CBIS'24
XX Congresso Brasileiro de Informática em Saúde
08/10 a 11/10 de 2024 - Belo Horizonte/MG - Brasil

Enhancing automated electrocardiogram (ECG) diagnosis through multimodal pre-training with text reports

Aprimoramento do diagnóstico automatizado de eletrocardiograma (ECG) por meio de pré-treinamento multimodal com laudos em texto

Mejora del diagnóstico automatizado de electrocardiograma (ECG) mediante preentrenamiento multimodal con informes de texto

Jose Geraldo Fernandes[1], Diogo Tuler[2], Gabriel Lemos[2], Pedro Robles Dutenhefner[2], Turi Rezende[2], Gisele Pappa[3], Gabriela Paixão[4], Antônio Ribeiro[4], Wagner Meira Jr.[3]

1 MSc, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte (MG), Brazil.
2 UGS, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte (MG), Brazil.
3 PhD, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte (MG), Brazil.
4 MD PhD, Telehealth Center from Hospital das Clínicas da Universidade Federal de Minas Gerais (UFMG), Belo Horizonte (MG), Brazil.

Corresponding author: MSc Jose Geraldo Fernandes Costa de Lima
E-mail: josegeraldof@ufmg.br
Links: github.com/cljosegfer/lesaude-clip

Abstract

Objective: Heart diseases are the leading cause of death worldwide, and the electrocardiogram (ECG) is the primary diagnostic tool for assessing cardiac activity. Automated and remote ECG diagnosis can help the healthcare system with timely and high-quality cardiac assessments, especially for peripheral regions and rural areas. Automatic ECG classification has been extensively researched, but it is still challenging to build accurate models for such a wide spectrum of scenarios. **Method:** This study enhances the performance of ECG deep learning classification models using a multimodal pre-training stage with physician's reports. **Results:** Our approach improves the state-of-the-art model and achieves a mean F1 score of 0.755 over six categories using the full dataset, which is a relevant improvement for a relatively larger unlabeled corpus. **Conclusion:** The results demonstrate the potential to improve automated cardiac assessment with text pretraining.

Keywords: Cardiology; Electrocardiography; Machine Learning

Resumo

Objetivo: Doenças cardíacas são a principal causa de morte globalmente, e o eletrocardiograma (ECG) é a principal ferramenta para avaliar a atividade cardíaca. O diagnóstico automatizado e remoto do ECG pode ajudar o sistema de saúde com avaliações cardíacas antecipadas e precisas, especialmente em regiões periféricas e áreas rurais. A classificação automática de ECG foi amplamente pesquisada, mas ainda é um desafio criar modelos precisos para um espectro tão amplo. **Método:** Este estudo aprimora o desempenho dos modelos de classificação de aprendizagem profunda de ECG usando um estágio de pré-treinamento multimodal com o laudo médico. **Resultados:** Nossa abordagem melhora o modelo estado-da-arte e atinge uma pontuação média de F1 de 0,755 em seis categorias usando o conjunto de dados completo, o que é uma melhoria relevante para um corpus não-rotulado relativamente grande. **Conclusão:** Os resultados demonstram o potencial de melhora da avaliação cardíaca automatizada com o pré-treinamento de texto.

Descritores: Cardiologia; Eletrocardiografia; Aprendizado de Máquina

Resumen

Objetivo: Cardiopatías son la principal causa de muerte globalmente, y el electrocardiograma (ECG) es la principal herramienta para evaluar la actividad cardiaca. El diagnóstico automatizado y remoto del ECG puede ayudar al sistema sanitario con evaluaciones cardiacas tempranas y precisas, especialmente en regiones periféricas y zonas rurales. La clasificación automática de ECG ha sido ampliamente investigada, pero sigue siendo un reto crear modelos precisos para un espectro tan amplio. **Método:** Este estudio mejora el rendimiento de los modelos de clasificación de aprendizaje profundo de ECG utilizando una etapa de preentrenamiento multimodal con el informe médico. **Resultados:** Nuestro enfoque mejora el modelo de vanguardia y alcanza una puntuación F1 media de 0,755 en seis categorías utilizando el conjunto de datos completo, lo que supone una mejora relevante para un corpus sin etiqueta relativamente grande. **Conclusión:** Los resultados demuestran el potencial de mejora de la evaluación cardiaca automatizada con preentrenamiento de texto.

Descriptores: Cardiología; Electrocardiografía; Aprendizaje Automático

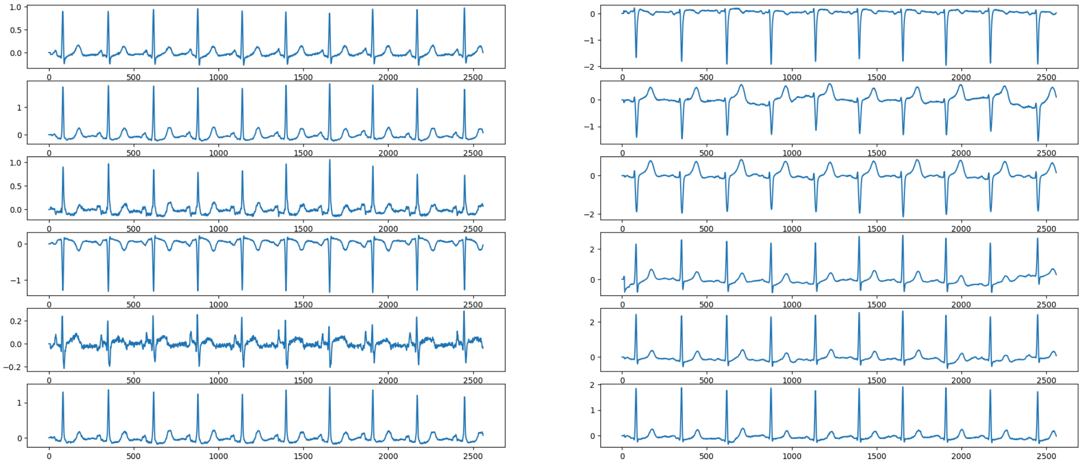
Introduction

Heart diseases are the leading cause of death worldwide[20] and an electrocardiogram (ECG) is currently the gold standard exam to diagnose several of these diseases. ECG is a non-invasive medical test that records the heart's electrical activity. Electrodes are placed on the skin to record the electrical impulses generated by the heart's natural pacemaker, the sinus node, as they travel through the heart muscle. This data is then displayed as a graph, showing the timing and strength of each electrical signal -- which in the case of Figure 1 represents the standard 12-lead ECG. Abnormalities in the ECG may indicate, for example, problems such as arrhythmia (irregular heart rhythms), myocardial infarction (heart attack), myocarditis (inflammation of the heart muscle), and heart valve problems. By analyzing ECG patterns, healthcare professionals can assess heart health, monitor ongoing conditions, and guide treatment strategies to manage heart diseases.

However, interpreting these signals to detect cardiac abnormalities presents significant challenges: it is necessary to identify specific and varied wave patterns that - within the appropriate context - may indicate underlying diseases. In this sense, the adoption of automated computational methods for analyzing ECGs has gained significant traction, especially with the integration of deep learning techniques[11,18].

Recent advancements in deep neural networks (DNNs) have achieved remarkable success across diverse domains, including computer vision[5], natural language processing (NLP)[16], and medical applications[8,22]. This success can be attributed, in part, to these models' capacity to extract meaningful and hierarchical features from massive datasets.

Figure 1 – Illustration of the 12-lead ECG of a normal exam.



Automated ECG diagnosis plays a crucial role in digital health by significantly improving access to healthcare services. Through digital platforms, individuals can be conveniently diagnosed remotely by specialists in different locations, eliminating geographic barriers and reducing the need for physical visits to centralized healthcare facilities, as the test is performed at local points and later diagnosed at a centralized facility. This accessibility is particularly important for peripheral regions and rural areas, allowing them to receive timely and high-quality cardiac assessments that support their primary health care system(1).

In addition, automated ECG analysis uses advanced algorithms to interpret ECG data, providing fast and accurate diagnoses. This efficiency not only improves patient outcomes, but also optimizes healthcare resources by streamlining the diagnostic process. Overall, digital health and automated ECG diagnostics work together to empower individuals by democratizing access to essential cardiac care services, ultimately fostering a more equitable healthcare landscape(10).

This study contributes to the research area of automated ECG diagnosis by building up on a current state-of-the-art model based on ECG signals by using multimodal data. Apart from the signal, text from ECG medical reports is given to the model, which leverages both types of data. In short, two models are combined under the Contrastive Language-Image Pre-training (CLIP) framework(17): the first is a signal encoder based on a Convolutional Neural Networks (CNN), previously proposed by (18), and the second is a specialized version of BERT to work with biological data written in Portuguese(21). CLIP is a contrastive learning method (a category of self-supervised learning), where the idea is to learn a joint representation of both data modalities such that similar instances are close together in the representation space, while dissimilar instances are far apart.

The model is tested in the CODE-15% dataset(19) to identify six different ECG abnormalities, and the results show an improvement in the state-of-the-art model, especially in low data context, i.e., where unlabeled data for pre-training is available but context labeled data for fine-tuning is scarce.

Related work

There has been a wide range of works that proposed deep learning techniques to diagnose or find abnormal patterns in ECG signals(11,18). However, most of these methods rely on ECG signal data while ignoring other interesting information, such as the medical report or patient information. In this direction, this paper proposes to use joint information from both ECG signals and medical reports. As there is a close relationship between medical exams and reports(14), (3) explored self-supervised learning (SSL) in vision-language processing by pre-training image encoder models with their corresponding reports, achieving state-of-the-art performance in various biomedical tasks.

Apart from the task of ECG classification, large language models (LLMs) hold immense potential in extracting medical knowledge from reports. However, existing models often face limitations such as being closed-source or having restricted scalability(6). To address this point, MEDITRON, an open-source suite of LLMs specifically tailored to the medical domain, was introduced, demonstrating significant performance gains across major medical benchmarks. Similarly, Med-Flamingo, a multimodal few-shot learner, was designed to excel in generative medical visual question-answering tasks(15).

The relevance of clinical natural language processing tasks is further emphasized by (21), who introduced BioBERTpt, a deep contextual embedding model trained on clinical narratives and biomedical-scientific papers in Brazilian Portuguese. Their evaluation showcased improved performance in biomedical named-entity recognition tasks, underscoring the importance of domain-specific literature in enhancing NLP models.

Overall, ongoing research in ECG classification aims to develop more interpretable architectures and techniques while integrating diverse data types, including text, to enhance medical knowledge within models[4]. This fosters trust and promotes knowledge exchange between machine learning experts and healthcare professionals, ultimately benefiting patient outcomes and healthcare delivery.

Methodology

This section describes the methodology proposed for multimodal ECG classification, illustrated in Figure (2). The process works in two phases. First, pairs of ECG signals and their textual reports are used to pre-train a deep neural network in an unsupervised manner. In this phase, the weights are updated according to the similarity of the ECG and text representations, as detailed later in this section. Next, the weights learned in this first phase are transferred and used for a second supervised classification task, which explicitly uses ECG labels, in a process we call fine-tuning.

Dataset

This paper used the CODE-15% dataset for ECG classification. This is a subset of 15% of patients from the Clinical Outcomes in Digital Electrocardiology (CODE) dataset, which contains more than 2 million exams labeled and analyzed by the Telehealth Network of Minas Gerais[2]. The dataset corresponds to 345,779 ECGs from 233,770 patients. The labels indicate the non-exclusive presence of six categories, each treated independently as six uncorrelated binary problems: 1aAVb (1st degree AV block), 1.8% of the samples; RBBB (right bundle branch block), 3.0% of the samples; LBBB (left bundle branch block), 1.9% of the samples; SB (sinus bradycardia), 1.7% of the samples; AF (atrial fibrillation), 2.4% of the samples; and ST (sinus tachycardia), 2.2% of the samples. Note that this is an unbalanced classification problem.

In addition to the signals, medical reports were used for pre-training the model with textual information. The reports are structured into three main parts, as illustrated in Figure 3: the description of the signal, the conclusion of the diagnosis, and the one-hot encoding of the classification. During model pre-training, only the description was used without any label information; during model fine-tuning, only the labels were used.

Figure 2 – Methodology proposed for multimodal ECG classification.

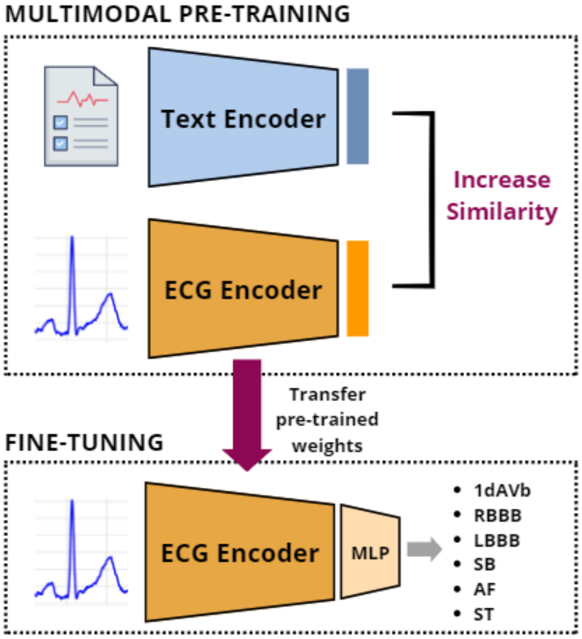
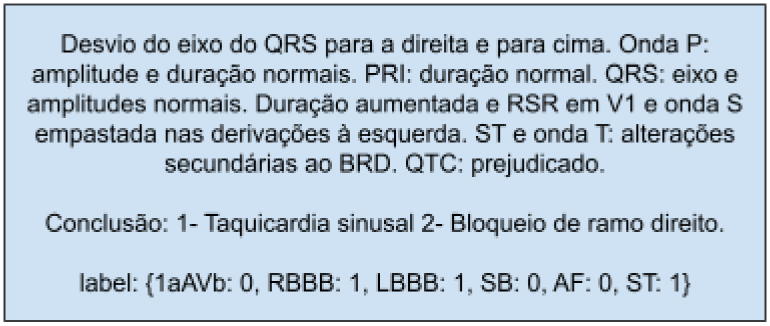


Figure 3 – Example of an ECG medical report (written in Portuguese).



Desvio do eixo do QRS para a direita e para cima. Onda P: amplitude e duração normais. PRI: duração normal. QRS: eixo e amplitudes normais. Duração aumentada e RSR em V1 e onda S empastada nas derivações à esquerda. ST e onda T: alterações secundárias ao BRD. QTC: prejudicado.

Conclusão: 1- Taquicardia sinusal 2- Bloqueio de ramo direito.

label: {1aAVb: 0, RBBB: 1, LBBB: 1, SB: 0, AF: 0, ST: 1}

The first part of the report, the description module, was automatically obtained from the ECG analysis program of the University of Glasgow(12,13). The report describes the main features of the ECG signal, including the amplitude and duration of the characteristic waves, and also gives a diagnostic interpretation. However, only the description was used to simulate unlabeled data, which is an important consideration for the pre-training stage.

Deep learning models

As shown in Figure 2, there are two different models to handle the two types of data: the ECG signal encoder and the text encoder. For the signal encoder, we used the state-of-the-art CNN modeled in (18), which is a ResNet[9] adapted for one-dimensional signals. For the text coder, BioBERTpt was used, which consists of a BERT model[7] pre-trained on the clinical domain for the Portuguese language. The learned representation of this model for the texts is expected to be coherent with the subsequent classification task.

In the pre-training phase, the representation of the signal H_s is approximated to the representation of the text H_t by a cost function borrowed from the multimodal CLIP model[17]. For each batch of size n, a distance matrix $n\,x\,n$ is constructed, where each element i, j corresponds to the cosine similarity of $H_s(i)$ and $H_s(j)$. The cost to be optimized is the cross-entropy of the diagonal of this distance matrix so that coincident pairs are placed closer and dissident pairs are placed far apart which is the principle of contrastive learning. In the fine-tuning phase, the cost function is the independent binary cross entropy for each category.

In contrast to the multimodal CLIP model, where both encoders are trained simultaneously, in this study only the signal encoder was trained, since BioBERTpt is already pre-trained in the medical domain. In this way, the learned representation of the text H_t is previously computed for each sample in the dataset. Since the text encoder's weight is virtually frozen, these embeddings (weights) serve as “supervisors” in the pre-training phase for the signal encoder, i.e., the model tries to replicate the same embedding for the ECG input.

Results and discussion

The experiments were performed using the CNN model proposed in (18), employing as parameters 4 blocks with (64, 128, 192, 256) channels and a dropout rate of 0.2. BioBERTpt was run with its default parameter values, which are 12 encoders with 12 bidirectional self-attention heads.

For text, we pre-trained the signal encoder with the previously computed text embedding H_t. This model is compared to a random initialization of the weights as in (18). All experiments were compared using four evaluation metrics in the test set: accuracy, which is the ratio of true positives (TP) to the total number of samples; precision, as the ratio of TP to the sum of TP and false positives (FP); recall, similar to precision, but the denominator is the sum of TP and false negatives (FN); and the F1 score, which is our focus because it balances precision and recall being the harmonic mean of these metrics.

Experimental Setup

We separated 10% of the full dataset for model validation and 5% for model testing, making sure that different sets did not share exams from the same patient. The validation set was used to determine the thresholds of the model's output logit function. Table 1 shows the experimental setup in terms of data usage. For pre-training, all experiments use the full training data (85% of the full dataset). For fine-tuning, as a baseline, we first used this same dataset for fine-tuning the model for the classification task. Next, as it is common in the self-supervised learning literature, we simulate the effect of using a large corpus of unlabeled data for multimodal pre-training while reducing the corpus for the fine-tuning stage, generating the scarce data setup.

In this latter experiment, we gradually reduce the training and validation sets used in fine-tuning, arbitrarily following a negative exponential curve to determine the proportions. Since there is a minor data leakage when using the same data in the multimodal pre-training and later fine-tuning, even though the labels are not shared, we also designed an *unseen data* experiment, where we fine-tune using half of the validation set, which was not seen in the multimodal pre-training. The other half of the validation set is used by default for thresholding and the old training set is ignored.

Table 1 – Percentage of the dataset used for multimodal pre-training and fine-tuning of ECG signals. Note that in the unseen data experiment, there is no sharing of samples, the training and validation sets are randomly sampled from the standard validation set*. The test data is the same for each experiment.

Experiment	Pre-training	Fine-tuning	
	Training data	Training	Validation
Full data	85%	85%	10%
Scarce data	85%	37%	3.7%
	85%	5%	0.5%
	85%	0.7%	0.07%
Unseen data	85%	5%*	5%*

Full dataset experiment

Table 2 shows the results for the full dataset being used to fine-tune the model (first line in Table 2), for each of the six classes considered in the CODE dataset. We compare the results of pre-training the model with both signals and text data (Multimodal) with those obtained by pre-training with signals only (Signal) in the test set. As expected, the gain is marginal, as the multimodal pre-training has clear advantages when there is a large amount of unlabeled data for pre-training the model and fewer instances for fine-tuning, the scenario simulated in the next experiment.

Scarce data

As previously explained, in this second experiment we gradually reduce the proportion of the training and validation sets used for fine-tuning. We sample 36.8%, 4.98% and 0.67% of the corpus arbitrarily following a negative exponential curve (see Table 2). Again, we assessed the four metrics on the test set, which is the same for all experiments. Tables 3, 4 and 5 show the results when reducing data for fine-tuning. Notice that the effect of the multimodal pre-training for preserving the model's performance even with an extremely small training set for fine-tuning is clear. Figure 4 shows the evolution of the mean F1 score for the four ratios tested, including the baseline experiment. The same figure shows the potential of multimodal pre-training in scenarios where labeled data may be scarce.

Table 2 – Evaluation metrics in the full dataset experiment when comparing the pre-training with only the signal and the multimodal approach proposed in the test set.

Class	Accuracy		Precision		Recall		F1 Score	
	Baseline	Pretrain	Baseline	Pretrain	Baseline	Pretrain	Baseline	Pretrain
1dAVb	0.987	0.988	0.532	0.574	0.621	0.652	0.573	0.610
RBBB	0.992	0.992	0.779	0.783	0.913	0.894	0.841	0.835

LBBB	0.995	0.995	0.808	0.837	0.823	0.780	0.816	0.808
SB	0.987	0.986	0.642	0.602	0.770	0.816	0.700	0.693
AF	0.994	0.995	0.868	0.863	0.727	0.802	0.791	0.831
ST	0.988	0.988	0.651	0.660	0.891	0.878	0.752	0.754
Mean	0.990	0.991	0.713	0.720	0.791	0.804	0.746	0.755

Figure 4 – Evolution of the mean F1 score over the six categories when reducing the proportion of the training and validation set.

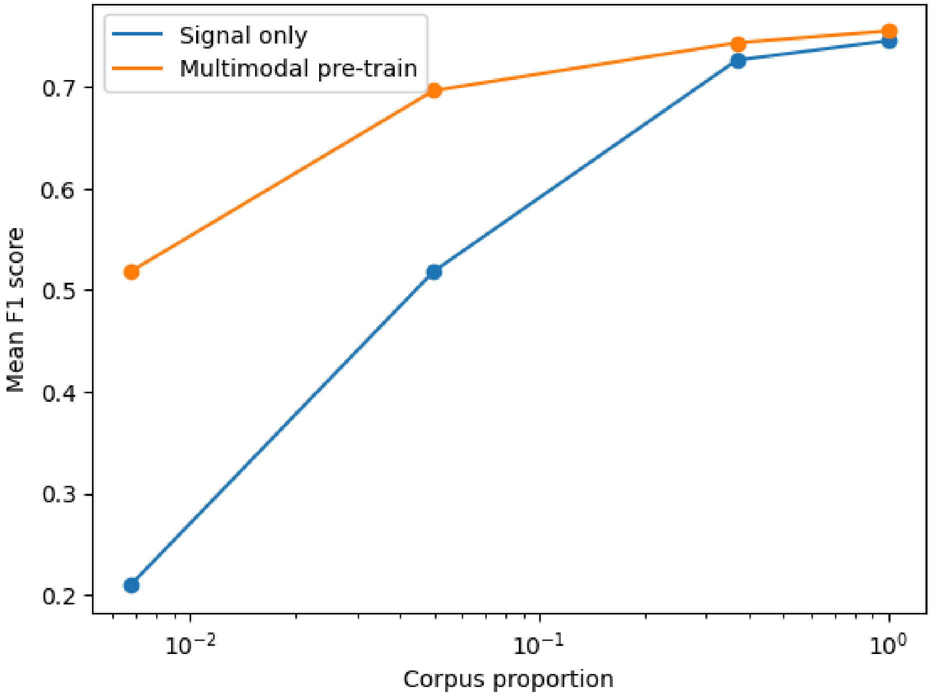


Table 3 – Evaluation metrics with approximately 37% of the dataset experiment when comparing the pre-training with only the signal and the multimodal approach proposed in the test set.

Class	**Accuracy**		**Precision**		**Recall**		**F1 Score**	
	Baseline	Pretrain	Baseline	Pretrain	Baseline	Pretrain	Baseline	Pretrain
1dAVb	0.984	0.986	0.459	0.514	0.658	0.671	0.541	0.582
RBBB	0.992	0.992	0.803	0.784	0.882	0.909	0.841	0.842
LBBB	0.994	0.994	0.770	0.798	0.835	0.817	0.801	0.807
SB	0.986	0.985	0.619	0.602	0.719	0.724	0.665	0.657
AF	0.993	0.994	0.846	0.760	0.686	0.884	0.759	0.817
ST	0.989	0.988	0.688	0.651	0.826	0.900	0.751	0.755
Mean	0.990	0.990	0.698	0.685	0.768	0.817	0.726	0.743

Table 4 – Evaluation metrics with approximately 5% of the dataset experiment when comparing the pre-training with only the signal and the multimodal approach proposed in the test set.

Class	**Accuracy**		**Precision**		**Recall**		**F1 Score**	
	Baseline	Pretrain	Baseline	Pretrain	Baseline	Pretrain	Baseline	Pretrain

1dAVb	0.957	0.988	0.084	0.586	0.199	0.553	0.118	0.569
RBBB	0.988	0.991	0.683	0.745	0.878	0.924	0.769	0.825
LBBB	0.992	0.992	0.832	0.667	0.604	0.878	0.700	0.758
SB	0.985	0.984	0.606	0.561	0.645	0.829	0.625	0.669
AF	0.965	0.993	0.197	0.876	0.419	0.616	0.268	0.724
ST	0.986	0.987	0.663	0.787	0.600	0.530	0.630	0.634
Mean	0.979	0.989	0.511	0.704	0.557	0.722	0.518	0.696

Table 5 – Evaluation metrics with approximately 0.7% of the dataset experiment when comparing the pre-training with only the signal and the multimodal approach proposed in the test set.

Class	**Accuracy**		**Precision**		**Recall**		**F1 Score**	
	Baseline	Pretrain	Baseline	Pretrain	Baseline	Pretrain	Baseline	Pretrain
1dAVb	0.986	0.986	0.000	0.875	0.000	0.043	0.000	0.083
RBBB	0.978	0.982	0.531	0.580	0.616	0.928	0.570	0.713
LBBB	0.977	0.990	0.341	0.865	0.610	0.390	0.438	0.538
SB	0.019	0.981	0.019	0.498	1.000	0.618	0.038	0.551
AF	0.969	0.976	0.071	0.376	0.081	0.878	0.076	0.526
ST	0.837	0.986	0.075	0.617	0.613	0.800	0.134	0.697
Mean	0.794	0.983	0.173	0.635	0.487	0.609	0.209	0.518

Unseen data

Since there is a small data leakage when pre-training and fine-tuning the models with the same dataset, in this final experiment we separate half of the validation set for fine-tuning the model, in a way that no data used for training is shown to the model again during fine-tuning. Table 6 shows the results, which are very similar to those of Table 4, the signal-only model fails to classify the 1dAVb category with an F1 score of 0.118, and the difference in the mean F1 score is close to 0.2 as well. This happens because half of the validation set corresponds to 5% of the full corpus. In this case, these similar results show that the model is as robust for unseen data.

Table 6 – Evaluation metrics for the unseen data experiments, half of the old validation set was used now as the training, this corresponds to 5% of the full corpus.

Class	**Accuracy**		**Precision**		**Recall**		**F1 Score**	
	Baseline	Pretrain	Baseline	Pretrain	Baseline	Pretrain	Baseline	Pretrain
1dAVb	0.977	0.984	0.124	0.452	0.093	0.640	0.106	0.530
RBBB	0.989	0.992	0.734	0.804	0.810	0.859	0.770	0.831
LBBB	0.992	0.994	0.747	0.760	0.720	0.829	0.733	0.793
SB	0.983	0.986	0.565	0.642	0.599	0.668	0.582	0.655

AF	0.981	0.991	0.316	0.677	0.215	0.756	0.256	0.714
ST	0.985	0.987	0.608	0.632	0.774	0.835	0.681	0.719
Mean	0.985	0.989	0.516	0.661	0.535	0.765	0.521	0.707

Conclusion

This paper proposed a multimodal approach to pre-training a model using both signals and text from ECGs. The results obtained show a great improvement in multimodal pre-training with unlabeled text data when compared with the use of only the signal of the ECG. The results of the proposed multimodal approach stand out in the scarce data setup, which reflects most cases of real-world scenarios. By using the same state-of-the-art CNN as the signal encoder, we observe a difference in the mean F1 score of 0.01, 0.02, 0.18, and 0.31, respectively, that is, the F1 improves as data scarcity for fine-tuning increases.

Deep learning models are already a reality in the medical universe and should not be used to replace medical professionals, but rather to optimize the workflow and eliminate possible human biases. The proposed model can be easily integrated into remote diagnosis systems for initial classification and queue prioritization.

The major limitation of our work is the simulation of a data-scarce scenario where, in reality, in the CODE dataset we know the labels. However, this is an important test to be executed before our next step. We are currently preparing medical reports from the Glasgow analysis program to test into fully unlabeled exams.

Acknowledges

This work was made possible by the availability of data from the Telehealth Network of Minas Gerais repository and our always productive discussions. We would also like to thank the Brazilian agencies CNPq, CAPES and FAPEMIG for their financial support.

References

1. M. Alkmim, A. Ribeiro, G. Carvalho, M. Pena, R. Figueira, and M. Carvalho. Success factors and difficulties for implementation of a telehealth system for remote villages: Minas telecardio project case in brazil. J Health Technol Appl, 5(3):197–202, 2007.
2. M. B. Alkmim, R. M. Figueira, M. S. Marcolino, C. S. Cardoso, M. P. d. Abreu, L. R. Cunha, D. F. d. Cunha, A. P. Antunes, A. G. d. A. Resende, E. S. Resende, et al. Improving patient access

to specialized health care: the telehealth network of minas gerais, brazil. Bulletin of the World Health Organization, 90:373–378, 2012.

3. S. Bannur, S. Hyland, Q. Liu, F. Perez-Garcia, M. Ilse, D. C. Castro, B. Boeck- ing, H. Sharma, K. Bouzid, A. Thieme, et al. Learning to exploit temporal structure for biomedical vision-language processing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 15016–15027, 2023.
4. I. Bica, A. Ili´c, M. Bauer, G. Erdogan, M. Boˇsnjak, C. Kaplanis, A. A. Grit- senko, M. Minderer, C. Blundell, R. Pascanu, et al. Improving fine-grained understanding in image-text pre-training. arXiv preprint arXiv:2401.09865, 2024.
5. J. Chai, H. Zeng, A. Li, and E. W. Ngai. Deep learning in computer vision: A critical review of emerging techniques and application scenarios. Machine Learning with Applications, 6:100134, 2021.
6. Z. Chen, A. H. Cano, A. Romanou, A. Bonnet, K. Matoba, F. Salvi, M. Pagliardini, S. Fan, A. K ¨opf, A. Mohtashami, et al. Meditron-70b: Scaling medical pretraining for large language models. arXiv preprint arXiv:2311.16079, 2023.
7. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.
8. A. Esteva, K. Chou, S. Yeung, N. Naik, A. Madani, A. Mottaghi, Y. Liu, E. Topol, J. Dean, and R. Socher. Deep learning-enabled medical computer vision. NPJ digital medicine, 4(1):5, 2021.
9. K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recogni- tion. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016.
10. P. A. Jennett, L. A. Hall, D. Hailey, A. Ohinmaa, C. Anderson, R. Thomas, B. Young, D. Lorenzetti, and R. E. Scott. The socio-economic impact of telehealth: a systematic review. Journal of telemedicine and telecare, 9(6): 311–320, 2003.
11. E. M. Lima, A. H. Ribeiro, G. M. Paix˜ao, M. H. Ribeiro, M. M. Pinto-Filho, P. R. Gomes, D. M. Oliveira, E. C. Sabino, B. B. Duncan, L. Giatti, et al. Deep neural network-estimated electrocardiographic age as a mortality predictor. Nature communications, 12(1):5117, 2021.
12. P. Macfarlane, B. Devine, S. Latif, S. McLaughlin, D. Shoat, and M. Watts. Methodology of ecg interpretation in the glasgow program. Methods of information in medicine, 29(04):354–361, 1990.
13. P. Macfarlane, B. Devine, and E. Clark. The university of glasgow (uni-g) ecg analysis program. In Computers in Cardiology, 2005, pages 451–454. IEEE, 2005.
14. P. Messina, P. Pino, D. Parra, A. Soto, C. Besa, S. Uribe, M. Andia, C. Tejos, C. Prieto, and D. Capurro. A survey on deep learning and explainability for automatic report generation from medical images. ACM Computing Surveys (CSUR), 54(10s):1–40, 2022.
15. M. Moor, Q. Huang, S. Wu, M. Yasunaga, C. Zakka, Y. Dalmia, E. Reis, P. Rajpurkar, and J. Leskovec. Med-flamingo: a multimodal medical few- shot learner (2023). URL: https://arxiv. org/abs/2307.15189, 2023.
16. D. W. Otter, J. R. Medina, and J. K. Kalita. A survey of the usages of deep learning for natural language processing. IEEE transactions on neural net- works and learning systems, 32(2):604–624, 2020.
17. A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, et al. Learning transferable visual models from natural language supervision. In International conference on machine learning, pages 8748–8763. PMLR, 2021.
18. A. H. Ribeiro, M. H. Ribeiro, G. M. Paix˜ao, D. M. Oliveira, P. R. Gomes, J. A. Canazart, M. P. Ferreira, C. R. Andersson, P. W. Macfarlane, W. Meira Jr, et al. Automatic diagnosis of the 12-lead ecg using a deep neural network. Nature communications, 11(1):1760, 2020.

19. A. H. Ribeiro, G. Paixao, E. M. Lima, M. H. Ribeiro, M. M. Pinto Filho, P. R. Gomes, D. M. Oliveira, W. Meira Jr, T. B. Schon, and A. L. P. Ribeiro. Code-15%: A large scale annotated dataset of 12-lead ecgs. Zenodo, Jun, 9, 2021.
20. G. A. Roth, C. Johnson, A. Abajobir, F. Abd-Allah, S. F. Abera, G. Abyu, M. Ahmed, B. Aksut, T. Alam, K. Alam, et al. Global, regional, and national burden of cardiovascular diseases for 10 causes, 1990 to 2015. Journal of the American college of cardiology, 70(1):1–25, 2017.
21. E. T. R. Schneider, J. V. A. de Souza, J. Knafou, L. E. S. e. Oliveira, J. Co- para, Y. B. Gumiel, L. F. A. d. Oliveira, E. C. Paraiso, D. Teodoro, and C. M. C. M. Barra. BioBERTpt - a Portuguese neural language model for clinical named entity recognition. In Proceedings of the 3rd Clinical Natural Language Processing Workshop, pages 65–72, Online, Nov. 2020. Association for Computational Linguistics. URL https://www.aclweb.org/anthology/2020.clinicalnlp-1.7.
22. S. Wu, K. Roberts, S. Datta, J. Du, Z. Ji, Y. Si, S. Soni, Q. Wang, Q. Wei, Y. Xiang, et al. Deep learning in clinical natural language processing: a methodical review. Journal of the American Medical Informatics Association, 27(3):457–470, 2020.