

# Medical Visual Question Answering

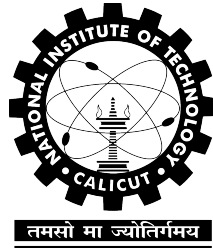
CS4099D Project  
End Semester Report

*Submitted by*

Nidhin Mohan (B180948CS)  
Shaaheen A M (B181134CS)  
Abhinav B Naik (B170297CS)

*Under the Guidance of*

Dr. Saidalavi Kalady  
Dr. S. Sheerazuddin

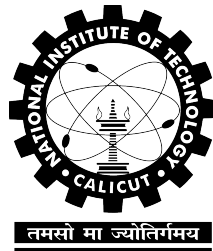


Department of Computer Science and Engineering  
National Institute of Technology Calicut  
Calicut, Kerala, India - 673 601

May, 2022

NATIONAL INSTITUTE OF TECHNOLOGY CALICUT  
KERALA, INDIA - 673 601

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING



## CERTIFICATE

*Certified that this is a bonafide report of the project work titled*

**MEDICAL VISUAL QUESTION ANSWERING**

*done by*

**Nidhin Mohan**

**Shaaheen A M**

**Abhinav B Naik**

*of Eighth Semester B. Tech, during the Winter Semester 2021-'22, in  
partial fulfillment of the requirements for the award of the degree of  
Bachelor of Technology in Computer Science and Engineering of the  
National Institute of Technology, Calicut.*

05-05-2022

**Date**

(Dr. Saidalavi Kalady)

(Dr. S. Sheerazuddin)

**Project Guide**

# DECLARATION

I hereby declare that the project titled, **Medical Visual Question Answering**, is our own work and that, to the best of our knowledge and belief, it contains no material previously published or written by another person nor material which has been accepted for the award of any other degree or diploma of the university or any other institute of higher learning, except where due acknowledgement and reference has been made in the text.

Place : NIT Calicut  
Date : 05-05-2022

Name : Nidhin Mohan  
Roll. No. : B180948CS  
Name : Shaaheen A M  
Roll. No. : B181134CS  
Name : Abhinav B Naik  
Roll. No. : B170297CS

## **Abstract**

The Visual Question Answering(VQA) domain explores the problem of answering questions regarding images. VQA in the medical domain(Med-VQA), has enormous potential and applications in the medical industry, such as assisting doctors in the field of radiology. Present Med-VQA frameworks employ deep learning approaches and image and question classification techniques. However, the quantity and diversity of questions and images in the medical domain, makes the task of Med-VQA challenging. Here we present our Med-VQA framework evaluated using the SLAKE dataset, which contains 642 images and 14,028 questions, which is larger compared to other datasets such as VQA-RAD which has 315 images and 3,515 questions. We use pre-trained VGG16 model for obtaining image features. GloVe word embeddings and LSTMs are used for obtaining question features. The questions are classified into two, the image and question features are then concatenated together, which is then passed to a classifier to obtain the answers. Our model has reached an accuracy of 64.66% and 61.39% for open and closed-ended questions respectively.