ADAPTIVE DISCRETE TOKENIZATION OF ELECTROCARDIOGRAMS FOR CLINICAL APPLICATIONS

Rohan Banerjee[1,2], **Jacques Delfrate**[1,2] **& Robert Avram**[1,2,3]
[1] Heartwise (heartwise.ai), Montreal Heart Institute, Montreal, Quebec, Canada
[2] Montreal Heart Institute, Department of Medicine, Montreal, Quebec, Canada
[3] Department of Biochemistry and Molecular Medicine, Faculty of Medicine, University of Montreal, Montreal, Quebec, Canada
banerjee.rohan98@gmail.com, robert.avram.md@gmail.com

1 INTRODUCTION

Electrocardiography (ECG) is a crucial non-invasive diagnostic tool in cardiology, renowned for its accuracy in detecting a broad range of cardiovascular diseases [Rafie et al. (2021)]. The increasing volume of ECG recordings, coupled with the limitations of manual analysis which demand considerable expertise of at least four years of medical training and six years of cardiology training and can be prone to variability and error rates reaching 25 percent [Cook et al. (2020)], highlights the need for accurate, automated clinical report generation. To leverage the power of modern Large Language Models (LLMs) for this task, effective methods for representing raw ECG signals in a discrete, tokenized format are essential.

Our work focuses on exploring tokenization methods to convert continuous ECG signals into discrete tokens. By learning meaningful discrete representations directly from raw ECG data, we aim to bridge the gap between complex waveform patterns and automated LLM applications such as clinical report generation and decision support in diagnostic assistance.

Tokenization converts complex data into discrete units, facilitating alignment of continuous signals with LLMs. In language, sentences are tokenized into discrete units enabling efficient processing by LLMs. In the audio domain, tokenization has revolutionized tasks such as speech recognition and synthesis, allowing for the representation of continuous audio waveforms as sequences of continuous or discrete acoustic tokens [Mousavi et al. (2024), Zhang et al. (2023)] leading to the development of multi-modal LLMs [Borsos et al. (2022)]. Similarly, in the physiological signals space including ECG, EEG, EHR, there have been works aligning continuous representations with LLMs [Cai et al. (2024), Oh et al. (2022), Duan et al. (2023), Lee et al. (2024)]. If we specifically focus on ECG, the literature exploring discrete tokenization remains limited. A recent work ECGByte [Han et al. (2024)] explored byte-pair encoding for tokenizing ECG signals by taking inspiration from how tokenization is done in LLMs [Radford et al. (2019)]. We draw inspiration from audio codecs [D'efossez et al. (2022), Zeghidour et al. (2021)] and use vector quantization (VQ) [Mammen & Ramamurthi (1990)], specifically QINCo [Huijben et al. (2024)] to tokenize ECG signals. QINCo has an adaptive residual quantization nature which would dynamically tailor its codebooks at each training epoch in order to capture the subtle and time-varying patterns present in ECG data.

2 DATA AND METHODS

We utilize the open-source MIMIC-IV dataset [Johnson et al. (2023)] for all trainings and evaluations. MIMIC-IV is a collection of 789,481 ECGs, each being a 10-second, 12-lead time series sampled at 500Hz. Prior to model training, we applied a signal-based normalization technique for preprocessing. To establish clinically relevant labels for linear probing (LP), two expert cardiologists, each with over four years of ECG interpretation experience, independently annotated 10,075 diagnostic statements (drawn from MIMIC-IV and other datasets) using a standardized set of 77 labels based on American Heart Association guidelines [Kligfield et al. (2007)]. These labels span six clinical categories—Rhythm Disorders, Conduction Disorders, Chamber Enlargement, Pericarditis, Infarction/Ischemia, and Other—and were finalized through consensus review in a standard 12-lead ECG format, achieving high inter-rater reliability (Cohen's kappa > 0.80).

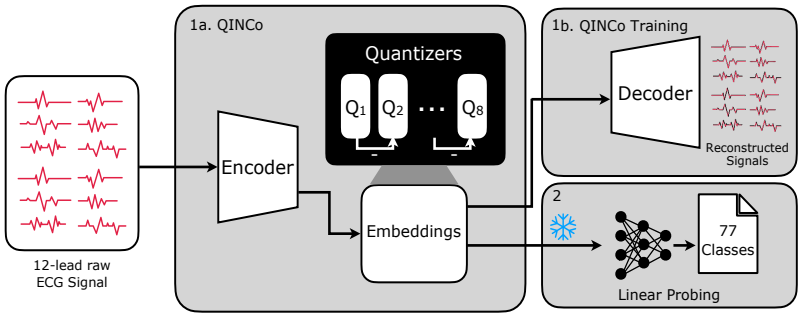

Figure 1: **Method Overview** 1a. and 1b. show the QINCo training with a reconstruction objective. 2. Linear probing on frozen embeddings

Our methodology for discrete ECG tokenization consists of two principal stages: **QINCo training** and **LP**. The initial stage involves **training the QINCo model**, which serves as our ECG tokenizer. This tokenizer is trained on randomly chosen 90% of the dataset, while the remainder is reserved for testing, enabling it to learn a discrete codebook representation with a codebook size of 1024 and 8 quantizers. The input signals of size 5000×12 are fed to the encoder, which consists of several 1-dimensional convolutional layers, and the decoder produces a reconstructed signal. The model is trained using the QINCo loss, which extends the standard MSE loss by iteratively refining the reconstruction over 8 quantization steps. In each step, the residual error from the previous quantization is further quantized, enabling the model to progressively capture increasingly finer details of the ECG signals. The training was performed for 2 epochs with a learning rate of $3 \times 10^{-4}$and a batch size of 4. In the **LP stage**, the encoder of the QINCo is frozen. Subsequently, the entire dataset is passed through this frozen encoder to generate discrete embeddings for each ECG signal. The embeddings are represented as codebook entries from the multiple quantizers hence converting the continuous signals into discrete tokens. The embeddings are divided into 70:10:20 ratio for train, validation and test sets respectively. We used a different split from the QINCo training stage, as QINCo was trained for reconstruction, while LP involves classification. We believe that the downstream performance is a pure reflection of the learned representations' quality. A single layer linear probe of dimension 256 is used to perform multi-label classification on the frozen discrete tokens. We performed hyperparameter optimization using WandB [Biewald (2020)] and train the probe with a learning rate of 1×10^{-5} with a CosineAnnealing learning rate scheduler for 100 epochs.

3 Results

For our tokenizer, we had a MSE loss of 0.028 on test set and 96.28% overall usage of the codebooks. This means that our tokenizer was able to discretize the signals without codebook collapse. For qualitative assessment, an expert cardiologist reviewed reconstructed test-set signals. We also evaluated performance on six diagnostic classes—Rhythm Disorders, Conduction Disorders, Chamber Enlargement, Pericarditis, Infarction/Ischemia, and Other Diagnoses using micro-averaged AUC and F1-score using the Youden Index for each label, with weighted averaging for overall AUC. The micro AUC values were 0.957, 0.893, 0.888, 0.705, 0.833, and 0.929, respectively, while the corresponding micro F1 scores were 0.911, 0.789, 0.791, 0.688, 0.719, and 0.846.

4 Discussion

In this study we introduce a novel way to create discrete tokens from ECG signals. To the best of our knowledge, this is the first work employing deep learning-based VQ for raw ECG signals. Through its reliable performance on six diagnostic categories, we demonstrate that the learned representations are clinically meaningful and potentially valuable for other downstream tasks. We note that LP performs worse on certain classes compared to others, likely due to their under-representation in

the public dataset MIMIC-IV. Although this work only pretrained QINCo on MIMIC-IV, we plan to scale up the pretraining and benchmark on other datasets like PTBXL, Code-15 and Montreal Heart Institute (MHI) dataset. We would like to explore other VQ methods to tokenize the ECG signals. This work is a part of our ongoing efforts to develop a clinical report generation model using LLMs. Once we complete the pipeline, we plan to validate the report generation performance across 10 sites spread across North America. We would also compare this work with the current tokenization and report generation models for ECG.

MEANINGFULNESS STATEMENT

We define a meaningful representation of life as one capable of conveying information from macro to micro scales of a living organism's state. Our exploration focuses on human heart health within this vast scope. We are investigating methods to compress and learn representations from ECG signals, ultimately seeking to translate these signals into human-readable language. This effort contributes to building AI systems that can understand and interpret complex biological data, paving the way for advanced, personalized medical care.

REFERENCES

Lukas Biewald. Experiment tracking with weights and biases, 2020. URL https://www.wandb.com/. Software available from wandb.com.

Zalán Borsos, Raphaël Marinier, Damien Vincent, Eugene Kharitonov, Olivier Pietquin, Matthew Sharifi, Dominik Roblek, Olivier Teboul, David Grangier, Marco Tagliasacchi, and Neil Zeghidour. Audiolm: A language modeling approach to audio generation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 31:2523–2533, 2022. URL https://api.semanticscholar.org/CorpusID:252111134.

Yifu Cai, Arvind Srinivasan, Mononito Goswami, Arjun Choudhry, and Artur Dubrawski. Jolt: Jointly learned representations of language and time-series for clinical time-series interpretation (student abstract). In *AAAI Conference on Artificial Intelligence*, 2024. URL https://api.semanticscholar.org/CorpusID:268713756.

David A. Cook, So Young Oh, and Martin V. Pusic. Accuracy of physicians' electrocardiogram interpretations: A systematic review and meta-analysis. *JAMA internal medicine*, 2020. URL https://api.semanticscholar.org/CorpusID:222152103.

Alexandre D'efossez, Jade Copet, Gabriel Synnaeve, and Yossi Adi. High fidelity neural audio compression. *ArXiv*, abs/2210.13438, 2022. URL https://api.semanticscholar.org/CorpusID:253097788.

Yiqun Duan, Jinzhao Zhou, Zhen Wang, Yu kai Wang, and Ching-Teng Lin. Dewave: Discrete eeg waves encoding for brain dynamics to text translation. *ArXiv*, abs/2309.14030, 2023. URL https://api.semanticscholar.org/CorpusID:262466081.

William Jongwon Han, Chaojing Duan, Michael A. Rosenberg, Emerson Liu, and Ding Zhao. Ecg-byte: A tokenizer for end-to-end generative electrocardiogram language modeling. *ArXiv*, abs/2412.14373, 2024. URL https://api.semanticscholar.org/CorpusID:274860152.

Iris Huijben, Matthijs Douze, Matthew Muckley, Ruud van Sloun, and Jakob Verbeek. Residual quantization with implicit neural codebooks. *ArXiv*, abs/2401.14732, 2024. URL https://api.semanticscholar.org/CorpusID:267301189.

Alistair E. W. Johnson, Lucas Bulgarelli, Lu Shen, Alvin Gayles, Ayad Shammout, Steven Horng, Tom J. Pollard, Benjamin Moody, Brian Gow, Li wei H. Lehman, Leo Anthony Celi, and Roger G. Mark. Mimic-iv, a freely accessible electronic health record dataset. *Scientific Data*, 10, 2023. URL https://api.semanticscholar.org/CorpusID:255439889.

Paul D. Kligfield, Leonard S. Gettes, James J. Bailey, Rory W. Childers, Barbara Deal, E. William Hancock, Gerard van Herpen, Jan A. Kors, Peter W. Macfarlane, David M. Mirvis, Olle Pahlm,

Pentti M. Rautaharju, Galen S. Wagner, Mark E. Josephson, Jay W. Mason, Peter M. Okin, Borys Surawicz, and Hein Maarten Wellens. Recommendations for the standardization and interpretation of the electrocardiogram: part i: the electrocardiogram and its technology a scientific statement from the american heart association electrocardiography and arrhythmias committee, council on clinical cardiology; the american college of card. *Journal of the American College of Cardiology*, 49 10:1109–27, 2007. URL https://api.semanticscholar.org/CorpusID:52798353.

Yoonhyung Lee, Younhyung Chae, and Kyomin Jung. Leveraging vq-vae tokenization for autoregressive modeling of medical time series. *Artificial intelligence in medicine*, 154:102925, 2024. URL https://api.semanticscholar.org/CorpusID:270826912.

C. P. Mammen and Bhaskar Ramamurthi. Vector quantization for compression of multichannel ecg. *IEEE Transactions on Biomedical Engineering*, 37:821–825, 1990. URL https://api.semanticscholar.org/CorpusID:19544790.

Pooneh Mousavi, Jarod Duret, Salah Zaiem, Luca Della Libera, Artem Ploujnikov, Cem Subakan, and Mirco Ravanelli. How should we extract discrete audio tokens from self-supervised models? *ArXiv*, abs/2406.10735, 2024. URL https://api.semanticscholar.org/CorpusID:270559888.

Jungwoo Oh, Hyunseung Chung, Joon myoung Kwon, Dongwoo Hong, and E. Choi. Lead-agnostic self-supervised learning for local and global representations of electrocardiogram. *ArXiv*, abs/2203.06889, 2022. URL https://api.semanticscholar.org/CorpusID:247446583.

Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. 2019. URL https://api.semanticscholar.org/CorpusID:160025533.

Nikita Rafie, Anthony H. Kashou, and Peter A. Noseworthy. Ecg interpretation: Clinical relevance, challenges, and advances. *Hearts*, 2021. URL https://api.semanticscholar.org/CorpusID:243466810.

Neil Zeghidour, Alejandro Luebs, Ahmed Omran, Jan Skoglund, and Marco Tagliasacchi. Soundstream: An end-to-end neural audio codec. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 30:495–507, 2021. URL https://api.semanticscholar.org/CorpusID:236149944.

Xin Zhang, Dong Zhang, Shimin Li, Yaqian Zhou, and Xipeng Qiu. Speechtokenizer: Unified speech tokenizer for speech large language models. *ArXiv*, abs/2308.16692, 2023. URL https://api.semanticscholar.org/CorpusID:261394297.