UNIVERSITAS
GADJAH MADA
MAMMOGRAM CLASSIFICATION WITH PROMPT-GUIDED MULTI-VIEW VISION LANGUAGE MODEL
Ahmad Alif Robit Al Hazmi, Moh. Edi Wibowo, S.Kom., M.Kom., Ph.D.;Professor Jia-Ching Wang
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

ABSTRACT

MAMMOGRAM CLASSIFICATION WITH PROMPT-GUIDED MULTI-VIEW VISION LANGUAGE MODEL

by

Ahmad Alif Robit Al Hazmi

23/527472/PPA/06680

In the era of artificial intelligence development in the medical field, mammogram image analysis demands a model capable of simultaneously processing visual information and clinical context to improve the accuracy of breast cancer diagnosis. Because conventional visual-only approaches often fail to optimally capture the semantic relationship between images and clinical descriptions, this study developed a Prompt-Guided Multi-View Vision Language Model. The goal is to improve multimodal representation by combining information from various image views and descriptive text guidance. This study used the VinDr-Mammo dataset containing 5,000 examinations from four main views (LCC, LMLO, RCC, RMLO) along with clinical information such as masses, calcifications, and BI-RADS. The method includes a pre-training stage with contrastive learning to align image and text embeddings, and a fine-tuning stage with a multi-view prompt scheme. These prompt texts are structured using a Large Language Model (LLM) to generate structured and clinically relevant visual descriptions. Evaluation with Area Under the Curve (AUC), Balanced Accuracy (B-ACC), and computational time metrics proves that the proposed model outperforms the baseline. In the zero-shot scenario, the model achieved an AUC of 0.85 and a B-ACC of 0.78 for the classification category. This performance improved further in the linear probe scenario to an AUC of 0.91 and a B-ACC of 0.85, while maintaining computational efficiency. In conclusion, rapid integration and multi-views have proven effective in improving the model's discrimination capabilities, despite challenges in cases with complex visual characteristics, such as those that still occur. This study emphasizes the importance of a multimodal approach to improving the accuracy and reliability of AI diagnostic systems in the real world.

Keyword: Computer Vision, Image Classification, Mammogram Analysis, Vision- Language Models, Prompt Learning

UNIVERSITAS
GADJAH MADA
MAMMOGRAM CLASSIFICATION WITH PROMPT-GUIDED MULTI-VIEW VISION LANGUAGE MODEL
Ahmad Alif Robit Al Hazmi, Moh. Edi Wibowo, S.Kom., M.Kom., Ph.D.;Professor Jia-Ching Wang
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

INTISARI

MAMMOGRAM CLASSIFICATION WITH PROMPT-GUIDED MULTI-VIEW VISION LANGUAGE MODEL

Oleh

Ahmad Alif Robit Al Hazmi

23/527472/PPA/06680

Di era perkembangan kecerdasan buatan di bidang medis, analisis citra mamogram menuntut sebuah model yang mampu memproses informasi visual dan konteks klinis secara bersamaan demi meningkatkan akurasi diagnosis kanker payudara. Karena pendekatan konvensional yang hanya berbasis visual kerap gagal menangkap hubungan semantik antara citra dan deskripsi klinis secara optimal, penelitian ini mengembangkan Prompt-Guided Multi-View Vision Language Model. Tujuannya adalah untuk meningkatkan representasi multimodal dengan memadukan informasi dari berbagai sudut pandang citra dan panduan teks deskriptif. Penelitian ini menggunakan dataset VinDr-Mammo yang berisi 5.000 pemeriksaan dari empat pandangan utama (LCC, LMLO, RCC, RMLO) beserta informasi klinis seperti massa, kalsifikasi, dan BI-RADS. Metodenya mencakup tahap pra-pelatihan dengan pembelajaran kontrastif untuk menyelaraskan embedding citra dan teks, serta tahap fine-tuning dengan skema prompt multi-view. Prompt teks ini disusun menggunakan Large Language Model (LLM) untuk menghasilkan deskripsi visual yang terstruktur dan relevan secara klinis. Evaluasi dengan metrik Area Under the Curve (AUC), Balanced Accuracy (B-ACC), dan waktu komputasi membuktikan bahwa model yang diusulkan konsisten melampaui baseline. Pada skenario zero-shot, model mencapai AUC 0,85 dan B-ACC 0,78 untuk kategori kalsifikasi. Performa ini kembali meningkat pada skenario linear probe menjadi AUC 0,91 dan B-ACC 0,85, sembari tetap mempertahankan efisiensi waktu komputasi. Kesimpulannya, integrasi prompt dan multi-view terbukti efektif meningkatkan kemampuan diskriminasi model, kendati tantangan pada kasus dengan karakteristik visual yang kompleks seperti massa masih ada. Studi ini menegaskan pentingnya pendekatan multimodal demi meningkatkan akurasi dan keandalan sistem diagnosis AI di dunia nyata.

Kata Kunci: Visi Komputer, Klasifikasi Citra, Analisis Mammogram, Model Bahasa-Visual, Pembelajaran Prompt