9th International Workshop on Innovative Simulation for Healthcare
17th International Multidisciplinary Modeling & Simulation Multiconference

ISSN 2724-0371 ISBN 978-88-85741-50-8 © 2020 The Authors.
DOI: 10.46354/i3m.2020.iwish.004

Multitask Learning for Medical Image Classification Using VGG Architecture

Yuan Yang[1,2,3], Lin Zhang[1,2,3,*], Lei Ren[1,2,3] and Yuanjun Lali[1,2,3]

[1]School of Automation Science and Electrical Engineering, No. 37 Xueyuan Road, Haidian District, Beijing, 100191, China
[2]Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, School of Medicine and Engineering,No. 37 Xueyuan Road, Haidian District, Beijing, 100191, China
[3] Key Laboratory of Big Data-Based Precision Medicine,Ministry of Industry and Information Technology, No. 37 Xueyuan Road, Haidian District, Beijing, 100191, China

[*]Corresponding author. Email address: johnlin9999@163.com

Abstract

In order to recognize different kinds of medical images under a single network structure background, a multi-task medical image recognition model based on the combination of transfer learning and automatic path search is proposed. Based on VGG-16 model, a neural network module is designed and evolutionary algorithm is used to select the path. Experiments were conducted on ECG data sets and pneumonia data sets respectively. Finally, a joint classification test was conducted on these 2 datasets. Ultimately, joint classification experiments on the ECG and pneumonia datasets resulted in an overall accuracy of 93% and a recall rate of 88%.

Keywords: transfer learning; multitask learning;neurological architecture search;medical image

1. Introduction

We perceive the world through a multimodal, and multitasking, lens. Observing objects with our eyes, listening to sounds with our ears, etc. In order for artificial intelligence to enable us to better understand the world around us, it is necessary to target this multimodal and multitasking signal to be learned. Multi-task learning has attracted more and more researchers' attention. Multitask learning is used in natural language processing, such as the task of understanding multiple languages(Liu et al., 2020). Multitask learning has applications in face recognition as well.

Xi et al. proposed and built a multi-task learning network to implement the main task of face recognition and three secondary tasks (pose/illumination/emotion) classification. It also makes the main task and the three auxiliary tasks mutually reinforcing(Yin and Member). At the same time, some tasks are inherently multitasking, such as automatic driving(Chen et al., 2017).

Machine learning techniques typically require a large number of training samples to learn and train a model. Deep learning has to introduce quite a lot of data and perform relevant calculations, so it is bound to be quite costly. In the meantime, more complex models have to be designed, and a lot of manpower, resources and time have to be introduced to complete these algorithms, design models and so on. At the same time, the model also needs to be trained by introducing quite a lot of training data to ensure that a certain level of accuracy is achieved. However, in some applications, such as

medical image analysis, this requirement cannot be met because labeled samples are difficult to collect. For this problem of insufficient data, multi-task learning is a good solution(Caruana, 1997).

With the development of IoT, cloud computing, and wearables, medical data is electronic and exploding. At the same time, it may be possible to provide this medical data from different hospitals, devices, and regions, etc. Therefore, it must be affirmed that if it is possible to smooth out these multiple sources, multiple formats, and explosive growth of these medical data analysis and processing, then these data can certainly improve the quality of care, ensure more patient safety, reduce medical risks, costs, etc(Dinov, 2016). The key contribution of the paper is given below:

- Propose a multi-task learning model for medical data, which can process different physiological signal data of human body, especially medical image data.Based on transfer learning, the purpose of fast training and diagnosis of diseases is realized by using mature trained network models. The last layer of the model uses the autoML framework of general artificial intelligence to construct automatic network search.

2. State of the art

Multitasking learning also has applications in medical image processing, with good results. Liu proposed a new characterization method that preserves complementary information between multimodalities. Using multimodal synthesis information to improve the diagnosis of Alzheimer's disease(Liu et al., 2014). Wang proposed to identify 36 different retinal diseases using region-specific multitasking recognition models(Wang et al., 2019). Sitthichok et al. Constructed into a single CNN model to simultaneously perform image classification and image Segmentation tasks(Chaichulee et al., 2017). The proposed multitasking network has a shared core network with two branches: a patient detection branch implemented using a global average pooling layer; and a skin segmentation branch implemented using feature maps from the entire shared core network. Chen et al. propose a multitasking network that optimizes both the segmentation task for supervised learning and the image for unsupervised learning Reconstruction. And two multitasking training strategies are also compared: joint training and alternate training(Chen et al.). Yan et al. proposed a multitasking universal lesion analysis network (MULAN), which can jointly detect, tagging and segment lesions in various parts of the body, i.e., the detection task, tagging task and image segmentation task can be performed simultaneously(Yan et al., 2019).

A common feature of the input data processed above is that the input raw data is the same, but different analysis tasks are performed on the same input data at the same time. The problem addressed in this paper is to train a generic network capable of classifying diseases from different image data.

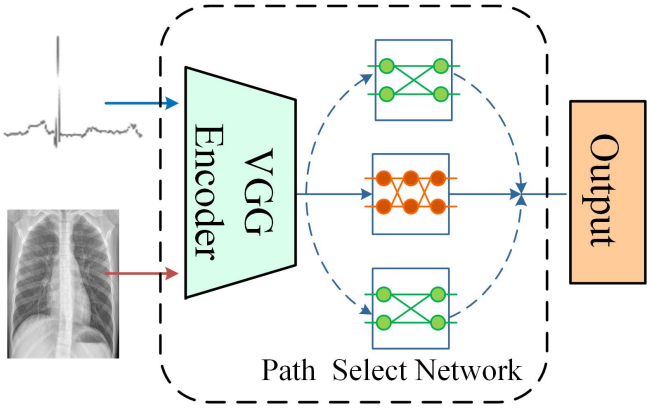

Figure 1. Main Body Architecture Of The Multitask Neural Network Model

3. Materials and Methods

The whole model architecture consists of two parts, one is encoder and the other is path select network which can realize multitask. Figure1 shows the main architecture of the multitasking neural network model. Here, the main function of the encoder is to extract the features from the image. The basic function of the path selection network is mainly to perform tasks that are in different classification contexts.

3.1. Encoder architecture

The encoder adopts the idea of transfer learning, using the VGG-16 model structure for transfer learning. Transfer learning refers to taking a convolutional neural network model trained on a task and briefly adjusting it to meet the requirements of a new task. These convolutional layers of the trained neural network can extract the features of the model, and the extracted feature vectors can be directly fed into these relatively simple structures of the full-connection layer, and ensure that the recognition and classification achieve the expected results.

In the article, I primarily use the VGG-16 basic network model for the entire training effort. VGG-16 is a network model consisting of the a neural network model developed by the Visual Geometry Group and other departments at Oxford University(Simonyan and Zisserman, 2015). It consists of: 13 convolutional layers, 5 pooling layers, and 3 fully connected layers. The 3 fully connected layers consist of the path selection network mentioned in the next section Substitution.

3.2. Path select network architecturee

In recent years, the neural architecture search (NAS) method has received a great deal of attention from the academic community as a new approach. Concerning the NAS algorithm, the neural structures it is used to search out are superior in terms of performance to those designed manually. First, a single neural structure is selected in a given search space according to a structure search strategy. The structure is then evaluated with the help of a performance evaluation method. The results obtained are returned to the relevant structure search strategy, which then adjusts the subsequent neural structure selection. This process is carried out in a cyclic and iterative manner until a neural structure is found that meets the performance criteria. In this process, performance evaluation methods, search space, and structured search strategy are the 3 key elements. It uses a certain Search Strategy, selects a model from Search Space, and then evaluates the effect of the model. Google Brain's NasNet paper lists some of the most basic network modules that can be used to build networks like building blocks(Zoph et al., 2018). PathNet is a network of neural networks, trained by random gradient descent and genetic selection(Fernando et al., 2017). PathNet consists of layers of modules, each of which can be any type of neural network-convolution network, feedforward network, etc.

The network structure used in this case is: a modular deep neural network that can be set up in L layers, each layer consisting of M modules. Each module is itself a small neural network, generally of only two types: convolutional or linear, the last of which is an activation function. For each layer, the output of the modules in that layer is summed and then passed to the next layer.

3.3. Path Selection Evolutionary Algorithm

After defining the representation of the network architecture, we want to extract the architecture with more advantageous performance through a larger search space . This process is often thought of as manipulating each node and optimizing connections. This process is also often referred to as hyper-parametric optimization. The most commonly used algorithm in hyperparametric optimization is the evolutionary algorithm. Evolutionary algorithms (EA) are essentially general population-based heuristic optimization algorithms, inspired by biological evolution. . The method is more similar to a global than previous calculus, or exhaustive, based algorithms, while the method It is relatively mature and its adaptability to the environment is high. The biggest advantage of its proposal is that it can deal with problems that are not handled by traditional methods and are not affected by the nature of the problem. Here, the more representative evolutionary algorithms specifically cover the following aspects: selection, crossover, variation, and update.

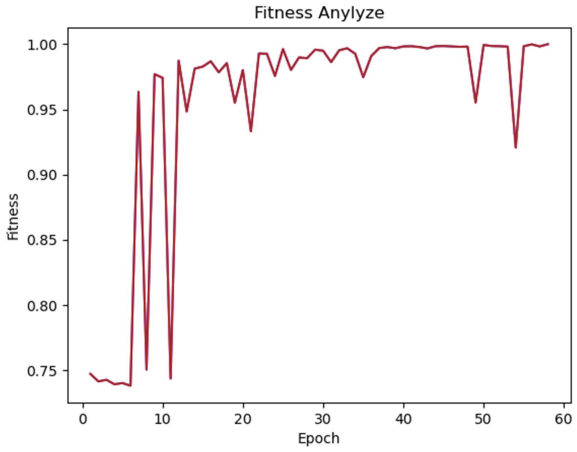

Figure 2. The Evolutionary Curve Of Population Adaptation

The selection of a single network is carried out by three different strategies. Fit selection is one of them, which means that the probability of the individual to be selected is positively related to its fit value. P genotypes are initialized randomly, each genotype is at most a n by l matrix of integers. This matrix represents the active modules. This is the process by which the network model is encoded. During the evolution process, the collection of network models will be maintained and expanded, and the fitness of these network models will be given by their accuracy on the verification set. During the process, two models will be randomly selected, the poor one will be killed (eliminated) directly, and the good one will become the parent node. This method is called tournament selection. Figure2 shows the evolutionary curve of population adaptation when the model was doing the electrocardiogram classification task.

4. Results and Discussion

4.1. Electrocardiogram Classification

Here this paper introduce the MIT-BIH arrhythmia database (https://www.physionet.org/content/mitdb/1.0.0/). It is frequently put to use in the study of arrhythmia diseases and in the evaluation of corresponding medical devices. Here, a total of about 48 ECG signal data were extracted, and each data file is specific to the patient's 0.5-h second-lead Recording of ECG information. In this paper, I use 41 types of annotated information for this aspect of the data, mainly: normal fluctuations, left bundle branching annotations for block, right bundle branch block, etc., as well as information on waveform changes in the ECG signal. It should be noted that although many abnormal heart rhythms are covered in the MIT-BIH, however, because the entire database contains a total of 48 data

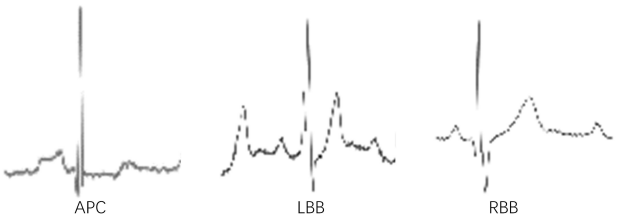

Figure 3. The Evolutionary Curve Of Population Adaptation

records, some arrhythmias are labeled with a small sample size.

Therefore, in this case has chosen to classify the arrhythmia diseases with relatively large sample sizes. In addition to the four arrhythmia diseases, such as: ventricular premature beats (VPC), left bundle branch block (LBB), and atrial premature beats (APC), and right bundle branch block (RBB), the number of disease labeling is less than a thousand, and so it is possible here to pass the These five diseases were used to build the model. In this paper, three diseases of APC, LBB and RBB are chosen to construct the classification model. Since some of the data records in the MIT-BIH heart rate database fail to maintain good quality, the three diseases that I have selected here are mostly from some signal stable records to perform the extraction operation of the features. In this paper, the annotation information is used to center the R wave, thus obtaining a series of ECG signals and converting the ECG signals into images. A Electrocardiogram classification involves distinguishing three classes of ECG signals from one another. The two-dimensional picture of the electrocardiographic waveform generated by the data preprocessing is shown in Figure 3.

The overall selection path network consists of L=3 layers. Each layer contains M=10 cells. The convolutional module is in the first layer and the fully connected module is in the next two layers. At this point, a new population of path genotypes is initialized and evolves on this task until perfect performance is achieved on the training set . The ECG dataset in which 20% of the data is used as a validation set. The purpose of the validation set is to improve the accuracy of model classification and to prevent model overfitting. The parameters contained in the optimal path evolved on this task are fixed. After the path selection is complete, the entire network is then trained and tested. Figure4 and Figure5 shows the performance of the training and test sets in relation to the number of epochs. In this paper, the performance of the proposed method is evaluated using the most commonly used metrics: accuracy and recall. The confusion matrix diagram of the testing process is shown in Figure 6. Table 1 shows that the proposed method achieves the best accuracy in the previous state of the art. The work of Fujita et al. achieved an accuracy of 94.9%(Acharya et al., 2017). The work

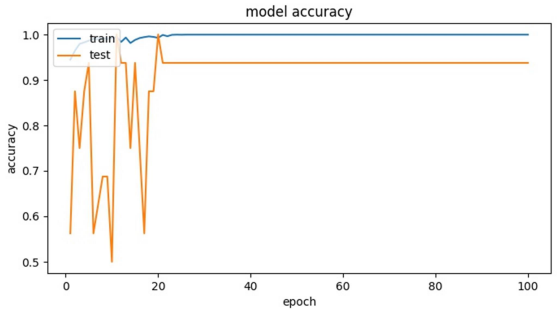

Figure 4. Model Accuracy

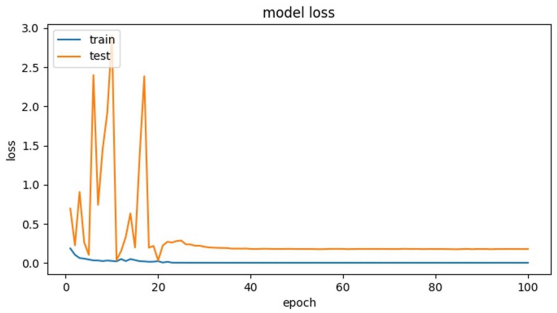

Figure 5. Model Loss

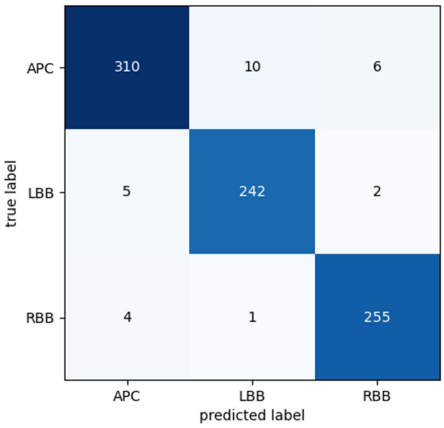

Figure 6. Confusion Matrix On The Test Set

of Inan et al. achieved an accuracy of 95.2%(Inan et al., 2006).

Results can be obtained from the confusion matrix. The method proposed in this paper was well applied on level 3 classification with an accuracy rate of 96.5% and a recall rate of 96.8%.

4.2. Chest-Xray-Pneumonia Classification

In this case the data set I used was the Kaggle Pneumonia Competition. The website is

Table 1. An example table.

Methods	Classifier	Class	Accuracy
Fujita et al.2017	CNN	2	94.90%
Inan et al. 2006	NN	2	95.20%
Our method	CNN	3	96.50%

https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia. The data consisted of 5863 images, divided into three parts: the training set, the test set and the validation set. Data enhancement is performed as a regularization mechanism to avoid immediate overfitting. The operations performed in real time on the training set are as follows:

1. Horizontal flip.
2. Scaling range is 10%.
3. Randomly rotate 0.1 degrees.

The pre-processed data set is divided into three parts: the training set, the test set and the validation set. The training set is used to train the model and the test set is used to validate the model. And to prevent the model from being inaccurate in the analysis of hard-to-identify images, this paper constructs a validation set to provide timely feedback on the model during the training process. In this paper, the post-test model predictions are presented in a confusion matrix, and the accuracy and recall are calculated from this. The model was evaluated using the test set and the final results were 88% accuracy and 76% recall. The model's performance in terms of recall means that it detects a large majority of patients, which can buy time for patients to seek early treatment. The accuracy suggests that the model incorrectly classifies some normal patients as patients, which in practice requires further judgment by the attending physician. Different from the previous ECG classification task, this task belongs to the second classification task, namely pneumonia and normal. The overall select path network has the same path configuration as the previous task and has not made any changes to the network structure. The confusion plot for the test process is shown in Figure7. From the confusion chart we can calculate that the accuracy can reach 88%.

4.3. Comparison experiment with no path selection network

The entire model architecture consists of two parts, one for the encoder and the other for the path selection network. In order to verify the role of the path selection network in this subject network, comparative experiments were done. Comparison experiments were performed on the electrocardiographic dataset and the pneumonia dataset, respectively, by adding a path selection network and not adding a path selection network. The experimental results on the 3-classified ECG dataset, i.e., the confusion matrix with no path selection network added are shown in Figure8. Comparing figure 8and figure6 you can see the difference is obvious. The effect of adding a path selection network is significantly better than the effect of not adding a path selection network. As can be calculated from figure8, the classification accuracy is 72% and the recall rate is 56%.Both the accuracy and recall rates are significantly lower than the results in figure 6.

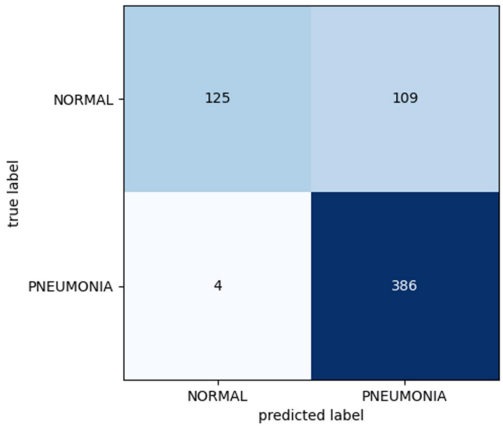

Figure 7. Confusion Matrix On The Test Set

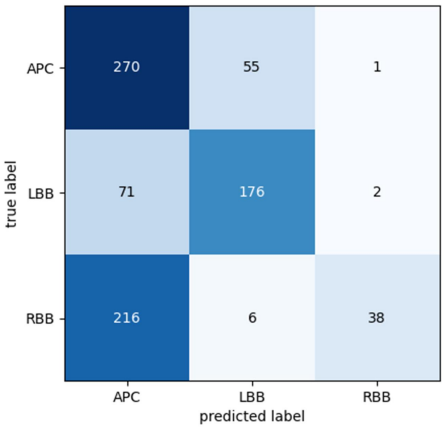

Figure 8. Confusion Matrix Results For Without Adding Path Selection Networks

The experimental results on the 2-classified pneumonia dataset, i.e., the confusion matrix with no path selection network added are shown in Figure9. Comparing Figures 9 and 7, the difference in results can be seen. The effect of adding a path selection network is better than the effect of not adding a path selection network. From Figure 9 it can be calculated that the classification accuracy on the pneumonia data set is 48% and the recall rate is 48%. Both the accuracy and recall rates are significantly lower than the results in

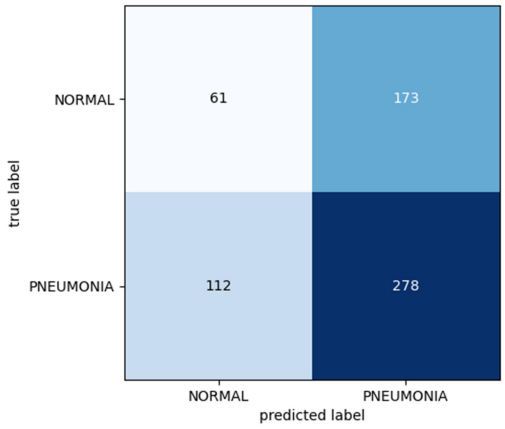

Figure 9. Confusion Matrix Results For Without Adding Path Selection Networks

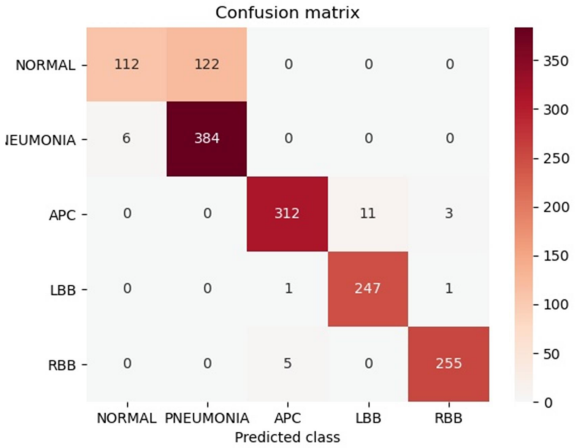

Figure 10. Confusion Matrix Results For joint classification experimental

Figure 7.

4.4. Comparison experiment with no path selection network

The above experiments on the ECG dataset and the pneumonia dataset are independent. In order for the model to achieve multitask classification, the network fixes the weight file and the corresponding path when conducting independent experiments. Thus, when throwing 2 data sets into the network model at the same time, the first step determines whether it belongs to the ECG data set or the Pneumonia data set, and the second step takes different paths and uses different weight files for different data. The final classification results were obtained.

Figure10 shows the experimental results of the joint classification on the 2 data sets. Comparing Figures10 and 7 , the difference in experimental results can be seen. In terms of the results on the pneumonia test set alone, the combined effect was slightly less than the effect of the experiment alone. But the accuracy rate, the recall rate also reached 85 percent and 73 percent, respectively. Comparing Figures10 and 6, the experimental results can be seen. As far as the results on the ECG test set alone are concerned, the combined effect is essentially equal to the effect of the experiment alone. Accuracy, recall rates were all 97% and 97% respectively. But the advantage of this network model is that it enables the processing of different kinds of medical data sets. From Figure10, it can be calculated that the overall classification accuracy on the combined data set is 93% and the overall recall rate is 88%.

5. Conclusions

In this work, we construct models that can handle many different medical image classifications. The modified model was experimented on the MIT ECG dataset and the kaggle pneumonia dataset, respectively. In the electrocardiographic classification, the electrocardiographic time series data were first transformed into 2-dimensional pictures, and finally the accuracy of the method was 97%. The lack of excessive pre-processing of pneumonia images resulted in poor classification on the final pneumonia dataset. But the main advantage of this model is that it can handle many different medical image classification tasks. Although we only demonstrated how to realize multi-task learning of medical data in a relatively small network, this principle can be extended to a larger neural network. This network structure allows expansion to realize multiple medical image classification tasks. The limitation of this network is that this network has to be trained in two phases. The first phase requires finding paths that are highly adaptable and the second phase requires training the network weighting parameters. Therefore, the training is time consuming.

6. Funding

This work was supported in part by Beijing Advanced Innovation Center for Big Data-Based Precision Medicine.Principal Investigator:Lin Zhang.

References

Acharya, U. R., Fujita, H., Lih, O. S., Hagiwara, Y., Tan, J. H., and Adam, M. (2017). Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network. *Information Sciences*, 405:81–90.

Caruana, R. (1997). Multitask Learning. *Machine Learning*, 28(1):41–75.

Chaichulee, S., Villarroel, M., Jorge, J., Arteta, C., Green,

G., McCormick, K., Zisserman, A., and Tarassenko, L. (2017). Multi-Task Convolutional Neural Network for Patient Detection and Skin Segmentation in Continuous Non-Contact Vital Sign Monitoring. *Proceedings - 12th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2017 - 1st International Workshop on Adaptive Shot Learning for Gesture Understanding and Production, ASL4GUP 2017, Biometrics in the Wild, Bwild 2017, Heteroge*, pages 266–272.

Chen, S., Bortsova, G., Juárez, A. G.-u., Tulder, G. V., and Bruijne, M. D. Learning for Medical Image Segmentation. pages 1–9.

Chen, X., Ma, H., Wan, J., Li, B., and Xia, T. (2017). Multi-view 3D object detection network for autonomous driving. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017-Janua:6526–6534.

Dinov, I. D. (2016). Methodological challenges and analytic opportunities for modeling and interpreting Big Healthcare Data. *GigaScience*, 5(1).

Fernando, C., Banarse, D., Blundell, C., Zwols, Y., Ha, D., Rusu, A. A., Pritzel, A., and Wierstra, D. (2017). PathNet: Evolution Channels Gradient Descent in Super Neural Networks.

Inan, O. T., Giovangrandi, L., and Kovacs, G. T. (2006). Robust neural-network-based classification of premature ventricular contractions using wavelet transform and timing interval features. *IEEE Transactions on Biomedical Engineering*, 53(12):2507–2515.

Liu, F., Wee, C. Y., Chen, H., and Shen, D. (2014). Inter-modality relationship constrained multi-modality multi-task feature selection for Alzheimer's Disease and mild cognitive impairment identification. *NeuroImage*, 84:466–475.

Liu, X., He, P., Chen, W., and Gao, J. (2020). Multi-task deep neural networks for natural language understanding. *ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, pages 4487–4496.

Simonyan, K. and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pages 1–14.

Wang, X., Ju, L., Zhao, X., and Ge, Z. (2019). Retinal abnormalities recognition using regional multitask learning. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11764 LNCS:30–38.

Yan, K., Tang, Y., Peng, Y., Sandfort, V., Bagheri, M., Lu, Z., and Summers, R. M. (2019). MULAN: Multitask Universal Lesion Analysis Network for Joint Lesion Detection, Tagging, and Segmentation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11769 LNCS:194–202.

Yin, X. and Member, X. L. Multi-Task Convolutional Neural Network for Pose-Invariant Face Recognition. pages 1–12.

Zoph, B., Vasudevan, V., Shlens, J., and Le, Q. V. (2018). Learning Transferable Architectures for Scalable Image Recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 8697–8710.