# Stenotic Nares Report

Disclaimer: all accuracy figures below are based on only 100 validation images, so take it with a pinch of salt.

Aim: create a Machine Learning algorithm that can distinguish between dog with and without stenotic nares

## Result:

Achieved overall accuracy of 83%, and in particular predicts images with stenotic nares with 94% accuracy.

## Key challenge:

Lack of data (I handpicked 400 images from google here) which is by far the biggest limiting factor to model’s performance. This has led to a number of compromises below.

1. Due to small dataset, I’ve categorised all degrees of severity of stenotic nares under stenotic nares category, even though they may carry drastically different risk level.
2. I’ve used a separate model to automatically crop the images of dog down to just the nose region. Mass cropping process is slow and model isn’t 100% accurate. However the algorithm does fine for small scale applications such as cropping a few images at a time for prediction.
3. For testing new images, ideally we want a frontal image of dog with face (and nose) clearly visible. The cropping model will do much better if the face is reasonably up right.
4. Due to exceptionally small dataset, more complex models have performed poorly in this case.
5. I’ve created 3 independent models, each with accuracy around 80-81%. Then I combined the predictions from all 3 to reach a final prediction on an image, with overall accuracy of 83%.
6. Introduced some conscious bias in the model so that it can predict images WITH stenotic nares with 94% accuracy. The thinking behind this is that if we think a dog with stenotic nares as health and write the policy, then we might make a loss in future when the dog needs an operation. But if it was the reverse, then we just don’t make a gain or a loss.

## How we can improve this model further:

1. MORE DATA. 400 images might just be inadequate for training an industrial standard deep neural network. It would help alleviate a lot of the problems above, and perhaps even remove the need for a separate network to performing auto cropping.
2. Better transfer learning technique which allows us to take other existing models and improve upon it. The implementation here is rather basic and non-optimal.

## Conclusion:

I believe we can indeed develop powerful deep-learning applications internally, given the right data and some support from the rest of the business. I hope through this project I’ve highlighted the value in collecting / retaining data such as images of the pet insured.