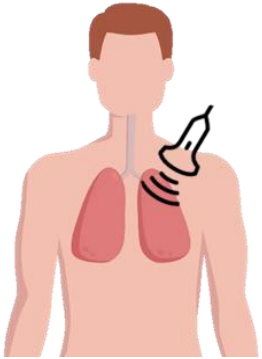
Systematic analysis of neural networks performance and generalization capabilities with application to the automatic assessment of Lung Ultrasound data from Covid-19 patients

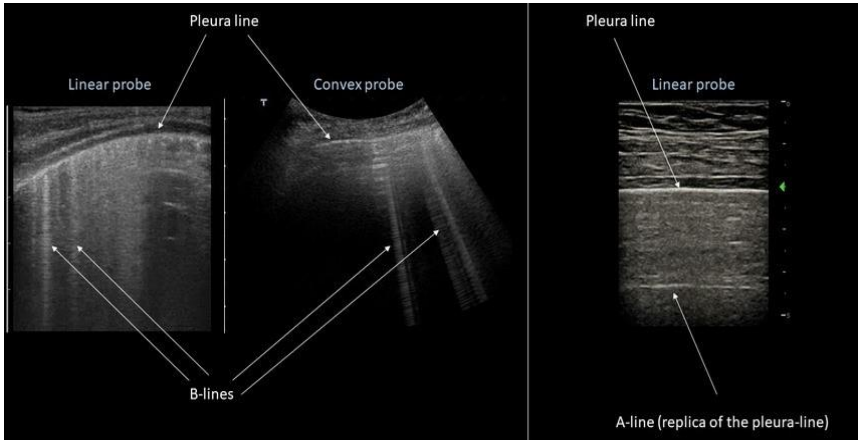
ZIHADUL AZAM

1. Introduction

Lung Ultrasound (LUS) and COVID-19

- Covid-19 virus can cause Pneumonia.
- Covid-19 Pneumonia: the lungs become filled with fluid and inflamed.
- Clinicians use LUS to assess the lungs condition by evaluating the presence of artifacts in images.





Role of Deep Learning

- During the pandemic different DL techniques have been employed to automate the evaluation process in order to support clinicians.
- But majority of them were solutions based on novel architecture or combination of different DL techniques.
- Only few studies have been conducted on the application of the state-of-the-art networks applied on LUS data.
- For this reason, this work is focused only on state-of-the-art models.

2. Goals

Goals

- How **State-of-the-art CNNs** perform with LUS data?
- Do they have a good **generalization capability**?
- **How much data** is required to build a model with compareable performance?
- **Explainable AI**?

3. Dataset

ICLUS-DB Dataset

05
Clinics

- Brescia
- Rome
- Lucca
- Tione
- Pavia

35
Patients

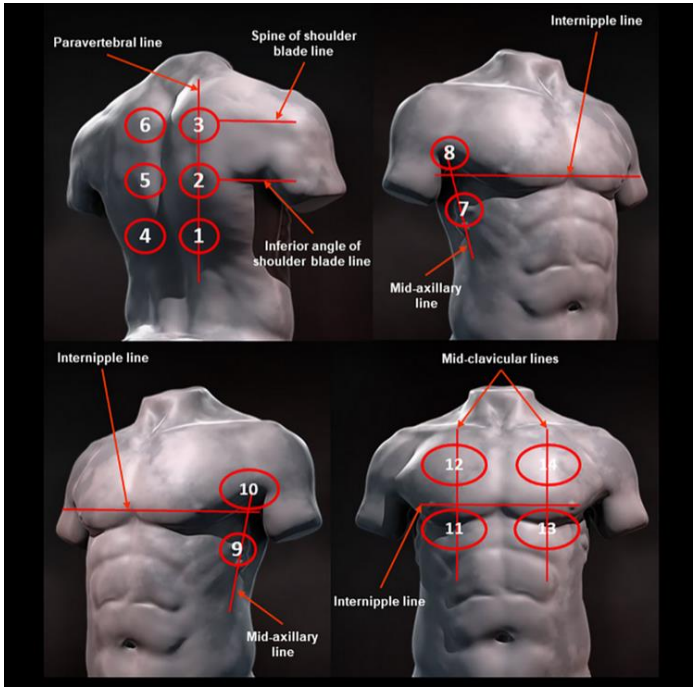
277
Videos

58’924
Frames

Standard Acquisition Protocol

14 areas:

- 4 front
- 4 lateral
- 6 back



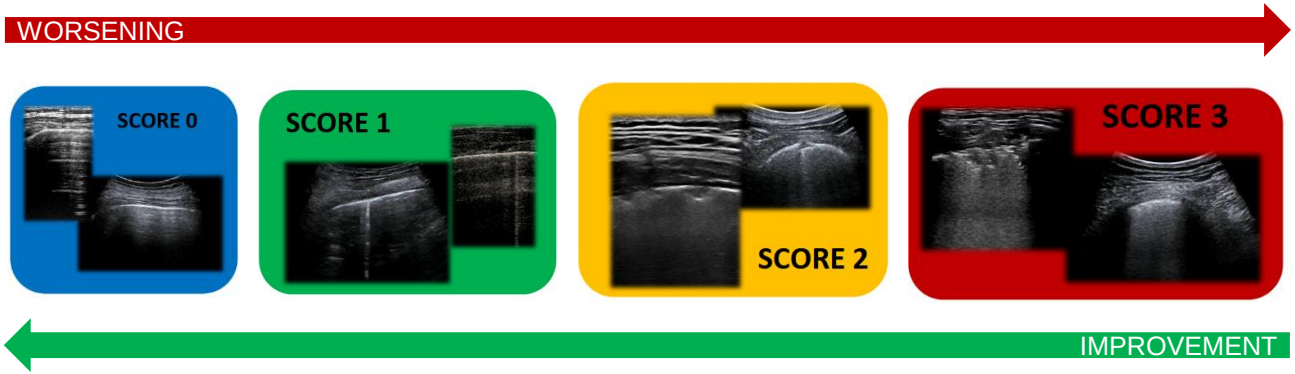
"Proposal for International Standardization of the Use of Lung Ultrasound for Patients With COVID-19"

- G. Soldati et al, 2020

Data labelling

4 level scoring system:

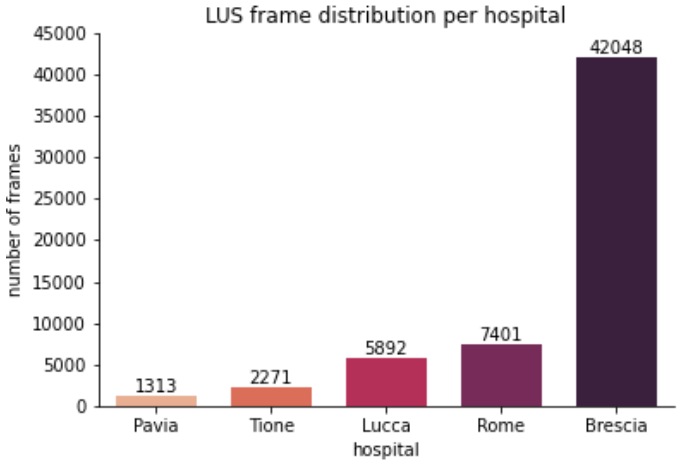
- Score 0
- Score 1
- Score 2
- Score 3

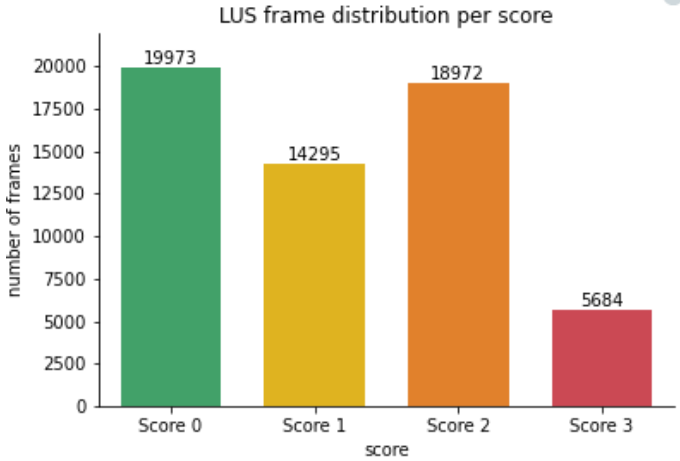


"Proposal for International Standardization of the Use of Lung Ultrasound for Patients With COVID-19"

- G. Soldati et al, 2020

Dataset statistics



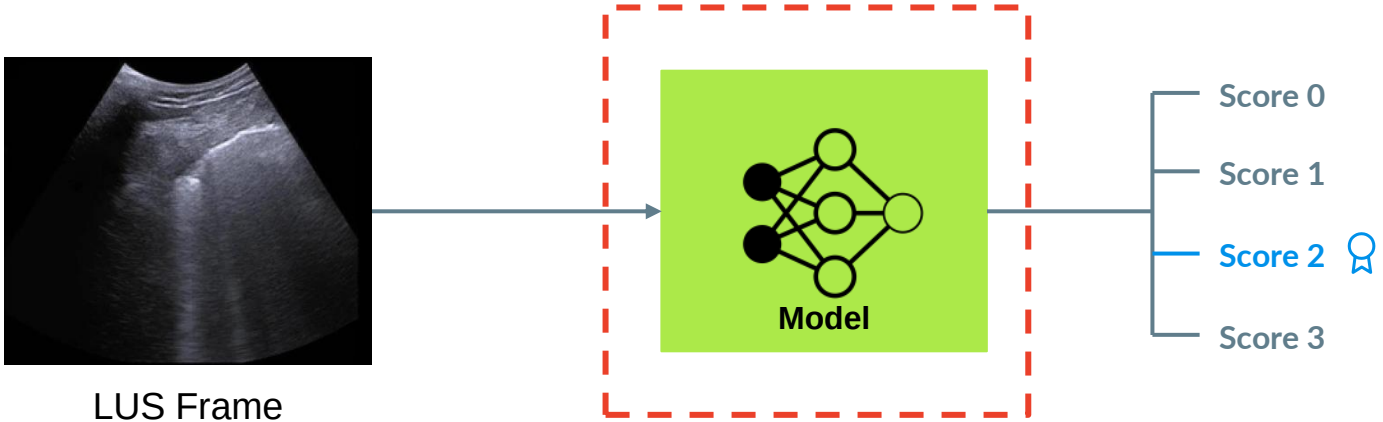


4. Methodologies & Results

Methodologies

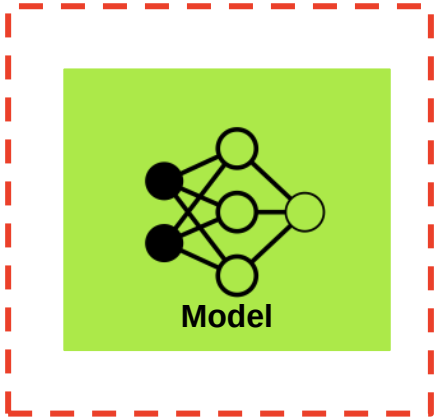
1. Frame-level scoring system
2. Grad-CAM algorithm
3. Generalization Capability across different medical centers

1. Frame-level scoring system



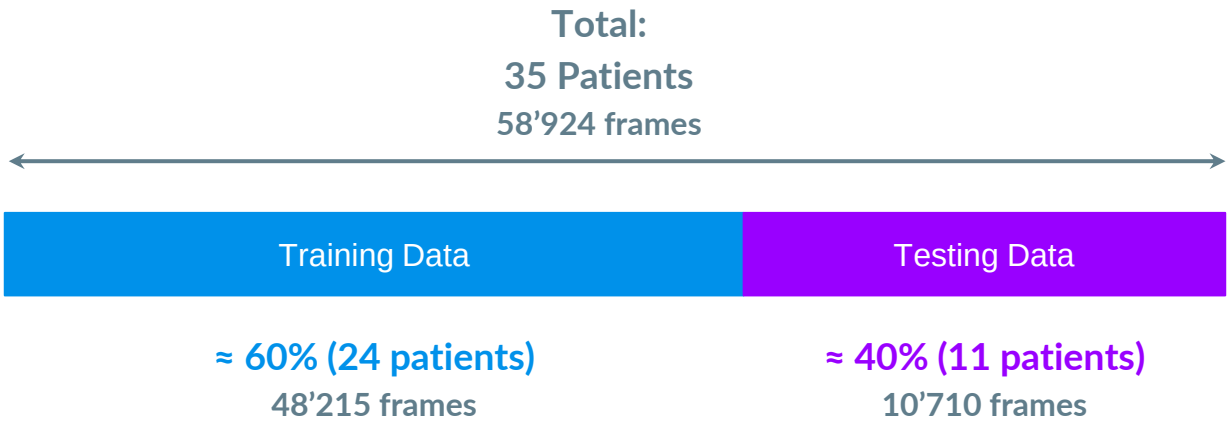
Model Architectures

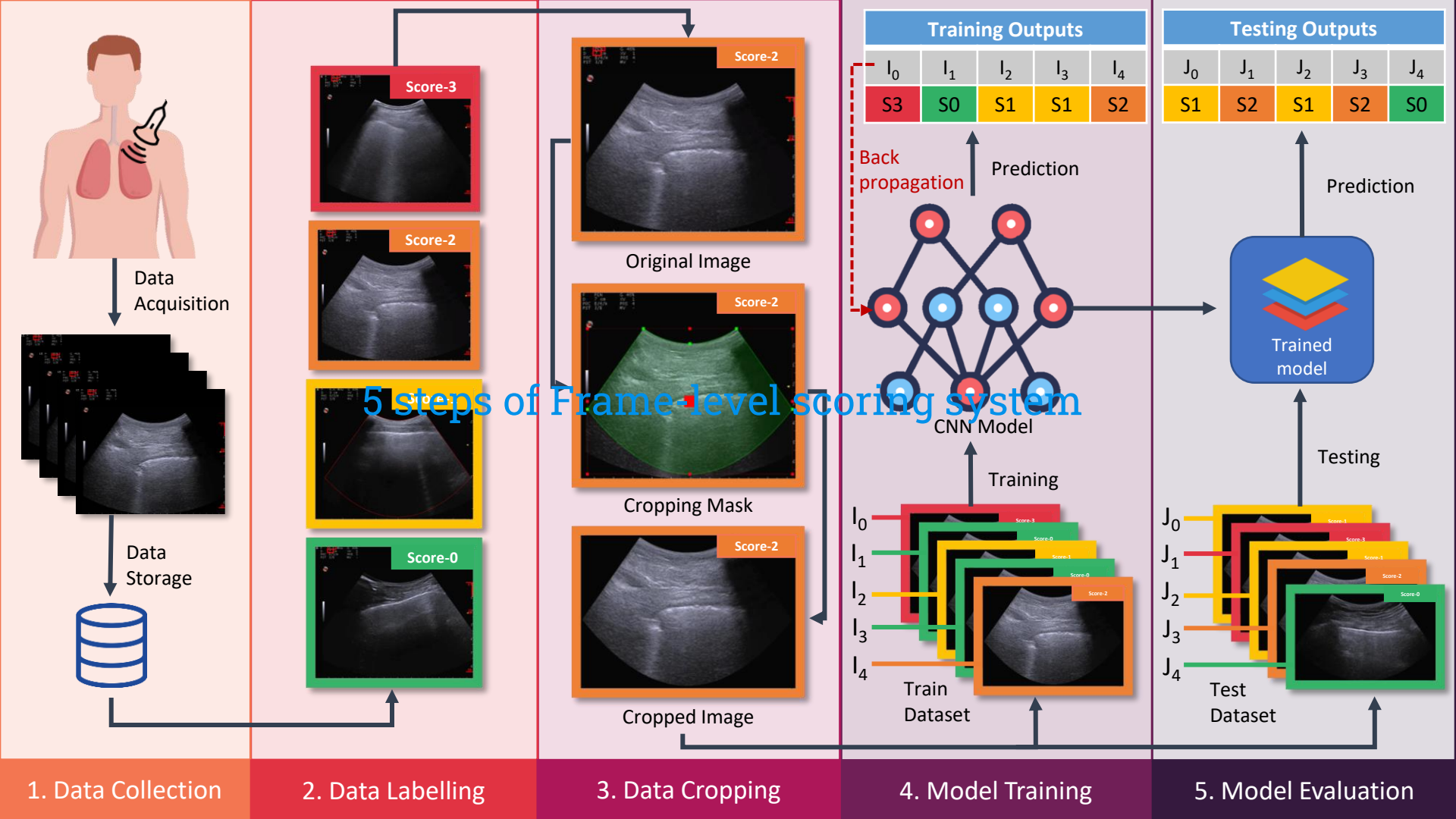
1. ResNet-18
2. ResNet-50
3. ResNet-101
4. DensNet-121
5. DensNet-201
6. InceptionV3



Data splitting

- Patient level




Data Acquisition
Data Storage
Score-3
Score-2
Score-1
Score-0
Original Image
Cropping Mask
Cropped Image
Training Outputs
I0 I1 I2 I3 I4
S3 S0 S1 S1 S2
Back propagation
Prediction
CNN Model
Training
Train Dataset
Testing Outputs
J0 J1 J2 J3 J4
S1 S2 S1 S2 S0
Prediction
Trained model
Testing
Test Dataset
5 steps of Frame-level scoring system
1. Data Collection
2. Data Labelling
3. Data Cropping
4. Model Training
5. Model Evaluation

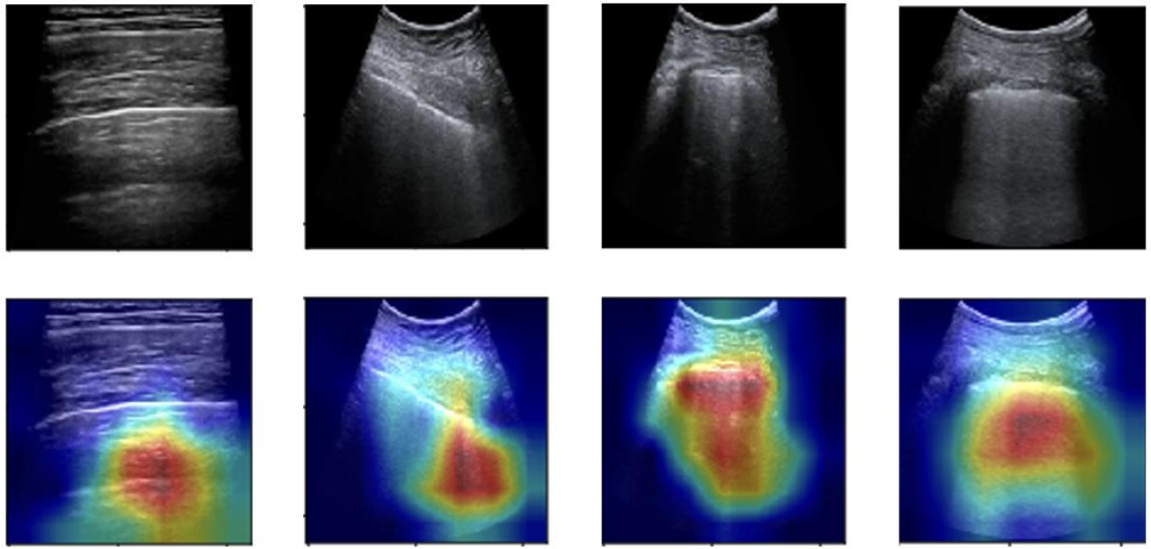
1. Frame-level scoring system: Results

Methodology	Employed Model and Technique	F1-Score
Base line	Reg + STN + CNN [1]	0.651
	ResNet-18 + Annotations [2]	0.688
Proposed	ResNet-18	0.659
	ResNet-50	0.655
	ResNet-101	0.651
	DensNet-121	0.6513
	DesnsNet-201	0.6517
	InceptionV3	0.612
Proposed Pre-Trained	ResNet-18	0.645

[1] S. Roy et al, Deep learning for classification and localization of covid-19 markers in point-of-care lung ultrasound, 2020

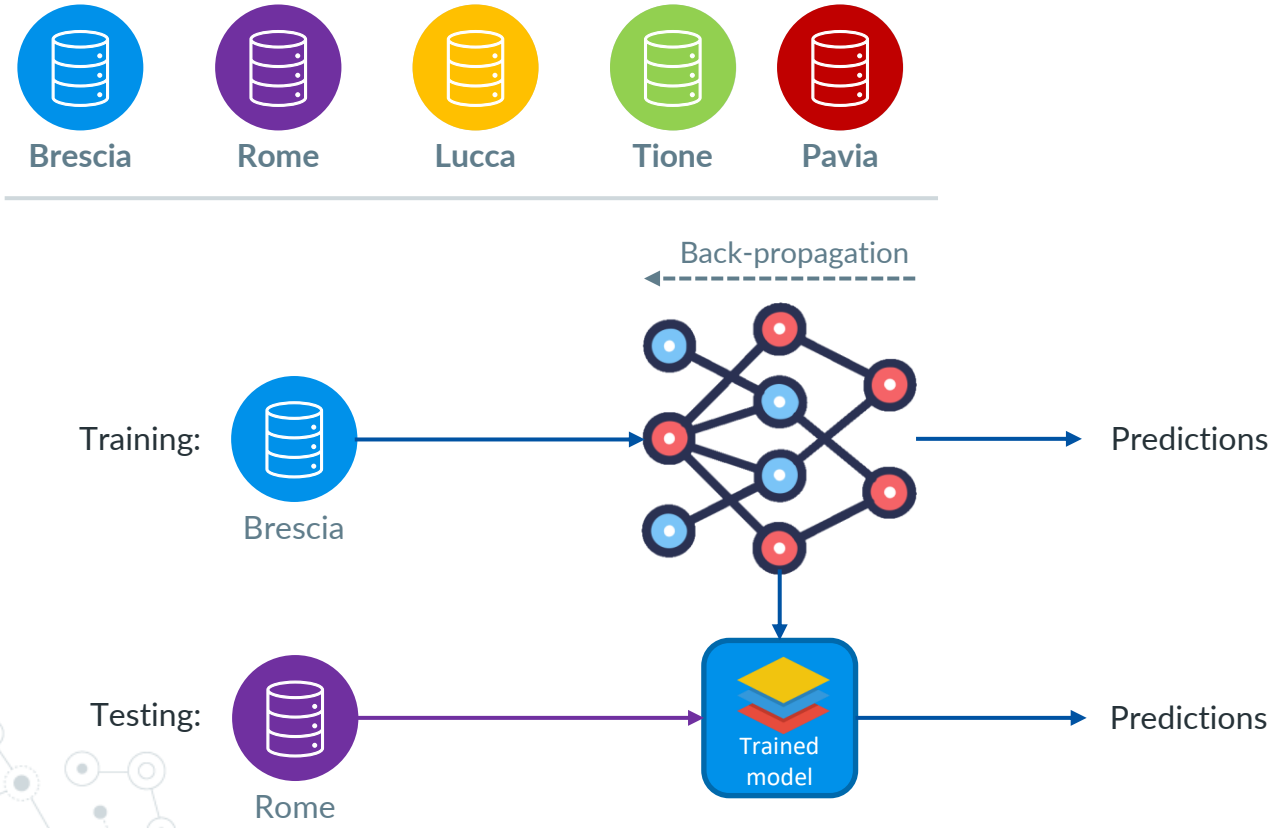
[2] O. Frank et al, Integrating domain knowledge into deep networks for lung ultrasound with applications to covid-19, 2021

2. Grad-CAM on ResNet-18 (trained from scratch)

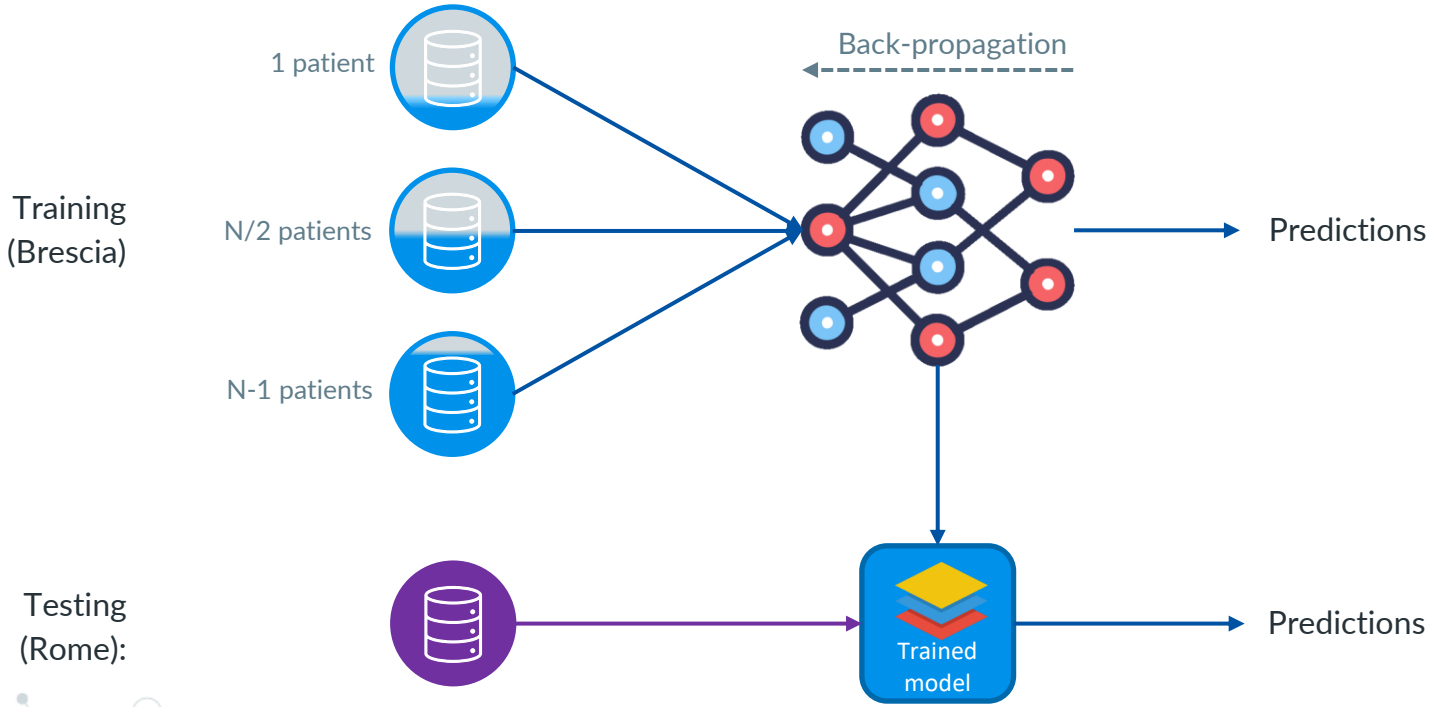


Score 0 Score 1 Score 2 Score 3

3. Generalization Capability – 1

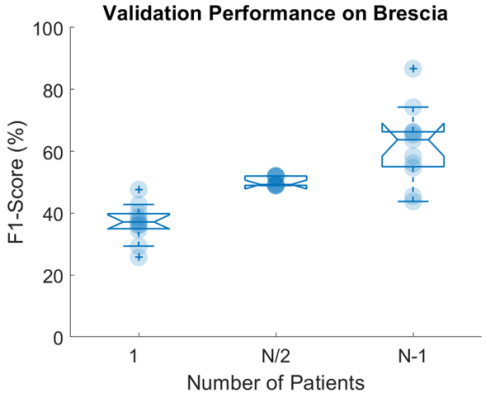


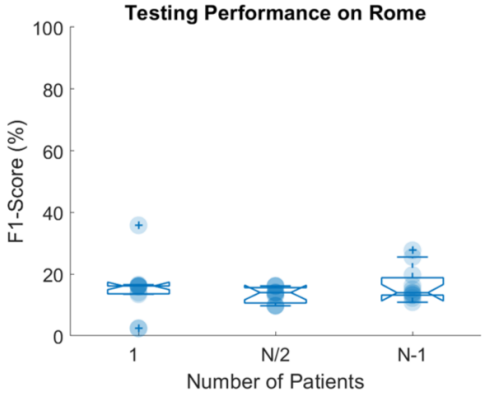
3. Generalization Capability – 2



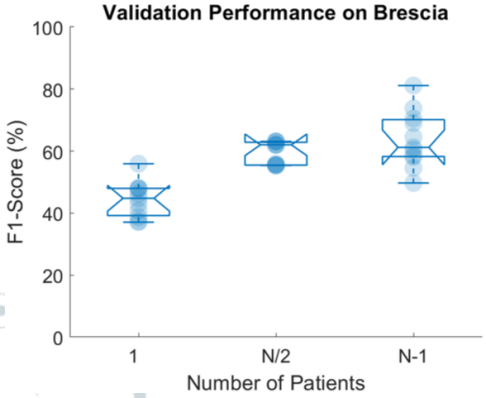
Generalization Capability: Results

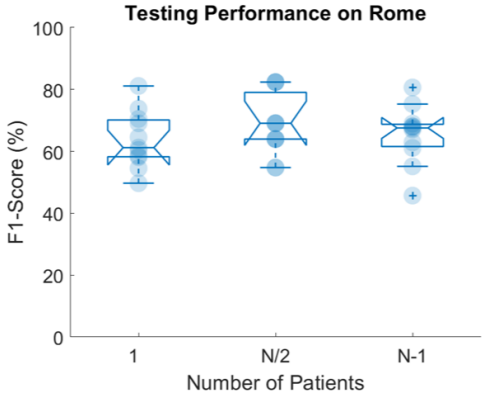
ResNet-18
Trained from scratch





ResNet-18
Pre-trained





Conclusion

- Outperforming results were achieved by using state-of-the-art CNN models.
- ResNet-18 was found to be the best-performing model with an F1-Score of 0.659.
- Grad-CAM algorithm helped to understand model behaviors.
- Pre-trained model has better generalization capabilities.
- Half of training patients were enough to achieve comparable performance.
- Future works: transformer-based models, video-level classification.

UNIVERSITÀ DI TRENTO

ULTRa
ULTRASOUND LAB TRENTO

Thank You !

For Your Attention