

แบบฟอร์มเสนอหัวข้อวิชาโครงการ

สำหรับนักศึกษาชั้นปีที่ 3 ปริญญาตรี คณะเทคโนโลยีสารสนเทศ สจล.

1. สาขาวิชา

- สาขาวิชาวิทยาการข้อมูลและการวิเคราะห์เชิงธุรกิจ

2. ชื่อหัวข้อโครงการ

ภาษาไทย: HIKARI: ระบบวินิจฉัยรอยโรคช่องปากอัจฉริยะแบบหลายภาษาด้วยโมเดลวิชันและภาษาแบบสามขั้นตอนร่วมกับกลไกค้นหาเชิงภาพเพื่อคำแนะนำทางคลินิกที่มีหลักฐาน

- Healthcare-oriented
- Intelligent
- Knowledge-Augmented
- Retrieval and
- Inference system

ภาษาอังกฤษ: HIKARI: A Three-Stage Vision-Language Model for Multilingual Oral Lesion Diagnosis with Retrieval-Augmented Visual Grounding and Evidence-Based Clinical Recommendations

- Healthcare-oriented
- Intelligent
- Knowledge-Augmented
- Retrieval and
- Inference system

3. ข้อมูลนักศึกษา

- | | |
|---|-----------------------|
| • นักศึกษาคนที่ 1: นายวาริน พรมฝ่าย | รหัสนักศึกษา 66070184 |
| • นักศึกษาคนที่ 2: นส.ภาวดีรา บุญประสาท | รหัสนักศึกษา 66070304 |

4. ชื่อ อาจารย์ที่ปรึกษา

[ชื่อ-นามสกุลอาจารย์ที่ปรึกษา]

5. วัตถุประสงค์

- เพื่อพัฒนาระบบ HIKARI ซึ่งเป็นโมเดลวิชันและภาษาแบบสามขั้นตอน (Three-Stage Vision-Language Model) ที่สามารถวินิจฉัยรอยโรคช่องปากได้อย่างแม่นยำผ่านการรับคำสั่งที่ขับข้อนในหลายภาษา โดยอาศัยเทคนิค Complex Instruction Grounding ที่ปรับปรุงจากแนวทาง Ground-V
- เพื่อพัฒนากลไก Knowledge-Augmented Retrieval and Inference (ส่วน K-A-R-I ของ HIKARI) ที่สามารถค้นหากรณีด้วยทางคลินิกที่คล้ายคลึงพร้อม segmentation masks และผลลัพธ์การรักษา เพื่อนำมาเป็นหลักฐานในการให้คำแนะนำทางคลินิกที่มีความน่าเชื่อถือ
- เพื่อสร้างระบบ Healthcare-oriented Intelligent system (ส่วน H-I ของ HIKARI) ที่มีความสามารถในการป้องกัน hallucination ทางการแพทย์ ประเมินความไม่แน่นอน (Uncertainty Quantification) และรู้จักเมื่อความแนวโน้มที่ปรึกษาผู้เชี่ยวชาญ เพื่อความปลอดภัยในการใช้งานทางคลินิก
- เพื่อสร้างชุดข้อมูลคำสั่งที่ขับข้อนสำหรับโรคช่องปากในหลายภาษา และประเมินประสิทธิภาพของระบบ HIKARI ในการทำงานข้ามภาษา (Cross-lingual Transfer) รวมถึงความสามารถในการทำงานแบบ zero-shot

6. ที่มาและความสำคัญ

การวินิจฉัยรอยโรคในช่องปาก เช่น oral leukoplakia, oral lichen planus, recurrent aphthous ulcer และ discoid lupus erythematosus เป็นกระบวนการที่ต้องอาศัยความเชี่ยวชาญสูง เนื่องจากลักษณะของรอยโรค มีความหลากหลายและคล้ายคลึงกัน การวินิจฉัยที่ผิดพลาดอาจส่งผลกระทบแรง โดยเฉพาะกรณีที่มีความเสี่ยงต่อการกลایพันธุ์เป็นมะเร็ง ซึ่งมีอัตราการกลایพันธุ์ 0.13-34% สำหรับ oral leukoplakia

ระบบช่วยวินิจฉัยที่มืออยู่ในปัจจุบันมีข้อจำกัดสำคัญหลายประการ ประการแรก ระบบส่วนใหญ่ใช้คำสั่งที่เรียนร่าง่าย เช่น "โรคนี้คืออะไร" ซึ่งไม่สะท้อนความขับข้อนของการวินิจฉัยจริงที่แพทย์ต้องพิจารณาหลายปัจจัย ประการที่สอง ระบบไม่สามารถจัดการกับคำสั่งที่ขับข้อน เช่น "segment บริเวณที่มีรอยโรคที่มีความเสี่ยงสูงต่อการกลایเป็นมะเร็ง" หรือ "แสดงบริเวณที่ต้องติดตามอย่างใกล้ชิด" ประการที่สาม มีปัญหา hallucination คือระบบอาจารย์วินิจฉัยโรคที่ไม่มีในภาพ ซึ่งเป็นอันตรายต่อผู้ป่วย ประการที่สี่ ขาดความ

สามารถให้คำแนะนำทางคลินิกที่มีหลักฐานสนับสนุนจากการณ์ตัวอย่างจริง และประการสุดท้าย รองรับเฉพาะภาษาเดียว ทำให้จำกัดการใช้งานในบริบทที่หลากหลาย

งานวิจัยล่าสุดในด้าน Vision-Language Models (VLMs) เช่น OralGPT ได้แสดงให้เห็นศักยภาพในการวินิจฉัยโรคช่องปาก แต่ยังมีข้อจำกัดในการทำ pixel-level grounding ภายใต้คำสั่งที่ขับช้อน งานวิจัย Ground-V ได้เสนอแนวทางในการสร้างชุดข้อมูลคำสั่งที่ขับช้อน 5 ประเภท ได้แก่ multi-granularity, multi-object, hallucination mitigation, reasoning และ part-level reference แต่ยังไม่มีการประยุกต์ใช้ในโดเมนการแพทย์ นอกจากนี้ ระบบที่มีอยู่ยังขาดกลไกในการให้คำแนะนำที่อิงจากหลักฐานทางคลินิกจริง

โครงการ HIKARI นี้จึงมุ่งพัฒนาระบบวินิจฉัยรอยโรคช่องปากที่ครบถ้วนและกว้างล้ำกว่าระบบเดิม โดย:

- **Healthcare-oriented (H):** ออกแบบเฉพาะสำหรับการใช้งานทางการแพทย์ โดยเน้นความปลอดภัยและความน่าเชื่อถือ
- **Intelligent (I):** ใช้เทคนิค Three-Stage Vision-Language Model ที่ผ่านการเรียนรู้แบบ classification, complex grounding และ captioning
- **Knowledge-Augmented Retrieval (K-A-R):** พัฒนากลไกการค้นหาที่ใช้ knowledge base ของกรณีทางคลินิกจริงพร้อมผลลัพธ์การรักษา
- **Inference system (I):** สร้างระบบอนุมานที่สามารถให้คำแนะนำทางคลินิกที่มีหลักฐานสนับสนุน พร้อมประเมินความแน่นอนของการวินิจฉัย

ระบบ HIKARI จะรองรับการทำงานในหลายภาษา สามารถทำ zero-shot cross-lingual transfer และมีกลไก hallucination mitigation ที่แข็งแกร่ง ซึ่งจะช่วยเพิ่มการเข้าถึงบริการสุขภาพช่องปากที่มีคุณภาพ สำหรับประชากรที่หลากหลาย และลดความเสี่ยงจากการวินิจฉัยที่ผิดพลาด

7. วิธีการดำเนินงาน (หลักการสำคัญ)

7.1 การเตรียมข้อมูลและสร้าง Knowledge Base

1. รวบรวมภาพถ่ายช่องปากจากแหล่งข้อมูลที่มีอยู่ พร้อม segmentation masks และ disease labels
2. สร้าง captions ทางคลินิกในหลายภาษาโดยใช้ Large Language Model (Claude) ร่วมกับการตรวจสอบโดยผู้เชี่ยวชาญตามสัดส่วนที่เหมาะสม (10-20%)
3. สร้างชุดคำสั่งที่ซับซ้อน (complex instructions) 5 ประเภทตามแนวทาง Ground-V โดยปรับให้เหมาะสมกับโภemen การแพทย์:
 - **Multi-granularity:** คำสั่งที่มีระดับความเฉพาะเจาะจงต่างกัน เช่น "รอยโรค" → "white patch" → "homogeneous leukoplakia on left buccal mucosa"
 - **Multi-object:** คำสั่งที่ต้องการ segment หลายบริเวณพร้อมกัน เช่น "segment all erosive areas requiring monitoring"
 - **Hallucination mitigation:** คำสั่งที่มีข้อมูลที่ไม่มีในภาพ เช่น "segment melanoma" เพื่อฝึกให้ระบบปฏิเสธได้
 - **Reasoning:** คำสั่งที่ต้องการการให้เหตุผลทางคลินิก เช่น "which lesions show high malignant potential?"
 - **Part-level reference:** คำสั่งที่อ้างถึงส่วนย่อยของรอยโรค เช่น "segment the erythematous border of the ulcer"
4. สร้าง Knowledge Base โดยจัดเก็บกรณีทางคลินิกพร้อม:
 - Visual features (จาก image encoder)
 - Segmentation mask features
 - Clinical descriptions
 - Treatment information
 - Clinical outcomes (improved/stable/worsened)

7.2 การพัฒนา HIKARI: Three-Stage Vision-Language Model

Stage 1 - Healthcare-oriented Classification (H-component):

1. ฝึกโมเดลด้วยคำสั่งแบบ binary classification ("Is this OK?" → "Yes/No")
2. เน้นการสร้างพื้นฐานความเข้าใจในลักษณะของโรคแต่ละชนิด
3. ใช้ loss function: Binary Cross-Entropy
4. Architecture: Vision Encoder (CLIP) + Text Encoder + Alignment Module + LLM (Qwen2.5-VL-7B)
5. Training: 10 epochs with LoRA fine-tuning

Stage 2 - Intelligent Complex Grounding (I-component):

1. Initialize จาก weights ของ Stage 1
2. ฝึกโมเดลด้วยชุดคำสั่งที่ซับซ้อนทั้ง 5 ประเภท
3. Model เรียนรู้ที่จะ generate segmentation tokens (<seg1>, <seg2>, etc.)
4. Segmentation tokens ถูกส่งไปยัง SAM decoder เพื่อสร้าง pixel-level masks
5. ฝึกให้โมเดลสามารถ:

- Understand instructions ที่มีความซับซ้อนหลายระดับ
 - Reject hallucinated references (ปฏิเสธเมื่อไม่มีในภาพ)
 - Reason about clinical implications
 - Segment multiple objects และ part-level details
6. Loss function: IoU Loss + Cross-Entropy
 7. Training: 5 epochs on multilingual complex instruction dataset

Stage 3 - Captioning with Clinical Context:

1. Initialize จาก weights ของ Stage 2
2. ฝึกโมเดลให้สามารถสร้าง clinical descriptions ที่มีรายละเอียดในหลายภาษา
3. Input: Image + Captioning prompt ("Please describe the symptoms shown in this image")
4. Output: Clinical narrative describing location, appearance, characteristics
5. Loss function: Cross-Entropy (next token prediction)
6. Training: 5 epochs with full-resolution images

7.3 การพัฒนา Knowledge-Augmented Retrieval (K-A-R component)

1. Encoding Module:

- Visual Encoder: CLIP/DINOv2 สำหรับ encode images
- Mask Encoder: สำหรับ encode segmentation patterns
- Text Encoder: BioBERT สำหรับ encode clinical descriptions

2. Vector Database (FAISS):

- สร้าง index สำหรับ visual features, mask features และ text features
- จัดเก็บ metadata: disease type, treatment, outcomes, follow-up duration

3. Retrieval Mechanism:

- Query encoding: ผสม features จาก image + predicted masks + instruction
- Similarity search: ค้นหา top-k similar cases โดยใช้:
 - Visual similarity (cosine similarity of image features)
 - Semantic similarity (text feature matching)
 - Cross-modal score fusion
- Adaptive k: ถ้า uncertainty สูง → retrieve มากรีน (k=10), ถ้าต่ำ → น้อยลง (k=3)

4. Evidence Aggregation:

- วิเคราะห์ outcomes จากกรณีที่ retrieve มา
- คำนวณ success rate, average follow-up time, complications
- สร้าง evidence strength score

7.4 การพัฒนา Inference System (I-component) with Uncertainty Quantification

1. Uncertainty Estimation:

- ใช้ Monte Carlo Dropout: forward pass หลายรอบด้วย dropout enabled
- คำนวณ variance ของ predictions
- Uncertainty score = standard deviation of predictions

2. Adaptive Inference:

- ถ้า uncertainty > threshold (0.7):
 - Retrieve มากขึ้น ($k=10$)
 - แนะนำให้ปรึกษาผู้เชี่ยวชาญ
 - ใช้ temperature ต่ำในการ generate (0.3) เพื่อความระมัดระวัง
- ถ้า uncertainty ต่ำ:
 - Retrieve น้อยลง ($k=3$)
 - Proceed with automated recommendation
 - ใช้ temperature ปกติ (0.7)

3. Evidence-Based Recommendation Generation:

- Input: Current case + Retrieved cases + Uncertainty score
- Process:
 - วิเคราะห์ treatment patterns จาก retrieved cases
 - ระบุ best practices จาก successful cases
 - สร้าง recommendation ที่มี evidence citation

Output format:

Diagnosis: [Disease name] (Confidence: XX%) Segmentation: [Areas with <seg> tokens]
Recommendation: Based on X similar cases (Y% success rate):

- Treatment: [Treatment name]
- Expected outcome: [Based on evidence]
- Monitoring: [Areas to watch]
- Evidence: [Case references]

[If uncertain: "Specialist consultation recommended"]

7.5 การประเมินประสิทธิภาพ HIKARI

1. Classification Performance:

- Metrics: Accuracy, Precision, Recall, F1-score
- Evaluation: Per-disease และ macro-average

2. Complex Grounding Performance:

- Segmentation metrics: IoU, gloU, cloU
- Evaluate แยกตาม 5 types of instructions
- Hallucination rate: N-Accuracy (ability to abstain correctly)

3. Multilingual Performance:

- Monolingual baseline: Thai-only, English-only
- Bilingual model: Train on both
- Zero-shot transfer: Train Thai → Test English (and vice versa)

4. Caption Quality:

- Lexical metrics: BLEU-1, BLEU-4, METEOR, ROUGE-L
- Semantic evaluation: Using DeepSeek-V3 or similar LLM judge
- Clinical accuracy: Expert evaluation

5. Retrieval Quality:

- Retrieval accuracy: % of relevant cases retrieved
- Evidence strength correlation: Does evidence match outcomes?

6. Clinical Validation:

- Expert evaluation of recommendations (50-100 cases)
- Agreement rate with expert suggestions
- Safety assessment: Check for harmful recommendations

7. Explainability Analysis:

- Attention visualization
- Similarity scores interpretation
- Evidence traceability

7.6 การพัฒนา User Interface

1. Web Interface (สำหรับบุคลากรทางการแพทย์):

- Upload image + enter instruction
- View segmentation overlay
- Access detailed report with similar cases
- Review evidence and confidence scores

2. LINE API Integration (สำหรับประชาชนทั่วไป):

- Send image via chat
 - Simple question-answer format
 - Receive preliminary screening
 - Guidance on when to see doctor
-

8. ขอบเขตของงาน

1. ข้อมูลและโรค:

- ใช้ภาพถ่ายช่องปากที่มีอยู่จากแหล่งข้อมูลต่างๆ ประมาณ ... ภาพ
- ชนิดของโรคขึ้นอยู่กับข้อมูลที่ได้รับจริง (อาจจะ OLK, OLP, ROU, DLE หรือโรคอื่นที่มีข้อมูลเพียงพอ)
- ทุกภาพต้องมี segmentation masks และ disease labels

2. ภาษาที่รองรับ:

- รองรับอย่างน้อย 2 ภาษา (ขึ้นอยู่กับข้อมูลที่มี อาจเป็น ไทย-อังกฤษ หรือคู่ภาษาอื่น)
- สร้าง captions และ instructions ในทั้ง 2 ภาษา
- ทดสอบ zero-shot cross-lingual transfer

3. สถาปัตยกรรม HIKARI:

Stage 1 - Classification:

- Input: Image + Simple binary prompt
- Architecture: Vision Encoder (CLIP-ViT) + Text Encoder + Aligner + LLM (Qwen2.5-VL-7B)
- Training: 10 epochs with LoRA (rank=16)
- Output: Binary classification (Yes/No)

4. Stage 2 - Complex Grounding:

- Input: Image + Complex instruction (5 types)
- Initialize from Stage 1 weights
- Add: SAM decoder for segmentation
- Training: 5 epochs on complex instruction dataset
- Output: Segmentation tokens → Pixel-level masks

5. Stage 3 - Captioning:

- Input: Image + Captioning prompt
- Initialize from Stage 2 weights
- Training: 5 epochs with full-resolution images
- Output: Clinical description in target language

6. K-A-R Module (Knowledge-Augmented Retrieval):

- Visual Encoder: CLIP/DINOv2
- Mask Encoder: Custom CNN
- Text Encoder: BioBERT
- Database: FAISS vector index
- Retrieval: Top-k similar cases with adaptive k
- Output: Retrieved cases + Evidence + Recommendation

7. ขดลักษณะที่สร้าง:

Captions: ~1,000 clinical descriptions (generated + validated)

- Complex instructions: 5,000-10,000 instruction-mask pairs
- 5 types: Multi-granular, Multi-object, Hallucination, Reasoning, Part-level
- ทั้งหมดมีทั้ง 2 ภาษา

8. การประเมินประสิทธิภาพ:

- Classification: Accuracy, Precision, Recall, F1-score (per-disease + macro-avg)
- Segmentation: IoU, gloU, cloU (per instruction type)
- Hallucination: N-Accuracy (ability to abstain)
- Caption: BLEU, METEOR, ROUGE, Semantic score
- Multilingual: Monolingual vs Bilingual vs Zero-shot transfer
- Retrieval: Retrieval accuracy, Evidence correlation
- Clinical: Expert validation (50-100 cases)

9. เทคนิคที่ใช้:

- Vision-Language Models: Qwen2.5-VL-7B (base LLM)
- Vision Encoders: CLIP-ViT, DINOv2 (**May be use DINOv3**)
- Segmentation: **SAM (Segment Anything Model)**
- Text Encoders: **BioBERT** (for medical text)
- Data Generation: **Claude 3 Sonnet** (for captions + instructions)
- Fine-tuning: LoRA (Low-Rank Adaptation) (**May be use QLoRA**)
- Uncertainty: **Monte Carlo Dropout**
- Retrieval: FAISS (vector similarity search)

10. การ Deploy:

- Web Interface: สำหรับบุคลากรทางการแพทย์
- LINE API: สำหรับประชาชนทั่วไป
- พิจารณา: Response time, User experience, Data privacy

11. ข้อจำกัด:

- ไม่รวมการเก็บข้อมูลผู้ป่วยจริงใหม่ (ใช้เฉพาะข้อมูลที่มีอยู่)
- ระบบเป็นเครื่องมือช่วยเหลือ (assistive tool) ไม่ทดแทนแพทย์
- จำนวนภาษาและโรคขั้นอยู่กับข้อมูลที่ได้รับ
- ไม่รับประกันความแม่นยำ 100%
- ต้องผ่านการตรวจสอบจากผู้เชี่ยวชาญก่อนใช้งานจริงในคลินิก

9. ประโยชน์ที่คาดว่าจะได้รับ

1. ด้านนวัตกรรมทางเทคนิค:

- ได้ระบบ HIKARI ซึ่งเป็น Three-Stage Vision-Language Model แรกที่ออกแบบเฉพาะสำหรับการวินิจฉัยโรคช่องปากแบบหลายภาษา
- ได้กลไก Knowledge-Augmented Retrieval (K-A-R) แบบใหม่ที่รวม visual exemplars พร้อม segmentation masks และผลลัพธ์ทางคลินิกเข้าด้วยกัน ซึ่งเป็นงานแรกที่ทำใน medical imaging domain
- ได้ชุดข้อมูล complex instructions สำหรับโรคช่องปากในหลายภาษา (~10,000 pairs) ที่สามารถนำไปใช้วิจัยต่อได้
- มีความรู้เชิงลึกเกี่ยวกับ hallucination mitigation และ uncertainty quantification ในระบบ AI ทางการแพทย์
- แสดงให้เห็นความสามารถของ zero-shot cross-lingual transfer ในโดเมนการแพทย์

2. ด้านการแพทย์และสาธารณสุข:

- เพิ่มความแม่นยำและความปลอดภัยในการวินิจฉัยเบื้องต้น โดย HIKARI สามารถ:
 - วินิจฉัยโรคจากคำสั่งที่ซับซ้อน (complex instructions)
 - ปฏิเสธการวินิจฉัยเมื่อไม่แน่ใจ (hallucination mitigation)
 - แนะนำให้ปรึกษาผู้เชี่ยวชาญเมื่อ uncertainty สูง
- ให้คำแนะนำทางคลินิกที่มีหลักฐานสนับสนุน (Evidence-based recommendations) เช่น "12/15 กรณีที่คล้ายกันหายด้วยวิธีการรักษา X"
- ลดภาระของบุคลากรทางการแพทย์ในการคัดกรองผู้ป่วยเบื้องต้น
- เพิ่มการเข้าถึงบริการประเมินสุขภาพช่องปากสำหรับประชาชนในหลายกลุ่มภาษา

3. ด้านวิชาการและการเผยแพร่:

- มีศักยภาพในการตีพิมพ์ในที่ประชุมวิชาการระดับนานาชาติชั้นนำ (A* conference) เช่น:
 - CVPR/ICCV (Computer Vision): Visual grounding + Retrieval
 - NeurIPS/ICML (Machine Learning): Multilingual learning + RAG
 - MICCAI (Medical Imaging): Clinical application + Validation
 - ACL/EMNLP (NLP): Multilingual instruction understanding
- เป็นงานวิจัยแรกที่รวม Ground-V methodology กับ medical imaging
- เป็นงานวิจัยแรกที่ทำ Retrieval-Augmented Grounding พร้อม clinical outcomes

4. ด้านการประยุกต์ใช้:

- สามารถ deploy จริงผ่าน Web interface และ LINE API
- เป็นดันแบบสำหรับการพัฒนา multilingual medical AI ในโรคอื่นๆ
- Framework ของ HIKARI สามารถปรับใช้กับโดเมนการแพทย์อื่นได้ เช่น:
 - Skin lesions (dermatology)
 - X-ray interpretation (radiology)
 - Fundus images (ophthalmology)

5. ด้านการเรียนรู้และพัฒนาทักษะ:

- ประสบการณ์ในการพัฒนาระบบ AI ที่มีความรับผิดชอบ(Responsible AI): - Safety: Uncertainty quantification, Hallucination mitigation - Explainability: Evidence-based recommendations, Attention visualization - Fairness: Multilingual support
- ทักษะในการทำงานกับ state-of-the-art models:
 - Vision-Language Models (Qwen2.5-VL, CLIP, SAM)
 - Large Language Models (Claude for data generation)
 - Retrieval systems (FAISS, Vector databases)
- ความเข้าใจในการพัฒนา three-stage training pipeline
- ประสบการณ์ในการสร้างและ validate ชุดข้อมูลทางการแพทย์

6. ผลกระทบต่อสังคม:

- ลดความเหลื่อมล้ำในการเข้าถึงบริการสุขภาพช่องปาก โดยเฉพาะในพื้นที่ห่างไกล
- ให้ความรู้และเพิ่มความตระหนักรู้กับสุขภาพช่องปากแก่ประชาชน
- สนับสนุนการตรวจสอบโรคเบื้องต้นได้ง่ายและรวดเร็ว
- เป็นเครื่องมือช่วยในการศึกษาของนิสิตแพทย์และทันตแพทย์

ลงชื่อ นักศึกษาผู้เสนอโครงการ วันที่ / /

ลงชื่อ นักศึกษาผู้เสนอโครงการ วันที่ / /

อาจารย์ที่ปรึกษา

ลงชื่อ ได้พิจารณาและอนุมัติหัวข้อดังกล่าวข้างต้น

(.....)

วันที่/...../.....

ผลการอนุมัติจากคณะกรรมการ

อนุมัติ

ไม่อนุมัติ

○