The T-F-M Battle

Manasa Gadiyaram

Department of Electrical Engineering, Texas A&M University, College Station, TX.

Abstract

Motivation: In medical imaging, we often have very less amount of data and the unavailability of GPU. This is a problem because oftentimes, machines are used to assist doctors and with less data available, the model trained on that becomes unreliable. So, there has been an ongoing research as to how can we get a model that can work well and aid the healthcare workers. So, many people have come up with methods and models to tackle this issue. But training a model from scratch may not be feasible in every situation. My aim in this project will be to analyse whether fine-tuning an existing model or training it from scratch will provide better results, in this case the ResNet50 model. I plan to do this by comparing the performance of these models using the Osteopathology dataset containing images of cancerous and non-cancerous tumors. I am also planning to compare these models with Random Forest Classifier and observe the accuracy and speed of results. The classifier will perform texture-based classification.

Results: Transfer learning provides better and faster results as compared to full training of the model ResNet50. Also, when compared to a machine learning model like Random Forest, the deep learning models are time consuming. Accuracy is good in both type of learning methods.

Availability: The code and the report are available on https://github.com/Manasa-projects/Bioinformatics-final-project
Contact: manasa97@tamu.edu

1 Introduction

Healthcare organizations are becoming increasingly interested in using AI to provide better healthcare services to people all over the world. Within AI, mainly Machine Learning and Deep Learning have proven to provide results that corroborate the results obtained from actual doctors. In this project, we are trying to get an understanding on how Machine Learning and Deep Learning actually help to get a better judgement when it comes to identifying a disease.

One of the main methods that doctors use to analyze and judge whether their patients have any particular disease is by running blood tests, biopsies etc. They study the images obtained from these tests and give an educated judgement. Many times, owing to the complex nature of the human body, it becomes very difficult to give a firm judgment about the final results. Also, these analyses take time, thus taking away time from treatment. Thus, ML and DL help to reduce this time and also give better answers with better confidence.

What distinguishes AI technology from traditional technologies in health care is the ability to gain information, process it and give a well-defined output to the end-user. AI does this through machine learning algorithms and deep learning. These algorithms can recognize patterns in behavior and create their own logic. In order to reduce the margin of error, AI algorithms need to be tested repeatedly. AI algorithms behave differently from humans in two ways: (1) algorithms are literal: if you set a goal, the algorithm can't adjust itself and only understand what it has been told explicitly, (2) and some deep learning algorithms are black boxes; algorithms can predict extremely precise, but not the cause or the why.

There are many diseases and there also many ways that AI has been used to efficiently and accurately diagnose them. Some of the diseases that are the most notorious such as Diabetes, and Cardiovascular Disease (CVD) which are both in the top ten for causes of death worldwide have been the basis behind a lot of the research/testing to help get an accurate diagnosis. Due to such a high mortality rate being associated with these diseases there have been efforts to integrate various methods in helping get accurate diagnosis'.

It has been demonstrated that there are several types of AI techniques that have been used for a variety of different diseases. Some of these techniques include: Support vector machines, neural networks, Decision trees, and many more. Each of these techniques is described as having a "training goal" so classifications agree with the outcomes as much as possible.

In this project, we are going to harness the usability of Random Forest algorithm and ResNet50 pre trained deep learning model to understand how they can provide excellent results when it comes to detecting diseased cells.

2 Methods

The model used in this project is the ResNet50. ResNet-50 is a convolutional neural network that is 50 layers deep. It is a pretrained version of the network trained on more than a million images from the ImageNet database. The pretrained network can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals. As a result, the network has learned rich feature representations for a wide range of images. The network has an image input size of 224-by-224. It has over 76 million parameters that can be obtained by training the model from scratch. This requires considerable amount of memory and data, both of which are not in abundance in healthcare industries. But, to get accurate results, the industries still rely on this method, even though it is time consuming and data demanding.

Another method of using pre-trained model is by performing transfer learning. This literally means transferring knowledge learnt from one network to another. This can be done by removing the last layer of a pre-trained model and adding our own dense layers. During training, only these layers will be trained and the rest of the model will be freezed or remained untouched. This helps because the initial layers of any model are for extracting bigger features, which can be transferred from natural images to medical images. Also, the time required to train is less and they require less memory.

In the day and age where deep learning is dominating all areas of life, Machine Learning has not lost its stand. It tends to provide faster results as compared to deep learning models. In this project, we have used Random Forest as the model as it not only provides good results but also does that in very less time. Also, this algorithm obtains results using various combination of features, making it more robust and diverse.

3 Dataset

The dataset consists of Ostepathology images of tumour cells taken from teenagers between the age of 12 and 19. There are a total of 1144 images which have been collected from the National Cancer Institute database. Apart from this, the dataset also contains an excel file capturing texture features of each of these images, which have a god potential to be harnessed. This dataset seemed the best to work with because it was balanced with good number of positive and negative cases. Also, this type of cancer is very prevalent in teenagers, thus this project may help to contribute in better research in this area.

4 Results

The three methods- transfer learning, machine learning and full training have been compared on the basis of accuracy and error. They are summarized as follows:

4.1 Transfer Learning

<u>Input:</u> Images are resized to (150,150,3) and given as the input. They are all converted to Numpy arrays. This size allows for the model to use less memory. Each of these images have been zoomed, sheared, rotated and translated to increase the dataset size.

<u>Method:</u> For transfer learning, two dense layers with two dropout layers with 'Relu' activation function were added. The output layer is a dense layer with the 'Sigmoid' activation function. The model used the 'mean-squared-error' loss with an 'sgd' optimizer.

<u>Hyperparameters:</u>

Epochs-35

Batch size-12

<u>Results:</u>

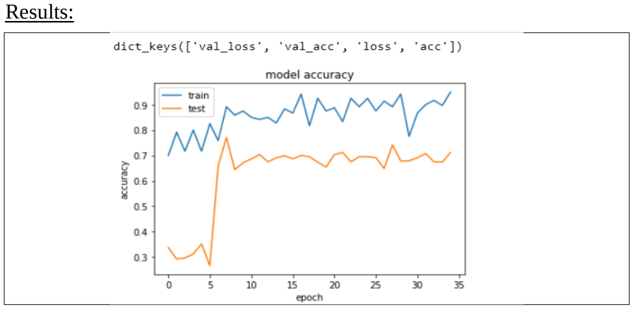

Fig. 1. Accuracy of transfer learning model.

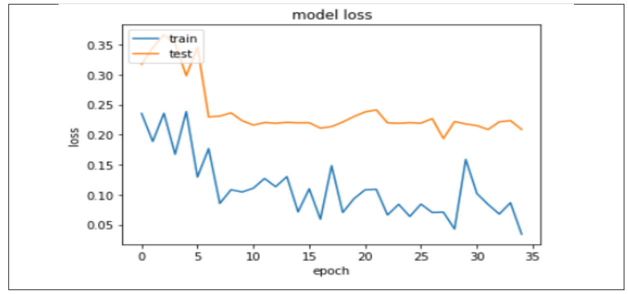

Fig. 2. Loss of transfer learning model

<u>Inference:</u>

We can see from the above graphs that the testing accuracy and training accuracy are almost nearby and high enough, reaching up to 70% for testing data. Thus, no underfitting or overfitting is taking place. The gap between both the accuracies can be decreased by taking more data.

4.2 Full Training

<u>Input:</u> Images are resized to (224,224,3) and given as the input. They are all converted to Numpy arrays. This size is the standard input size used for the ResNet50 model. Each of these images have been zoomed, sheared, rotated and translated to increase the dataset size.

<u>Method:</u> In this experiment, 2 fully connected layers have been added to the output layer and no layers have been freezed. Thus, a total of 76 million parameters have to be trained to fully train the ResNet50 model. The two dense layers with two dropout layers with 'Relu' activation function were added. The output layer is a dense layer with the 'Sigmoid' activation function. The model used the 'mean-squared-error' loss with an 'sgd' optimizer.

<u>Hyperparameters:</u>

Epochs-35

Batch size-12

<u>Results:</u>

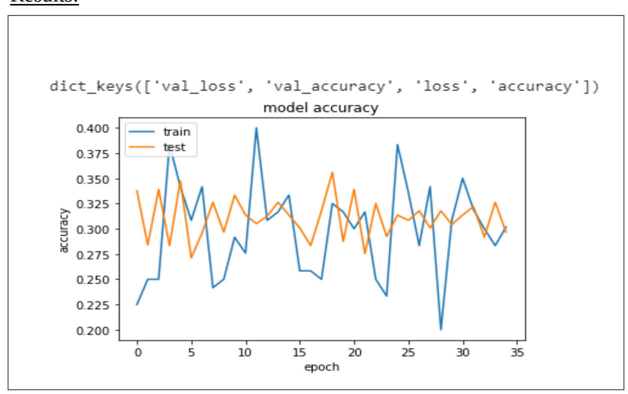

Fig. 3. Accuracy of Full training model

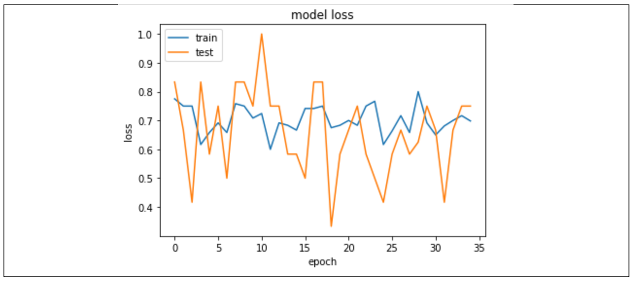

Fig. 4. Loss of Full Training model

<u>Inference:</u>
From above, it can be observed that the accuracy/loss of both training and testing data is very close, so that means that at least overfitting is not taking place. But, the value of accuracy is too less, around 30% (70% for the loss). This means underfitting is taking place. Thus, we need a lot of data to achieve better accuracy since there are 76 million parameters to be determined.

4.3 Machine Learning – Random Forest

<u>Input:</u> Texture information is read from the excel sheet. There are 53 features which define the texture of each of the image. Thus, in total there are (53*1144) datapoints. This information is fed into the Random Forest model.
<u>Method:</u> The Random Forest model with a depth of 5 has been taken to avoid overfitting.
<u>Results:</u>

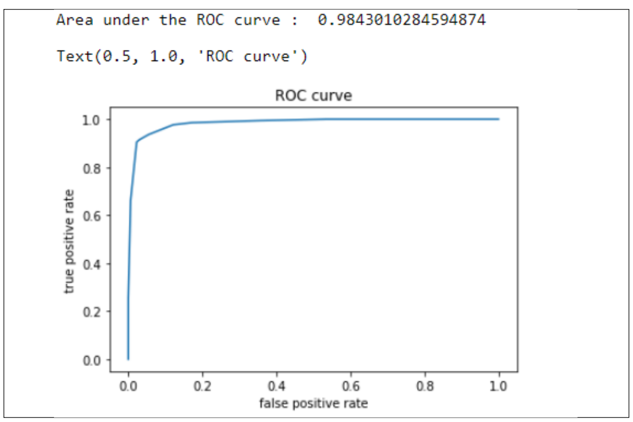

Fig. 5. AUC curve for Random Forest Classifier

<u>Inference:</u>
It can be observed from above that the ROC curve has large area under it, which proves that this is an accurate classifier. Also, time taken to predict results was less than a second. Thus, this seems to overpower deep learning models.

Observations

From the detailed analysis above, we observed the following points:

1. The time taken to train the complete ResNet50 network from scratch is around 1 hour. This also requires high powered GPU's. In this project, this network could not be trained on local machines like laptops, even with ones with GPU. We used Google Colab, which took 1 hour to train. Even after this time to train, the accuracy of results was around 30%.
2. The transfer learning was completed in 30 minutes, on the local machine itself. The accuracy was around 70%, which is not bad sinceResNet50 was initially trained on natural images. Thus, the network adapted to biological images quite quickly.
3. The area under the curve for Random Forest model is 0.984. This means that the accuracy is extremely high for this model. Also, this model tried out different combinations of features randomly, thus also making the model more robust and freer from unnecessary features.

Conclusion

From the detailed analysis above, we can conclude the following points:

1. The time taken to train the ResNet50 network from scratch shows that it is a time-consuming process. This happened when the dataset was small, around 1144 images, so we can imagine the time it will take for more data. It may take weeks or months to train. This is not exactly desirable in healthcare industries since, medical professional require assistance immediately and with precision. Thus, too much time can become a hinderance.
2. The GPU requirement for ResNet50 training from scratch is massive. Thus, healthcare industries will have to spend extra on cloud services or GPU services to train their networks. This make the process dependent on other aspects. Also, GPU availability is not present in many machines throughout the world. Many people still use machines with only CPU. Thus, complete training is not economically friendly since investing in GPU's takes a lot of incentives.
3. The data availability is not huge in healthcare industries, thus the poor performance and high error rate of fully training a model shows that it is not suitable for the medical world.
4. The transfer learning model takes less time to train, thus speed-wise, it is a good option to use.
5. The accuracy is around 70%. This is good enough for natural images, but not for the medical world. We need very precise answers as we do not want false positives or worse false negatives. Thus, we have to tune the network better to get good results. More data can also help in this process.
6. It would be better to develop a model like ResNet50 which can work as a pre-trained model for medical problems. It may be possible that we can get higher accuracy if we use such a model for these issues. But sadly, a model like that is yet to be developed.
7. The Random Forest Classifier works the best in this case. It gives almost ideal answers with a high true positive rate and low false positive rate. But this model is very feature oriented and also discards some features randomly. This may not be helpful when certain features which are very important ten to get discarded.
8. Thus, it can be concluded that the transfer learning model and machine learning model work the best. Even among those, it is better to go with transfer learning model as it takes into account all the features which it learns from the data and does not discard anything.

Acknowledgements

I would like to thank Dr. Yang Shen for giving me guidance on how to select a project based on my interests and for being there to clear my doubts at all times. This project would have been impossible without him. I would also like to thank my parents and friends for supporting me through my journey of not only this project but also though my life in the US.

References

Andrearczyk, Vincent & Whelan, Paul. (2016). Deep Learning for Biomedical Texture Image Analysis.

Justin Ker, Lipo Wang, Jai Rao, Tchoyoson Lim, "Deep Learning Applications in Medical Image Analysis", Access IEEE, vol. 6, pp. 9375-9389, 2018.

Fangyan Ye, Jian Pu, Jun Wang, Yuxin Li, Hongyuan Zha, "Glioma grading based on 3D multimodal convolutional neural network and privileged learning", Bioinformatics and Biomedicine (BIBM) 2017 IEEE International Conference on, pp. 759-763, 2017

Marcia Hon, Naimul Mefraz Khan, "Towards Alzheimer's disease classification through transfer learning", Bioinformatics and Biomedicine (BIBM) 2017 IEEE International Conference on, pp. 1166-1169, 2017.

Haipeng Wan, Lei Chen, Hong Song, Jian Yang, "Dorsal hand vein recognition based on convolutional neural networks", Bioinformatics and Biomedicine (BIBM) 2017 IEEE International Conference on, pp. 1215-1221, 2017.

Agnese Marchesi, Alessandro Bria, Claudio Marrocco, Mario Molinara, Jan-Jurre Mordang, Francesco Tortorella, Nico Karssemeijer, "The Effect of Mammogram Preprocessing on Microcalcification Detection with Convolutional Neural Networks", Computer-Based Medical Systems (CBMS) 2017 IEEE 30th International Symposium on, pp. 207-212, 2017.v

Xiaogang Li, Tiantian Pang, Biao Xiong, Weixiang Liu, Ping Liang, Tianfu Wang, "Convolutional neural networks based transfer learning for diabetic retinopathy fundus image classification", Image and Signal Processing BioMedical Engineering and Informatics (CISP-BMEI) 2017 10th International Congress on, pp. 1-11, 2017.

Hak Gu Kim, Yeoreum Choi, Yong Man Ro, "Modality-bridge transfer learning for medical image classification", Image and Signal Processing BioMedical Engineering and Informatics (CISP-BMEI) 2017 10th International Congress on, pp. 1-5, 2017.