# Using A Novel Semi-Supervised Machine Learning Method to Improve Image-based Lung Cancer Diagnostic Algorithms

Wang, Alexander (School: Trinity Preparatory School)

In the field of computer vision, a large amount of annotated data is the key to success in deep learning because these computer models need a wider variety of data to “train” and be able to discern certain discrepancies in all types of possible data. Because of the powerful learning capabilities of deep learning, breakthroughs have been seen in convolutional neural networks (CNNs) in many areas, including medical imaging. CNNs using supervised learning can be used to train a computer model (with images where specific regions have already been meticulously labeled by a doctor) to label these unlabeled images. However, obtaining annotated data can be difficult and expensive since acquiring accurate annotations is time consuming and requires physician expertise, particularly for CT images. In order to help combat these inconveniences in medical imaging, a method called semi-supervised learning can be used which uses a small amount of labeled data and large amount of unlabeled data to improve the accuracy of the model. In this project, this method is explored through CT lung cancer image segmentation. A semi-supervised learning method called "mean teacher" is utilized to automatically learn consistent representations from unlabeled data. The core of this approach is to obtain an updated teacher model by averaging the weights of the student models. Training the model with consistency loss and classification loss allows the student model to learn more meaningful information from the teacher model. The dataset used was divided into labeled and unlabeled parts. This new semi-supervised learning approach by far exceeds the results obtained by training with only 110 labeled data images, proving that meaningful information can be extracted from unlabeled data.

**Awards Won:**