POLITECNICO
MILANO 1863

Lung Cancer Detection

False Positive Reduction

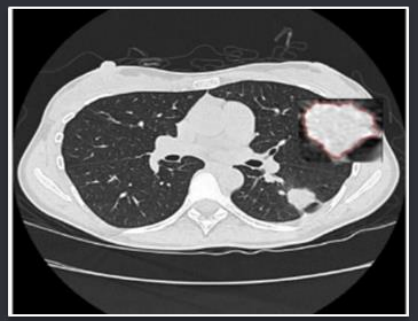
Brégeat Joseph
Bucciarelli Vittoria
Gerletti Sebastian Edoardo
Kassem Karim
Mazzucchelli Umberto

1 INTRODUCTION

LUNG CANCER

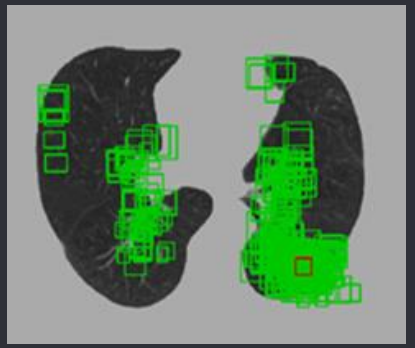
1.3 million deaths annually

Need for early diagnosis from **LDCT scans**

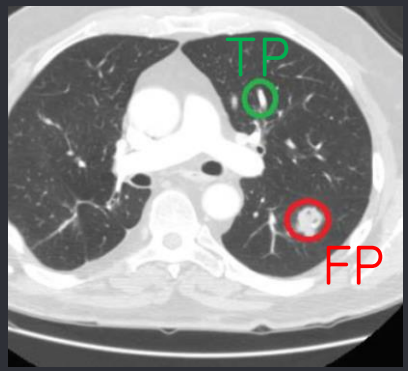


Computer aided detection system

Candidate screening

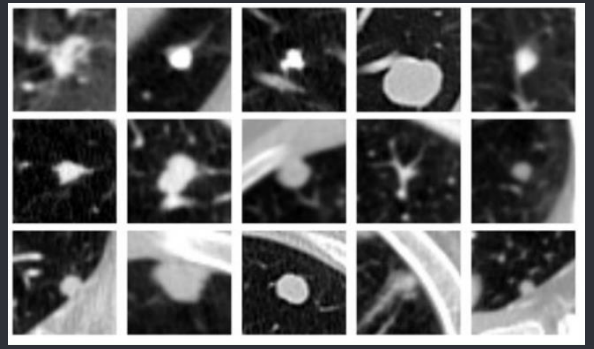


False positive reduction

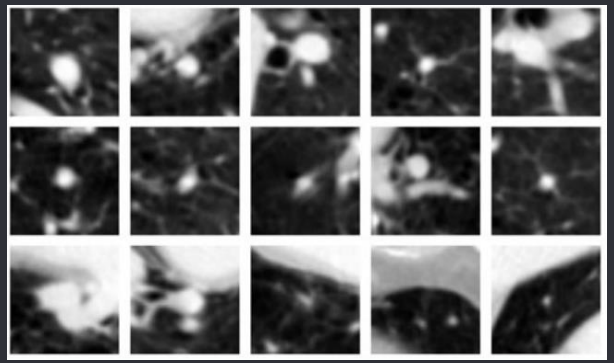


FALSE POSITIVE REDUCTION

- Similar appearance of FP and TP nodules
- **Noisy** background
- Overdiagnosis
- **Heterogeneous** size, shape and location

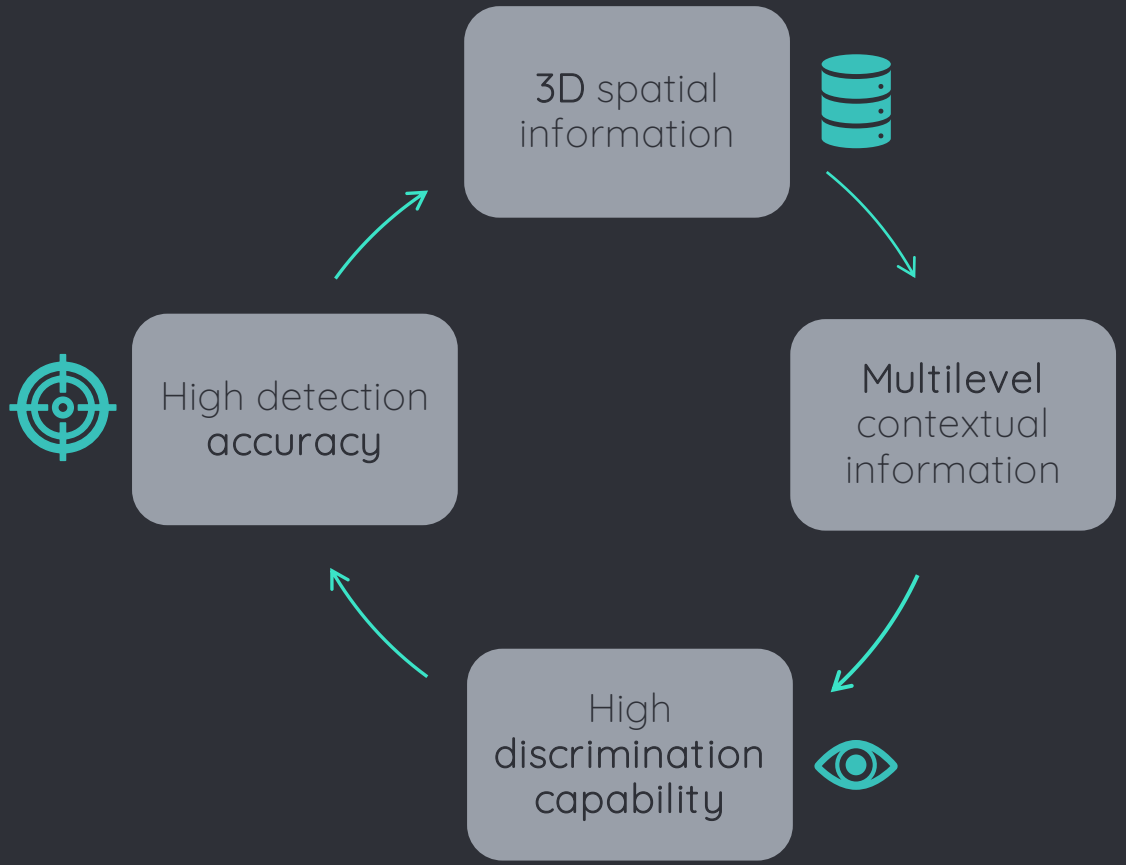


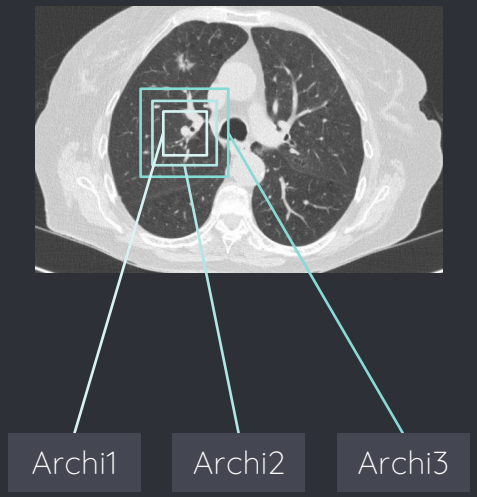
True positives



False positives

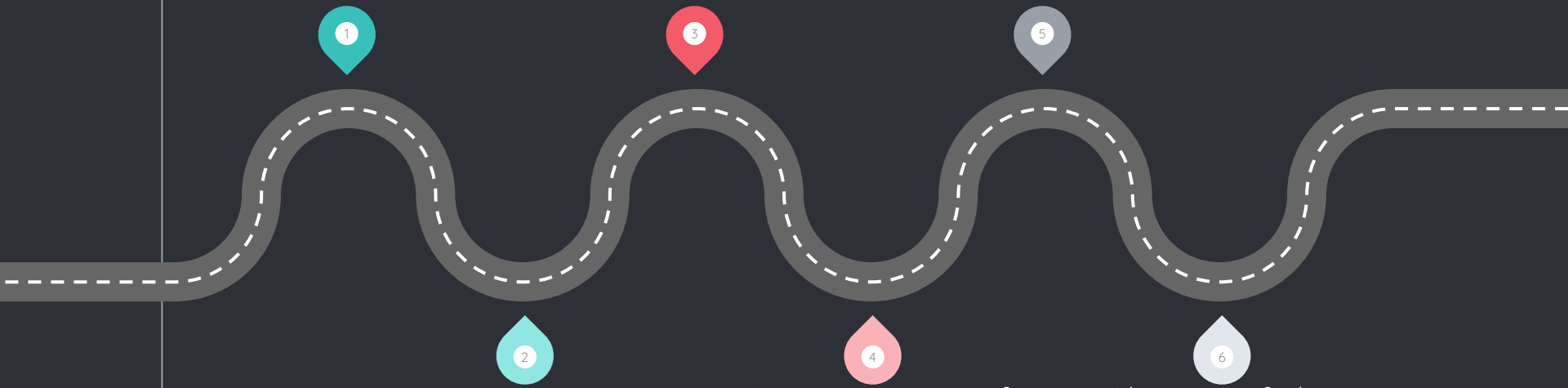
WHY 3D CNNS?


3D spatial information
Multilevel contextual information
High discrimination capability
High detection accuracy


Archi1
Archi2
Archi3

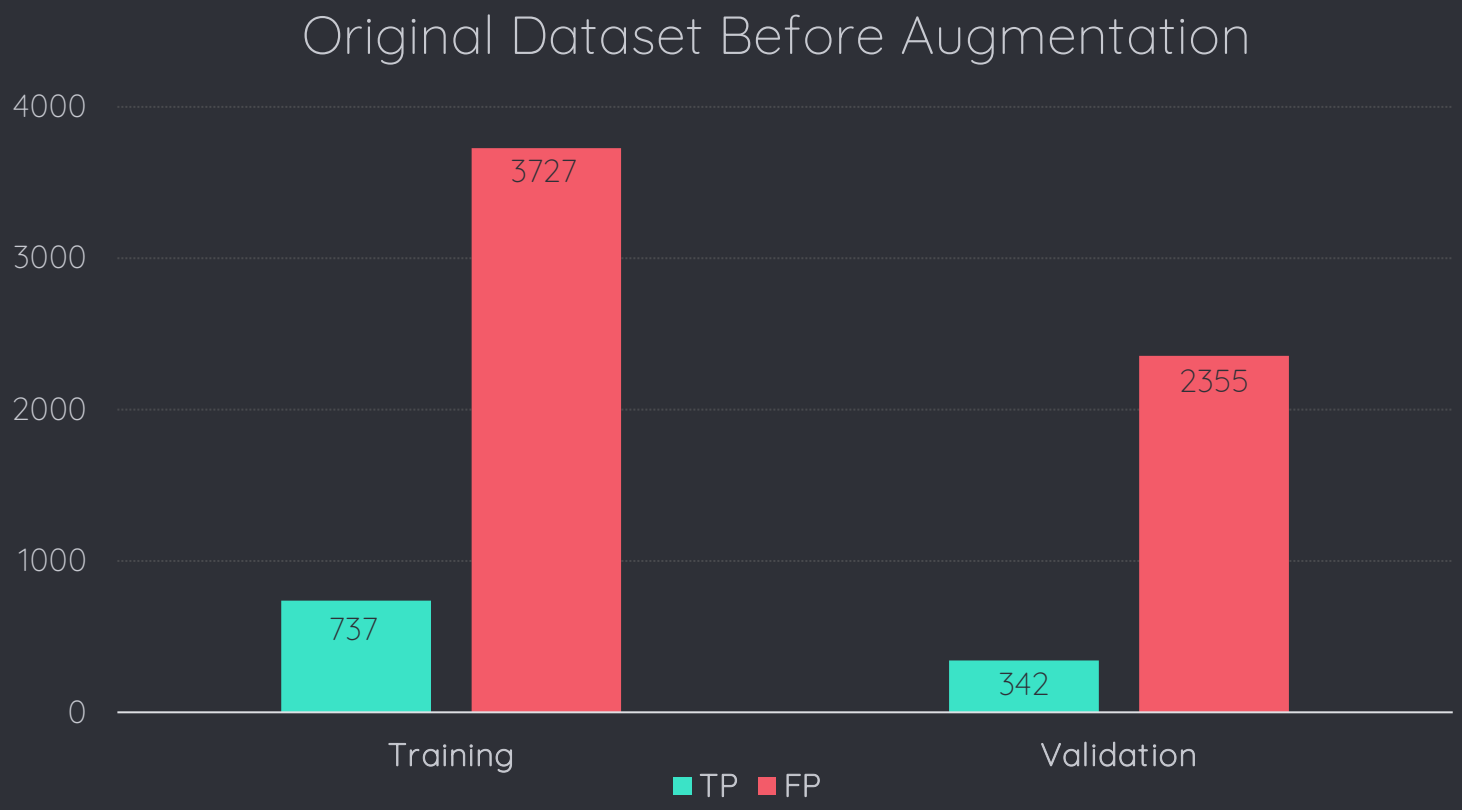
WHAT WE DID

1. Implementation of Archi1, Archi2 and Archi3[1]
2. Optimization of the 3 networks separately
3. New image processing technique
4. Comparison of the performances with the old and new processing technique
5. Networks combination and performance evaluation
6. Choice of the best model



1: Dou et al., Multilevel Contextual 3-D CNNs for False Positive Reduction in Pulmonary Nodule Detection, 2017

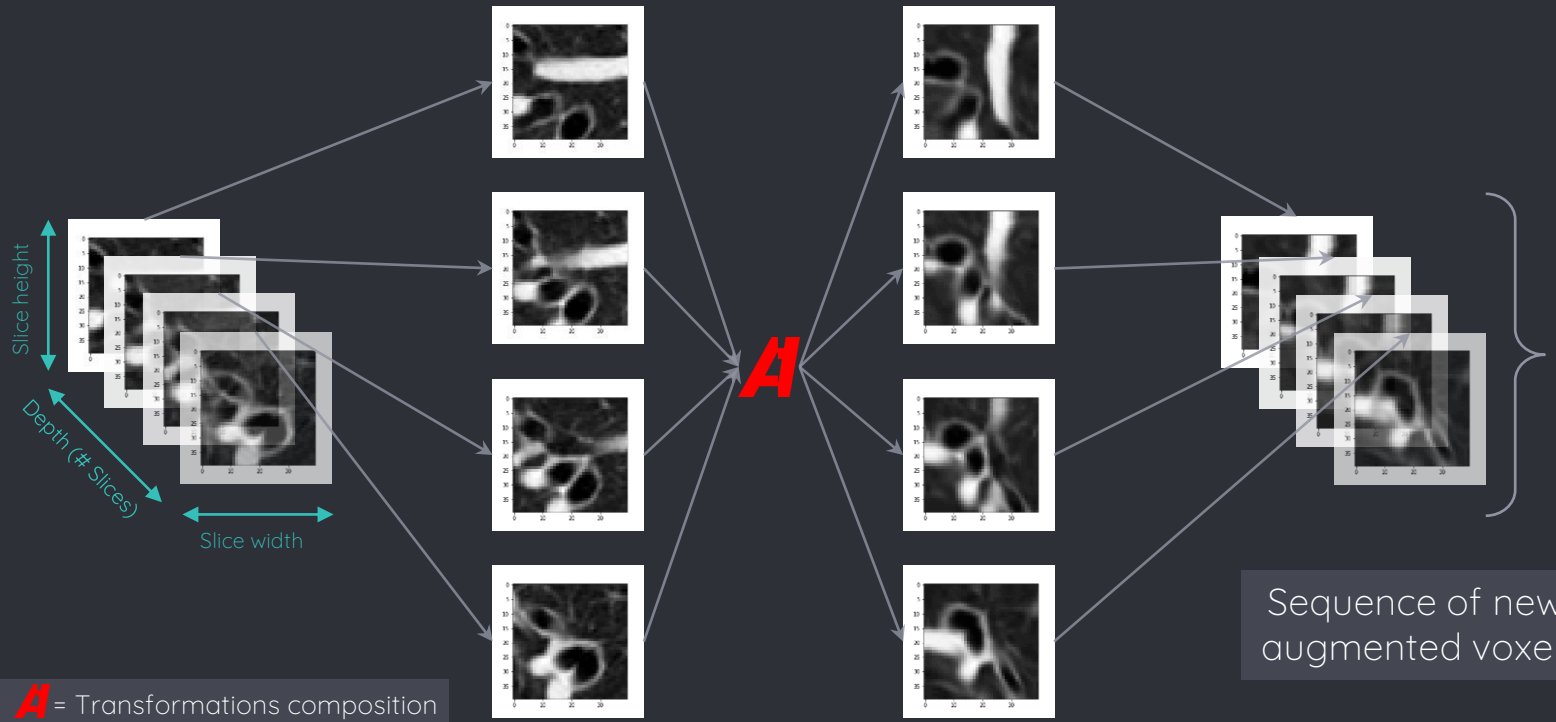
2 IMAGE PROCESSING TECHNIQUE



DATASET FEATURES

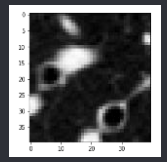
DATA AUGMENTATION

Original Voxel | Voxel Decomposition | Augmentation | Voxel Reconstruction



TRANSFORMATIONS

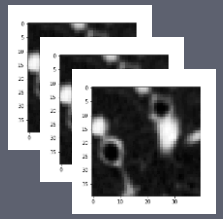
Original image



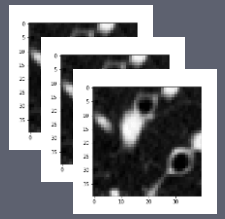
Applied to:

- Archi 1
- Archi 2
- Archi 3

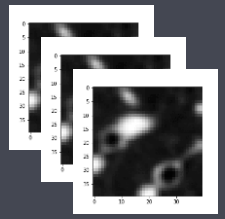
Rotate



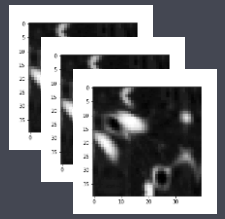
Flip - Transpose



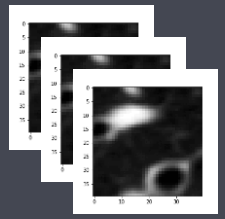
Blur



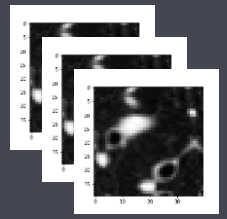
Grid Distortion



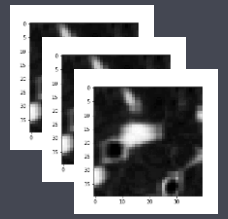
Perspective



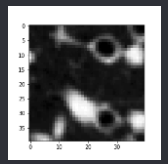
Optical Distortion



Elastic Transform



Augmented image

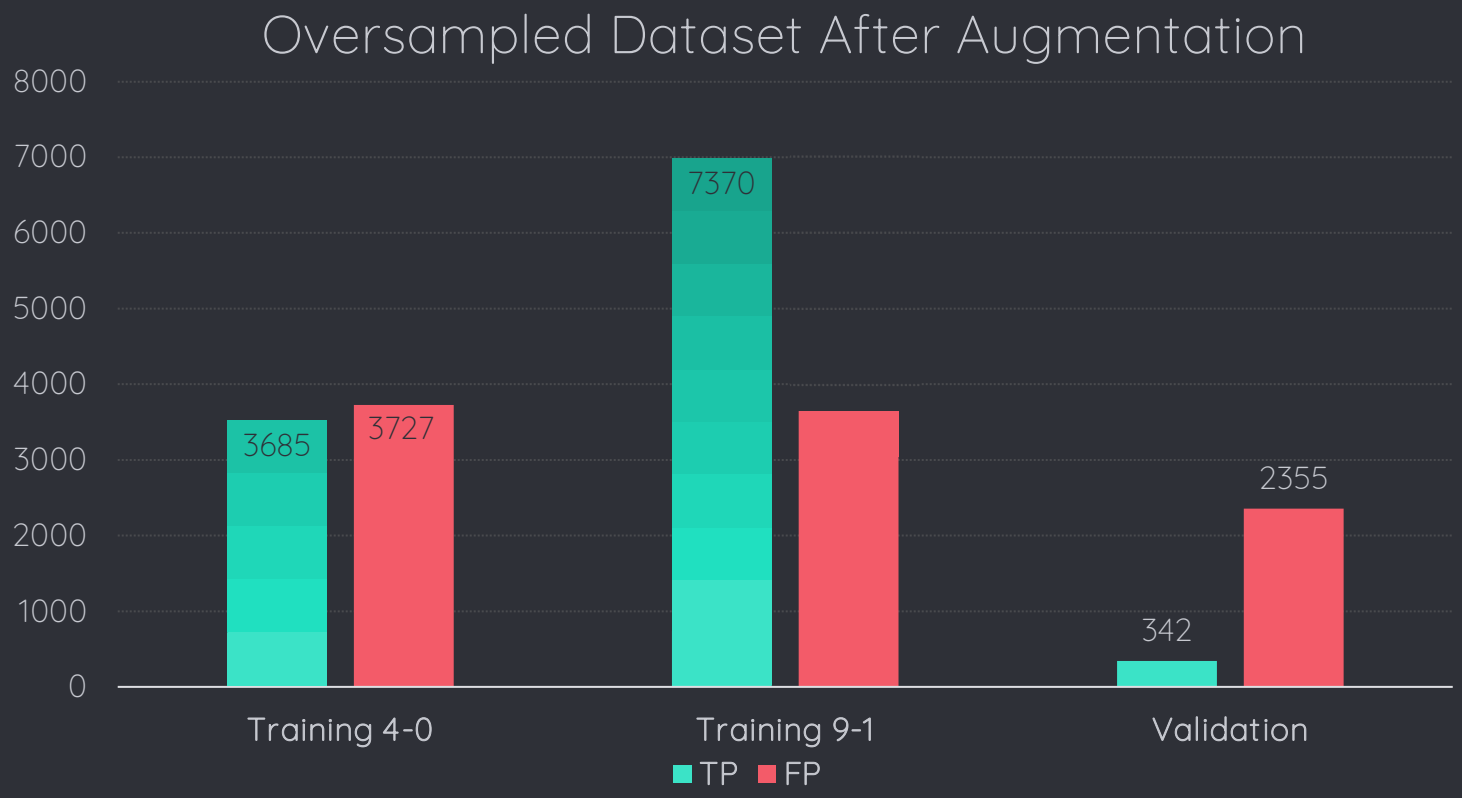


Standard — NumPy

Advanced — Albumentation

Joint Probability

Exclusive Probability

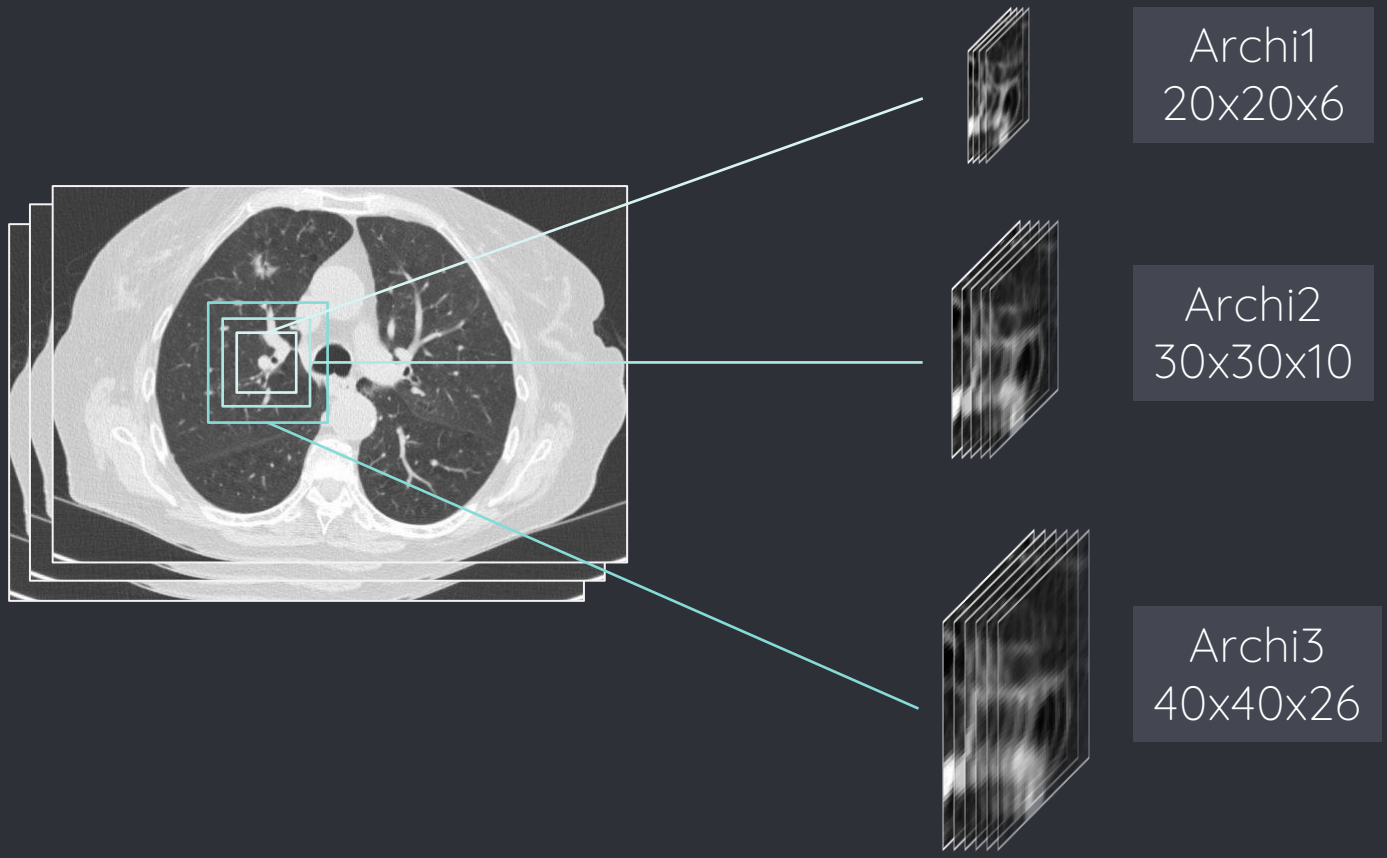


DATASET FEATURES

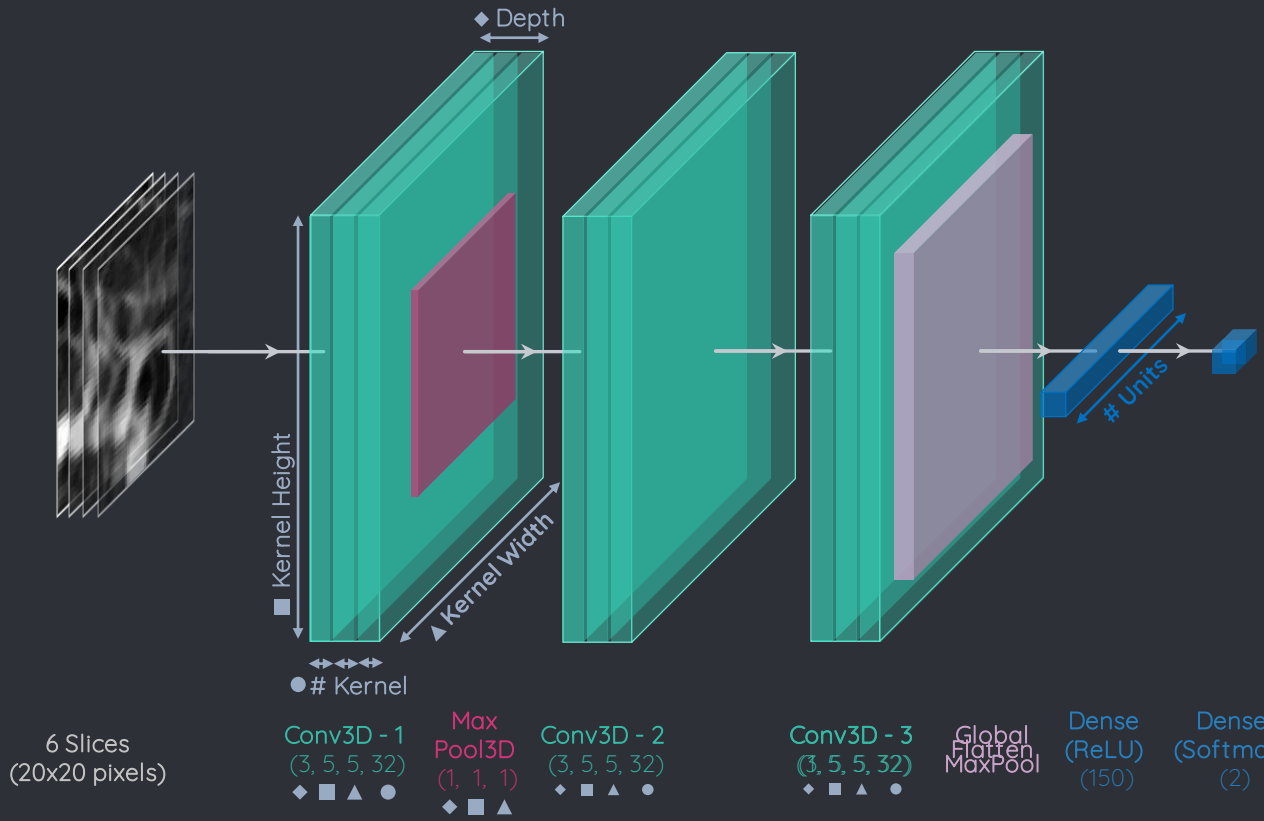
3

ARCHITECTURES EVALUATION & COMPARISON

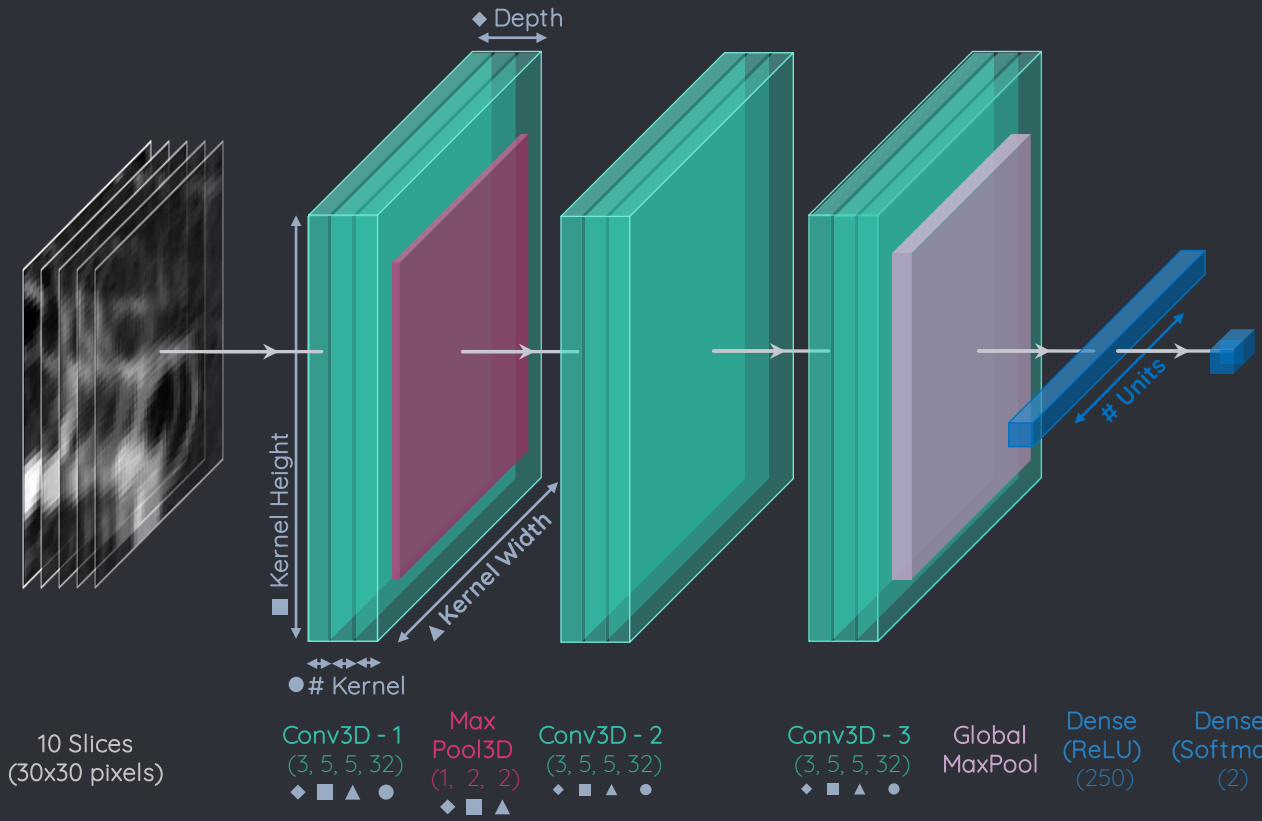
RECEPTIVE FIELD



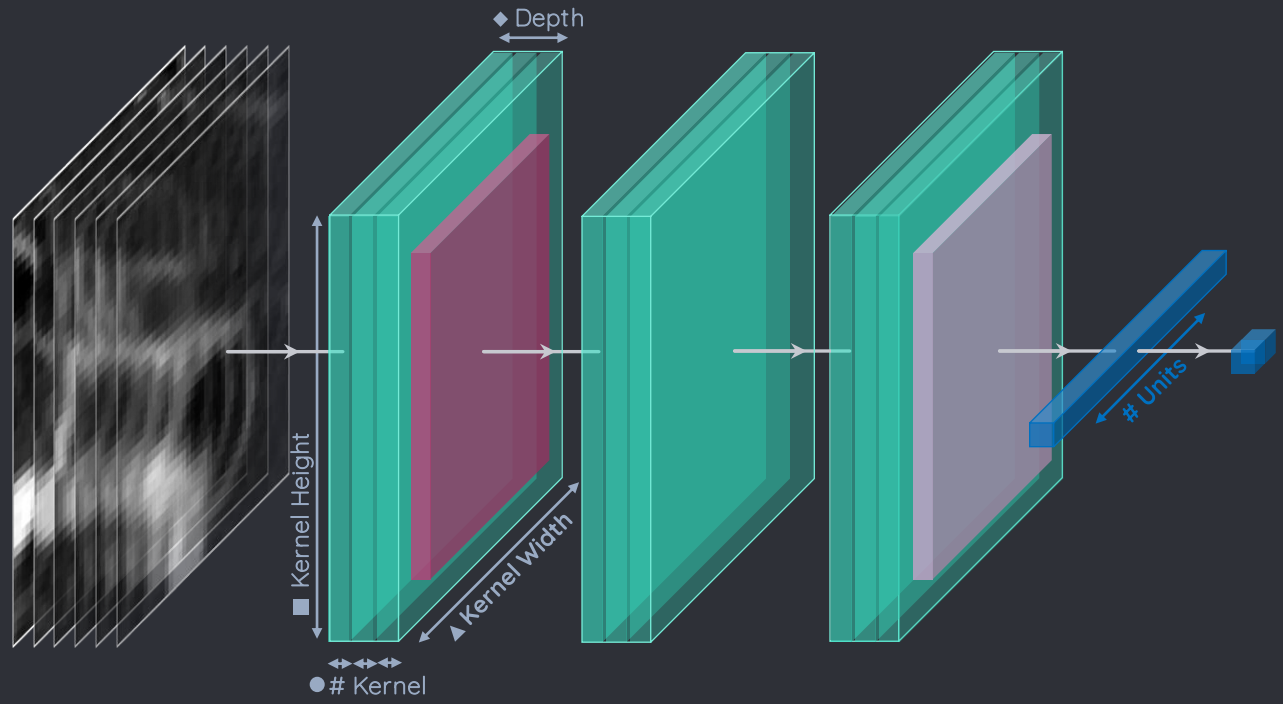
ARCHI 1 - STRUCTURE



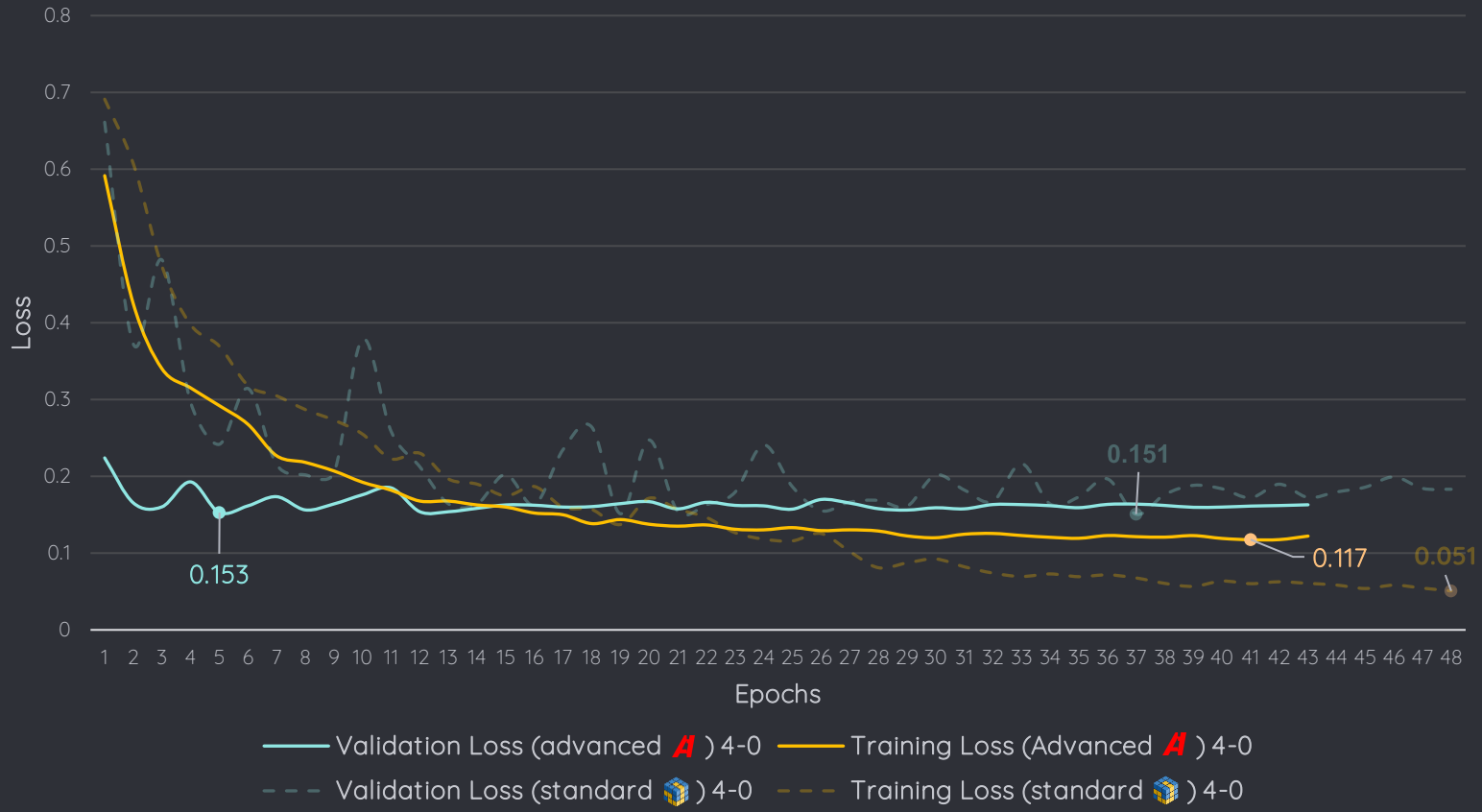
ARCHI 2 - STRUCTURE



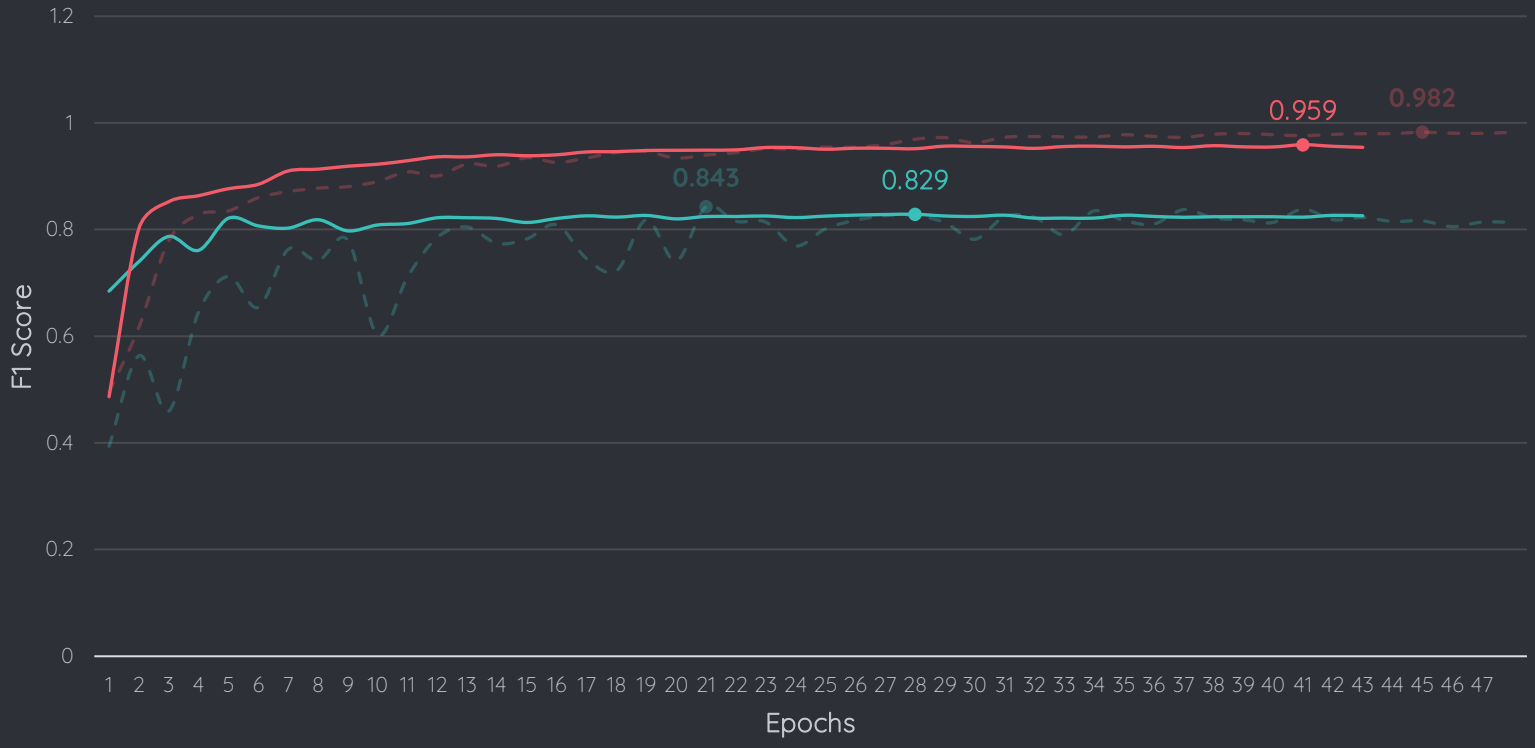
ARCHI 3 - STRUCTURE



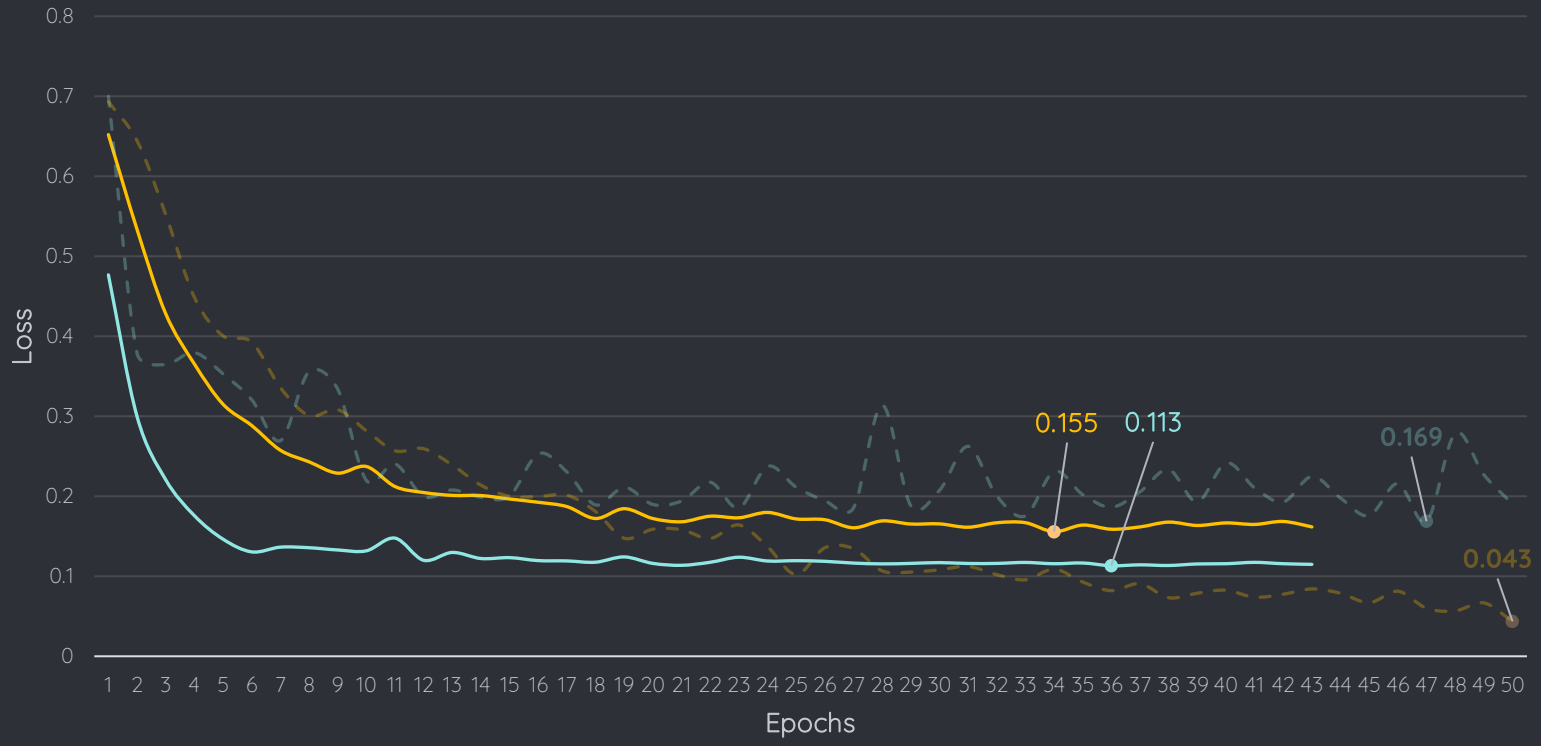
ARCHI 1 – LOSS


0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Loss
0.153
0.151
0.117
0.051
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48
Epochs
Validation Loss (advanced AI) 4-0
Training Loss (Advanced AI) 4-0
Validation Loss (standard) 4-0
Training Loss (standard) 4-0

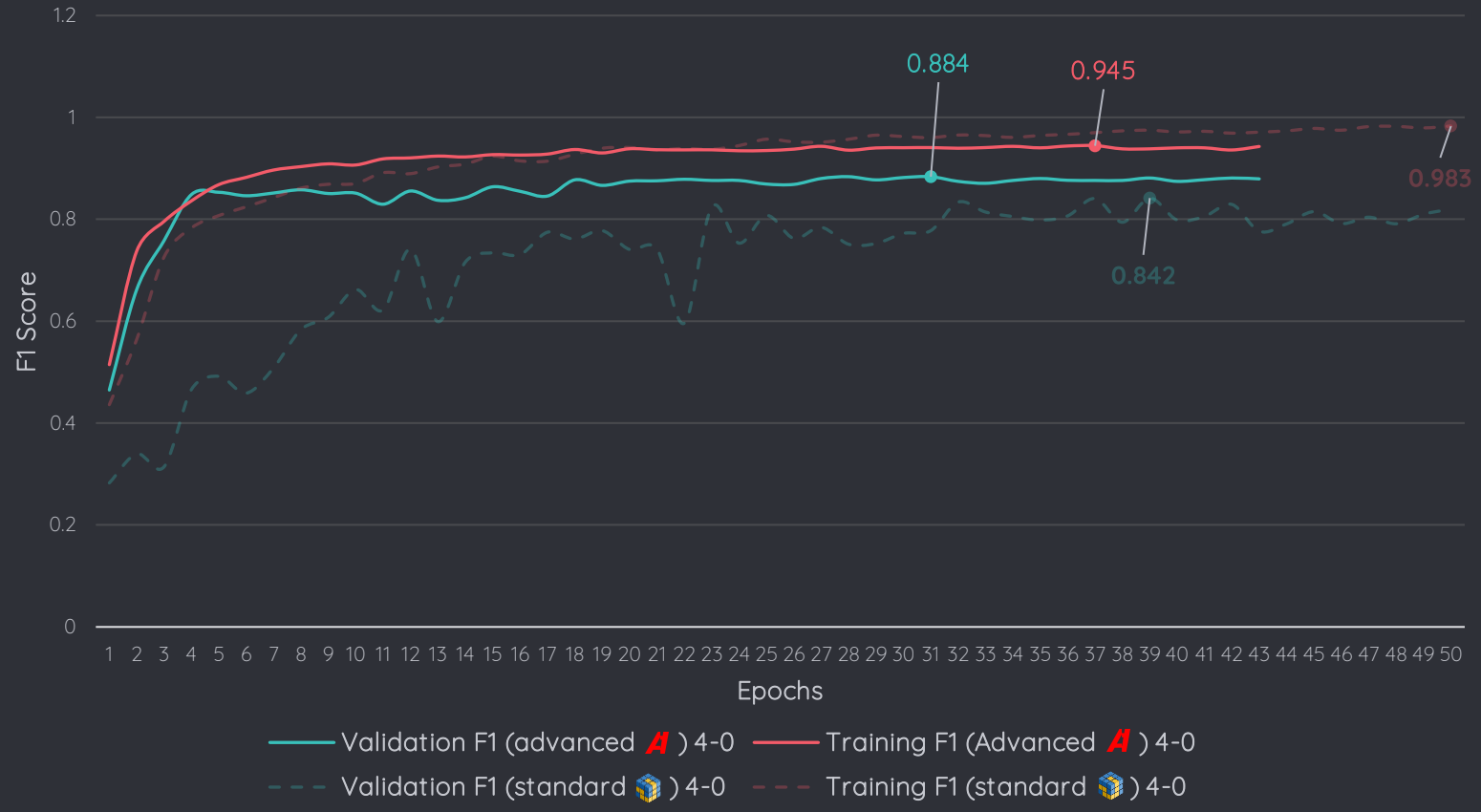
ARCHI 1 – F1 SCORE



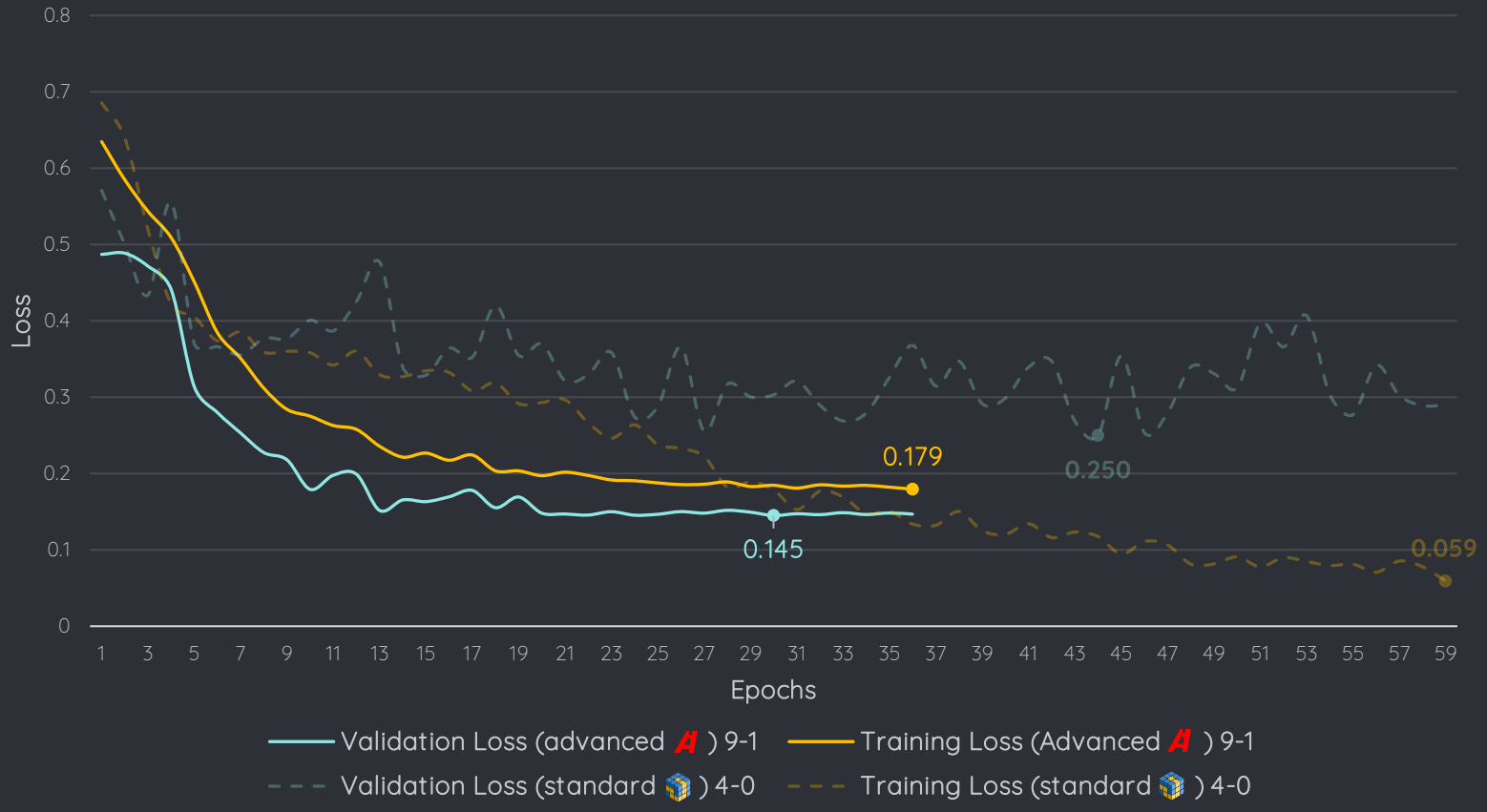
ARCHI 2 – LOSS


0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Loss
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50
Epochs
0.155
0.113
0.169
0.043
Validation Loss (advanced) 4-0
Training Loss (Advanced) 4-0
Validation Loss (standard) 4-0
Training Loss (standard) 4-0

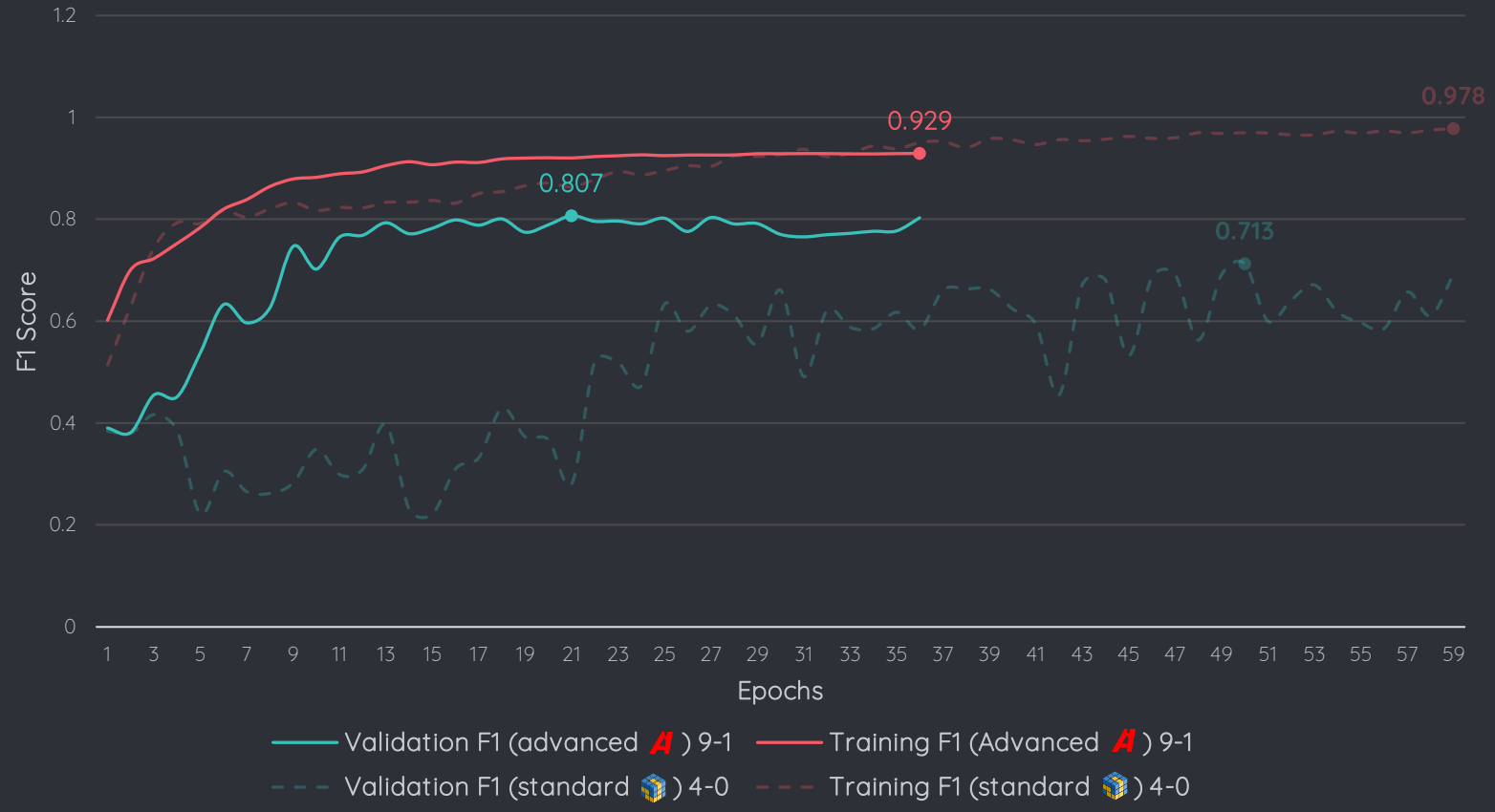
ARCHI 2 – F1 SCORE



ARCHI 3 – LOSS


0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Loss
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59
Epochs
0.179
0.250
0.145
0.059
Validation Loss (advanced A) 9-1
Training Loss (Advanced A) 9-1
Validation Loss (standard) 4-0
Training Loss (standard) 4-0

ARCHI 3 – F1 SCORE



ARTICLE'S ARCHIs vs OURS

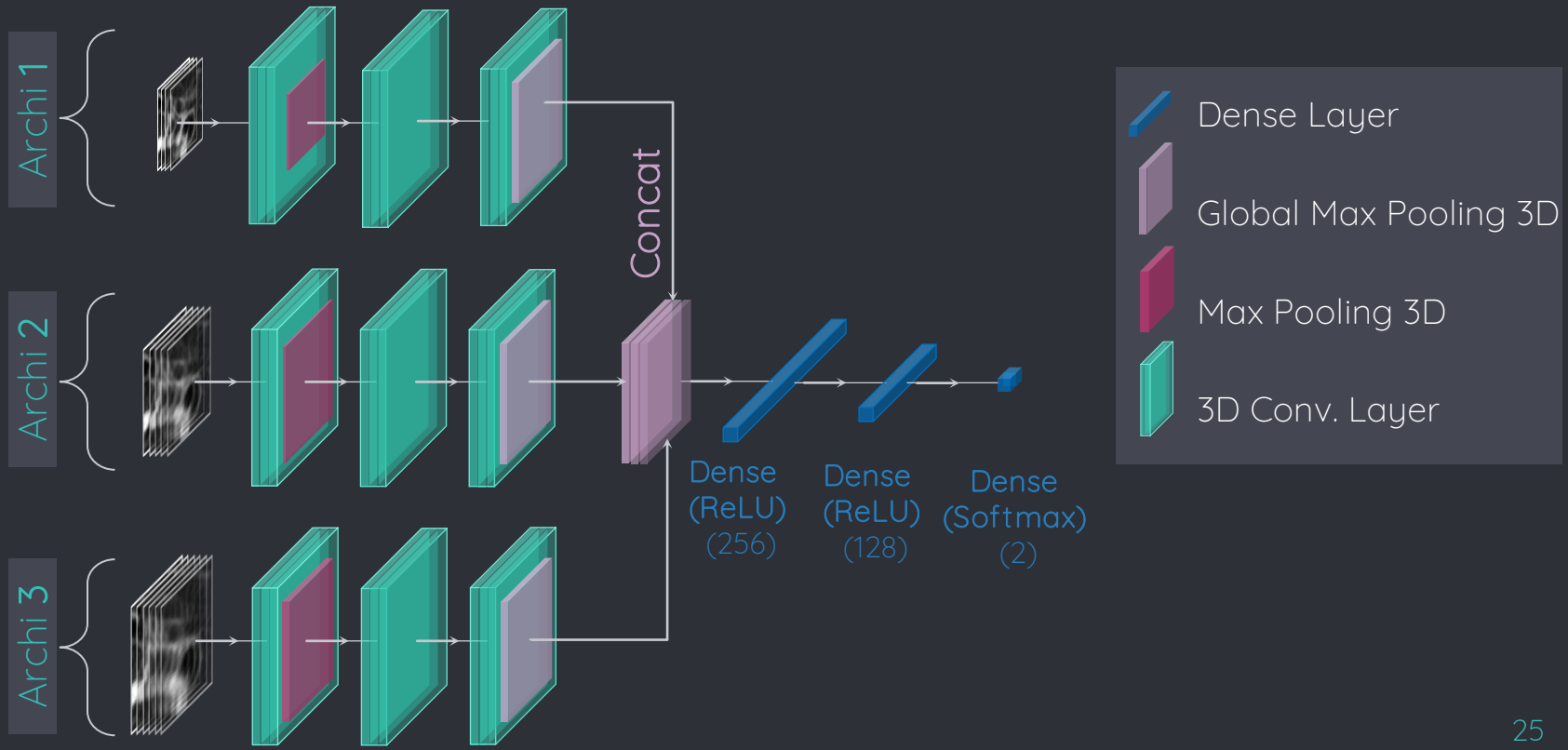
	Archi 1	Archi2	Archi3
F1 score – Article* *(Dou's model structure)*	0,7714	0,7929	0,7431
F1 score – Our best	0,8290	0,8836	0,8066
Improvement *(percentage point)*	+ 5,76%	+ 9.07%	+ 6,35%

* score obtained by applying the article's architecture to our dataset

4 MULTIMODAL NETWORKS

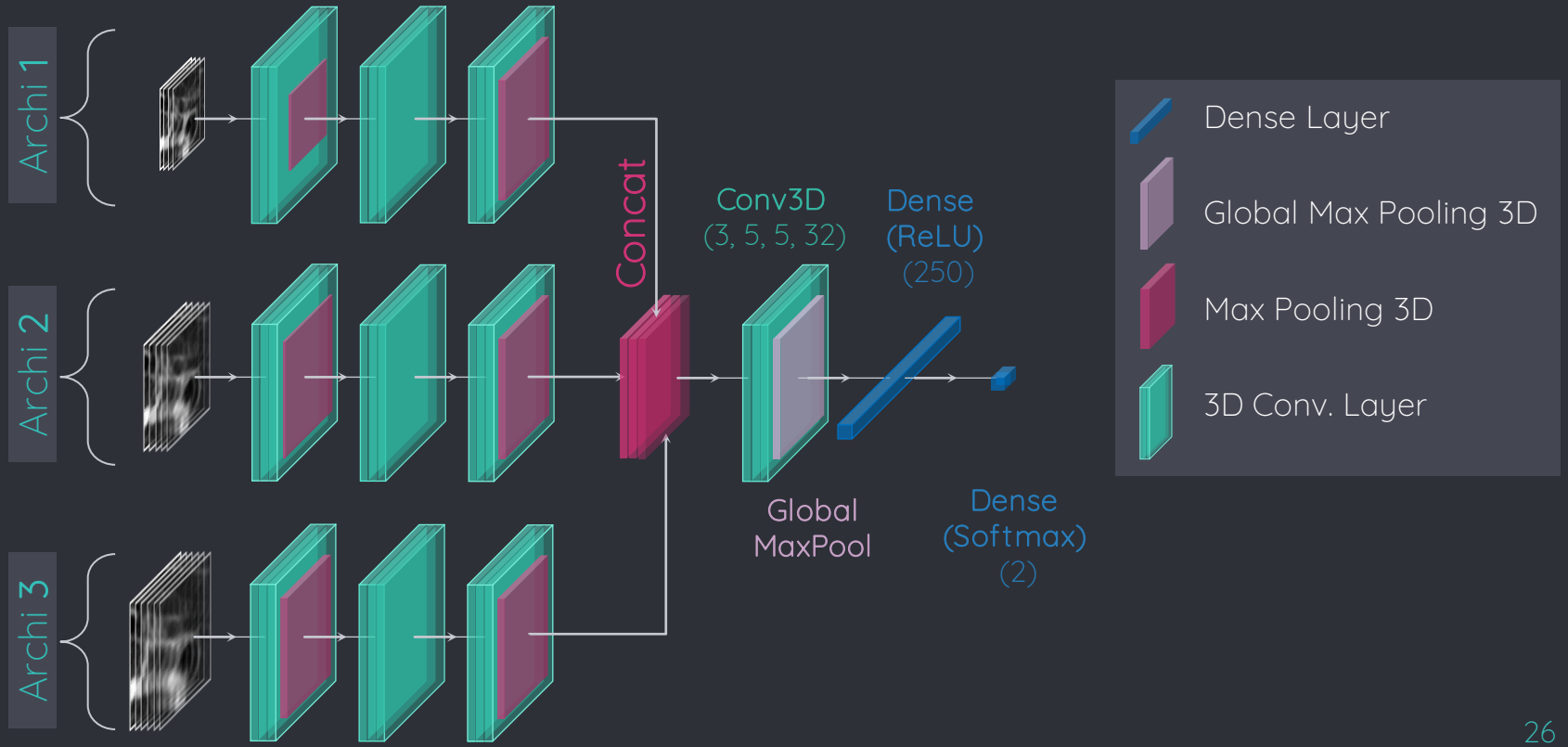
MULTIMODAL NETWORKS

1) Multimodal network with parallel convolutions and common classifier

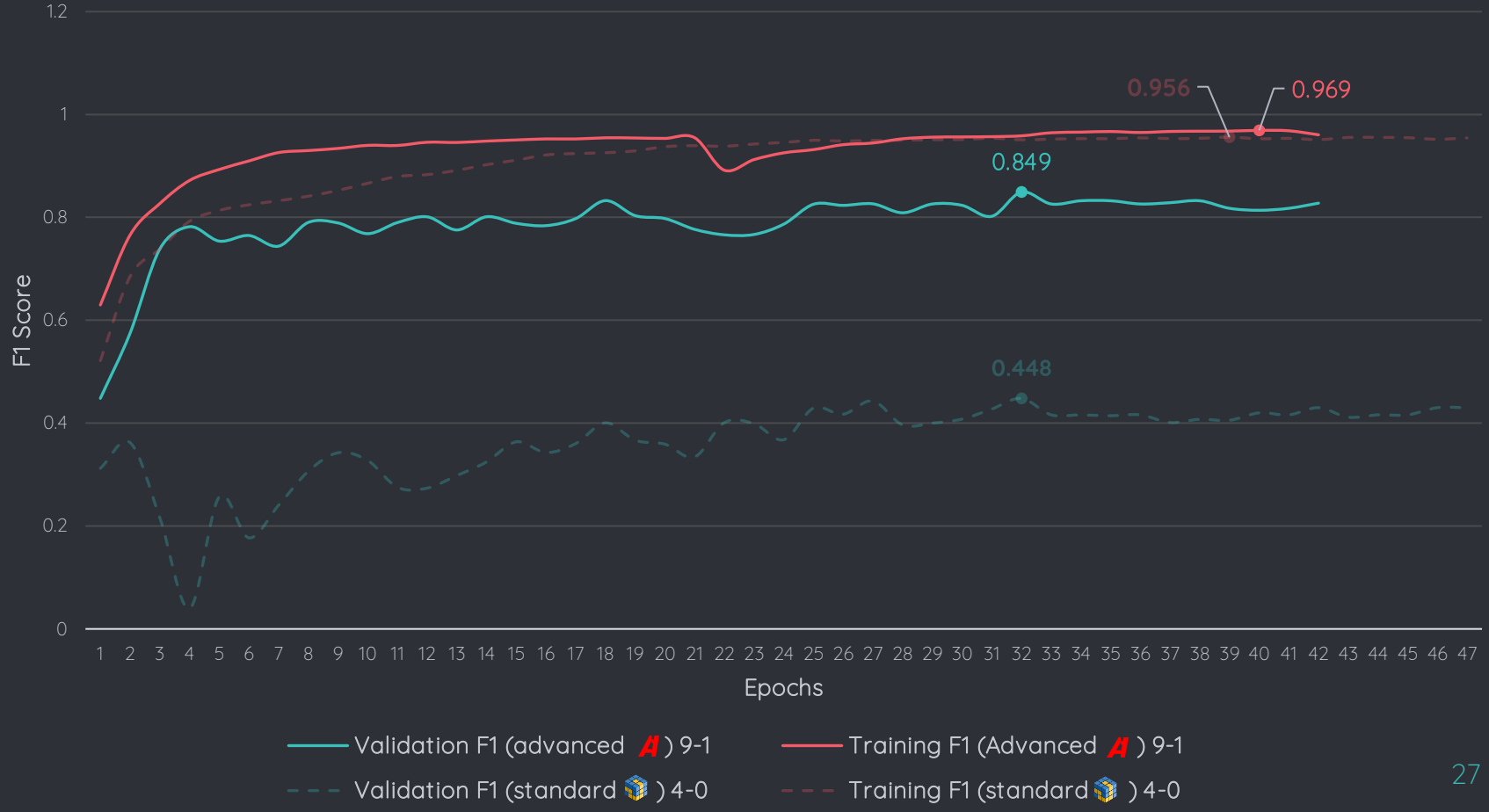


MULTIMODAL NETWORKS

2) Multimodal network with parallel convolutions and common convolution



COMMON CONVOLUTIONS



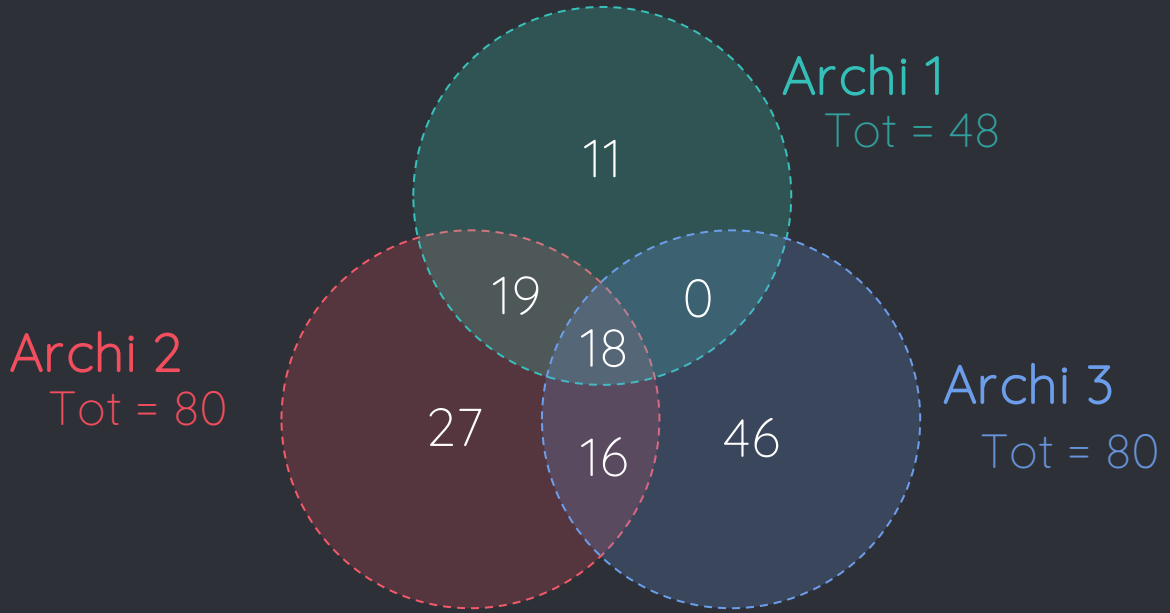
MULTIMODAL #2

	F1 score	Recall	Precision
Standard	0.448	0,527	0,442
Advanced (10-2)	0,849	0,941	0,792
Improvement	+ 40%	+ 41.4%	+ 35%

5 ENSEMBLE MODELS

WHY ENSEMBLES?

Advanced Dataset (***A***):



Number of missclassified samples

ENSEMBLE BUILDING APPROACHES

Gs-W) Gridsearch **weights** for each ARCHI for best aupr

Gs-T) Gridsearch prediction **threshold** for all ARCHIs for best f1

F1-W) Weight each ARCHI based on **f1score** on validation set

DT) Apply **decision tree** to all ARCHIs outputs

LR) Apply **logistic regression** to all ARCHIs outputs

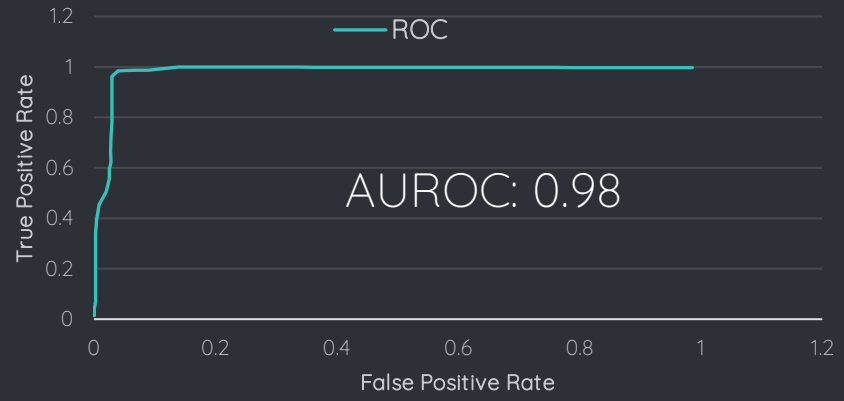
RESULTS FOR:
5 TYPES OF ENSEMBLES – 3 DATASETS

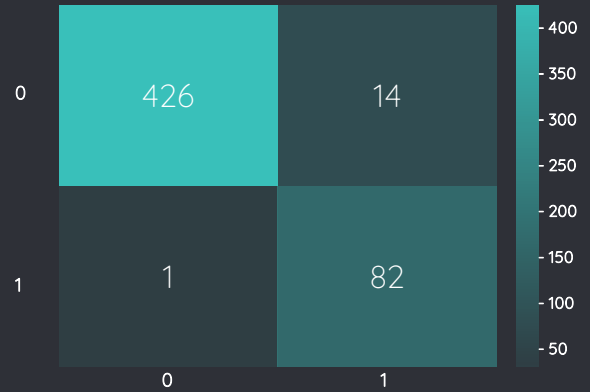
F1 score	Gs-W	Gs-T	F1-W	DT	LR
Standard	0.84	0,83	0,83	0,79	0,80
Advanced (19-3)	0,73	0,85	0,81	0,73	0,78
Advanced (10-2) (only A1,2)	0,92	0,85	0,91	0,90	0,90

BEST MODEL:
linear combination of A1,2 on A1 (10-2)

Advanced (10-2): wA1: 0,21 wA2: 0,79 AUPRC: 0.83

	F1 score	Recall	Precision
Advanced (10-2) (only A1,2)	0,92	0,99	0,85





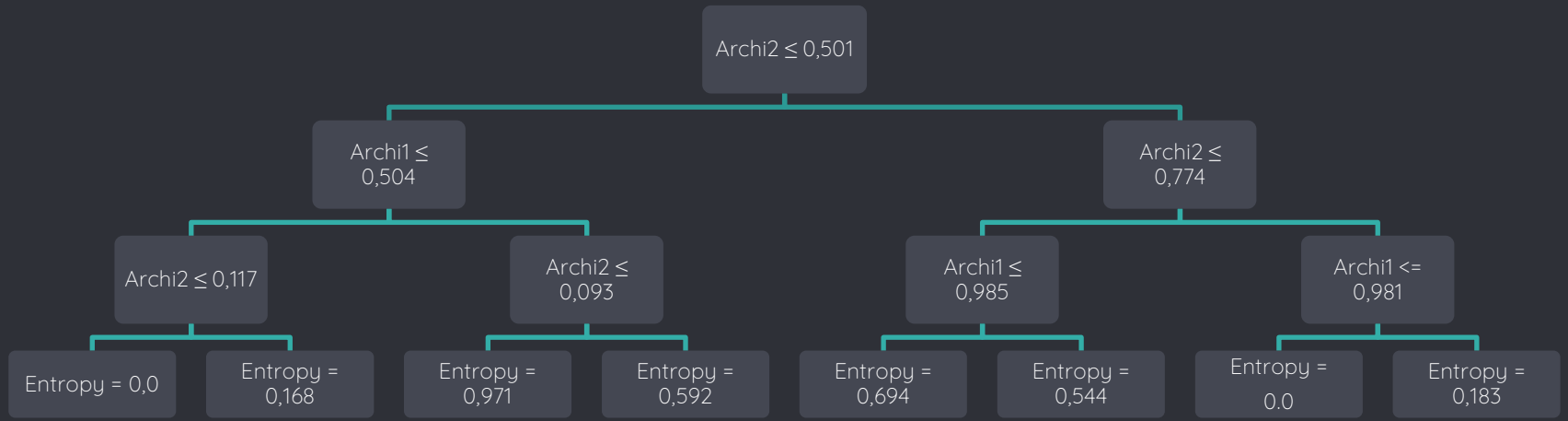
BEST ENSEMBLE ANALYSIS

F1 based: f1 A1: 0,86 f1 A2: 0,90

A (10-2)

Logistic regression: wA1: 3,34 wA2: 6,56

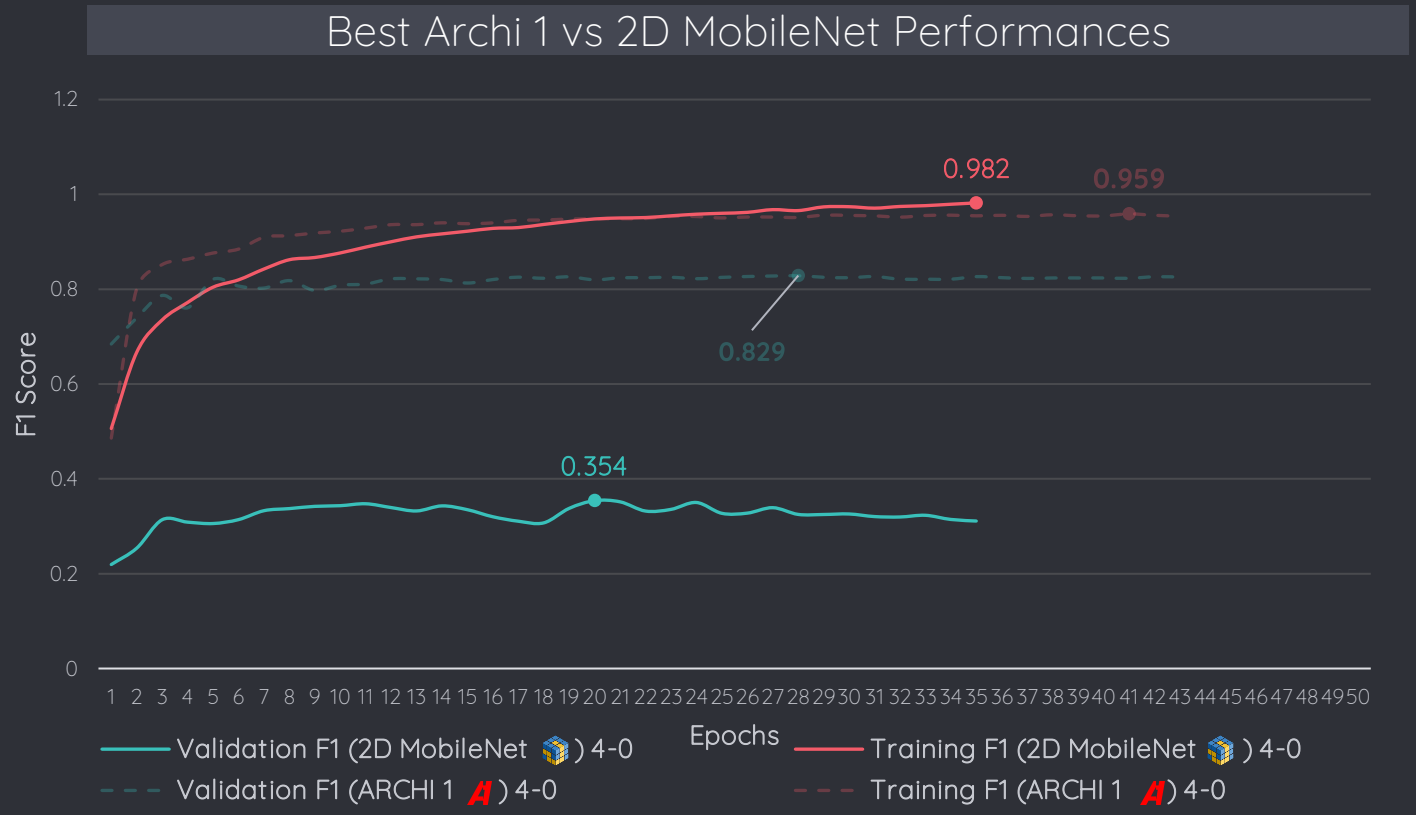
Decision tree:



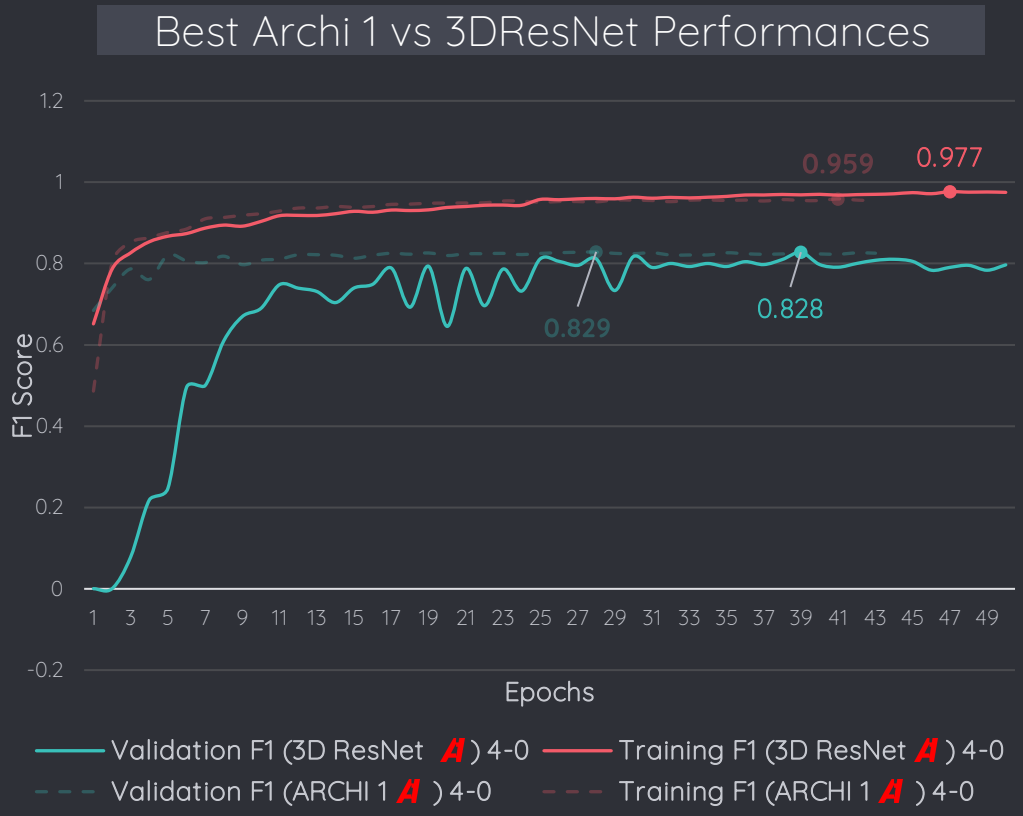
6 CONCLUSIONS

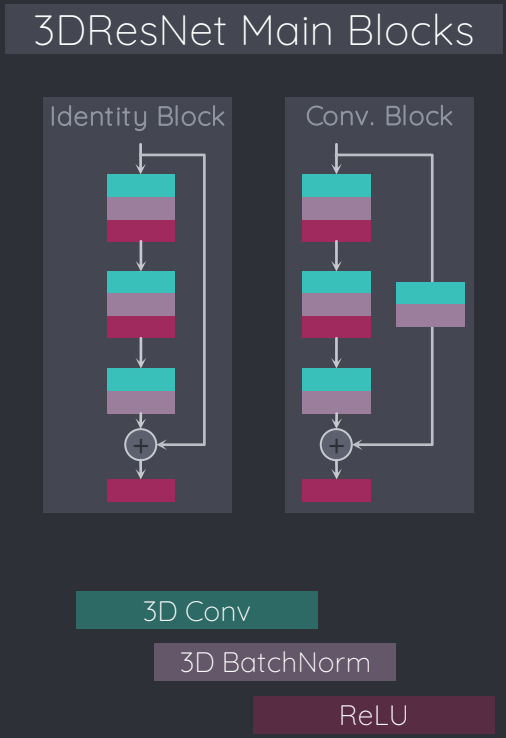
- Balanced loss instead of data augmentation → Low data variability
- Augmentation → Best results
- Multimodal networks → Too many parameters

2D Network – MobilNet (Transfer Learning & Fine Tuning)



3D Network – Custom 3D ResNet



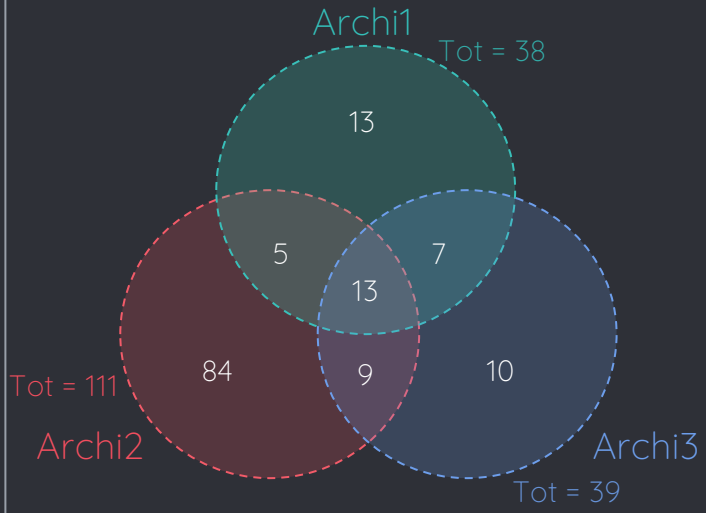


THANK YOU FOR YOUR ATTENTION!

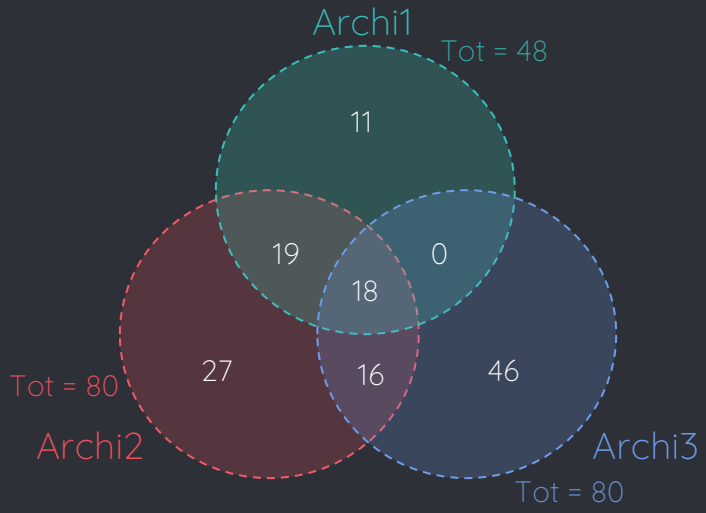
ADDITIONAL RESULTS

DIFFERENCES IN WHAT EACH ARCHI LEARNS:

Standard Dataset ( ):

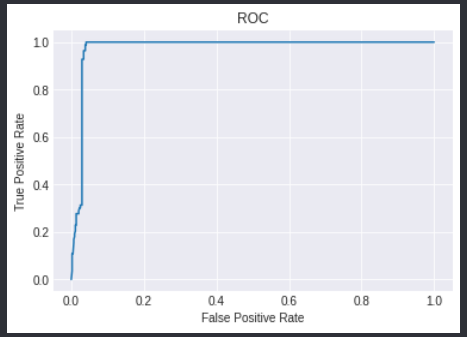


Advanced Dataset ( ):

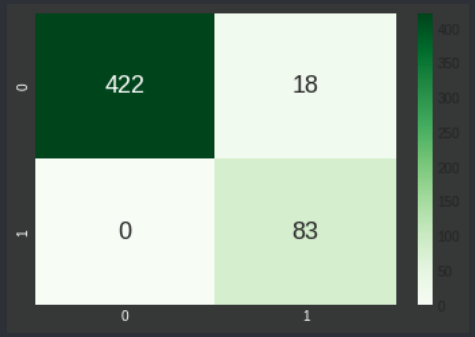


Number of common missclassified samples for Archi 1 – 2 - 3

MULTIMODAL WITH PARALLEL AND COMMON CONVOLUTIONS

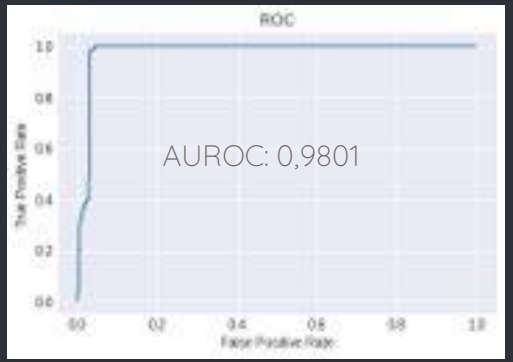


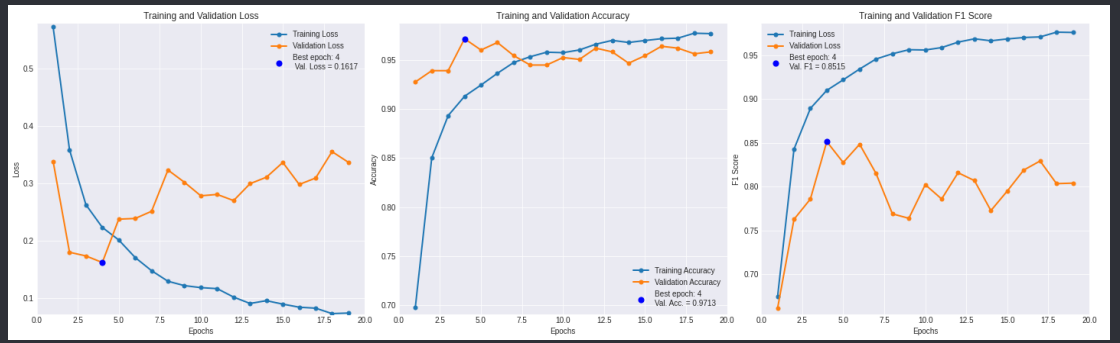
AUROC: 0,9766



MULTIMODAL WITH PARALLEL CONVOLUTIONS AND COMMON CLASSIFIER

	F1 score	Recall	Precision
Standard	0.816	0,941	0,736
Advanced (10-2)	0,852	0,929	0,807
Improvement	+ 3,6%	-1,2%	+ 7,1%





GRIDSEARCH PREDICTION THRESHOLD FOR BEST F1SCORE

Standard: thr: 0.43

Advanced (19-3): thr: 0.66

Advanced (10-2): thr: 0.36

	F1 score	Recall	Precision
Standard	0.83	0,90	0,77
Advanced (19-3)	0,85	0,87	0,83
Advanced (10-2) (only A1 and A2)	**0,85**	**0,87**	**0,83**

WEIGHT EACH ARCHI BASED ON F1SCORE FOR SEPARATE VALIDATION SET

Standard: f1 A1: 0,80 f1 A2: 0,33 f1 A3: 0,76

Advanced (19-3): f1 A1: 0,81 f1 A2: 0,75 f1 A3: 0,66

Advanced (10-2): f1 A1: 0,86 f1 A2: 0,90

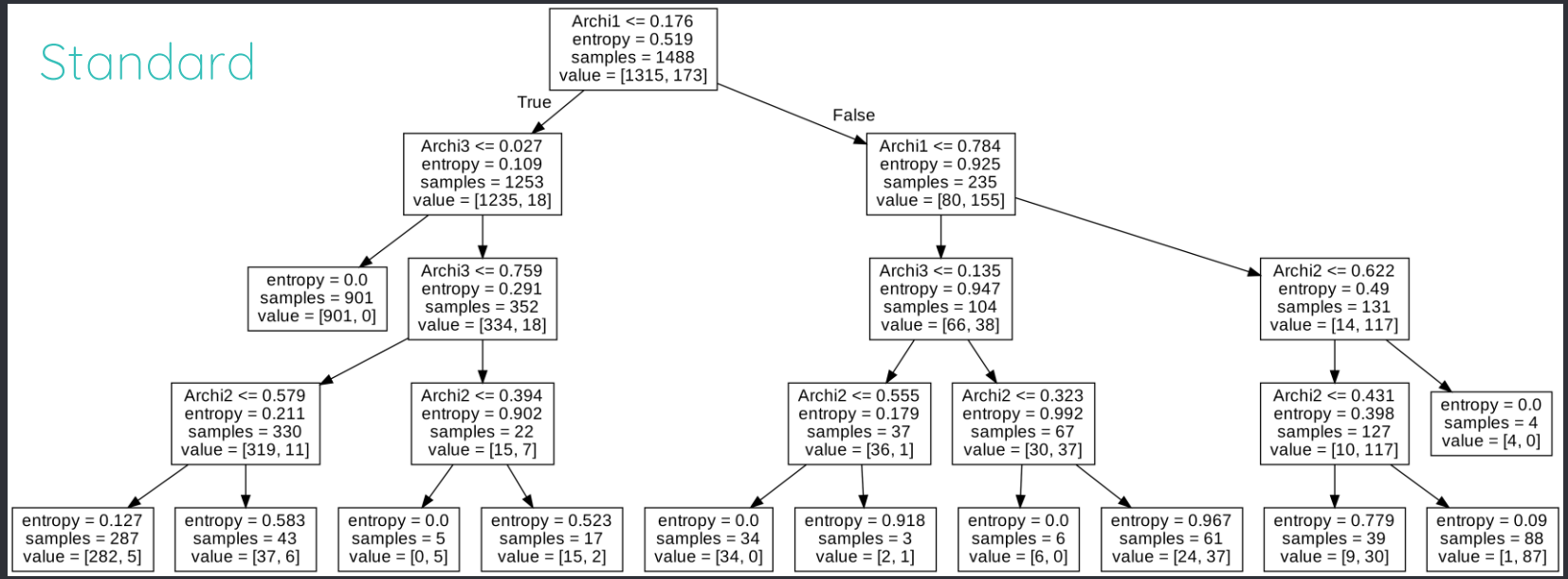
	F1 score	Recall	Precision
Standard	0.83	0,88	0,79
Advanced (19-3)	0,81	0,99	0,69
Advanced (10-2) (only A1 and A2)	0,91	0,98	0,85

APPLY DECISION TREE TO ALL ARCHIS OUTPUTS (INTERPRETABLE)

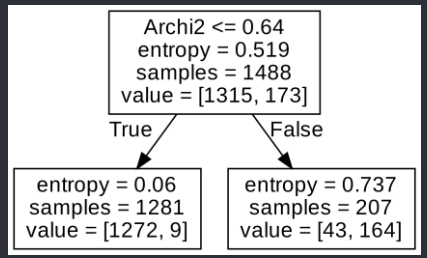
	F1 score	Recall	Precision
Standard	0.79	0,90	0,71
Advanced (19-3)	0,73	0,96	0,58
Advanced (10-2) (only A1 and A2)	0,90	0,96	0,85

APPLY DECISION TREE TO ALL ARCHIS OUTPUTS (INTERPRETABLE)

Standard

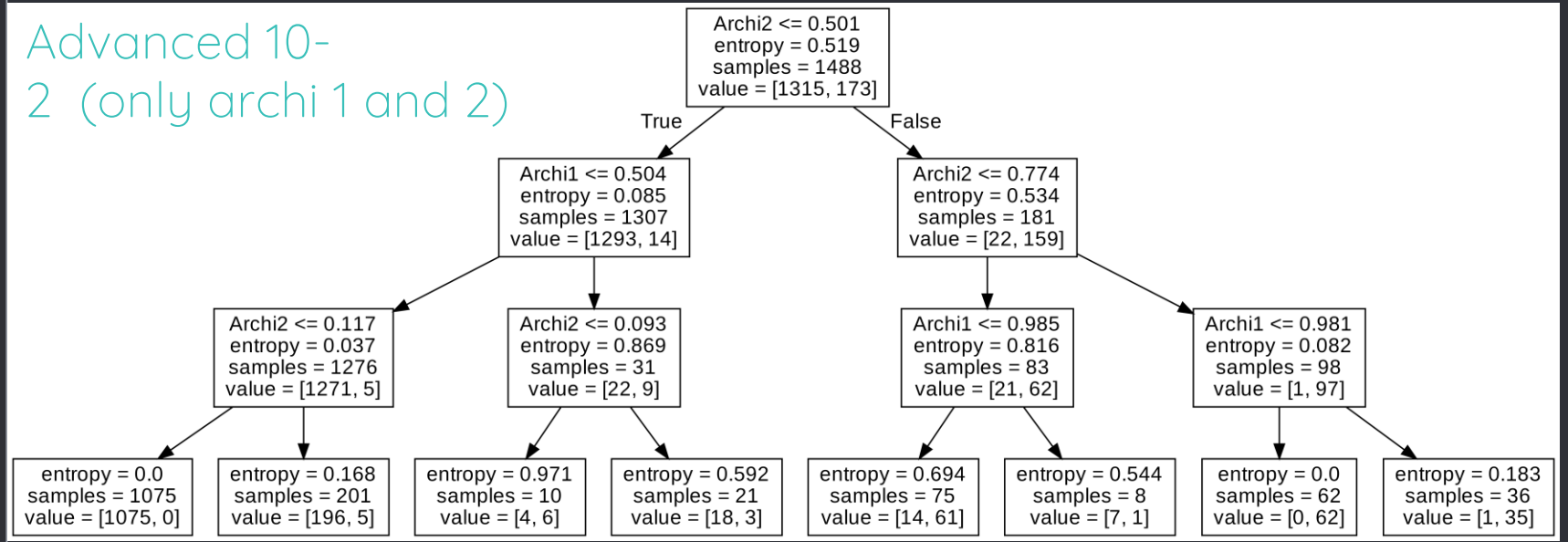


APPLY DECISION TREE TO ALL ARCHIS OUTPUTS (INTERPRETABLE)



Advanced 19-3

Advanced 10-2 (only archi 1 and 2)



APPLY LOGISTIC REGRESSION TO ALL ARCHIS OUTPUTS (INTERPRETABLE)

Standard: wA1: 4,02 wA2: 4,47 wA3: 2,76

Advanced (19-3): wA1: 1,76 wA2: 5,48 wA3: 3,04

Advanced (10-2): wA1: 3,34 wA2: 6,56

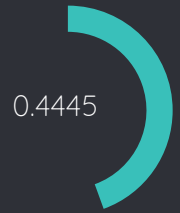
	F1 score	Recall	Precision
Standard	0.80	0,77	0,82
Advanced (19-3)	0.78	0,73	0,82
Advanced (10-2) (only A1 and A2)	0,90	0,96	0,85

2D NETWORK: MOBILENET

- TRANSFER LEARNING AND FINE TUNING FROM IMAGENET WEIGHTS USING THE 3 CENTRAL SLICES

F1 score

Recall



Precision