UM6P | College of Computing

FEBRUARY 2024

KNOWLEDGE DISTILLATION FOR ON DEVICE AUDIO CLASSIFICATION DATA SCIENCE PROJECT

Data science Project

Presented by Jabbar Samir

Members of the jury:
Rahhal Errattahi
Ali Idri

TABLE
of contents

01. Introduction

02. Objectives

03. Literature review

04. Data

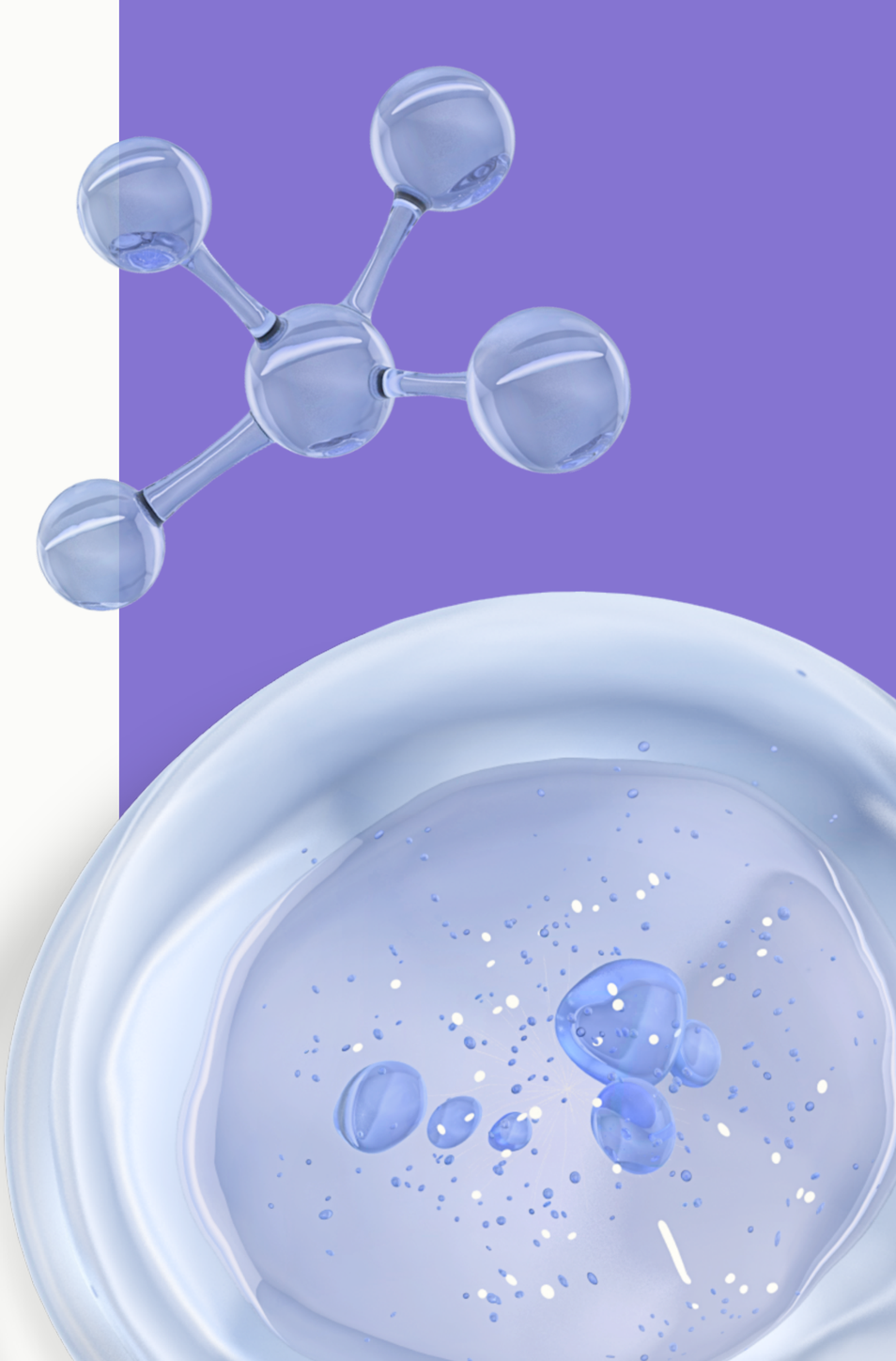
05. Methodology

06. Results

07. Conclusion

Introduction

COVID-19 is an infectious disease caused by the recently found virus known as SARS-CoV-2 (or coronavirus).
Before the outbreak originated in Wuhan, China in December 2019, there was no information about this virus.
Researchers around the world are working to fight against COVID-19 and reduce its spread and prevent it.



Objectives

1 Develop a Machine Learning Model (Wav2Vec):

- Implement the Wav2Vec model architecture to effectively process audio data.
- Train the model to recognize patterns and features indicative of respiratory symptoms, particularly associated with COVID-19.

2 Utilize Audio Data:

- Explore the richness of audio data as a novel and informative source for respiratory health screening.
- Leverage the unique characteristics of audio signals to extract meaningful insights relevant to COVID-19 symptom detection.

3 Compress the Model using Knowledge Distillation:

- Employ knowledge distillation techniques to distill knowledge from a large, complex teacher model (Wav2Vec) into a smaller, more lightweight student model.
- Reduce computational requirements while maintaining performance, enabling efficient deployment in resource-constrained environments.

4 Contribute to Public Health Efforts:

- Deploy the developed model to production, making it accessible on mobile devices for widespread use.
- Enable early detection and screening of COVID-19 symptoms, aiding public health initiatives and facilitating timely interventions to mitigate the spread of the virus.

Literature Review:

Literature Review: Audio Classification in Healthcare

Audio classification in healthcare has emerged as a promising field leveraging the power of machine learning and signal processing techniques. With advancements in audio processing algorithms and the availability of large-scale datasets, researchers are exploring the use of audio signals for various healthcare applications.

Existing Models for COVID-19 Audio Classification

In response to the COVID-19 pandemic, researchers have actively investigated the feasibility of audio-based screening methods for detecting potential COVID-19 cases. Several studies have explored the use of machine learning models to analyze respiratory sounds, cough patterns, and voice characteristics associated with COVID-19 infections.r.

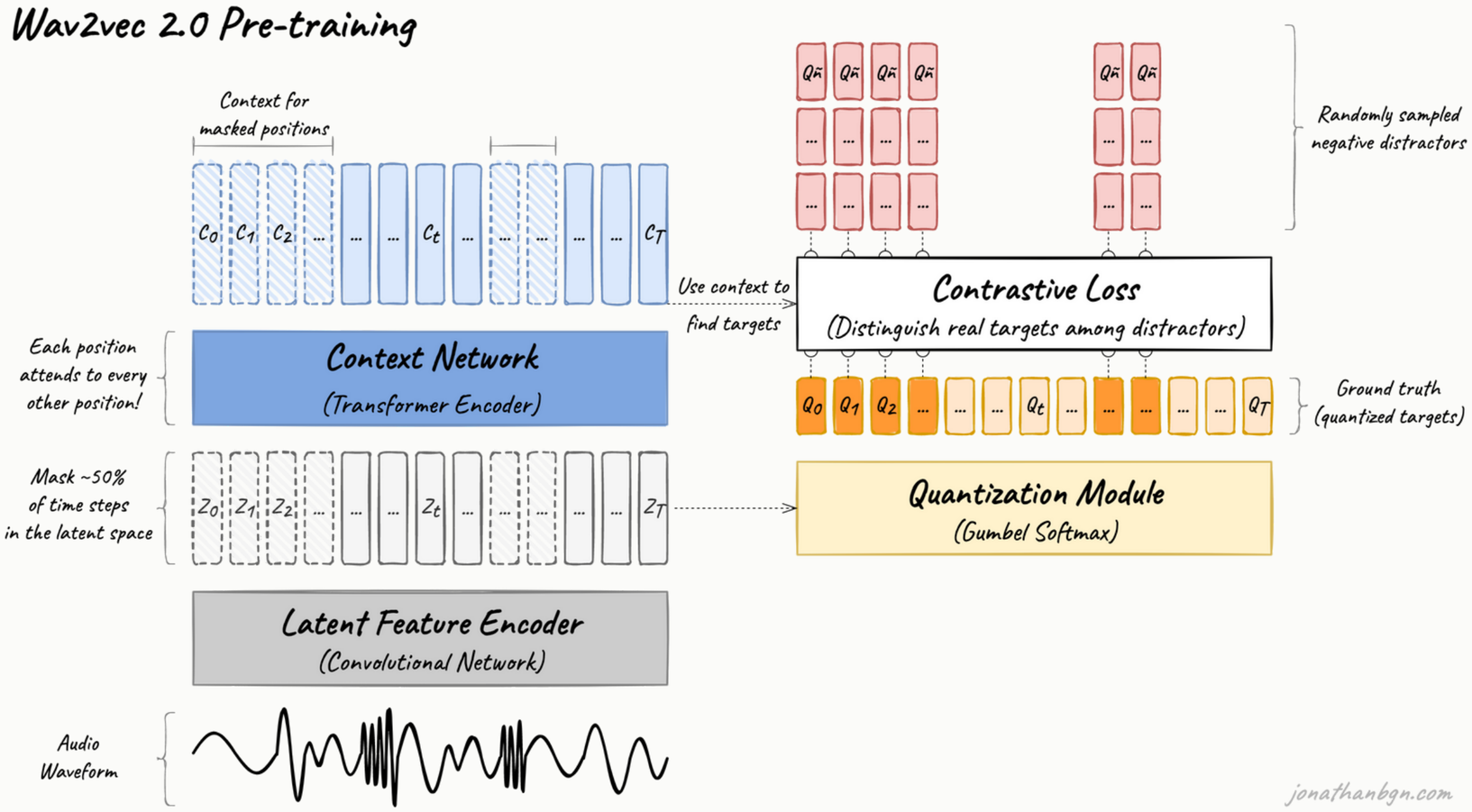
Knowledge Distillation in Audio Classification

Knowledge distillation techniques offer a promising approach to compressing large, complex machine learning models into smaller, more lightweight versions suitable for deployment in resource-constrained environments. In the context of audio classification, knowledge distillation involves transferring knowledge from a large, teacher model to a smaller, student model while maintaining performance.

Methodology

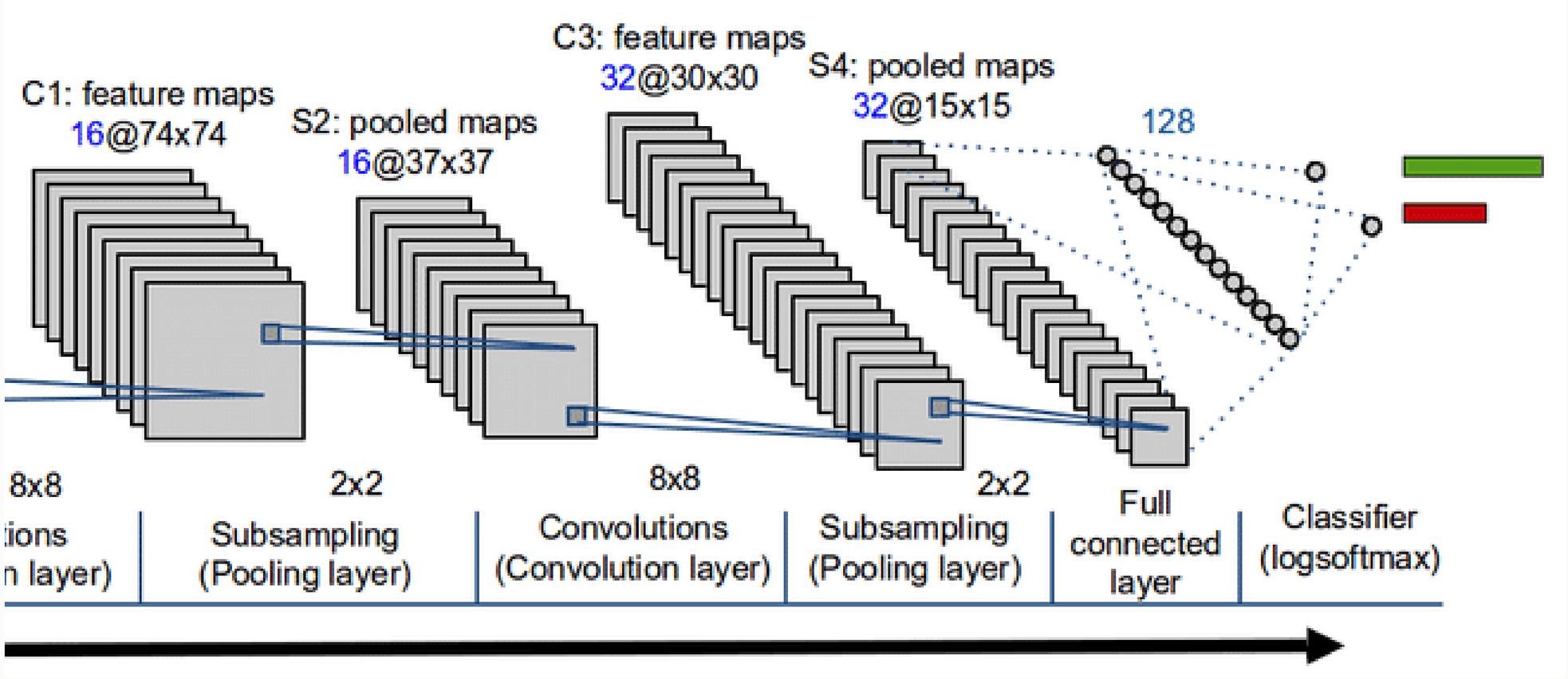
Teacher model architecture

- Wav2Vec is a state-of-the-art model designed for speech recognition and audio classification tasks.
- Its architecture is based on convolutional neural networks (CNNs) for feature extraction and transformer layers for contextual understanding.
- The model employs a masked language modeling objective similar to BERT's, adapted for speech signals.



1st student model architecture

Develop a lightweight student model through knowledge distillation from the teacher model for prototyping purposes



Covid
No Covid

2nd student model architecture

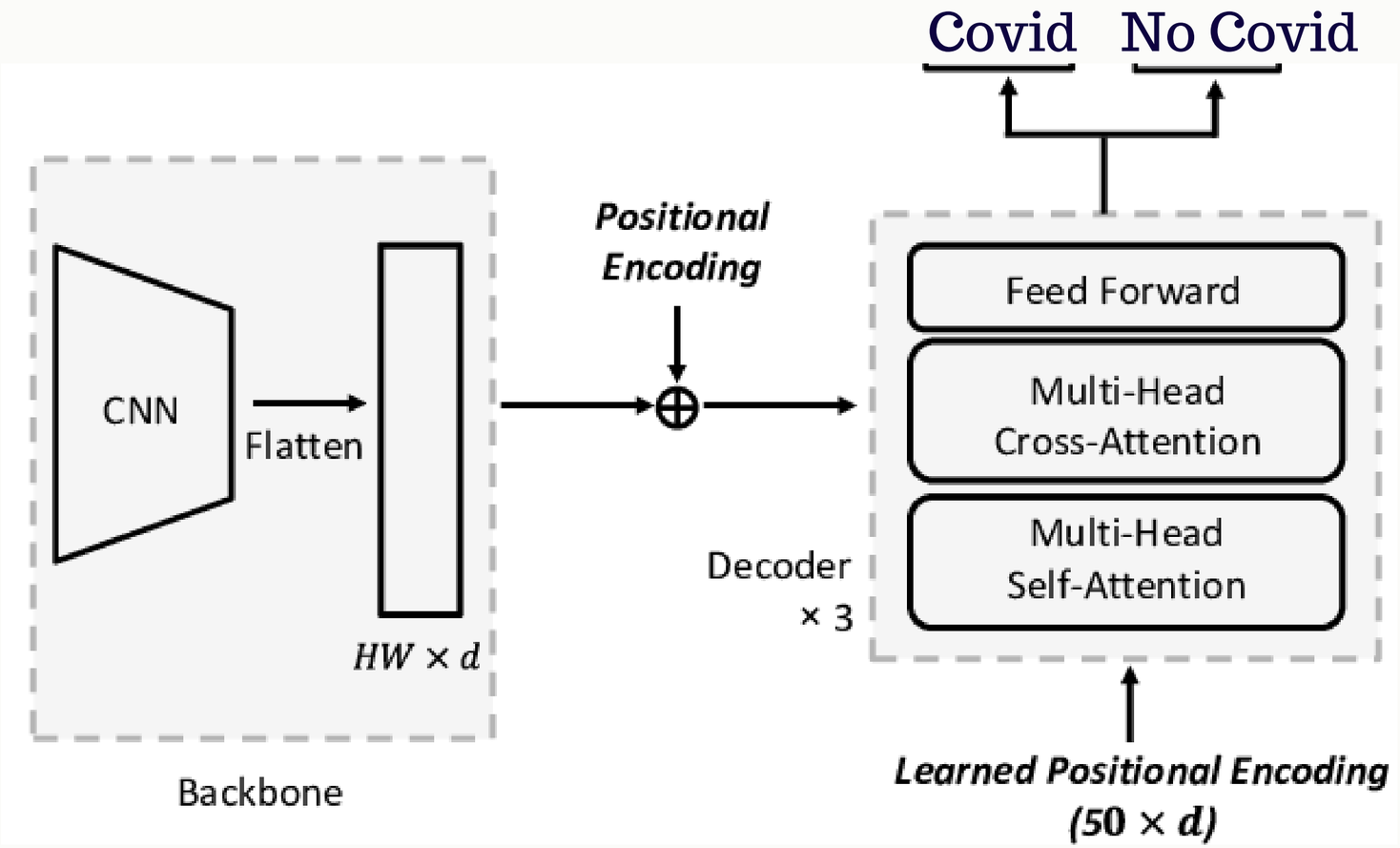
- Objective:

Develop a lightweight student model through knowledge distillation from the teacher model while maintaining performance.

- Architecture:

Retains the same number of CNN layers as the teacher model but reduces the number of transformer layers for efficiency.

Employs a simplified version of the transformer network to capture contextual information from the encoded features.



Results

Comparison between Simple Student Model and Teacher Model

	Loss	Accuracy
Teacher Model Evaluation	0.70	0.50
Simple Student Model Evaluation	0.81	0.30

For the Complex Student Model, there will be a slight degradation in performance compared to the teacher model due to the reduced complexity of the student model.
The performance degradation will be minimal and acceptable for practical applications.

Model Name	# of Parameters	Model Size	CPU (8 cores) Inference Time	GPU Inference Time	WER
wav2vec_big_960h	317 M	1262 mb	4433 s	123 s	2.63%
Student wav2vec 2.0	65 M	262 mb	1560 s	51 s	9.51%

Conclusions

In this project, we embarked on a journey to explore the potential of knowledge distillation in enhancing the efficiency and effectiveness of speech classification models, particularly in the context of COVID-19 classification using the Wav2Vec 2.0 architecture.

- Continual refinement and optimization of the student model architecture and distillation process
- Experimentation with different distillation techniques, regularization methods, and hyperparameter configurations
- Leveraging larger and more diverse datasets for enhanced effectiveness of the distilled model
- Integration of the distilled model into real-time applications and healthcare systems
- Exploration of advanced techniques such as pruning, quantization, and knowledge distillation cascades
- Collaborative efforts with domain experts to advance the state-of-the-art in speech classification and combat the COVID-19 pandemic