AIDL Final Project: Self-supervised learning for medical imaging

Francesc Garcia, Pol Llopart, Sergio Rodriguez

April 25, 2022

Contents

1 Introduction

The main objective of this project is to apply a self-supervised model to classify medical images. Generating medical labeled datasets is very expensive, due to the needs of having field experts involved in the process. It is for this reason that self-supervised learning is a very promising approach to improve medical imaging tasks.

Our team explored several self-supervised models (e.g. SimClr, MoCo, BYOL) but we discovered Barlow Twins in a recent publication. Barlow Twins architecture is very resource efficient and we could not find any reference on the use of it on medical imaging on the literature.

Barlow Twins is a Self-supervised Deep Learning method (Fig. 1). Its objective function (Eq. 2) measures the cross-correlation matrix between the embeddings of two identical networks fed with distorted versions of a batch of samples. λ_{BT} is a positive constant trading off the importance of the diagonal and the off-diagonal components of the cross-correlation matrix in the loss.

It is for this reason that we have decided to apply the Barlow twins architecture to Chest X-ray images. The main goals for the project are:

- Pre-train a model using Barlow Twins architecture on Chest X-ray images.
- Use the self-supervised pre-trained model in several downstream tasks and compare its performance to untrained models.
- Observe whether Barlow Twins could be applied to medical images for a classification and evaluate its labels efficiency.
- Interpret how the model works: understand what parts of the image (visual patterns) are key to determine the predictions.
- Compare Barlow Twins' performance on medical images with state of the art, and similar self-supervised models.

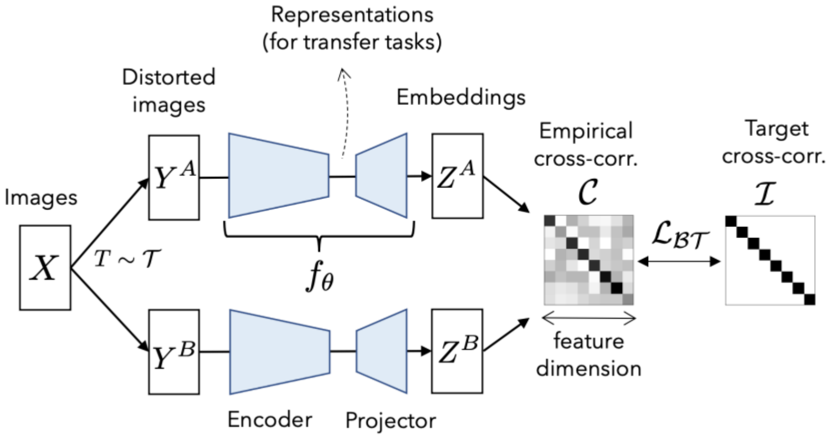


Figure 1: Representation of the Barlow Twins architecture.

$$\mathcal{L_{BT}} \triangleq \underbrace{\sum_i (1 - \mathcal{C}_{ii})^2}_{\text{invariance term}} + \lambda \underbrace{\sum_i \sum_{j \neq i} {\mathcal{C}_{ij}}^2}_{\text{redundancy reduction term}}$$

Figure 2: Barlow Twins Loss. λ_{BT} is a positive constant trading off the importance of the diagonal and the off-diagonal components of the cross-correlation matrix C in the loss.

2 Datasets

Since the scope of this project was wide open at the beginning, one of the first tasks performed was to select the dataset to be used to execute the project. During the research, many kind of datasets were found: MRI, X-RAY, CT, Echography and Image. Out of these types of medical imaging, most of them were localized on specific parts of the body: Chest, skin, extremity, bone, brain, full body and knee.

Given the amount of possibilities, we evaluated the following criteria: Number of samples, complexity of the imaging type, resources required.

Evaluating these criteria, we discarded all datasets found but two: COVID X-RAY [6] dataset and CheXpert X-Ray dataset [7] . These two datasets have similarities, as both are compilations of X-Ray images focused in the chest area of the patient.

Parameter	***COVID***	***CheXpert***
Number of samples	21,164	224,316
Resolution	299x299	389x320
Labels	4	14 (Multi-label annotation)
View	Frontal	Frontal and Lateral

Table 1: Comparison between COVID and CheXpert datasets.

The team decided to use the COVID dataset because the classification task is simpler and the resolution is smaller, which reduces the resources required to execute the tests.

The COVID dataset is a compilation of X-ray images selected from different sources and compiled into a single database on the platform 'Kaggle'. This dataset contains four classification labels:

Category	***Number of samples***
COVID	3615
Lung Opacity	6012
Normal	10192
Viral Pneumonia	1345
TOTAL	***21164***

Table 2: Number of samples by category in the COVID dataset in the kaggle platform.

This dataset also contains images in the form of masks that point out where the lungs are within the x-ray image. This mask will be discarded in this project.

During the course of this project, it was discussed the idea of training the barlow twins model with different data than the COVID dataset. The rationale behind it is that even if the labels will not be used during the self supervised part of the training, there may be some kind of bias in the COVID dataset. To avoid this kind of contamination, the team decided to use the CheXpert dataset as a source of huge amount of chest x-ray images only for the self-supervised part of the training.

3 Study of resources

Before starting tuning and training the self-supervised Barlow Twins model, we performed a study to find out the most optimal configuration, given the limited resources available. The three main variables that impact the training time and the needed hardware are:

- Resolution of the chest X-ray images
- Batch size
- Encoder/backbone model used in the Barlow Twins architecture
- Number of input channels (1 or 3 channels)

The GPU resources available by our team is a Nvidia RTX 2070 Super Graphics Card with 8GB of memory. Figure 3 displays the results from scanning the previously commented parameters with our available hardware.

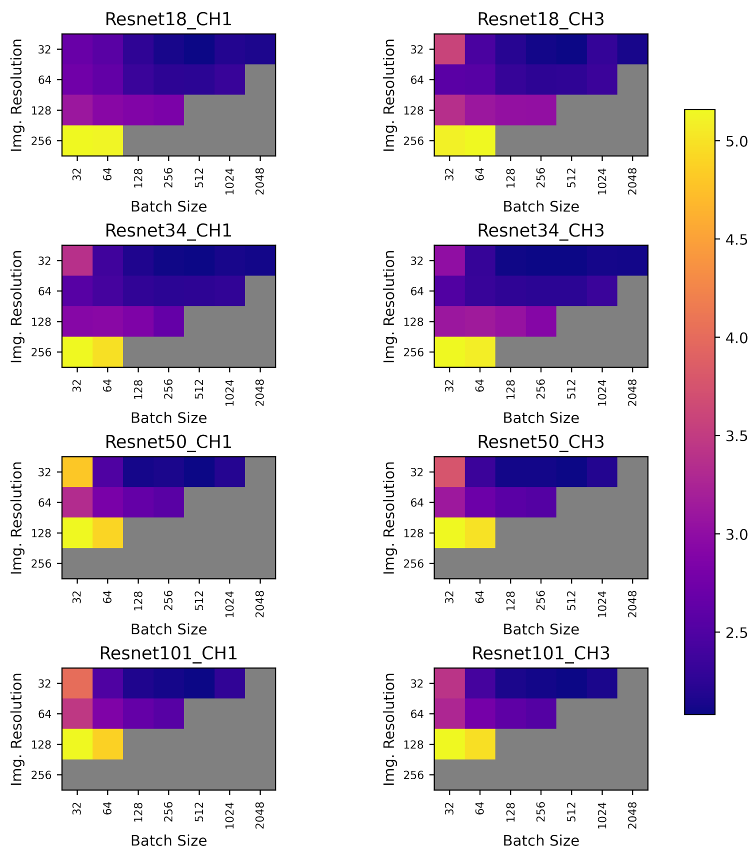


Figure 3: Set of color maps that picture the number of hours that it would take to train 100 epochs of the Covid datset with the Barlow Twins architecture, by changing model, image resolutions, batch size and number of channels. The gray areas correspond to GPU Out of Memory.

Apart from evaluating several Resnet networks (Resnet18, Resnet34, Resnet50, Resnet101) Efficientnet networks were also tested. Unfortunately, our GPU ran Out Of Memory with the lightest network from the Efficientnet family (Efficientnet b0). Figure 4 displays a plot of the memory footprint from Resnet and Efficientnet models.

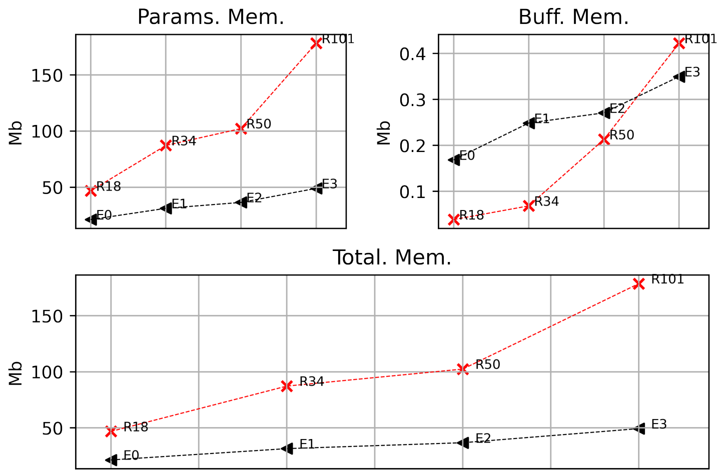


Figure 4: Memory footprint of several Resnet and Efficientnet models.

As described in this blog by Google, Efficientnet networks are indeed, more efficient in terms of trainable parameters. Despite taking less parameters, they are more complex in the number and type of layers that they use. It is for this reason why we think that we have not been able to load and train with Efficientnet networks.

Given the available hardware and training durations, we decided to use the following parameters for all the self-supervised related tasks:

- Resolution of the chest X-ray images: **224**
- Batch size: **128**
- Encoder/backbone model used in the Barlow Twins architecture: **Resnet-18**
- Number of input channels (1 or 3): **1 single channel**

Figure 5 represents the structure of the Resnet-18 network. The Encoder part of the network is the one used as backbone in the Barlow Twins architecture.

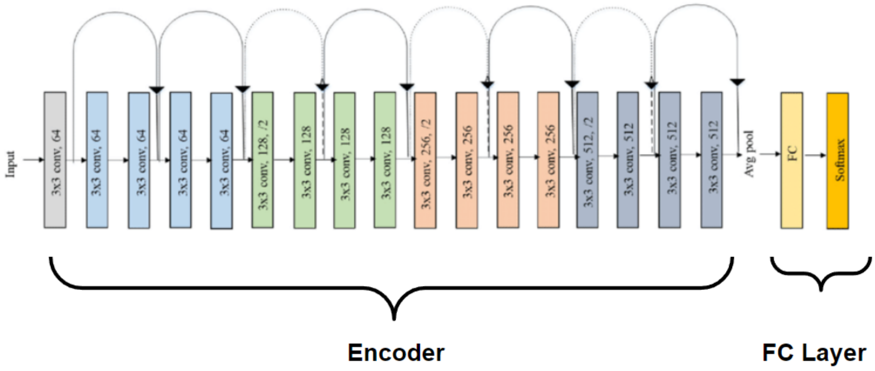


Figure 5: Visual representation. of the Resnet-18 architecture.

4 Self - Supervised experiments

Before training the Barlow Twins self-supervised model for many epochs on the Covid and CheXpert datsets, we have performed a set of hyperparameter searches. Due to the number of parameters to tune, and our limited resources, we have decided to split the hyperparameter search into smaller, independent tasks, which are described in the following sections 4.1, 4.2, 4.3. As a starting point for our experiments, the following parameters found in the literature (other self-supervised model applied to Chext X-ray images [1] [2], [3], [4]) have been used, :

- Optimizer: **Adam**
- Weight decay: **1e-5**
- Barlow Twins Lambda: **5e-5**
- Barlow Twins projector layers: **512 x 512**

For each one of the parameter searches performed (scan best lr, scan best transformations, scan best Barlow Twins hyperparameters), we have used the knowledge from previous experiments to perform the next ones. This is a sub-optimal greedy search but, due to the lack of resources, our team does not have the capabilities to train and test all the combinations independently.

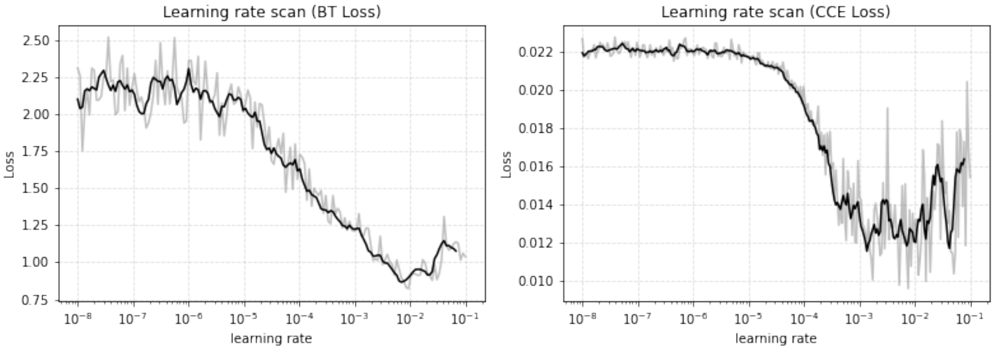
To train and evaluate one single instance from one experiment these steps are followed:

1. Train the self-supervised Barlow Twins model and save the backbone weights
2. Load the weights into a Resnet-18 model
3. Train the Resnet-18 model (full network or a linear separator) in a supervised way, for several epochs
4. Evaluate the accuracy of this model, to get feedback on the hyperparameters used

Bayesian Optimization has not been used in the parameters search, due to the complexity of the two-step (self-supervised and supervised) training process. The feedback signal is not instantaneous.

4.1 Learning rate scan & scheduler

Figure 14 displays two learning rate scans (Barlow Twins model on the left, and Resnet-18 with CCE Loss on the right) performed on one epoch of the Covid dataset, with the objective of finding the best learning rate starting point for the self-supervised and supervised training instances.



(a) Self-supervised training (b) Supervised training

Figure 6: Learning rate scan for self-supervised training, using Barlow Twins loss (Left) and for supervised classification, using Categorical Cross Entropy loss (Right).

From the results in Fig. 14, we decided to initialize the learning rate for the self-supervised experiments at $lr_{selfsup} = 2e - 3$ and the learning rate for the supervised training iterations at $lr_{sup} = 1e - 4$.

For the self-supervised training, three learning rate schedulers were tested: constant lr, linear lr, cosine annealing. As seen in table 3 cosine annealing turned out to get the lowest learning rate after 10 epochs. Setting a constant lr turned out to be almost as good as the cosine scheduler, but we aimed to train the self-supervised Barlow Twins model for many epochs, and it was expected that decreasing the learning rate would be more beneficial in the high number of epochs regime.

	Constant lr	***Linear lr***	***Cosine annealing***
Final lr (after 10 epochs)	32.8685	37.4496	32.0759

Table 3: Final Barlow Twins loss (after 10 epochs) when using different learning rate schedulers.

4.2 Exploring image transformations

It is well known that the set of transformations applied to the Images fed to self-supervised (contrastive-like) models are critical for their optimal performance.

All the transformations used by [2], [3], [4] papers (where they use self-supervised learning on X-ray chest images) have been evaluated, together with an extra transformation: RandomPerspective.

RandomResizedCrop and RandomHorizontalFlip transformations were used by all three papers [2], [3], [4]. We decided to use these two transformations as our base set of transformations (Fig. 7), applied to all instances, and keep adding extra transformations on top of them (Fig. 8).

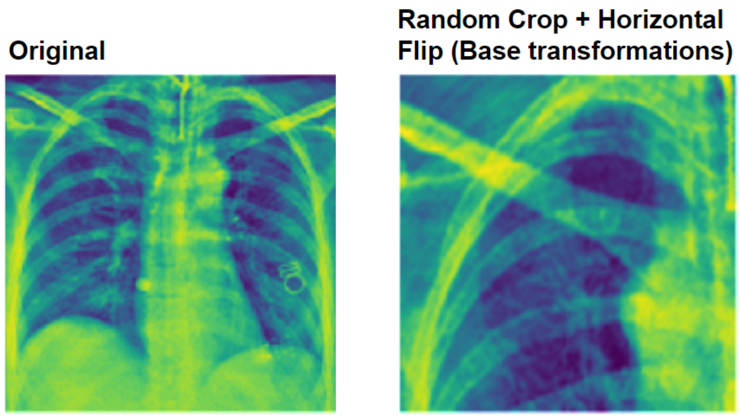


Figure 7: Original X-ray image (Left) and the same image with the base transformations applied (Right).

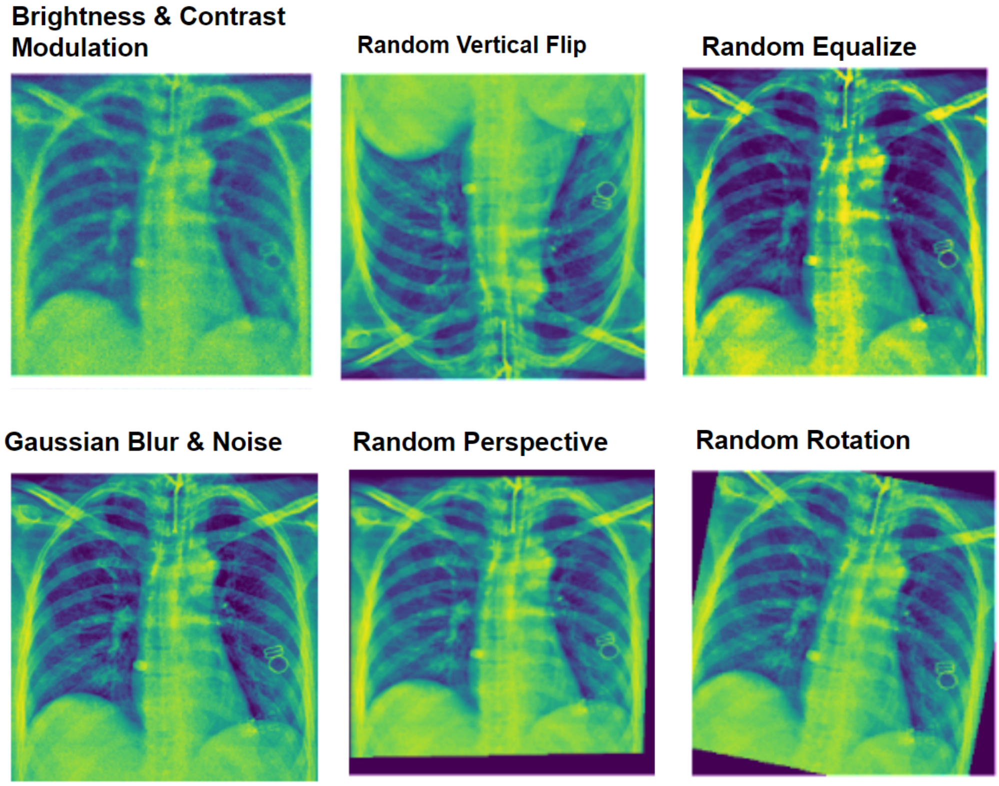


Figure 8: Extra transformations used to find out the best combination of transformations to be applied.

In the first scan, we applied the **Base transformations + one extra transformation**. First, the self-supervised model was trained for 20 epochs on the Covid dataset and saved the backbone weights. Then, the Resnet-18 backbone weights were loaded and trained for 5 epochs in two different supervised ways:

- Linear projector: Freezing all model weights but the last linear layer.
- Full network: Training all the parameters, the backbone and the last linear layer.

1st Transformations scan Validation Accuracy	*Linear projector*	*Full network*
Base transformations	0.6044	**0.8879**
Base transformations + GaussianBlur + GaussianNoise	0.6053	0.8833
Base transformations + RandomEqualize	**0.6136**	0.8686
Base transformations + RandomRotation	0.6030	0.8860
Base transformations + Brightness & Contrast modulation	0.5782	0.7786
Base transformations + RandomPerspective	0.6049	0.8686
Base transformations + RandomVerticalFlip	0.6131	0.8824

Table 4: Accuracy achieved by each set of transformations after pre-training the model for 20 epochs in a self-supervised approach, and training the model for 5 epochs in a supervised approach, with a linear separator and training the full network.

From the results in table 4, we decided to discard the **Brightness Contrast modulation** transformation, since it was performing very poorly. The Base Transformation alone is among the best performing combinations. Despite this, we have decided to sort all the transformations by their accuracy and to perform a second scan where we keep adding one extra transformation at a time (results displayed in table 5).

2nd Transformations scan Validation Accuracy	***Linear projector***	***Full network***
Base transformations + RandomVerticalFlip + RandomEqualize	**0.6646**	**0.8952**
Base transformations + RandomVerticalFlip + RandomEqualize + RandomPerspective	0.6619	0.8934
Base transformations + RandomVerticalFlip + RandomEqualize + RandomPerspective + GaussianBlur + GaussianNoise	0.6596	0.8920
Base transformations + RandomVerticalFlip + RandomEqualize + RandomPerspective + GaussianBlur + GaussianNoise + RandomRotation	0.6586	0.8879

Table 5: Accuracy achieved by each set of transformations after pre-training the model for 20 epochs in a self-supervised approach, and training the model for 5 epochs in a supervised approach, with a linear separator and training the full network.

One important learning is that the addition of more than one extra transformation leads to a big increase in Accuracy (particularly the combination of VerticalFlip and RandomEqualize). Eventually, we have decided to use all the transformations together except the RandomRotation (Fig. 9), since it is a particular case of the RandomPerspective , hence redundant. We expected that when training for many epochs, the model would be able to take advantage of the extra set of transformations.

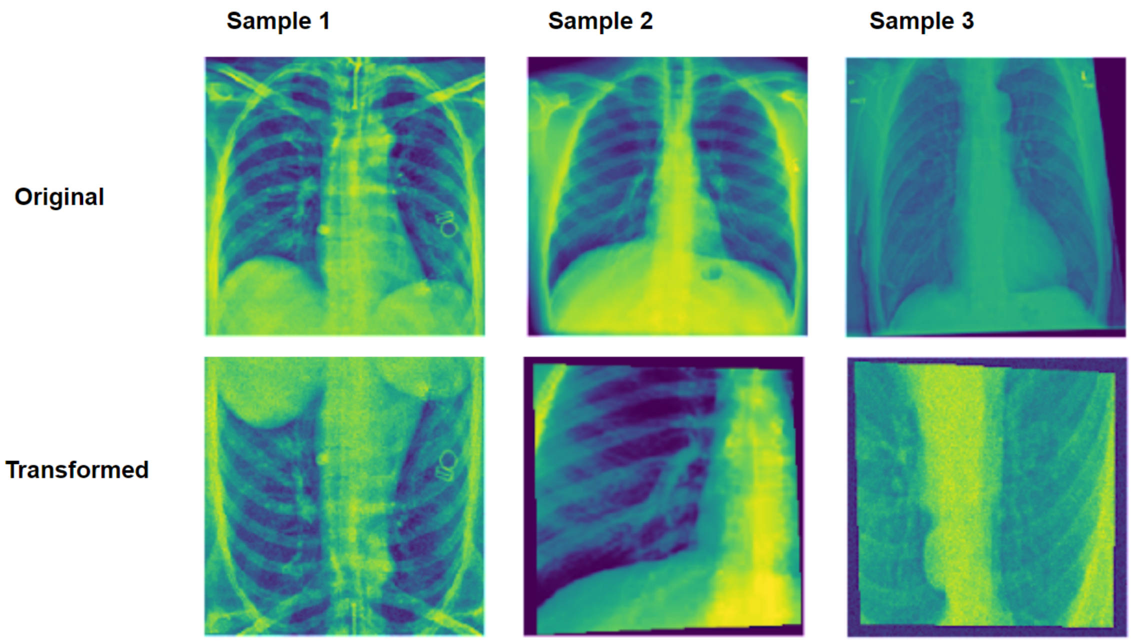


Figure 9: Three examples of images under the final set of transformations selected to train our self-supervised model.

4.3 Hyperparameter tuning

Once selected the set of transformations used for the self-supervised task, we proceeded to explore the best projector (for the Barlow Twins architecture) and the best Barlow Twins Loss Lambda λ_{BT}.

The parameters used up to this moment:

- Barlow Twins Lambda: **5e-5 (defined in the original paper** [1]**)**
- Barlow Twins projector layers: **512 x 512 (two layers with 512 units)**

Projector FC dimensionality	*Linear projector*	*Full network*
512	**0.6908**	**0.8975**
1024	0.6848	0.8718
2048	0.6320	0.8263
512-512	0.5888	0.8548
512-1024	0.6550	0.8814
512-2048	0.6481	0.8676

Table 6: Accuracy achieved by changing the model's projector unit after pre-training the model for 20 epochs in a self-supervised approach, and training the model for 5 epochs in a supervised approach, with a linear separator and training the full network.

Table 6 shows that the best accuracy is achieved when using a single dense layer with 512 units as a linear projector. The original Barlow Twins paper [1], states that the bigger the projection head the better. We have not been able to reproduce this result, we expect that this is due to the small amount of training epochs (20 epochs) and the small size of the Covid dataset ($\sim$ 20k images) are not enough to take advantage of projector heads with bigger size.

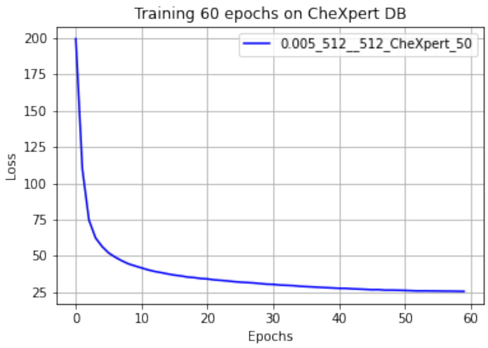
Barlow Twins Lambda λ_{BT}	*Linear projector*	*Full network*
1e-4	0.6311	0.8755
5e-4	0.6398	0.8814
1e-3	0.6476	0.8796
5e-3	0.6550	0.8548
1e-2	**0.6632**	**0.8851**
5e-2	0.5667	0.7835
1e-1	0.4938	0.7546

Table 7: Accuracy achieved by changing the Barlow Twins Lambda λ_{BT} after pre-training the model for 20 epochs in a self-supervised approach, and training the model for 5 epochs in a supervised approach, with a linear separator and training the full network.

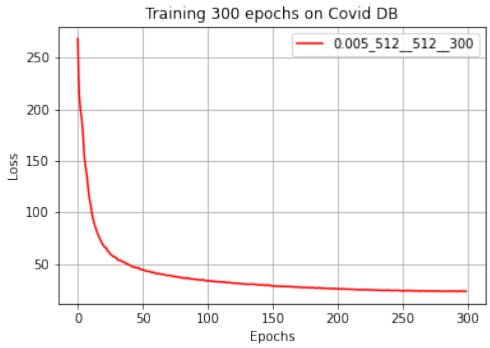
In table 7 one can see that Accuracy decreases substantially for values of the Barlow Lambda greater than $5e-2$. Despite finding the best accuracy at $\lambda_{BT} = 1e-2$, we decided to stick to the original paper's $\lambda_{BT} = 5e-3$, since there was not a clear trend in the range $1e-2 \geq \lambda_{BT} \geq 1e-4$ when training the full network.

4.4 Final training

With the hyperparameters found in sections 3, 4.1, 4.3 and the transformations 4.2, we have trained the Barlow Twins model for 300 epochs on the Covid dataset ($\sim 20k$ images) and for 60 epochs on the CheXpert dataset ($\sim 190k$ images).



(a) CheXpert dataset



(b) Covid dataset

Figure 10: Loss as a function of number of epochs, from the training of the Barlow Twins model on the CheXper dataset (Left) and the Covid dataset (Right).

It took approx. 15 hours to train 300 epochs on the Covid dataset and approx. 32 hours to train 60 epochs on the CheXpert dataset. We trained more time on the CheXpert dataset since we found out that we had more time than expected when planning the resources to perform all trainings.

5 Supervised experiments

5.1 Hyperparameter tuning

For the downstream tasks, once trained the self-supervised models, we performed a hyperparameter scan by running 55 trials with different configurations of parameters. The best configuration we found, which we have used on the next sections are:

- Learning rate: **0.00153**
- Batch Size: **96**
- Number of epochs: **15**
- Weight decay: **9.870e-6**
- Augmentations: **Horizontal flip with p=0.5**

5.2 Scanning % of labels

Once the backbone is trained using self supervision techniques, it is possible to evaluate the improvement of the neural network over other networks. In the best case scenario, the upstream task has been useful for the network to learn to recognize patterns about the information fed (In this case X-rays from the chest area).

To evaluate this behavior, in the proposed experiment four Resnet-18[5] are trained (supervised) in parallel. For the first Resnet-18, we load the weights and biases that we have obtained from self-supervised training using the COVID dataset. On the second Resnet-18 we load weights and biases that we have obtained from self-supervised training using the CheXpert dataset. The third one is a non-pretrained resnet18 and the last one is initialized to the pre-trained values of the resnet18 model (Trained with Imagenet).

Since the main goal of self supervised techniques is to minimize the 'labelling' tasks, we train the networks using a variable number of samples: 1%, 5%, 10%, 15%, 20%, 25%, 50% and 100% of the available (labelled) samples.

The expected results are that the pre-trained network is capable of classifying the X-ray images with a smaller train set.

Results

The results of the experiment are summarized in the table 8. With an extremely low level of samples (1% = 169 samples), neither network can predict accurately the labels. However, after augmenting slightly the number of samples (10% of the total samples) used for the training, it is possible to reach a good level of accuracy with the networks that are pre-trained using self supervised learning. On the other hand, the not pre-trained network and the default pre-trained networks, even if they increment the accuracy, they don't reach state of the art results.

There is minimum variation between using the same dataset or different one as a source for the upstream task. This reinforces the idea that self supervised learning is ideal for the medical field since we can reach state of the art results by training on a small labelled dataset while we pre-train the networks using unlabelled datasets.

Samples	***Barlow Twins*** (CheXpert)	***Barlow Twins*** (COVID)	***Not Pre-Trained***	***PreTrained***
1%	0.830	0.865	0.770	0.738
5%	0.895	0.925	0.815	0.883
10%	0.935	0.930	0.835	0.884
15%	0.930	0.945	0.860	0.914
20%	0.915	0.955	0.890	0.918
25%	0.940	**0.960**	0.900	0.926
50%	0.929	0.939	0.902	0.931
100%	**0.947**	0.946	**0.922**	**0.941**

Table 8: Incremental scanning of samples

Note: The results of 50% and 100% of samples were only trained for 15 epochs due to the lack of time and resources. For this reason, the best results for the Barlow Twins (COVID) are found on the test run with 25% of the labels.

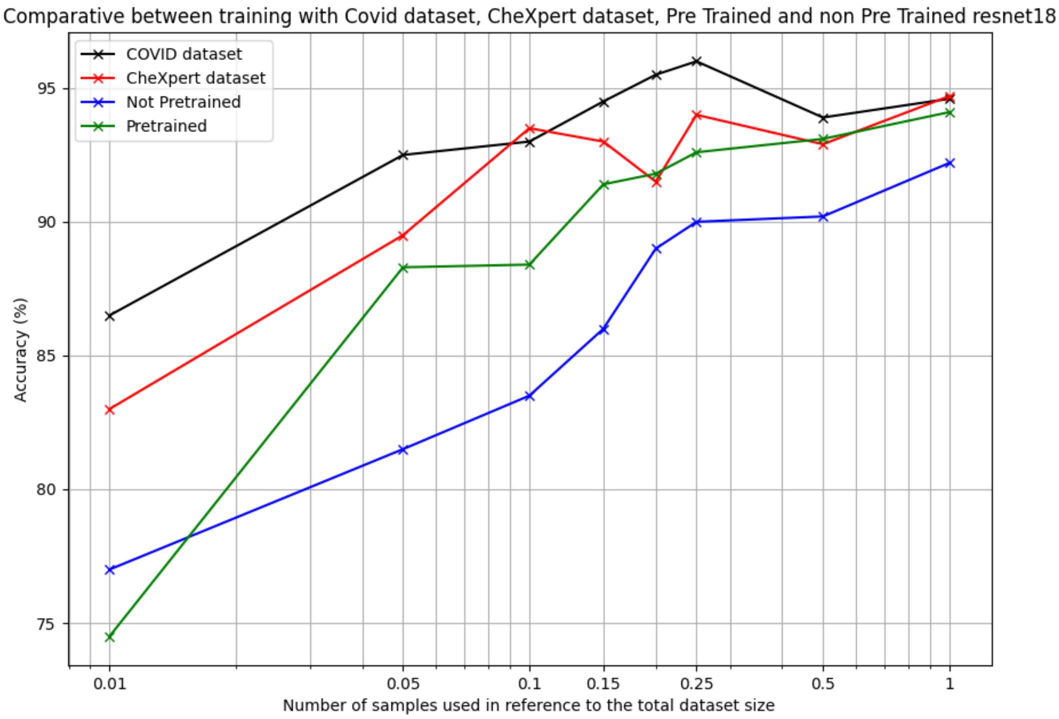


Figure 11: Incremental scanning of samples.

Note: The results of 50% and 100% of samples were only trained for 15 epochs due to the lack of time and resources. For this reason, the best results are found on the test run with 25% of the labels.

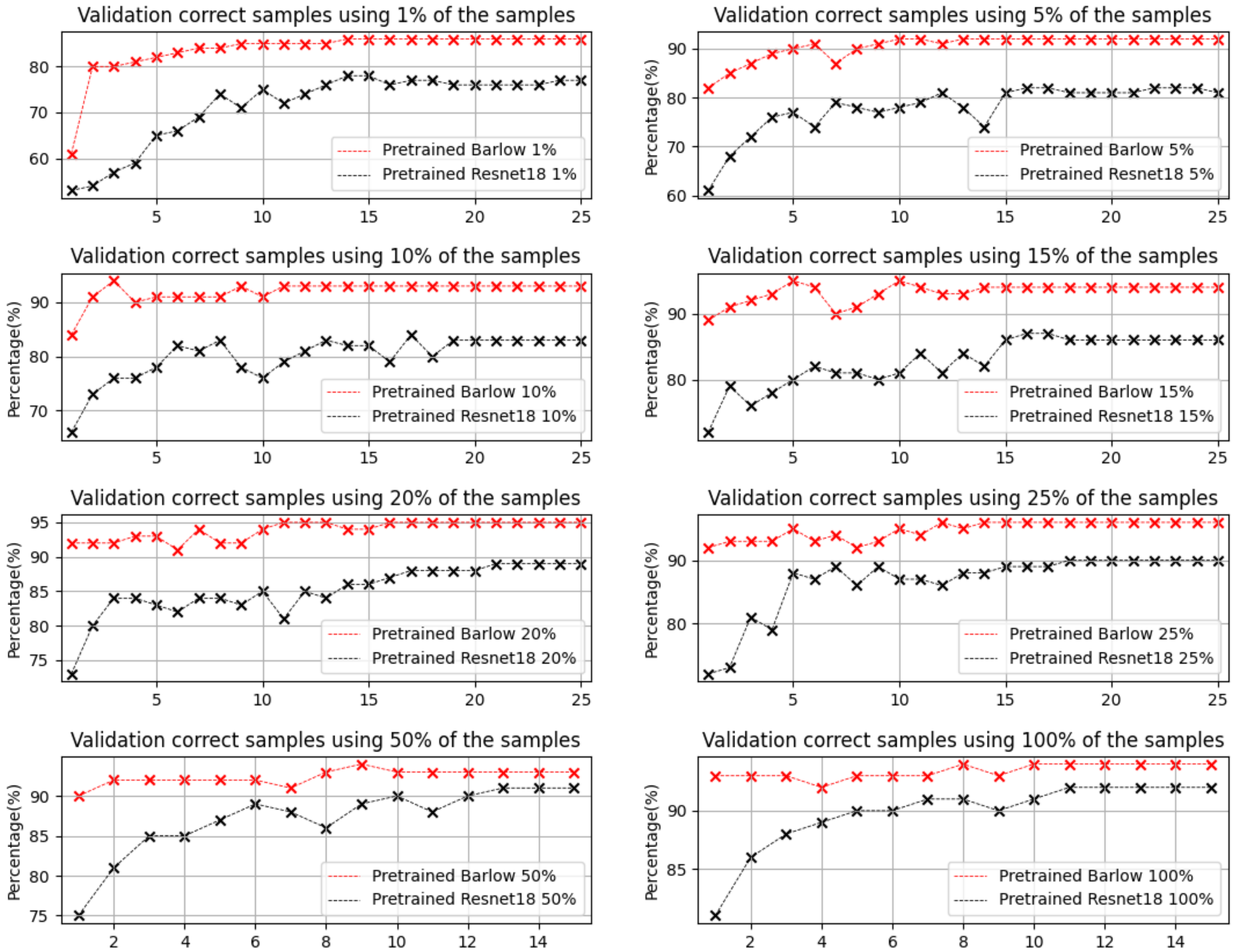


Figure 12: Evolution of training validation while executing supervised training using a network pre-trained with the COVID dataset (Comparison with non pre-trained network)

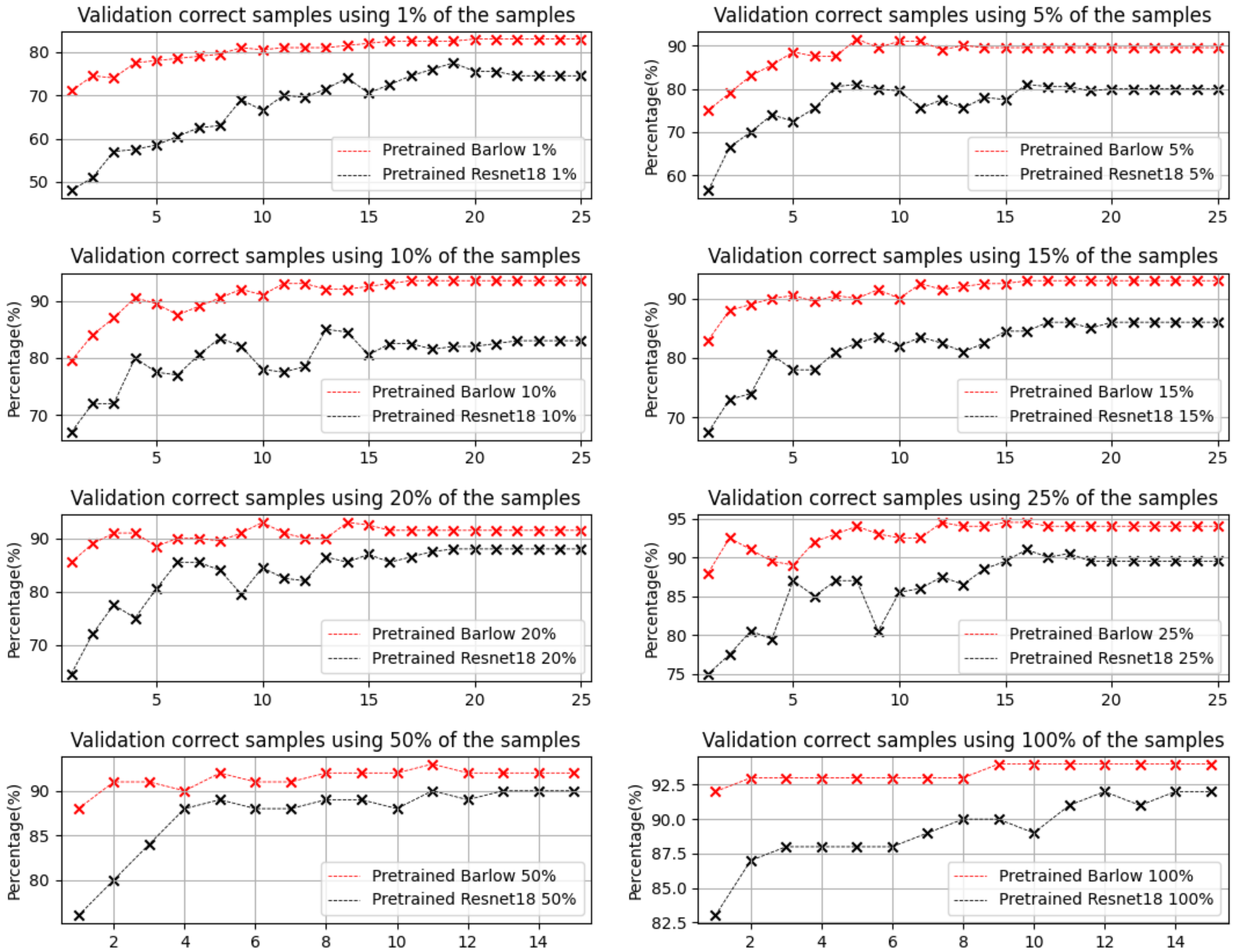
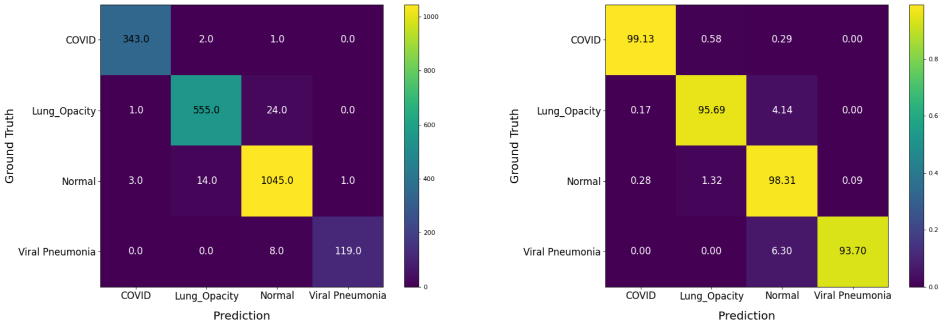


Figure 13: Evolution of training validation while executing supervised training using a network pre-trained with the CheXpert dataset (Comparison with non pre-trained network)



(a) Absolute results

(b) Normalized results (%)

Figure 14: Confusion matrix of the best accuracy model, using absolute numbers (Left) and normalized percentages to each category's amount of samples (Right).

5.3 Freezing weights: a transfer learning scenario

In addition, we thought of an scenario where the self-supervised pretrained model with Barlow Twins was used to make predictions using different datasets, as in a case of transfer learning. We considered that it would be an interesting experiment to freeze different layers of the model architecture in order to identify the convolutional layer blocks that have the most important parameters to train the network. To do so, we decided to build five test models by freezing the pretrained self-supervised network at different parts; starting form the top (Fully Connected layer projector head) and additionally freezing every convolutional layer block until the first one (Layer-1 of the ResNet architecture). Then, we trained these five models to analyze how they performed. For this task, we loaded the weights of our self-supervised trained model with the CheXpert dataset to the frozen models, and we used the COVID X-RAY dataset for their training.

By observing the results shown in Figure 15, we can conclude that the performance is different depending on the frozen part. We noticed that the models with frozen architectures that do not freeze the first three convolutional layer blocks were performing better than the ones that did so. Therefore, starting from the bottom of the model architecture we suggest to freeze at most until the 2nd convolutional block, to archive accuracy values higher than 90% with less than a 15 epochs training.

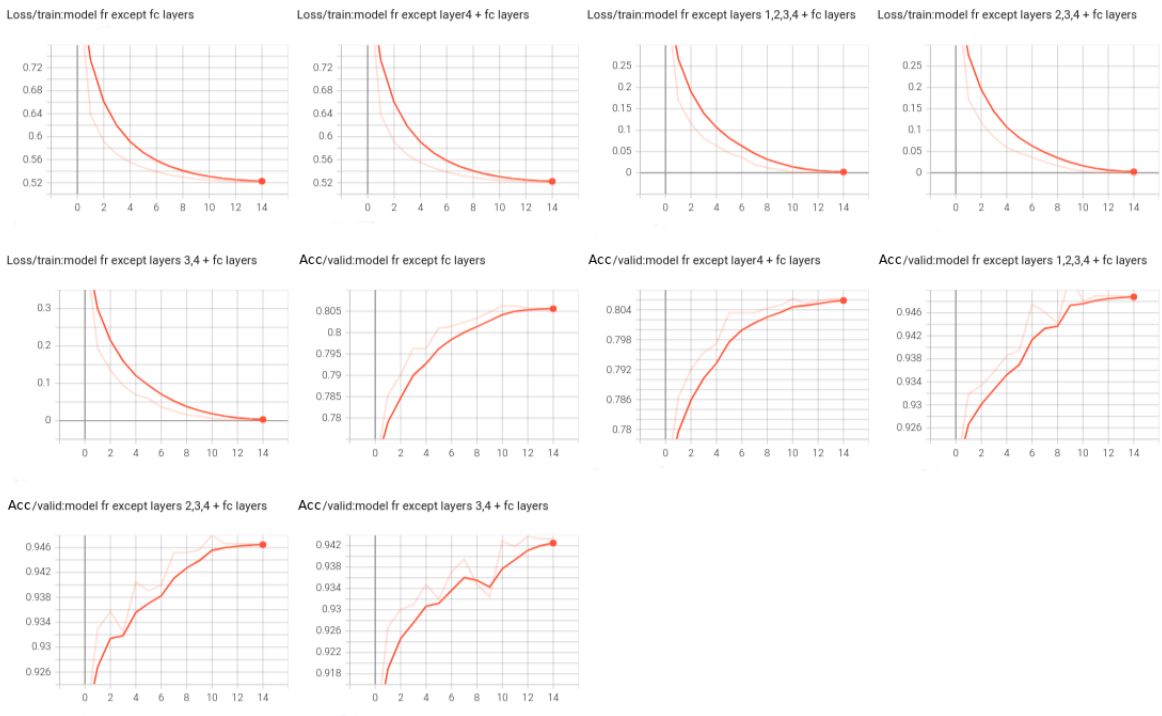


Figure 15: Loss and Accuracy values of the frozen test models at every epoch (15 epochs). The loss values depicted correspond to the loss values for the training step, and the accuracy values correspond to the ones obtained at the validation step, both metrics obtained at each epoch.

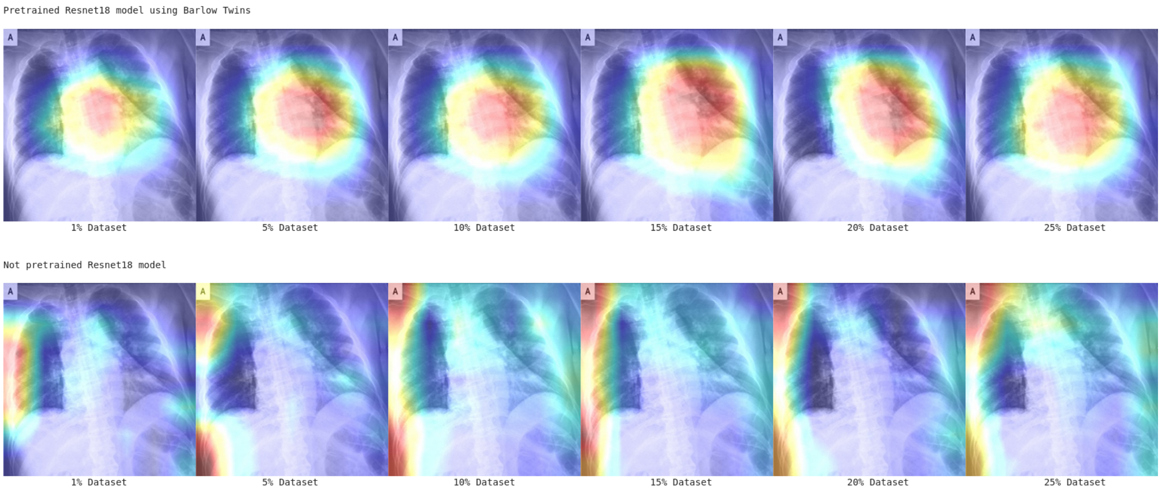
6 Model interpretability: Grad-CAM

We decided to apply Grad-CAM as the interpretability method for our model and results. By looking at figure 16 one can observe images corresponding to samples coming from patients with different diagnoses (COVID, Lung Opacity, Viral Pneumonia and Normal (healthy)). To help interpreting our model's performance we decided to run a test to spot the most significant regions of the images that are used by the model to make a prediction. To contrast this results we applied this test both using our pretrained Resnet18 model with Barlow Twins and non-pretrained Resnet-18 model, the two of them trained only for 15 epochs for the downstream task using the COVID X-RAY dataset.

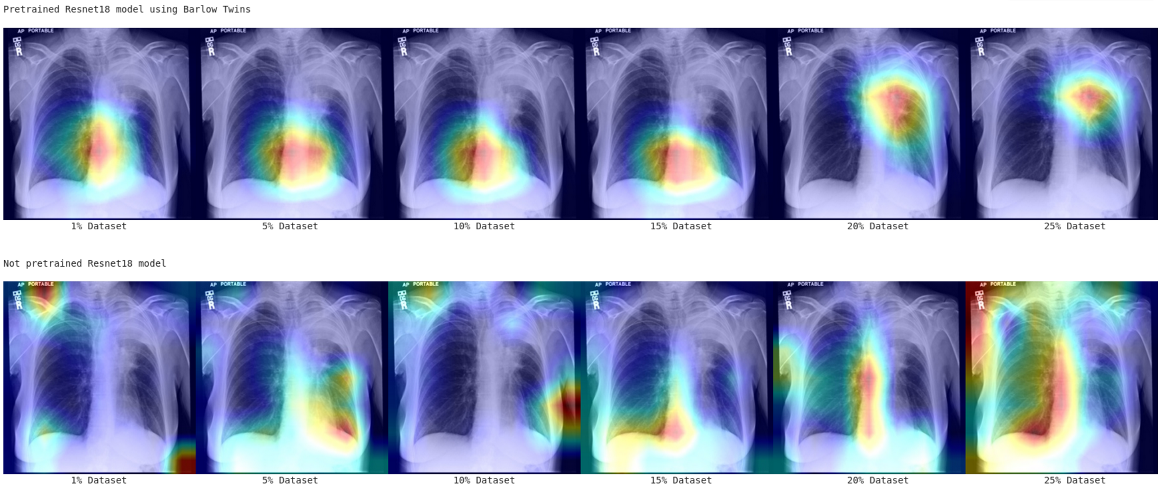
Interestingly, after performing the test, we observe that predictions made by the pretrained model focus on an image region close to the center of the image, including the area in which lungs are found. It is remarkable that as the percentage of samples used to train the model are increased, the model focuses on patterns that appear on the most affected lung, discriminating every other possible diagnose (predicted label), as we can observe in the image corresponding to Lung Opacity.

On the other hand, by observing the performance of the non-pretrained model we can conclude that it does not seem to be focusing on any significant visual pattern in order to discriminate the labels. It may be interesting to notice that, although the models trained with 1% of the samples seem not to focus on any significant part, when using 25% of the samples it seems to look for patterns closer to the center of the image.

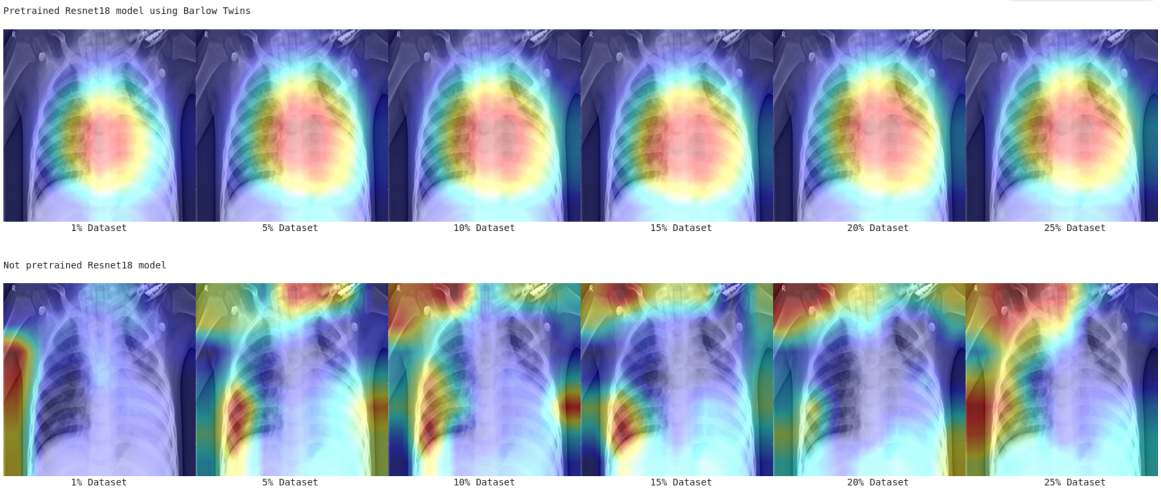
COVID



Lung Opacity



Viral Pneumonia



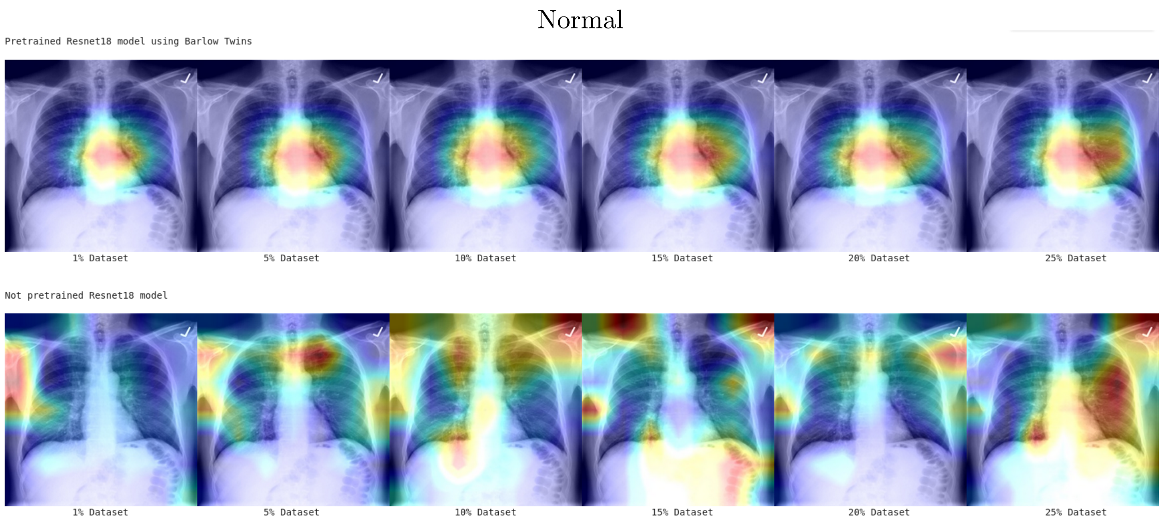


Figure 16: Grad-CAM results of an X-ray image corresponding to patients diagnosed with COVID, Lung Opacity, Viral Pneumonia and a healthy patient ('Normal' label). At the top we observe the results of forwarding the image to the model pre-trained with Barlow Twins and at the bottom we see the case of doing so with the non-pretrained Resnet-18 model. The heat maps projected onto the original image are depicted in a blue-red color scale (blue indicating low interesting regions and red the high ones).

7 Benchmarking

The following table 9 contains a list of papers found that have tackled the Covid detection problem using the Covid dataset found on Kaggle. This list is a compilation of found papers, not necessarily complete but accurate enough for this work.

Consulted references: [8][9][10][11][12][13][14][15][16]

Study	**Precision**
Automatic detection of coronavirus disease [...]	96%-99%
CovXNet	89.6% - 97.4%
Lightweight Neural Network for COVID-19 Detection	95.8%
Covid-19: automatic detection[...] transfer learning	95%
Multi-Channel Transfer Learning [...]	94%
COVID-Net	93.3%
Using X-ray images and deep learning ...	92.18%
CoroNet	89.6%
Automated detection of COVID-19 [...]	87.0%

Table 9: State of the art using COVID-19 dataset

Study	**Precision**
DeepAUC-v1 ensemble	93%
Hierarchical-Learning-V1 (ensemble)	93%
Conditional-Training-LSR ensemble	92.9%
Hierarchical-Learning-V4 (ensemble)	92.9%
YWW(ensemble)	92.9%
Conditional-Training-LSR-V1 ensemble	92.9%
Hierarchical-Learning-V0 (ensemble)	92.9%
Multi-Stage-Learning-CNN-V3 (ensemble)	92.8%

Table 10: State of the art using CheXpert dataset

The next table 10 contains a list of papers that have tackled the Covid detection problem using the CheXpert dataset. It is only kept as a reference, since in this project, the CheXpert dataset will not be used for classifying its multiple classes.

This information is extracted from the CheXpert competition front page [17]

8 Conclusions

The main task of this project was to evaluate the viability of the Barlow Twins architecture for the task of self-supervised learning on the medical imaging field. After performing many studies on the resource requirements of this architecture, we have concluded that this architecture is well suited for achieving state of the art results without major resources.

In the self supervised task, it has been observed that when adding more than one extra image transformations, the accuracy of the linear projector task improves substantially.

Contrary to the conclusions of the Barlow Twins paper [1], we have found that the best architecture for the projector head is a single 512 neuron linear layer. Our hypothesis is that the simplest architecture has given best results because of the small size of the dataset and the low number of training epochs.

On the supervised part of the training, we have observed that the self supervised pre-trained models do not improve drastically the accuracy of the classification problem when we use all the labels available on the dataset (∼16k). However, when the number of samples used to train is scarce (∼1.6k-2k) the pre-trained models outperform the not pre-trained model and the ImageNet pre-trained models.

Additionally, we have observed that the accuracy for the pre-trained models (one with Covid dataset and the other one with the CheXpert dataset) does not change significantly.

Freezing different layers of the model architecture we observed that the performance of the model is affected remarkably. The best results are obtained when freezing only the initial layers (Layer-1 and Layer-2 of the Resnet model). If additional layer are frozen, we have noticed that in order to improve the accuracy of the model, more training epochs are required.

Observing the results of Grad-CAM, in the pre-trained models we can observe the significant patterns in the area of the lungs even at lower percentage of samples. On the other hand, the not pre-trained models do not focus on the lungs until the number of seen samples is high enough.

As a final conclusion, we can state that the Barlow Twins model is suitable for the medical imaging field. With fewer resources we have been able to reach results close to state of the art. We find that this type of analysis can be suitable for the task of diagnosis based on medical imaging.

References

[1] Zbontar, Jure, et al. "Barlow twins: Self-supervised learning via redundancy reduction." International Conference on Machine Learning. PMLR, 2021.

[2] Sriram, Anuroop, et al. "Covid-19 prognosis via self-supervised representation learning and multi-image prediction." arXiv preprint arXiv:2101.04909 (2021).

[3] Chen, Xiaocong, et al. "Momentum contrastive learning for few-shot COVID-19 diagnosis from chest CT images." Pattern recognition 113 (2021): 107826.

[4] Azizi, Shekoofeh, et al. "Big self-supervised models advance medical image classification." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun. "Deep Residual Learning for Image Recognition". Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778

[6] Muhammad E. H. Chowdhury et al. (2022). COVID-19 Radiography Database. Retrieved March 2022 from https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database.

[7] Irvin, Jeremy and Rajpurkar, Pranav and Ko et al. (2019). CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison. Retrieved from https://arxiv.org/abs/1901.07031

[8] Sanida, T.; Sideris, A.;Tsiktsiris, D.; Dasygenis, M. Lightweight Neural Network for COVID-19 Detection from Chest X-ray Images Implemented on an Embedded System. Technologies 2022, 10, 37. https://doi.org/10.3390/

[9] T. Mahmud, M.A. Rahman, S.A. Fattah. CovXNet: a multi-dilation convolutional neural network for automatic COVID-19 and other pneumonia detection from chest X-ray images with transferable multi-receptive feature optimization Comput. Biol. Med., 122 (2020), p. 103869

[10] Wang, L., Lin, Z.Q. Wong, A. COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. Sci Rep 10, 19549 (2020). https://doi.org/10.1038/s41598-020-76550-z

[11] T. Ozturk, et al. Automated detection of COVID-19 cases using deep neural networks with X-ray images Comput. Biol. Med., 121 (2020), p. 103792

[12] Apostolopoulos, I.D., Mpesiana, T.A. Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. Phys Eng Sci Med 43, 635–640 (2020). https://doi.org/10.1007/s13246-020-00865-4

[13] Misra, S.; Jeon, S.; Lee, S.; Managuli, R.; Jang, I.-S.; Kim, C. Multi-Channel Transfer Learning of Chest X-ray Images for Screening of COVID-19. Electronics 2020, 9, 1388. https://doi.org/10.3390/electronics9091388

[14] Khalid El Asnaoui and Youness Chawki. Using X-ray images and deep learning for automated detection of coronavirus disease. Taylor Francis 2021. https://doi.org/10.1080/07391102.2020.1767212

[15] Narin, A., Kaya, C. Pamuk, Z. Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks. Pattern Anal Applic 24, 1207–1220 (2021). https://doi.org/10.1007/s10044-021-00984-y

[16] Asif Iqbal Khan, Junaid Latief Shah, Mohammad Mudasir Bhat, CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images, Computer Methods and Programs in Biomedicine, Volume 196, 2020, 105581, ISSN 0169-2607, https://doi.org/10.1016/j.cmpb.2020.105581.

[17] Stanford ML Group. A Large Chest X-Ray Dataset And Competition. Leaderboard consulted on April 2022 on https://stanfordmlgroup.github.io/competitions/chexpert/

9 Appendix

9.1 Self-supervised task: transformations scan

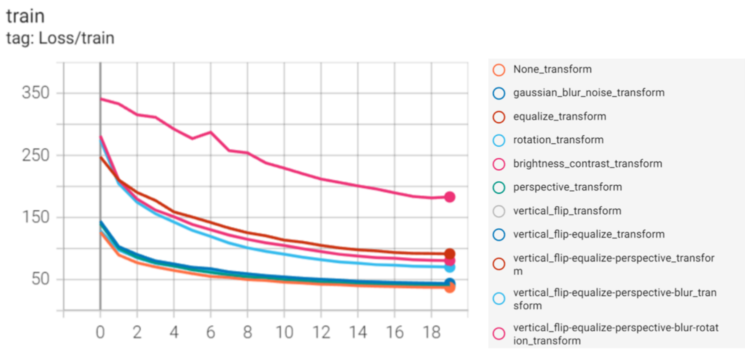


Figure 17: Losses from the training of the self-supervised models on several combinations of transformations.

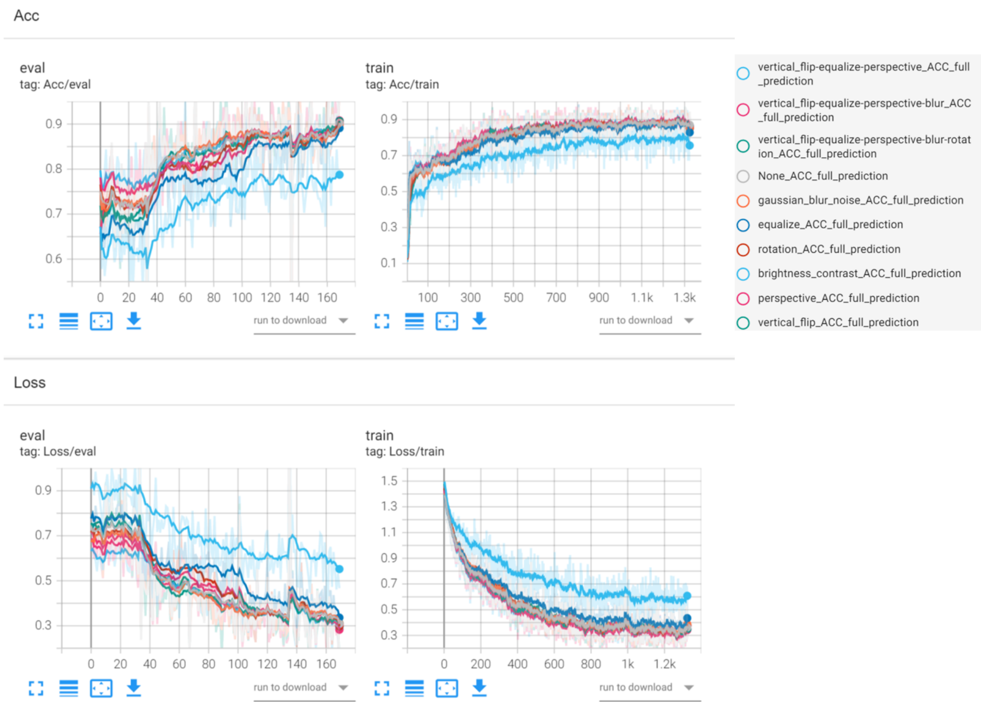


Figure 18: Losses and Accuracies from the training of the supervised, pre-trained models on several combinations of transformations, on the Covid dataset.

9.2 Self-supervised tasks: hyperparameters scan

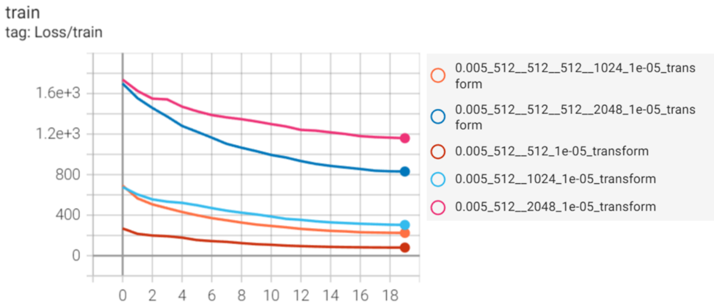


Figure 19: Losses from the training of the self-supervised models on several combinations of projector heads.

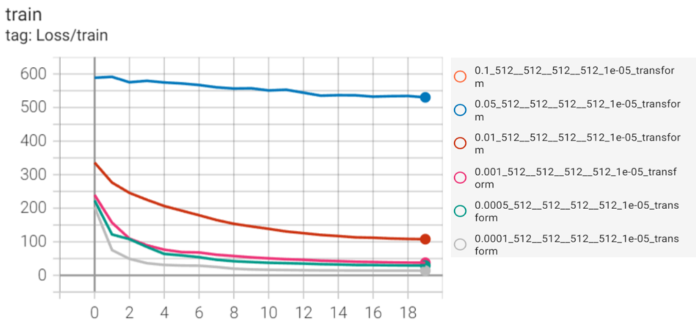


Figure 20: Losses from the training of the self-supervised models on several combinations of λ_{BT}.

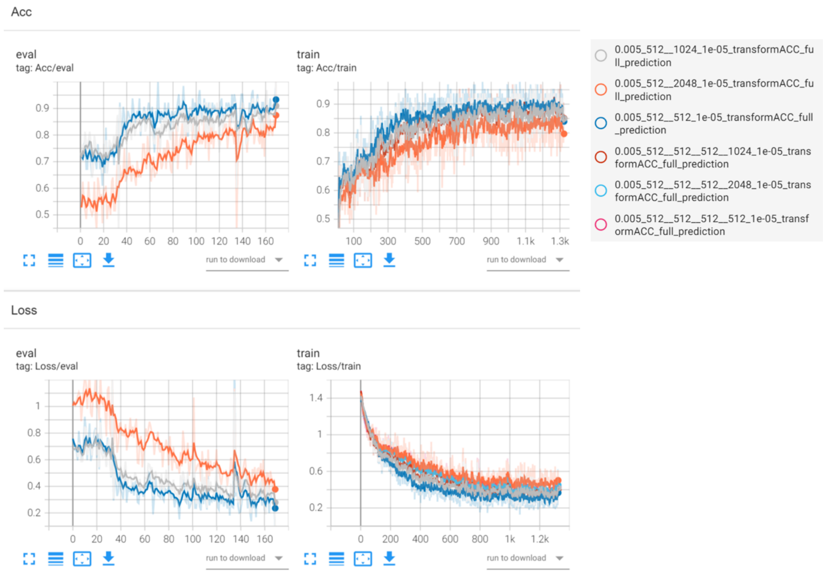


Figure 21: Losses and Accuracies from the training of the supervised, pre-trained models on several combinations of projector heads, on the Covid dataset.

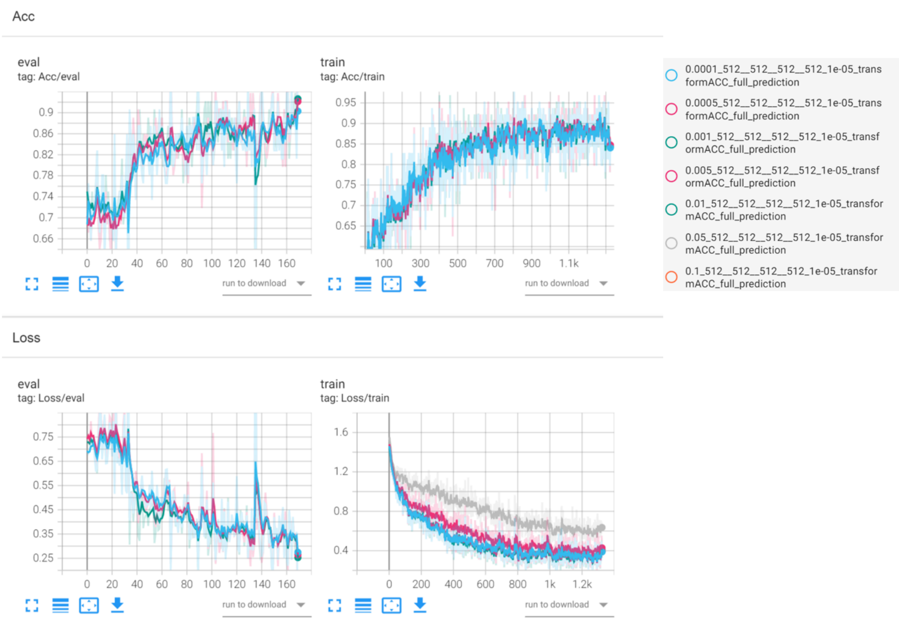


Figure 22: Losses and Accuracies from the training of the supervised, pre-trained models on several combinations of λ_{BT}, on the Covid dataset.