DeepECG Project

Main Project IV

2020-2021

DeepECG Project Summary

- After DeepCAS and DeepTAS projects, I led the DeepECG project (total 4 people) which detects heart diseases such as heart failure, myocardial infarction, and arrhythmia by utilizing ECG.
- Our team assumed that it was worth knowing whether heart disease will occur or what kind of heart disease have by using ECG signals. To do this, I put my effort into this project by defining study design, managing the dataset, and designing an entire framework. This project was mainly collaborated with Ajou University.
- To obtain outperform performance, I applied several novel technologies to our model. I tried to make a variety of data transformations, implement a variety of experiments for our baseline models, handle long-range ECG signals (about 48 hours), and utilize unlabeled data based on a pre-training approach.
- I also contributed to creating an ECG measuring device, which collects people's ECG data and will provide proper information about their heart-related disease.
- Now, this project was verified by the Korean FDA. In addition to this development, I published several papers and patents for this project.
- Based on this experience, I not only contributed to developing our ECG model, but also I understood the perspective to combine diverse fields with ML via a macro mindset.

DeepECG Project : Part Lead
Summary

- **Contribution**
  - Develop heart disease early prediction/classification model
    - Study design, data processing, architecture design, and evaluation in ECG data domain
- **Development**
  - Python, Tensorflow 2, Pytorch
- **Issues**
  - Data imbalance, preprocessing issues (filtering, centering, etc), and model design
  - Entire work to proceed an early stage of business development
- **Achievement**
  - 4th Int. Conf. of the ESC Council on Stroke: Stroke patient prediction (co-first))
  - Create a prototype to start B2C service (device UI/UX, hardware)

DeepECG Project
Introduction

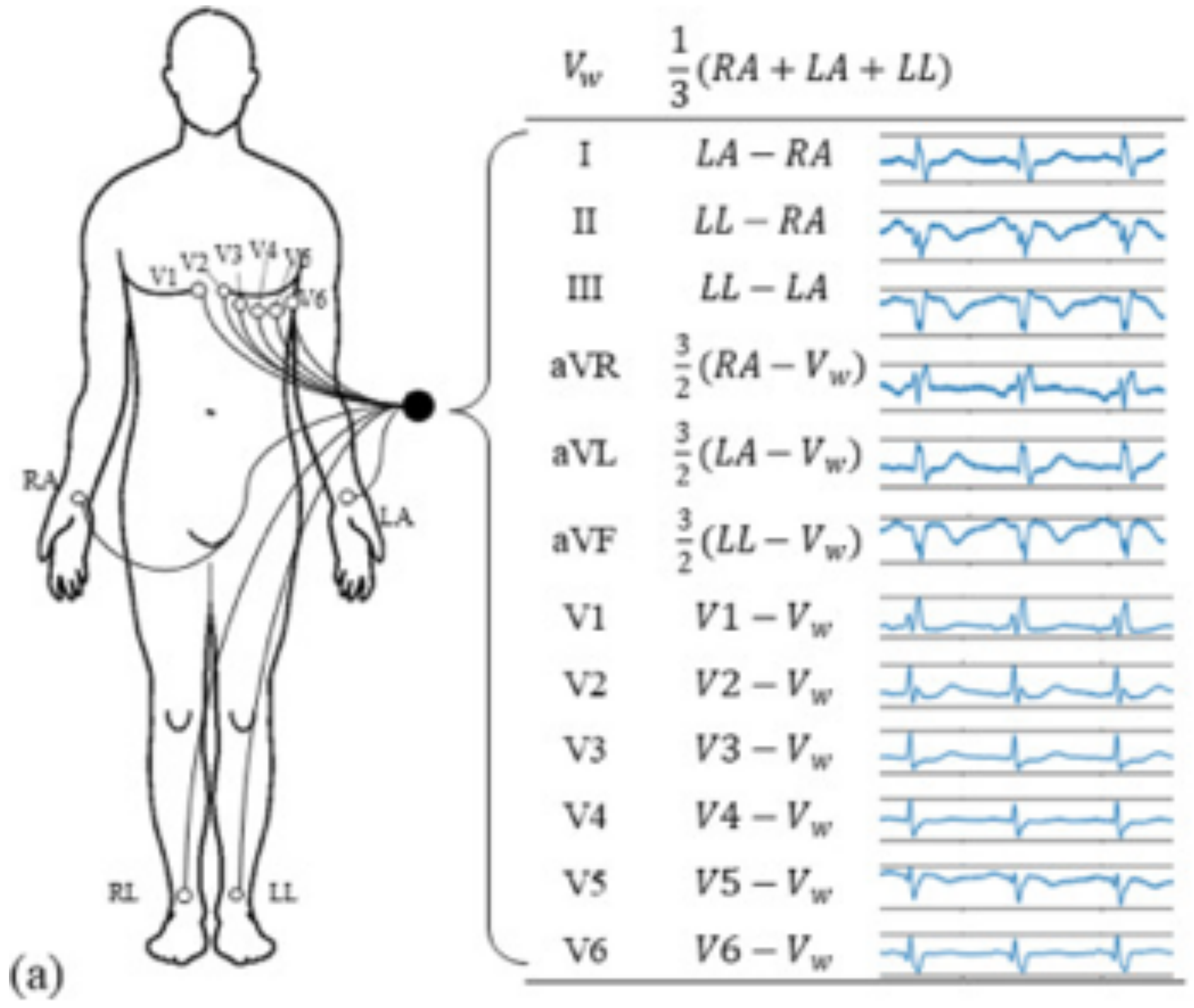
Input & Output definition

- Input: Continuous time-series data (mainly 12-lead ECG data)
- Output: Heart diseases (Arrhythmia, Myocardial Infarction, Heart Failure, ..) predict/classify

Motivation

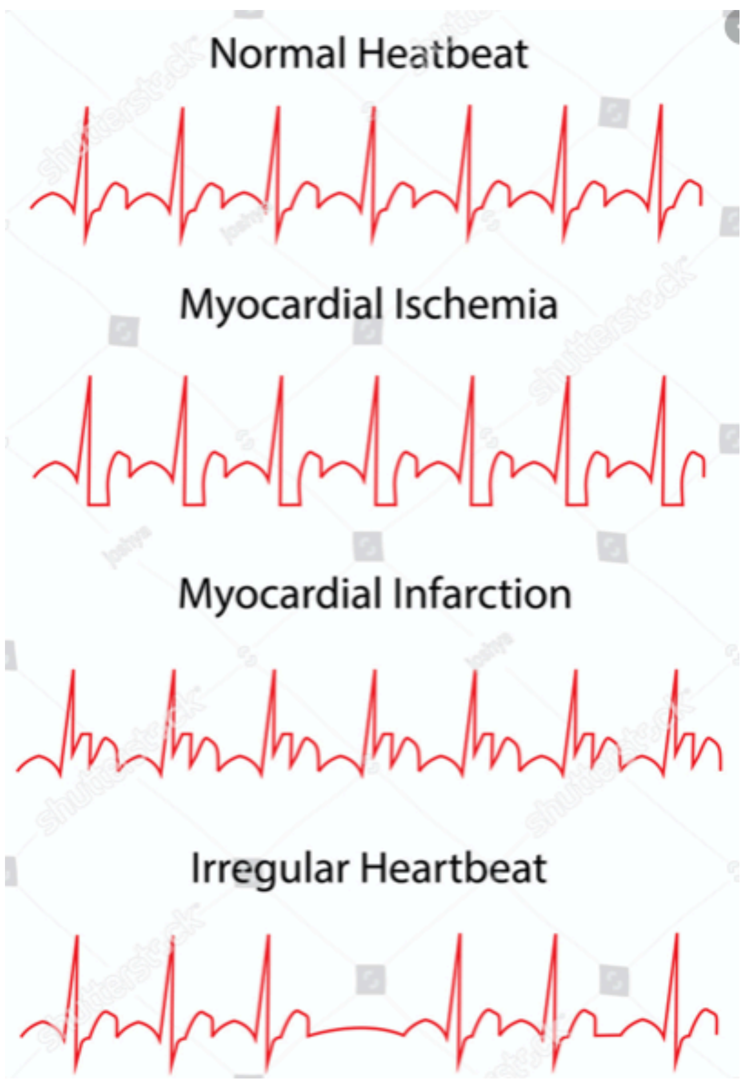
- Developing an intelligent AI-based medical solution for utilizing ECG signals

Input



Model

Output



DeepECG Project
Prediction model development

- First, I built an atrial fibrillation prediction algorithm based on a previous Lancet (2019) paper
    - They developed artificial intelligence (AI)-enabled electrocardiograph (ECG) to detect the electrocardiographic signature of atrial fibrillation present during normal sinus rhythm using standard 10-second, 12-lead ECG
- However, when I used this paper's method, training was not properly worked. Thus, I tried to utilize previous DeepCARS training knowledge. For example, I recalled that the construction of a reliable validation data set is one of the most important means by which to make training stable. Plus, I designed CNN framework suited for this task.
- Finally, our team predicted AFIB (event) during sinus (normal) ECG showing 0.78 AUROC, which gave a clue for future performance improvement.
- By using our baseline model, we built stable performance to provide an early prediction for arrhythmia. Ongoing progress for performance development based on previous research experience.

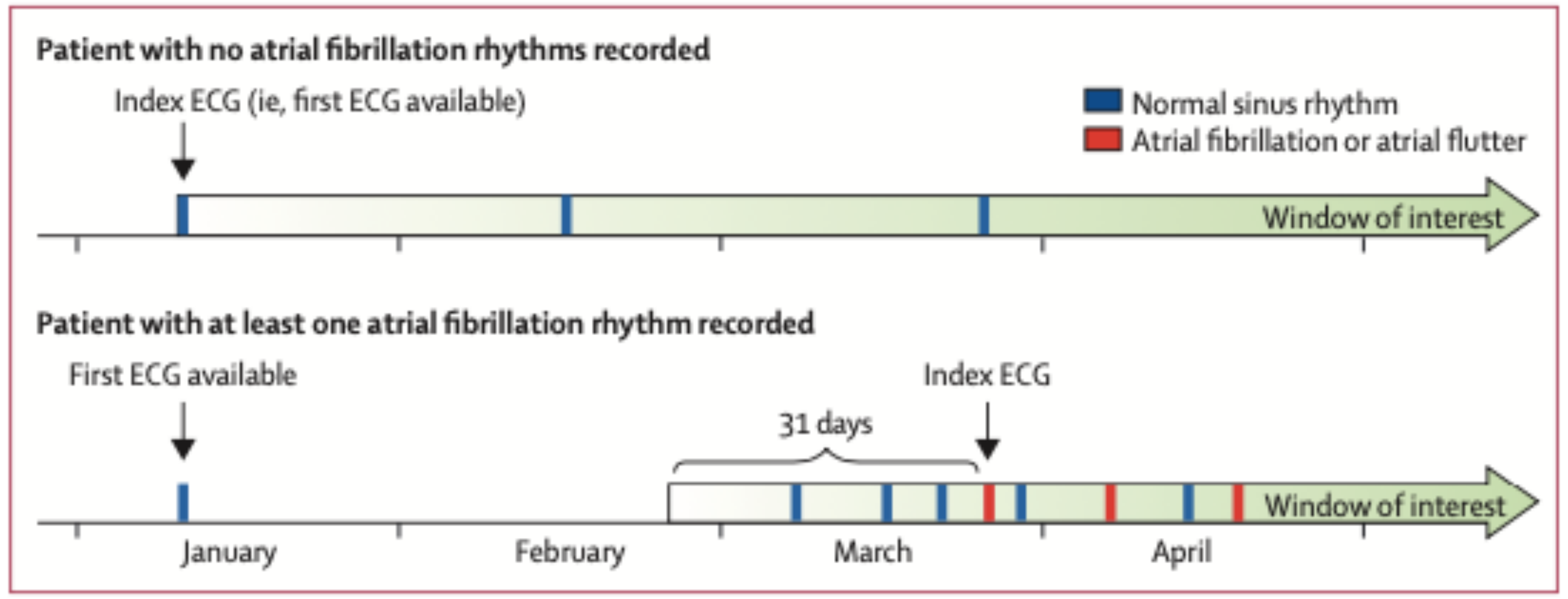


Figure 1: **ECG selection and windows of interest for patients with multiple ECGs**

An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction, 2019 Lancet

frontiers | Frontiers in Cardiovascular Medicine

ORIGINAL RESEARCH
published: 08 April 2022
doi: 10.3389/fcvm.2022.865852

Check for updates

Evaluating the Risk of Paroxysmal Atrial Fibrillation in Noncardioembolic Ischemic Stroke Using Artificial Intelligence-Enabled ECG Algorithm

OPEN ACCESS

Changho Han [1†], Oyeon Kwon [2†], Mineok Chang [2], Sunghoon Joo [2], Yeha Lee [2], Jin Soo Lee [3], Ji Man Hong [3], Seong-Joon Lee [3] and Dukyong Yoon [1,4,5*]*

[1] Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Yongin, South Korea, [2] VUNO Inc., Seoul, South Korea, [3] Department of Neurology, Ajou University School of Medicine, Suwon, South Korea, [4] Center for Digital Health, Yongin Severance Hospital, Yonsei University Health System, Yongin, South Korea, [5] BUD.on Inc., Seoul, South Korea

Edited by:
Shivnarayan Patidar,
National Institute of Technology, India

Reviewed by:

Frontiers in cardiovascular medicine

0 citations (on 2022.11.14), co-first author

DeepECG Project
Project Insight

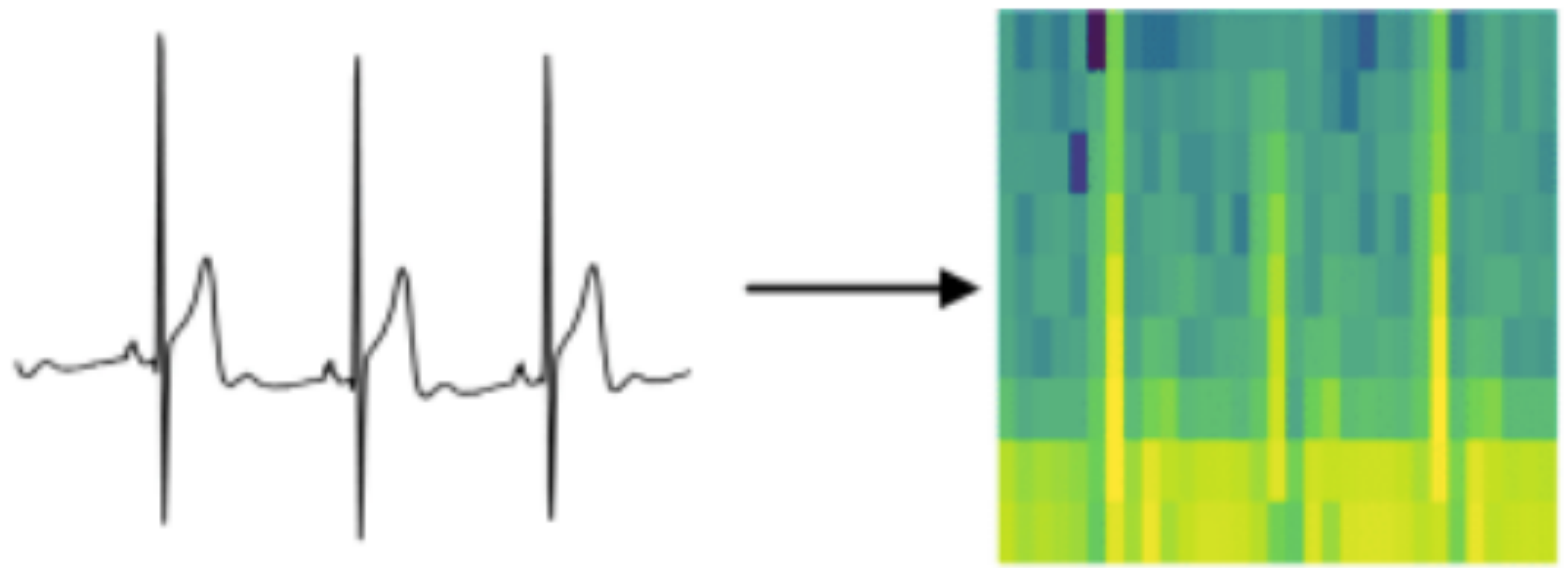
VUNO

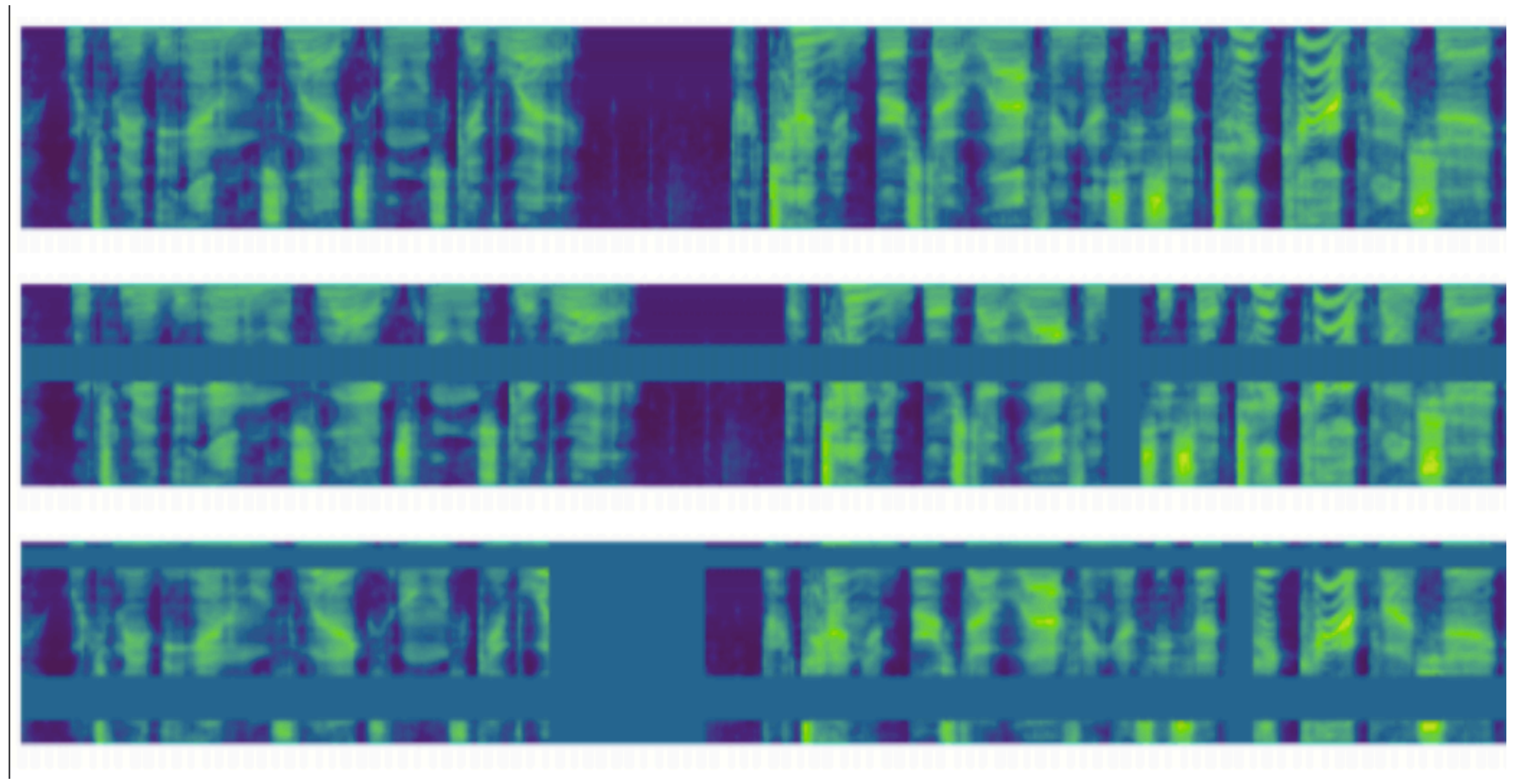
- During DeepECG project, I try different techniques to improve model performance. This was because our prototype will be launched in the industry.
  - Input Variation
    - 1D, 2D input transformation
    - Data augmentation
  - Model Variations
    - Baseline model setup
    - Long-range data analysis
    - Use of unlabeled data

- In the meantime, I considered not only high performance but also accurate decisions (meaning, reducing the false positive rate and improving the true negative rate).
  - I thought that this field is a little bit sensitive area unlike traditional computer science fields such as classifying images that require only high accuracy.
- That is why I applied a variety of approaches to know what kinds of approaches will reduce the low false positive rate and high false positive while providing high classification performance.

DeepECG Project
Input transformation

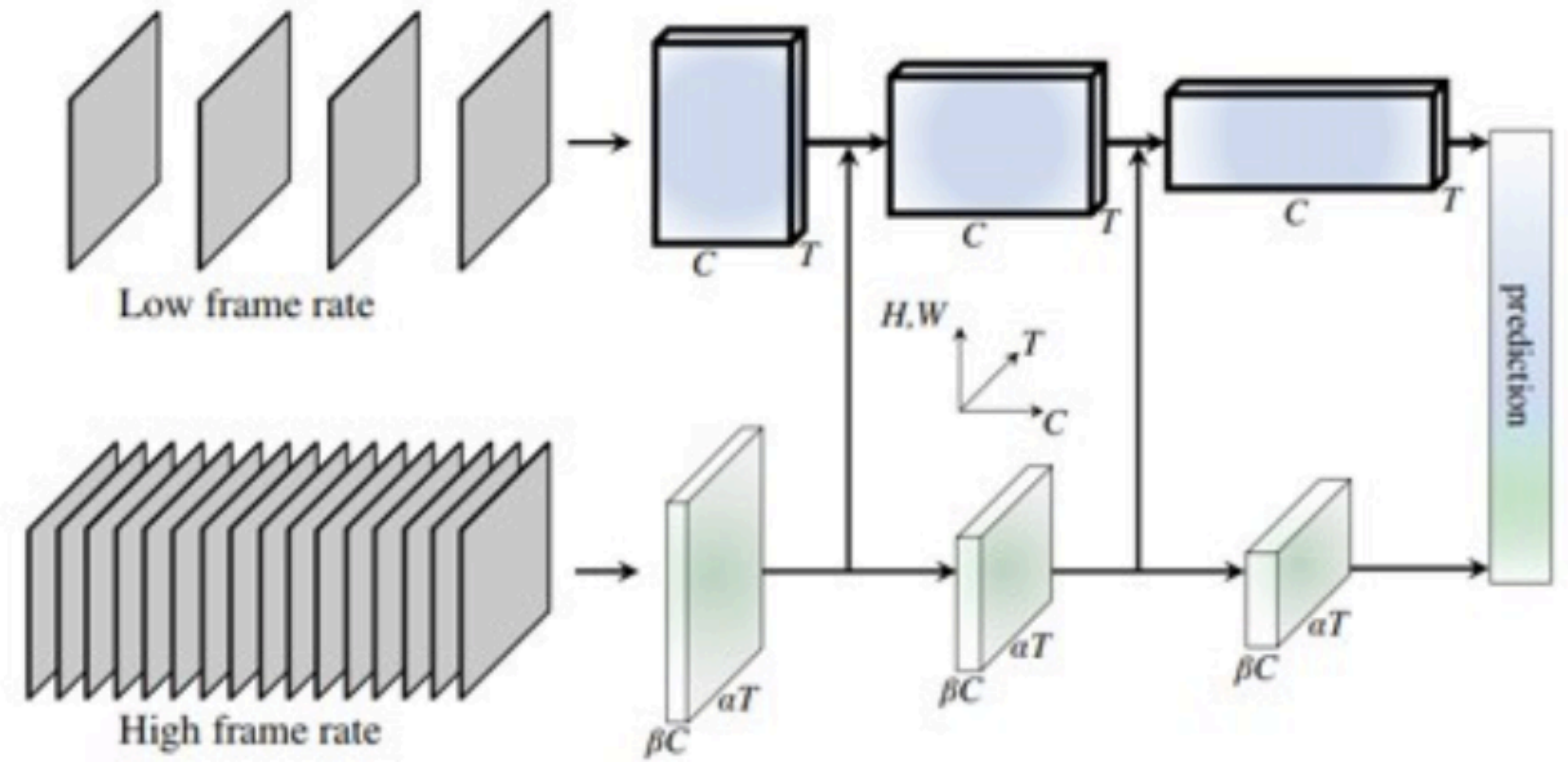
- First of all, I tried to understand what kinds of input formats will give a robust and superior performance for ECG tasks.
  - Previously (from a zero-calibration study), I understood the input format was one of the significant factors to build a great framework.
- Here, I transformed 1-D input to time-frequency inputs based on STFT, and consider combining different data augmentation techniques.
  - I believed 2D input formats in a way might be robust in a noise environment based on the fact that 2D architecture (i.e., CNN) is more robust to input setting.
- Moreover, I did several experiments only considering 1D input or 2D input or 1-D & 2D combination.
- Through this approach, I believed that low false positive will be reduced because we can classify noise data.



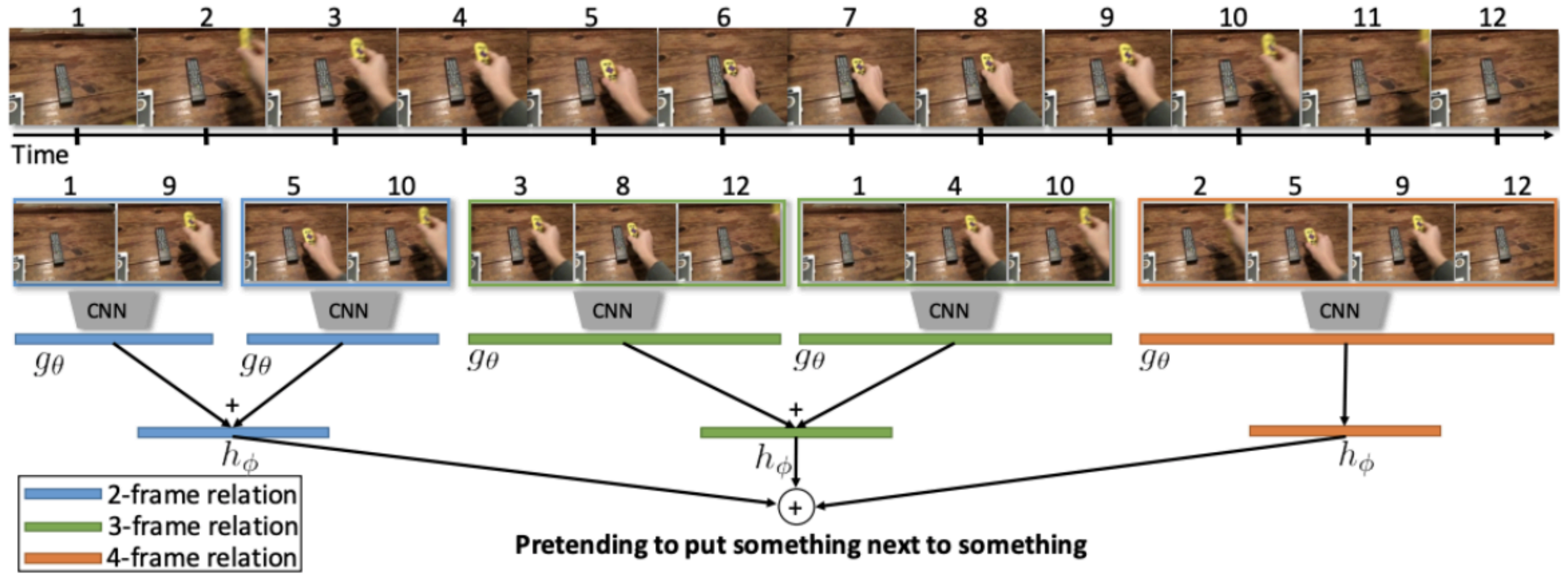


DeepECG Project
Model Implementation

- Next, I experimented in a different way to find our baseline models.
    - This is because there were no baseline models unlike computer vision fields in these fields so I should find out a proper network for this task.
    - For example, we implemented basic classification models (ResNet, Inception, Xception, ResNext, DenseNet, EfficientNet, ..)
- Furthermore, I believed that ECG's high and low-frequency information may have different characteristics, thus we should consider this information in designing networks.
    - Therefore, I borrowed the architecture concept in video classification like Slow Fast Framework & Temporal Relation Networks to consider low and high frames.
- Through this approach, I tried to find out the proper framework giving the low false positive rate and high true negative.



SlowFast Network(Facebook)



Temporal Relation Networks(MIT)

DeepECG Project
Long-range signal

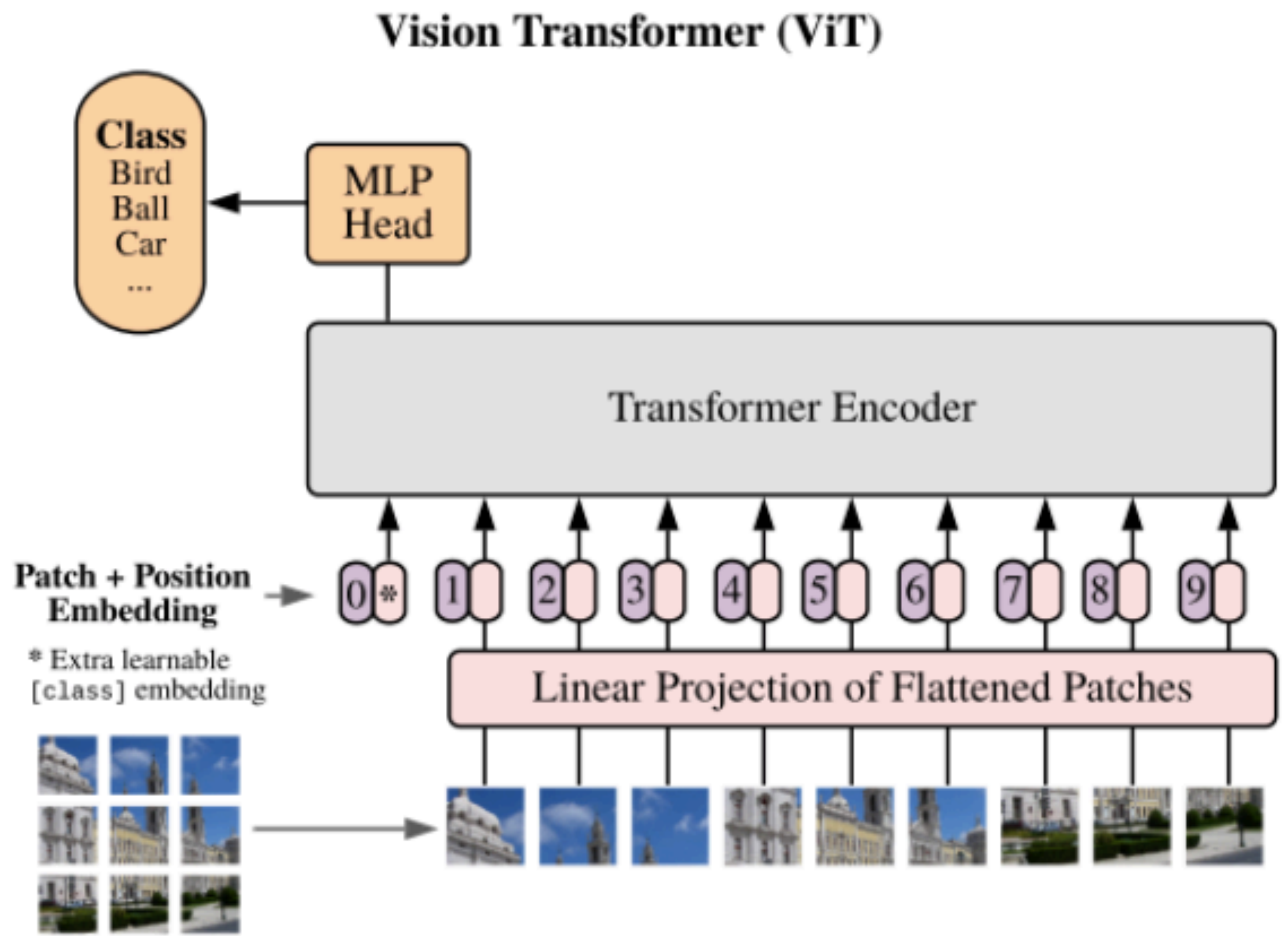
- Sometimes, our team should handle long-range ECG signals (about 48 hours). It was difficult to train these long-range signals in a single training due to numerous computational costs.

- Thus, we tried to compress raw ECG signals by using CNN and utilizing embedded vectors for the input of transformer or multi-layer perceptrons (based on previous studies).

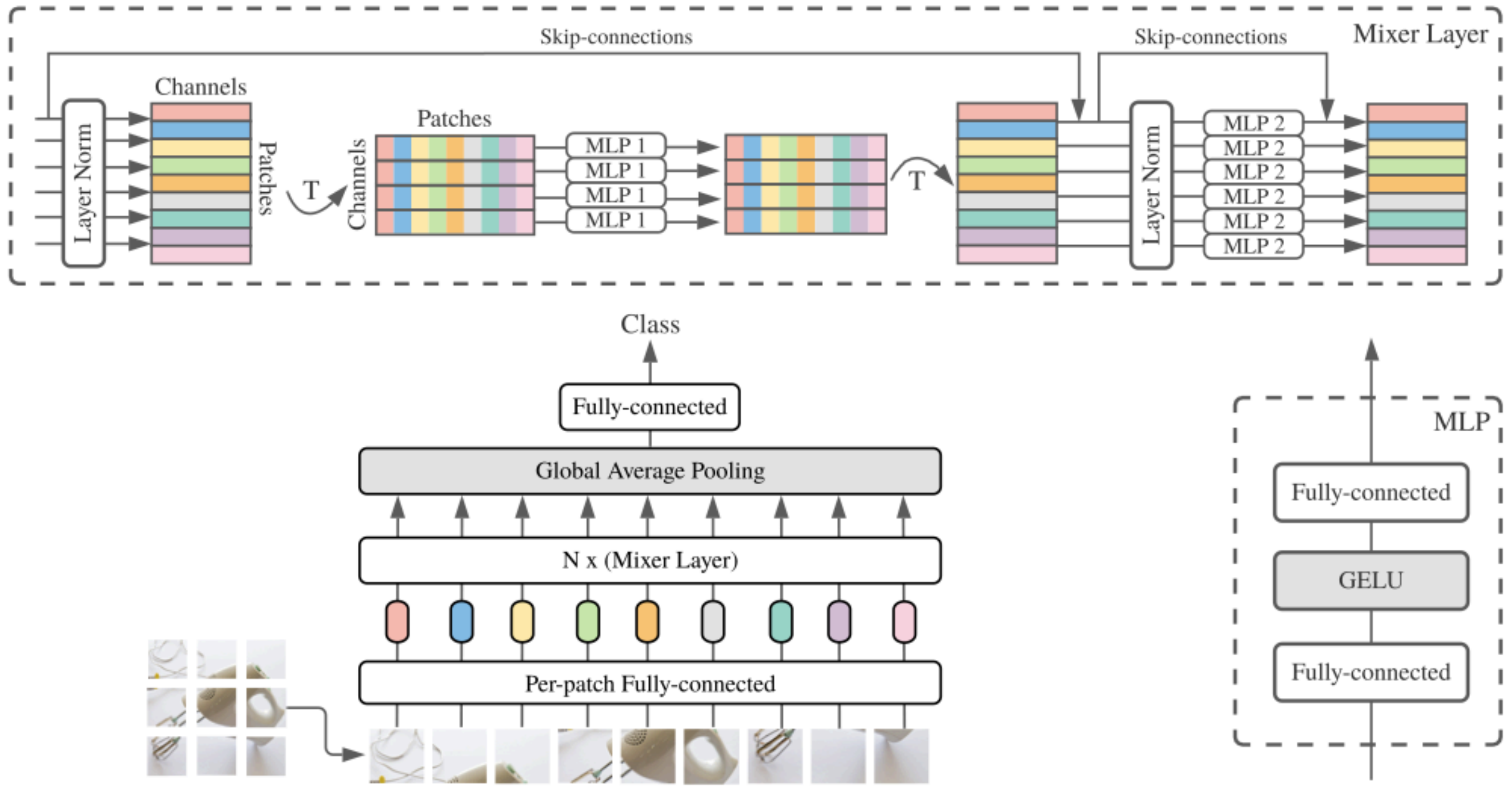
  - The major different point was how to augment input signals. Our input was not image-like input so we need to consider input transformation and augmentation.

  - Here, we tried to divide our input into an hourly signal and transform it into an embedding vector.

  - Now, we tried to expand our research to developing other data augmentation techniques.



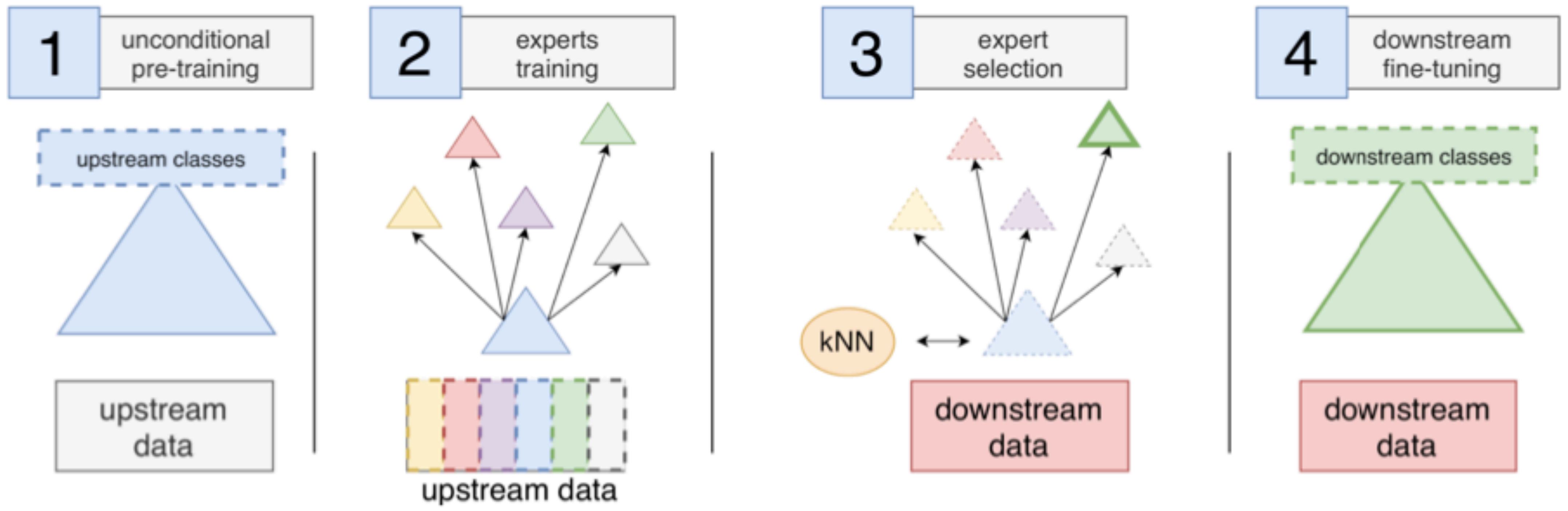
Vision Transformer (google)



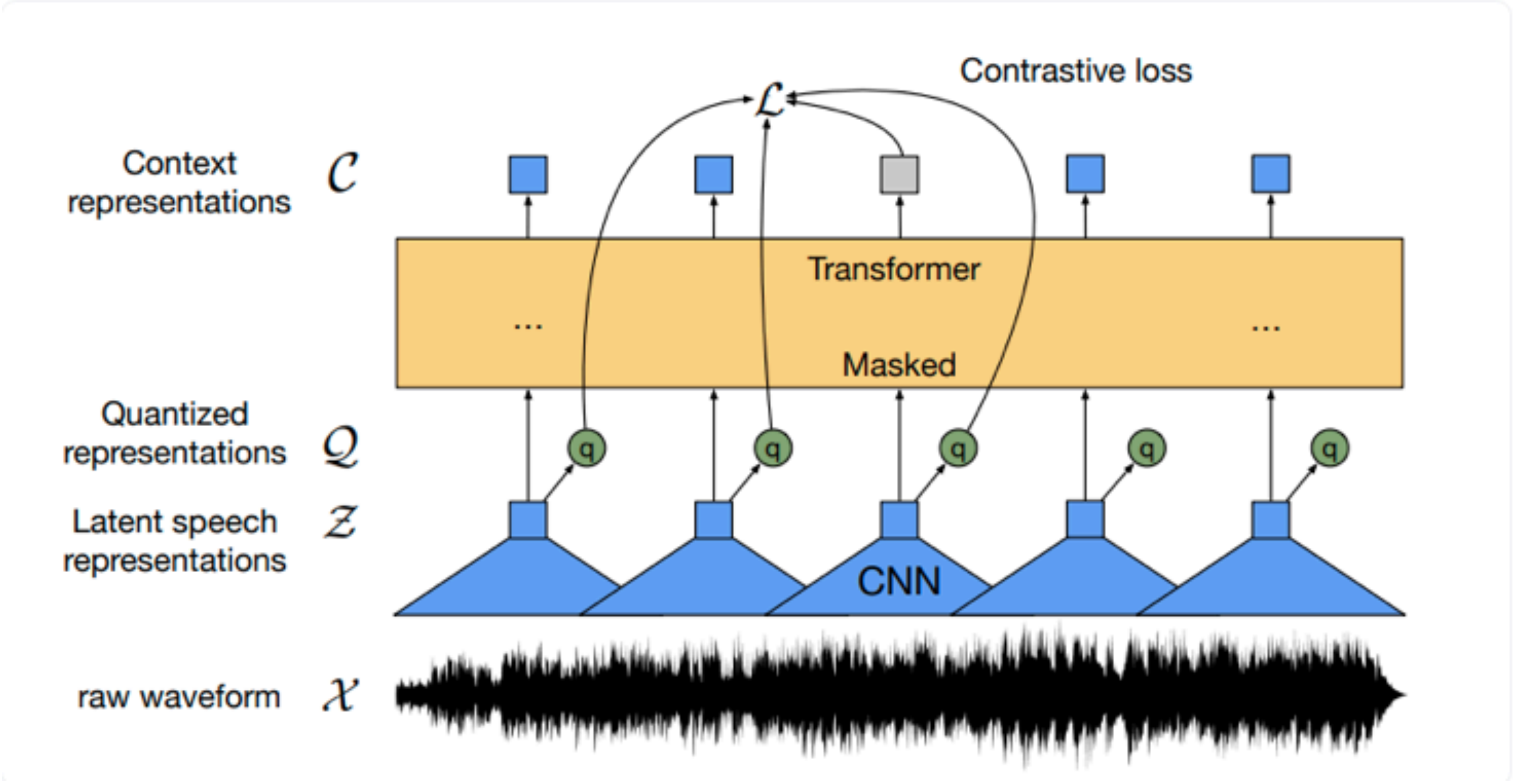
MLP-Mixer (google)

DeepECG Project
Unlabeled data

- Finally, we understood our company had many unlabeled data, so we tried to utilize those unlabeled data for our training.
  - We believe that it would be a good way to borrow the power of data by accumulating data.
- Therefore, we borrowed previous research concepts that label unlabeled data or use unlabeled-data representation.
  - In fact, in the field of ECG, there was scarce label information, especially in the case of minor heart diseases, it would be hard to get label information. Thus, we decided to label unlabeled data, and then we gave these label information to the medical director in the company for exact confirmation.
  - By building multi-label pre-trained models based on *label-hierarchy and* creating self-supervised learning methods for ECG tasks while preserving ECG inherent characteristics



Scalable Transfer Learning (google)



Wav2Vec (Facebook)

DeepECG Project Device

- We created a simple and portable ECG measurement, called Hativ P30.
- Hativ P30 is a simple and portable ECG measurement medical device that analyzes ECG data and provides analysis results such as normal sinus rhythm, atrial fibrillation, bradycardia, and tachycardia.
- Hativ can easily measure heart signals within 30 seconds anytime, anywhere. Analysis results can be checked at a glance through a connected mobile app.
- I led this project, collaborated with other departments (design, business, etc), and made a prototype.
    - We believed that this product will be one of the paths to deliver medical service for anyone.

Our product

DeepECG Project : Part Lead
Conclusion

- Through these studies, I learned the knowledge to apply engineering to the healthcare industry.
- In particular, I think the process of analyzing, modeling, and utilizing ECG data, making a device myself, and combining them is a special and unique experience.