

Medical Image Retrieval Based on Ensemble Learning using Convolutional Neural Networks and Vision Transformers

Y. Ahmed Yahya^{1*}, D. Khaled², W. Khaile Al-Azzawi³, T. Alghazali⁴, H. Sabah Jabr⁵, R. Madhat Abdulla⁶, M. Kadhim Abbas Al-Maeni⁷, N. Hussin Alwan⁸, S. Saad Najeeb⁹, Kh. T. Falih¹⁰

1- Department of nursing, Al-Hadba University College, Iraq.

Email: yahya@hcu.edu.iq (Corresponding author)

2 Al-Manara College for Medical Sciences, Maysan, Iraq.

Email: dalyakhaled@uomanara.edu.iq

3- Medical Lab. Techniques department, College of Medical Technology, Al-Farahidi University, Iraq.

Email: waleed.khalid@alfarahidiuc.edu.iq

4- College of Media, Department of Journalism, The Islamic University in Najaf, Najaf, Iraq.

Email: gazali.tawfeeq@gmail.com

5- Anesthesia Techniques Department, Al-Mustaqbal University College, Babylon, Iraq.

Email: huda.sabah@mustaqbal-college.edu.iq

6- The University of Mashreq, Baghdad, Iraq.

Email: rusul.madhat@uom.edu.iq

7- Al-Nisour University College, Baghdad, Iraq.

Email: mohammed.kadhimaam@gmail.com

8- Department of Nursing, Al-Zahrawi University College, Karbala, Iraq.

Email: natherahussin.alwan@g.alzahu.edu.iq

9- Al-Esraa University College, Baghdad, Iraq

Email: salma@esraa.edu.iq

10- New Era and Development in Civil Engineering Research Group, Scientific Research Center, Al-Ayen University, Thi-Qar, Iraq.

Email: khaldoon@alayen.edu.iq

Received: April 2022

Revised: June 2022

Accepted: July 2022

ABSTRACT:

The rapid increase in the number of medical image repositories nowadays has led to problems in managing and retrieving medical visual data. This has proved the necessity of Content-Based Image Retrieval (CBIR) with the aim of facilitating the investigation of such medical imagery. One of the most serious challenges that require special attention is the representational quality of the embeddings generated by the retrieval pipelines. These embeddings should include global and local features to obtain more useful information from the input data. To fill this gap, in this paper, we propose a CBIR framework that utilizes the power of deep neural networks to efficiently classify and fetch the most related medical images with respect to a query image. Our proposed model is based on combining Vision Transformers (ViTs) and Convolutional Neural Networks (CNNs) and learns to capture both the locality and also the globality of high-level feature maps. Our method is trained to encode the images in the database and outputs a ranking list containing the most similar image to the least similar one to the query. To conduct our experiments, an intermodal dataset containing ten classes with five different modalities is used to train and assess the proposed framework. The results show an average classification accuracy of 95.32 % and a mean average precision of 0.61. Our proposed framework can be very effective in retrieving multimodal medical images with the images of different organs in the body.

KEYWORDS: Content-Based Image Retrieval, Medical Image Retrieval, Ensemble Learning, Convolutional Neural Networks, Vision Transformers, Deep Learning, Similarity-Based Visual Search.