Mapana – Journal of Sciences
2024, Vol. 23, Special Issue 3, 103-111
ISSN 0975-3303|https://doi.org/10.12723/mjs.sp3.6

SSCMIRG: Self-Supervised Contrastive Learning for Medical Images with Report Generation

Rahul Kumar*, Arulkumar N†

Abstract

Machine learning has been used to determine what medical images mean in recent years. This is a promising trend for a wide range of medical applications. Image captioning is not meant to replace radiologists but to give accurate results in places without radiologists. Radiologists and patients can utilize this information to aid in diagnosis and therapy, while the potential of self-supervised learning can transform computer vision and medical imaging. Researchers may expect even more impressive results in the future if they keep studying and improving. This study focuses on the importance of self-supervised contrastive learning for medical images, which involves categorizing unlabeled data to generate reports. The technique draws inspiration from human categorization learning. The general classification of medical images, which predicts or groups certain parts of the images, helps doctors make better diagnoses and plan better treatments. In this model, medical report formation creates written descriptions of medical images based on what they show. This idea makes keeping medical records more accurate and consistent. Contrastive learning helps models learn broad dataset features without labels by finding similarities and differences between data points. These methods are helpful in many fields and are now essential to machine learning.

* Department of Computer Science, CHRIST (Deemed to be University). Bangalore - 560029, INDIA.

† Department of Statistics and Data Science, CHRIST (Deemed to be University). Bangalore - 560029, INDIA.
Corresponding Author: arul.kumar@christuniversity.in

Keywords: Self-Supervised Learning, Medical Image Classification, Medical Report Generation, Contrastive Learning

1. Introduction

Self-supervised contrastive learning has become one of the best ways to learn from data that has yet to be labeled [1]. By looking at how similar and different the data points are, this method can find out what the main features of a dataset are without having to label each point. This method has been used to put medical images into groups in medical imaging analysis. It can be used with report generation to turn the content of medical images into textual descriptions.

2. A Literature Review

Self-supervised learning is a powerful artificial intelligence technique that enables machines to learn without human supervision [2, 3]. Google has made much progress in self-supervised learning research. It has reached the top level of performance on several tasks using massive datasets and complicated algorithms. In a recent blog post, Google's AI researchers talked about how self-supervised learning can change computer vision, natural language processing, and other areas of AI [4]. The findings of the related works are given in Table 1.

Table 1: Findings of Related Works

Sl. No.	Title	Findings
1	Self supervised contrastive learning for digital histopathology [5] [Elsevier, 2022]	• Based on the contrastive self-supervised method of digital histopathology. • A large-scale study with 57 histopathology datasets without labels was conducted • Focuses on differences between natural-scene and histopathology images • Combining multiple multi-organ datasets improves task performances
2	Review on self-supervised image recognition using deep neural networks [6] [Elsevier, 2021]	• Pretraining • Self-supervised learning using contrastive learning and clustering methods that outperform supervised learning • Visual feature learning from images using self-supervised approaches.

3	Big Self-Supervised Models Advance Medical Image Classifications [7] [IEEE, 2021]	• Using pre-trained images • Self-supervised followed by supervised fine-tuning • The accuracy has improved using the new method of Multi-instance contrastive learning • Issues and challenges considered are reliability, cost, and scalability.
4	Self-supervised learning from 100 Million Medical Images [8] [Arxiv, Cornell Univeristy, 2022]	• Using 100 million medical images of various modularities. • Use the features to guide model training in hybrid and self-supervised. • It uses Contrastive learning and online feature clustering • AUC increased by 3-7 % for the detection of abnormalities in chest radiography
5	TSRNet: Diagnosis of COVID-19-based self-supervised learning and hybrid ensemble model [Elsevier, 2022]	• Proposed a new pre-training method based on transfer learning with self-supervised learning (TS) • A new convolutional neural network based on attention mechanism and deep residual network (RANet) was proposed to extract features • Hybrid ensemble model

3. Self-supervised Learning in Medical Images

Self-supervised learning could train deep learning models without labeled input. Because it is hard to put notes on medical images, there has yet to be much self-supervised learning in medical imaging [10, 11]. Much medical imaging data has yet to be labeled to train an unsupervised learning model. The standard way to mark up medical images is to have skilled radiologists or clinicians do it manually. This process is time-consuming, expensive, and demands specialized knowledge and abilities. Even with expert annotations, there is usually much difference in how people interpret medical images.

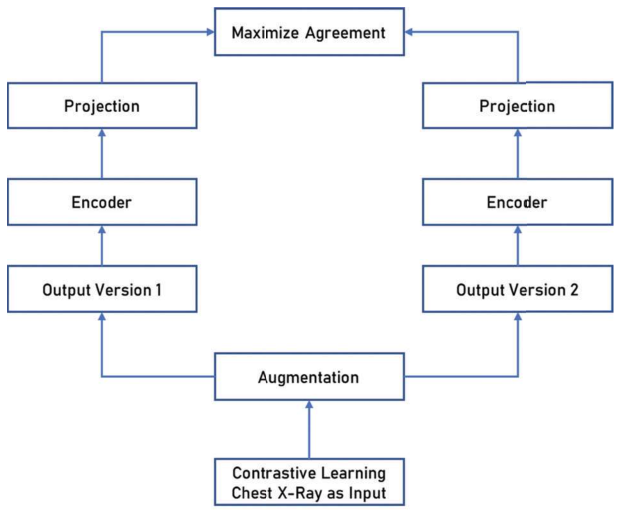

Figure 1: Example of Self-Supervised Learning in the Medical Image

To make self-supervised learning more common in medical imaging, a precise and cost-effective way is needed to label data that has yet to be labeled. This may involve the creation of new, user-friendly annotation tools requiring minimal training. Instead, it could involve using existing data sources, like electronic health records, which have clinical data that can be used to mark up medical photos. The goal is to make a system that can make accurate annotations quickly, on a large scale, and without much manual work. This will help get the most out of self-supervised learning in medical imaging so patients can be diagnosed and treated more accurately. An example of self-supervised learning in the medical image is given in Figure 1.

4. Contrastive Learning under Self-Supervised

For deep learning, models that use self-supervised learning techniques can find similarities and differences between objects and learn the high-level properties of the world [12]. A model can learn to distinguish between a dog and a cat, for instance, based on the dog's protruding nose and the cat's flat nose. Recent improvements in research on contrastive learning have made it much easier to make self-supervised learning models that work better. Google's state-of-the-art contrastive learning engine, SimCLR, has delivered exceptional outcomes on various tasks. Facebook has also developed its plan, MoCo, which has also been successful.

Self-supervised learning is comprised of three fundamental stages.

- Data Augmentation
- Encoding into a vector representation
- Loss Minimization

Stage 1: In the first step, "data augmentation," the unprocessed input data are turned into new training instances. This could involve techniques like cropping, scaling, color distortion, grayscale, and others that change the original data.

Stage 2: The second stage is encoding the additional data into vector form. This is performed by extracting high-level image data using a convolutional neural network (CNN). After a series of thick layers are added to the CNN's output, the data is changed into a latent space and vector representation.

Stage 3: In the third and final stage, loss minimization, the model is taught to reduce the distance between sets of data that are similar and to increase the distance between sets of different data. This is often done by measuring the angle between two vectors in vector space using cosine similarity. The model learns to tell the difference between different objects by making the loss function as small as possible and pulling out high-level features for later tasks.

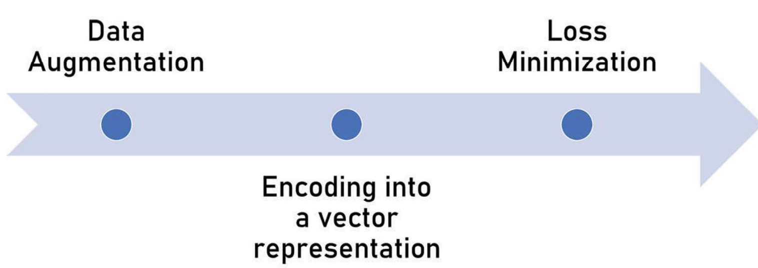

Figure 2: Stages of Self-supervised Learning

4. Report Generation

Creating textual descriptions of an image's content is called report generation or image captioning [13]. Convolutional Neural Networks

(CNNs) and Recurrent Neural Networks (RNNs) are often used together by advanced deep learning models (RNNs) to create image captions [14]. CNN is used to get high-level information from a picture, while RNN lists words that describe the picture. In image captioning, the encoder-decoder architecture is often used. CNN is the encoder, and RNN is the decoder. In addition to LSTM, attention methods are used to make the model focus on certain parts of the picture while it writes the caption. This is a more advanced method.

Learning the image captioning model requires many medical photos with notes on them. This dataset is used to train a computer to create correct picture captions. The model is then fine-tuned using a smaller set of photos with known anomalies to find and explain any new ones. Once the image captioning model has been trained, it could be used in places with few radiologists but many people who need them. The model can quickly make reports that describe any oddities in the photos, making diagnosis and treatment faster and more accurate. Radiologists can also use the reports to evaluate the images used as input. This improves the accuracy and validity of the diagnostic process.

The proposed algorithm for this research work is below:

1. ***Import:*** *Necessary libraries and modules*
2. ***Load:*** *Medical image dataset*
3. ***Define:*** *Data augmentation techniques*
4. ***Implement:*** *CNN to extract high-level features from the images*
5. ***Train:*** *The model is trained using contrastive learning*
6. ***Test:*** *The model is tested on a validation dataset*
7. ***Calculate:*** *Evaluation metrics*
8. ***Implement:*** *Medical image classification*
9. ***Implement:*** *Medical report generation*
10. ***Explore:*** *Finding the possibilities of using Self-supervised contrastive learning for medical images*

The flowchart of the proposed algorithm SSCMIRG is given in figure 3. Here, Deep Learning is used in this proposed algorithm to develop a way to classify medical pictures and make reports. This algorithm has several steps, like importing the necessary libraries and modules, loading the medical image dataset, defining data augmentation techniques, using a CNN to extract high-level features from the images, training the model with contrastive learning, testing the model on a validation dataset, calculating evaluation metrics, and implementing medical image classification and report generation.

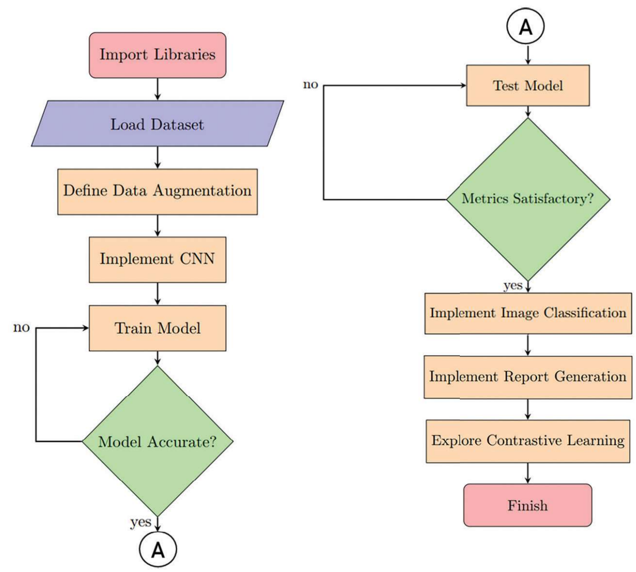

Figure 3: The Proposed SSCMIRG Algorithm

5. Conclusion

Recent research suggests that automatic report generation has the potential to make medical imaging much better and easier to use. Future models and applications that are far more complicated can be predicted if research and development efforts continue. In medical imaging, the proposed SSCMIRG model with report generation could make reports about any information in the images. The proposed technique comprises the stages of medical image analysis using self-supervised contrastive learning. This research will develop a CNN-based model capable of extracting high-level features from medical pictures and testing and evaluating that model using a validation dataset. The model may be used to classify and report on medical photographs. In addition, the study investigates the possibility of self-supervised contrastive learning for medical pictures, which may have consequences for enhancing the accuracy of medical diagnoses and treatment results.

References

1. *Albelwi, S. (2022). Survey on self-supervised learning: auxiliary pretext tasks and contrastive learning methods in imaging. Entropy, 24(4), 551.*
2. *Saeed, A., Ozcelebi, T., & Lukkien, J. (2019). Multi-task self-supervised learning for human activity detection. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 3(2), 1-30.*
3. *Liu, X., Zhang, F., Hou, Z., Mian, L., Wang, Z., Zhang, J., & Tang, J. (2021). Self-supervised learning: Generative or contrastive. IEEE Transactions on Knowledge and Data Engineering, 35(1), 857-876.*
4. *Self-Supervised Learning Advances Medical Image Classification*. (2021, October 13). Google AI Blog. Retrieved March 1, 2023, from https://ai.googleblog.com/2021/10/self-supervised-learning-advances.html
5. Ciga, O., Xu, T., & Martel, A. L. (2022). Self supervised contrastive learning for digital histopathology. *Machine Learning with Applications*, *7*, 100198.
6. Ohri, K., & Kumar, M. (2021). Review on self-supervised image recognition using deep neural networks. *Knowledge-Based Systems*, *224*, 107090.
7. Azizi, S., Mustafa, B., Ryan, F., Beaver, Z., Freyberg, J., Deaton, J., ... & Norouzi, M. (2021). Big self-supervised models advance medical image classification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 3478-3488).
8. Ghesu, F. C., Georgescu, B., Mansoor, A., Yoo, Y., Neumann, D., Patel, P., ... & Comaniciu, D. (2022). Self-supervised learning from 100 million medical images. *arXiv preprint arXiv:2201.01283*.
9. Sun, J., Pi, P., Tang, C., Wang, S. H., & Zhang, Y. D. (2022). TSRNet: Diagnosis of COVID-19 based on self-supervised learning and hybrid ensemble model. *Computers in biology and medicine*, *146*, 105531.

10. Chen, L., Bentley, P., Mori, K., Misawa, K., Fujiwara, M., & Rueckert, D. (2019). Self-supervised learning for medical image analysis using image context restoration. *Medical image analysis*, *58*, 101539.

11. Taleb, A., Lippert, C., Klein, T., & Nabi, M. (2021, June). Multimodal self-supervised learning for medical image analysis. In *Information Processing in Medical Imaging: 27th International Conference, IPMI 2021, Virtual Event, June 28–June 30, 2021, Proceedings* (pp. 661-673). Cham: Springer International Publishing.

12. Jing, L., & Tian, Y. (2020). Self-supervised visual feature learning with deep neural networks: A survey. *IEEE transactions on pattern analysis and machine intelligence*, *43*(11), 4037-4058.

13. Chun, P. J., Yamane, T., & Maemura, Y. (2022). A deep learning-based image captioning method to automatically generate comprehensive explanations of bridge damage. *Computer-Aided Civil and Infrastructure Engineering*, *37*(11), 1387-1401.

14. Monshi, M. M. A., Poon, J., & Chung, V. (2020). Deep learning in generating radiology reports: A survey. *Artificial Intelligence in Medicine*, *106*, 101878.