GaitForeMer: Self-Supervised Pre-Training of Transformers via Human Motion Forecasting for Few-Shot Gait Impairment Severity Estimation

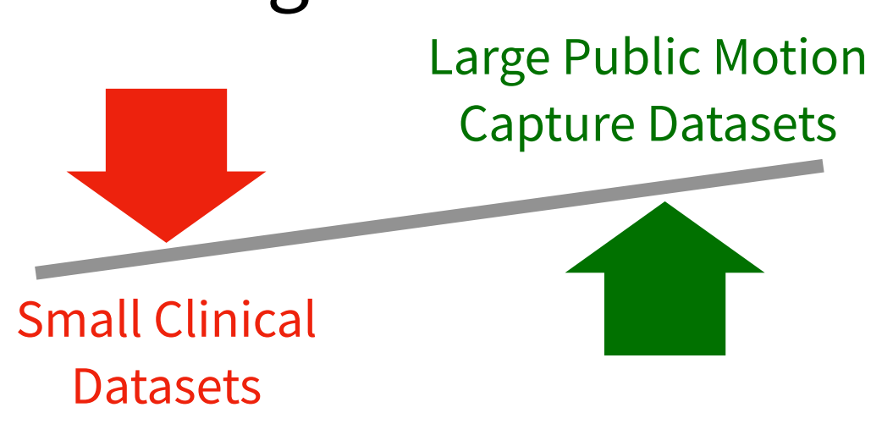
Mark Endo[1], Kathleen L. Poston[1], Edith V. Sullivan[1], Li Fei-Fei[1], Kilian M. Pohl[1,2], and Ehsan Adeli[1]

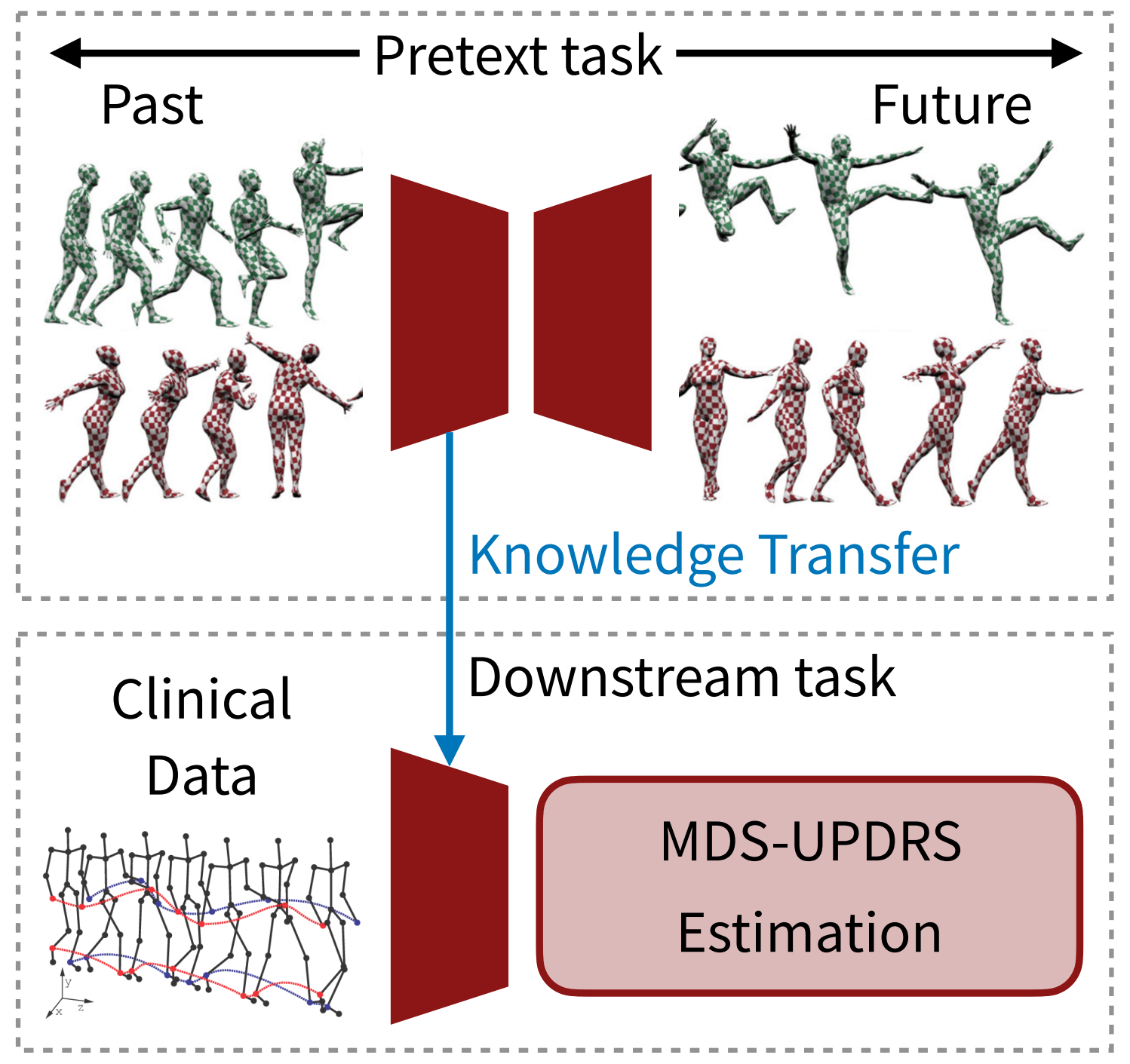
[1] Stanford University [2] SRI International

MICCAI2022 Singapore

Background

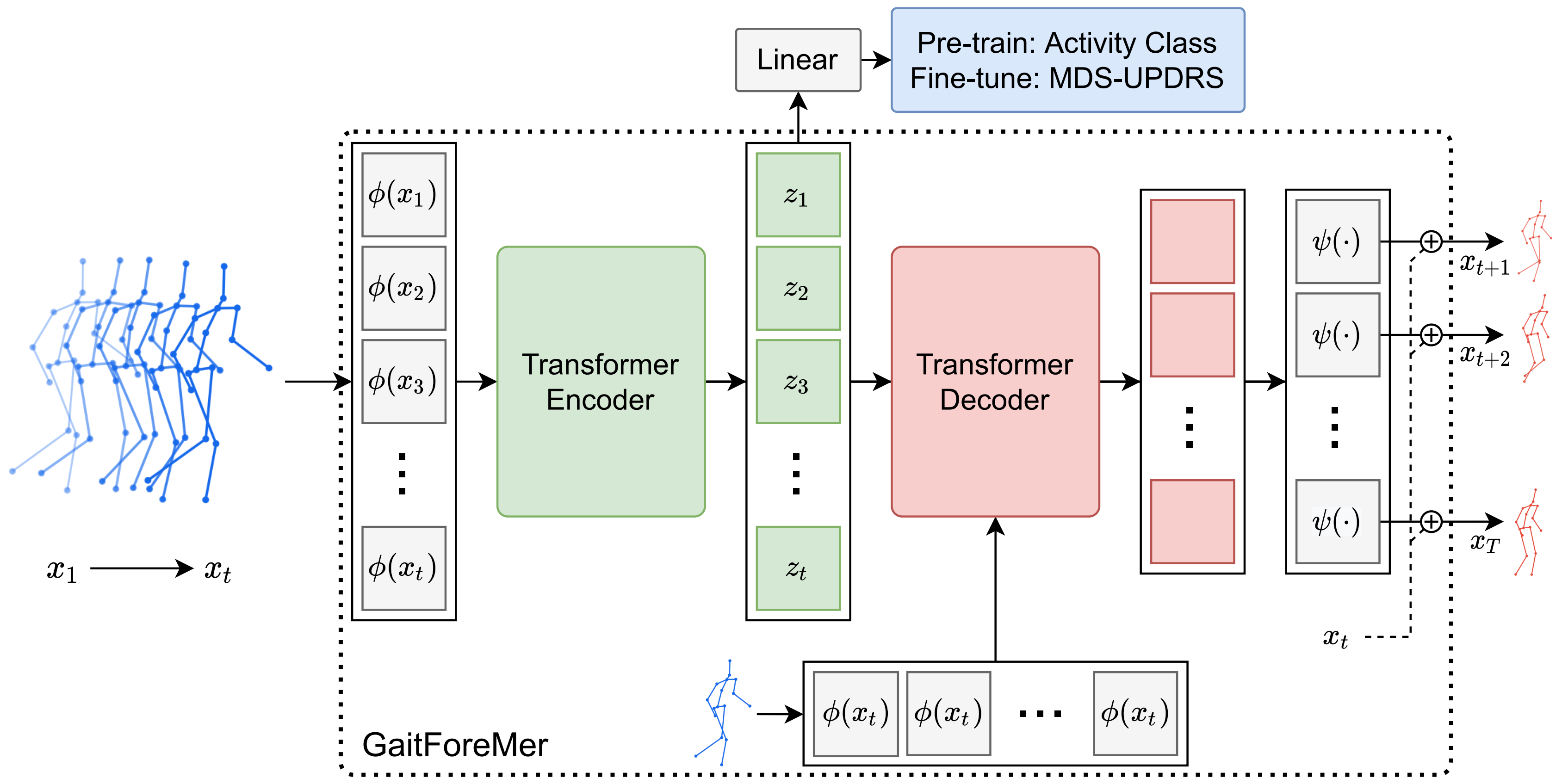
- Parkinson's disease is a chronic, progressive brain disorder with degenerative effects on mobility and muscle control
- **Task:** Prediction of motor impairment severity from videos of gait examinations of PD patients
- Clinical datasets are often limited in size; we can take advantage of large 3D motion capture datasets





- Recent advances in machine learning can allow us to take advantage of these datasets and translate them for clinical use
- **Goal:** learn good motion representations from large public dataset using the pretext task of motion forecasting and transfer knowledge for downstream task of gait impairment severity prediction

GaitForeMer

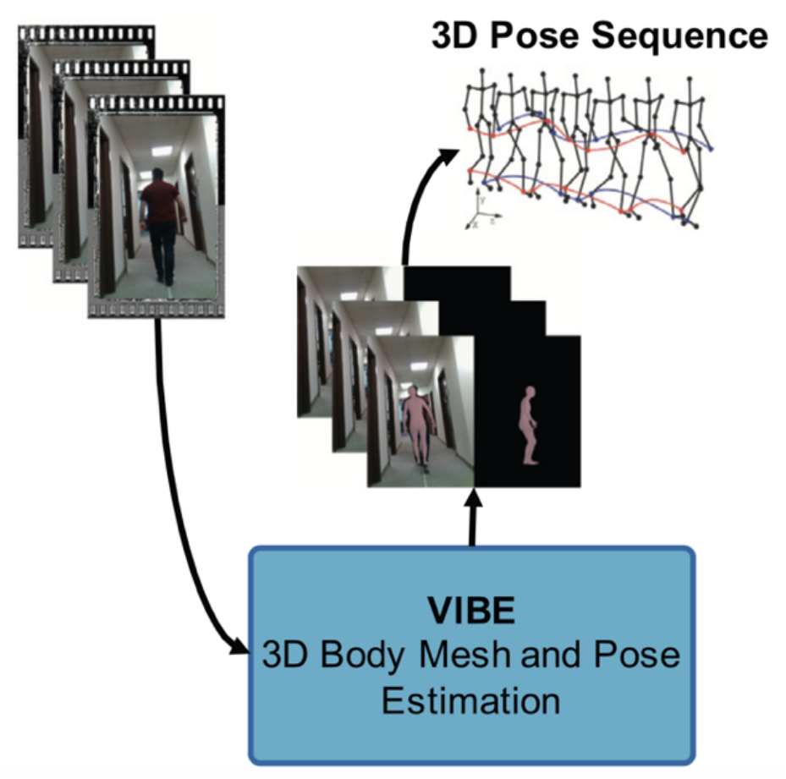


- We propose **GaitForeMer** (Gait Forecasting and impairment estimation transforMer) which forecasts motion and gait (pretext task) while estimating impairment severity (downstream task)
- Given a sequence of t 3D skeletons $\mathbf{x}_{1:t}$, we predict the next M skeletons $\mathbf{x}_{t+1:T}$ and the motion class y (either activity or MDS-UPDRS score)
- After pre-training the model components on a public dataset, we adapt the model to estimate MDS-UPDRS scores on our clinical data

Data

NTU RGB+D Dataset[1]: Large human motion capture dataset used to pre-train model

MDS-UPDRS Dataset: Gait recordings from 54 participants processed using Video Inference for Body Pose and Shape Estimation (VIBE)[2] to extract 3D skeletons



([1] Shahroudy et al., 2016; [2] Kocabas et al., 2020)

Results

- Results reported via leave-one-out cross-validation
- Compared methods:
  - GaitForeMer without pre-training (GaitForeMer-Scratch), Hybrid Ordinal Focal DDNet (OF-DDNet)[3], Spatial-Temporal Graph Convolutional Network (ST-GCN)[4], DeepRank[5], Support Vector Machine (SVM)[6]

Method	F_1	Pre	Rec
GaitForeMer (Ours)	**0.76**	**0.79**	**0.75**
GaitForeMer-Scratch (Ours)	0.60	0.64	0.58
OF-DDNet*	0.58	0.59	0.58
ST-GCN*	0.52	0.55	0.52
DeepRank*	0.56	0.53	0.58
SVM*	0.44	0.49	0.40

Our pre-trained GaitForeMer model results in best performance

* indicates statistical difference at ($p < 0.05$) compared with our method, measured by the Wilcoxon signed rank test

- Our GaitForeMer method pre-trained on a public dataset results in significantly improved accuracy over training the model from scratch and other baselines trained on the MDS-UPDRS dataset

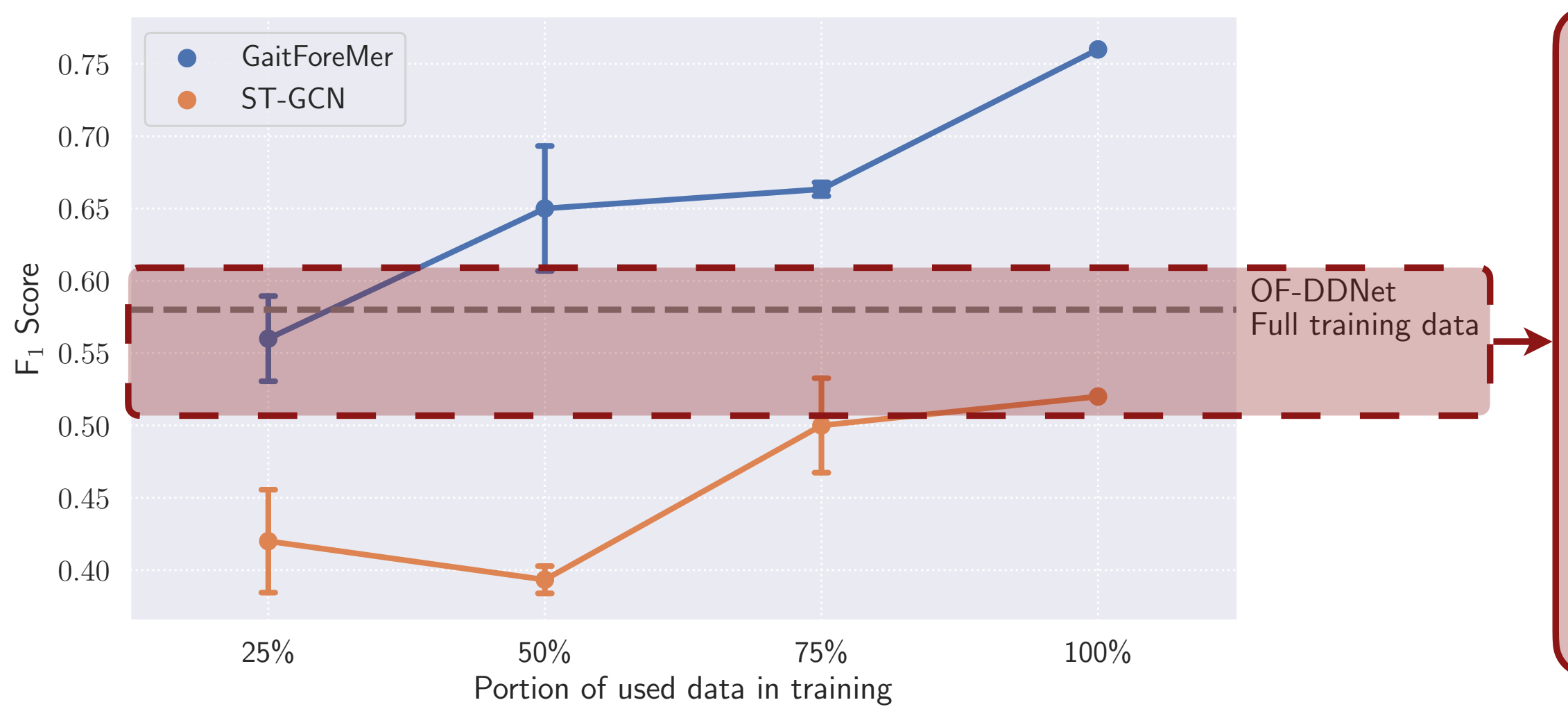
([3] Lu et al., 2021 ; [4] Yan et al., 2018; [5] Pang et al., 2017; [6] Weston et al., 1999)

Fine-tuning Setup

Pre-trained	Fine-tune strategy	F_1	Pre	Rec
Yes	Both branches then class branch	**0.76**	**0.79**	**0.75**
Yes	Both branches	0.72	0.75	0.71
Yes	Class branch	0.66	0.72	0.63
No		0.60	0.64	0.58

- We compare different training/fine-tuning strategies of our method
- First fine-tuning both branches then additionally fine-tuning the MDS-UPDRS prediction branch yields best results
- The relatively poor performance of only fine-tuning the class branch could be due to the data shift between the NTU RGB+D and MDS-UPDRS datasets that requires training of the motion forecasting branch

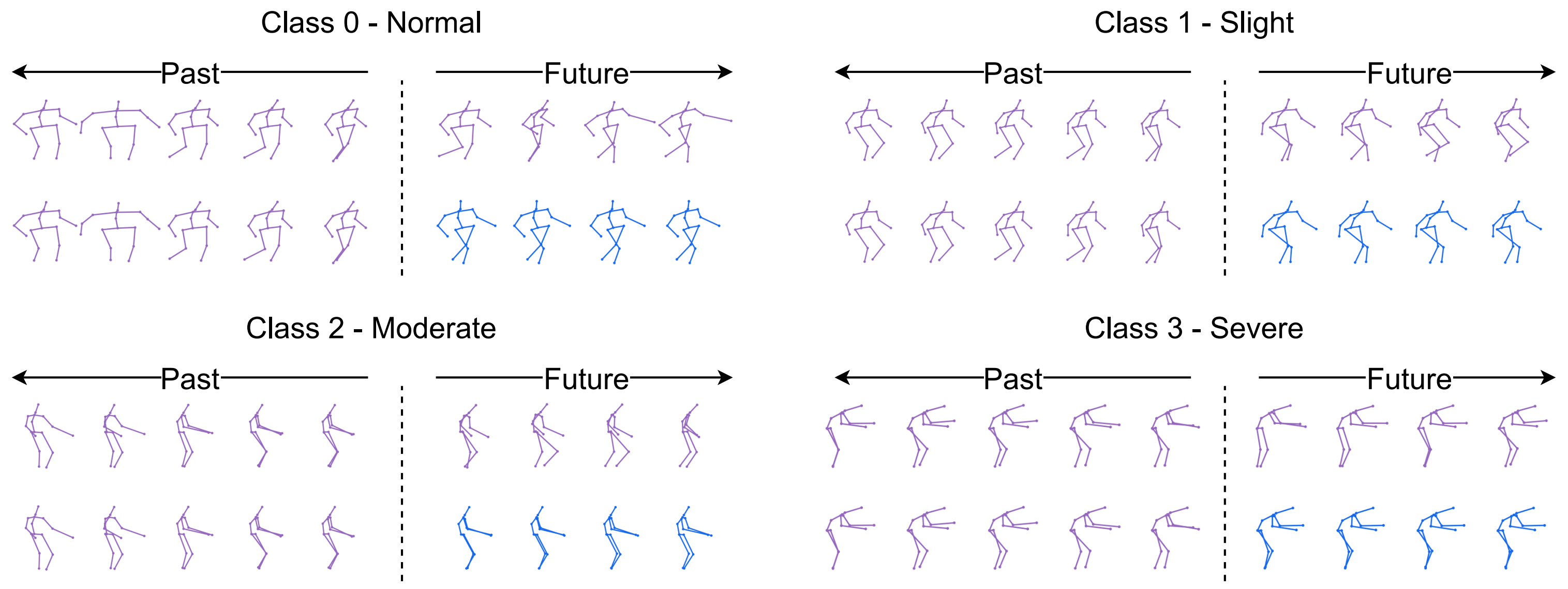
Few-shot Learning



With 25% training data, GaitForeMer outperforms ST-GCN using 100% training data and is comparable to OF-DDNet using 100% training data

- We find that our GaitForeMer method maintains relatively strong performance with only a fraction of the data
- This shows the power of using motion forecasting as a self-supervised pre-training task for few-shot gait impairment severity estimation

Motion Forecasting Visualization



The purple skeletons are ground-truth and the blue ones are predictions

- Accurate motion forecasting verifies that the model is able to properly predict motion that encodes motor impairments

Conclusion

- Human motion forecasting serves as an effective pre-training task
- Pre-trained model significantly outperformed models trained from scratch
- Approach demonstrates utility of using motion pre-training tasks in data-limited settings

Acknowledgments

This work was supported in part by NIH grants (AA010723, NS115114, P30AG066515), the Michael J Fox Foundation for Parkinson’s Research, UST (a Stanford AI Lab alliance member), and the Stanford Institute for Human-Centered Artificial Intelligence (HAI) Google Cloud credits.