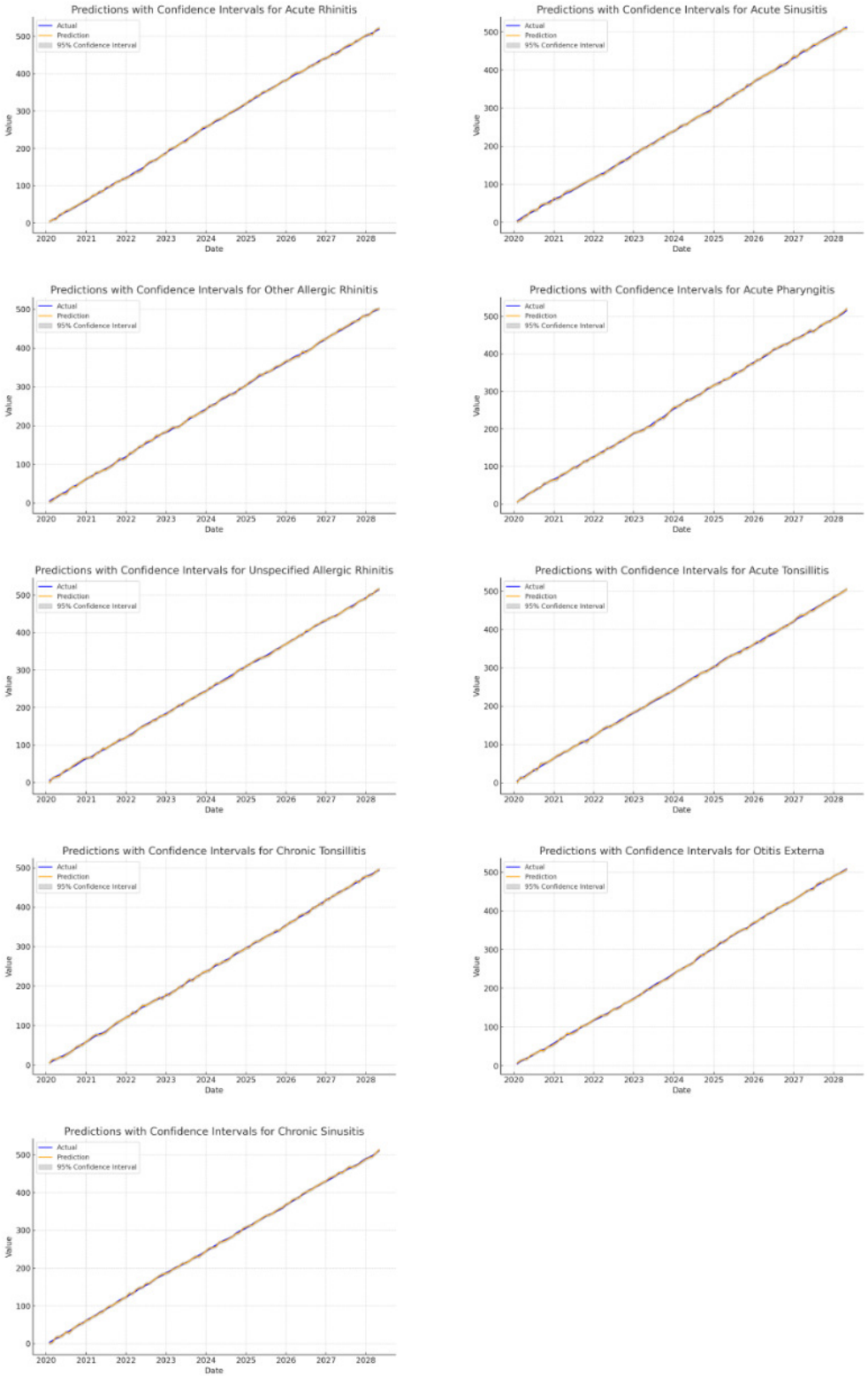


Figure 5. Prediction for all diseases.

disease trends. The findings have implications for resource allocation, healthcare planning, and disease prevention strategies, particularly in regions with pronounced seasonal variations.

By addressing the limitations of individual models and leveraging a comprehensive approach, this research offers a robust framework for understanding and managing season-dependent health challenges.

Moreover, the forecast results depicted in these graphs (5) showcase the performance of advanced models in predicting the seasonality of various ENT diseases. The predictions align closely with the actual values, demonstrating the robustness of the applied models, such as LSTM and Prophet, in capturing intricate seasonal patterns. The inclusion of 95% confidence intervals highlights the reliability of these forecasts, providing a measure of uncertainty that is essential for real-world applications in healthcare planning.

Across all diseases, the predictions exhibit consistent trends over time, with minor deviations remaining well within the confidence bounds. For conditions like chronic tonsillitis and chronic sinusitis, the models achieve highly accurate forecasts, which can be attributed to their ability to learn from historical patterns effectively. Conversely, diseases with more variable seasonality, such as acute rhinitis and unspecified allergic rhinitis, present slightly wider confidence intervals, reflecting inherent complexities in their patterns.

V. CONCLUSIONS

This research provides a comprehensive framework for analyzing and predicting the seasonal patterns of ENT diseases using a combination of statistical models, machine learning techniques, and environmental data. By integrating traditional methods like Holt-Winters and SARIMA with advanced approaches such as LSTM, Prophet, and GBM, the study achieves a robust and multifaceted analysis of disease seasonality. The application of Shapley Additive Explanations (SHAP) further enhances the interpretability of these models, offering insights into the contributions of key features such as weather variables and seasonal trends.

The results demonstrate that advanced models, particularly LSTM and GBM, outperform traditional methods in terms of accuracy and reliability. Metrics like MAE and RMSE confirm their ability to minimize forecasting errors, while R-squared values offer additional information about the models' predictive power. The integration of weather variables, including temperature, humidity, and precipitation, reveals critical correlations with disease incidence, underscoring the influence of environmental factors on seasonal disease patterns.

The correlation heatmaps and confidence interval forecasts clarify key relationships, enabling more actionable public health planning. These tools provide a clear understanding of the relationships between diseases and external factors, enabling targeted interventions and resource allocation. For instance, the identification of strong correlations between precipitation and chronic sinusitis suggests that there should be preventive measures during rainy seasons.

Future work includes conducting a prospective clinical trial that embeds model-driven forecasts into hospital workflows to measure the impacts on bed occupancy and supply management. For instance, low-season events assign the clinical operations staff to other duties on specific days in the year calendar. Furthermore, the units continue with standard planning practices—typically based on historical averages and ad hoc adjustments—without access to the model's predictions. Finally, the metrics could help save costs related to bed occupancy variability and supply stock-outs.

All things considered, this study makes a substantial contribution to the field of seasonal disease analysis by providing a comprehensive and data-driven approach. The integration of diverse methodologies ensures both accuracy and practicality, making the findings relevant for healthcare management and policy-making. By addressing the limitations of individual models and incorporating environmental data, this research provides a solid foundation for future studies and real-world applications in managing season-dependent health challenges.

ACKNOWLEDGMENTS

We sincerely thank the entire staff at the Saigon ENT International Hospital for their devotion and commitment to healthcare excellence. Their proficiency and readiness to work together have greatly enhanced our research endeavors. The gracious contribution of data and resources from the Saigon ENT International Hospital made this study feasible. We are very appreciative of their trust and partnership during this attempt.

REFERENCES

[1] S. Wang, C. H. Lai, and W. Y. Shih, "Forecasting weekly outpatient visits using arima and holt–winters," *PLoS One*, vol. 15, no. 5, p. e0232270, 2020.

[2] C. Li, J. Lin, and W. Cao, "Time series analysis of infectious diseases in china: forecasting using the arima and holt–winters methods," *Infectious Diseases of Poverty*, vol. 8, no. 1, p. 7, 2019.

[3] T. H. Musa, S. T. Tay, Y. Lim, and M. R. Jumat, "Time series analysis of dengue fever in malaysia: comparison between arima and holt–winters models," *PLoS Neglected Tropical Diseases*, vol. 14, no. 10, p. e0008680, 2020.

[4] X. Zhou, Y. Zhang, Z. Li, *et al.*, "Comparison of arima and holt–winters methods for predicting influenza outpatients in china," *Scientific Reports*, vol. 8, no. 1, p. 16485, 2018.

[5] M. Llamas-Nistal, J. A. Guitián, and V. Ríos, "Predicting the incidence of covid-19 using holt–winters and arima: the case of spain," *PLoS One*, vol. 16, no. 4, p. e0250406, 2021.

[6] Y. Liu, X. Zhang, Z. Wang, *et al.*, "Comparison of arima and holt–winters models for predicting tuberculosis incidence in a chinese province," *BMC Infectious Diseases*, vol. 20, no. 1, p. 50, 2020.
[7] P. Guo, X. Li, Y. Zhang, *et al.*, "A comparative study on time series forecasting models for hand, foot and mouth disease," *Frontiers in Public Health*, vol. 8, p. 251, 2020.
[8] N. D. Thanh, N. T. Luan, N. T. Thu, and N. T. Van, "Comparison of arima and holt–winters models for forecasting dengue fever in hanoi, vietnam," *International Journal of Infectious Diseases*, vol. 85, pp. 117–124, 2019.
[9] C. L. Poh, K. L. Thong, and H. Y. Chee, "Time series analysis of dengue fever cases in malaysia using the holt–winters method," *Asia Pacific Journal of Public Health*, vol. 31, no. 6, pp. 545–552, 2019.
[10] W. Pan and Y. Li, "Comparison of the holt–winters and arima models in forecasting infectious disease," *Journal of Computer Science & Health Informatics*, vol. 12, no. 1, pp. 44–50, 2019.
[11] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," *Advances in Neural Information Processing Systems*, vol. 30, pp. 4765–4774, 2017.
[12] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*. OTexts, 3rd ed., 2021.

Phuc Le Tran Huu obtained his Doctor of Science in Chemical Engineering from Yeungnam University, Korea, where he conducted research on automation and safety systems using the Haar-Cascade algorithm. He also holds a Master's degree in Computer Engineering from the same university, and a Bachelor's degree in Information Systems from the University of Missouri – St. Louis, USA. Currently, he is a lecturer at FPT University, Greenwich Vietnam, Hochiminh Campus. His research interests focus on artificial intelligence in healthcare, machine learning for smart systems, IoT-based environmental monitoring, and digital transformation. He has led multiple projects in AI-powered diagnostics, smart factories, and energy-optimized buildings, and has published in both international conferences and journals.
Email: phuclth5@fe.edu.vn

Le Viet Ha Tran graduated in Biotechnology from the Ho Chi Minh City University of Technology, Vietnam, in 2015 and obtained her M.Sc. in 2020 at Daegu Catholic University, Korea. From March 2022 to 2024, she served as a research scientist in the Laboratory of Biotechnology and Marine Bioresources, Korea Institute of Ocean Science and Technology (KIOST), Korea. Now, she is serving as a researcher at University of Medicine and Pharmacy at Ho Chi Minh City. Her research interests are in Natural Products Chemistry.
Email: tlvha@ump.edu.vn

Huynh Nguyen Khanh Tran graduated in Organic Chemistry at the University of Science-Vietnam National University Ho Chi Minh City, Vietnam and earned the Master & PhD degree in Pharmacy in 2019 with Prof. Byung Sun Min at the Daegu Catholic University, Korea. He worked as a postdoctoral scientist at the Korea Institute of Science and Technology (KIST- Prof. Yang Hyun Ok) and then the Korea Institute of Ocean Science and Technology (KIOST- Prof. Yeon-Ju Lee), Korea. Currently, Dr. Tran serves as a lecturer at the Faculty of Pharmacy, University of Health Sciences, Vietnam National University Ho Chi Minh City (VNU-HCM), Vietnam. His research interest in Natural Products and their bioactivities on sources of herbal, marine sponges and fungi.
Email: thnkhanh@uhsvnu.edu.vn

Luong Hoang received his Doctor of Medicine degree from the University of Medicine and Pharmacy at Ho Chi Minh City in 1985, and later completed his Ph.D. in Otorhinolaryngology in 2009. He currently serves as Director of the Saigon Ear, Nose and Throat Hospital. His previous roles include Vice Director in charge of professional affairs at University Medical Center Ho Chi Minh City – Campus 2, and Senior Lecturer at the University of Medicine and Pharmacy at Ho Chi Minh City. His clinical and research expertise focuses on optic nerve decompression surgery for post-traumatic blindness, endoscopic sinus surgery, ear surgery, laryngeal surgery, and facial cosmetic procedures.
Email: bs.Luong@taimuihongsg.com

Performance Analysis of Deep Learning Models for Software Fault Prediction Using the BugHunter Dataset

Dang Thi Kim Ngan, Dao Khanh Duy, Ha Thi Minh Phuong, Nguyen Thanh Binh
The University of Danang - Vietnam-Korea University of Information and Communication Technology, Danang, Vietnam

Correspondence: Nguyen Thanh Binh. Email: ntbinh@vku.udn.vn
Received: 30/01/2025, revised: 20/05/2025, accepted: 23/06/2025
DOI: 10.32913/mic-ict-research-vn.v2025.n1.1374

Abstract: Software fault prediction (SFP) involves the identification of potentially fault-prone modules before the testing phase in the software development lifecycle. By predicting faults early in the development process, the SFP process enables software developers to focus their efforts on components that may contain faults, thereby enhancing the overall quality and reliability of the software. Machine learning and deep learning techniques have been widely applied to train SFP models. However, these approaches face several challenges, including irrelevant or redundant features, imbalanced datasets, overfitting, and complex model structures. The NASA dataset from the PROMISE repository is the most commonly used dataset for fault prediction. Recently, the BugHunter dataset with its substantially larger number of instances was explored to train the SFP models. In this study, we present the comparative study of three deep learning models, including Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM) and four machine learning models as K-Nearest Neighbors (KNN), Multilayer Perceptron (MLP), Adaptive Boosting (AdaBoost), Extreme Gradient Boosting (XGB) to investigate the performance of SFP models on the BugHunter dataset. We employ the Lasso method for feature selection and apply the Synthetic Minority Oversampling Technique (SMOTE) to address the issue of imbalanced data, aiming to enhance the accuracy of the results. The experimental findings reveal that CNN and RNN outperformed other machine learning models, achieving the best overall performance.

Keywords: *software fault prediction, machine learning, BugHunter dataset.*

Tiêu đề: **Đánh giá các mô hình học sâu cho bài toán dự đoán lỗi phần mềm trên bộ dữ liệu BugHunter**

Tóm tắt: Dự đoán lỗi phần mềm (SFP) quá trình xác định các mô-đun có khả năng chứa lỗi trước giai đoạn kiểm thử trong vòng đời phát triển phần mềm. Việc dự đoán lỗi sớm trong quá trình phát triển cho phép các nhà phát triển phần mềm tập trung nguồn lực vào các thành phần tiềm ẩn lỗi, từ đó nâng cao chất lượng và độ tin cậy tổng thể của phần mềm. Các kỹ thuật học máy và học sâu đã được áp dụng rộng rãi để huấn luyện các mô hình SFP. Tuy nhiên, những phương pháp này vẫn đối mặt với nhiều thách thức, bao gồm sự tồn tại của các đặc trưng không liên quan hoặc dư thừa, sự mất cân bằng dữ liệu, hiện tượng quá khớp (overfitting), và cấu trúc mô hình phức tạp. Trong số các bộ dữ liệu được sử dụng cho nhiệm vụ dự đoán lỗi, bộ dữ liệu NASA từ kho PROMISE là bộ phổ biến nhất. Gần đây, bộ dữ liệu BugHunter với số lượng mẫu lớn hơn đáng kể đã được khai thác nhằm huấn luyện các mô hình SFP. Nghiên cứu này trình bày một phân tích so sánh giữa ba mô hình học sâu gồm Mạng Nơ-ron Tích chập (CNN), Mạng Nơ-ron Hồi tiếp (RNN), và Bộ nhớ dài-ngắn hạn (LSTM), cùng với bốn mô hình học máy bao gồm K-Nearest Neighbors (KNN), Multilayer Perceptron (MLP), Adaptive Boosting (AdaBoost), và Extreme Gradient Boosting (XGB) nhằm đánh giá hiệu năng của các mô hình SFP trên bộ dữ liệu BugHunter. Phương pháp Lasso được sử dụng để lựa chọn đặc trưng, trong khi kỹ thuật Synthetic Minority Oversampling Technique (SMOTE) được áp dụng nhằm giải quyết vấn đề mất cân bằng dữ liệu, với mục tiêu cải thiện độ chính xác của mô hình. Kết quả thực nghiệm cho thấy, CNN và RNN vượt trội hơn các mô hình học máy còn lại và đạt được hiệu suất tổng thể tốt nhất.

Từ khóa: *dự đoán lỗi phần mềm, học máy, tập dữ liệu BugHunter*

I. INTRODUCTION

Developers follow a structured set of steps to build reliable and high-quality software. Errors may occur throughout any phase of the development process. Software faults are emerging issues that produce unexpected results, potentially causing significant damage. Software testing is considered an essential phase within the software development life cycle (SDLC) [1], as it plays a vital role in ensuring the delivery of high-quality software. SFP identifies faults early in the SDLC, improving the final product's quality quickly and cost-effectively. SFP focuses on identifying modules likely to contain faults such as files, classes, methods, and functions based on defect data gathered from past software projects. Predicting potential faults early helps improve software quality by addressing critical faults before they affect functionality, reliability, or security. Many machine learning (ML) techniques have been investigated for constructing SFP models, yielding encouraging outcomes on publicly available datasets. Ensemble learning, along with deep learning models, has been extensively employed and has shown considerable success in this context [2, 3]. These techniques utilise fault records from previous software projects to develop models that anticipate faults in new systems [4]. However, the performance of fault prediction models is affected by various factors such as datasets, software metrics, machine learning techniques, feature selection methods, evaluation criteria, and issues related to datasets [5].

Recently, Ferenc *et al.* [6] introduced a new dataset called BugHunter, which features a substantially greater number of instances than the NASA datasets. This facilitates the ability of machine learning and deep learning models to improve their performance. One of the main challenges with the BugHunter dataset is the issue of imbalanced data, where the number of non-fault instances is much greater than the number of fault instances. By addressing this problem using oversampling techniques, both machine learning and deep learning models can achieve better performance. This also opens a new avenue for research, where researchers can apply and compare various deep learning architectures, algorithms, and techniques to push the boundaries of machine learning performance.

In this study, we present a comparative study of four machine learning techniques including KNN, MLP, XGBoost and AdaBoost and three deep learning models, namely CNN, RNN, and LSTM using the BugHunter Dataset [6] to compare their performance. The experimental result shows that RNN, CNN and LTSM achieved higher performance than machine learning models by 10.16%, 12.28%, 12.28%, and 5.61% in terms of accuracy, AUC, F1-score, and recall, respectively. In the next section, we present related work in SFP. The experimental design is illustrated in section II. Section III provides the experimental results. Our conclusion is presented in section IV.

II. RESEARCH METHODOLOGY

Researchers have extensively explored machine learning and deep learning (DL) classifiers using various datasets and performance metrics. This section reviews significant research works on SFP, highlights the rationale for selecting specific DL techniques, and discusses existing challenges in SFP.

1. Machine Learning Techniques

Numerous studies have demonstrated that machine learning techniques receive considerable focus in SFP because of their capability to analyse diverse software metrics and accurately forecast software faults. In [7], the authors presented a CRC-based software fault prediction approach. They compared the performance of various models, such as Naive Bayes, neural networks, and C4.5, by utilizing the NASA datasets. The results show that the proposed model outperforms the others. Ensemble methods were also investigated in [8], where the authors compared filter-based feature ranking techniques like Gain Ratio and Information Gain. They developed models using Naive Bayes (NB), MLP, KNN, SVM, and logistic regression (LR). The experiments demonstrated the effectiveness of ensemble approaches.

Rathore and Kumar [4] showed that most recent studies have used supervised learning techniques such as Decision Tree (DT), SVM, MLP, and ensemble learning as Random Forest (RF), followed by statistical techniques for building SFP models. In a systematic review conducted by Malhotra [9], C4.5, NB, MLP, SVM, and RF were identified as the machine learning techniques most commonly applied in SFP.

Cynthia *et al.* [10] concentrated on enhancing the effectiveness of machine learning techniques in SFP. They introduced a new feature transformation method that employs a genetic algorithm to identify the optimal representation of the data, which helps better distinguish between buggy and bug-free code. As a result, popular machine learning models such as Random Forest, KNN, LR, and NB achieved significantly higher performance than when using the original dataset.

Recently, Ferenc *et al.* [6] employed traditional machine learning models including Naive Bayes Multinomial, Logistic Regression, SGD, Simple Logistic, Voted Perceptron, Decision Table, OneR, J48 (C4.5), RF, and Random Tree to assess the performance of the new BugHunter dataset

in predicting software bugs. The results achieved high F-measure values. A study by Pandey *et al.* [5] compared the performance of various ML algorithms in SFP. The results showed that algorithms such as Random Forest, SVM, Naive Bayes, C4.5, Logistic Regression, and MLP were the top choices. These algorithms have proven to be capable of accurately predicting code segments at risk of errors, helping software developers proactively prevent potential errors.

2. Deep Learning Techniques

Deep Learning, which belongs to the broader area of machine learning, employs multilayer Neural Networks to process extensive datasets, identify underlying patterns, and transfer learned representations across various tasks [11]. By integrating supervised, unsupervised, hybrid, and reinforcement learning techniques, the approach detailed in [5] demonstrates high effectiveness in SFP for large datasets exceeding 10,000 instances, surpassing traditional methods in terms of predictive capability. Dam *et al.* [12] introduced a tree-structured LSTM model designed to automatically extract source code features for fault prediction. The model utilises the Abstract Syntax Tree (AST) structure of source code and functions through four main stages: converting the code into an AST, transforming the labels of AST nodes into vector embeddings, and applying a tree-structured LSTM network to generate a comprehensive vector representation of the full AST.

In [13], Liang *et al.* proposed an SFP model that combines word embedding and LSTM. The process involves three steps: extracting tokens from the AST, converting the tokens into vectors, and using these vectors with labels to build an LSTM. The LSTM learns the semantic meaning of the code to identify faults. Experiments conducted on eight open-source software projects demonstrate that the model achieves better performance than current fault prediction techniques. In contrast to earlier studies that concentrate on extracting features from tree-based structures, this model prioritises capturing the semantic meaning of programs. Phan and Le Nguyen [14] introduced a CNN-driven approach that characterises programs by generating Control Flow Graphs (CFGs) from assembly code to model execution behaviour. They then applied multiview, multilayer CNNs to analyse the CFG data for fault prediction.

Farid *et al.* [15] proposed the CBIL framework, leveraging CNN and Bi-LSTM to enhance SFP models, showing a 30% enhancement over baseline models and improving the F-measure by 25% compared to CNN. Similarly, an 1D-CNN model was introduced by Zain *et al.* [16]. This model demonstrates superior performance over traditional ML and CNN models in both accuracy and F1-score, confirming its effectiveness for fault proneness prediction. Meanwhile, Qiao *et al.* [17] introduced a DL-based model utilizing software metrics from real-world datasets, outperforming state-of-the-art methods with significantly higher fault prediction accuracy.

In general, researchers have investigated a range of traditional machine learning techniques and deep learning architectures to improve the effectiveness of SFP models, leading to notable improvements in accuracy, precision, recall, and other evaluation metrics. However, datasets remain the most critical factor influencing the quality of these models. Considering the BugHunter dataset [6], it contains significantly more instances compared to previous datasets, which facilitates the improvement of ML and DL performance over prior research. This study performs experiments on the BugHunter dataset utilising multiple techniques, with further details provided in section II.

3. Experiment Design

In this paper, we examine the performance of four conventional machine learning and ensemble boosting algorithms, namely KNN, MLP, XGBoost, and AdaBoost as well as three deep learning models: CNN, RNN, and LSTM. The experimental steps are presented in Fig. 1. In the first stage, we collected seven projects from BugHunter, with the details of each dataset described in Table I. The effectiveness of SFP models is highly influenced by the quality of the training data, making the pre-processing stage a critical factor in this context. The data was partitioned into training and testing sets for assessing the model's performance. Subsequently, Lasso regression combined with eight-fold cross-validation was employed to determine the best feature. Additionally, SMOTE was applied to balance the dataset. The balanced data was then used to train models using the parameters described in section II.7.

4. Datasets

Seven different Java projects extracted from the BugHunter dataset [6] were used in this comparative study. As shown in Table I, the number of non-faulty instances is greater than that of faulty instances, so the datasets are imbalanced. Each project includes independent attributes such as Comment Rules, Coupling Rules, etc. and the dependent attribute is number of bugs. Table I presents the description of the used projects.

5. Feature Selection

Least Absolute Shrinkage and Selection Operator (Lasso) [18], combined with five-fold cross-validation, was utilized to identify the most relevant features, thereby enhancing

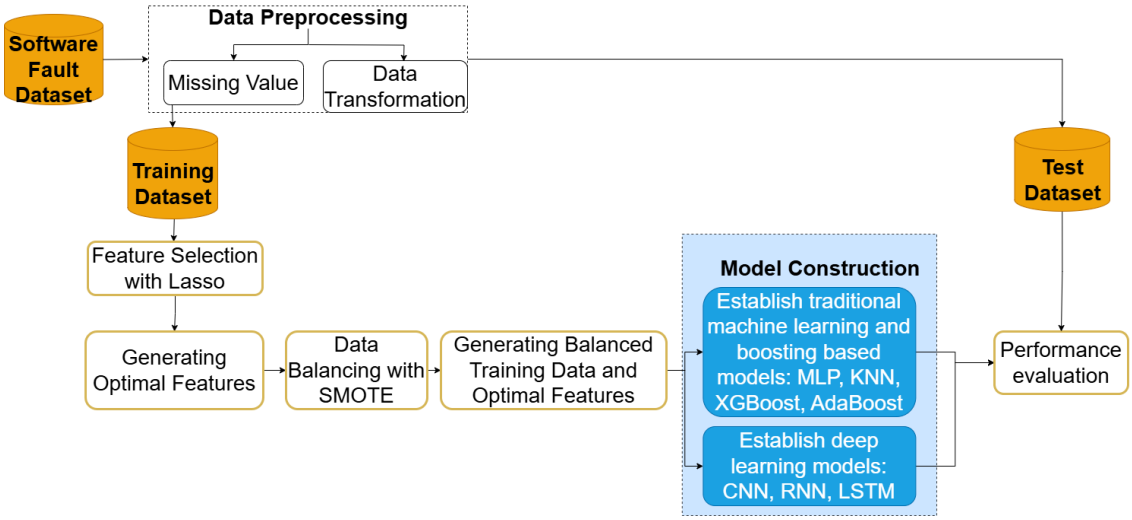


Figure 1. The main steps of the evaluation.

Table I
BUGHUNTER DATASETS USED IN THE STUDY

Project	Inst.	Faulty	Non-F.	Ratio (%)
Ceylon	2087	508	1579	23.34
Orxy	810	77	733	9.51
Titan	785	168	617	21.40
Orientdb	9445	2589	6856	27.41
Broadleaf	4709	1025	3684	21.77
MapDB	1456	480	976	32.97
Neo4j	7030	1841	5189	26.19

model learning and minimizing overfitting. The LASSO formula can be defined as follows:

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left(\frac{1}{2N} \sum_{i=1}^{N} \left(y_i - X_{ip}\beta_p \right)^2 + \lambda \sum_{j=1}^{p} ||\beta_j|| \right) \quad (1)$$

where X_{ip} is pth predictor (feature), y_i is the value of the response and β_j is the regression coefficient of the pth feature.

6. Handling Imbalanced Data

The BugHunter dataset presents a challenge due to its imbalanced nature. Therefore, in this study, SMOTE [19] was utilized. To maintain the reliability and validity of our model, the dataset was split into training and testing subsets, applying SMOTE solely to the training portion instead of the full dataset.

7. Parameter Configuration

In this study, the hyperparameter optimisation Optuna [20] was utilised for tuning hyperparameters in each machine learning model. Furthermore, we applied 10-fold cross-validation to assess the performance of the comparative methods. During each iteration, the model was trained on eight segments of the dataset, with the remaining segment set aside for evaluation. Tables II-IV describe some parameters and their values in deep learning models (CNN, RNN, LSTM).

8. Machine Learning and Deep Learning algorithms

As presented in sections II.1 and II.2, researchers have explored and applied various ML and DL approaches to improve the effectiveness of SFP models. Among the numerous options, we selected the following algorithms for this comparative study.

1) KNN represents a fundamental classification algorithm, particularly suitable in scenarios where limited or no prior knowledge regarding the data distribution is available [21]. The algorithm assigns a class label to a query instance by analyzing the labels of its K nearest neighbors in the feature space. Specifically, the predicted label corresponds to the class that appears most frequently among the K nearest neighbors.
2) MLP belongs to the family of Artificial Neural Networks (ANN) and features several layers of neurons arranged sequentially. This architecture aims to

Table II
PARAMETER CONFIGURATION FOR CNN

Parameter	Definition	Value
layer_1_filters	Number of filters in first convolutional layer	50
layer_2_filters	Number of filters in second convolutional layer	20
kernel_size	Size of the convolutional kernel	3
pool_size	Size of max pooling window	2
dense_units	Number of units in hidden dense layer	64
batch_size	Number of samples per gradient update	32
epochs	Number of complete passes through the training dataset	200
learning_rate	Optimizer learning rate	0.01
optimizer	Optimization algorithm	Adam
loss_function	Loss function for model training	Binary Cross-entropy
activation_function	Activation function in hidden layers	ReLU
output_activation	Activation function in output layer	Sigmoid

Table III
PARAMETER CONFIGURATION FOR RNN

Parameter	Definition	Value
rnn_units	Number of units in RNN layer	32
dense_1_units	Number of units in first dense layer	32
batch_size	Number of samples per gradient update	32
epochs	Number of complete passes through the training dataset	200
learning_rate	Optimizer learning rate	0.01
optimizer	Optimization algorithm	Adam
loss_function	Loss function for model training	Binary Cross-entropy
activation_rnn	Activation function in RNN layer	ReLU
activation_dense	Activation function in dense layer	ReLU
output_activation	Activation function in output layer	Sigmoid

Table IV
PARAMETER CONFIGURATION FOR LSTM

Parameter	Definition	Value
layer_1_dim	Number of units in first LSTM layer	128
layer_2_dim	Number of units in second LSTM layer	64
batch_size	Number of samples per gradient update	64
epochs	Number of complete passes through the training dataset	200
learning_rate	Optimizer learning rate	0.001
dropout_rate	Dropout probability for regularization	0.2
l2_regularization	L2 regularization coefficient	0.01
optimizer	Optimization algorithm	Adam
loss_function	Loss function for model training	Binary Cross-entropy
output_activation	Activation function in output layer	Sigmoid

replicate the neural processing functions observed in the human brain. The MLP generally comprises an input layer, multiple hidden layers, and an output layer. The hidden layers which are unobservable from outside the network [22]-are essential for capturing and learning intricate patterns within the data.

3) AdaBoost [23] is a member of the Boosting family of algorithms. This method operates iteratively by prioritizing misclassified samples during training, adjusting the sample distribution accordingly, and continuing this process until the weak classifier has been trained a predefined number of times.
4) XGB [24] is also a part of the Boosting family of algorithms. The aim of XGB is to minimize the loss function by constructing decision trees sequentially, where each new tree corrects the errors made by the previous trees. XGB enables regularization (L1 and L2), parallel processing, and automatic handling of missing data.
5) CNN is a prominent deep learning framework inspired by the visual perception systems found in biological entities [25]. Its architecture commonly includes essential layers such as convolutional, activation, pooling, fully connected, dropout, and batch normalization layers.
6) RNN is a neural network architecture specifically developed for handling sequential data. Its strength lies in a high-dimensional hidden state governed by nonlinear dynamics, allowing it to retain and utilise past information effectively [26].
7) LSTM was introduced to address the exploding/vanishing gradient problems that typically arise when

learning long-term dependencies [27] by using memory cells and gating mechanisms to retain or forget information over long sequences selectively.

9. Performance Metrics

In order to evaluate and deeply understand the performance of the models, we examined them based on four main metrics: accuracy, F1-score, AUC, and recall.

- True Positive (TP): The number of positive samples classified correctly.
- True Negative (TN): The number of negative samples classified correctly.
- False Positive (FP): The number of negative samples classified as positive samples.
- False Negative (FN): The number of positive samples classified as negative samples.

Precision is the percentage of correctly predicted positive samples out of the total predicted positive samples. It is represented as:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

Accuracy is the ratio of correct predictions to the total instances in the target attribute. The value of accuracy ranges from 0 to 1.

F1-score is calculated as:

$$F1\text{-}score = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{3}$$

The F1-score increases when both precision and recall are high, providing a balanced measure of a model's performance, especially in handling imbalanced datasets.
Recall measures the percentage of correctly predicted positive samples out of all actual positive samples. It is represented as:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4}$$

The AUC (Area Under the Curve) is used to evaluate the efficiency of the SFP model. A curve that approaches the upper-left corner of the plot indicates good model performance, whereas a curve that deviates significantly from this area suggests poorer performance.

III. EXPERIMENT RESULTS

This section presents the experimental findings from a comparative analysis of machine learning and deep learning models across various projects within the BugHunter dataset, evaluated using Accuracy, AUC, F1-score, and Recall metrics.

As shown in Table V, CNN, RNN, and LSTM demonstrated superior performance in most projects. In the Ceylon-ide-eclipse project, CNN achieved the highest performance in all metrics, achieving an value of 63.4%, an AUC value 67.8%, an F1-score value 62.7%, and recall value 64.3%. Meanwhile, RNN demonstrated better performance compared to traditional machine learning models, achieving an accuracy value 62%, an AUC value 66.4%, F1-score value 61.8%, and recall value 63.9%. LSTM outperformed traditional machine learning models only in terms of AUC, with a value value 63.2%. RNN showed exceptional performance on the Orxy project, achieving the highest accuracy (91.0%) and recall (96.2%) values while CNN achieved the best AUC (95.0%).

For the Titan project, the RNN model consistently outperformed all other models by significant margins, achieving the highest performance across all evaluated metrics, including accuracy (65.6%), AUC (73.5%), F1-score (63.2%), and recall (62.0%). Similarly, in the Neo4j project, the CNN model demonstrated superior performance across all metrics, attaining an accuracy of 56.0%, an AUC of 57.8%, an F1-score of 59.5%, and a recall of 65.3%.

In the MapDB project, deep learning models demonstrated superior performance across multiple metrics. The CNN achieved the highest accuracy value of 62.5%, the RNN attained the highest AUC value of 66.2%, while the LSTM exhibited strong performance in terms of an F1-score value of 68.0% and a recall value of 90.7%.

On the other hand, traditional machine learning models achieved better performance than deep learning models in some cases. For instance, in the Broadleaf-Commerce project, KNN achieved superior performance across most metrics, with an accuracy value of 68.6%, an F1-score value of 68.4%, and a recall value of 69.2%. The highest AUC, however, was attained by the RNN model, with a value of 75.6%. Similarly, in the OrientDB project, the MLP model demonstrated robust performance, achieving the highest F1-score and recall values of 59.5% and 59.5%, while the RNN model achieved the best AUC value of 65.4%.

Moreover, drawing upon the study by Ferenc *et al.* [6], we compare their best performing techniques with our experimental results in terms of F1-score as summarized in Table VI. For the MapDB project, RNN slightly outperforms CNN, with F1-score values of 0.643 and 0.604, respectively. Meanwhile, Random Forest demonstrates competitive performance, achieving the highest F1-score value of 0.736 on the BroadleafCommerce project, while also performing well on Orientdb (0.623). Lastly, LSTM showed varied performance, with its best score of 0.832 on the Oryx project. Overall, deep learning models (CNN, RNN and LSTM) achieved better performance compared with the

Table V
PERFORMANCE COMPARISON OF MACHINE LEARNING AND DEEP LEARNING MODELS

Project	Metric	KNN	MLP	AdaBoost	XGB	CNN	RNN	LSTM
Ceylon-ide-eclipse	Accuracy	0.538	0.544	0.571	0.583	**0.634**	0.620	0.581
	AUC	0.543	0.582	0.568	0.587	**0.678**	0.664	0.632
	F1-score	0.530	0.537	0.562	0.578	**0.627**	0.618	0.562
	Recall	0.541	0.547	0.577	0.586	**0.643**	0.639	0.548
Orxy	Accuracy	0.688	0.713	0.650	0.613	0.908	**0.910**	0.834
	AUC	0.750	0.757	0.770	0.725	**0.950**	0.945	0.905
	F1-score	0.669	0.712	0.621	0.544	0.911	**0.914**	0.832
	Recall	0.742	0.712	0.720	0.782	0.954	**0.962**	0.832
Titan	Accuracy	0.494	0.530	0.500	0.530	0.617	**0.656**	0.570
	AUC	0.513	0.500	0.493	0.500	0.680	**0.735**	0.613
	F1-score	0.488	0.528	0.474	0.528	0.573	**0.632**	0.532
	Recall	0.495	0.532	0.500	0.531	0.558	**0.620**	0.536
Orientdb	Accuracy	0.576	0.594	0.592	0.580	0.600	0.602	**0.616**
	AUC	0.599	0.634	0.630	0.628	0.641	**0.654**	0.650
	F1-score	0.571	**0.595**	0.592	0.575	0.565	0.557	0.576
	Recall	0.579	**0.595**	0.592	0.584	0.524	0.505	0.522
Broadleaf-Commerce	Accuracy	**0.686**	0.673	0.650	0.672	0.639	0.683	0.623
	AUC	0.725	0.732	0.709	0.748	0.702	**0.756**	0.693
	F1-score	**0.684**	0.667	0.646	0.665	0.584	0.680	0.584
	Recall	**0.692**	0.688	0.657	0.688	0.513	0.676	0.534
MapDB	Accuracy	0.586	0.603	0.588	0.584	**0.625**	0.615	0.575
	AUC	0.622	0.632	0.619	0.609	0.655	**0.662**	0.620
	F1-score	0.560	0.592	0.586	0.582	0.604	0.643	**0.680**
	Recall	0.614	0.614	0.690	0.586	0.584	0.706	**0.907**

Table VI
PERFORMANCE COMPARISON OF DEEP LEARNING MODELS WITH BEST-PERFORMING TECHNIQUES IN [8]

Project	Random Forest	SimpleLogistic	SGD	CNN	RNN	LTSM
Ceylon-ide-eclipse	0.539	–	–	**0.627**	0.618	0.562
Oryx	0.667	–	–	0.911	**0.914**	0.832
Orientdb	**0.623**	–	–	0.565	0.557	0.576
MapDB	–	0.561	–	0.604	0.643	**0.680**
BroadleafCommerce	**0.736**	–	–	0.584	0.680	0.584
Titan	–	–	0.579	0.573	**0.632**	0.532

ML models on 5 out of 7 projects in terms of F1-score.

The results imply that RNN and CNN models demonstrate reliable effectiveness, confirming their relevance to this domain. However, traditional machine learning approaches remain viable alternatives in specific scenarios, demonstrating competitive performance in certain metrics.

IV. CONCLUSIONS

Choosing models in SFP plays a critical role in determining overall performance. Although machine learning models have shown considerable effectiveness, deep learning approaches possess the capability to address the inherent limitations of traditional machine learning techniques. Our objective is to evaluate and contrast the performance of various machine learning and deep learning models on the BugHunter dataset, using Accuracy, F1-score, AUC, and Recall as evaluation metrics. In general, deep learning models demonstrated their superiority by achieving higher performance than machine learning models on most projects. In particular, RNN, CNN, and LSTM outperformed traditional machine learning models, yielding respective improvements of 10.16%, 12.28%, 12.28%, and 5.61% in Accuracy, AUC, F1-score, and Recall. In the future, we plan to evaluate more state-of-the-art deep learning models on the BugHunter dataset and improve the performance of SFP models by handling imbalanced data and irrelevant or redundant feature issues. We expect these models to deliver even higher performance, providing significant advantages in the early prediction of software faults.

REFERENCES

[1] S. Najihi, S. Elhadi, R. A. Abdelouahid, and A. Marzak, "Software testing from an agile and traditional view," *Procedia Computer Science*, vol. 203, pp. 775–782, 2022.

[2] A. Balaram and S. Vasundra, "Prediction of software fault-prone classes using ensemble random forest with adaptive synthetic sampling algorithm," *Automated Software Engineering*, vol. 29, no. 1, p. 6, 2022.

[3] M. Mangla, N. Sharma, and S. N. Mohanty, "A sequential ensemble model for software fault prediction," *Innovations in Systems and Software Engineering*, vol. 18, no. 2, pp. 301–308, 2022.

[4] S. S. Rathore and S. Kumar, "A study on software fault prediction techniques," *Artificial Intelligence Review*, vol. 51, pp. 255–327, 2019.

[5] S. K. Pandey, R. B. Mishra, and A. K. Tripathi, "Machine learning based methods for software fault prediction: A survey," *Expert Systems with Applications*, vol. 172, p. 114595, 2021.

[6] R. Ferenc, P. Gyimesi, G. Gyimesi, Z. Tóth, and T. Gyimóthy, "An automatically created novel bug dataset and its validation in bug prediction," *Journal of Systems and Software*, vol. 169, p. 110691, 2020.

[7] C. Jin, S.-W. Jin, and J.-M. Ye, "Artificial neural network-based metric selection for software fault-prone prediction model," *IET software*, vol. 6, no. 6, pp. 479–487, 2012.

[8] S. Wang, T. Liu, J. Nam, and L. Tan, "Deep semantic feature learning for software defect prediction," *IEEE Transactions on Software Engineering*, vol. 46, no. 12, pp. 1267–1293, 2018.

[9] R. Malhotra, "A systematic review of machine learning techniques for software fault prediction," *Applied Soft Computing*, vol. 27, pp. 504–518, 2015.

[10] S. T. Cynthia, B. Roy, and D. Mondal, "Feature transformation for improved software bug detection models," in *Proceedings of the 15th Innovations in Software Engineering Conference*, pp. 1–10, 2022.

[11] C. Zhang, P. Patras, and H. Haddadi, "Deep learning in mobile and wireless networking: A survey," *IEEE Communications surveys & tutorials*, vol. 21, no. 3, pp. 2224–2287, 2019.

[12] H. K. Dam, T. Pham, S. W. Ng, T. Tran, J. Grundy, A. Ghose, T. Kim, and C.-J. Kim, "A deep tree-based model for software defect prediction," *arXiv preprint arXiv:1802.00921*, 2018.

[13] H. Liang, Y. Yu, L. Jiang, and Z. Xie, "Seml: A semantic lstm model for software defect prediction," *IEEE Access*, vol. 7, pp. 83812–83824, 2019.

[14] A. V. Phan and M. Le Nguyen, "Convolutional neural networks on assembly code for predicting software defects," in *2017 21st Asia Pacific Symposium on Intelligent and Evolutionary Systems (IES)*, pp. 37–42, IEEE, 2017.

[15] A. B. Farid, E. M. Fathy, A. S. Eldin, and L. A. Abd-Elmegid, "Software defect prediction using hybrid model (cbil) of convolutional neural network (cnn) and bidirectional long short-term memory (bi-lstm)," *PeerJ Computer Science*, vol. 7, p. e739, 2021.

[16] Z. M. Zain, S. Sakri, N. H. Asmak Ismail, and R. M. Parizi, "Software defect prediction harnessing on multi 1-dimensional convolutional neural network structure.," *Computers, Materials & Continua*, vol. 71, no. 1, 2022.

[17] L. Qiao, X. Li, Q. Umer, and P. Guo, "Deep learning based software defect prediction," *Neurocomputing*, vol. 385, pp. 100–110, 2020.

[18] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 58, no. 1, pp. 267–288, 1996.

[19] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[20] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 2623–2631, 2019.

[21] S. B. Imandoust, M. Bolandraftar, *et al.*, "Application of k-nearest neighbor (knn) approach for predicting economic events: Theoretical background," *International journal of engineering research and applications*, vol. 3, no. 5, pp. 605–610, 2013.

[22] M.-C. Popescu, V. E. Balas, L. Perescu-Popescu, and N. Mastorakis, "Multilayer perceptron and neural networks," *WSEAS Transactions on Circuits and Systems*, vol. 8, no. 7, pp. 579–588, 2009.

[23] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of computer and system sciences*, vol. 55, no. 1, pp. 119–139, 1997.

[24] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794, 2016.

[25] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, *et al.*, "Recent advances in convolutional neural networks," *Pattern recognition*, vol. 77, pp. 354–377, 2018.

[26] I. Sutskever, J. Martens, and G. E. Hinton, "Generating text with recurrent neural networks," in *Proceedings of the 28th international conference on machine learning (ICML-11)*, pp. 1017–1024, 2011.

[27] G. Van Houdt, C. Mosquera, and G. Nápoles, "A review on the long short-term memory model," *Artificial Intelligence Review*, vol. 53, no. 8, pp. 5929–5955, 2020.

Dang Thi Kim Ngan obtained a Bachelor degree in Information Management at the University of Danang -University of Economics in 2011. She got an MSc degree of at the Chinese Culture University, Taiwan in 2014. She is currently serving as a lecturer at the University of Danang - Vietnam-Korea University of Information and Communication Technology. Her research focuses on software engineering.
Email: dtkngan@vku.udn.vn.

Dao Khanh Duy is currently pursuing a Bachelor of Engineering at the University of Danang - Vietnam-Korea University of Information and Communication Technology. He has been gaining valuable hands-on experience in research and development through internships at the King Mongkut's University of Technology North Bangkok and MGM technology partners. His current research focuses on software testing and software effort estimation.
Email: duydk.22git@vku.udn.vn.

Ha Thi Minh Phuong received a Bachelor degree in Information Technology from the University of Danang-University of Science and Technology in 2010. She earned M.Sc. degree in Computer Science at Yuan Ze University, Taiwan in 2013. She is a lecturer at the University of Danang - Vietnam-Korea University of Information and Communication Technology. She is currently pursuing the Ph.D. degree in Computer Science at the University of Danang. Her current research focuses on software testing, deep learning.
Email: htmphuong@vku.udn.vn.

Nguyen Thanh Binh graduated in Information Technology from the University of Danang - University of Science and Technology in 1997. He received a PhD. degree in Information Technology at Grenoble Institute of Technology, France in 2004. He has been qualified as Associate Professor since 2013. He is currently working at the University of Danang - Vietnam-Korea University of Information and Communication Technology. His research interests include software engineering and software quality.
Email: ntbinh@vku.udn.vn.

THÔNG TIN DÀNH CHO TÁC GIẢ

Các công trình nghiên cứu, phát triển và ứng dụng Công nghệ Thông tin và Truyền thông - gọi tắt là Công nghệ TTTT, phiên bản tiếng Anh là ICT Research – là chuyên san khoa học của Tạp chí Thông tin và Truyền thông, do Bộ Thông tin và Truyền thông xuất bản từ năm 1999. Chuyên san Công nghệ TTTT có phản biện, mở trực tuyến (open access) và được phát hành hàng năm với Ấn phẩm tiếng Việt (ISSN: 1859-3526) có 2 số và Ấn phẩm tiếng Anh (ISSN: 1859-3534) – Research and Development on Information and Communication Technology (ICT Research) – có 2 số.

Chuyên san Công nghệ TTTT là một diễn đàn dành cho các nhà nghiên cứu và các chuyên gia phổ biến các ý tưởng mới và sáng tạo trong các lĩnh vực của Công nghệ Thông tin, Truyền thông và Điện tử ở Việt Nam và trên toàn thế giới. Hội đồng biên tập của Chuyên san CNTT-TT bao gồm các nhà khoa học uy tín đến từ các đơn vị nghiên cứu ở Việt Nam và các quốc gia khác.

HƯỚNG DẪN DÀNH CHO CÁC TÁC GIẢ

Các tác giả nên tìm hiểu cách viết bài báo kỹ thuật, chẳng hạn như các hướng dẫn trong tài liệu "How to write for technical periodicals & conferences" của Viện Kỹ sư Điện, Điện tử Hoa Kỳ (IEEE).

Chuyên san Công nghệ TTTT chỉ nhận các bản thảo được nộp trực tuyến trên trang web của Chuyên san Công nghệ TTTT tại https://ictmag.vn/cntt-tt/ Tác giả cần chế bản bản thảo theo mẫu mà Chuyên san Công nghệ TTTT hướng dẫn.

Các tác giả phải chịu trách nhiệm trong việc đảm bảo rằng bản thảo này chưa từng được công bố trước đó, cũng như hiện đang không trong quá trình phản biện hoặc chuẩn bị xuất bản ở nơi khác. Tác giả liên hệ cũng phải chịu trách nhiệm đảm bảo sự đồng ý của tất cả các đồng tác giả khi nộp bài cho Chuyên san Công nghệ TTTT. Bản thảo cần được nộp dưới dạng PDF. Để hỗ trợ phản biện kín hai chiều, tác giả cần nộp hai phiên bản của bản thảo, một phiên bản có đầy đủ thông tin của tất cả các tác giả, một không có thông tin gì.

Các bản thảo nộp lại sau khi chỉnh sửa cũng cần tuân theo các chỉ dẫn như của bản thảo nộp lần đầu, nhưng cần kèm theo một văn bản giải trình các ý kiến của phản biện, trong đó giải thích các nội dung được chỉnh sửa và/hoặc trả lời các ý kiến của các phản biện. Sau khi một bản thảo được chính thức chấp nhận xuất bản, các tác giả cần phải chuẩn bị bản thảo cuối cùng bằng LATEX theo mẫu trình bày cho IEEE Transactions.

ĐẠO ĐỨC TRONG XUẤT BẢN

Chuyên san Công nghệ TTTT cam kết tuân thủ các quy định, tiêu chuẩn và hướng dẫn được thiết lập bởi COPE (Committee on Publication Ethics) trong việc đảm bảo tính liêm chính của tạp chí cũng như xử lý các trường hợp vi phạm. Tất cả các bên tham gia (Biên tập viên, Tác giả, Phản biện và Nhà xuất bản) cần phải hiểu và tuân thủ các quy định, tiêu chuẩn và hướng dẫn được công bố trên trang web của COPE.

Ban biên tập của Chuyên san Công nghệ TTTT sẽ sàng lọc tất cả các bản thảo trước khi công bố. Tất cả các bản thảo sẽ được kiểm tra về mức độ tương tự với các công trình khác sử dụng các công cụ trực tuyến có được. Mọi hình thức đạo văn không được chấp nhận và bản thảo sẽ bị từ chối.

Các kết luận phản biện gồm có: (A) Chấp nhận, (B1) Chấp nhận với yêu cầu chỉnh sửa nhỏ, (B2) Chỉnh sửa và nộp lại (bản thảo cần phải viết lại với các thay đổi lớn và sau đó cần qua vòng phản biện lần thứ hai), và (C) Từ chối.

Mỗi vòng phản biện có thể cần khoảng hai tháng. Trong trường hợp các tác giả được Biên tập viên yêu cầu chỉnh sửa bản thảo, phiên bản được chỉnh sửa cần được nộp trong vòng hai tháng đối với yêu cầu chỉnh sửa lớn và một tháng với yêu cầu chỉnh sửa nhỏ. Tác giả nào cần nhiều thời gian hơn xin vui lòng liên hệ với Biên tập viên.

Mỗi bài báo được chấp nhận sẽ được xuất bản trực tuyến ngay sau đó, với số DOI, sau khi được biên tập và chế bản theo các quy tắc biên tập của Chuyên san Công nghệ TTTT. Đồng thời, khi có đủ một số lượng bài phù hợp, Chuyên san Công nghệ TTTT sẽ tập hợp lại và xuất bản một số in, kèm với mục lục các bài báo.

THÔNG TIN VỀ BẢN QUYỀN