

DINO-Ortho: Self-Supervised Pretrained Neural Network for Orthopedic Image Classification

Rohan Bagchi¹, Nicholas J. Yee^{1,3,4}, Soheil Ashkani-Esfahani^{1,2}, Atta Taseh^{1,2}

INSTITUTIONS:

¹Foot & Ankle Research and Innovation Lab (FARIL), Department of Orthopaedic Surgery, Mass General Brigham, Harvard Medical School, Boston, MA

²Division of Foot and Ankle, Department of Orthopaedic Surgery, Mass General Brigham, Harvard Medical School, Boston, MA

³Division of Orthopaedic Surgery, Department of Surgery, Temerty Faculty of Medicine, University of Toronto, Toronto, ON

⁴Institute of Biomedical Engineering, Faculty of Applied Science and Engineering, University of Toronto, Toronto, ON

AUTHOR DISCLOSURES: None

ABSTRACT INTRODUCTION: Current deep learning models developed for orthopedic imaging use pre-trained models on ImageNet, a large dataset of general images from the Internet. While it is advantageous in improving performance and training times, in-domain model weight initialization on orthopaedic imaging may offer further improvements. Our aim is to develop a foundational self-supervised learning (SSL) model for orthopaedic radiographs, DINO-Ortho, and evaluate its added value to standard supervised approaches pretrained on ImageNet.

METHODS: The DINOv1 Resnet-50 convolutional neural network backbone was pretrained using SSL on the MURA and FracAtlas datasets, containing about 44,000 bone x-rays total. MURA is an upper extremity X-ray dataset, whereas FracAtlas contains X-rays from the hand, leg, hip, and shoulder. Both datasets are publicly available and contain both normal and pathologic X-ray images. After SSL, the pretrained image weights were trained on both datasets using supervised learning (SL). Training and validation accuracies were computed based on the model's ability to classify an X-ray as normal or abnormal. Both training steps used Python 3.13. DINO-Ortho's performance was compared to two other models: Original DINOv1 Resnet-50 and a standard Resnet-50 model pretrained on ImageNet using SL only. Metrics evaluated included sensitivity, specificity, area under the receiver-operating characteristic curve (ROC AUC), and area under the precision-recall curve (PRC AUC). Performance across the three experiments was compared using Friedman's test.

RESULTS SECTION: Of the three experiments, DINO-Ortho demonstrated the highest performance for ROC AUC and PRC AUC when tested on FracAtlas ($P < 0.05$). The original DINOv1 demonstrated the highest performance for ROC AUC and PRC AUC when tested on MURA ($P < 0.05$).

DISCUSSION: Adding the SSL step to DINO can potentially lead to higher performance metrics when tested on musculoskeletal (MSK) X-ray datasets. Future research will include testing DINO-Ortho on other specific MSK image datasets, to evaluate the model's performance for certain orthopaedic image classification tasks.

SIGNIFICANCE/CLINICAL RELEVANCE: DINO-Ortho can be used as a base model in all orthopaedic deep learning applications to improve model performance, training speed, and ultimately clinicians' ability to detect pathology in plain radiographs.

IMAGES and TABLES:

Table 1. Performance metric comparison between a standard Resnet-50 model pretrained without self-supervised learning, the original DINOv1 Resnet-50 model, and DINO-Ortho

	MURA				FracAtlas			
Experiment	ROC AUC	PRC AUC	Sens	Spec	ROC AUC	PRC AUC	Sens	Spec
SL	0.865	0.822	0.760	0.876	0.920	0.754	0.602	0.980
DINO + SL	0.894	0.897	0.776	0.874	0.929	0.844	0.657	0.978
DINO + DINO-Ortho + SL	0.875	0.878	0.748	0.864	0.986	0.957	0.880	0.996

Abbreviations: ROC AUC, area under receiver-operating characteristic curve; PRC AUC, area under precision-recall curve; Sens, sensitivity; Spec, specificity