# Biomedical Concept Detection in Medical Images: MQ-CSIRO at 2019 ImageCLEFmed Caption Task

Sonit Singh[1,3], Sarvnaz Karimi[3], Kevin Ho-Shon[2], and Len Hamey[1]

[1] Department of Computing, Macquarie University, Sydney, Australia
[2] Macquarie University Health Sciences Centre, Sydney, Australia
[3] DATA61, CSIRO, Sydney, Australia
{sonit.singh}@hdr.mq.edu.au

**Abstract.** We describe our concept detection system submitted for the ImageCLEFmed Caption task, part of the ImageCLEF 2019 challenge. The advancements in imaging technologies has improved the ability of clinicians to detect, diagnose, and treat diseases. Radiologists routinely interpret medical images and summarise their findings in the form of radiology reports. The mapping of visual information present in medical images to the condensed textual description is a tedious, time-consuming, expensive, and error-prone task. The development of methods that can automatically detect the presence and location of medical concepts in medical images can improve the efficiency of radiologists, reduce the burden of manual interpretation, and also help in reducing diagnostic errors. We propose a Convolutional Neural Network based multi-label image classifier to predict relevant concepts present in medical images. The proposed method achieved an F1-score of 0.1435 on the held-out test set of the 2019 ImageCLEFmed Caption Task. We present our proposed system with data analysis, experimental results, comparison, and discussion.

**Keywords:** Medical Imaging · Concept Detection · Caption Prediction · Computer Vision · Convolutional Neural Network · Multi-label classification.

## 1 Introduction

Medical images contain rich semantic information in the form of concepts, attributes, and their interaction. Modelling the rich semantic information and its dependencies is essential for understanding medical images. Due to the rapid increase in big data, continuous evolution of medical imaging technologies, and the rise of electronic health records, medical imaging data is accumulating at a very fast pace. Automated understanding of medical images is highly beneficial for clinicians to provide useful insights and reduce the significant burden of the overall clinical workflow. Motivated by this need of automated image understanding methods in the healthcare domain, ImageCLEF[4] [16] organised its first concept detection and caption prediction tasks in 2017. The main objective of the *concept detection* task is to automatically find relevant clinical concepts present in medical images. Concept detection is also important for improving various downstream tasks such as knowledge discovery, medical report generation, question answering, and clinical decision making. Figure 1 shows sample images and their corresponding relevant clinical concepts present in the training set provided by the challenge organisers.

ImageCLEF is an evaluation campaign organised as a part of the Conference and Labs of the Evaluation Forum (CLEF) initiative. In 2019, the *ImageCLEFmedical* proposed three tasks namely, *Visual Question Answering* [3], *Caption Analysis* [21], and *tuberculosis* [9]. This paper describes the participation of the MQ-CSIRO (Macquarie University and CSIRO, Sydney) team participation in the $3^{rd}$ edition of ImageCLEFmed Caption task 2019. The task consists of identifying the UMLS (Unified Medical Language System) Concept Unique Identifiers (CUI) [5] present in each sample image. Each medical image can be annotated with multiple concepts, making it a *multi-label* image classification task. Compared to *single-label* classification where an image is associated with a single label from a finite set of disjoint labels, multi-label classification associates a single image with multiple labels which may have semantic dependencies between them. We identified the relevant concepts present in medical images based on a multi-label classification model using Convolutional Neural Network (CNN). In section 2, we describe work in multi-label image classification. Section 3 describes building blocks of a convolutional neural network. In section 4, we describe our data exploration, experimental settings, and analysis of results. Finally, section 5 provides conclusion and future work.

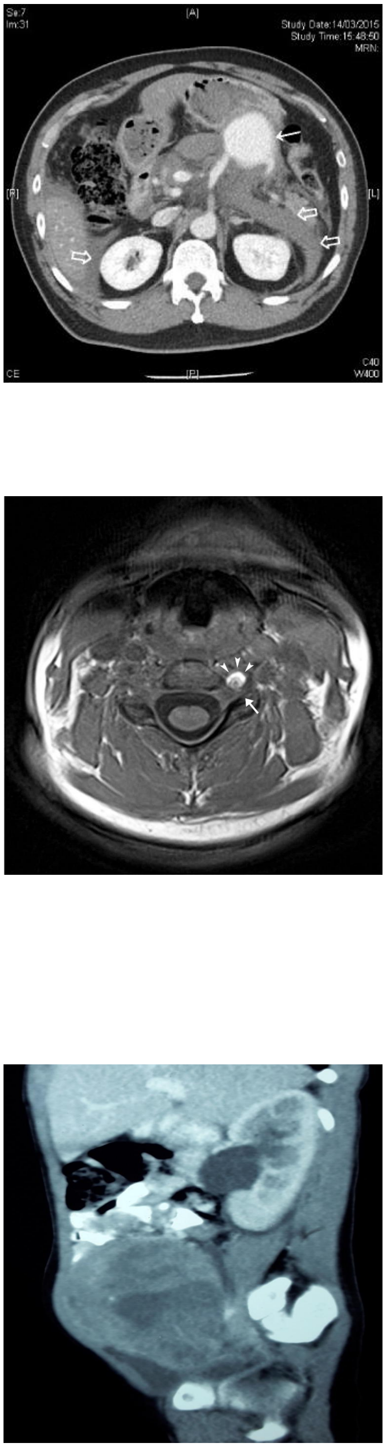
## 2 Related Work

Multi-label image classification is a fundamental task towards general visual understanding. Both medical images and natural images contain diverse semantic content that need multiple visual concepts to classify [19]. Compared to single-label classification, multi-label image classification is more challenging due to the association of concepts with semantic regions and capturing the semantic dependencies among concepts. In the following subsections, we explore work related to multi-label image classification in natural and medical images.

### 2.1 Multi-label image classification

The performance of image classification has recently experienced a rapid progress due to the establishment of large-scale hand-labeled datasets such as ImageNet [24]

[4] https://www.imageclef.org/



**Concepts present**:
C0019066: non-traumatic hemoperitoneum
C0162868: false aneurysm
C0037993: lien
C0607422: abdoman
C0025474: mesenteric membrane
C0009924: materials
C0441633: diagnostic scanning
C0003842: arteri
C0449900: contrasting

**Concepts present**:
C0015252: surgical removal procedure
C0007876: caesarean section (c-section) delivery
C0542560: degrees
C0021815: discus intervertebralis
C0056663: cyanmethaeglobin
C1552858: section
C1318154: root [a body part]
C0546660: methemoglobin (methb) level test
C0965970: et combination
C0728940: excisional
C0251244: alexanian protocol
C0442106: intervertebral
C0052142: ap combination
C0549207: bone tissue of vertebra
C0005847: blood vessel structure
C0184905: bisection
C0003842: arteri

**Concepts present**:
C0086972: separated status
C0022646: nephros
C0227665: kidneys bilateral
C0030797: region

Fig. 1: Sample medical images and their corresponding relevant concepts [8].

and MS-COCO [18], and the fast development of deep Convolutional Neural Networks [25,14]. Due to their great success on binary and multi-class image classification, research has been towards extending deep convolutional networks for multi-label image classification. Multi-label image classification is a fundamental and practical task in Computer Vision where the aim is to identify the set of objects present in an image.

A simple approach for multi-label image classification is to train independent binary classifiers for each label or class. However, this method does not consider the relationship among labels, and the number of predicted label combinations will grow exponentially as the number of categories increase. For instance, if a dataset contains 20 labels, then the number of predicted label combination could be more than 1 million (*i.e.*, $2^{20}$). Besides, this baseline method ignores the topology structure among labels, which can be an important regulariser for the co-occurrence patterns of objects. For example, the combination of *sand*, *trees*, *sky*, *boats*, and *clouds* is plausible to appear in the physical world, but some combinations of labels are almost impossible such as *glacier*, *rain forest*, and *sun*. There is a possibility that artificial or partly artificial images can violate such natural dependencies.

In order to regularise the prediction space, many researchers have attempted to capture label dependencies. Gong *et al.* [12] proposed three multi-label ranking losses to adapt convolutional neural networks for the multi-label problem. These losses were namely, *softmax*, *pairwise ranking*, and *weighted approximate ranking* (WARP). They found that the WARP loss function performs significantly better than the other two loss functions. Wang *et al.* [28] proposed a joint framework combining a convolutional neural network and a recurrent neural network in order to learn the semantic dependencies among labels. Zhu *et al.* [33] proposed a unified framework that captures both semantic and spatial relations of labels using a Spatial Regularisation Network (SRN). The network learns an attention map for each label, which associates relevant image regions to each label. By learning convolutions on the attention maps of all labels, the SRN captures the underlying semantic and spatial relations between labels and acts as a spatial regularisation for multi-label output. In order to use object detection methods to provide region proposals, Wei *et al.* [30] proposed the Hypothesis-CNN-Pooling (HCP) network, it first finds region proposals using object detection techniques such as Edge Boxes [34] to produce a set of candidates. These selected hypothesis are fed to a shared CNN to compute confidence vectors. The confidence vectors are combined through a fusion layer with max-pooling to generate the final multi-label predictions. Wang *et al.* [29] proposed a recurrent memorised-attention module that combines a spatial transformer layer and an LSTM to capture global contextual dependencies among labels and to avoid the additional computational cost of predicting region proposals.

Recently, Durand *et al.* [11] proposed a partial binary cross-entropy (partial-BCE) loss function and used curriculum learning to train a multi-label image classification model with partial labels, which reduces the cost of annotating all labels in each image. To improve the performance by capturing and exploring label dependencies, Chen *et al* [6] proposed a Graph Convolutional Network which learned to map the label graph into a set of inter-dependent object classifiers.

### 2.2 Concept Detection in Medical Images

The goal of concept detection is to find relevant clinical concepts in medical images. Automatic identification of relevant medical concepts in medical images is vital for indexing and retrieval, report generation, and clinical decision support systems [26]. Concept detection can be solved as a classification problem where a mapping function is learned between low-level visual features and high level semantic concepts based on the annotated training data.

Dimitris and Ergina [10] proposed the use of the ResNet50 [14] model for predicting biomedical concepts for the ImageCLEF 2017 caption prediction task. Abacha *et al.* [1] used CNN and Binary Relevance [31] Decision Tree for concept detection. Since the distribution of concepts is uneven with large number of concepts present in only a few images, they build two different training subsets targeting the most frequent concepts having frequency greater than 400 and 1500, respectively. The Binary Relevance approach has limitations in terms of computational cost since a different classifier is trained for each concept present in the dataset. Hasan *et al.* [13] proposed an attention based encoder-decoder framework for concept detection for ImageCLEF 2017 caption prediction. The encoder is a VGG-19 [25] model and the decoder is a Long-Short Term Memory (LSTM) [15] network with a soft attention mechanism. The dependencies have been captured by hidden states of the LSTM. This approach treated concept detection as a sequence generation task which lacks in identifying the dependency of different concepts. Because concepts are not inherently ordered into a sequence, capturing dependencies by the hidden states presents a problem.

Pinho and Costa [23] proposed an adversarial network for feature learning and training a multi-label classifier using the extracted features to predict medical concepts. They showed that the use of deep learning methods outperformed more traditional representations. Valanavis and Kalamboukis [27] proposed a k-Nearest Neighbour (kNN) based approach for concept detection. Images are represented using two models namely, Bag of Visual Words (BoVW) and generalised Bag of Colours (QBoC). Using the extracted image visual representation, for each test image, training images are sorted based on their similarity score and the concepts of the top matched image are assigned to the test image. In an another approach Zhang *et al.* [32] proposed *retrieval* and *topic-modelling* based methods for concept detection in the ImageCLEF 2018 challenge. They used Lucene Image Retrieval (LIRE) [20] for retrieving the most similar images and their corresponding clinical concepts from the training set to assign concepts to the test images. Also, Latent Dirichlet Allocation (LDA) [4] was used to analyse the topic distribution of clinical concepts present in the retrieved similar images from the training set. Although, the above approaches were simple, they suffer from computational complexity and lack novelty in identifying concepts in unseen images. Singh *et al.* [26] also did similar study in classifying the modality of images and finding relevant medical concepts on a publicly available dataset, and found that convolutional neural networks are better for feature extraction when compared to the traditional approaches. Motivated by the success of Convolutional Neural Networks (CNNs) for various computer vision task, we use a CNN model for finding relevant medical concepts present in an image.

## 3 Convolutional Neural Network

With the rapid collection of large-scale datasets and rapid development of high performance computing devices, Convolutional Neural Networks (CNNs) are increasingly drawing attention from both research and industry [25,14,28]. The common building blocks of Convolutional Neural Networks are *Convolutional layer*, *activation layer*, *pooling layer*, *flattening layer*, and *fully-connected layer*.

### Convolutional Layer

This is the main building block of Convolutional Neural Networks. The main role of the convolutional layer is to detect features by applying an affine filter (or kernel) over the image pixels. The early convolutional layers in a CNN extract low-level features whereas the later convolutional layers are responsible for extracting higher level semantic features.

### Activation layer

The goal of an activation layer is to pass the output of the convolutional layer through an activation function. This layer is also called a non-linearity layer because we pass the output through some non-linear function such as *sigmoid*, *tanh*, or $ReLU$ to get feature maps. The activation layer does not change the dimensions of the feature maps.

### Pooling layer

The main functionality of a pooling layer is to reduce the spatial dimensions of the feature maps and provide some spatial invariance to distortions and translations. Apart from this, pooling layers are also responsible for reducing the number of parameters and computation in the network. Various pooling operations include: *max pooling*, *average pooling*, or *L2-norm pooling*. Pooling helps reduce overfitting, which would occur if the CNN is given too much information, especially if that information is not relevant to classify an image.

### Flattening layer

The goal of a flattening layer is to transform the entire pooled feature map matrix into a single column which is then fed to the neural network for processing.

Table 1: Statistics of ImageCLEFmed Caption Task.

| **Data Set** | **No. of images** |
|---:|---|
| Training set | 56629 |
| Validation set | 14157 |
| Test set | 10000 |
| Total | 80786 |

**Fully-connected layer**

After flattening, output of the network is fed through fully connected layers similar to an ordinary neural network. With the fully connected layers, we combine the extracted features together to create a model which performs high-level reasoning. After the final layer, we apply an activation function such as *softmax* or *sigmoid* to produce the classifier output.

## 4 Experimental Setup

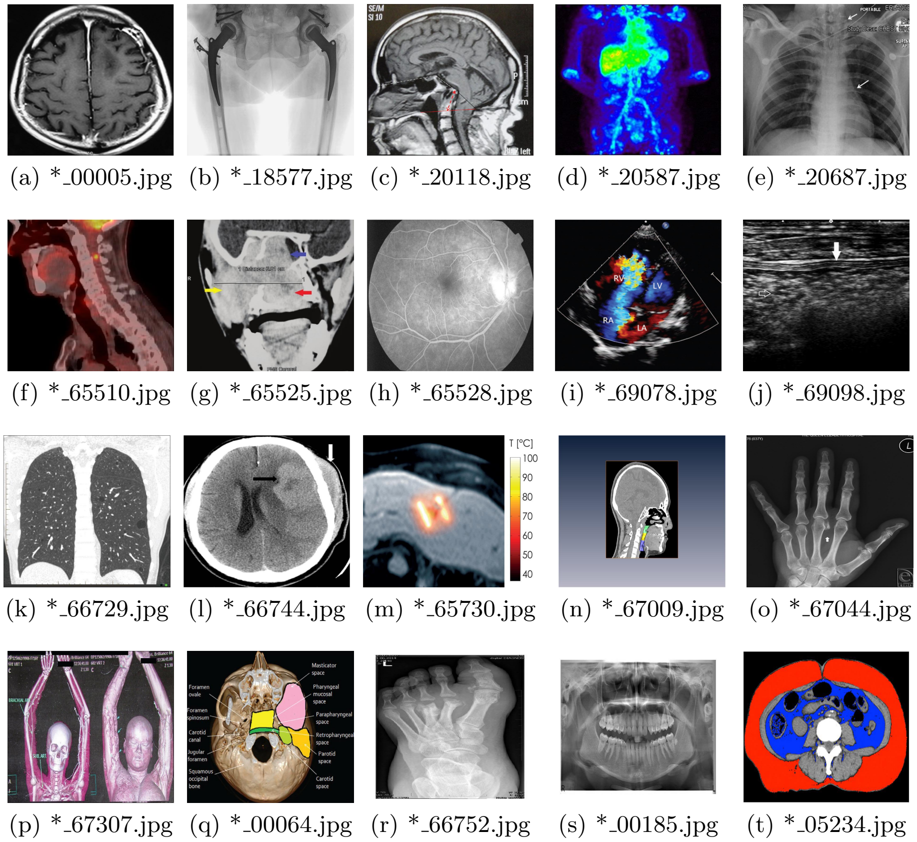
### 4.1 Notation

Concept detection in medical images can be formulated as a multi-label image classification problem where each class corresponds to a concept label. The multi-label classification aims at associating a given instance $x_i \in \mathcal{X}$ with a set of labels $Y_i = y_{i1}, y_{i2}, \ldots, y_{iN}$. For medical concept detection, $x_i$ is a given medical image, $Y_i$ refers to a set of clinical concepts relevant to the medical image, and $N$ refers to number of concepts relevant to that particular image.

### 4.2 Dataset

The dataset provided in the ImageCLEFmed Caption task is collected from the PubMed [5] Open Access subset containing $1,828,575$ archives, having a total of $6,031,814$ image-caption pairs. Automatic filtering using deep learning and manual revisions have been applied to focus on radiology images and non-compound figures, giving a reduced dataset of $70,786$ radiology images of various medical imaging modalities. The official split of data in the form of training, validation, and test is provided by the challenge organisers. Table 1 shows the statistics of the datasets. The ground-truth concepts are provided for the training and validation set, whereas only images are provided for the test set in order to provide a fair evaluation.

[5] https://www.ncbi.nlm.nih.gov/pubmed



(a) *_00005.jpg (b) *_18577.jpg (c) *_20118.jpg (d) *_20587.jpg (e) *_20687.jpg

(f) *_65510.jpg (g) *_65525.jpg (h) *_65528.jpg (i) *_69078.jpg (j) *_69098.jpg

(k) *_66729.jpg (l) *_66744.jpg (m) *_65730.jpg (n) *_67009.jpg (o) *_67044.jpg

(p) *_67307.jpg (q) *_00064.jpg (r) *_66752.jpg (s) *_00185.jpg (t) *_05234.jpg

Fig. 2: Diversity in terms of different modalities and anatomy present in the ImageCLEFmed Caption dataset. * in the image names denotes ROCO_CLEF.

### 4.3 Data Exploration

The dataset in the ImageCLEFmed caption task has huge diversity. Figure 2 shows sample data highlighting various modalities such as X-ray, MRI, ultrasound, and PET, and different anatomies such as hands, feet, brain, chest, and teeth. Apart from this, the images differ in terms of contrast, pixel dimensions, and resolution.

A data analysis shows that there are in total 5216 unique clinical concepts present in the training set. The validation set has a total of 3233 unique clinical concepts present. We found that there are 312 concepts that are present in the validation set but not present in the training set. So, to train our model on all the concepts, we combine the data of the training and validation sets, having a total of 5528 unique clinical concepts present in the dataset. Figure 3 shows the distribution of concepts present in the entire dataset. There are 2,655

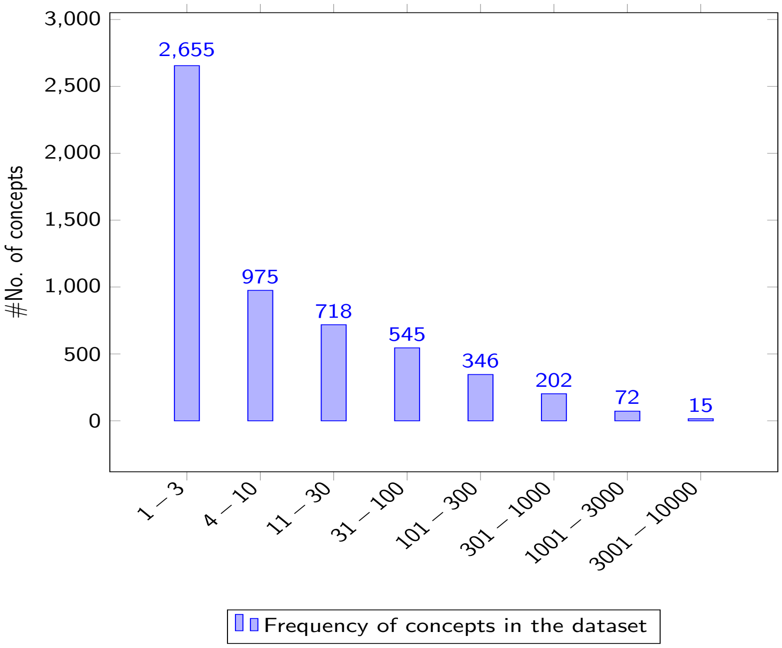


Fig. 3: Number of concepts versus frequency of their occurrence in the dataset.

clinical concepts that occur in less than four images in the dataset. Out of 5528 concepts, 5441 concepts occur less than or equal to 1000 times in the dataset whereas only 87 concepts are present in more than 1000 images. Given that a deep learning model needs at least 1000 samples per class to perform adequately, the distribution of concepts shows the difficulty in training such a model on rare concepts present in the dataset.

Top 20 clinical concepts present in the dataset in terms of their occurrence is show in Table 2. We can clearly see that the top 10 concepts refer to the type of imaging study undertaken. Table 3 shows example of clinical concepts that are found in the dataset but are not visually represented in the images, making it challenging for the model to learn to predict these concepts.

### 4.4 Evaluation Metrics

The challenge organisers provide code for evaluating the performance of the model in terms of *F1* scores, which is the official evaluation metric to rate submissions from different teams. The F1 score is the weighted average of the precision and recall, where an F1 score of 0 indicates the worst score and 1 indicates the best score. As the task is *multi-label classification*, the final F1 score is the average of the F1 scores of each class with *binary* weighting method.

Table 2: Top 20 concepts with their count in the training set.

| | |
|---|---|
| 8425: C0441633 (diagnostic scanning) | 4445: C0003842 (arteri) |
| 7906: C0043299 (X-ray procedure) | 4022: C0024109 (lungs pair) |
| 7902: C1962945 (radiogr) | 3627: C0449900 (contrasting) |
| 7697: C0040395 (tomogr) | 3534: C0009924 (materials) |
| 7564: C0034579 (pantomogr) | 3257: C0041618 (medical sonography) |
| 7470: C0817096 (thoracics) | 2983: C0231881 (resonance) |
| 7164: C0040405 (X-ray CAT) | 2872: C0751437 (adenohypophyseal dis) |
| 6428: C1548003 (radiograph) | 2840: C0000726 (abdominopelvis) |
| 5678: C0221198 (visible lesion) | 2707: C0935598 (sagittal planes set) |
| 5677: C0772294 (alesion) | 2668: C0002978 (x-ray of the blood vessel) |

Table 3: Some of the CUI clinical concepts present in the dataset that are not represented in the medical images, making it difficult for the model to predict directly from the images.

| | | |
|---|---|---|
| C0949214: advertisement | C1561610: signed | C1561611: improved |
| C1552850: start | C1552852: prev | C1552856: copyright |
| C1578434: spouse | C1507394: studyprotocol | C0549649: misuse |
| C3813540: pineapple | C0007306: cartoon | C1550655: patient |
| C1550473: business | C0332148: likely | C3244316: medication |
| C0871472: t-test | C0969625: methodology | C0038435: stressed |
| C4049977: satisfied | C0016538: projected | C0552371: citations |
| C0332219: not at all | C2346845: approval | C1096774: letter |
| C0560453: jump | C1550043: identified | C0034975: registry |

### 4.5 Experimental settings

We build our Convolutional Neural Network for multi-label image classification model in Python using Keras [7] with a Tensorflow backend [2]. Figure 4 shows the architecture of the CNN used in this study. The input to the network is given as a $400 \times 400 \times 3$ representing the Red, Green, and Blue (RGB) values of the input image. The input unsigned byte pixels are *normalised* by dividing them by 255. The first convolutional layer uses a local receptive field (or kernel) of size $5 \times 5$ with a stride of 1 pixel to extract 16 feature maps, followed by a max-pooling operation conducted over $2 \times 2$ regions. The second, third, and fourth convolutional layers produce 32, 64, and 128 feature maps respectively. All convolutional layers use *Rectified Linear Units* (ReLUs) as the activation function. After each convolutional layer, max-pooling with size of $2 \times 2$ and dropout of 0.25 is applied to avoid overfitting of the model. After four blocks of Convolution, max-pooling, and dropout, we flatten the activation map, and apply the fully connected layers. The final fully connected layer consists of 5528 neurons corresponding to the total number of concepts in our dataset. We use the *sigmoid* activation function instead of *softmax* at the output layer of the network to get the probability of each class $c_j$ as *Bernoulli distribution*. The

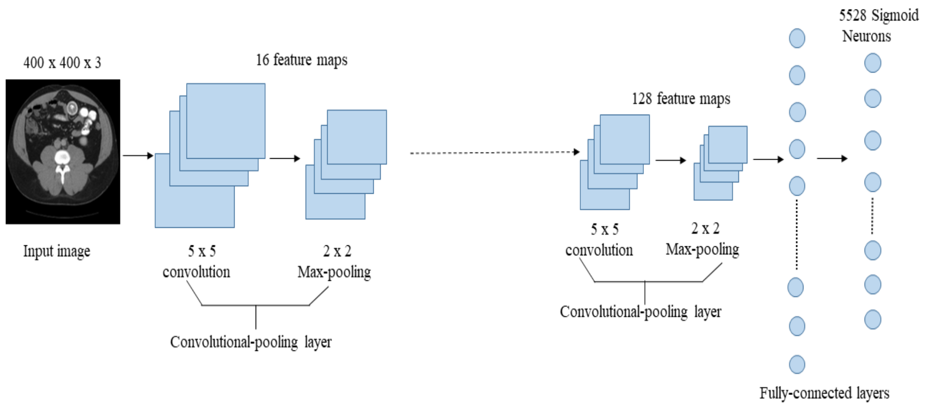


Fig. 4: Schematic of the proposed Convolutional Neural Network for multi-label classification.

motivation is to get the probability of each concept independent of the other concept probabilities so that by using a threshold $\theta$ we can predict whether a particular clinical concept is present in a medical image or not.

The network was trained with the stochastic gradient descent (SGD) algorithm, namely *Adam* [17] with a *binary-crossentropy* loss function. We use *binary-crossentropy* loss instead of *categorical-crossentropy* to penalise each output node independently. Deep neural networks are highly sensitive to *hyper-parameters*, so we tune our model hyper-parameters by selecting a range of value for each parameter and tuning in a coarse to fine search. The *batch size* (BS) is set to 32 and the initial *learning rate* ($\eta$) is set to 0.0001 with a decay of $1 \times e^{-6}$. The model is trained for 10 epochs and the best model based on the accuracy score is saved as the final model. In order to predict concepts on the test data, we set a *threshold* ($\theta$) of 0.1 based on the performance of the model on the validation set.

### 4.6 Results and Discussion

The proposed method convolutional neural network is trained in an end-to-end manner to predict relevant medical concepts on the test set images. Although, three different runs are evaluated internally, only the best run is sub-

Table 4: Performance of our proposed method compared to other teams at 2019 ImageCLEFmed Caption task. The results of the best run by each team is selected for comparison as provided by the organisers on the challenge web page. Source: https://www.imageclef.org/2019/medical/caption/.

| **Team Name** | **Run Name** | **F1 score** |
|---|---|---|
| AUEB NLP Group | s2_results.csv | 0.2823094 |
| damo | ensemble_avg.csv | 0.2655099 |
| GuaJing | 06new_F1Top1.txt | 0.2265250 |
| ImageSem | F1TOP1.txt | 0.2235690 |
| UA.PT_Bioinformatics | simplenet.csv | 0.2058640 |
| richard_ycli | testing_result.txt | 0.1952310 |
| Sam Maksoud | TRIAL_1.txt | 0.1749349 |
| AI600 | ai600_result_weighting_1557061479.txt | 0.1656261 |
| **MacUni-CSIRO** | **run1FinalOutput.txt** | **0.1435435** |
| pri2si17 | submission_1.csv | 0.0496821 |
| AILAB | results_V3.txt | 0.0202243 |
| LIST | denseNet_pred_all_0.55.txt | 0.0013269 |

mitted to the evaluation server for the challenge. Table 4 shows the performance of our proposed approach under the name MacUni-CSIRO with the run name run1FinalOutput.txt having F1 score of 0.1435435. We performed an error analysis on the validation set to figure out the reasons for the low performance of the model. As highlighted in Figure 3 that majority of concepts are rare and are not present in at least 1000 instances (or data points) which makes the task quite challenging. When comparing the results of the multi-label classification model on generic datasets and the ImageCLEFmed caption dataset, we found that the low performance is also attributable to the large number of medical concepts (5528 in the ImageCLEFmed caption task) and the difficulty of obtaining a bounding box annotation for each medical concept present in the medical image. Although the ImageCLEFmed caption 2019 dataset is of a smaller size and is focused on radiology images only (compared to the previous version of the challenge), there is still a huge diversity in images in terms of modality, anatomy, and contrast. Further, during data exploration we found that there are many concepts that do not correspond to any visual data present in the medical images, making the task more difficult. Finally, we feel the need to have a more robust evaluation metric so that partial correct concepts predicted by the model can be considered since current evaluation metric don't take into account of the partial correct concepts predicted by the model.

## 5 Conclusions

This paper presents our experiments for detecting concepts in medical images submitted for the 2019 ImageCLEFmed caption task. The proposed convolutional neural network as a multi-label classifier achieved an F1 score of 0.1435435.

No external resources are used in our experiments. The best model achieved an F1 score of 0.2823094 which is still far from the required performance for these systems to be deployed in a real-world setting. In future, we aim to incorporate domain knowledge so that the performance of these systems can further be improved.

## Animal and Human Research Ethics

The de-identified dataset in the form of medical images and their relevant medical concepts is provided by the challenge organisers [16]. The dataset provided is also a subset of the Radiology Objects in COntext (ROCO) dataset [22]. The detailed description about how the original dataset is given in [21].

## Acknowledgement

This work is supported by an international Macquarie University Research Excellence Scholarship and the DATA61 CSIRO top-up scholarship. This research is undertaken with the assistance of resources and services form the National Computational Infrastructure (NCI), supported by the Australian Government.

## Declaration of Conflicting Interest

The Authors declare that there is no conflict of interest.

## References

1. Abacha, A.B., Herrera, A.G.S.d., Gayen, S., Demner-Fushman, D., Antani, S.: Nlm at imageclef 2017 caption task. In: CLEF2017 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org <http://ceur-ws.org>, Dublin, Ireland (September 11-14 2017)
2. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D.G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., Zheng, X.: TensorFlow: A System for Large-scale Machine Learning. In: Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation. pp. 265–283. OSDI'16, Savannah, GA, USA (2016)
3. Ben Abacha, A., Hasan, S.A., Datla, V.V., Liu, J., Demner-Fushman, D., Müller, H.: VQA-Med: Overview of the medical visual question answering task at imageclef 2019. In: CLEF2019 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org, Lugano, Switzerland (September 09-12 2019)
4. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. Journal of Machine Learning Research **3**, 993–1022 (Mar 2003)
5. Bodenreider, O.: The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic acids research **32**(Database issue), D267–D270 (Jan 2004)

6. Chen, Z., Wei, X., Wang, P., Guo, Y.: Multi-Label Image Recognition with Graph Convolutional Networks. CoRR **abs/1904.03582** (2019)
7. Chollet, F., et al.: Keras. https://keras.io (2015)
8. Demner-Fushman, D., Kohli, M.D., Rosenman, M.B., Shooshan, S.E., Rodriguez, L., Antani, S., Thoma, G.R., McDonald, C.J.: Preparing a collection of radiology examinations for distribution and retrieval. Journal of the American Medical Informatics Association **23**(2), 304–310 (2016)
9. Dicente Cid, Y., Liauchuk, V., Klimuk, D., Tarasau, A., Kovalev, V., Müller, H.: Overview of ImageCLEFtuberculosis 2019 - automatic ct-based report generation and tuberculosis severity assessment. In: CLEF2019 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org <http://ceur-ws.org>, Lugano, Switzerland (September 9-12 2019)
10. Dimitris, K., Ergina, K.: Concept detection on medical images using deep residual learning network. In: CLEF2017 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org <http://ceur-ws.org>, Dublin, Ireland (September 11-14 2017)
11. Durand, T., Mehrasa, N., Mori, G.: Learning a Deep ConvNet for Multi-label Classification with Partial Labels. CoRR **abs/1902.09720** (2019)
12. Gong, Y., Jia, Y., Toshev, A., Leung, T., Ioffe, S.: Deep convolutional ranking for multilabel image annotation. In: International Conference on Learning Representations (2014)
13. Hasan, S.A., Ling, Y., Liu, J., Sreenivasan, R., Anand, S., Arora, T.R., Datla, V., Lee, K., Qadir, A., Swisher, C., Farri, O.: Prna at imageclef 2017 caption prediction and concept detection tasks. In: CLEF2017 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org <http://ceur-ws.org>, Dublin, Ireland (September 11-14 2017)
14. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778 (2016)
15. Hochreiter, S., Schmidhuber, J.: Long Short-Term Memory. Neural Computation **9**(8), 1735–1780 (Nov 1997)
16. Ionescu, B., Müller, H., Péteri, R., Cid, Y.D., Liauchuk, V., Kovalev, V., Klimuk, D., Tarasau, A., Abacha, A.B., Hasan, S.A., Datla, V., Liu, J., Demner-Fushman, D., Dang-Nguyen, D.T., Piras, L., Riegler, M., Tran, M.T., Lux, M., Gurrin, C., Pelka, O., Friedrich, C.M., de Herrera, A.G.S., Garcia, N., Kavallieratou, E., del Blanco, C.R., Rodríguez, C.C., Vasillopoulos, N., Karampidis, K., Chamberlain, J., Clark, A., Campello, A.: ImageCLEF 2019: Multimedia retrieval in medicine, lifelogging, security and nature. In: Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the 10th International Conference of the CLEF Association (CLEF 2019), LNCS Lecture Notes in Computer Science, Springer, Lugano, Switzerland (September 9-12 2019)
17. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: International Conference on Learning Representations. San Diego, California, United States (2015)
18. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft COCO: Common Objects in Context. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) Computer Vision – ECCV 2014. pp. 740–755. Springer International Publishing (2014)
19. Liu, Y., Sheng, L., Shao, J., Yan, J., Xiang, S., Pan, C.: Multi-label image classification via knowledge distillation from weakly-supervised detection. In: Proceedings of the 26th ACM International Conference on Multimedia. pp. 700–708. MM '18 (2018)

20. Lux, M., Marques, O.: Visual Information Retrieval using Java and LIRE. Morgan Claypool (2013)
21. Pelka, O., Friedrich, C.M., García Seco de Herrera, A., Müller, H.: Overview of the ImageCLEFmed 2019 concept prediction task. In: CLEF2019 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org, Lugano, Switzerland (September 09-12 2019)
22. Pelka, O., Koitka, S., Rückert, J., Nensa, F., Friedrich, C.M.: Radiology Objects in COntext (ROCO): A Multimodal Image Dataset. In: Stoyanov, D., Taylor, Z., Balocco, S., Sznitman, R., Martel, A., Maier-Hein, L., Duong, L., Zahnd, G., Demirci, S., Albarqouni, S., Lee, S.L., Moriconi, S., Cheplygina, V., Mateus, D., Trucco, E., Granger, E., Jannin, P. (eds.) Intravascular Imaging and Computer Assisted Stenting and Large-Scale Annotation of Biomedical Data and Expert Label Synthesis. pp. 180–189. Springer International Publishing (2018)
23. Pinho, E., Costa, C.: Feature Learning with Adversarial Networks for Concept Detection in Medical Images: UA.PT Bioinformatics at ImageCLEF 2018. In: CLEF2018 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org <http://ceur-ws.org>, Avignon, France (September 10-14 2018)
24. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L.: ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision **115**(3), 211–252 (Dec 2015)
25. Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition. CoRR
26. Singh, S., Ho-Shon, K., Karimi, S., Hamey, L.: Modality classification and concept detection in medical images using deep transfer learning. In: International Conference on Image and Vision Computing New Zealand. pp. 1–9 (2018)
27. Valavanis, L., Kalamboukis, T.: IPL at ImageCLEF 2018: A kNN-based Concept Detection Approach. In: CLEF2018 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org <http://ceur-ws.org>, Avignon, France (September 10-14 2018)
28. Wang, J., Yang, Y., Mao, J., Huang, Z., Huang, C., Xu, W.: CNN-RNN: A Unified Framework for Multi-label Image Classification. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 2285–2294 (2016)
29. Wang, Z., Chen, T., Li, G., Xu, R., Lin, L.: Multi-label image recognition by recurrently discovering attentional regions. In: 2017 IEEE International Conference on Computer Vision. pp. 464–472 (2017)
30. Wei, Y., Xia, W., Lin, M., Huang, J., Ni, B., Dong, J., Zhao, Y., Yan, S.: Hcp: A flexible cnn framework for multi-label image classification. IEEE Transactions on Pattern Analysis and Machine Intelligence **38**(9), 1901–1907 (Sep 2016)
31. Zhang, M., Zhou, Z.: A review on multi-label learning algorithms. IEEE Transactions on Knowledge and Data Engineering **26**(8), 1819–1837 (2014)
32. Zhang, Y., Wang, X., Guo, Z., Li, J.: ImageSem at ImageCLEF 2018 Caption TaskL Image Retrieval and Transfer Learning. In: CLEF2018 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org <http://ceur-ws.org>, Avignon, France (September 10-14 2018)
33. Zhu, F., Li, H., Ouyang, W., Yu, N., Wang, X.: Learning spatial regularization with image-level supervisions for multi-label image classification. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 2027–2036 (2017)
34. Zitnick, C.L., Dollár, P.: Edge boxes: Locating object proposals from edges. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) Computer Vision – ECCV 2014. pp. 391–405. Springer International Publishing (2014)