# Multitask learning for haemodynamics

Titouan Brossy, Francesco Sala, Nicolò Viscusi
Supervisor: Riccardo Tenderini
CS-433 - Project 2
*EPFL, Lausanne, Switzerland*

***Abstract*—Supervised machine learning techniques can be applied in medicine to predict the insurgence of disease such as myocardial infarctions related to stenotic formation. In this project we investigated the accuracy of different multitask learning approaches (WDA and OL-AUX) with respect to the dimension of the available dataset (from $5\%$ to $20\%$) while optimizing a selection of related hyperparameters. The results obtained show that multitask learning techniques allow to significantly reduce the prediction error when using a reduced-size dataset. Moreover, we show that only two layers of the adopted MTL neural network architecture need to be task-specific in order to attain the smallest test error.**

## I. Introduction

This project is based on the work of Betti et al. [3]. In collaboration with the Centre Hospitalier Universitaire Vaudois (CHUV), a machine learning model was implemented to predict myocardial infarction (MI) risk. An *in-silico* dataset of steady-state CFD solutions was generated and post-processed in order to mimick angiographic images of a patient-specific femoropopliteal bypass. This procedure was necessary due to the well-known issue (especially in clinical applications) of data availability: patients rarely undergo imaging exams before having an MI episode. For each of the images generated, relevant parameters that identify the stenosis geometry were stored. Furthermore, an artificial MI risk function was designed, using some of the aforementioned parameters and based on literature results. A neural network architecture was then developed in order to predict MI risk. Three different scenarios of MI risk prediction were investigated based on the type and amount of data at disposal. In the first scenario, it was assumed that 100% of labels (selected parameters) was available and Single Task Learning (STL) was employed. Since it could be argued that this is not a realistic scenario, given the previously mentioned data availability issue, two other scenarios were considered. In the first, it was assumed to have at disposal some labels for every data point and MI labels for 20% of the entire dataset. In the second instead, it was assumed that only the 20% of the dataset has all the labels available. The present work focuses only on the third scenario exploiting MTL techniques; Sections until Section II-F are a summary of the work carried out by Betti et al [3].

## II. Models and methods

### A. Dataset generation and preprocessing

The dataset generation was carried out by first solving the linear elasticity problem on a simulated femoropopliteal bypass, in order to deform a reference geometry, adding artificial stenotic formations. Subsequently, blood flow was simulated by solving the Navier-Stokes equations: in this way, 8100 solutions were obtained for different values of the stenosis depth $A$, the stenosis diameter $r$, the flow rate $Q$ and the stenosis location $c$. Finally, the dataset was generated by taking B&W snapshots and adding Poisson random noise; each datapoint is made out of two images, taken $\approx 30°$ apart. An example is shown in fig. 1. Data augmentation was performed by taking snapshots from 5 different angles, yielding a total of 40500 images. The dataset was then divided into a training (60%), validation (20%) and test (20%) set .

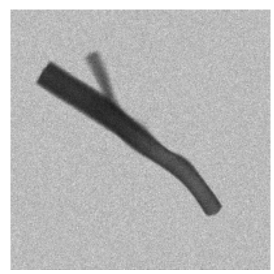

*Fig. 1: Example of datapoint*

### B. Parameters and labels selection

For each of the generated images 8 different quantities were stored: the parameters used to run the simulations ($A$, $r$, $Q$, $c$), velocity, pressure, wall shear stress (WSS) fields and fractional flow rate (FFR). For MTL, the stenosis depth ($A$), the stenosis location ($c$) and the FFR were selected as labels and used as auxiliary tasks for MI risk prediction.

### C. MI risk function design

The MI risk function was designed as follows:

$$\mathrm{MI} = \tanh\left(\sqrt{\frac{R^2}{2G}}e^A\right) \tag{1}$$

where $A$ is the diameter of the stenosis, $R$ a risk factor (used to help the model learn the stenosis location) and $G = \left(1 + \left|\frac{WSS}{WSS_0}\right|^2\right)^{-1}$ is a factor used to describe plaque growth models, with $WSS$ and $WSS_0$ denoting the mean wall shear stress with and without stenosis respectively (see [2]). Subsequently, machine learning models were trained to predict the MI risk from the generated angiography-like images.

### D. Neural network architecture for multitask learning

The neural network architecture used for the MI risk prediction based on the generated angiographic-like images is the classification model ResNet18 (provided by the PyTorch library Torchvision). The network is separated into shared layers, common to all the taks, and task-specific layers as sketched in fig. 2. The auxiliary tasks are hence used to improve both the prediction of the the main task and generalization capabilities.

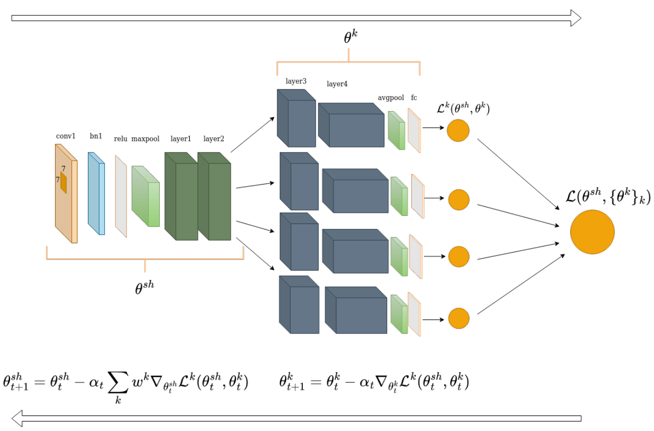


*Fig. 2: Neural network architecture*

### E. Objective and weights update design

For the given neural network architecture, it is possible to distinguish between shared parameters $\theta^{sh}$ and task-specific parameters $\theta^k$. Training aims at solving the following optimization problem:

$$\min_{\theta^{sh}, \theta^1, \dots, \theta^K} \sum_{k=1}^{K+1} w^k \mathcal{L}^k(\theta^{sh}, \theta^k) \tag{2}$$

where $w^k \geq 0, \sum_{k=1}^{K+1} w^k = 1$, and $K$ is the number of auxiliary tasks. In this work, stochastic gradient descent (SGD) is used to solve the problem 2 and the parameters are updated as follows:

$$\theta^{sh}_{t+1} = \theta^{sh}_t - \alpha_t \sum_k w^k \nabla_{\theta^{sh}_t} \mathcal{L}^k(\theta^{sh}_t, \theta^k_t) \tag{3}$$

$$\theta^k_{t+1} = \theta^k_t - \alpha_t w^k \nabla_{\theta^{sh}_t} \mathcal{L}^k(\theta^{sh}_t, \theta^k_t) \tag{4}$$

The weights $w^k$ are also updated at each iteration step. Two different strategies have been taken into account to update them. The first one, called Weighted Dynamical Average (WDA), aims at updating the weights assuming no hierarchical structure among tasks. Conversely, in the Online Learning for Auxiliary losses algorithm (OL-AUX), introduced by Lin et al. [1], a main task is individuated and jointly-trained auxiliary tasks are used to obtain a more effective representation of the data. For the considered problem, the prediction of MI risk is the main task. The OL-AUX model is characterized by two hyperparameters, $N$ and $\beta$. $N$ refers to the number of steps used to update of the weights (see [1] for details), while $\beta$ is the learning rate used in the gradient descent update of the weights.

### F. Hyperparameters selection

The objective was to select and optimize different hyperparameters. The two first hyperparameters considered are related to the neural network architecture: a suitable partition of the neural network trainable parameters ($\theta$) into a set of shared parameters, $\theta^{sh}$ and a set of task-specific parameters $\theta^k$, and the combination of auxiliary tasks ($A$, FFR, $c$) to use as inductive bias for MI risk prediction. Subsequently, the previously mentioned weighting strategies WDA and OL-AUX were compared. Although OL-AUX proved to be the method with the highest accuracy [3], both algorithms were taken into account in order to measure their effectiveness when varying the dataset size.

### G. Network architecture optimization

The hyperparameters related to ResNet18 architecture were optimized only considering WDA weighting strategy and using 20% of the dataset. This choice was made under the assumption that optimal hyperparameters related to the network architecture are independent from the dataset size and weighting strategy.

As shown in Fig. 2, the ResNet18 architecture features ten different sequential layers. However in this work we focused only on six combinations for time reasons. Finally, since there are three different auxiliary tasks (stenosis depth, stenosis location and FFR), theoretically there could exist seven different combinations. However, only four were selected, again for time reasons, under the assumption that, with respect to MI risk, FFR is the most relevant label and stenosis location is the least relevant.

### H. Effects of dataset size variation

After optimizing the hyperparameters of the network architecture, the performance of the two weighting strategies was assessed comparing the mean absolute test error of MI risk predictions for models trained on different ratios of the dataset. Furthermore, the parameter $N$ and the weight learning rate $\beta$ appearing in the OL-AUX weighting strategy were also tuned.

## III. Results

Table I summarizes the optimal hyperparameters of the neural network architecture considering the WDA approach and using 20% of the available dataset. In Fig. 3 the mean absolute validation error is shown for different intervals of the prediction of MI risk, using the WDA approach on the 20% of the dataset the four different combinations of auxiliary tasks. The error bars show the standard deviation of the mean absolute error. Fig. 4 shows the error distribution obtained for the STL case trained on the whole dataset, for comparison with the multitask learning results.

| Hyperparameter | Best hyperparameter |
|---|---|
| Task combination | MI-FFR-A |
| Common layers | Up to layer4 (see Fig. 2) |
| Specific layers | avgpool-fc |

*TABLE I: Best hyperparameters for MTL network architecture based on MI validation loss for WDA weighting strategy trained on 20% of the dataset.*

Fig. 6 shows the MI mean absolute test error for different weighting strategies against the dimension of the training dataset while in Fig. 7 the performance of the different models is displayed against the dimension of the training dataset.

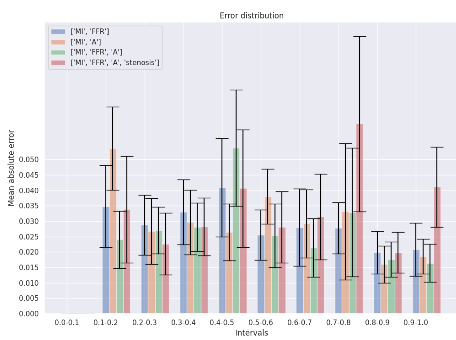


*Fig. 3: MI mean absolute validation error for different task combinations for WDA and for different intervals of MI risk.*

Fig. 5 shows how the absolute test error behaves, using the WDA approach, when varying the number of parameters in each task-specific layer, and the linear regression of the found results. We remark that the smallest error is obtained considering only few task-specific parameters, yielding a leaner network and significant computational savings. Finally, Fig. 8 shows MI mean absolute test error against the variation of hyperparameters that characterize OL-AUX models. In this case, both $N$ and $\beta$ tuning is displayed, along with the linear regression of both data. Note the two different $y$-axes for the two hyperparameters.

## IV. Discussion

First of all, we highlight that the optimal combination of auxiliary tasks is FFR-A. This suggests that the stenosis location $c$ is not that relevant when predicting MI risk. Additionally, we noted that using only FFR as auxiliary task allows to perform almost as well as FFR-A combination. This suggests that FFR label alone could be sufficient for a good prediction in a real-world scenario, where the value of $A$ could be missing or difficult to obtain. As predictable, STL has smaller errors since it is trained on a bigger dataset.

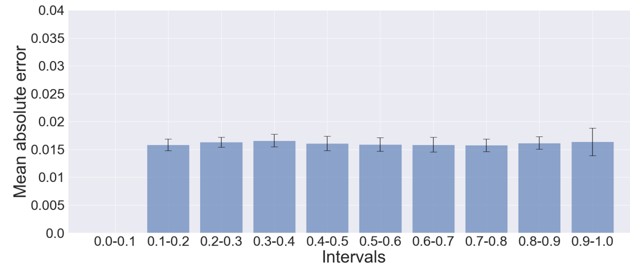


*Fig. 4: MI mean absolute validation error distribution for STL on the whole dataset (as in [3]).*

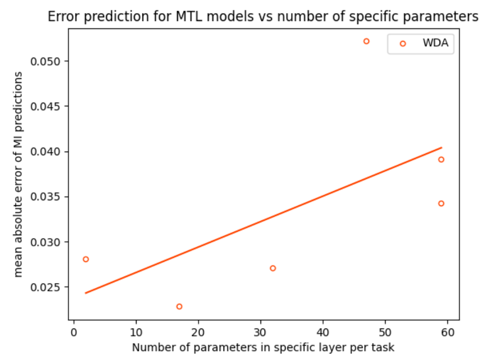


*Fig. 5: MI mean absolute test error against number of parameters in specific layers for WDA approach (with optimal task combination).*

However, it is remarkable that MTL almost matches STL accuracy, despite working on less data. Furthermore, it can be observed that the variability of the results is higher for multitask methods. Therefore it is possible to hypothesize that training models on smaller datasets makes the models less robust. For what concerns the usage of a reduced-size dataset, Fig. 6 clearly shows that the MTL approach is always more accurate than the STL approach when the dimension of the available dataset is restrained. However, the error increases dramatically when the dimension of the dataset is too small, going from $\approx 0.05$ to $0.15$ for the worst OL-AUX configuration ($N = 40$). Additionally, it is possible to see that OL-AUX models perform better than WDA for suitable values of $N$. Fig. 7 allows to further highlight not only how the error increases when the ratio is smaller than $10\%$, but also how robustness is affected because of the bigger variability. The same robustness issue is visible in Fig. 3: the error bars in some case highlight a standard deviation comparable to the average value of the mean absolute error.

The WDA approach seems to perform better when the number of parameters in the task-specific layers is small. Fig. 5 shows an increasing error for an increasing number of parameters. Such hyperparameter tuning should be performed with more different numbers of parameters, in order to be able to better capture the trend of the error. That said, it must be noted that specialization for the different auxiliary tasks is performed on only a small fraction of the layers in

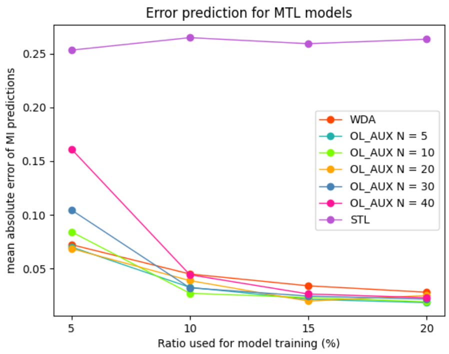


*Fig. 6: Mean absolute error for different weighting strategies against dimension of dataset. The hyperparameters used are those shown in Table I. For OL-AUX models, we set* $\beta = 0.5$.

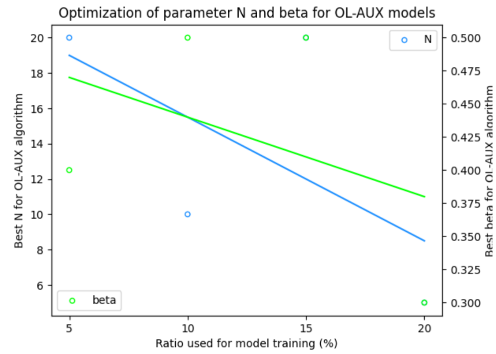


*Fig. 8: Optimal parameters N and* $\beta$ *for OL-AUX MTL method and different ratios. The optimal value of N for different ratios was assessed with a value of* $\beta = 0.5$ *and the optimal value of* $\beta$ *with a value of* $N = 10$.

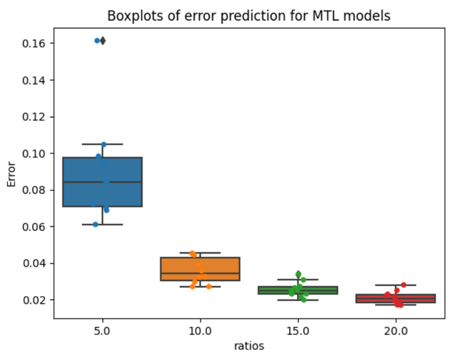


*Fig. 7: Boxplots for error prediction with respect to percentage of dataset used.*

the neural network.

Finally, for what concerns the variation of the optimal hyperparameters $\beta$ and $N$ for the OL-AUX weigthing strategy, it is possible to observe in Fig. 8 that their optimal values tend to decrease when training is performed on a bigger dataset. In [1] the authors point out that updating the weights every $N$ epochs rather than at every epoch allows to consider the objective of reducing the MI loss on the long term and to achieve less noisy gradients by averaging gradients of auxiliary tasks: this suggests that a bigger $N$ can compensate for a smaller training dataset for what concerns producing less noisy gradients.

## V. Summary

In general, the best MTL model proved to be OL-AUX. When reducing the size of the dataset, the mean absolute error obtained with properly tuned MTL algorithms remains decidedly lower than the one got with the STL algorithm trained on the same datasets, and is comparable with the error obtained through STL on the whole dataset. As expected, the error increases as the dataset size is reduced, while the robustness of the models decreases. Moreover, the choice of the auxiliary tasks FFR-A (or FFR in case of missing stenosis geometry characterization) allows to reduce the mean absolute error when the size of the available dataset is small. Optimization procedure highlighted that the use of MTL techniques should be carried out using only two task-specific layers, which leads to a smaller neural network architecture. Future work could focus on performing the architecture optimization using OL-AUX approach rather then WDA, to prove the correctness of the assumption that the optimal architecture does not depend on the chosen weighting strategy. Furthermore, hyperparameter tuning could be carried out using WDA, to observe whether such algorithm performs better than OL-AUX using specific hyperparameters combinations. In conclusion, in a real scenario we suggest the use of OL-AUX algorithm with FFR as auxiliary task and an increasing value of $N$ and $\beta$ for decreasing sizes of the available dataset.

## References


[1] X. Lin, H. Baweja, G. Kantor, and D. Held, "Adaptive auxiliary task weighting for reinforcement learning," in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds., vol. 32, Curran Associates, Inc., 2019. [Online]. Available: https://proceedings.neurips.cc/paper/2019/file/0e900ad84f63618452210ab8baae0218-Paper.pdf.

[2] T. Silva, W. Jäger, M. Neuss-Radu, and A. Sequeira, "Modeling of the early stage of atherosclerosis with emphasis on the regulation of the endothelial permeability," *Journal of Theoretical Biology*, vol. 496, p. 110 229, 2020, ISSN: 0022-5193. DOI: https://doi.org/10.1016/j.jtbi.2020.110229. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022519320300849.

[3] F. Betti, O. Y. Senouf, R. Tenderini, E. Abbé, A. Buffa, and S. Deparis, "Transfer learning and multitask learning for mi risk prediction from in-silico steady-state data," Unpublished, 2022.