

CODE DESCRIPTION

Medical Visual Question Answering (MVQA)

Medical Visual Question Answering (MVQA) is a multidisciplinary field combining computer vision and natural language processing to interpret visual data and respond to related questions in natural language. The MVQA focuses on analyzing medical images (e.g., X-rays, CT scans, MRI scans) and providing relevant answers to queries posed by medical practitioners or researchers. This capability aims to assist in diagnosis, decision-making, and streamlining workflows in clinical environments.

Fetch required inputs from the datasets

The MVQA dataset is divided into three subsets: training, validation, and test. Each subset consists of medical images paired with corresponding Question-Answer Pairs (QA-Pairs). The preparation of these inputs and computation of their required values is facilitated by the *data_utils.py* script.

Two key issues arise within the QA-Pairs and images:

1. **Imbalanced Labels:** Certain labels have disproportionately large or small sample sizes.
2. **Insufficient Samples:** Some answers have very few associated samples, affecting representativeness.

These challenges are addressed through two processes:

1. **Improvement of QA-Pairs:** Balancing and augmenting QA-Pair labels to ensure better distribution.
2. **Improvement of Medical Images:** Enhancing the dataset by increasing the sample representativeness of underrepresented labels.

Finally, the improved QA-Pairs and medical images are concatenated based on their `image_id`, resulting in an enhanced and balanced dataset ready for model training and evaluation.

Improvement of QA-Pairs

This module addresses the two primary issues in the dataset through the following steps:

Resolving the First Issue (Imbalanced Labels)

1. **Computation of Question Counts:** Calculate the number of questions associated with each label.
2. **Hard-Sample Threshold Calculation:** Determine a threshold value to identify underrepresented labels (hard samples).
3. **Selection of an Average Value:** Choose an appropriate average value to guide balancing efforts.

Resolving the Second Issue (Insufficient Samples)

New samples are generated for underrepresented labels based on the computed hard-sample threshold, ensuring a more balanced distribution across the dataset.

These steps collectively improve the quality and representativeness of the QA-Pairs and associated labels, contributing to a more robust dataset for model training and evaluation.

Improvement of Medical Images

This module addresses the dataset's key issues through the following approaches:

Resolving the First Issue (Imbalanced Labels)

1. **Image Count per Label:** Calculate the number of images associated with each label to identify imbalances.
2. **Hard-Sample Threshold:** Determine a threshold value to identify underrepresented labels.
3. **Average Value Selection:** Select an appropriate average value to guide the balancing process.

Resolving the Second Issue (Insufficient Samples)

The second issue is addressed using:

1. **Mixup:** A data augmentation technique that generates new samples by combining pairs of images and their labels, enhancing diversity.
2. **Label Smoothing:** Regularizing the labels to reduce overconfidence in the model and improve generalization.

These methods collectively enhance the dataset's balance and quality, ensuring a more effective training process.

MVQA Model Creation

From the improved dataset, features are extracted and combined to build the model.

1. Feature Extraction

- **Image Features:** Medical images are preprocessed, and features are extracted using VGGNet. These features are stored in separate files for training, validation, and testing: *train_features.pkl*, *valid_features.pkl*, and *test_features.pkl*.
- **Text Features:** Word embeddings for QA-pairs are generated using GloVe vectors (stored in the "Glove" folder). Using these embeddings, LSTM generates a sequence of words for the answers based on the timestamps.

2. Feature Combination

The extracted image features and text features are concatenated using LSTM to form a unified representation.

3. Model Generation and Prediction

A model is trained using the concatenated feature vectors. This trained model is then used to predict answers for the test set.

4. Validation

The predicted answers are evaluated using quantitative metrics to measure the model's performance.

Layerwise Relevance Propagation eXplainable Artificial Intelligence (LRP XAI)

The reasoning behind the generated answers is visualized using Layer-wise Relevance Propagation (LRP) Explainable AI (XAI). The process involves:

1. Heatmap Generation

- Gradients are computed between the most contributing layer in the model and the final output layer.
- These gradients are used to create a heatmap that highlights the regions in the input contributing most to the model's decision.

2. Superimposed Images

- The heatmap is overlaid on the input image to visualize significant regions, resulting in superimposed images.
- Examples of these visualizations for samples from the Path-VQA dataset under various scenarios are shown in *Figs. 1 and 2*.

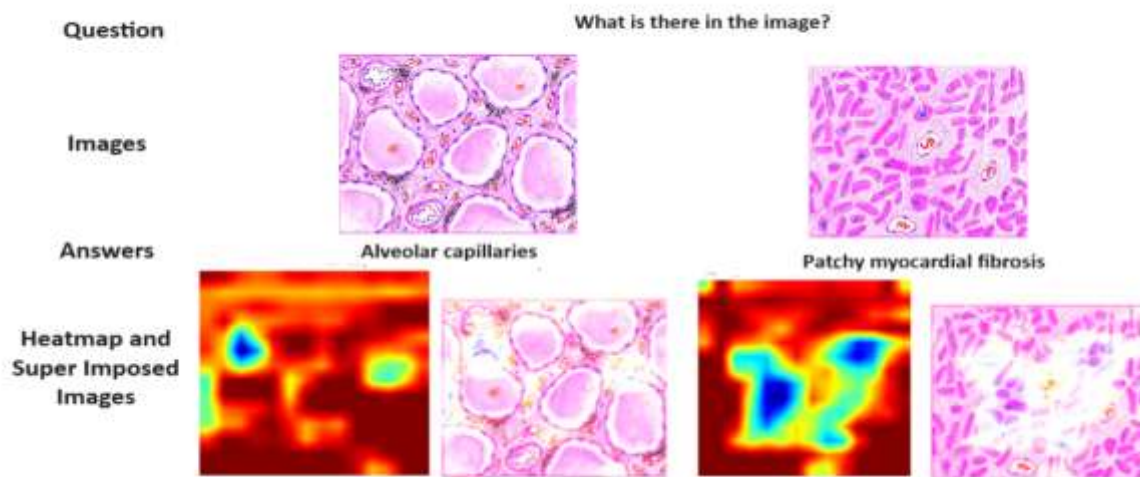


Fig 1. Visualization of heatmaps and super imposed images for different samples

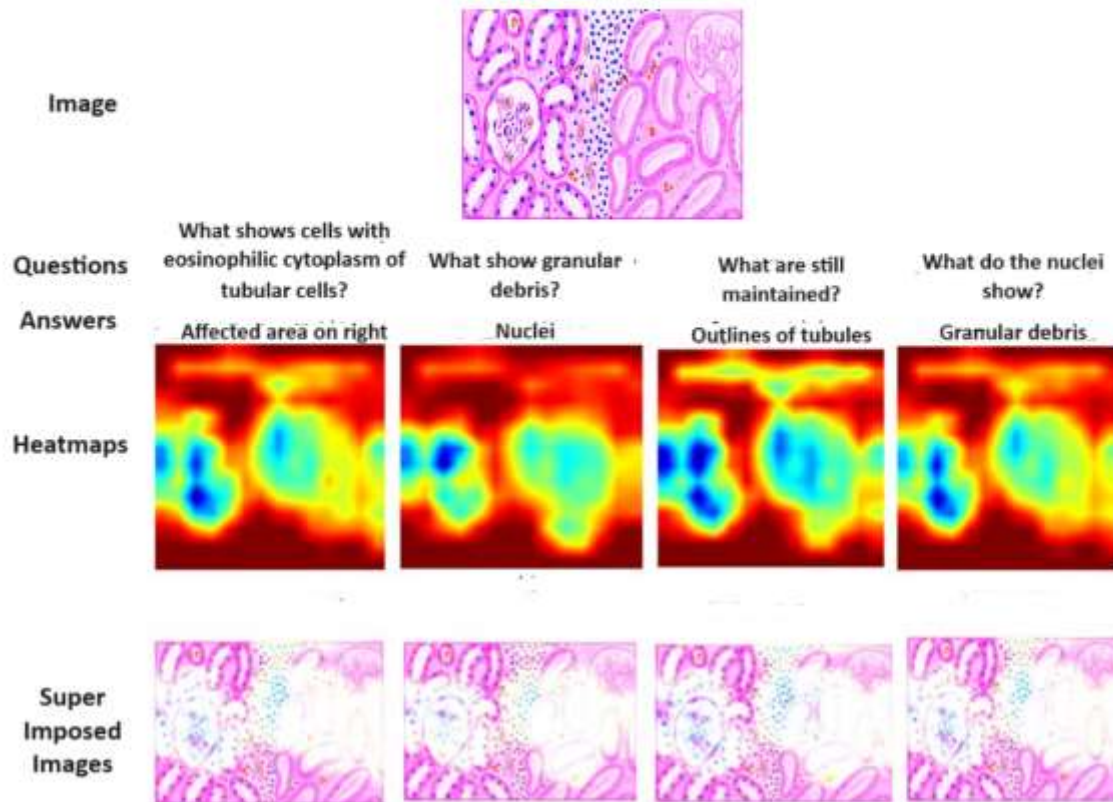


Fig 2. Visualization of heatmaps and super imposed images for four different QA pairs

3. Effect of Augmentation and Visualization

- The correctly and incorrectly classified samples from five MVQA datasets are analyzed.
- **Figs. 3 to 7** showcase the impact of data augmentation and visualization techniques on answer prediction accuracy.

This visualization approach provides insights into the model's decision-making process, enhancing interpretability and understanding of its predictions.


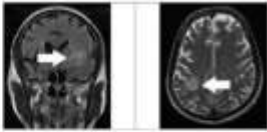

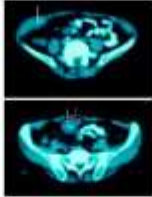


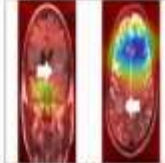

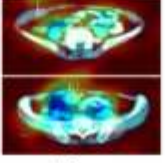
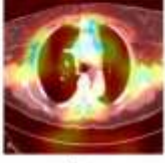
Correctly Classified Samples				Wrongly Classified Samples	
Image					
Question	What does a ct scan of the abdomen show?	What does mri demonstrate?	What is most alarming about this x-ray?	What does ct show?	What shows narrowing of the svc?
Expected Answer	Multiple low density lesions on the spleen and liver	Tuberculoma	Radial head fracture	Localisation of drain that migrated in right iliac fossa	Contrast of the chest
Predicted Answer (BDI)	Non enhancing lesion	Tuberculoma	Radial head fracture	Two enlarged nodes	CT scan of chest
Predicted Answer (ADI)	Multiple low density lesions on the spleen and liver	Tuberculoma	Radial head fracture	Meningioma fossa	Lesion
Generated XAI Output					
	(a)	(b)	(c)	(d)	(e)

Fig 3. Visualization of samples from VQA-MED 2018 dataset w.r.t predicted answer





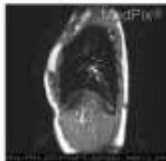
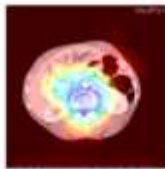

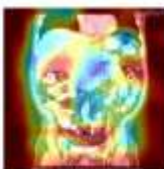
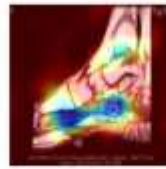
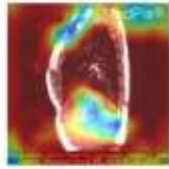
Correctly Classified Samples				Wrongly Classified Samples	
Image					
Question	What type of contrast did this patient have?	What imaging modality is used?	Which plane is this ct scan taken in?	What is the primary abnormality in this image?	What organ system is displayed in this mri?
Expected Answer	GI and IV	XR- Plain film	Coronal	Plantar fibromatosis	Heart and great vessels
Predicted Answer (BDI)	Upper GI	XR- Plain film	Coronal	Primary intraocular lymphoma	Lung, mediastinum, pleura
Predicted Answer (ADI)	GI and IV	XR- Plain film	Coronal	Bone Cyst	Lung, mediastinum, pleura
Generated XAI Output					
	(a)	(b)	(c)	(d)	(e)

Fig 4. Visualization of samples from VQA-MED 2019 dataset w.r.t predicted answer

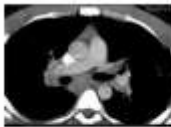




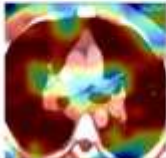


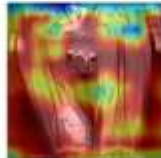
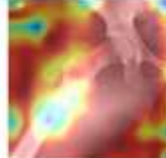
Correctly Classified Samples				Wrongly Classified Samples	
Image					
Question	What is abnormal in the ct scan?	What abnormality is seen in the image?	What is abnormal in the ct scan?	What is abnormal in the mri?	What abnormality is seen in the image?
Expected Answer	Sarcoid	Abdominal abscess	Pulmonary embolism	Aneurysmal bone cyst	Stress fracture
Predicted Answer (BDI)	Empyema	Abdominal abscess	Pulmonary embolism	Osteoporosis	Enchodroma
Predicted Answer (ADI)	Sarcoid	Abdominal abscess	Pulmonary embolism	Osteoporosis	Lisfranc fracture-dislocation
Generated XAI Output					
	(a)	(b)	(c)	(d)	(e)

Fig 5. Visualization of samples from VQA-MED 2020 dataset w.r.t predicted answer






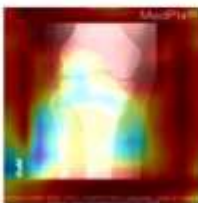
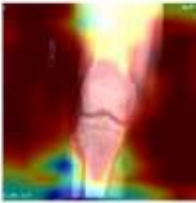
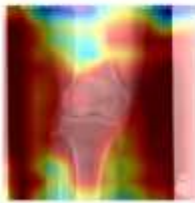
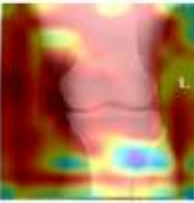
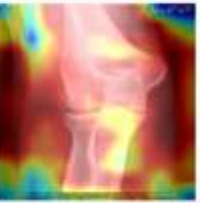
Correctly Classified Samples				Wrongly Classified Samples	
Image					
Question	What is most alarming about this x-ray?	What is the primary abnormality in this image?	What is most alarming about this x-ray?	What is abnormal in the x-ray?	What is the primary abnormality in this image?
Expected Answer	Osteochondroma	Osteosarcoma	Radial head fracture	Chondrocalcinosis	Osteochondroma
Predicted Answer (BDI)	Osteochondroma	Osteochondroma	Osteosarcoma	Osteosarcoma	Page's disease
Predicted Answer (ADI)	Osteochondroma	Osteosarcoma	Radial head fracture	Osteosarcoma	Fat necrosis
Generated XAI Output					
	(a)	(b)	(c)	(d)	(e)

Fig 6. Visualization of samples from VQA-MED 2021 dataset w.r.t predicted answer


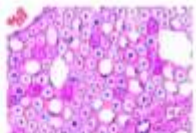
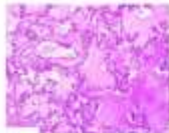

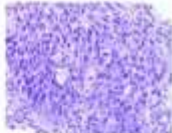

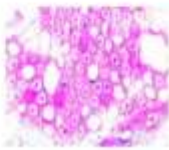



Correctly Classified Samples				Wrongly Classified Samples	
Image					
Question	Are the nuclei of affected tubules pale ?	What show multiple small vacuoles in the cytoplasm microvesicles ?	What are congested ?	What is there increased in the red pulp, capsule and the trabeculae ?	What are the atypical dysplastic squamous cells confined to ?
Expected Answer	Yes	Hepatocytes	Alveolar capillaries	Fibrosis trabeculae	Layers of mucosa
Predicted Answer (BDI)	No	Hepatocytes	Alveolar capillaries	Sinuses	Cytoplasm
Predicted Answer (ADI)	Yes	Hepatocytes	Alveolar capillaries	Acute inflammatory	Cytoplasm
Generated XAI Output					
	(a)	(b)	(c)	(d)	(e)

Fig 7. Visualization of samples from Path-VQA dataset w.r.t predicted answer