LUS Images classification with uncertainty detection and image similarity

Medical Imaging Diagnostic

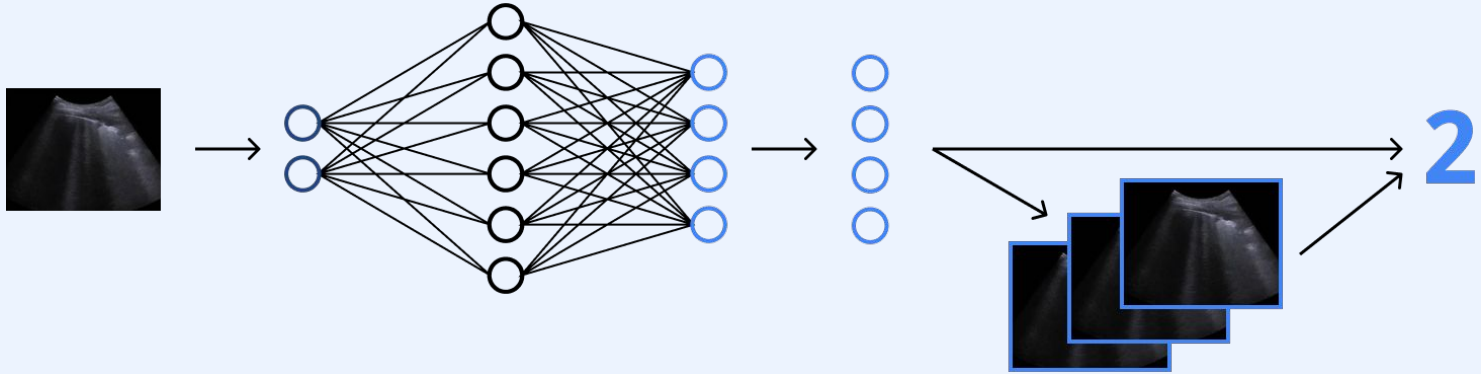
Davide Modolo - Università degli Studi di Trento

0

Quick overview

1 2 3

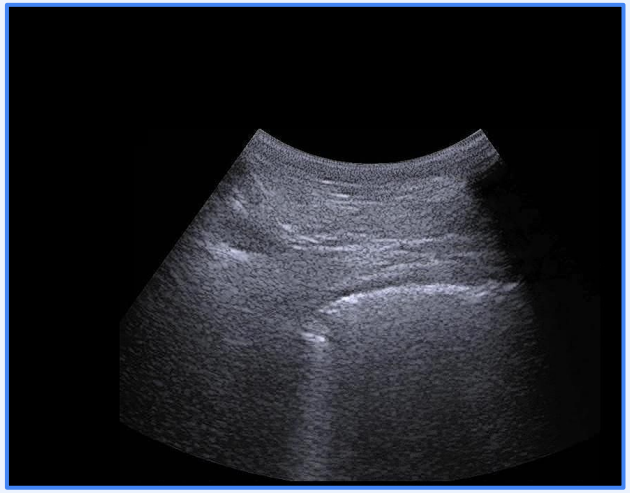
1. Frame into a classification model
2. Softmax output is analyzed by a classifier
3. If *false*, analyze the closer images score
4. Final score



1 Briefly on data

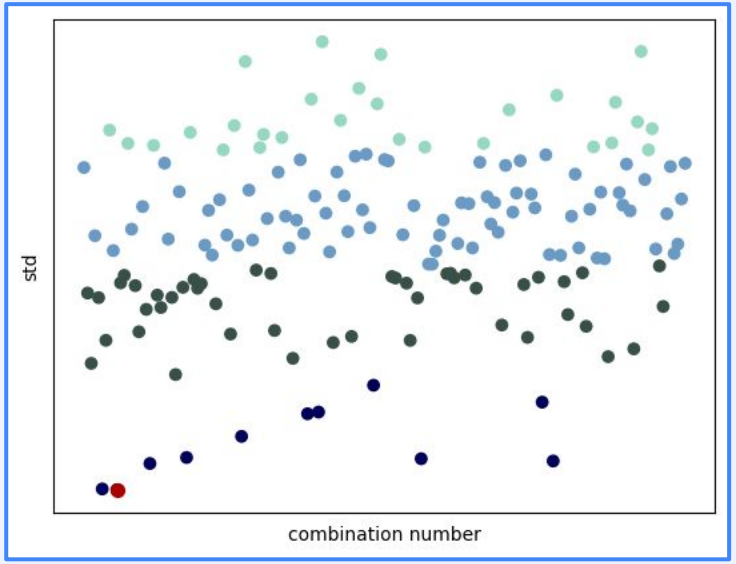
Lung Ultrasound Images

1. Images are scored from 0 to 3
2. This project works frame-wise
3. Augmentation was required
4. They seems in grayscale, but images are RGB



Dividing patients

To balance training, computed all possible 8-patients combinations stds between the number of frames per score -> took the smallest one



0	1	2	3
7888	7540	7189	7592

2

First classifier

3 4 5

Choosing the model

- ResNet18
- VGG16
- SqueezeNet
- CNN from scratch

Added layers to fine-tune, but easy overfitting. Working solution was to only add a layer to output the 4 classes

Augmentation

“Deep learning for classification and localization of covid-19 markers in point-of-care lung ultrasound”

E. Torri, R. Inchingolo, A. Smargiassi, G. Soldati, P. Rota, A. Passerini, R. J. G. van Sloun, E. Ricci, and L. Demi

- affine transformations
- multiplication with a constant
- Gaussian blurring
- horizontal flipping

3

Confidence

456

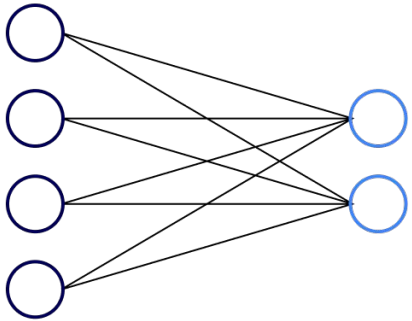
Threshold?

Independently from correct or wrong prediction and from the scores, the highest confidence in the softmax values were very similar

Understand the behaviour

Saving the softmax values and the prediction correctness

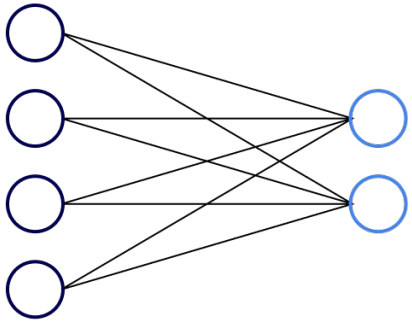
Neural Network



Sigmoid activation function

Basic model, any addition layer resulted in worse performance

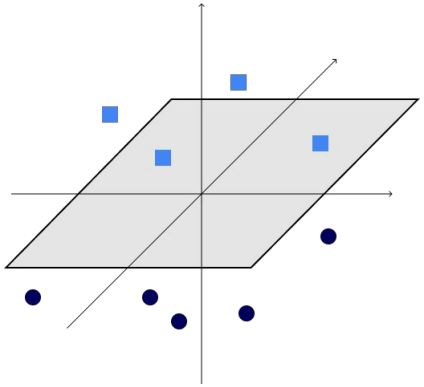
Neural Network



Sigmoid activation function

Basic model, any addition layer resulted in worse performance

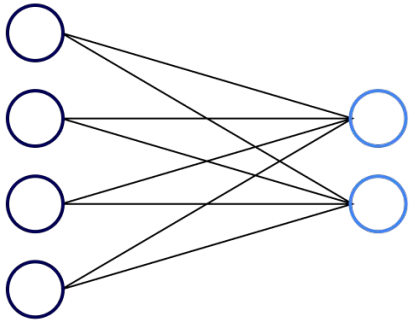
SVC



Support Vector Classifier

It required perfectly balanced dataset

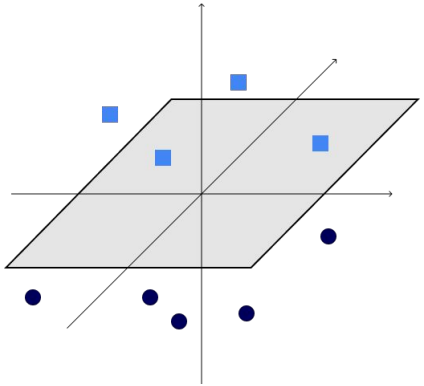
Neural Network



Sigmoid activation function

Basic model, any addition layer resulted in worse performance

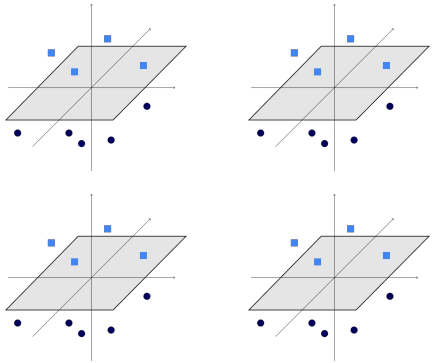
SVC



Support Vector Classifier

It required perfectly balanced dataset

4 SVCs



One for each class

The SVC used depended on the predicted score

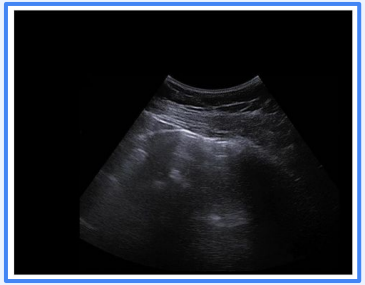
4 Similarity

Searching for similar images

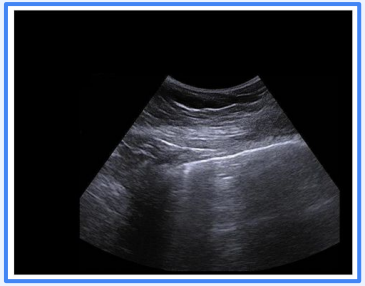
Only a portion of the available images from the train set has been used

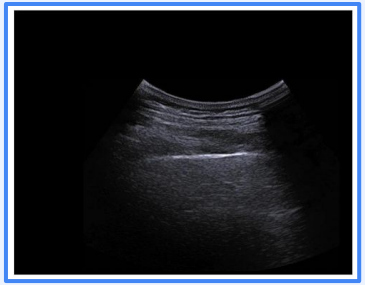
The same number of images for each score for each patient from the training set and again in the test set

QUERY



ANSWERS

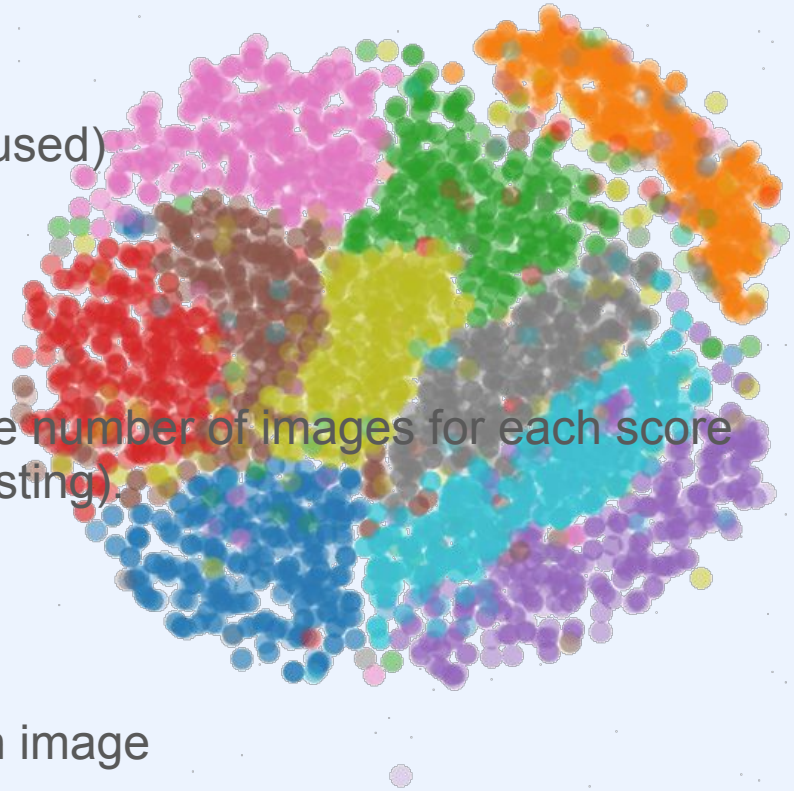




Near Duplicate Image Search

t-SNE

- similarity-wise (lighter, entire training set used)
- image-wise
- resnet18 embedding



First, this is not a classification method, is a…

Due to high time demanding, I chose the same number of images for each score for each training patients (and the same for testing).

Averaged the most X close images scores

For sim-wise ~Y%, for image-wise Z%

for a single frame, it required ~30 sec for each image

t-SNE

Raw Images

Softmax

Embeddings

Embeddings alone

5

Final results

Results

all possible configurations

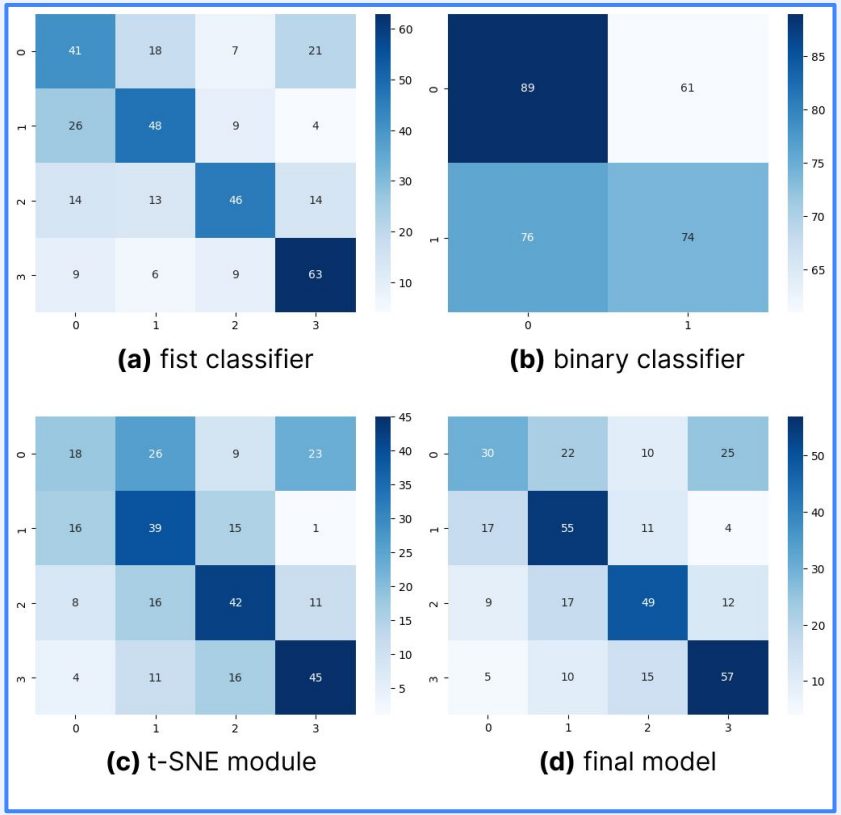
of the three components were tested. Although the behavior of each set of three modules was quite similar, the configuration that achieved the best accuracy was the one consisting of the three models that individually produced the best results.

The three components used in the final model are:

- ResNet18
- Single SVC
- t-SNE behaviour version

Results

verage accuracy of the final model
on this new test set is very similar to the one of just the
classification model. Starting from a 56.9% accuracy of the
classifier, with the 54.33% of the binary and with 48% of the
t-SNE module, the final model achieved a 54.89% accuracy
in the end.



Results

the behaviour of the
binary model in the final model was analyzed
and it showed
that, probably due to the little amount of data
used to train
it, it only got the right prediction on 3059/17512
images (the
entire test set), getting almost always the
correct result for the
label 1, less than 25% of the time with label 2
and very few
times with labels 0 and 3

Results

the similarity module uses only a small subset of the training images and it could be that most of the time the binary model states the first one is not a correct prediction it overlaps with the wrong results from the analysis of the similar images

Results

While there isn’t a single component responsible for the model’s lack of success, it seems that having multiple potential points of failure could have contributed to the final performance. Overall, it appears that the proposed model needs further improvements and adjustments to achieve better or more valu-
able results. However, it is important to note that the project
was still valuable as it provided insights and learning experi-
ences, but the effectiveness of the proposed model in practical
applications is questionable.

6

Extra?

Example screen of similar images

It can also show the similar images, so in a “real world scenario”, if there is uncertainty in a prediction, the X closest images can be plot:

And that’s it, thank you

Medical Imaging Diagnostic

Davide Modolo - Università degli Studi di Trento