The Role of Transfer Learning on the Performance of Machine Learning Models Across Racial/Ethnic Groups

Chi Chun Steve Tsang, PhD[1]; Yan Cui, PhD[2]; William C. Cushman, MD[2]; Carmen C. Coleman, EdD[1]; Junling Wang, PhD[1]
Presenting author: Chi Chun Steve Tsang, PhD, Research Associate, Department of Clinical Pharmacy and Translational Science. E-mail: ctsang@uthsc.edu
1. University of Tennessee College of Pharmacy 2. University of Tennessee College of Medicine

BACKGROUND

- In the Enhanced Medication Therapy Management (MTM) demonstration initiated by the Centers for Medicare & Medicaid Services, Part D plans used predictive modeling with machine learning to identify MTM-eligible individuals.[1]
- Machine learning shows stronger predictive performance for racial/ethnic majority populations,[2] raising concerns about inequitable access.[3,4]
- Transfer learning may improve model performance for racial/ethnic minority populations by adapting information from larger to smaller subgroups.[5]

OBJECTIVES

To assess whether transfer learning improves model performance for identifying high-cost Medicare beneficiaries in racial/ethnic minority groups.

METHODS

- **Study design:** A retrospective, cross-sectional analysis of 2019 Medicare administrative claims linked to the Area Health Resources Files (AHRF).
- **Study sample:** A 10% sample of traditional fee-for-service Medicare beneficiaries with continuous Parts A, B, and D enrollment in 2019 and alive at the end of the year.
- **Conceptual framework:** Gelberg-Anderson behavioral model for vulnerable populations.[6]
- **Outcome:** Binary indicators for top-quartile inclusion in healthcare and medication costs as proxies for potential MTM eligibility.
  - Top-quartile thresholds reflected the expected 25% MTM enrollment benchmark used by the Centers for Medicare & Medicaid Services.[7]
  - Healthcare costs were derived from Parts A & B claims and medication costs from Part D claims, both measured from a health sector perspective.
- **Data Analysis:**
  - Assessment of racial/ethnic differences in predictive performance.
    - A multi-layer perceptron neural network (MLPNN) with three hidden layers was developed and trained separately for each cost outcome.
    - A stratified training (75%) and test (25%) split was conducted within each racial/ethnic group before combining data for the pooled model training.
    - Model performance was evaluated using 3 metrics: the area under the receiver operating characteristic curve (AUC-ROC); the area under the precision-recall curve (AUC-PR); and Brier scores.
    - Pairwise metric differences between White and minority groups were computed, with 95% confidence intervals (CIs) estimated using 1,000 nonparametric bootstrap replicates.
  - Assessment of transfer learning improvements for minority groups.
    - An identical MLPNN was trained on data from White beneficiaries (source) and adapted to each minority group (target) using a fine-tuning algorithm.
    - Prediction metrics from transfer learning and pooled models were compared for each minority group, with a similar bootstrapping procedure used to estimate CIs and assess the significance of performance differences.

RESULTS

- The analytic cohort included 1,848,654 beneficiaries, with 1,386,489 assigned to the training set and 462,165 to the test set. Racial and ethnic differences in patient characteristics were similar across datasets (Table 1).
- Pooled prediction models showed significant racial/ethnic differences in model performance for both outcomes (Table 2). For healthcare costs, minority groups demonstrated better predictive performance than White beneficiaries, including higher AUC metrics and lower Brier scores, with the largest gaps observed between Black/Hispanic and White groups. Differences for medication costs were more modest, with minority groups generally having higher AUC-PR values and Brier scores than White beneficiaries.
- Transfer learning generally led to performance gains for the minority groups (Table 3). Black, Hispanic, and Asian groups experienced small but significant increases in AUC-ROC (range: 0.0024 to 0.0069) and AUC-PR (range: 0.0087 to 0.0164), along with decreases in Brier scores (range: -0.0015 to -0.0031).

TABLES/FIGURES

Table 1. Characteristics of training population by race/ethnicity population by race/ethnicity

Characteristics	White (n = 1,116,901; 80.56%)	Black (n = 111,329; 8.03%)	Hispanic (n = 76,119; 5.49%)	Asian/Pacific Islander (n = 39,939; 2.88%)	Other (n = 42,201; 3.04%)
Predisposing factors					
Age, mean (SD)	73.05 (11.21)	66.54* (15.01)	69.44* (14.12)	74.43* (11.15)	68.54* (12.02)
Male, n (%)	468,222 (41.92)	44,874* (40.31)	32,347 (42.50)	16,598 (41.56)	23,363* (55.36)
Original reason for Medicare enrollment, n (%)					
OASI	879,276 (78.72)	53,687* (48.22)	46,705* (61.36)	34,136* (85.47)	32,697* (77.48)
DIB	234,229 (20.97)	53,910 (48.42)	27,545 (36.19)	5,379 (13.47)	9,100 (21.56)
ESRD	1,762 (0.16)	1,806 (1.62)	1,026 (1.35)	239 (0.60)	239 (0.57)
DIB and ESRD	1,634 (0.15)	1,926 (1.73)	843 (1.11)	185 (0.46)	165 (0.39)
Low-income subsidy, n (%)	242,284 (21.69)	69,068* (62.04)	51,798* (68.05)	22,808* (57.11)	12,684* (30.06)
Percentage of non-White population, mean (SD)	30.91 (19.30)	47.82* (18.26)	55.69* (21.95)	54.85* (18.78)	38.15* (21.58)
Proportion of married-couple families, mean (SD)	72.26 (6.10)	65.37* (7.96)	68.40* (6.29)	70.34* (6.24)	71.25* (7.02)
Proportion with education ≥ high school, mean (SD)	88.70 (4.69)	86.95* (4.71)	83.87* (7.38)	87.22 (5.22)	88.69 (4.79)
Per capita income (in $1,000), mean (SD)	54.94 (18.84)	54.55 (19.62)	57.64 (23.17)	70.42* (26.20)	58.91* (22.56)
Proportion without health insurance, mean (SD)	10.35 (4.65)	11.11* (4.52)	12.61* (6.57)	9.16* (4.20)	10.31 (4.83)
Enabling factors					
Metropolitan statistical area, n (%)	850,421 (76.14)	94,586* (84.96)	67,792* (89.06)	38,406* (96.16)	33,317* (78.95)
Health Professional Shortage Area, n (%)	1,021,134 (91.43)	106,629* (95.78)	73,537* (96.61)	37,747 (94.51)	39,238* (92.98)
Census region, n (%)					
Northeast	224,749 (20.12)	17,396* (15.63)	13,203* (17.35)	7,086* (17.74)	9,237* (21.89)
Midwest	285,721 (25.58)	20,649 (18.55)	7,254 (9.53)	4,074 (10.20)	9,060 (21.47)
South	410,712 (36.77)	64,807 (58.21)	24,633 (32.36)	7,822 (19.58)	12,116 (28.71)
West	195,719 (17.52)	8,477 (7.61)	31,029 (40.76)	20,957 (52.47)	11,788 (27.93)
Need factor					
Risk adjustment summary score, mean (SD)	1.22 (1.19)	1.56* (1.69)	1.44* (1.50)	1.26 (1.22)	1.08* (1.20)

Abbreviation: SD = standard deviation; OASI = old age and survivors insurance; DIB = disability insurance benefits; ESRD = end-stage renal disease.
* Indicates a statistically significant difference from non-Hispanic White patients ($P < 0.05$).

TABLES/FIGURES

Table 2. Pooled model performance metrics and differences in predictive power across racial and ethnic groups

Costs, racial/ethnic groups	AUC-ROC			AUC-PR			Brier score		
	PE	ED	95% CI, p-value	PE	ED	95% CI, p-value	PE	ED	95% CI, p-value
Healthcare									
White	0.7839			0.5776			0.1446		
Black	0.8353	0.0514	[0.0467, 0.0566], < 0.0001	0.7437	0.1661	[0.1584, 0.1748], < 0.0001	0.1425	-0.0021	[-0.0045, -0.00004], 0.0440
Hispanic	0.8287	0.0448	[0.0389, 0.0509], < 0.0001	0.7043	0.1267	[0.1162, 0.1373], < 0.0001	0.1398	-0.0048	[-0.0075, -0.0021], < 0.0001
Asian/Pacific Islander	0.8140	0.0301	[0.0202, 0.0387], < 0.0001	0.6130	0.0354	[0.0179, 0.0528], < 0.0001	0.1273	-0.0173	[-0.0207, -0.0137], < 0.0001
Other	0.8012	0.0173	[0.0081, 0.0264], < 0.0001	0.6098	0.0322	[0.0154, 0.0499], < 0.0001	0.1263	-0.0183	[-0.0217, -0.0146], < 0.0001
Medication									
White	0.8049			0.5647			0.1420		
Black	0.8077	0.0028	[-0.0020, 0.0076], 0.2680	0.6386	0.0739	[0.0629, 0.0842], < 0.0001	0.1567	0.0147	[0.0126, 0.0168], < 0.0001
Hispanic	0.8136	0.0087	[0.0030, 0.0147], 0.0040	0.6438	0.0791	[0.0664, 0.0915], < 0.0001	0.1540	0.0120	[0.0099, 0.0145], < 0.0001
Asian/Pacific Islander	0.8036	-0.0013	[-0.0099, 0.0071], 0.7800	0.6277	0.0630	[0.0456, 0.0799], < 0.0001	0.1590	0.0170	[0.0138, 0.0204], < 0.0001
Other	0.8103	0.0054	[-0.0023, 0.0139], 0.1600	0.5707	0.0060	[-0.0099, 0.0245], 0.4880	0.1398	-0.0022	[-0.0056, 0.0011], 0.1700

Abbreviation: AUC-ROC = the area under the receiver operating characteristic curve; AUC-PR = the area under the precision-recall curve; PE = point estimate; ED = estimated difference; CI = confidence interval.
Note: Estimated difference refers to the racial/ethnic difference in each model performance metric, calculated as the minority group value minus the White value.

Table 3. Transfer learning model performance and corresponding differences from pooled model across racial and ethnic minority groups

Costs, racial/ethnic groups	AUC-ROC			AUC-PR			Brier score		
	PE	ED	95% CI, p-value	PE	ED	95% CI, p-value	PE	ED	95% CI, p-value
Healthcare									
Black	0.8367	0.0014	[0.0007, 0.0020], < 0.0001	0.7498	0.0061	[0.0046, 0.0075], < 0.0001	0.1409	-0.0016	[-0.0019, -0.0012], < 0.0001
Hispanic	0.8315	0.0028	[0.0017, 0.0037], < 0.0001	0.7102	0.0059	[0.0039, 0.0080], < 0.0001	0.1385	-0.0013	[-0.0018, -0.0010], < 0.0001
Asian/Pacific Islander	0.8147	0.0007	[-0.0008, 0.0021], 0.3800	0.6158	0.0028	[0.000049, 0.0053], 0.0500	0.1267	-0.0006	[-0.0011, -0.0002], 0.0060
Other	0.8010	-0.0002	[-0.0018, 0.0013], 0.7240	0.6104	0.0006	[-0.0019, 0.0031], 0.6400	0.1261	-0.0002	[-0.0006, 0.0002], 0.3180
Medication									
Black	0.8124	0.0047	[0.0036, 0.0057], < 0.0001	0.6550	0.0164	[0.0131, 0.0198], < 0.0001	0.1540	-0.0027	[-0.0031, -0.0022], < 0.0001
Hispanic	0.8160	0.0024	[0.0012, 0.0035], < 0.0001	0.6525	0.0087	[0.0051, 0.0128], < 0.0001	0.1525	-0.0015	[-0.0020, -0.0011], < 0.0001
Asian/Pacific Islander	0.8105	0.0069	[0.0042, 0.0092], < 0.0001	0.6438	0.0161	[0.0103, 0.0216], < 0.0001	0.1559	-0.0031	[-0.0041, -0.0021], < 0.0001
Other	0.8116	0.0013	[-0.0005, 0.0030], 0.1620	0.5778	0.0071	[0.0019, 0.0118], 0.0080	0.1391	-0.0007	[-0.0013, -0.0001], 0.0240

Abbreviation: AUC-ROC = the area under the receiver operating characteristic curve; AUC-PR = the area under the precision-recall curve; PE = point estimate; ED = estimated difference; CI = confidence interval.
Note: Estimated difference refers to the difference in each model performance metric, calculated as the transfer learning value minus the pooled model value for each minority group.

STUDY LIMITATIONS

- Administrative claims and AHRF data did not capture patient behavioral characteristics and provider-level factors, which may have contributed to residual confounding and influenced predictive performance across racial and ethnic groups.
- Binary top-quartile cost outcomes simplified the continuous nature of expenditure and may not fully capture variation in spending patterns across racial/ethnic groups among Medicare beneficiaries.
- The analysis focused only on cost-based MTM eligibility and did not incorporate other existing criteria, such as chronic conditions or medication use, potentially limiting the representation of populations selected by plans.
- Findings may not generalize beyond fee-for-service Medicare beneficiaries to Medicare Advantage, commercially insured, uninsured, or other populations.
- The study evaluated racial/ethnic differences in predictive performance but did not assess causal mechanisms underlying these differences, limiting interpretation of factors contributing to the observed patterns.

CONCLUSIONS

- Despite concerns regarding algorithmic inequities across racial/ethnic groups, disparities in predictive performance were not identified, with minority populations generally demonstrating comparable or better performance in cost-based MTM eligibility models than White beneficiaries.
- Given the modest gains observed, transfer learning could represent a promising strategy for improving equity in predictive MTM eligibility models for racial and ethnic minority groups, although its utility may be greater in settings with smaller subgroup sample sizes than in this study.
- The effectiveness of transfer learning may depend on the predictive strength of the source population and similarity of cost-related patterns across racial/ethnic groups, emphasizing the need to explore bidirectional and alternative transfer learning strategies to improve predictive performance across all population groups.

ACKNOWLEDGEMENT

This research was supported by the National Institute On Aging of the National Institutes of Health under Award Number R01AG040146.

REFERENCES

1. Centers for Medicare and Medicaid Services. Part D Enhanced Medication Therapy Management model. Available at: https://innovation.cms.gov/innovation-models/enhancedmtm. Accessed April 22, 2026.
2. Huang J, Galal G, Etemadi M, et al. Evaluation and mitigation of racial bias in clinical machine learning models: scoping review. *JMIR Med Inform*. 2022;10(5):e36388.
3. Browning JA, Tsang CCS, Zeng R, et al. Racial/ethnic disparities in the enrollment of medication therapy management programs among Medicare beneficiaries with Alzheimer's disease and related dementias. *Curr Med Res Opin*. 2022;38(10):1715–1725.
4. Garuccio J, Tsang CCS, Wan JY, et al. Racial and ethnic disparities in the enrolment of Medicare medication therapy management programs. *J Pharm Health Serv Res*. 2023;14(2):188–197.
5. Gao Y, Cui Y. Deep transfer learning for reducing health care disparities arising from biomedical data inequality. *Nat Commun*. 2020;11(1):5131.
6. Gelberg L, Andersen RM, Leake BD. The Behavioral Model for Vulnerable Populations: application to medical care use and outcomes for homeless people. *Health Serv Res*. 2000;34(6):1273–1302.
7. Centers for Medicare and Medicaid Services. Medicare program; contract year 2015 policy and technical changes to the Medicare Advantage and the Medicare prescription drug benefit programs; proposed rule. *Fed Regist*. 2014;79:1917–2073.