Transfer Learning for ECG Classification using PTB-XL

CSC2541 COURSE PROJECT

NIKHIL VERMA
OMKAR DIGE
DEEPKAMAL GILL

DATE: 03 DEC 2021

Agenda

- Introduction
- Transfer Learning
- Related Work
- Data Summary
- Methodology
- Experiments and Results
- Conclusion
- Limitations and Future scope

Introduction

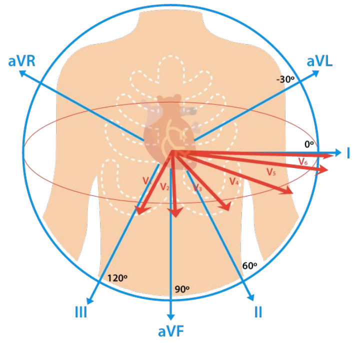
- Cardiovascular diseases account for the death of every **1 in 4 persons** in the US[1]
- An ECG (electrocardiography) records the electrical activity of heart at rest
- Provides information about
  - heart rate and rhythm
  - enlargement of the heart due to high blood pressure (hypertension)
  - evidence of a previous heart attack (myocardial infarction)
- Automated classification of ECG signals can aid in
  - Early diagnosis of heart diseases

[1]https://www.cdc.gov/heartdisease/facts.htm

Challenges with ECG Data

- Limited labeled data

- Datasets recorded at different configurations
  - Signal frequency
  - Time duration
  - Diagnostic code formats
  - Only specific leads available
    - A **12-lead ECG** paints a complete picture of the heart's electrical activity by recording information through **12** different perspectives.



Solution?

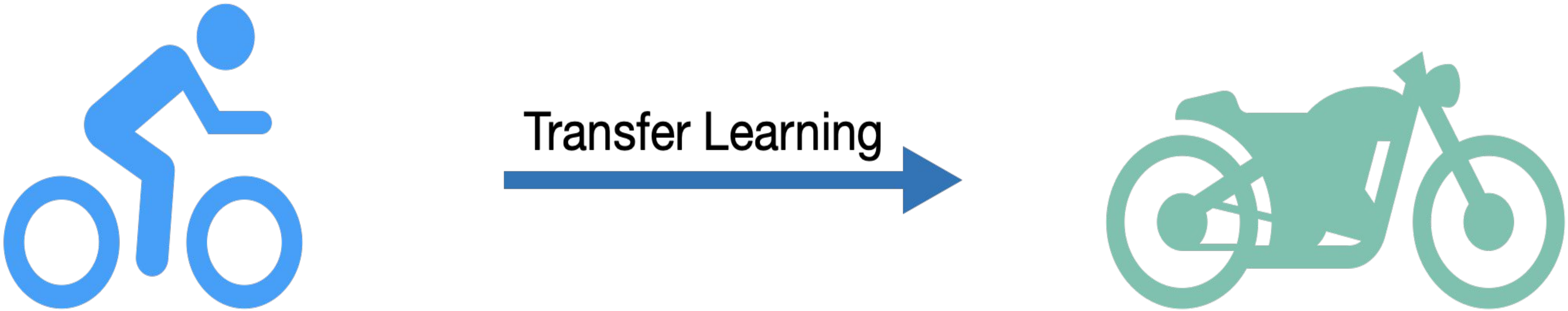
- Data Augmentation
- Semi-supervised Learning
- **Transfer Learning**

What is Transfer Learning?

Transfer learning is make use of knowledge gained while solving one problem and applying it to related problem

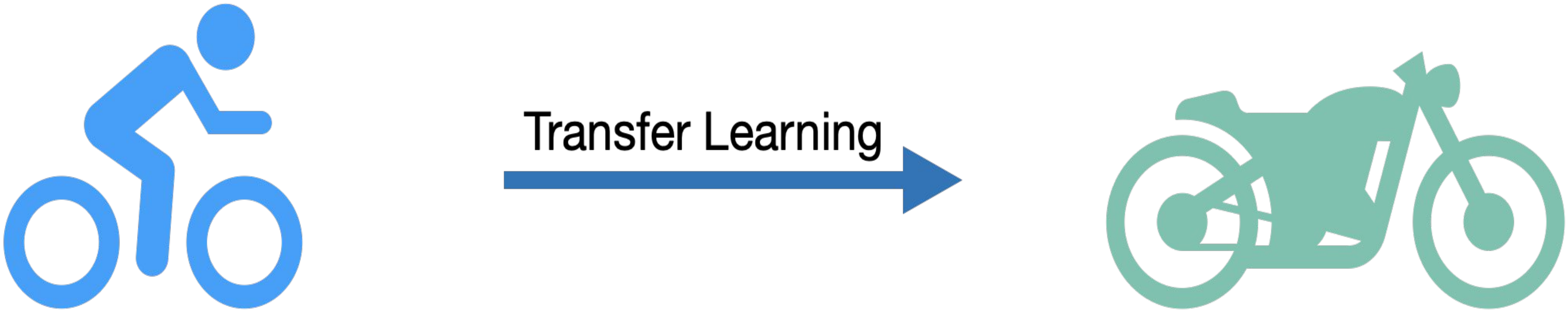
What is Transfer Learning?

Transfer learning is make use of knowledge gained while solving one problem and applying it to related problem



What is Transfer Learning?

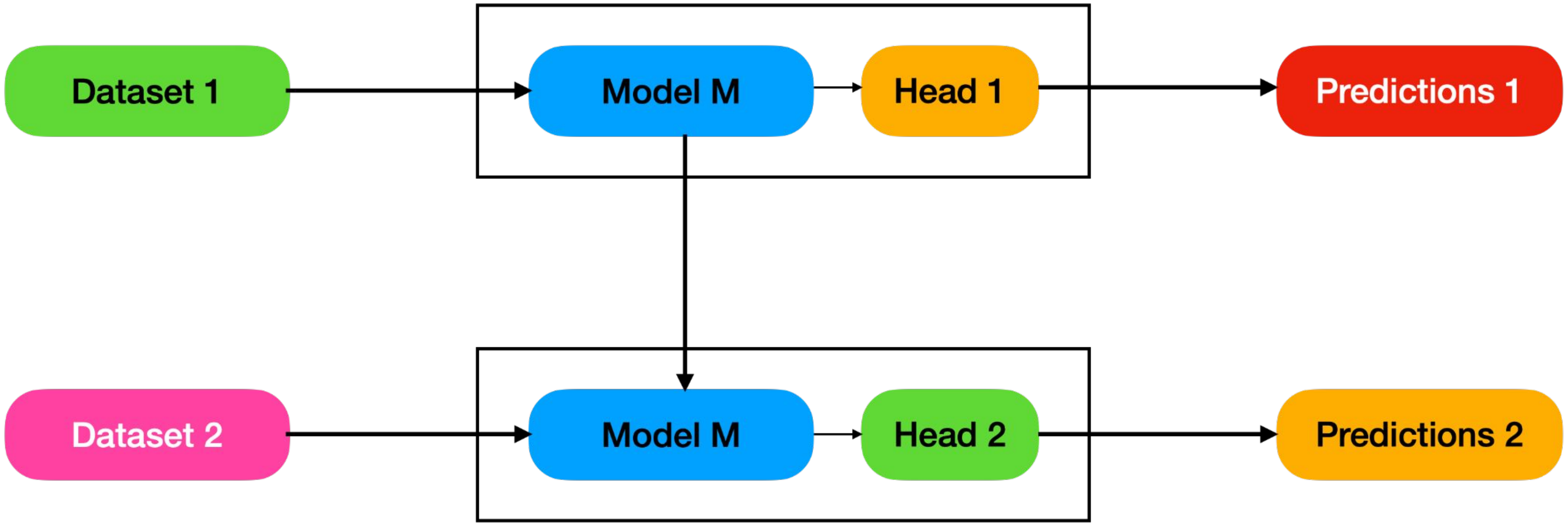
Transfer learning is make use of knowledge gained while solving one problem and applying it to related problem



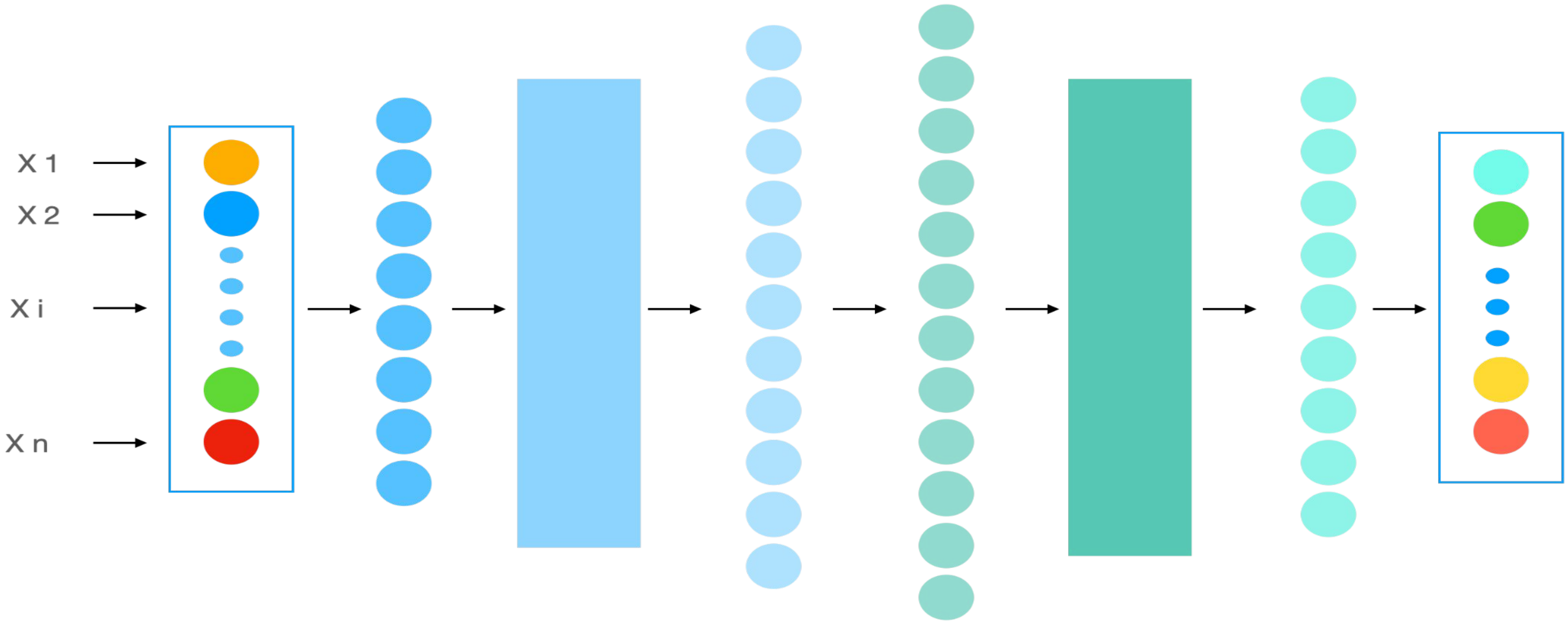
A model trained on one task/dataset is re-purposed for a second related task/dataset

What is Transfer Learning?

A model trained on one task/dataset is re-purposed for a second related task/dataset

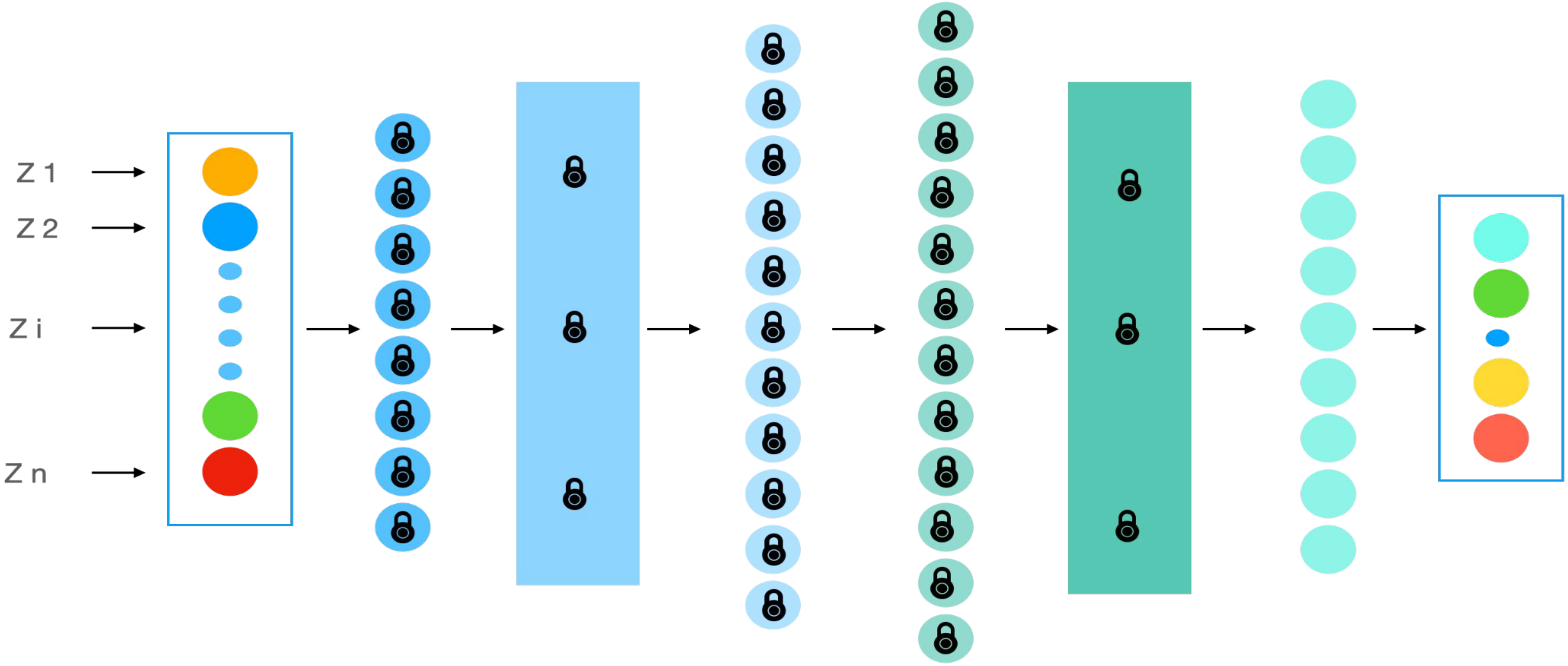


Gradual Unfreezing



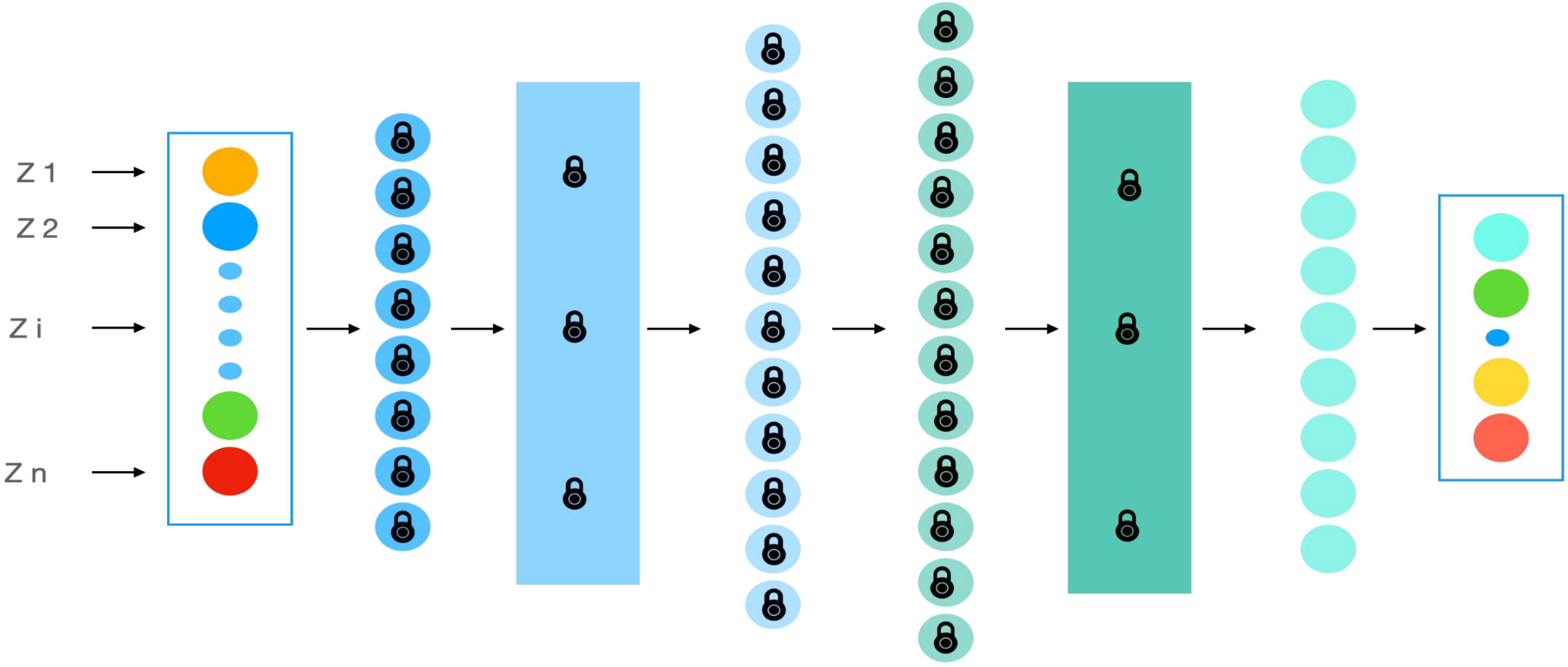
Gradual Unfreezing

Gradual Unfreezing



Gradual Unfreezing

Freezing a layer prevents its weights from being modified.



Related Work

- Most current research focuses on transfer learning for classification of specific heart abnormalities
  - Weimann et al. **"Transfer learning for ECG classification."** *Scientific reports* (2021)

- This paper hypothesized that PTB-XL, largest 12-lead ECG dataset released recently, could serve as a prospective base dataset for transfer learning
  - Strodthoff, Nils, et al. "**Deep learning for ECG analysis: Benchmarks and insights from PTB-XL**." *IEEE Journal of Biomedical and Health Informatics* (2020)

- Evaluation on data subsets to verify the potential of transfer learning
  - Jang, Jong-Hwan et al. **"Effectiveness of Transfer Learning for Deep Learning-Based Electrocardiogram Analysis."** *Healthcare informatics* (2021)

Main Idea

Goal: Test whether PTB-XL is a good dataset for transfer learning for ECG classification

	PTB-XL[1] (Base Dataset)	G12EC[2] (Evaluation Dataset)
No. of Records	21,837	10,344
No. of Leads	12	
Frequency	500 Hz	
Time Duration	10 seconds	

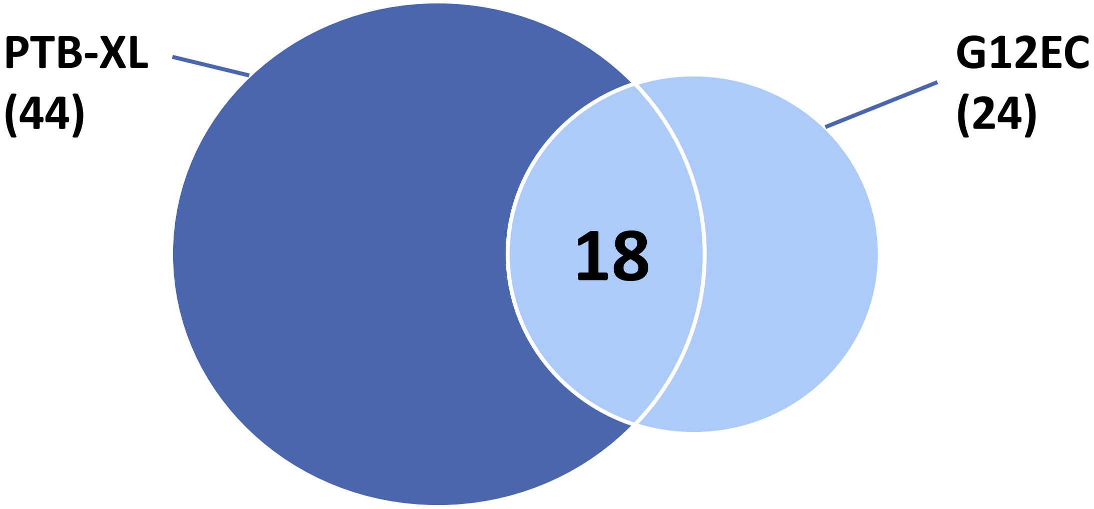
Key Contributions:

- Pre-train ECG classification model using PTB-XL
- Finetune and evaluate performance on G12EC
- Compare with the best standalone model trained on G12EC

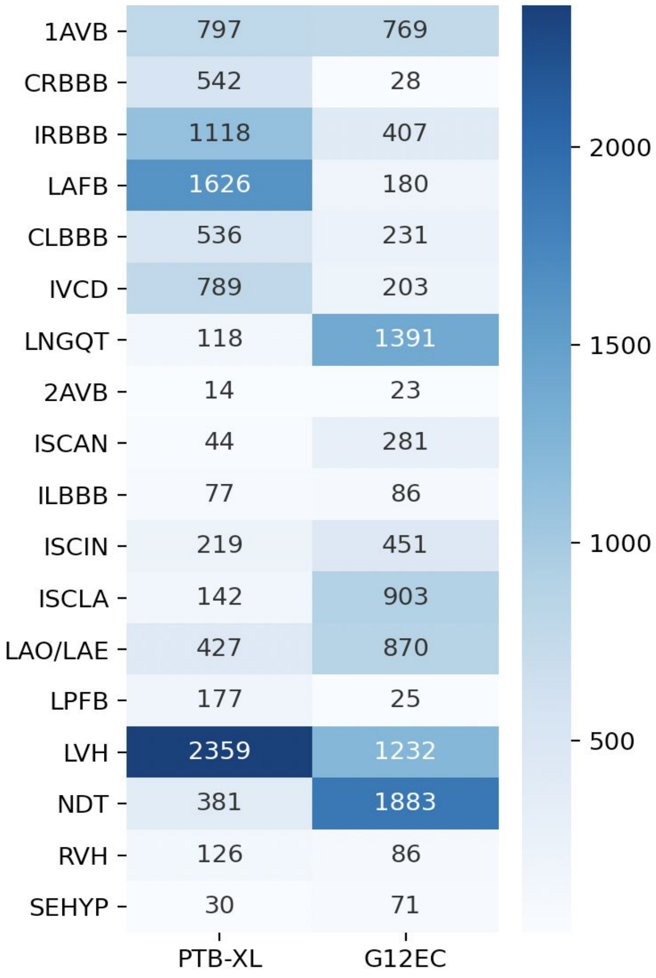
[1] https://physionet.org/content/ptb-xl/1.0.1/
[2] https://www.kaggle.com/bjoernjostein/georgia-12lead-ecg-challenge-database

Data Summary

- PTB-XL has 5 diagnostic superclasses and 44 subclasses

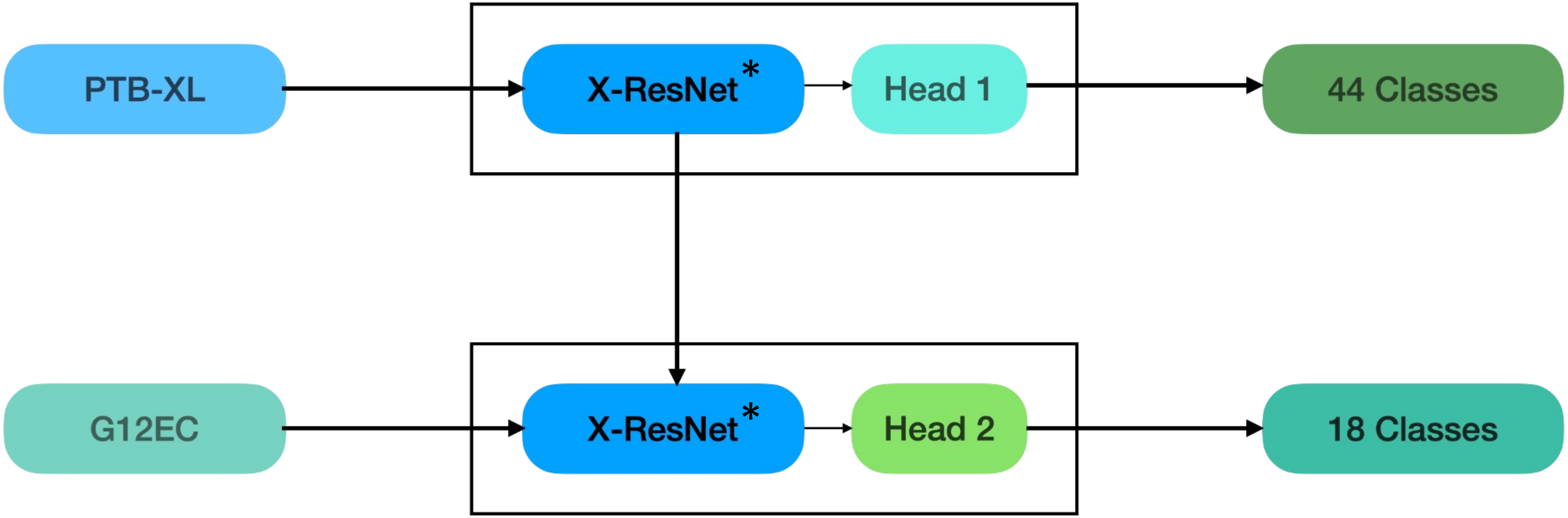


- Label distribution for both datasets is quite different
- Will transfer learning still work?

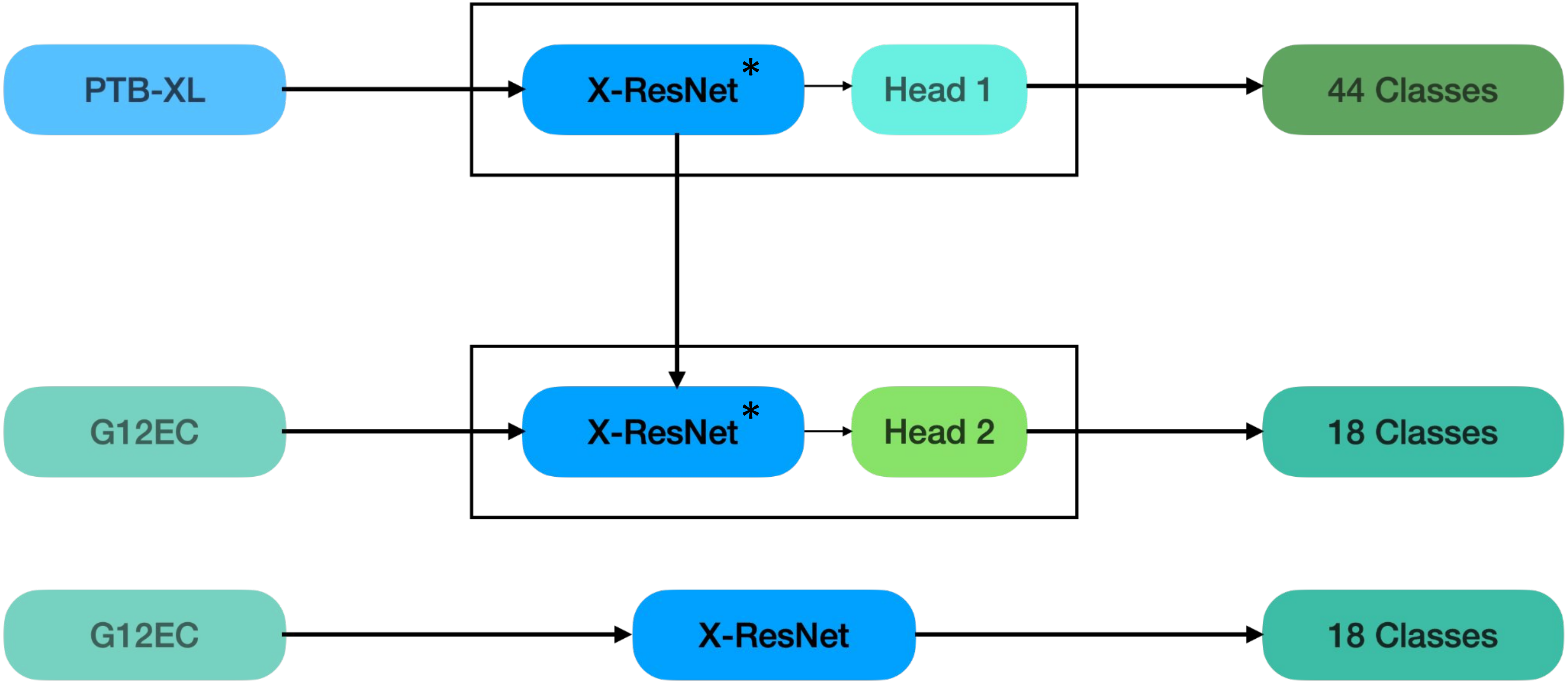


Methodology

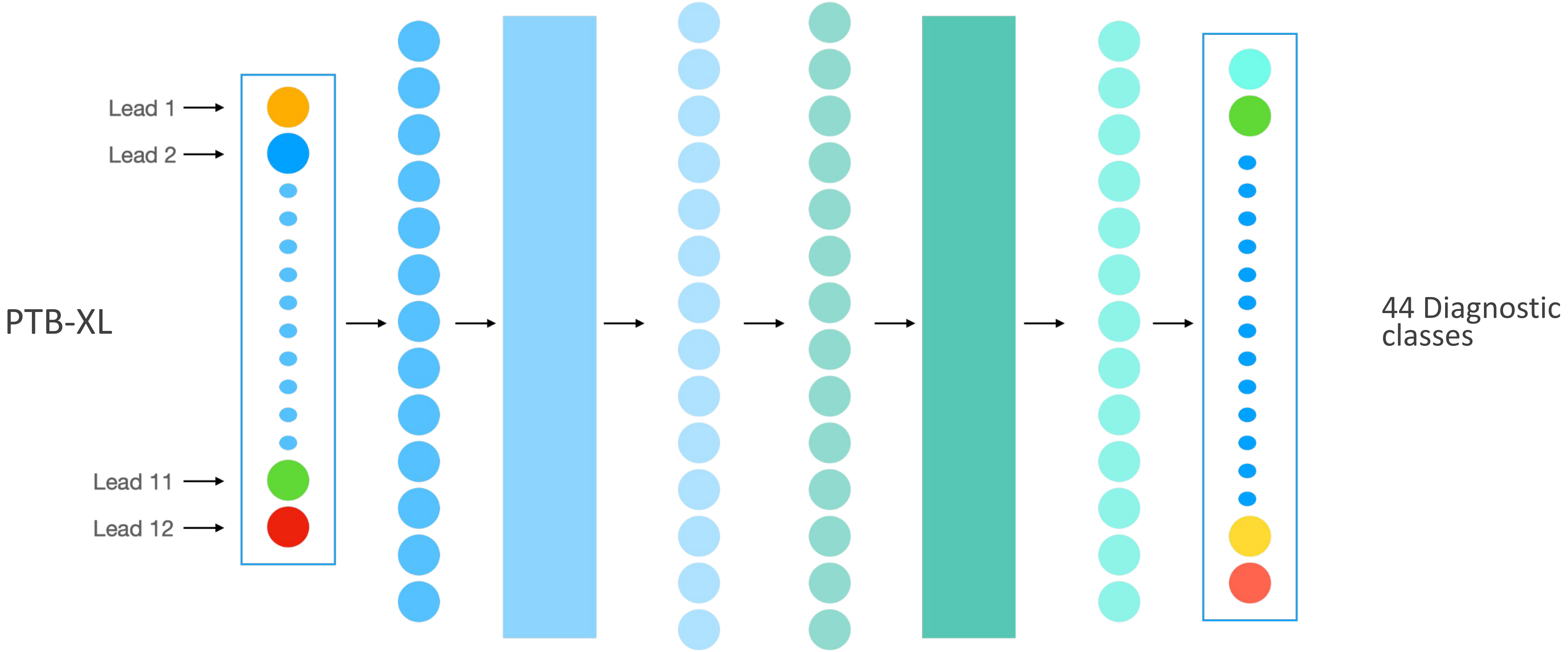
Methodology



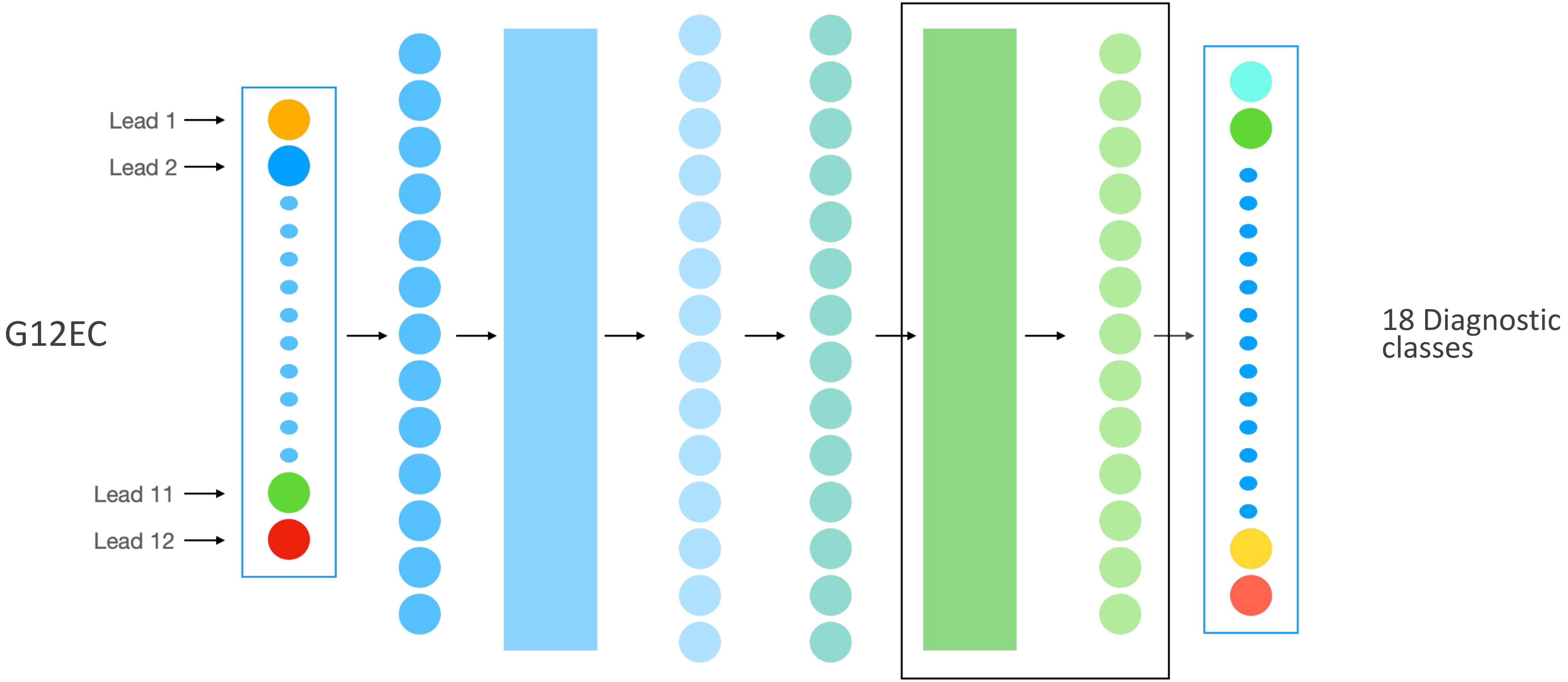
Methodology



Methodology - Pre-training on PTB-XL



Methodology - Fine-tuning on G12EC



Results

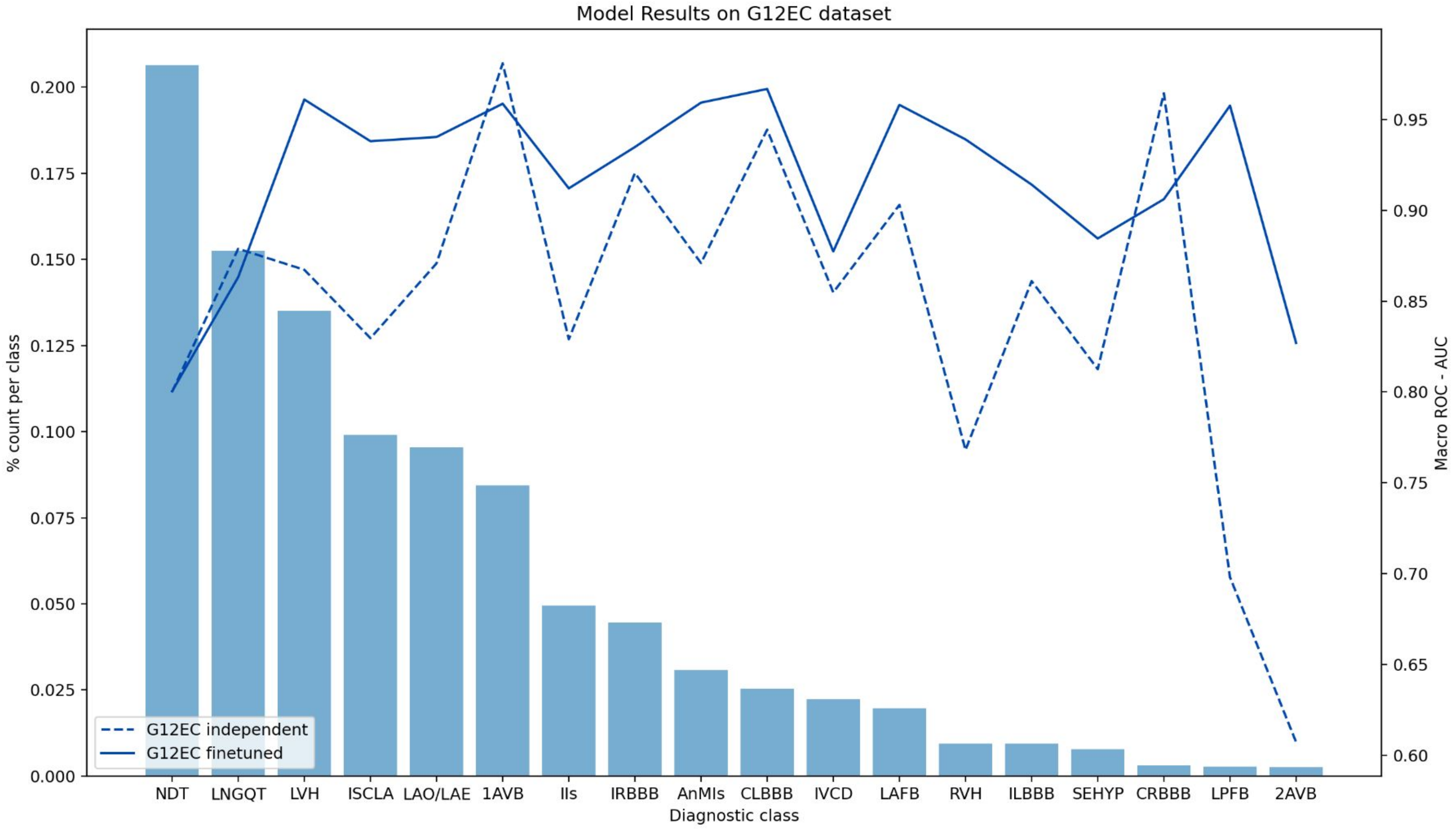
Performance of different models trained only on G12EC dataset (independent)

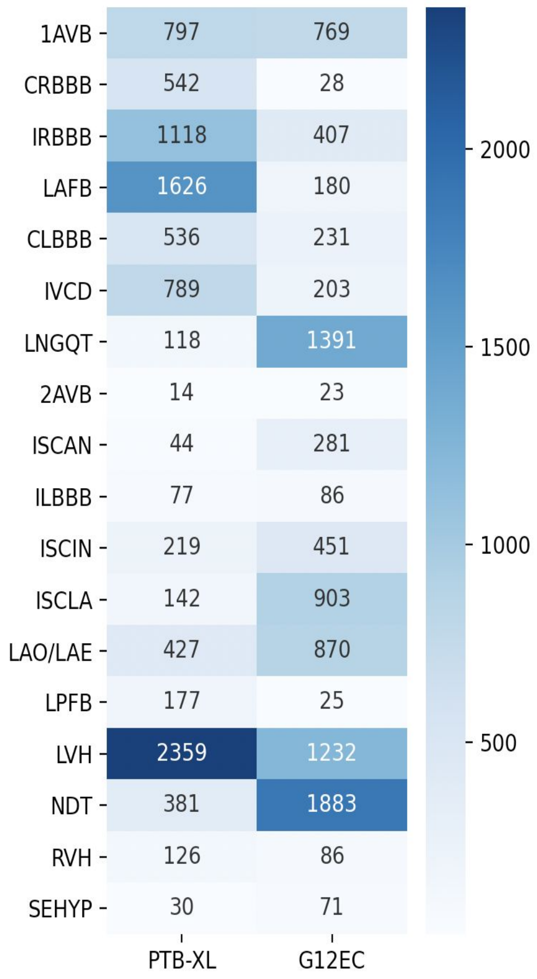
Model	Macro AUC
XResNet	**0.843**
ResNet	0.785
InceptionNet	0.685

Performance of G12EC fine-tuned using model pre-trained on PTB-XL

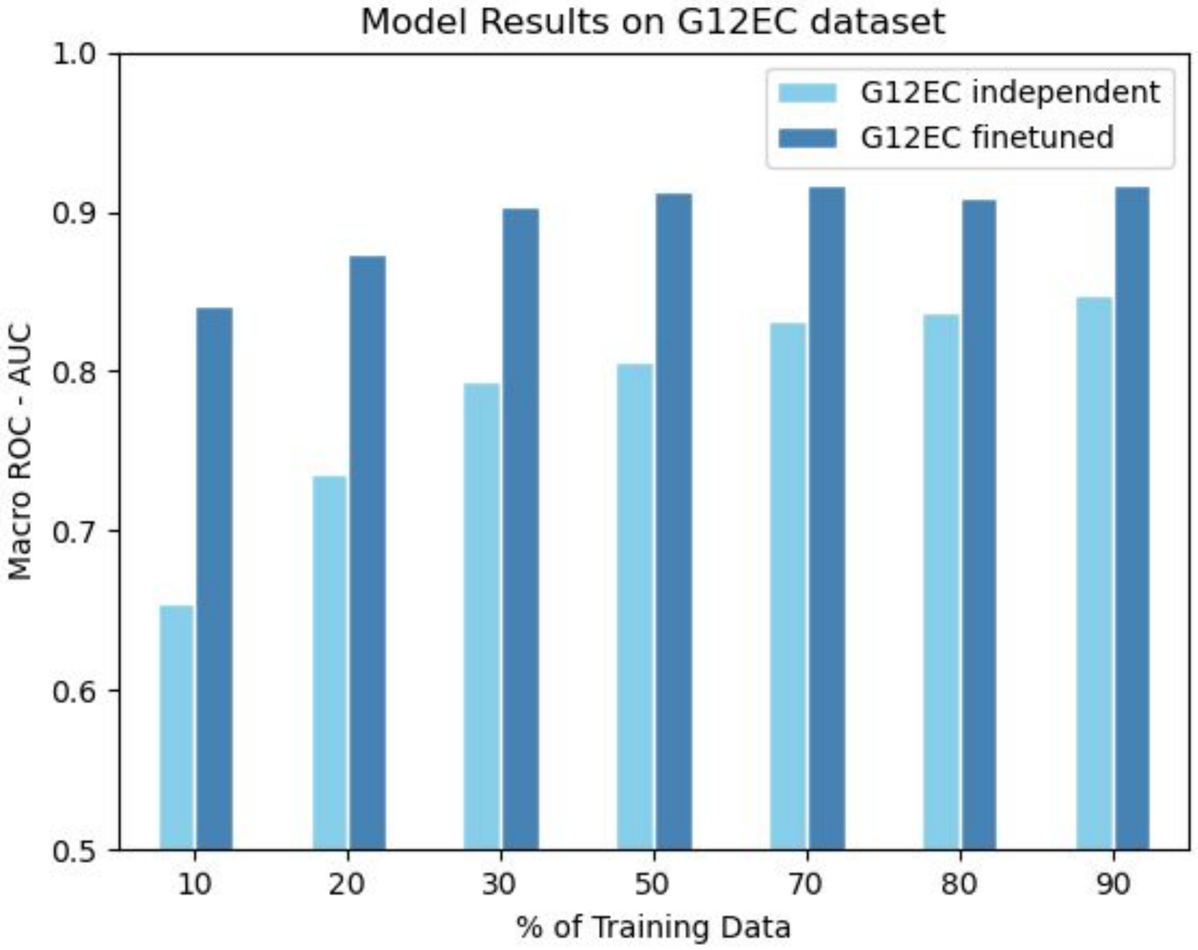
Task	Macro AUC
PTB-XL Pre-training	0.947
G12EC Fine-tuning	**0.916 (+0.073)**

Results: Class-wise performance



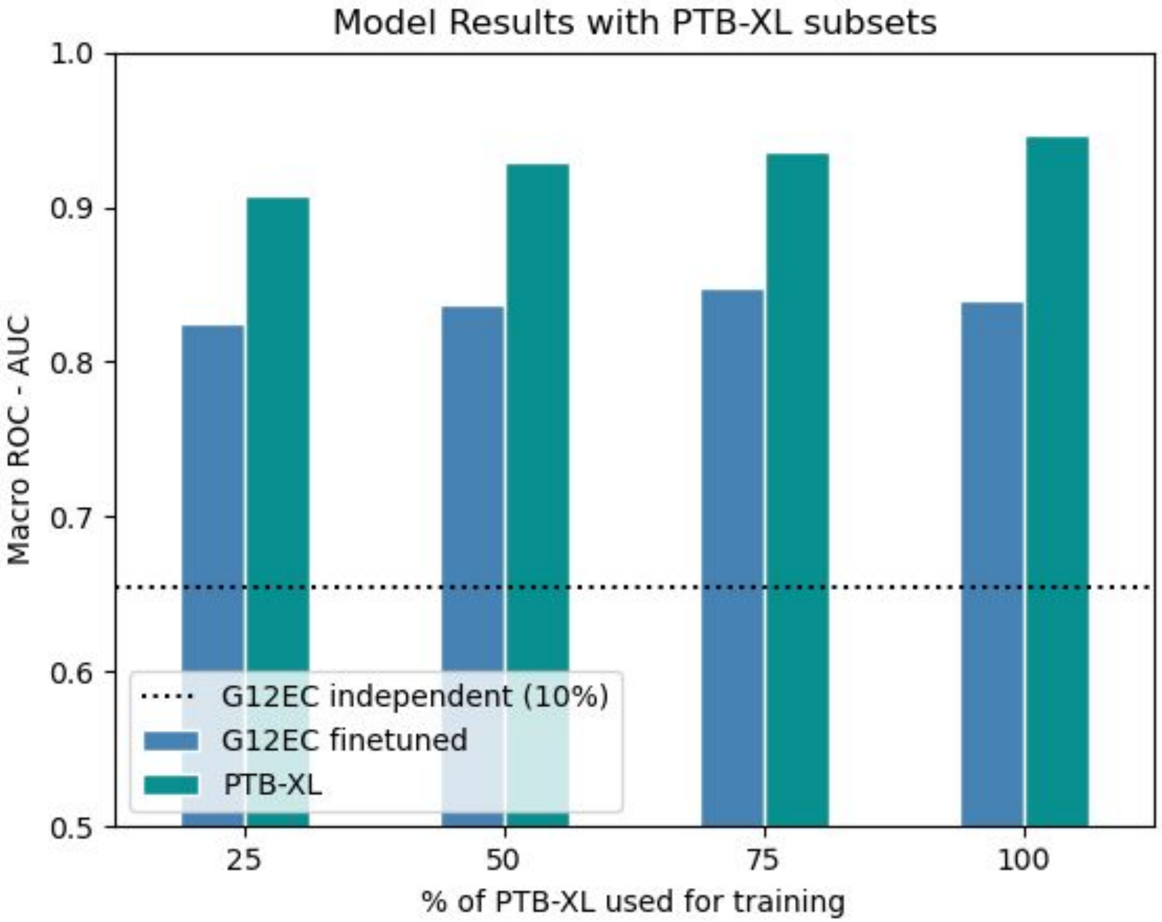


Experiment 1: Performance using G12EC subsets



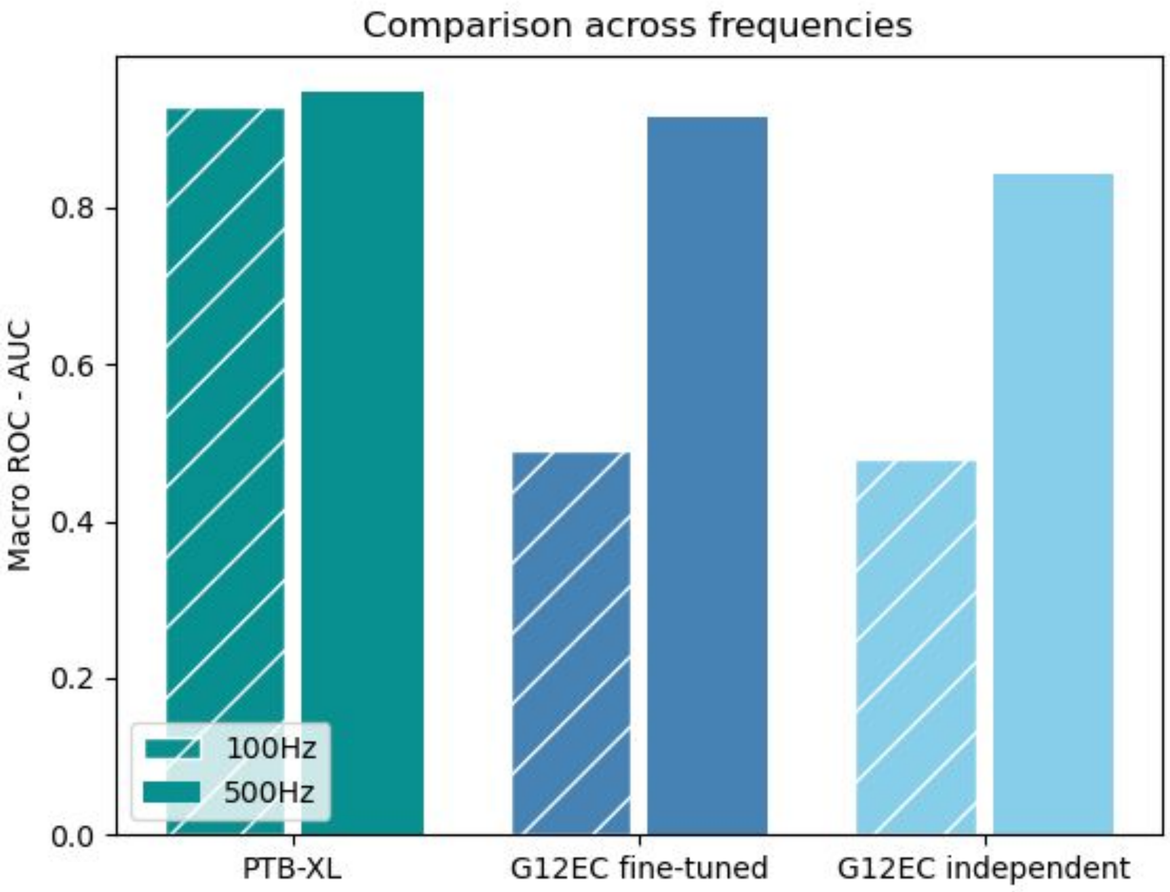
- Fine-tuned model performs consistently better than independent model
- Gap increases as training data for fine-tuning is reduced
- Test AUC of fine-tuned model only drops by ~0.07 when the training data is reduced to 10%

Experiment 2: Performance by PTB-XL subsets



- Used 10% of G12EC for a fair comparison
- As expected model performance drops with decrease in data used for pre-training
- Performance of fine-tuned model is better even after pre-training on 25% PTB-XL data

Experiment 3: Frequency downsampling

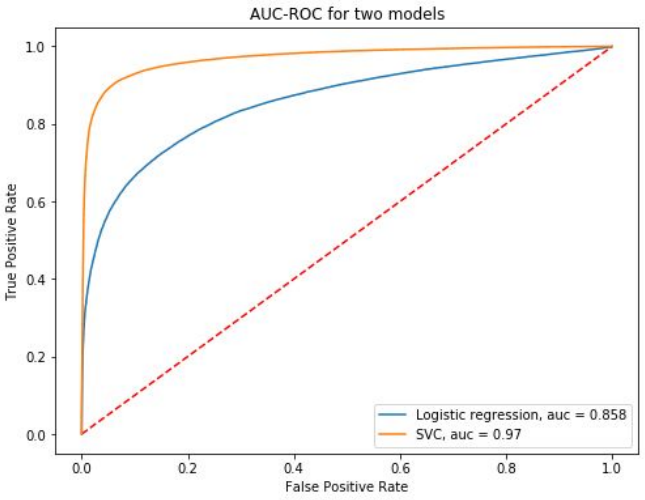


- PTB-XL performance remained almost same on lower frequency
- G12EC performance dropped significantly
- G12EC loses significant information at lower frequencies

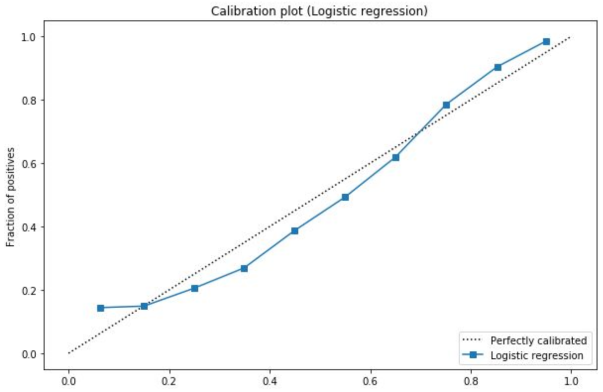
Experiment 4: Calibration

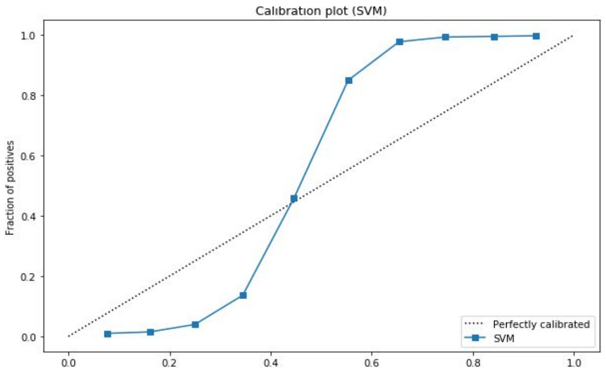
What is calibration?

A model is perfectly calibrated if, for any probability value p, a prediction of a class with confidence p is correct 100*p per cent of the time

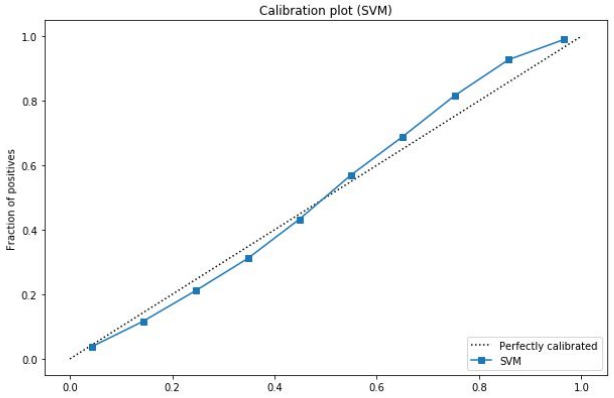


Logistic regression Vs SVM AUC-ROC curves

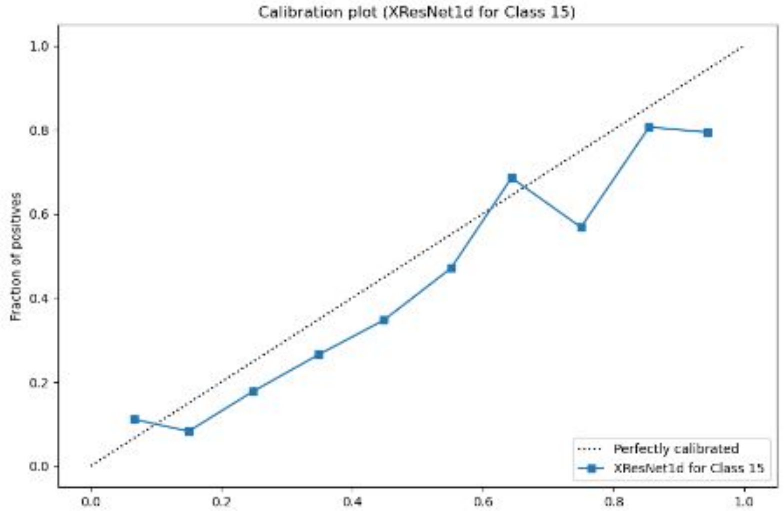


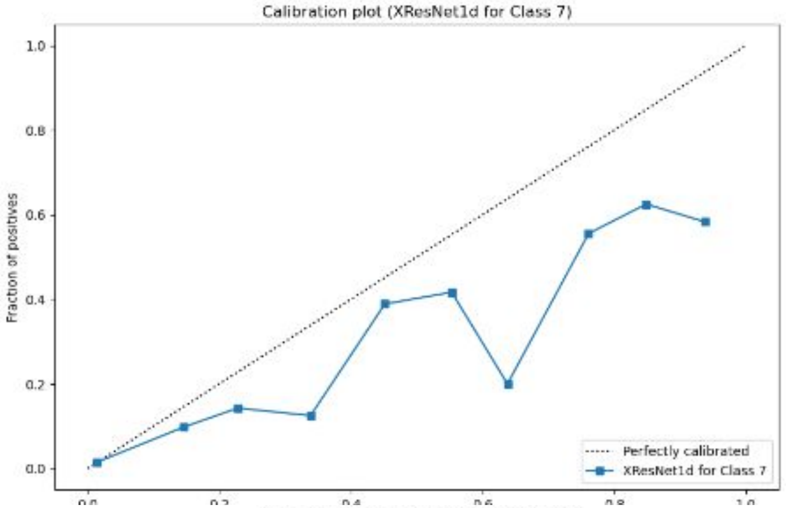


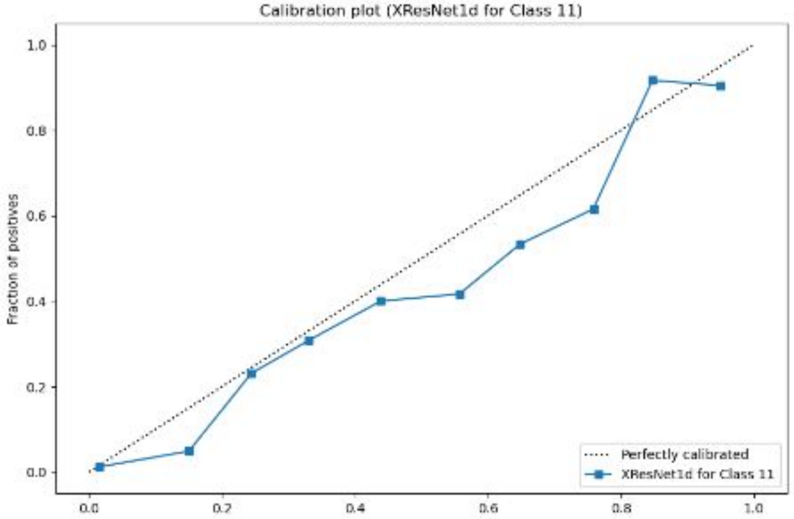
Platt’s Scaling

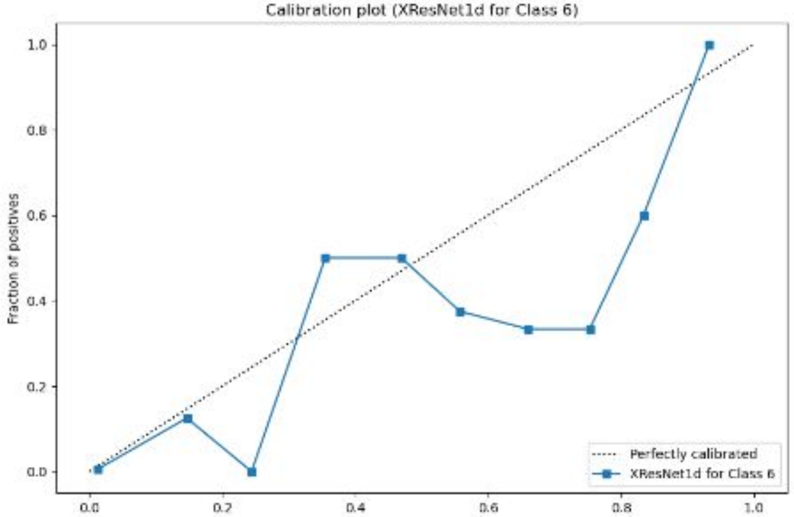


Experiment 4: Calibration (WIP)









Experiment 5: Frequency Domain Analysis

Component	PTB-XL	G12EC Fine-tuned
Absolute	0.668	0.555
Real	0.664	0.508
Imaginary	0.614	-

- Macro AUC on PTB-XL model drops significantly
- This effect on performance propagates to the fine-tuned model on G12EC
- Further work is required to understand the reasons behind the drop in model performance

Conclusion

- PTB-XL is a good dataset for transfer learning: fine-tuning achieves ~7% increase in performance over independently trained model
- Even with different label distributions the transfer learning performed well
- Existing literature focuses on a single or a certain class of diagnosis
- Implication: milestone for classification on thousands of smaller datasets currently available

Limitations / Future Work

- Model might not perform well in cases when labels for both datasets are different
  - Use techniques to match label distribution which could potentially improve finetune performance

- Performance drops significantly on lower frequency data
  - Find techniques to preserve most useful information in lower frequencies

- Current datasets were quite similar, can test the technique on diverse datasets

- Test if information from one lead or a particular waveform contributes more to the performance

- Improve the underlying architecture that can
  - Learn from periodic patterns in ECG signals
  - Demonstrate invariance to frequency changes

Thank You

Questions?