

Non-invasive mean Pulmonary Artery Pressure Prediction using Multi-Modal Feature Fusion of Chest X-ray and ECG

Siyeop Yoon¹, Jerome Charton¹, Michal Grzeszczyk¹, Arkadiusz Sitek¹, Hui Ren¹, Quangzheng Li¹, and Kei Nakata^{2,†}

¹ Department of Radiology & Center for Advanced Medical Computing and Analysis, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

² Sapporo Medical University
k.nakata@sapmed.ac.jp

Abstract. Pulmonary hypertension (PH) is a life-threatening condition marked by elevated mean pulmonary arterial pressure (mPAP), with high morbidity and mortality. Right heart catheterization (RHC) is the gold standard for mPAP measurement because it provides direct and accurate hemodynamic assessment. However, RHC necessitates specialized facilities and continuous monitoring, which limits its accessibility, especially in community hospitals. This study introduces a deep learning model that leverages DINOv2 to estimate mPAP from chest X-ray and ECG images. The DINOv2-based chest X-ray encoder is fine-tuned to extract high-dimensional representations followed by feature fusion with those extracted from ECG images using a light-weight convolutional neural network, enabling the model to generate accurate mPAP predictions. The model was trained on 290 RHC invasive mPAP measurements from 163 patients and subsequently tested on 71 measurements from 38 patients at a town-based hospital. Performance evaluation using Bland-Altman analysis and regression correlation with invasive mPAP measurements showed low bias (-1.55 mmHg, limits of agreement = [-21.34, 18.24]), and moderate agreement ($R^2 = 0.43$). Moreover, the model demonstrates the potential for tracking long-term disease progression trajectories by correlating longitudinal changes in imaging features with mPAP variations. The model is deployable as a web tool, enabling scalable, non-invasive PH screening and monitoring with routine CXR and ECG, particularly in settings with limited RHC access.

Keywords: Pulmonary Hypertension · Multi-Modal Deep Learning · Vision Foundation Models · Low-Resource Healthcare

1 Introduction

Pulmonary hypertension (PH) is a progressive and life-threatening cardiovascular disorder defined by elevated mean pulmonary artery pressure (mPAP) equal

[†] Corresponding Author.

or higher than 20 mmHg at rest. PH proceeds to right ventricular failure and mortality if undiagnosed or untreated [10]. Early identification of PH is critical since targeted therapies can substantially improve patient outcomes. Nevertheless, diagnosis is frequently delayed due to non-specific symptoms and limited access to advanced diagnostic tools in many healthcare settings [6].

The current gold standard for PH diagnosis and mPAP measurement is right heart catheterization (RHC). Although RHC provides a direct hemodynamic assessment, its invasive nature, associated procedural risks, high costs, and limited availability—especially in low-resource environments—pose significant challenges [7]. In non-invasive clinical practice, transthoracic echocardiography (TTE) is commonly employed to estimate pulmonary pressures via Doppler-derived tricuspid regurgitation velocity. However, TTE suffers from operator dependency, unreliability in up to 30% of cases, and a tendency to overestimate pulmonary pressures in patients with concurrent lung disease [4,11].

Recent advances in deep learning have demonstrated the feasibility of predicting PH and elevated pulmonary pressures from non-invasive data. For instance, machine learning models applied to 12-lead ECGs have identified latent markers of pulmonary vascular pathology, while deep learning models trained on cardiac magnetic resonance imaging (CMR) have achieved accurate PH classification that slightly outperforms standard diagnostic metrics, delivering results within seconds [8,5]. Although resource intensive CMR was used, these studies underscore the potential of AI-driven, non-invasive PH detection. Moreover, previous models have typically relied on a single imaging modality, which limits their ability to capture complementary information that could be obtained by integrating multiple data sources. A promising solution lies in multi-modal AI approaches that integrate both structural and electrophysiological markers, such as those derived from chest X-ray (CXR) and ECG data. Although CXR and ECG lack the specificity of traditional diagnostics, they can capture secondary disease markers—including pulmonary artery enlargement, right ventricular hypertrophy, and electrical conduction abnormalities—that are often challenging to quantify manually. We hypothesize that deep learning models can extract and interpret these subtle patterns, correlating them with hemodynamic abnormalities and disease progression [11].

The field of computer vision has been transformed by foundation models—large-scale, self-supervised models that learn general-purpose visual representations. Among these, DINOv2, a self-supervised Vision Transformer trained on 142 million images, has exhibited strong generalization across diverse tasks without the need for labeled data during pre-training [3,2,9]. The robustness and scalability of DINOv2 render it particularly attractive for medical imaging applications, where labeled data are frequently limited.

In this study, we introduce a multi-modal deep learning framework for PH screening and monitoring that combines chest X-rays and images of ECG signals to predict continuous mPAP values non-invasively. Our model is designed for deployment in low-computation environments, utilizing standard X-ray and scanned ECG images, and it can be accessed publicly via web platforms and

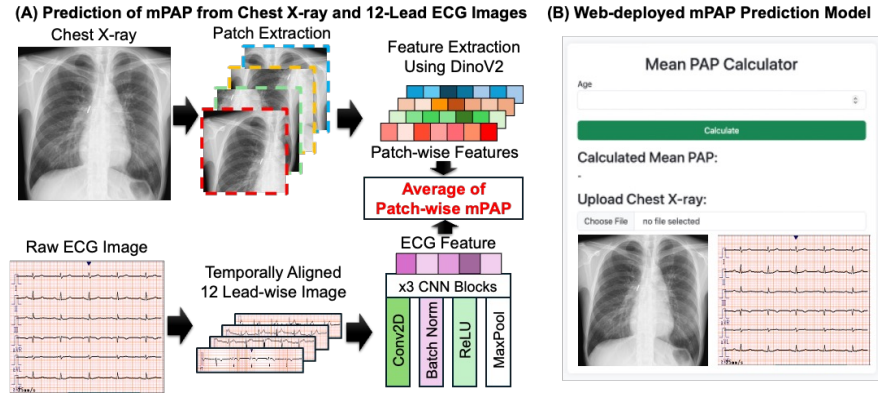


Fig. 1. Multi-Modal Deep Learning Framework for mPAP Prediction and Web-Based Deployment. (A) The proposed pipeline for estimating mean pulmonary arterial pressure (mPAP) from chest X-ray and 12-lead electrocardiogram (ECG) images. A pre-trained foundation model (DINOv2) extracts high-level features from the chest X-ray, while a convolutional neural network (CNN) processes temporally aligned ECG lead images. The extracted feature representations are combined to predict mPAP. (B) A web-based mPAP prediction tool that allows users to upload chest X-ray and ECG images for real-time estimation of mPAP, demonstrating the feasibility of deploying the model for potential clinical use in resource-limited settings.

open-source repositories, ensuring its availability in resource-limited settings (Fig. 1). We propose a DINOv2-based foundation model for chest X-ray analysis, integrated with a convolutional ECG encoder, to directly estimate mPAP. Model performance is validated using Bland-Altman analysis, linear regression, and subgroup analyses (e.g., smokers versus non-smokers) to assess predictive consistency. Additionally, we demonstrate the model’s capacity to track estimated mPAP trajectory over time and compare with repeated invasive measurements in PH patients. We assess the feasibility of deploying the model via a web-based prototype, enabling broader testing and feedback from the various clinical and research settings.

2 Methods

2.1 Data Description

This study utilized a retrospective dataset from a Sapporo Medical University Hospital. This study was approved by the Institutional Review Board (IRB) (Protocol: 2358678) and utilized retrospectively collected data. Our dataset consists of patients clinically diagnosed with PH who underwent both chest X-ray and ECG examinations close to the time of invasive hemodynamic testing. Inclusion criteria required that a frontal chest radiograph (posteroanterior view) and

a standard 12-lead ECG were available within a 7-days of a right heart catheterization measuring mPAP. The final dataset included 201 PH patients (age 65 ± 16 years, 54% female), with a total of 361 records of mPAP measurements, X-rays, and ECG. PH etiologies in the cohort were mixed, including Group 1 PAH, Group 2 PH (due to left heart disease), and Group 3 PH (due to lung disease) [1], reflecting a real-world case mix.

For each patient, we obtained the digital chest X-ray image and a digitized ECG recording. The 12-lead ECGs (originally recorded on paper or electronically) were converted into a scanned image format for input into the vision model. For training, the images underwent preprocessing: chest X-rays were rescaled to a uniform size (512x512 pixels) and normalized, and 12-lead ECG images were similarly resized to a uniform size (224x48) per lead, and concatenated to channel dimension. We paired each X-ray with its corresponding ECG image and the ground-truth mPAP value (measured by RHC). The 361 records were split into training, validation, and test sets (70% train, 10% val, 20% test) ensuring that each patient appears in only one set.

2.2 Multi-Modal mPAP Regression Model

The proposed model estimates mPAP by integrating information from chest X-rays and 12-lead ECG images. The chest X-ray branch employs a DINOv2-based Vision Transformer, pre-trained on 142 million images, as a robust feature extractor for pulmonary and cardiovascular structures. In our implementation, the DINOv2 backbone is wrapped within a custom fine-tuning module that reshapes its output into a 1024-dimensional embedding.

During **training**, the DINOv2 is optimized on 224×224 chest patches obtained by randomly cropping from 512×512 resized images. When ECG images are provided, the ECG feature extractor branch extracted a 128-dimensional vector via a series of convolutional layers. The first layer accepts an input with 36 channels (representing the 12-lead ECG image with RGB channels) and outputs 64 feature maps using a kernel size of 3×3 , stride 1, and padding 1. This is followed by batch normalization, ReLU activation, and a 2×2 max-pooling operation. The channel depth increases to 128 and 256 for the second and third blocks, respectively.

The final ECG-features are projected with an adaptive average pooling layer reducing the spatial dimensions to 1×1 , and a fully connected layer projects the 256-dimensional output to a 128-dimensional feature vector. This branch, therefore, yields a 128-dimensional embedding that captures key waveform features indicative of right ventricular strain and electrical conduction abnormalities.

For each 224×224 chest patch, the DINOv2 backbone produces a 1024-dimensional feature, which is concatenated with the ECG feature. The individual feature vectors of the patch are concatenated into a single fused feature vector with a total dimensionality of $d_{\text{fused}} = d_{\text{chest}} + d_{\text{ECG}} = 1024 + 128 = 1152$. This fused representation is processed by a regression head that first reduces the dimensionality to 592 through a fully connected layer, applies a ReLU activation,

and finally maps the features to a single scalar output representing the mPAP (in mmHg).

The networks are trained at the same time using the Mean Squared Error (MSE) loss defined as $L = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2$, where $\hat{y}_i$ is the predicted mPAP and y_i is the ground-truth measurement obtained from right heart catheterization. Optimization is performed for 20 epochs using the Adam optimizer with a learning rate of 5×10^{-5} .

In **validation/testing**, the full 512×512 chest X-ray is partitioned into a 3×3 grid of 224×224 patches. A patch-level regressor then predicts an mPAP value for each patch; the final chest-based mPAP estimate is computed as the average of patch-wise predictions.

For the **model deployment**, we developed a web-based mPAP prediction tool that enables users to upload chest X-ray and 12-lead ECG images for real-time estimation of mPAP. The web-based mPAP prediction tool leverages a multi-modal deep learning model that integrates DINOv2-based Vision Transformer with complementary ECG-derived features, estimating mPAP. This tool utilizes the Hugging Face API for deployment in resource-limited settings, thereby offering a cost-effective, rapid, and accessible solution for clinical decision-making and patient management (<https://github.com/siyeopyoon/PulmoFusion-mPAP> and <https://jcharton-mean-pap-api.hf.space>).

Evaluation Metrics To assess the impact of different chest X-ray feature extractors on mPAP prediction accuracy, we conducted experiments comparing DINOv2, ResNet34, and VGG16 backbones. Moreover, we evaluated the performance of the multi-modal framework both with and without the inclusion of the ECG encoder. Evaluation of the model was performed using a suite of quantitative analyses to assess the agreement and accuracy of the predicted mPAP values against the reference measurements obtained via RHC. A Bland-Altman plot was generated to identify systematic biases and to define the limits of agreement between the two methods. In addition, the Mean Absolute Error (MAE) was computed as $MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|$, providing a direct measure of the average discrepancy (in mmHg) between predictions and true values. Finally, a linear regression analysis was conducted between the predicted and catheterized mPAP measurements, yielding key statistical parameters such as the coefficient of determination (R^2) and the regression slope.

3 Results

Figure 2 presents a Bland-Altman plot that compares the differences between predicted and measured mPAP values against their averages. The plot includes a bias line representing the mean difference (-1.55 mmHg) and dashed lines that indicate the 95% limits of agreement (from -21.34 to 18.24 mmHg). Table 1 details the performance of different deep learning models for mPAP prediction. Standalone imaging models (VGG16, Resnet34, DinoV2) achieved regression slopes of 0.15, 0.03, and 0.09 with correlation coefficients (R^2) of 0.13, 0.02, and

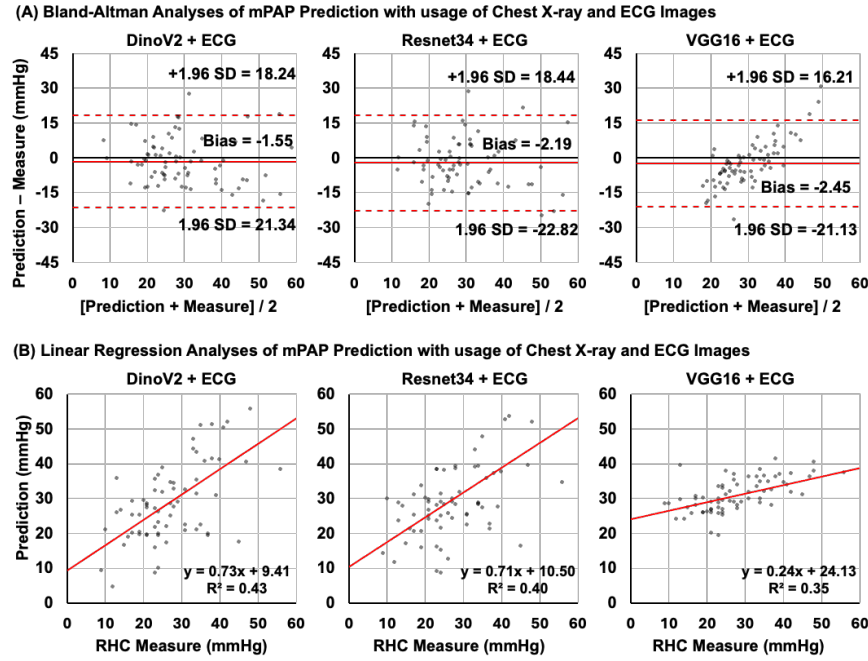


Fig. 2. The measurement agreement of mean pulmonary artery pressure between right heart catheterization and AI-based prediction with the usage of Chest X-ray and ECG image. (A) Bland-Altman plots show bias and limits of agreement (± 1.96 SDs), the solid red line and red dotted lines, respectively. In (B) linear regression the correlation coefficient (R^2) and slope (y) are shown.

0.06, respectively. The incorporation of ECG data led to improved performance; notably, the DinoV2+ECG model reached a regression slope of 0.73 and an R^2 of 0.43, along with a mean absolute error of 8.20 ± 6.03 and a bias of -1.55 mmHg (limits of agreement: -21.34 to 18.24 mmHg). These numerical results demonstrate that the integration of ECG with chest X-ray data enhances the accuracy of non-invasive mPAP estimation.

To assess the model’s ability to monitor changes over time, we selected a patient from our cohort who underwent multiple imaging studies and corresponding RHC measurements over a period from Day 0 to Day 4513. For this case, our trained model was applied to each paired chest X-ray and ECG image, and predicted mPAP values were recorded at each time point. As illustrated in Figure 3, the patient initially showed high mPAP values together with clear ECG indicators of right ventricular strain (highlighted by the red circle in panel (B)). Over time, as the ECG features indicative of RV strain diminished, the predicted mPAP values decreased correspondingly. This observation provides a concrete example of how changes in RV strain may be associated with alterations in pulmonary arterial pressure.

Table 1. Performance evaluation of deep learning models for mPAP prediction. Reported are mean absolute error (MAE $\pm$ SD), regression slope, R^2 , and bias with limits of agreement for standalone imaging models (VGG16, Resnet34, DinoV2) and their counterparts augmented with ECG data.

Method	MAE $\downarrow$	Regression Slope	Correlation R^2	Bias [LoA%]
VGG16	8.77 ± 6.63	0.15	0.13	-2.41 [-26.52, 18.69]
Resnet34	10.14 ± 6.46	0.03	0.02	-3.92 [-26.52, 18.69]
DinoV2	9.05 ± 6.70	0.09	0.06	-1.28 [-23.31, 20.75]
VGG16+ECG	7.53 ± 6.26	0.24	0.35	-2.45 [-21.13, 16.21]
Resnet34+ECG	8.62 ± 6.35	0.71	0.40	-2.19 [-22.82, 18.44]
DinoV2+ECG	<u>8.20 ± 6.03</u>	<u>0.73</u>	<u>0.43</u>	-1.55 [-21.34, 18.24]

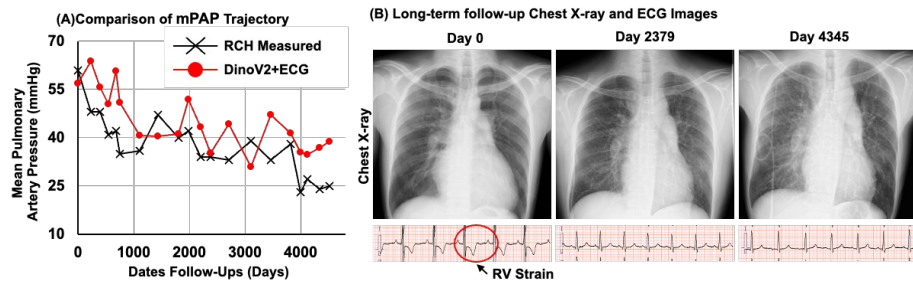


Fig. 3. Comparison of measured mean pulmonary artery pressure (mPAP) from right heart catheterization (black crosses) and predicted mPAP from the proposed DINOv2+ECG model (red circles) over 4513 days, illustrating the framework’s ability to track disease progression. Chest X-ray and ECG images at Day 0, 2379, and 4345 reveal progressive cardiopulmonary changes, including right ventricular strain (red circle) captured by ECG morphology. These findings demonstrate how our proposed model can non-invasively estimate mPAP from routine clinical imaging and follow pulmonary hypertension trajectories using chest X-ray and ECG.

4 Conclusion

We presented a novel approach for non-invasive pulmonary hypertension assessment using a DINOv2 vision foundation model to analyze chest X-ray and ECG images for mPAP prediction. The proposed model demonstrated high accuracy in estimating mean pulmonary arterial pressure, with strong correlation and close agreement to invasive measurements. Through Bland-Altman analysis, we showed minimal bias and clinically acceptable limits of agreement between the model’s predictions and right heart catheterization values. We further illustrated the model’s value by tracking mPAP changes in longitudinal case studies, highlighting potential use in monitoring disease progression or treatment response. It could also be applied retrospectively to prior images to see trends (for example, reading old X-rays to estimate what mPAP might have been before). Importantly, the method is non-invasive and repeatable, making it safe for serial use.

Our model could be integrated into hospital PACS systems or web-based access, as shown in the example website, Figure 1(B).

While the results are encouraging, our study has several limitations. The dataset size (201 patients) is relatively modest, drawn from a single center. This raises concerns about overfitting to specific population characteristics or image acquisition protocols. Although using DINOv2 should improve generalizability, the model should be validated on external cohorts (multi-center data) to ensure it works broadly. Future studies will focus on validating our model across broader populations and integrating additional data modalities to further improve robustness and precision. With additional refinement, this foundation model-based approach has the potential to be implemented as a low-cost, widely accessible solution for improving pulmonary hypertension care. In essence, our study highlights the promise of cross-domain AI models in bridging the gap between simple diagnostic tests and complex invasive measurements for PH management.

Acknowledgments. This study was partially supported by NIH R01HL159183.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Anderson, J.J., Lau, E.M.: Pulmonary hypertension definition, classification, and epidemiology in asia. *JACC: Asia* **2**(5), 538–546 (2022)
2. Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., Joulin, A.: Emerging properties in self-supervised vision transformers. In: *Proceedings of the IEEE/CVF international conference on computer vision*. pp. 9650–9660 (2021)
3. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020)
4. Frea, S., Centofanti, P., Pidello, S., Giordana, F., Bovolo, V., Baronetto, A., Franco, B., Cingolani, M.M., Attisani, M., Morello, M., et al.: Noninvasive assessment of hemodynamic status in heartware left ventricular assist device patients: validation of an echocardiographic approach. *JACC: Cardiovascular Imaging* **12**(7 Part 1), 1121–1131 (2019)
5. Grzeszczyk, M.K., Korzeniowski, P., Alabed, S., Swift, A.J., Trzciński, T., Sitek, A.: Tabmixer: Noninvasive estimation of the mean pulmonary artery pressure via imaging and tabular data mixing. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 670–680. Springer (2024)
6. Hambly, N., Alawfi, F., Mehta, S.: Pulmonary hypertension: diagnostic approach and optimal management. *CMAJ* **188**(11), 804–812 (2016)
7. Hasan, B., Hansmann, G., Budts, W., Heath, A., Hoodbhoy, Z., Jing, Z.C., Koestenberger, M., Meinel, K., Mocumbi, A.O., Radchenko, G.D., et al.: Challenges and special aspects of pulmonary hypertension in middle-to low-income regions: Jacc state-of-the-art review. *Journal of the American College of Cardiology* **75**(19), 2463–2477 (2020)

8. Johns, C.S., Kiely, D.G., Rajaram, S., Hill, C., Thomas, S., Karunasaagarar, K., Garg, P., Hamilton, N., Solanki, R., Capener, D.A., et al.: Diagnosis of pulmonary hypertension with cardiac mri: derivation and validation of regression models. *Radiology* **290**(1), 61–68 (2019)
9. Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., et al.: Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193* (2023)
10. Simonneau, G., Montani, D., Celermajer, D.S., Denton, C.P., Gatzoulis, M.A., Krowka, M., Williams, P.G., Souza, R.: Haemodynamic definitions and updated clinical classification of pulmonary hypertension. *European respiratory journal* **53**(1) (2019)
11. Thenappan, T., Prins, K.W., Pritzker, M.R., Scandurra, J., Volmers, K., Weir, E.K.: The critical role of pulmonary arterial compliance in pulmonary hypertension. *Annals of the American Thoracic Society* **13**(2), 276–284 (2016)