Comparison of data balancing techniques for vertebral compression fracture detection

Sejin Yi Yoo

Multiple myeloma (MM) is a rare hematological disease that affects the skeletal system. In the worst cases it results in vertebral compression fractures (VCFs). These lesions cause a great amount of pain and increase vertebral instability. To avoid the worsening of the symptoms surgery can be carried out. However, the sooner these VCFs are detected and treated the more effective the results will be.

To help with this detection, machine learning models can be employed. These models can be trained to learn what parts of an image hold valuable information to correctly identify VCFs. However, they need great quantities of training example images (both healthy and VCF) to be able to learn accurately. Example images of healthy spines are numerous but since MM is a rare disease VCF example images are hard to find. The lack of images from one class with respect to the other is called the data imbalance problem. The predictive ability achieved by a machine learning VCF detection model will be greatly affected by this data imbalance.

There are several techniques used to mitigate the effects of data imbalance. For instance, data augmentation, which makes small changes to the existing images to create new ones. Another example would be cost sensitive learning which makes the model pay more attention and learn more from the example images of the VCFs than the healthy ones. A recent addition to these data balancing techniques are diffusion models. They learn from the original set of VCF example images and create completely new fake examples that resemble the real VCFs. These fake images can then be used along with the real ones to train the VCF detection model. However, there is no knowledge about their efficacy in comparison to the other explained techniques.

In this study, a machine learning model was trained to learn to detect VCFs. Then, the mentioned techniques to mitigate the effects of data imbalance were applied. For each technique it was observed whether after their usage the model got better at detecting VCFs and which of them helped the model the most.

As a result, the findings indicated that no technique actually helped the detection model. This was a surprising finding since these techniques have been proven to work for other types of images. However, when looking at the literature data augmentation had only been used for spine images in a few instances and it did not help in the model's learning. This might indicate that data augmentation does not work well for spine images. On the other hand, the diffusion models are also dependent on the quantity of example images they learn from. This means that without enough training example images, diffusion models might have created fake example images that were not accurate enough. These inaccurate fake images might confuse the model instead of helping it understand which images are VCFs. As for cost sensitive learning, there are different ways of deciding how much attention the model should pay to each image class so the one chosen might not have been the most suitable.