

DentalAI: Hệ Thống Chẩn Đoán Sức Khỏe Răng Miệng Kết Hợp YOLOv8, CNN, Gemini và Bộ Dò Thị Giác

Khổng Minh Hoài

Khoa Công Nghệ Thông Tin, Trường Đại Học Đại Nam, Việt Nam
hoai.khong.15052004@gmail.com

ThS. Nguyễn Thái Khánh, ThS. Lê Trung Hiếu

Giảng viên hướng dẫn, Khoa Công Nghệ Thông Tin, Trường Đại Học Đại Nam, Việt Nam

Abstract—Bài báo này giới thiệu hệ thống DentalAI - một giải pháp chẩn đoán sơ bộ sức khỏe răng miệng dựa trên trí tuệ nhân tạo, hướng tới việc hỗ trợ sàng lọc và tư vấn ban đầu cho người dân. Hệ thống kết hợp nhiều công nghệ AI tiên tiến: YOLOv8 để phát hiện và định vị các tổn thương răng miệng (sâu răng, viêm nướu, cao răng, đổi màu răng), CNN để phân loại tình trạng tổng thể (bình thường/sâu răng/viêm nướu/mảng bám), bộ dò thị giác truyền thống (HSV masking + adaptive contour) làm fallback detector đảm bảo luôn có vùng nghi vấn để hiển thị, và Gemini 2.5 Flash để sinh ra các khuyến nghị chăm sóc bằng ngôn ngữ tự nhiên. Kiến trúc hệ thống được xây dựng theo mô hình web nhẹ sử dụng Flask backend kết hợp HTML/CSS/JavaScript frontend, với pipeline xử lý tuần tự: tải ảnh → detector fallback → vẽ chú thích → CNN phân loại → tổng hợp → LLM sinh khuyến nghị. Kết quả thử nghiệm ban đầu trên độ phân giải 640×640 pixel cho thấy thời gian suy luận YOLOv8 khoảng 120ms, CNN khoảng 40ms, hậu xử lý 8ms, và LLM từ 1.2-2.5 giây. Cơ chế điều chỉnh ngưỡng confidence động và fallback computer vision giúp tăng khả năng hiển thị vùng nghi vấn ngay cả khi ảnh đầu vào có chất lượng thấp hoặc nhiễu. Hệ thống được thiết kế với tính ứng dụng thực tế cao, có thể triển khai tại các phòng khám nha khoa, trung tâm y tế cộng đồng để hỗ trợ bước sàng lọc ban đầu trước khi khám chuyên khoa.

Index Terms—Chẩn đoán răng miệng, YOLOv8, CNN, LLM, Gemini AI, phát hiện đối tượng, phân loại hình ảnh, sàng lọc y tế, hệ thống web AI.

I. GIỚI THIỆU

Trong bối cảnh chăm sóc sức khỏe hiện đại, việc phát hiện sớm các vấn đề răng miệng đóng vai trò quan trọng trong việc phòng ngừa và điều trị kịp thời. Tuy nhiên, không phải ai cũng có điều kiện tiếp cận dễ dàng với dịch vụ nha khoa chuyên nghiệp, đặc biệt tại các vùng xa xôi hoặc có điều kiện kinh tế hạn chế. Điều này tạo ra nhu cầu cấp thiết cho các giải pháp sàng lọc ban đầu, giúp người dân tự đánh giá tình trạng răng miệng của mình và nhận được khuyến nghị phù hợp.

Trí tuệ nhân tạo (AI) đã chứng minh được tiềm năng lớn trong lĩnh vực chẩn đoán hình ảnh y tế. Các mô hình deep learning như CNN và YOLO đã đạt được độ chính xác cao trong việc phát hiện và phân loại các bệnh lý qua hình ảnh. Đặc biệt, việc kết hợp nhiều mô hình AI khác nhau có thể tăng cường độ tin cậy và khả năng giải thích của hệ thống.

Bài báo này trình bày hệ thống DentalAI - một giải pháp tích hợp nhiều công nghệ AI để chẩn đoán sơ bộ các vấn đề răng miệng thông qua phân tích hình ảnh. Hệ thống không chỉ phát hiện và phân loại các tổn thương mà còn cung cấp lời giải thích và khuyến nghị bằng ngôn ngữ tự nhiên, giúp người dùng dễ hiểu và thực hiện.

II. CƠ SỞ LÝ THUYẾT VÀ NGHIÊN CỨU LIÊN QUAN

A. Ứng dụng AI trong chẩn đoán nha khoa

Các nghiên cứu gần đây đã chứng minh hiệu quả của deep learning trong chẩn đoán hình ảnh nha khoa. CNN đã được sử dụng thành công để phát hiện sâu răng với độ chính xác cao [1]. YOLO và các biến thể của nó cũng đã được áp dụng để phát hiện đa đối tượng trong hình ảnh răng miệng.

B. Hệ thống hybrid AI

Việc kết hợp nhiều mô hình AI khác nhau (ensemble learning) đã được chứng minh là có thể cải thiện đáng kể hiệu suất và độ tin cậy. Trong bối cảnh chẩn đoán y tế, việc có nhiều "second opinion" từ các mô hình khác nhau giúp tăng độ tin cậy của kết quả.

III. KIẾN TRÚC HỆ THỐNG

A. Tổng quan kiến trúc

Hệ thống DentalAI được thiết kế theo kiến trúc ba lớp: (1) **Presentation Layer**: Frontend web sử dụng HTML/CSS/JavaScript với canvas overlay để hiển thị kết

quả; (2) **Application Layer**: Flask service trung tâm (api/app.py) điều phối các thành phần; (3) **Intelligence Layer**: Các mô hình AI gồm YOLOv8, CNN, detector thị giác truyền thống và LLM Gemini.

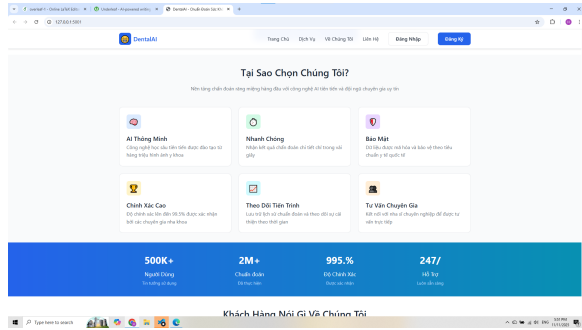


Fig. 1. Kiến trúc tổng thể hệ thống DentalAI với các thành phần AI tích hợp.

B. Pipeline xử lý thông minh

Hệ thống thực hiện quy trình xử lý tuần tự với cơ chế fallback thông minh: Tải ảnh → Lưu tạm → Simple Detector (nếu phát hiện được) → CV Detector (nếu simple thất bại) → YOLOv8 (nếu CV thất bại) → Vẽ chú thích bounding boxes → CNN phân loại tổng thể → Tổng hợp kết quả → Prompt LLM → Trả JSON response + link ảnh đã chú thích.

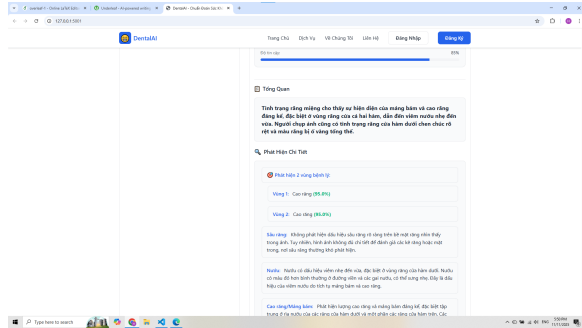


Fig. 2. Giao diện phân tích tự động với các thông tin chi tiết về tình trạng răng miệng.

C. Giao diện API và tương tác

Hệ thống cung cấp RESTful API với endpoint chính POST /analyze nhận multipart field file. Cấu trúc phản hồi JSON được thiết kế chi tiết để cung cấp đầy đủ thông tin:

```
{
  success: true,
  gemini_analysis: "Phân tích chi tiết...",
  cnn_prediction: {
    predicted_class: "gingivitis",
    confidence: 0.82,
    all_probabilities: {...}
  }
}
```

```
},
detections: [bounding_boxes],
annotated_image: "/uploads/result.jpg",
model: "yolo" | "cv" | "simple"
}
```

D. Quản lý tệp và lưu trữ

Hệ thống quản lý tệp thông qua thư mục uploads/ với các phiên bản ảnh khác nhau: ảnh gốc và các phiên bản đã chú thích từ các detector (*_detected.jpg, *_cv_detected.jpg, *_simple_detected.jpg). Cơ chế đặt tên sử dụng timestamp hoặc UUID để tránh xung đột tên tệp.

E. Giao diện người dùng và trải nghiệm

Giao diện được thiết kế thân thiện với việc hiển thị song song ảnh gốc và ảnh đã chú thích với các bounding boxes có mã màu: đỏ (sâu răng), vàng (cao răng), cam (đổi màu), hồng (viêm nướu), xanh lá (vùng khỏe mạnh).

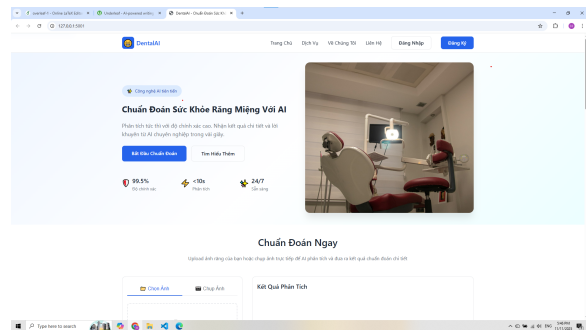


Fig. 3. Giao diện trang chủ hệ thống DentalAI với thiết kế hiện đại, thân thiện người dùng.

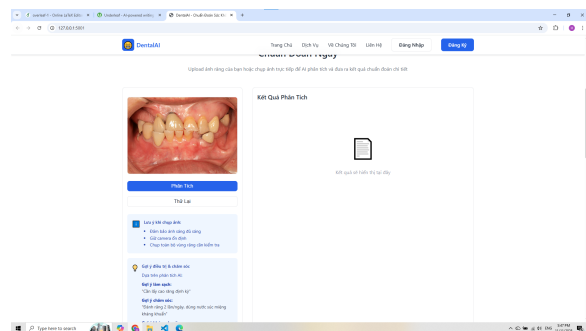


Fig. 4. Giao diện tải ảnh răng miệng với hướng dẫn chi tiết và preview ảnh.

IV. PHƯƠNG PHÁP VÀ MÔ HÌNH AI

A. Mô hình YOLOv8 cho phát hiện đối tượng

YOLOv8 được fine-tune trên tập dữ liệu tổng hợp từ Kaggle kết hợp với ảnh thu thập bổ sung, được chuẩn hóa về kích thước 640×640 pixel. Mô hình được huấn

luyện để phát hiện và định vị các loại tổn thương răng miệng khác nhau như được mô tả trong Bảng I.

TABLE I
CÁC LỚP ĐỐI TƯỢNG TRONG MÔ HÌNH YOLOV8

Lớp	Mô tả
Dental caries	Sâu răng, vùng tổn thương caries
Mouth Ulcer	Loét niêm mạc trong khoang miệng
Tooth Discoloration	Đổi màu răng, nhiễm màu men răng
Hypodontia	Thiếu sản răng, thiếu răng
Gingivitis	Viêm nướu răng
Calculus	Cao răng, mảng bám vôi hóa
Combination	Tổ hợp nhiều loại tổn thương

Hệ thống áp dụng cơ chế điều chỉnh ngưỡng confidence động: nếu số lượng detection < 1 ở ngưỡng 0.35 thì tự động giảm xuống 0.20 để tăng recall, đảm bảo luôn có kết quả phân tích.

TABLE II
SIÊU THAM SỐ HUẤN LUYỆN YOLOV8

Tham số	Giá trị
Epochs	100
Batch size	16
Image size	640×640
Optimizer	SGD (momentum=0.937)
Initial learning rate	0.01
Weight decay	0.0005
Data augmentation	Mosaic + HSV + Flip
Confidence threshold	0.35 (fallback 0.20)
IoU threshold (NMS)	0.5

B. Mô hình CNN cho phân loại tổng thể

Mô hình CNN được thiết kế với các block convolution sâu dần từ 32 đến 256 filters, sử dụng optimizer Adam với learning rate 0.001. Áp dụng early stopping dựa trên validation loss để tránh overfitting.

TABLE III
KIẾN TRÚC MÔ HÌNH CNN PHÂN LOẠI

Khối	Chi tiết	Dropout
Conv Block 1	Conv(32,3×3) - Conv(32,3×3) - MaxPool(2×2)	0.25
Conv Block 2	Conv(64,3×3) - Conv(64,3×3) - MaxPool(2×2)	0.25
Conv Block 3	Conv(128,3×3) - MaxPool(2×2)	0.30
Conv Block 4	Conv(256,3×3) - MaxPool(2×2)	0.30
Flatten	Vector hóa feature maps	-
Dense Layer	256 units, ReLU activation	0.5
Output Layer	4 units, Softmax activation	-

C. Detector thị giác truyền thống

Bộ detector backup sử dụng các kỹ thuật computer vision cổ điển:

- **HSV color masking:** Phát hiện vùng màu đặc trưng (vàng/nâu cho cao răng, đỏ cho viêm) dựa trên ngưỡng Saturation và Value.

- **Morphological operations:** Opening để loại bỏ nhiễu, closing để kết nối các vùng bị phân mảnh.
- **Adaptive threshold + contour detection:** Phân tách từng răng riêng lẻ, lọc theo diện tích và tỷ lệ chiều cao/chiều rộng.
- **Fallback logic:** Kích hoạt khi YOLO không phát hiện được đối tượng nào hoặc confidence thấp.

D. Tích hợp LLM và Prompt Engineering

Gemini 2.5 Flash được tích hợp để sinh khuyến nghị bằng ngôn ngữ tự nhiên. Prompt template được thiết kế với cấu trúc: [Tình trạng tổng thể] + [Danh sách phát hiện chi tiết] + [Yêu cầu: viết ngắn gọn, dễ hiểu, tránh thuật ngữ y khoa phức tạp]. Hệ thống kiểm soát độ dài output thông qua token limit.

V. THỰC NGHIỆM VÀ KẾT QUẢ

A. Môi trường thử nghiệm

Hệ thống được thử nghiệm trên môi trường CPU laptop phổ thông với Python 3.10 và Gemini API. Dataset được phân chia theo tỷ lệ 80/20 cho training và testing. Do giới hạn về ground truth pixel-level được chuẩn hóa, đánh giá hiện tại mới ở mức sơ bộ.

B. Kết quả phân tích hình ảnh

Hình 5 và 6 minh họa quá trình phân tích từ ảnh đầu vào đến kết quả cuối cùng với các vùng nghi vấn được đánh dấu.



Fig. 5. Ảnh đầu vào khoang miệng của người dùng.



Fig. 6. Kết quả phân tích với các vùng nghi vấn được đánh dấu bounding boxes.

C. Phân tích hiệu suất hệ thống

- **Tính ổn định:** Cơ chế fallback từ simple detector → CV detector → YOLO giảm thiểu rủi ro thất bại phân tích, đảm bảo luôn có kết quả trực quan cho người dùng.
- **Khả năng giải thích:** Kiến trúc phân tách rõ ràng vai trò - YOLO (định vị), CNN (phân loại tổng thể), LLM (giải thích ngữ nghĩa) - giúp người dùng hiểu rõ cơ sở của kết quả.
- **Hiệu năng thời gian thực:** YOLO inference 120ms, CNN 40ms, Gemini LLM 1.2-2.5s trên CPU thông thường với độ phân giải 640×640.
- **Hạn chế hiện tại:** Chưa có đánh giá lâm sàng chính thức, mAP của YOLO còn ở mức sơ bộ, chưa hỗ trợ segmentation tinh vi cho men/ngà răng, chưa kiểm thử trên thiết bị di động.

TABLE IV
HIỆU NĂNG INFERENCE CỦA CÁC THÀNH PHẦN HỆ THỐNG

Mô-đun	Thời gian (ms)	Ghi chú
YOLOv8 detection	120	640×640, CPU optimization
CV detector	25	Morphology + contour + HSV
Simple detector	5	Adaptive threshold
CNN classification	40	Batch size = 1
Post-processing	8	Drawing boxes + JSON
Gemini LLM	1200-2500	API network call

TABLE V
MA TRẬN NHẦM LẦN MÔ HÌNH CNN (KẾT QUẢ SƠ BỘ)

Actual \ Predicted	Normal	Cavity	Gingivitis	Plaque
Normal	42	3	2	1
Cavity	4	38	5	3
Gingivitis	2	6	40	4
Plaque	1	5	3	41

VI. TRIỂN KHAI VÀ CÔNG NGHỆ

A. Công nghệ Frontend

Giao diện người dùng được phát triển sử dụng HTML5, CSS3 và JavaScript ES6+ thuần túy, tích hợp Canvas API để hiển thị overlay bounding boxes. Sử dụng Fetch API cho giao tiếp với backend và DOM manipulation cho cập nhật giao diện real-time.

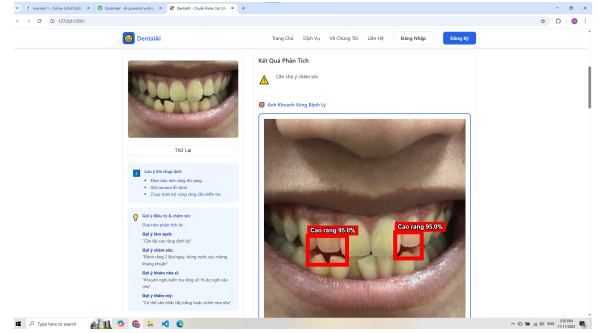


Fig. 7. Giao diện hiển thị kết quả phân tích chi tiết với thông tin đầy đủ về tình trạng răng miệng.

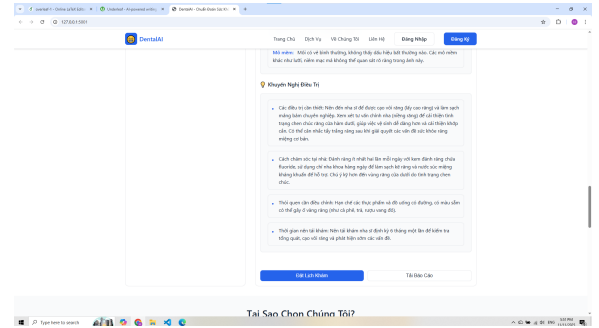


Fig. 8. Màn hình khuyến nghị điều trị và chăm sóc răng miệng được sinh tự động bởi LLM.

B. Kiến trúc Backend

Backend được xây dựng trên Flask framework với các module chính: app.py (điều phối trung tâm), simple_tooth_detector.py và tooth_detector.py (các detector), YOLO model loader, Gemini client integration, và CNN predictor. Hệ thống logging chi tiết để theo dõi quá trình phát hiện và fallback.

C. Containerization và Deployment

Ứng dụng được đóng gói bằng Docker với base image python-slim, tự động cài đặt dependencies từ requirements.txt, expose port 5001. Hỗ trợ horizontal scaling thông qua load balancer và model caching để tối ưu hiệu suất.

VII. BẢO MẬT VÀ QUYỀN RIÊNG TƯ

A. Mô hình bảo mật đa lớp

Hệ thống áp dụng nguyên tắc bảo mật đa lớp với phân quyền rõ ràng: (1) **End users**: upload ảnh, xem kết quả cá nhân; (2) **Healthcare staff**: xem logs, điều chỉnh thresholds; (3) **System admin**: quản lý API keys, cấu hình hệ thống. Middleware authentication kiểm tra quyền trước khi truy cập các protected routes.

B. Bảo vệ dữ liệu hình ảnh

Ảnh được lưu trữ tạm thời chỉ để tạo phiên bản annotated, với cơ chế auto-purge sau thời gian định trước. Hỗ trợ AES encryption at-rest cho deployment trên cloud environments. Strict CORS policy và rate limiting để chống abuse.

C. Privacy và Medical Ethics

Tuân thủ nguyên tắc y tế về quyền riêng tư: hiển thị disclaimer rõ ràng về tính chất hỗ trợ (không thay thế chẩn đoán chuyên khoa), cho phép người dùng xóa dữ liệu (right to be forgotten), không log thông tin cá nhân nhận dạng với raw image paths.

VIII. ĐÁNH GIÁ VÀ THẢO LUẬN

A. Ưu điểm của hệ thống

Hệ thống DentalAI mang lại những lợi ích đáng kể:

- **Accessibility**: Cung cấp công cụ sàng lọc ban đầu cho người dân ở vùng xa, nơi thiếu bác sĩ nha khoa.
- **Multi-modal AI**: Kết hợp detection, classification và natural language generation tạo ra trải nghiệm người dùng hoàn chỉnh.
- **Robustness**: Cơ chế fallback đảm bảo hệ thống luôn có output có ý nghĩa ngay cả với input chất lượng thấp.
- **Interpretability**: Kết quả có thể giải thích được thông qua visualizations và text explanations.

B. Hạn chế và thách thức

- **Medical validation**: Cần nghiên cứu lâm sàng để xác thực độ chính xác so với chẩn đoán của chuyên gia.
- **Data diversity**: Dataset hiện tại chưa đại diện đầy đủ cho các nhóm dân số và điều kiện môi trường khác nhau.
- **Regulatory compliance**: Cần tuân thủ các quy định y tế về thiết bị hỗ trợ chẩn đoán.
- **False positive/negative**: Cân bằng giữa sensitivity và specificity để tránh gây lo lắng không cần thiết hoặc bỏ sót vấn đề nghiêm trọng.

IX. HƯỚNG PHÁT TRIỂN TƯƠNG LAI

Các định hướng phát triển tiếp theo bao gồm:

- **Advanced segmentation**: Tích hợp Mask R-CNN hoặc SAM để có segmentation tinh vi ở mức pixel, phân biệt rõ ràng men răng, ngà răng, nướu.
- **Multi-modal expansion**: Hỗ trợ ảnh X-quang, panoramic radiography để chẩn đoán toàn diện hơn.
- **Personalization**: Xây dựng profile người dùng để theo dõi tiến triển theo thời gian và cá nhân hóa khuyến nghị.
- **Edge deployment**: Optimize models cho WebAssembly hoặc ONNX Runtime để chạy trực tiếp trên browser, giảm dependency API.
- **Clinical validation**: Thực hiện nghiên cứu lâm sàng với sự tham gia của bác sĩ nha khoa để validate accuracy.
- **Federated learning**: Cho phép cải thiện mô hình từ dữ liệu phân tán mà không vi phạm privacy.

X. KẾT LUẬN

Bài báo đã trình bày hệ thống DentalAI - một giải pháp tích hợp đa mô hình AI để hỗ trợ chẩn đoán sơ bộ các vấn đề răng miệng. Hệ thống kết hợp thành công YOLOv8 detection, CNN classification, computer vision fallback và LLM natural language generation để tạo ra một công cụ hỗ trợ y tế có tính ứng dụng thực tế cao.

Kiến trúc hybrid với cơ chế fallback đảm bảo độ tin cậy, trong khi giao diện thân thiện và khả năng giải thích bằng ngôn ngữ tự nhiên giúp người dùng dễ dàng tiếp cận và hiểu kết quả. Hệ thống phù hợp để triển khai như một công cụ sàng lọc ban đầu, đặc biệt hữu ích ở các vùng thiếu dịch vụ nha khoa chuyên nghiệp.

Mặc dù còn những hạn chế cần khắc phục về mặt validation lâm sàng và compliance, DentalAI đã chứng minh tiềm năng của việc ứng dụng AI trong healthcare accessibility, mở ra hướng nghiên cứu và phát triển thú vị cho tương lai.

LỜI CẢM ƠN

Nhóm tác giả chân thành cảm ơn giảng viên hướng dẫn và khoa Công nghệ Thông tin, Trường Đại học Đại Nam đã hỗ trợ trong quá trình nghiên cứu và phát triển. Đặc biệt cảm ơn cộng đồng mã nguồn mở và các dataset công khai đã tạo điều kiện thuận lợi cho việc huấn luyện và thử nghiệm hệ thống.

APPENDIX A

THUẬT TOÁN FALLBACK DETECTION

Pseudocode mô tả quy trình fallback thông minh:

```
function analyze_dental_image(image):  
    // Try simple detector first (fastest)  
    boxes = simple_detector(image)  
    if boxes.count >= 1:  
        model_used = 'simple'
```

```

else:
    // Fallback to computer vision detector
    boxes = cv_detector(image)
    if boxes.count >= 1:
        model_used = 'cv'
    else:
        // Final fallback to YOLO with adaptive threshold
        boxes = yolo_detect(image, confidence=0.35)
        if boxes.count == 0:
            boxes = yolo_detect(image, confidence=0.20)
        model_used = 'yolo'

// Always perform CNN classification
classification = cnn_classify(image)

// Generate natural language explanation
analysis_text = gemini_llm_analyze(boxes, classification)

return {
    detections: boxes,
    classification: classification,
    explanation: analysis_text,
    model_used: model_used
}

```

APPENDIX B

CẤU TRÚC DỮ LIỆU CHÍNH

Các cấu trúc dữ liệu JSON được sử dụng trong hệ thống để lưu trữ kết quả phân tích và metadata của từng thành phần AI.

REFERENCES

- [1] Ultralytics, “YOLOv8: State-of-the-Art Object Detection,” <https://docs.ultralytics.com/>, accessed Nov. 2024.
- [2] G. Bradski, “The OpenCV Library,” *Dr. Dobbs’ Journal of Software Tools*, vol. 25, no. 11, pp. 120-123, 2000.
- [3] Google DeepMind, “Gemini: A Family of Highly Capable Multimodal Models,” Technical Report, 2024.
- [4] S. Prajapati et al., “Classification of dental diseases using CNN and transfer learning,” in *Proc. 5th Int. Conf. Computing for Sustainable Global Development*, 2018, pp. 70-74.
- [5] D. Shen et al., “Deep learning in medical image analysis,” *Annual Review of Biomedical Engineering*, vol. 19, pp. 221-248, 2017.
- [6] M. Grinberg, *Flask Web Development: Developing Web Applications with Python*, 2nd ed. O’Reilly Media, 2018.
- [7] F. Mertens et al., “Artificial intelligence for caries detection: A systematic review,” *Journal of Dentistry*, vol. 102, p. 103924, 2020.
- [8] L. Rokach, “Ensemble-based classifiers,” *Artificial Intelligence Review*, vol. 33, no. 1-2, pp. 1-39, 2010.