Figure 1. Problem Statement

- **Task**: automatically determine breast density (A, B, C, D) from mammography data

- **Data**: Cancer Imaging Archive mammogram of 2864 images, with ground truth breast density

- **Summary**: 2D image multiple classification

- **Approach**: load DICOM, resize to 224 x 224, and design a modified VGG-16 network

Figure 2. Network Architecture Design & Model Training

*If accepted, all code will be made available on Github.

```python
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import dicom
import os, sys
from keras.applications.vgg16 import VGG16
from keras.models import Model
from keras.layers import GlobalAveragePooling2D, Conv2D, Dense, Flatten
from keras_tqdm import TQDMNotebookCallback
from keras.optimizers import SGD, Adam, Nadam, RMSprop

### Start Here with pre-loaded images
X = np.load('mdata_X.npy')
y = np.load('mdata_y.npy')

### base VGG-16 model
model_base = VGG16(include_top=False, weights='imagenet', input_shape=(224,224,3))

### Add additional custom-designed top-layers
x = model_base.layers[-1].output
x = Flatten()(x)
x = Dense(512, init='orthogonal', activation='relu')(x)
x = Dense(4, init='orthogonal', activation='softmax')(x)

### Train, Validate, and Test the Model
model2 = Model(model_base.inputs, x)
model2.compile(loss='sparse_categorical_crossentropy', metrics=['accuracy'], optimizer=Adam(lr=0.0001))
model2.fit(X, y, batch_size=64, epochs=10, validation_split=0.2,
    verbose=0,callbacks=[TQDMNotebookCallback(leave_inner=True)])
y_pred = model2.predict(X)
```

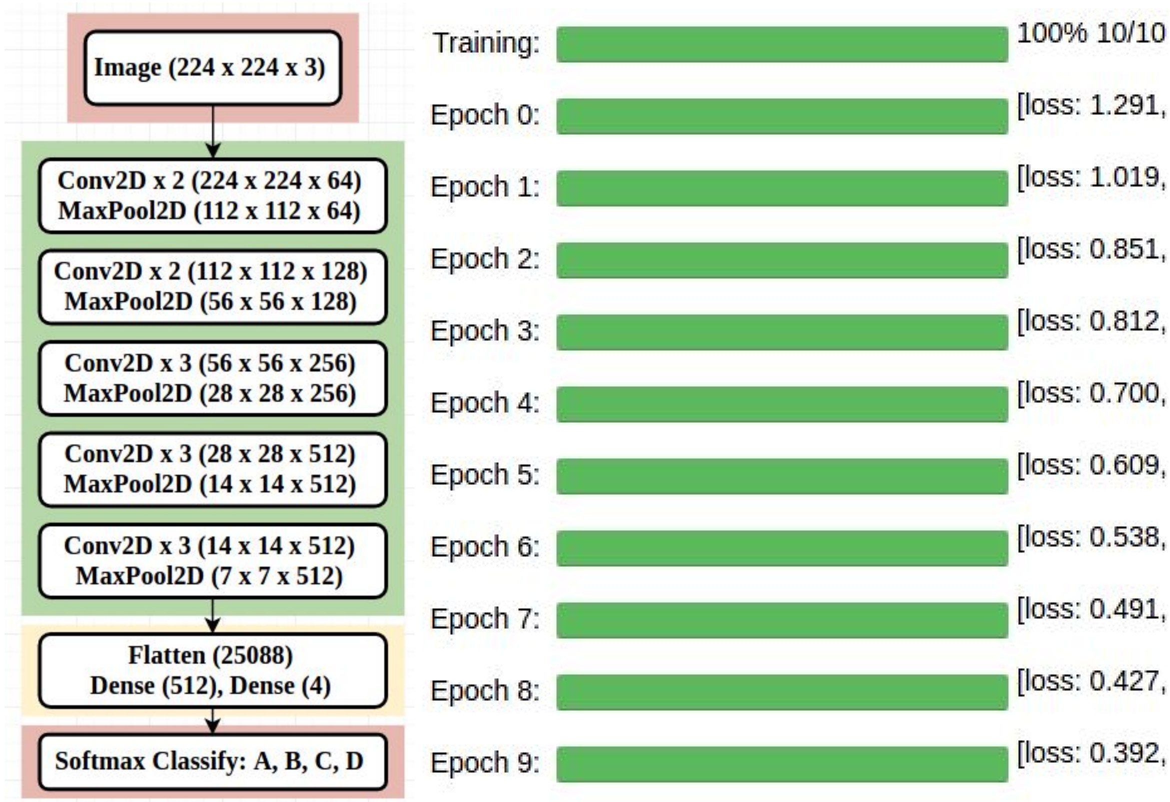
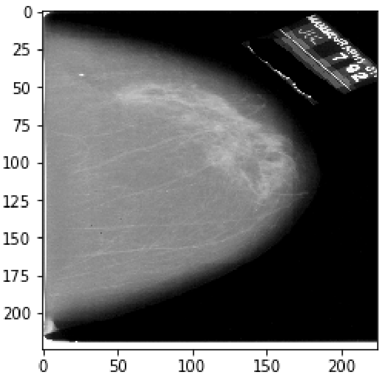


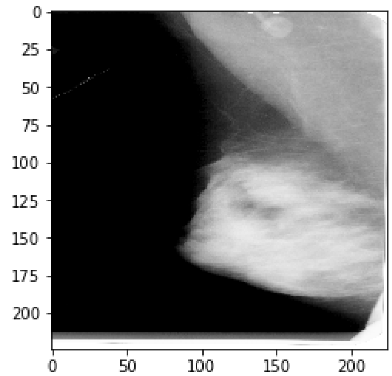
Figure 3. Error Analysis

Analyzing these three categories is a quick way of understanding why an algorithm fails when it fails. Labels: Truth (ground truth density from radiologist), Predict (algorithm-predicted density), and % is algorithm's degree of confidence.

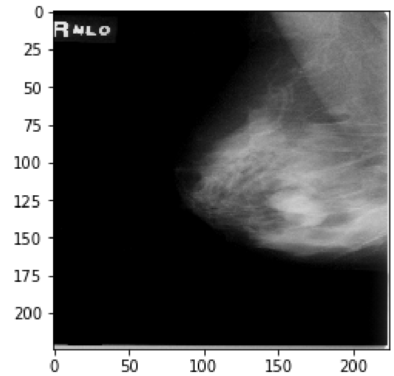
<Confident and Correct>



Truth: B, Predict: B (99.7%)

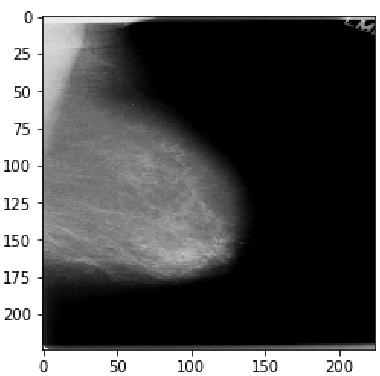


Truth: D, Predict: D (98.3%)

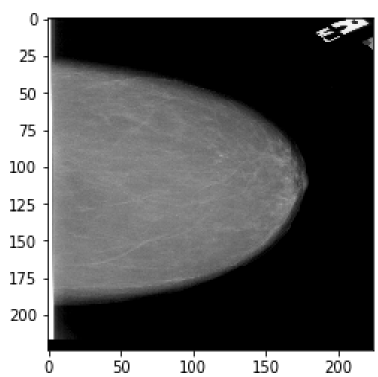


Truth: C, Predict: C (99.2%)

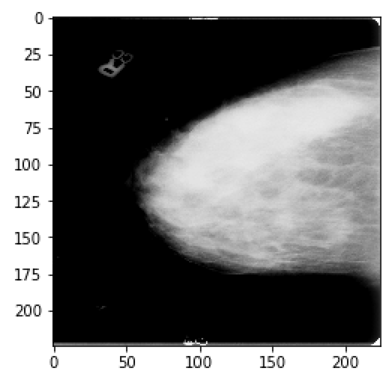
<Confident but Incorrect>



Truth A, Predict B (93.8%)

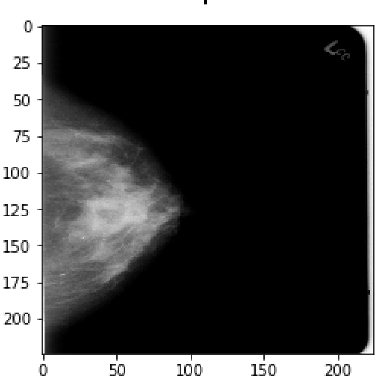


Truth B, Predict A (98.8%)

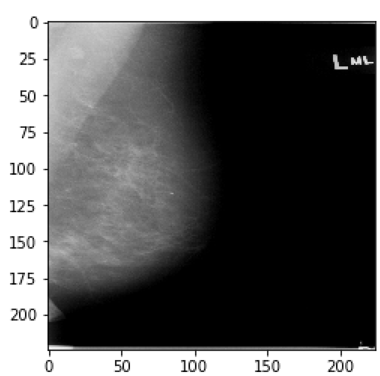


Truth C, Predict D (92.1%)

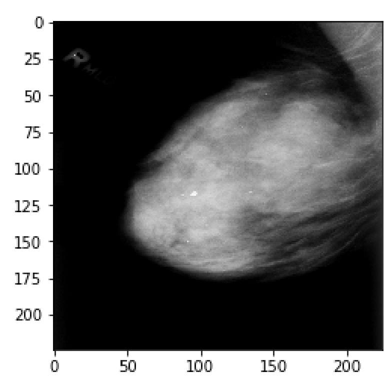
<Random Samples>



Truth C, Predict C (98.1%)



Truth B, Predict B (77.2%)



Truth D, Predict C (53.5%)

Figure 4. Confusion Matrix

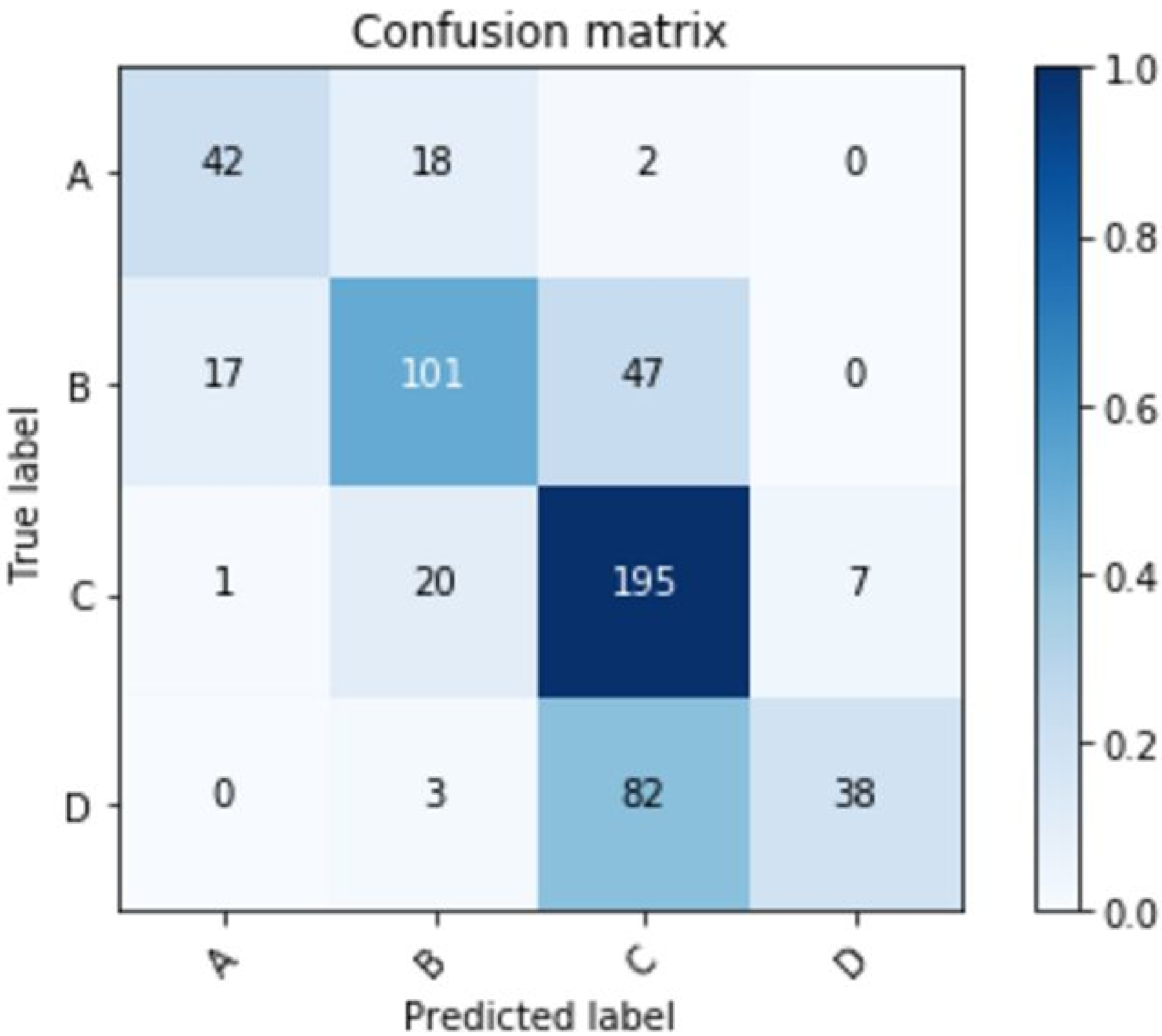
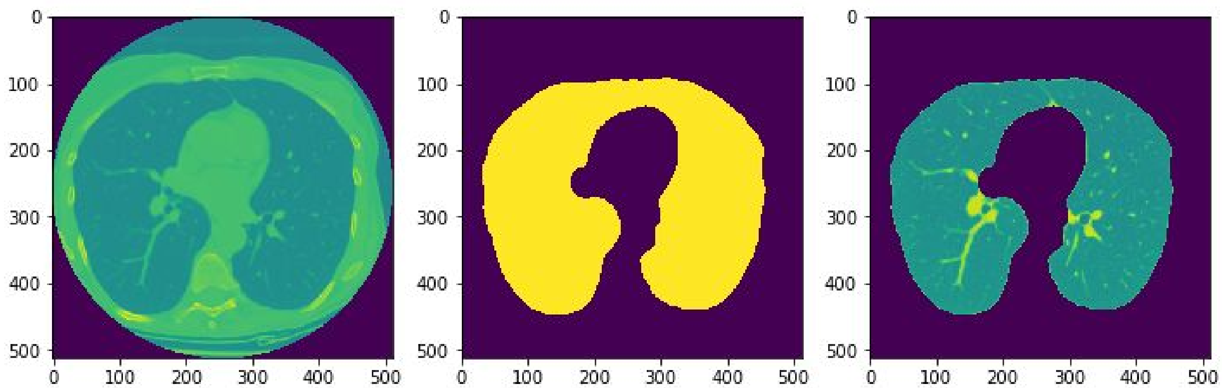


Figure 5. Generalization to 3D binary classification problems

Problem 4. Lung Nodule Malignancy Risk Classification

<Preprocessing>



<Algorithm-Predicted Non-cancer Examples>