Comparison of InceptionV3 and ResNet18 Deep Neural Networks for Regression in Medical Imaging

Jacob Barkovitch

Department of Computer Science, Binghamton University

CS301: Ethical, Social, and Global Issues in Computing

Dr. George Weinschenk

April 16, 2021

Abstract

Machine learning in the medical imaging field is a fast-growing billion-dollar industry. The need for faster automation, and more accurate diagnoses, leads researchers to try and find the best machine learning models for medical imaging. A comparison of two such models is useful to evaluate their uses for regression tasks in medical imaging. The two deep neural network models are InceptionV3 and ResNet18. Both these models are pre-trained on ImageNet, making them ideal for medical imaging tasks that have minimal training samples. A comparison of these models takes place through analyzing the models' architectures, similar uses in literature, and results on a synthetic image dataset. The generation of the dataset utilizes Poisson-disc sampling which mimics random distributions in nature and biology, such as in human cell distributions. The dataset comprises twenty thousand images with the number of points in each image as labels. The points mimic cell centroid placements which cell recognition networks from other research papers label in real cell tissue stains. The models use the synthetic dataset for an accurate prediction of their scalabilities to real cell tissue samples. The results are that InceptionV3 performs 320% better than ResNet18 for predicting cell counts in the synthetic dataset. Current research in regression for medical imaging also utilizes models more related to InceptionV3 than ResNet18. Based on these findings, the InceptionV3 surpasses ResNet18 in medical imaging regression tasks.

Keywords: deep neural networks, medical imaging, regression, cell count prediction

Comparison of InceptionV3 and ResNet18 Deep Neural Networks for Regression in Medical Imaging

Correctly identifying cancer in a medical image means the difference between life and death for a patient. Neil Savage (2012) states in a *Communications of the ACM* article, that roughly 12%-30% of lung cancers in medical scans go unnoticed by radiologists. He also says how recent deep neural network models trained for detecting lung cancer in medical images only miss cancer 3% of the time. This shows how medical practitioners can use machine learning models to save up to ten times as many lives.

Starting in the 1980s, computer scientists have been trying to develop machine learning models for recognizing different diseases by using feature-based models (Savage, 2012). Savage explains these models use texture, and morphological features to identify distinct types of tissue in an image, but that this comes at the cost of obstructions like bones getting in the way. Savage's article says that computer scientists were able to overcome the problem through more advanced computing power. The increased computing power, as described by Savage, allowed models to look at intensities of individual pixels to recognize certain tissues which bypass obstructions.

These advances in computing power have also spurred on the development of extremely accurate classification networks for ImageNet (a dataset of over a thousand classes and a million images). Two notable models that have trained on this dataset and achieved top-of-the-line results are InceptionV3 and ResNet18. Computer scientists can freeze the weights and biases of these networks to transfer their representations to other computer vision tasks, such as regression in medical imaging. This paper proposes that the InceptionV3 algorithm surpasses the ResNet18 algorithm in transfer learning regression tasks for medical imaging; accurate prediction of cell count, patient age, and organ boundaries in medical images is a step towards faster treatments and diagnosis of patients. Examining the alternative technologies to these models in the medical field shows the strengths of using a pre-trained deep neural network versus more traditional methods.

Precedents & Related Technology

Multiple instance learning (MIL) is one alternative method of deep learning that researchers have used in identifying congenital abnormalities of the kidney and urinary tract (CAKUT) as shown in Yin's paper (2020). The paper explains that roughly 50-60% of chronic kidney disease in children are a result of CAKUT; accurate identification of abnormalities in ultrasounds would allow doctors to treat CAKUT sooner. Yin and his team combine MIL with a transfer learning approach that takes advantage of a pre-trained VGG16 network. Their method can compute a subject's classification label (whether they are a patient with CAKUT or not) based on the pre-trained model's ability to learn discriminative features from an image in conjunction with MIL outputs. In Yin's paper the formulation of the MIL model is a classifier of subjects *X* that contain 2D kidney images denoted, $x_i (i = 1, \dots, I)$, where each image x_i is an 'instance' and all $x_i \in X$ are a 'bag'. The scoring function for the classification of a bag *X* is:

$$P(X) = g\left(\sum\nolimits_{x_i \in X} f(x_i)\right),$$

Where f and g are transformation functions. Yin's team implement f as a deep learning model that for each x_i, obtains a probability of CAKUT denoted, $p_i = f(x_i)$. They implement g as an outlier invariant mean pooling operation. The combination of the MIL and transfer learning models allowed their method to achieve accuracies of 92% on their ultrasound dataset containing over six thousand images. The accuracy of the model can lead to faster treatment of CAKUT in children.

Researchers have not only used machine learning for aiding in the diagnosis of diseases, but also in the preprocessing of medical images. One such research paper (Chen et al., 2017), proposes a method that uses a convolutional neural network (CNN) to denoise low-dose CAT scans (CT). The paper explains that the purpose of using low-dose CT is to reduce the amount of radiation exposure that patients go through which would lower the risk for contracting side effects like cancer. The problem though, as described by Chen's team, is that low-dose CT comes at the cost of lower image quality and noise-filled results. They propose to use a CNN to denoise these images to achieve high-dose CT image quality.

The group's model for noise reduction starts with $X \in R^{m \times n}$ being a low-dose CT image, $Y \in R^{m \times n}$ being a normal-dose equivalency of X, and their relationship denoted:

$$X = \sigma(Y)$$

Where, $\sigma : R^{m \times n} \rightarrow R^{m \times n}$ is the process that distorts normal-dose CTs. They use this reduction formula to find a function f:

$$f = \frac{argmin}{f} \|f(X) - Y\|_2^2$$

Where a deep neural network approximates f to find the best approximation of σ^{-1}. The noise reduction on low-dose CTs allows Chen's team to achieve comparable results to normal-dose CTs. Their findings will allow patients to undergo lower doses of radiation and receive the same accuracy of diagnosis. These papers show just a couple of ways that researchers use successful machine learning in the medical field and demonstrates that a strong pre-trained model applied to regression is promising. The inceptionV3 and Resnet18 models portray their differences by running on a dataset resembling those in the medical field.

Support

The paper by Szegedy (2016) introduces the InceptionV3 architecture. The model consists of 94 convolutional layers, 11 average pooling layers, four max-pooling layers, and two fully connected layers. InceptionV3 uses SoftMax to compute its loss, and batch normalization to stabilize exploding gradients. More layers result in a denser network. To make the network lighter Szegedy's team use smaller asymmetrical convolutions. A three-by-one layer followed by a one-by-three layer replaces a three-by-three layer that replaces a five-by-five layer. InceptionV3 uses auxiliary classifiers between layer groups to regularize the network, which results in an accurate and computationally efficient model. The model still has many layers though which comes at the cost of speed; therefore, another model with lower computational cost is a good comparison.

The ResNet18 model comes from the paper by He (2016) and uses 18 convolutional layers, two average pooling layers, one max-pooling, and one fully connected layer. The network uses stochastic gradient descent for optimization and no dropout layers for improved speed. The model's advantage is a residual mapping technique. He's team says that this technique solves the degradation problem where very deep neural networks start to perform worse as they get deeper. They describe a residual layer as consisting of two weighted layers, two ReLu activation functions, and an identity connection mapping. The group formulates many residual blocks to allow the network to become incredibly deep while maintaining high accuracy and low computational cost. This model is more lightweight than InceptionV3, but less accurate since it scored lower on a uniform dataset.

The InceptionV3 and ResNet18 models train and test on ImageNet so their accuracies are comparable. InceptionV3 achieves a top-1 error of 21.2% on ImageNet with ResNet18 at 27.88%. Their accuracies on ImageNet demonstrate the two models' performances on a classification task with millions of images. Medical scan datasets contain a small set of samples since the images are expensive and private.

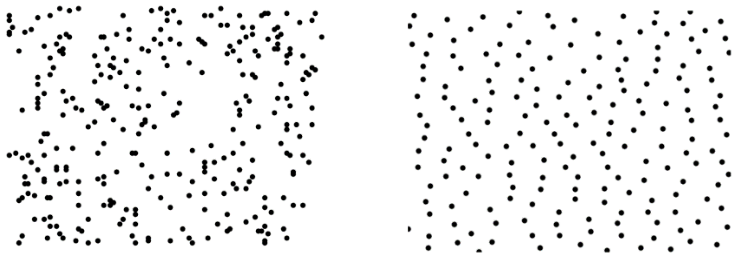
Applying the models on a synthetic dataset demonstrates their accuracies on a regression task that applies to medical imaging.

Dataset

Current research in human tissue cell count prediction uses images of real tissue stains such as in Xie’s research (2018) and He’s paper (2021). The methods, as done in Xie’s (2018) and He’s (2021) papers, consist of passing these images of real tissue stains through a computer vision model that recognizes all the cells in each image and puts a dot on the cell centroids. These dots then pass over to a regression network that counts all the dots and reports the total cell count of the image (He, 2021). To test two models for their applicability for cell count prediction, the models only need images of the centroid dot placements; therefore, a computer-generated synthetic dataset of just the cell centroid placements can be a suitable substitution for the real samples. A random 2D coordinate generator is not sufficient to mimic the placements of cell centroids because it results in large whitespace uncommon to cell tissues. Poisson-disc sampling is more suitable for organic distributions because it localizes each point based on the points around it (see Figure 1).

Figure 1.

Random Versus Poisson-disc Generation

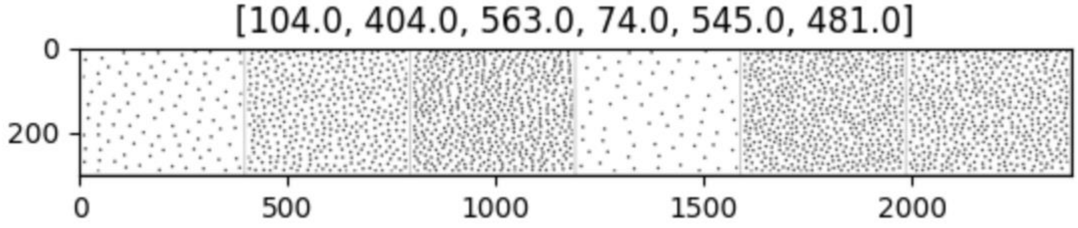


Note. Randomly generated points (left) versus Poisson-disc generated points (right). The randomly generated points result in overlapping and large areas of white space. The Poisson-disc generated points more closely resemble the organic random distribution of cells in certain parts of human tissue (Lanaro, 2020).

Lanaro's team says that Poisson-disc sampling allows for a uniform random distribution of points where no two points overlap or are too far apart. A sparser image of Poisson-disc samples would correlate to a zoomed-in image of cells' centroids (see Figure 2). The following personally made dataset is available upon request.

Figure 2.

Six Random Poisson-disc Dataset Samples



Note. The numbers above each image label the cell counts. The denser images represent a zoomed-out image of cell tissue while the sparser images represent the opposite.

InceptionV3 and ResNet18 test on a generated set of 20,000 Poisson-disc sampler images. The 20,000 images have randomized parameters for minimum distance between points as well as width, and height. The average, median, and mode of labels for the dataset are 361, 325, and 213 respectively. The dataset contains fifteen thousand images for testing and five thousand for validation. The labels are normalized using the Min-Max scalar and un-normalized using an inverse transformation. What follows are the details used for training the two models on this dataset.

Methods

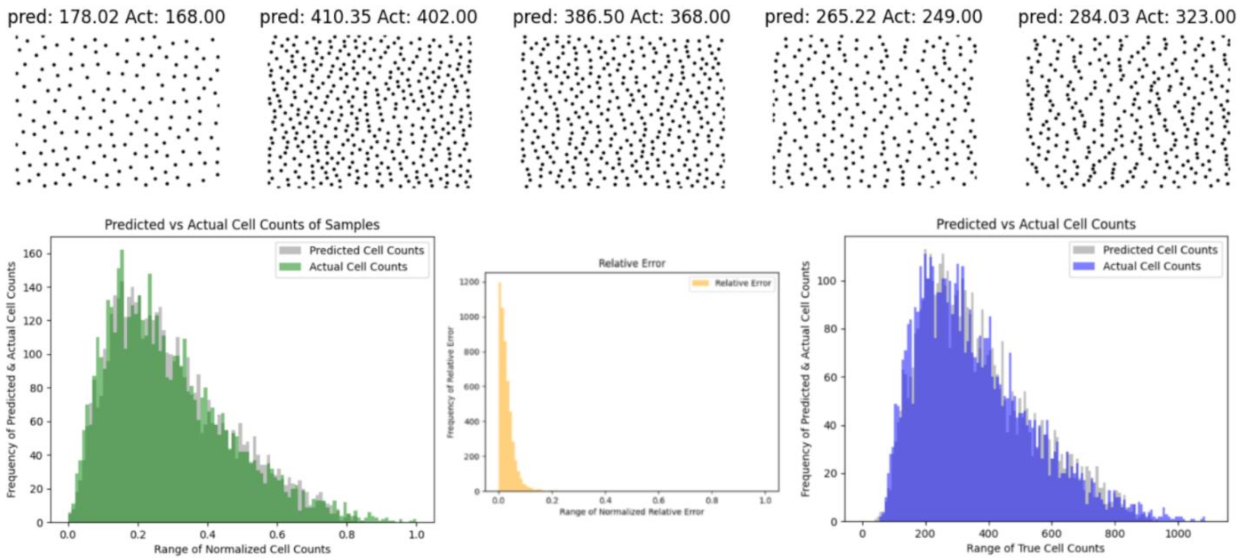
InceptionV3 and ResNet18 are both trained on the Poisson-disc sampled dataset for 20 epochs with 128 batch sizes and a cosine-annealing learning rate scheduler starting at 0.001. The two models use mean squared error to calculate the loss and use the Adam optimizer for sequential improvements. Feature extraction and CUDA are enabled in both models for testing against the normalized labels. The respective architectures of both models remain the same as described previously. The models' weights and biases are frozen for transfer learning to take place. These specifications come from iterative testing and evaluation. It is less important what the specific parameters are than it is for both models to test with the same ones. The only difference in the following results is which model architecture runs on the training and validation dataset.

Results

The findings are that the InceptionV3 algorithm performs 320% better than the ResNet18 algorithm based on their respective total relative errors summed over the length of the validation set. InceptionV3 had all relative errors below 0.3 while ResNet18 had relative errors as high as 0.9. The InceptionV3 model can fit the distribution of labels accurately (see Figure 3) while the ResNet18 model is leaning towards the mean of the data (see Figure 4).

Figure 3.

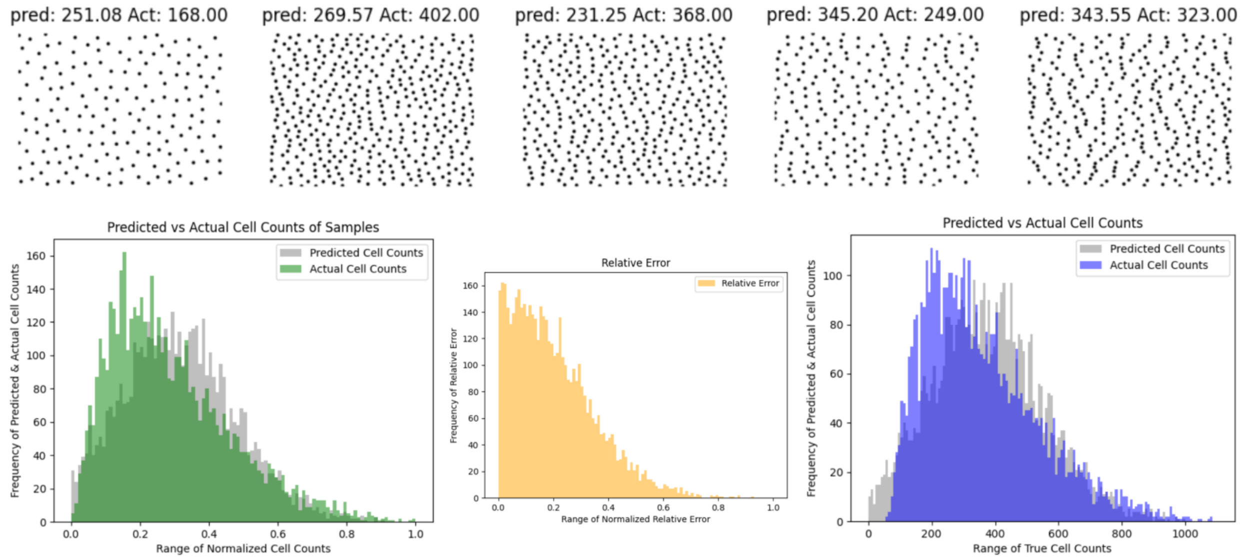
InceptionV3 Results



Note. InceptionV3 prediction samples given inputted images (top), predictions vs normalized labels (left), relative error over the validation set (middle), and predictions vs actual values (right). The model is matching the bell curve of the actual values with predictions only off by double digits in the given samples. The average error over the validation set is 24.56%.

Figure 4.

ResNet18 results



Note. ResNet18 outputs of similar format to Figure 3. The model is only loosely matching the bell curve of the actual values and predictions are off by triple digits in most samples. The average error over the validation set is 78.63%.

Looking at figures three and four the InceptionV3 algorithm outperforms the ResNet18 algorithm for all prediction accuracy metrics. One area where ResNet18 outperforms InceptionV3 is in training times. ResNet18 takes 5.39 minutes to train while InceptionV3 takes 16.02 minutes. Training is performed on an NVIDIA Titan V GPU with CUDA enabled. The training times are inconsequential though when it comes to medical imaging since prediction accuracy overshadows computational cost. Therefore, the InceptionV3 model surpasses ResNet18 when it comes to regression tasks that mimic those used in medical imaging. Analyzing the social impact of these algorithms shows the importance of these findings.

Social Impact

One of the most important things in society involves the ability of society to save lives. Healthcare and medicine companies make it their goal to find ways to save more lives faster. Research into medical imaging improves on both these frontiers. Neil Savage (2012) states that machine learning models can take the rate of missing cancer in patients from 30% to 3%. Implementing machine learning models into medical processes means radiologists no longer need to spend hours looking over a medical scan for signs of cancer. Instead, they can spend time doing more productive tasks like treatment. Thus, healthcare companies have already used machine learning models to save lives, costs, and time.

Pratik Shah (2017), the principal investigator of the health 0.0 research lab at MIT, also discusses the societal benefits of machine learning in healthcare. In his TED talk, he explains how expert physicians currently diagnose diseases by ordering expensive medical scans such as CTs, and MRIs, to do their evaluations. He also explains that other expert physicians must look over the CT and MRI scans before giving their analysis back to the first physician. Pratik describes how this is a time-consuming and expensive process that machine learning has the potential to fix. His team at the MIT Media Lab develop a machine learning model that cuts the required amount of training samples from 10,000 expensive medical images to just fifty. The group's model can accurately diagnose patients based on fewer training samples making it a viable tool in healthcare applications. Pratik's team's method of cutting down on training samples is like transfer learning models such as InceptionV3 and ResNet18 which, require fewer training samples because of their pre-trained weights and biases. Pratik's work shows that transferable models like InceptionV3 are an effective way to save more lives in the medical field.

Another similar paper involves decreasing times and improving accuracy for cell recognition. Xie (2018) and his team propose a model to decrease diagnosis time and proper identification of diseases with machine learning. Their model tackles cell recognition for cancer cell detection and cancer cell count calculation. They explain how manual cell labeling by physicians is time-consuming and prone to errors. Their model aims specifically to improve the detection of cancer cells in microscopy images. Accurate detection and counting of cancer cells can give vital information for a patient's diagnosis and potentially save their life. Their model works by having humans manually label the training set with dots over the cell centroids so that the model can know where to look to recognize cells. Once they train their model to recognize the cells, they can test it on un-labeled data to detect cancer cells, and cell count. They structure their model similar to the InceptionV4 architecture which improves on the InceptionV3 model. Xie's paper shows the applicability of deep neural networks for use in saving lives based on regression tasks.

Shi Yin's (2019) paper is another example of a research article tackling regression problems in medical imaging. Yin and his team specifically propose a method of automatic kidney segmentation in ultrasound images which medical physicians currently do manually. The manual segmentations are once again time-consuming and error-prone, which can interfere with the proper diagnosis. Yin’s team’s methods consist of using a convolutional neural network pre-trained on ImageNet, just like InceptionV3 and ResNet18, to learn high-level features of the ultrasound images. The features then pass into a regression model that learns boundary masks for the kidney images. The specific transfer learning model they use is VGG16 and is closely related to the architecture of InceptionV3. Yin and his team achieved accuracies of 94% for correct kidney recognition and segmentation. Their model improves on all the drawbacks of manual labeling making it a viable option to help expert physicians in their work.

The review of papers and talks show the benefits that machine learning models have in the medical industry. Machine learning models can more save lives so their incorporation into the medical field is vital. If healthcare facilities incorporate these models, they will allow medical practitioners to focus on tasks other than diagnosis. The models would also save hospital costs by being faster, and more accurate. Therefore, since the InceptionV3 algorithm is a machine learning model, and it surpasses the ResNet18 model, the InceptionV3 algorithm can help more save lives than ResNet18 can. If medical businesses want to improve on their machine learning networks, then they should choose InceptionV3 over ResNet18.

Conclusion

Machine learning can improve healthcare services by facilitating more accurate diagnoses, faster treatment, and cheaper costs. An improvement in machine learning algorithms means an improvement in the benefits they provide. Finding improved algorithms is a process of searching literature and documentations for transferable and enhanced models. InceptionV3 and ResNet18 are object classifiers, but they apply to regression tasks because their structures are adjustable. The medical imaging field neglects these models because of their dissimilarity to required tasks. Recognizing the models' transferability through analyzing literature means that better medical imaging models are obtainable. The related technologies in machine learning show the different possibilities that the models apply to. Finding the superior algorithm between InceptionV3 and ResNet18 involves reviewing their uses in literature and comparing their results on a correlated dataset.

Multiple papers using models like InceptionV3 show it is a more recognized and trusted algorithm. To show that their beliefs are well-grounded both models test against a synthetic dataset that mimics cell tissue centroid placements. The findings solidify InceptionV3 as the superior algorithm with an average error 320% times lower than ResNet18's. The findings also show that machine learning in other research improves diagnosis times and accuracies, resulting in the ability to save more lives. Based on these findings and the review of literature, the InceptionV3 algorithm surpasses the ResNet18 algorithm in transfer learning regression tasks for medical imaging. The more accurate prediction of cell count, patient age, and organ boundaries is a step towards faster treatments and diagnosis of patients.

Some implications are that InceptionV3 and ResNet18 are not the most recent versions of their architectures. There are newer variations of the ResNet architecture such as ResNet34, ResNet50, and even up to ResNet-1202. InceptionV4 is the most recent version of the inception architecture, but because it is so recent there is a lack of extensive research on it yet. The abundance of literature on InceptionV3 and ResNet18 means that they are easier to compare. The implication is that the outcome of comparing the older models is unable to indicate the outcome of comparing newer variations. of these architectures would follow the conclusion that the Inception architecture is better. The only takeaway is that InceptionV3 surpasses ResNet18 specifically; moreover, based on the results, InceptionV3 is a promising model for regression in medical imaging which could help save lives.

References

Bridson, R. (2007). Fast Poisson disk sampling in arbitrary dimensions. *ACM SIGGRAPH 2007 Sketches on - SIGGRAPH '07*. https://doi.org/10.1145/1278780.1278807

Chen, H., Zhang, Y., Zhang, W., Liao, P., Li, K., Zhou, J., & Wang, G. (2017). Low-dose CT denoising with convolutional neural network. *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)* 143-146. https://doi.org/10.1109/isbi.2017.7950488

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/cvpr.2016.90

He, S., Minn, K. T., Solnica-Krezel, L., Anastasio, M. A., & Li, H. (2021). Deeply-supervised density regression for automatic cell counting in microscopy images. *Medical Image Analysis*, *68*, 101892. https://doi.org/10.1016/j.media.2020.101892

Lanaro, M. P., Perrier, H., Coeurjolly, D., Ostromoukhov, V., & Rizzi, A. (2020). Blue-noise sampling for human retinal cone spatial distribution modeling. *Journal of Physics Communications*, *4*(3), 035013. https://doi.org/10.1088/2399-6528/ab8064

Parekh, S. (2020, September 9). *AI in medical imaging market to reach $1.5B by 2024*. Imaging Technology News. https://www.itnonline.com/article/ai-medical-imaging-market-reach-15b-2024

Pratik, S. (2017, August). *How AI is making it easier to diagnose disease* [Video]. TED Conferences. www.ted.com/talks/pratik_shah_how_ai_is_making_it_easier_to_diagnose_disease

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., &amp; Wojna, Z. (2016). *Rethinking the inception architecture for computer vision*. 2016 IEEE Las Vegas Conference on Computer Vision and Pattern Recognition (CVPR). doi.org/10.1109/cvpr.2016.308

Savage, N. (2012, January). Better medicine through machine learning. *Communications of the ACM*, *55*(1), 17–19. doi.org/10.1145/2063176.2063182

Xie, Y., Xing, F., Shi, X., Kong, X., Su, H., & Yang, L. (2018). Efficient and robust cell detection: A structured regression approach. *Medical Image Analysis*, *44*, 245–254. doi.org/10.1016/j.media.2017.07.003

Yellott, J. (1983). Spectral Consequences of Photoreceptor Sampling in the Rhesus Retina. *Science*, *221*(4608), 382-385. http://www.jstor.org/stable/1691744

Yin, S., Peng, Q., Li, H., Zhang, Z., You, X., Fischer, K., Furth, S., Tasian, G., Fan, Y (2020). Automatic kidney segmentation in ultrasound images using subsequent boundary distance regression and pixelwise classification networks. *Medical Image Analysis, 60*, 101602. doi.org/10.1016/j.media.2019.101602

Yin, S., Peng, Q., Li, H., Zhang, Z., You, X., Fischer, K., Furth, S., Tasian, G., Fan, Y. (2020). Computer-Aided Diagnosis of Congenital Abnormalities of the Kidney and Urinary Tract in Children Using a Multi-Instance Deep Learning Method Based on Ultrasound Imaging Data. *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)* 1347-1350. https://doi.org/10.1109/isbi45749.2020.9098506