Hi, my name is Su Kara. I’m a junior at Capistrano Valley High School. The title of my project is Assessment of Breast Density Using Unsupervised Variational Autoencoders.

1 in 8 U.S. women develops breast cancer over the course of their lifetime. Women with extremely dense breasts have a sixfold greater risk of developing breast cancer. Breast density is about how much fibro-glandular tissue or FGT there is in a breast.

An MRI shows breast density clearly, but it’s an expensive and time-consuming procedure. Besides, image annotation for supervised algorithms requires the time of radiologists. So, I was wondering whether an unsupervised deep-learning algorithm can predict breast density.

Variational Autoencoder, or VAE, is an unsupervised deep-learning algorithm. It can be trained on mammograms with no annotations of breast or FGT segmentation. VAE doesn’t require the supervision of FGT/breast ratio as a number either. My hypothesis was that if I train VAE on mammograms with no annotations, then it will predict breast density as the ratio of FGT to breast.

I used 10 GeForce RTX Titan servers at the UCI lab to train the VAE models with different hyperparameters in parallel. I used Python, TensorFlow, and Keras libraries for development. I had 6,987 mammograms of 734 patients to train on.

After training, I ran the encoder model prediction on each 512x512 input image. It generated a 16x16 latent feature matrix for a patch shape of 32x32. I cleaned up latent feature matrix by applying a mask generated from a separate U-Net study. Then I collapsed the masked latent matrix into a single value such as mean. Finally, I correlated them with the FGT/breast ratios from another U-Net study on MRIs.

Pearson correlation between the mean of masked latent and the ground-truth FGT/breast ratios was 0.68. Linear regression showed a mean absolute error of 0.05. The breast density can be formulated as y = 0.23x + 0.38, where x is the mean of masked latent and y is the breast density.

The performance could be further improved with higher latent dimensions to learn more features. Similarly, it could be trained and tested with different patch shapes, which will directly affect the latent matrix size and resolution.

The data supports my hypothesis that an unsupervised deep learning algorithm such as VAE can be used to predict breast density. Even though supervised algorithms such as regular CNN and U-Net models still provide higher accuracy, radiologists need to spend a lot of time to annotate MRIs, CT scans, and mammograms manually. Unsupervised techniques will be the next breakthrough in the use of AI in medical diagnosis as there will be no need to annotate images anymore.

Radiologists can start using this solution to measure breast density without any manual annotations or segmentations. This study can also serve as a general framework for researchers who plan to use unsupervised deep learning solutions in other domains including image processing and medical diagnosis.