Evaluating pain disparities in underserved populations

Koushik Sridhar

Department of Computer Science, University of North Carolina at Chapel Hill
sridhark@cs.unc.edu

Abstract. Differences in racial and socioeconomic status result in strong variations in reported pain. Pain disparities persist even after accounting for objective severity measures of osteoarthritis based on medical image analysis, suggesting that other factors — such as differences in pain perception, reporting behavior, or unmeasured population characteristics — may contribute to these disparities. We implement a machine learning approach to predict osteoarthritis severity by using knee x-rays to predict patients' experienced pain. We utilize these predictions as a measure of severity and find that this approach results in reductions in unexplained disparities in pain on racial and socioeconomic lines. We demonstrate the effectiveness of this method in experiments conducted on data from the Osteoarthritis Initiative (OAI) and the Johnston County Osteoarthritis Project (JOCO). The results of this study indicate that an algorithmic severity measure may better capture pain in underserved populations and show potential to reduce disparity when informing treatment decisions.

1 Introduction

In recent years, machine learning methodologies have afforded unique opportunities to improve decision-making across disciplines, in part due to the accessibility of expansive datasets. This increase in data availability extends to medical data analysis, where high-quality machine learning models rely on appropriate, accurate, and granular clinical information. However, this comes with a unique set of opportunities and challenges. Medical datasets often contain a wide range of metrics—including diagnostic imaging, patient-reported outcomes, lab results, and physician assessments—which together enable rich observational analyses. Yet, some of these metrics may reflect or exacerbate existing disparities across socioeconomic groups, ultimately influencing how decisions are made in clinical settings.

This issue becomes particularly salient when attempting to understand and model pain. For instance, in the context of knee osteoarthritis—a condition that affects approximately 10% of men and 13% of women over the age of 60 in the United States [12]—subjective pain reports are a commonly used metric. However, these reports can vary systematically across populations: people of color, for example, often report significantly higher levels of knee pain than white individuals with similar radiographic severity [1]. This discrepancy suggests that patient-reported pain, while clinically relevant, may be influenced by factors beyond biological pathology, such as socioeconomic status, access to care, or cultural norms around expressing pain. Recognizing and addressing these disparities is essential, as such pain assessments directly inform treatment decisions, including eligibility for interventions like physical therapy, medication, or surgery.

In this work, we develop and evaluate a machine learning approach to predict pain in a way that reduces disparities across racial and socioeconomic groups. Our goal is to create a pain prediction model that better captures patient-reported pain compared to traditional radiographic severity measures, which assess osteoarthritis severity strictly based on physical features visible in x-ray images and do not account for external factors such as socioeconomic status or lived experience. We use radiographic severity as a comparison benchmark to evaluate whether our model can provide more equitable pain assessments. To assess this, we measure the extent to which our model-based pain predictions reduce disparities (e.g., by race or socioeconomic class) in comparison to disparities observed using radiographic severity or direct comparisons of raw pain score distributions. Here, we define pain disparity as the portion of reported pain that is not explained by our severity measures (model predictions or KLG scores), but instead correlates with racial or socioeconomic factors—implying that these factors influence how pain is experienced or reported beyond what structural severity alone can capture. This disparity metric serves as a way to evaluate a model's ability to equitably assess pain across racial and socioeconomic groups: lower disparity values indicate that the model's predictions are less influenced by these demographic factors and better aligned with reported pain across populations. Importantly, this measure does not assume inherent differences in pain levels between groups, but rather evaluates whether the model reduces the degree to which race or socioeconomic status is predictive of unexplained variation in pain. We conduct this analysis using data from two osteoarthritis datasets — the Osteoarthritis Initiative (OAI) [6] and the Johnston County Osteoarthritis Project (JOCO) [3] — and evaluate performance using validated patient-reported pain scores.

2 Related Works

Efforts to objectively understand pain in the context of knee osteoarthritis have traditionally focused on radiographic analysis, particularly through scoring systems based on x-ray images. A widely used measure in this domain is the Kellgren–Lawrence grade (KLG), which classifies the structural severity of osteoarthritis into five categories ranging from 0 (no signs) to 4 (severe joint damage) [5]. Importantly, KLG does not directly assess pain but rather evaluates physical joint degradation observable via imaging.

Prior studies have highlighted disparities in KLG scores across racial groups, with Black patients tending to have higher KLG grades than white patients, even when controlling for other health factors [11]. Yet, a significant and welldocumented challenge in osteoarthritis research is the disconnect between radiographic severity and experienced pain: some individuals with low KLG scores report high levels of pain, while others with high structural severity report minimal discomfort. This has led to ongoing debate about the adequacy of radiographic measures in fully capturing the lived experience of osteoarthritis-related pain. Moreover, KLG was developed based on predominantly white British populations and may therefore overlook structural patterns or disease presentations more prevalent in other racial or ethnic groups [2].

Recent work has proposed machine learning-based severity scores as an alternative to radiographic grading, with some approaches demonstrating improvements in equity. For example, models trained to predict patient-reported pain — such as Knee Injury and Osteoarthritis Outcome Score (KOOS) values — directly from knee x-rays using a ResNet-18 architecture have shown promise. In these models, disparities observed in treatment-relevant severity assessments are reduced when compared to traditional KLG-based evaluations. Crucially, this does not imply that the model eliminates disparities present in the underlying pain reports. Rather, it suggests that traditional measures like KLG may underestimate disease burden in underserved populations, and that machine learning-based approaches may better align severity assessment with patients' self-reported experiences. Building on this line of work, we expand evaluation to include Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain scores from both the OAI and JOCO datasets. While KOOS and WOMAC pain assessments both employ self-administered Likert scales (0–4 per item) and share common questions, they differ in interpretation: KOOS pain scores are transformed to a 0–100 scale (100 = no pain), whereas WOMAC pain scores typically use raw sums (0–20) or transformed scales where higher values indicate worse pain [8].

3 Methods

We replicate the methodology and analysis found in [7]. In our analysis, we utilize data from OAI and JOCO datasets as part of our analysis, performing some data preprocessing on each dataset to ensure consistent analysis. We train a ResNet-18 model on each dataset with x-ray image input with the goal of predicting KOOS or WOMAC pain score based on the analysis case and evaluate its effectiveness via Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Pearson R^2. We fit a linear regression with true pain score as the dependent variable, with binary race or socioeconomic group and a measure of severity (KLG or predicted pain). We define pain disparity as the coefficient on binary race or socioeconomic group, that being the gap in mean pain between racial or socioeconomic groups while controlling for severity. This methodology was followed for OAI Data with KOOS pain scoring, OAI Data with WOMAC pain scoring, and JOCO Data with WOMAC scoring. The methodology is illustrated in Figure 1

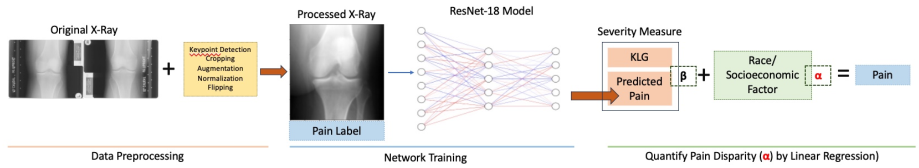

Fig. 1. Methods were broken down into Data Preprocessing, Network Training, and Pain Disparity Quantification steps. Data preprocessing standardized images prior to network training, and removed any incomplete samples. Network training utilized x-ray images and associated pain labels to fine-tune a ResNet-18 model to predict pain scores. The resulting predicted pain score was used as a severity measure in a linear regression where severity measure and binary race/socioeconomic factor were regressed against pain score. The resulting coefficient on the race/socioeconomic factor term (α) quantified the pain disparity. We compare the pain disparities when using predicted pain and KLG as severity measures to determine effectiveness in evaluating bias.

3.1 Datasets

Both OAI and JOCO datasets are structured similarly, featuring longitudinal data from participants aged 45 and older who had or were at high risk of developing knee osteoarthritis. All study data was anonymized prior to usage. Each dataset was handled in the same fashion as follows. Summary statistics on the dataset can be found in Table 1.

For the OAI dataset, pain data were analyzed at five time points: baseline and follow-ups at 12, 24, 36, and 48 months, using both KOOS and WOMAC pain scores. For the JOCO dataset, pain was assessed at four time points (T1, T2, T3, and T4), with approximately 6.5 years between each follow-up, using WOMAC pain scores only. Each observation in the dataset corresponds to one knee for one person at one time point. Observations were removed if they were missing pain scores, KLG, age, race, sex, socioeconomic status, or did not have an applicable knee x-ray. We randomly divided the data at the patient level into training, validation, and testing sets, where 70% of patients were in the training set, 10% were in the validation set, and 20% of patients were in the testing set. The training set was used to optimize model weights, the validation set was used to conduct a hyperparameter search and save the best model, and the test set was used as a final evaluation after all training and validation were complete.

3.2 Data Preprocessing

Our dataset consists of knee X-rays that needed to be standardized prior to model input. The X-rays encompass substantial areas of the femur and tibia. To extract the joint region of interest, we employ the method utilized in [10]. Through this, a region of 140mm $\times$ 140mm is extracted. All images are resized to 128 $\times$ 128, with augmentations including random rotation (up to 15 degrees), Gaussian noise, and contrast adjustment. Gaussian noise with a standard deviation of $\sigma = 0.01$ is added with 50% probability to simulate acquisition variability.

Table 1. Dataset summary statistics for OAI and JOCO datasets. OAI Dataset features more individuals and observations. Key differences are found in the percentage of black patients, non-college graduates, and the fraction of knees with KLG ≥ 2. Economic data were not available for JOCO patients, and thus, income disparities will not be measured when evaluating with WOMAC pain scores.

	OAI		**JOCO**	
	Train+Val	**Test**	**Train+Val**	**Test**
Sample				
# Individuals	2,877	1,295	1,942	482
# Observations	25,049	11,320	8,098	1,957
Demographics				
Black	17%	16%	33%	30%
Lower-income (<$50,000)	38%	39%	–	–
Non-College Graduates	39%	38%	90%	90%
Female	58%	56%	65%	64%
Fraction of Knees w/ KLG $\geq$ 2				
All	45%	46%	32%	31%
Black	60%	56%	37%	39%
Non-College Graduates	52%	49%	33%	31%

To mitigate potential noise introduced during image acquisition, we normalize the radiographs by linearly scaling the intensities such that the smallest 99% of values map to [0, 0.99]. We also apply horizontal flipping to all right knees to maintain anatomical consistency. Random contrast adjustments are applied with 50% probability by scaling intensities around the mean by a factor randomly sampled from [0.75, 1.25].

3.3 Network Training

A convolutional neural network was trained to predict pain scores for each knee, using a single knee x-ray as input. This means that one patient will have two observations at a time point, one for each knee. The model attempted to minimize a loss function defined as:

$$Loss = \sum_{i=1}^{D} (Y_{true} - Y_{predicted})^2 \tag{1}$$

where Y is a pain score for an observation. We fine-tune a ResNet-18 model for this task with network weights pretrained on ImageNet. Deeper layers of the network were fine-tuned on the chosen dataset, with the best model having the final 12 convolutional layers specifically unfrozen similar to our reference work, enabling the model to relearn task-specific features relevant to the dataset. Adam [4] was chosen as the optimizer, and the one cycle scheduler [9] with a max learning rate of 1e-5 was used for pain prediction. Models were trained for 30 epochs with a batch size of 8. The network parameters resulting in the best RMSE score on the validation set were selected for testing.

3.4 Quantifying Pain Disparities

We seek to quantify disparities with respect to race, education, and income (only for the OAI dataset). To do this, we turn race, education, and income into binary values so that we compare as follows:

1. **Race:** Patients who are black versus not black
2. **Education:** Patients who had a college education versus those who did not have a college education
3. **Income:** Patients earning $\geq$ \$50,000 versus $<$ \$50,000

Differences in pain scores across groups were first quantified without controlling for osteoarthritis severity, using the mean KOOS pain score between groups. For example, racial pain disparity was defined as the difference in mean pain between Black and non-Black patients. We then computed racial and socioeconomic pain disparities while controlling for a measure of osteoarthritis severity (either KLG or predicted pain). To do this, we fit a linear regression with KOOS pain score as the dependent variable and two independent variables: binary race or socioeconomic group and a measure of osteoarthritis severity. We operationalize pain disparity as the coefficient associated with a binary group indicator, representing either race or socioeconomic status, in a linear model of reported pain. Specifically, we model pain as:

$$Pain = \alpha G + \beta S + C \tag{2}$$

where G denotes the binary group variable (e.g., race or socioeconomic status), S represents the measure of osteoarthritis severity (KLG or predicted pain score), and α, β are the corresponding regression coefficients. C represents the baseline level of pain for the reference group at the reference severity level, serving as the model's intercept. In this formulation, a smaller absolute value of α indicates reduced disparity in pain between groups, implying that the severity measure S more fully accounts for the differences in reported pain. To ensure a fair comparison between KLG and predicted pain severity in terms of explanatory power, KLG was coded as a categorical variable rather than a continuous one. This approach avoids assuming a linear relationship between KLG levels and pain severity, allowing for more accurate estimation of each KLG category's distinct contribution when regressing and computing pain disparity coefficients.

4 Experimental Results

4.1 Reducing pain disparities

We compute pain disparity reductions on OAI Data with KOOS pain scoring, OAI Data with WOMAC pain scoring, and JOCO Data with WOMAC pain scoring. We sought to replicate results found from our reference work [7], which used

OAI Data with KOOS pain scoring. The main difference between the methodology employed in the previous work and this work is that the model employed in the original work (ensemble of five ResNet-18 models trained to also predict other radiographic features in addition to patient pain scores) is different to the model used in this work (finetuned ResNet-18 model predicting patient pain scores). We transfer the methodology to work with WOMAC data in the OAI and JOCO datasets. In our results, the pain disparity is the coefficient on the race/socioeconomic term of the linear regression, where lower coefficients indicate lower disparity between binary groups.

OAI Data with KOOS Pain Scoring. Table 2 shows the pain disparity reductions when using OAI Data with KOOS Pain scoring. Across all groups of comparison, we find that our model reduces pain disparities better than KLG severity measures. This trend in behavior falls in line with the reference work, however, the percent reduction does not reach the level of the previous work. One possible explanation is that ensembling and the features learned during training to predict other radiographic characteristics may contribute to improved performance. However, confirming this hypothesis will require further ablation studies to investigate the contributions of individual layers or sub-models to pain prediction.

Table 2. Reported Pain Disparities in OAI Data with KOOS Pain Scoring across race, education, and income groups. The columns None, KLG, and Model report racial and socioeconomic pain disparities (in KOOS points measured from 0-100) without any controls for severity (equivalent to the difference in mean pain scores between groups), when controlling for the KLG severity measure, and when controlling for the model-predicted pain severity measure, respectively. The KLG Reduction and Mod. Reduction columns show the percent difference between the None column with the KLG and model-predicted pain severity measures, respectively. The Lit. Reduction column shows the reduction obtained from the methods of the original work. We observe that using model predictions as a severity measure outperformed KLG when it comes to reducing pain disparities in the OAI dataset, however not to the same level found in the reference work.

Group	None	KLG	Model	KLG Reduction	Mod. Reduction	Lit. Reduction
Race	7.60	6.52	5.20	13%	32%	43%
Education	7.52	6.82	6.51	9%	14%	30%
Income	5.43	4.71	4.31	13%	20%	32%

OAI Data with WOMAC Pain Scoring. Table 3 shows the pain disparity reductions when using OAI Data with WOMAC Pain scoring. Across all groups of comparison, we find that our model continues to reduce pain disparities better than KLG severity measures.

Table 3. Reported Pain Disparities in OAI Data with WOMAC Pain Scoring across race, education, and income groups. The table is structured in the same way as Table 2 with pain values within the WOMAC scale (0-20). We see that the model severity measure reduces pain disparities better than the KLG severity measure.

Group	None	KLG	Model	KLG Reduction	Model Reduction
Race	1.48	1.24	0.96	16%	35%
Education	1.34	1.18	1.09	12%	20%

JOCO Data with WOMAC Pain Scoring. Table 4 shows the pain disparity reductions when using JOCO Data with WOMAC Pain scoring. Across all groups of comparison, we find that our model performs worse than KLG Reduction, however returns a percentage of reduction similar to performance in OAI with KOOS and OAI with WOMAC Pain Scoring tests (25% - 35%), establishing its consistency. A potential explanation for the higher KLG reduction in pain disparity observed when using the JOCO dataset, compared to the OAI dataset, may relate to the different distributions of KLG scores across racial groups, as shown in Figure 2. In the OAI dataset, KLG = 0 is the most prevalent class, whereas in the JOCO dataset, KLG = 0 and 1 appear in more comparable proportions. Importantly, the JOCO dataset shows a notably higher proportion of Black or African American patients in the KLG = 1 category compared to OAI. Since our analysis regresses the true WOMAC pain score as the dependent variable against our race/socioeconomic factor and KLG score, this distributional difference may influence the extent to which disparity in pain perception across racial groups can be captured and subsequently reduced. While Figure 2 highlights these distributional differences, it does not by itself establish causality. To support this hypothesis, one potential direction for validation would be to simulate or reweight the OAI dataset to match the joint distribution of race and KLG scores in the JOCO dataset, and then assess whether similar reductions in pain disparity emerge. Such an analysis could help confirm whether the observed effect is indeed driven by differences in the underlying data structure.

Table 4. Reported Pain Disparities in JOCO Data with WOMAC Pain Scoring across race, education, and income groups. The table is structured in the same way as Table 2 with pain values within the WOMAC scale (0-20). We see that the model severity measure reduces pain disparities in the 25% range but fails to perform better than the KLG severity measure.

Group	None	KLG	Model	KLG Reduction	Model Reduction
Race	0.83	0.52	0.61	37%	26%
Education	1.36	0.85	1.01	37%	27%

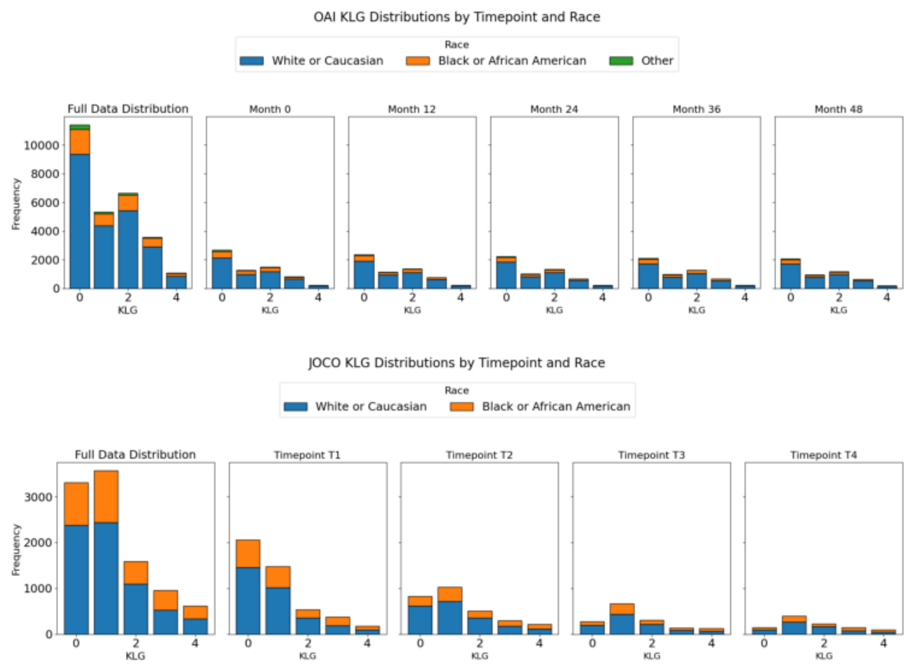

Fig. 2. KLG distributions across different timepoints for the OAI and JOCO datasets. The top panel shows the KLG distributions for the OAI dataset, including the full dataset and individual distributions at Months 0, 12, 24, 36, and 48. The bottom panel displays the KLG distributions for the JOCO dataset, including the full dataset and individual distributions at Timepoints T1 through T4. Both plots also display racial distributions for each KLG score found in the data. Across both datasets and all timepoints, the majority of subjects have lower KLG scores (0 or 1), with higher scores (3 or 4) being less frequent. Additionally, White or Caucasian participants make up the largest proportion in each KLG category, while Black or African American participants are less represented. Notably, the proportion of Black or African American participants is visibly higher in the JOCO dataset compared to the OAI dataset, especially at lower KLG scores.

Table 5. Performance Metrics for KOOS and WOMAC Predictions across OAI and JOCO datasets. Metrics from the reference work are also listed as Lit. Value. We note that the metrics from the reference work and OAI KOOS are similar to each other, and that the metrics in OAI WOMAC and JOCO WOMAC have a difference in about 1 WOMAC unit for RMSE and MAE, but have similar Pearson R^2 values.

Metric	Lit. Value	OAI KOOS	OAI WOMAC	JOCO WOMAC
RMSE	14.9	16.40	2.94	4.12
MAE	11.3	13.01	2.18	3.32
Pearson $\mathbf{R}^2$	0.16	0.20	0.30	0.30

4.2 Model Metrics

Table 5 summarizes the RMSE, MAE, and Pearson R^2 values of the models across the evaluated datasets. The metrics for the OAI KOOS test were consistent with prior work, so we adopted the same methodology and model setup for predicting WOMAC pain scores, retraining on those respective datasets. Notably, we observe differences in performance between the OAI WOMAC and JOCO WOMAC datasets, with approximately a 1-point difference in both RMSE and MAE metrics. One possible explanation for this performance gap is that the JOCO pain score labels[1] may exhibit greater variability or noise, which could challenge the model's ability to fit or generalize accurately. While we have not directly quantified label noise in this work, this remains a plausible hypothesis that could be tested in future analyses—for instance, by examining inter-rater reliability, variance in repeated measures, or error consistency across demographic strata.

We observe in Figure 3 that predictions across all datasets and pain score types (WOMAC and KOOS) tend to occupy a narrower range than the true label distributions. Specifically, predicted values for WOMAC pain scores often fall within the 0–10 range, despite the full scale extending to 20, and KOOS predictions tend to cluster between 50–100 rather than covering the full 0–100 spectrum. This compression of predicted values suggests that the model may be underfitting the full variability of the pain score distributions. Notably, the observed 25%–35% reduction in disparities when comparing average pain scores across race or socioeconomic groups occurs despite this underfitting—not because of it. In fact, we expect that more accurate predictions, which would better reflect the full range of true pain scores, would lead to even lower pain disparity values, since the model would be better capturing individual-level variation and less reliant on demographic residuals. Thus, while the disparity reduction is promising, it is likely a conservative estimate of the model's potential to improve equity. Further investigation is needed to determine whether the com-

[1] Timepoint T1 in Johnston County reported pain score as a singular value applied to both knees, Timepoints T2-T4 conducted their evaluations so that each knee could have an individual pain score. This may have led to the noisiness of data as a patient's knees are not guaranteed to have the same pain.

pressed output range merely dampens variability that contributes to disparities, or if it meaningfully addresses the mismatch between radiographic severity and reported pain.

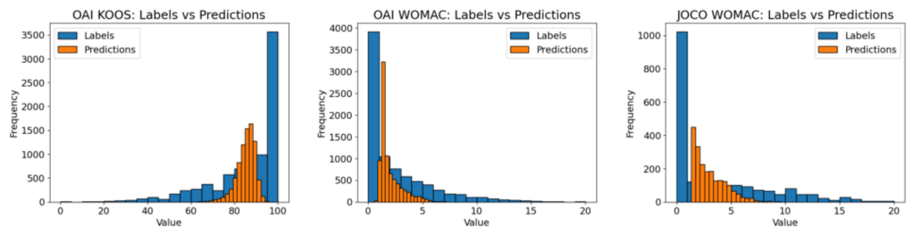

Fig. 3. Distributions of Labels and Predictions across OAI KOOS, OAI WOMAC, and JOCO WOMAC data. The distributions of labels for JOCO WOMAC and OAI WOMAC are heavily skewed towards lower values, while for OAI KOOS, the labels are skewed towards higher values. Predictions on all three datasets face domain-shrinkage toward the skewed half of the domain.

4.3 Potential Shortcoming of Approach

In evaluating the methodology used in this work, we identify a potential shortcoming in how pain disparities are quantified, especially in relation to prediction performance. In our analysis, we define pain as a linear function of a severity measure (either KLG score or model-predicted pain) and a binary indicator for race or socioeconomic group. Under this formulation, disparity is captured by the magnitude of the coefficient for the group variable: a coefficient of zero implies no disparity, meaning that the severity measure alone is sufficient to explain the pain score, irrespective of group membership.

Now consider a theoretical case where the model used for prediction achieves perfect accuracy—i.e., predicted pain exactly matches the observed pain labels for every individual. In this scenario, the severity measure (predicted pain) would be identical to the true pain, and the group variable would no longer contribute explanatory power. This would result in a group coefficient of zero, implying no disparity.

However, in practice, our models do not achieve perfect predictions. More importantly, as shown in Figure 3, model predictions exhibit domain shrinkage, where predictions cover only a subset of the full label range (e.g., 0–10 instead of 0–20 for WOMAC). Despite this limitation, we observe a substantial reduction (25%–35%) in the disparity coefficient when using predicted pain as the severity measure, compared to using KLG scores.

This leads to an important concern: it becomes difficult to determine whether the observed reduction in disparity reflects a meaningful improvement in fairness or whether it is merely an artifact of limited prediction variance. A model that fails to predict extreme values, particularly if those values are more common in one group than another, may inadvertently compress group-level differences, artificially lowering the disparity coefficient. In such cases, the appearance of fairness may be misleading.

In short, while minimizing the group coefficient is the objective, care must be taken in interpreting disparity reductions that arise alongside limited prediction range or accuracy. Further analysis is needed to determine whether these reductions represent real improvements or simply reflect bias suppression.

5 Conclusion

In this work, we utilize a method to quantify and reduce racial disparities via machine learning on more datasets and pain scoring systems. Through analysis on OAI and JOCO datasets, results indicate that our approach is consistent across datasets, resulting in disparity reduction around 25%-35% across socioeconomic groups. Future directions would include investigating how adding further views (MRI, tabular data, thickness maps) may impact pain disparity reduction across racial and socioeconomic groups. While this paper focuses race, education, and income, further research can focus on additional socioeconomic categories, including sex, occupation, and access to healthcare.

6 Acknowledgements

The OAI knee data were obtained from the controlled-access datasets distributed from the Osteoarthritis Initiative, a data repository housed within the NIMH Data Archive. The JOCO knee data was obtained from The Johnston County Osteoarthritis Project (JOCO) courtesy of the Nelson Lab. The author would like to thank Dr. Marc Niethammer, Dr. Jorge Silva, and Boqi Chen for their advice and support throughout this project.

References

1. Bolen, J., Schieb, L., Hootman, J.M., Helmick, C.G., Theis, K., Murphy, L.B., Langmaid, G.: Differences in the prevalence and impact of arthritis among racial/ethnic groups in the united states, national health interview survey, 2002, 2003, and 2006. Preventing chronic disease **7**(3), A64 (2010)
2. Hoffman, K.M., Trawalter, S., Axt, J.R., Oliver, M.N.: Racial bias in pain assessment and treatment recommendations, and false beliefs about biological differences between blacks and whites. Proceedings of the National Academy of Sciences **113**(16), 4296–4301 (2016)
3. Jordan, J.M., Helmick, C.G., Renner, J.B., Luta, G., Dragomir, A.D., Woodard, J., Fang, F., Schwartz, T.A., Nelson, A.E., Abbate, L.M., et al.: Prevalence of hip symptoms and radiographic and symptomatic hip osteoarthritis in african americans and caucasians: the johnston county osteoarthritis project. The Journal of rheumatology **36**(4), 809–815 (2009)

4. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
5. Kohn, M.D., Sassoon, A.A., Fernando, N.D.: Classifications in brief: Kellgren-lawrence classification of osteoarthritis. Clinical Orthopaedics and Related Research® **474**, 1886–1893 (2016)
6. Nevitt, M., Felson, D., Lester, G.: The osteoarthritis initiative. Protocol for the cohort study **1**, 2 (2006)
7. Pierson, E., Cutler, D.M., Leskovec, J., Mullainathan, S., Obermeyer, Z.: An algorithmic approach to reducing unexplained pain disparities in underserved populations. Nature Medicine **27**(1), 136–140 (2021)
8. Roos, E.M., Toksvig-Larsen, S.: Knee injury and osteoarthritis outcome score (koos)–validation and comparison to the womac in total knee replacement. Health and quality of life outcomes **1**, 1–10 (2003)
9. Smith, L.N., Topin, N.: Super-convergence: Very fast training of neural networks using large learning rates. In: Artificial intelligence and machine learning for multi-domain operations applications. vol. 11006, pp. 369–386. SPIE (2019)
10. Tiulpin, A., Melekhov, I., Saarakkala, S.: Kneel: Knee anatomical landmark localization using hourglass networks. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. pp. 0–0 (2019)
11. Vina, E., Ran, D., Ashbeck, E., Kwoh, C.: Natural history of pain and disability among african–americans and whites with or at risk for knee osteoarthritis: A longitudinal study. Osteoarthritis and cartilage **26**(4), 471–479 (2018)
12. Zhang, Y.: Jordan jm. Epidemiology of Osteoarthritis Clin. Geriatr Med Aug **26**(3), 355–369 (2010)