

1                   **A Claims-Based Machine Learning Classifier of Modified Rankin Scale**  
2   **in Acute Ischemic Stroke**

3  
4   *Mamoon Habib, MSc<sup>a</sup>; Rafaella Cazé de Medeiros, MD<sup>a</sup>; Syed Muhammad Ahsan, MBBS, MD*  
5   *<sup>a</sup>; Aidan McDonald Wojciechowski, BS<sup>a</sup>; Maria A. Donahue, MD<sup>a</sup>; Deborah Blacker, MD, ScD<sup>b</sup>,*  
6   *<sup>d</sup>; Joseph P. Newhouse, PhD<sup>c, e, f, i</sup>; Lee H. Schwamm, MD<sup>g</sup>; M. Brandon Westover, MD, PhD<sup>h\*</sup>;*  
7   *Lidia MVR Moura, MD, PhD, MPH<sup>a\*</sup>*

8  
9   *\*co-senior authors*

10  
11   *<sup>a</sup> Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston,*  
12   *Massachusetts*

13   *<sup>b</sup> Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School, Boston,*  
14   *Massachusetts*

15   *<sup>c</sup> Department of Health Care Policy, Harvard Medical School, Boston, Massachusetts*

16   *<sup>d</sup> Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston,*  
17   *Massachusetts.*

18   *<sup>e</sup> Department of Health Policy and Management, Harvard T.H. Chan School of Public Health,*  
19   *Boston, Massachusetts.*

20   *<sup>f</sup> Harvard Kennedy School, Cambridge, Massachusetts*

21   *<sup>g</sup> Yale School of Medicine, New Haven, Connecticut.*

22   *<sup>h</sup> Department of Neurology, Beth Israel Lahey Health Medical System, Boston, Massachusetts.*

23   *<sup>i</sup> National Bureau of Economic Research, Cambridge, Massachusetts.*

24  
25   Corresponding Author:

26   Lidia MVR Moura

27   [lidia.moura@mgh.harvard.edu](mailto:lidia.moura@mgh.harvard.edu)

28   617.726.3311

29   55 Fruit Street, Wang 739D

30   Boston, MA 02114

31  
32  
33  
34   Word Count: 5,503

35 **ABSTRACT**

36 **Background:**

37 We developed a classifier to infer acute ischemic stroke (AIS) severity from Medicare claims  
38 using the Modified Rankin Scale (mRS) at discharge. The classifier can be utilized to improve  
39 stroke outcomes research and support the development of national surveillance tools.

40 **Methods:**

41 This was a multistate study included all participating centers in the Paul Coverdell National  
42 Acute Stroke Program (PCNASP) database from nine U.S. states. PCNASP was linked to  
43 Medicare data sets for patients hospitalized with AIS, employing demographics, admission  
44 details, and diagnosis codes to create unique patient matches. We included Medicare  
45 beneficiaries aged 65 and older who were hospitalized for an initial AIS from January 2018 to  
46 December 2020. Using Lasso-penalized logistic regression, we developed and validated a  
47 binary classifier for mRS outcomes and as a secondary analysis we used ordinal regression to  
48 model the full mRS scale. Performance was evaluated on held-out test data using ROC AUC,  
49 ROC Precision-Recall, sensitivity, and specificity.

50 **Results:** We analyzed data from 68,636 eligible patients. The mean age was 79.5 years old.  
51 77.5% of beneficiaries were White, 14% were Black, 2.6% were Asian, and 2% were Hispanic.  
52 The classifier achieved an ROC AUC score of 0.85 (95%CI: 0.85-0.86), sensitivity of 0.81  
53 (95%CI: 0.80-0.81), specificity of 0.73 (0.72 - 0.74), and Precision-Recall AUC of 0.90 (95%CI:  
54 0.90-0.91) on the test set.

55 **Conclusion:** Among Medicare beneficiaries hospitalized for AIS, the claims-based classifier  
56 demonstrated excellent performance in ROC AUC, Precision-Recall AUC, sensitivity, and  
57 acceptable specificity for mRS classification.

58 **Key Words:** Acute Ischemic Stroke (AIS), Classifier, Medicare, Modified Rankin Scale (mRS),  
59 Paul Coverdell National Acute Stroke Program (PCNASP)

60

61

## 62 **Clinical Perspective**

63 What Is New?

64 - Developed a novel claims-based classifier to infer acute ischemic stroke (AIS) severity using  
65 the Modified Rankin Scale (mRS) at discharge.

66 - Integrated Medicare claims with clinical data from the stroke registry, utilizing penalized  
67 logistic regression for both binary and ordinal classification.

68 What Are the Clinical Implications?

69 - Provides a robust tool for assessing stroke severity, which can enhance stroke outcomes  
70 research and quality improvement initiatives.

71 - Supports the development of national surveillance tools, potentially guiding clinical decision-  
72 making and resource allocation in stroke care.

## 73 **Research Perspective**

74 What New Question Does This Study Raise?

75 - How can claims-based severity classifiers be effectively integrated into existing stroke  
76 research and clinical practice to enhance outcome measurement?

77 - To what extent is the classifier generalizable to diverse populations beyond Medicare  
78 beneficiaries?

79 What Question Should be Addressed Next?

80 - Future research should evaluate the impact of incorporating such classifiers into risk  
81 adjustment processes and their effect on long-term stroke outcomes.

82 - Investigate whether similar modeling approaches can be adapted for other patient groups  
83 and healthcare settings to improve surveillance and treatment strategies.

84

## 85 INTRODUCTION

86 Every 40 seconds, someone in the United States (U.S.) has a stroke.<sup>1</sup> Stroke is one of  
87 the leading causes of long-term disability, affecting about 795,000 people in the U.S. annually.<sup>2</sup>  
88 Acute ischemic stroke (AIS) severity can be variable, with a significant portion of discharged  
89 patients presenting with declining functionality, leading to increased needs for rehabilitation and  
90 admission to nursing facilities.<sup>3</sup> Both modifiable (i.e., obesity, diabetes, cardiovascular disease,  
91 certain medications, physical inactivity, etc.) and non-modifiable stroke risk factors (i.e., age,  
92 sex, race/ethnicity, genetics) can help determine prognosis, which is crucial for early tailored  
93 intervention.<sup>4</sup>

94 Functional outcome prediction in AIS impacts the quality of patient care decisions.<sup>5,6</sup>  
95 Recent advances in computational and software technologies have greatly impacted the rise of  
96 Machine learning (ML) studies, offering more precise outcome measures.<sup>7-9</sup> ML models have  
97 identified several crucial factors to predict and classify functional outcomes, such as an initial  
98 National Institutes of Health Stroke Scale (NIHSS) score, age, fasting blood glucose, creatinine  
99 levels, and the modified Rankin Scale (mRS).<sup>10,11</sup> mRS has been widely used to assess AIS  
100 severity and clinical prognosis in electronic health records (EHRs) and registries.<sup>12</sup> The creation  
101 of models and classifiers can be personalized to assess outcomes in AIS patients, including the  
102 classification of mRS.<sup>8,9,13</sup> However, limited valid measures of stroke severity have hindered  
103 national, large-scale, claims-based studies.<sup>14</sup>

104 Despite this limitation, claims data may offer indirect clues about a patient's level of  
105 disability based on the types of claims filed. Leveraging a dataset that links claims to mRS  
106 scores, we explored whether supervised ML could develop a classifier to infer mRS from claims  
107 information. Such a model could enable the personalization of outcome assessments for AIS  
108 patients and the classification of mRS in large, claims-based studies, thereby configuring a tool  
109 for national surveillance of stroke severity.

110 We linked the Paul Coverdell National Acute Stroke Program (PCNASP) and Medicare  
111 claims-based inpatient data of older adults presenting with AIS to develop and validate the mRs  
112 classifier of stroke severity at discharge.

113

## 114 **METHODS**

115 The Medicare data supporting this study's findings are collected routinely by The  
116 Centers for Medicare & Medicaid Services (CMS) for billing purposes and were made available  
117 by CMS with no direct identifiers. All results were aggregated following CMS Cell Suppression  
118 Policies. Restrictions apply to the availability of these data, which were used under license for  
119 this study. Medicare data are available through CMS with their permission. PCNASP data are  
120 available through the CDC with their permission.

121 This study was approved by the Mass General Brigham Institutional Review Board's  
122 (IRB) ethical guidelines and followed the Strengthening the Reporting of Observational Studies  
123 in Epidemiology (STROBE) guidelines for observational studies<sup>15</sup> and the transparent reporting  
124 of multivariable prediction models developed or validated using clustered data (TRIPOD)<sup>16</sup> and  
125 the updated guidance for reporting clinical prediction models that use regression or machine  
126 learning methods (TRIPOD-AI).<sup>17</sup>

127

### 128 **Study Design**

129 We conducted a retrospective analysis of claims data from AIS patients using a sample  
130 from nine large U.S states. We aimed to develop and validate a classifier based on claims data  
131 that infers mRS at discharge.

132

### 133 **Data Source**

134 We accessed data from the PCNASP registry and Medicare Claims data. PCNASP  
135 collects data on stroke cases and captures discharge mRS scores reported by clinicians or

136 hospital staff.<sup>18</sup> The PCNASP registry includes information from 2008 to 2020 from the following.  
137 states: California; Georgia; Massachusetts; Michigan; Minnesota; New York; Ohio; Washington;  
138 and Wisconsin.

139 We then matched the PCNASP data on individuals aged 65 or older with data from fee-  
140 for-service Medicare, a national health insurance program administered by the Centers for  
141 Medicare & Medicaid Services (CMS).<sup>19</sup> The Medicare Provider Analysis and Review  
142 (MEDPAR) files contain extensive information about these beneficiaries, including patient  
143 demographics, admission and discharge dates, diagnosis, procedure codes, provider identifiers,  
144 and comorbidities.<sup>20</sup>

145

## 146 **Study Population**

147 We analyzed Medicare claims data for beneficiaries aged 65 and older hospitalized for  
148 AIS from January 2018 to December 2020. We included beneficiaries who were enrolled in  
149 traditional Medicare Part A (inpatient hospital insurance; care in a skilled nursing facility,  
150 hospice care, and some home health care) and Part B (physician and other medical provider  
151 services; outpatient care, medical supplies, and preventive services) who had mRS values  
152 documented in the PCNASP clinical database (based on ICD-10 code information).

153 We used a multi-step exclusion and inclusion process to refine our patient population.  
154 First, we excluded patients with missing mRS scores and deceased patients in the PCNASP  
155 data and then linked the remaining data with Medicare claims data. We found patients with a  
156 diagnosis of AIS in the Medicare claims data during 2018-2020 and used only their first stroke  
157 encounter. We next created two groups based on the availability of an mRS score for any stroke  
158 (Supplemental Figure 1). The first group included patients admitted to the hospital with a  $\geq 90\%$   
159 or more completion rate of mRS, while the second group included patients admitted to hospitals  
160 with less than  $< 90\%$  of mRS completion. We used 20% of the first group and all of the second

161 group as a training sample; the remaining 80% of the first group was set aside as an  
162 independent test sample.

163

## 164 **Linking Databases**

165 Because there were no unique patient identifiers common to both databases, we applied  
166 a matching strategy to link individuals in the PCNASP and Medicare datasets.<sup>21</sup> For this linkage  
167 we used variables such as age, gender, admission and discharge dates, diagnosis code,  
168 hospitals, and state. After linkage, we retained patients with unique matches, excluding cases  
169 where PCNASP IDs corresponded to multiple Medicare Beneficiary IDs and vice versa. Due to  
170 limited access to baseline institutionalized (non-outpatient) data, we excluded patients  
171 transferred from another hospital, skilled nursing facility (SNF), or other healthcare facilities.

172

## 173 **Variables**

174 We included demographic variables, medical history, treatments, and discharge  
175 outcomes. Most variables were extracted from the MEDPAR files. Those not included in  
176 MEDPAR were extracted from hospital level data by linking MEDPAR data with provider-level  
177 data and included variables such as bed size and hospital location, category and level. We  
178 included two stroke-related variables for inpatient conditions and procedures such as tissue  
179 plasminogen activator (tPA) and endovascular treatment. We used the value “1” if the condition  
180 or procedure was present and the value “0” if not. For continuous variables such as age and  
181 length of stay, we standardized their values. Categorical variables, such as race and admission  
182 type, were converted into dummy variables for use in the model. We used the variables included  
183 in the Chronic Conditions Warehouse (CCW) algorithms from Medicare to determine  
184 comorbidities and relevant patient medical history in our patient population.<sup>22</sup> CCW flagged 27  
185 chronic conditions for each beneficiary within the study period, which we used to determine if

186 the beneficiary had any comorbidities. We selected the first-ever criteria a beneficiary met for  
187 the chronic condition.

188

### 189 **Construct of Interest (Endpoint)**

190 Our primary endpoint was the accurate classification of mRS at discharge. We  
191 dichotomized the mRS scale into “favorable” if valued as equal or less than 2 (from no  
192 symptoms to slight disabilities) and “unfavorable” if the mRS score was  $> 2$  (interval from  
193 moderate disability to death).<sup>12,23</sup>

194 As a secondary analysis, we developed ordinal classifiers using the previous sampling  
195 approach to obtain more granularity among mRS categories. The two approaches of ordinal  
196 classification consist of a full mRS scale, one represented by 0: no symptoms; 1: no significant  
197 disabilities, despite symptoms; 2: slight disabilities; 3: moderate disability; 4: moderate to severe  
198 disability; 5: severe disability; 6: death.<sup>23, 24</sup> The second ordinal model consists of the same full  
199 scale but excludes the death category.

200

### 201 **Model Development**

202 *Primary analysis - Binary Classifier:* The binary classifier outputs probabilities for each  
203 class. A threshold of 0.5 was used to convert the probabilities into binary values. Predictions  
204 with a probability greater than or equal to 0.5 were assigned to the unfavorable mRS category,  
205 and those below 0.5 to the favorable class.

206 For development of our binary classifier, binary logistic regression with a lasso penalty  
207 was trained to predict the binary mRS category (favorable vs unfavorable). The best  
208 hyperparameters were determined through a grid-search hyperparameter tuning process. The  
209 hyperparameters included a range of the inverse regularization strength C ( $10^{-4}$  to 100),  
210 tolerance values (1e-4 to 1e-1), maximum iterations (5000 to 50000), solver methods ('liblinear'  
211 and 'saga'), and class weight settings (None and Balanced). The hyperparameters that

212 generated the largest area under the receiver operator characteristic curve (ROC AUC) were  
213 chosen. Stratified 5-fold cross-validation was used to evaluate the classifier's performance  
214 within the training set. The model was separately evaluated on the test set, which was not used  
215 in model development.

216 *Secondary analysis - Ordinal Classifier:* We also trained a classifier on the full-scale  
217 mRS values using ordinal regression. The ordinal regression model outputs probabilities for  
218 each class. To assign class labels, we selected the class with the maximum predicted  
219 probability.

220 We fitted the model as a parallel classifier with a logit link and Lasso L1 penalty using  
221 the ordinalNet R package. Grid-search hyperparameter tuning was performed on the training  
222 dataset to select the best model based on lambda and family values. We defined a sequence of  
223 lambda values (ranging from 0.001 to 0.01) and multiple family values (cumulative, acat, sratio,  
224 cratio).

225 For each family type in the classifier, models were fitted across a range of lambda  
226 values and log-likelihood was used to evaluate model performance. The optimal lambda for  
227 each family type was selected as the value that achieved the highest log-likelihood, once we  
228 selected the optimal family type and lambda value, we refitted the final classifier on the training  
229 data with the chosen parameters. We tested the refitted model on the test dataset to check for  
230 its generalizability.

231

## 232 **Performance Metrics**

233 For both primary and secondary analyses, we evaluated classifier's performance using  
234 ROC AUC and Area Under the Precision-Recall Curve (PR AUC) to assess the model's ability  
235 to distinguish between classes. Sensitivity and specificity, were included to evaluate the model's  
236 ability in identifying true positives and true negatives.

237 To calculate confidence intervals (CI) for our performance metrics, we performed 10,000  
238 iterations of bootstrap random sampling with replacement in each iteration. We created a  
239 distribution for each metric and calculated 95% confidence intervals to show the classifier's  
240 performance variability.

241

## 242 **RESULTS**

### 243 **Characteristics of the samples**

244 We assessed 295,241 hospital admissions for AIS between January 2018 and  
245 December 2020 for eligibility. After applying our inclusion and exclusion criteria, our sample  
246 included 68,636 unique Medicare beneficiaries who were 65 years old or older with a first  
247 admission for AIS and available discharge mRS scores. We obtained distinctive patient hospital  
248 encounters with  $<$  or  $\geq$  90% completion of the mRS (N= 33,654 and N= 34,982, respectively)  
249 (Supplemental Figure 1).

250 The mean age for the full sample was 79.53 (SD 8.7), and 77.5% of beneficiaries were  
251 White, 14% were Black or African American, 2.7% were Asian, and 2% were Hispanic (Table 1).  
252 The mean age for our test data was 79.76 (SD 8.7). Approximately 91% of our patient sample  
253 was admitted through emergency care. Regarding discharge disposition, the test set data was  
254 more evenly distributed between home, SNFs, and inpatient rehabilitation facilities with 28%,  
255 23%, and 19%, respectively, followed by interventions, such as receipt of tissue plasminogen  
256 activator and endovascular intervention. The remaining percentage was distributed between  
257 approximately 100 other discharge disposition variables. Concerning comorbidities, 71% of  
258 beneficiaries had hypertension, 39% diabetes, and 29% congestive heart failure. A further  
259 breakdown of the full sample, training, and test set demographics can be found in Table 1. We  
260 used 63 covariates to predict a scale score, such as demographics, medical history, treatments,  
261 and discharge outcomes (a list can be found in Figure 1 and Supplement Table 5).

262

## 263 **Binary Classifier**

264 On the held-out test data, our binary classifier achieved an ROC AUC score of 0.85  
265 (95%CI: 0.85 – 0.86, Figure 2), sensitivity of 0.81 (95%CI: 0.80 – 0.81), specificity of 0.73 (0.72  
266 - 0.74), and Precision-Recall AUC of 0.90 (95%CI: 0.90 – 0.91, Figure 3). Figure 1 shows the  
267 model's feature coefficients sorted/ranked by their contribution to its predictions. Palliative care  
268 was the strongest predictor (2.02) of unfavorable mRS outcomes. Similarly, coded hemiplegia  
269 (0.71), and the use of ventilator during the AIS hospitalization (0.61) were strong predictors of  
270 unfavorable outcomes. Several features were also associated with a lower likelihood of  
271 unfavorable outcomes. For instance, binary discharge disposition (home vs others) had the  
272 strongest negative coefficient (-1.95), suggesting that favorable discharge outcomes strongly  
273 predict better recovery. Transesophageal echocardiogram (-0.31), and tPA administration (-  
274 0.25), were associated with favorable outcomes.

275

## 276 **Ordinal Classifier**

277 For our secondary analysis, the ordinal model's overall performance on the test data is  
278 presented in Table 2. The model demonstrates a stronger ability to distinguish between mRS  
279 scores 0 (No Symptoms) and 5/6 (Severe Disability/Death) compared to its performance in  
280 differentiating intermediate outcomes (1–4) [see Supplementary Figure 2].

281 Classes 2 (Slight Disability) and 3 (Moderate Disability) showed the lowest ROC AUC  
282 and PR AUC scores. Supplementary Figure 4 presents a box plot of grouped probabilities,  
283 highlighting how the model conflates mRS scores 2 and 3 with mRS score 4. The model's ability  
284 to distinguish between mRS scores 0 (No Symptoms) and 5/6 (Severe Disability/Death) is  
285 higher compared to its performance in differentiating intermediate outcomes (1–4) [see  
286 Supplementary Figure 2].

287 Additionally, we excluded death to evaluate whether the model's performance improves  
288 in predicting intermediate outcomes 2 and 3, however, no significant changes in performance

289 were observed. The model's performance is presented in the supplementary section. The  
290 coefficients from both ordinal models (see Supplementary Tables 6 and 7) were consistent with  
291 those observed in the binary model. For instance, in the full-scale mRS ordinal model, discharge  
292 disposition [i.e., discharged home] (coefficient = 1.99) increased the odds of falling into a lower  
293 (better) mRS category, whereas palliative care (coefficient = -2.72) increased the odds of a  
294 higher (worse) category.

295

## 296 **DISCUSSION**

297         Considering the clinical burden of AIS and its influence on patient mortality, rate of  
298 disability, medical complications, and healthcare expenditures, it is fundamental to monitor the  
299 impact, severity, and prognosis of this condition.<sup>1,31,34</sup> Our interpretation of the identified factors  
300 driving the classification highlights their strong face validity and consistency with existing  
301 literature as they align with clinical expectations and prior studies. Palliative care, hemiplegia,  
302 endotracheal intubation, and feeding device usage were strong predictors of unfavorable mRS  
303 outcomes, which is consistent with established knowledge on poor prognostic factors in acute  
304 ischemic stroke. Similarly, favorable discharge disposition (e.g., discharged home), tPA  
305 administration and brain imaging (CT or MRI) were associated with better outcomes, reinforcing  
306 the importance of early and effective stroke management.

307         We developed and validated a claims-based classifier to accurately identify stroke  
308 severity measured by mRS at discharge in patients aged 65 or older who experienced AIS. By  
309 leveraging administrative claims data, our classifier demonstrates strong predictive performance,  
310 achieving excellent accuracy for categorizing stroke severity. This tool holds significant potential  
311 for facilitating large-scale research on stroke outcomes and improving national surveillance  
312 efforts, enabling more effective monitoring of stroke care quality and recovery outcomes.  
313 Validated claims-based classifiers for AIS surveillance are also important for observing

314 geographic trends and are essential for population health research, which in turn can inform  
315 public health policy and national guidelines to improve clinical practice.<sup>3</sup>

316 Previous studies have utilized ML methods for stroke functional outcome  
317 assessment.<sup>5,13,28</sup> Joon Nyung Heo et al. measured mRS 90 days after hospital discharge using  
318 three learning algorithm models: deep neural network, random forest, and logistic regression.  
319 The study had similar results with the logistic regression model (AUC 0.85), while the best  
320 performance was by the deep neural network model (AUC 0.88)<sup>28</sup> In our study, logistic  
321 regression for mRS classification at discharge yielded positive results with the ROC AUC score  
322 of 0.85, reiterating the results seen in other models.<sup>5,13,28,31</sup>

323 Most importantly, the previous studies were limited by selection bias due to their  
324 sampling from single regions of the US.<sup>5,13,28,31</sup> Our study overcomes this challenge by including  
325 a national, large-scale sample with representation of patients and practices from nine U.S.  
326 states spanning all regions of the US. Therefore, our cohort provides a more robust, inclusive,  
327 and representative claims-based classifier for beneficiaries with AIS than has been heretofore  
328 available.

329 Prior studies creating mRS stroke-severity classifiers used a random assignment  
330 approach within hospitals to create training and test sets.<sup>9,13</sup> This approach is potentially biased  
331 because random sampling does not account for hospital-level patterns in patient intake and  
332 reporting. We addressed this by categorizing the training and test data sets depending on  
333 whether hospitals reported < or ≥ 90% mRS completion. We only used data from those with  
334 ≥90% mRS completion as the test set, with a random 20% allocated to the training set for  
335 representativeness, allowing the classifier to be trained and tested with higher-quality data and  
336 partially accounting for potential bias in random sampling.

337 Furthermore, our study used binary and ordinal regression methods to classify the mRS  
338 score in AIS patients. Binary analyses yield results that are more easily interpreted by  
339 examining the absolute risk reduction between the two groups but do not exploit the within-

340 group variation.<sup>23</sup> We therefore also implemented an ordinal approach to achieve better use of  
341 the dataset.<sup>23,32</sup> The use of the ordinal method increased statistical power and decreased loss of  
342 information when compared with previous studies.<sup>5,33</sup>

343 Other research groups have focused on validating admission stroke severity, such as  
344 electronic health record (EHR)-based classifiers of NIHSS at admission.<sup>25</sup> This is important work,  
345 as classifiers of stroke severity at admission can inform resource allocation while patients are  
346 admitted and guide other care measures. However, we focused on leveraging claims data to  
347 classify stroke functional outcomes at discharge using the mRS. The mRS is important because  
348 it provides information on patient functional outcomes, which can inform the prioritization of  
349 post-discharge stroke care allocation and predictions of long-term outcomes, among other  
350 applications.<sup>26,27</sup> The score's ability to predict the level of functionality makes it an essential tool  
351 for national-level surveillance using administrative databases.<sup>5</sup>

352

### 353 **Limitations**

354 While we used a nationally representative stroke registry covering nine U.S. states and  
355 its major stroke centers linked to administrative claims data, results may not be generalized to  
356 states not included in our data set or smaller community healthcare centers. In addition, our  
357 selection of older adults  $\geq 65$  covered by fee-for-service Medicare may not represent other  
358 patient populations. Slightly over half of eligible Medicare beneficiaries are now enrolled in  
359 Medicare Advantage "Part C" instead of traditional Medicare. Beneficiaries must also be  
360 enrolled in Parts A and B, as well as Part B premium. Recent studies have shown that  
361 enrollment in lower-cost Medicare Advantage plans has increased among low-income and  
362 racial/ethnic minorities.<sup>35</sup> Future studies assessing these groups would benefit these  
363 populations.

364 We excluded also 12,894 patients transferred from another hospital, skilled nursing  
365 facility (SNF), or other healthcare facilities from the analytical sample, which may have omitted a

366 subset of the AIS population with a higher burden of baseline comorbidities. We selected this  
367 approach due to limited access to predictor data from these groups. Including these patients  
368 could have enhanced classifier representativeness and performance by increasing the sample  
369 size and introducing greater variability. Nevertheless, our classifier demonstrated high  
370 performance while capturing a broad and still nationally representative segment of the AIS  
371 population.

372 We were limited by data availability for the Medicare and PCNASP datasets. While  
373 utilization of administrative claims linked to data registries represents a vast source of  
374 information for research purposes,<sup>36</sup> some inherent limitations (e.g., human-type errors of  
375 scores and clinical scales and missing data e.g., missing mRS scores and other stroke-related  
376 variables) are surely present. Despite these limitations, national administrative claims data  
377 remains valuable in representing large-sized populations and their reflections.<sup>37,38</sup>

378 Lastly, the replicability of our classifier can present some challenges, for example,  
379 requiring at least two databases to perform linkage of common unique identifiers and extract  
380 multiple variables. Users looking to replicate should have experience in Python and R  
381 Programming and can refer to the GitHub link in the Supplementary Material for replication.

382

### 383 **Conclusion**

384 We developed a claims-based classifier to identify stroke severity in AIS patients using  
385 discharge mRS. Importantly, we partially addressed potential bias by accounting for hospital-  
386 level patterns in sampling using mRS completion rates. Our classifier has expanded on previous  
387 research by using PCNASP and Medicare-linked data from several states to assess stroke  
388 severity.

389 **ACKNOWLEDGEMENTS**

390 **FUNDING**

391 This study was supported by the NIH (1R01AG073410-01)

392 **DISCLOSURES OF CONFLICT OF INTEREST**

393 M.H., R.C.M., S.M.A., A.M.W., and M.A.D. have no conflict of interest to disclose.

394 D.B. receives support from the National Institute of Neurological Disorders and Stroke and the  
395 National Institute on Aging and reports no conflict of interest.

396 J.P.N. reports no conflict of interest.

397 M.B.W. is a co-founder, scientific advisor, and consultant to Beacon Biosignals and has a  
398 personal equity interest in the company.

399 L.M.S. receives research support from the National Institute of Neurological Disorders and  
400 Stroke and the National Institute on Aging and reports no conflict of interest.

401 L.M.V.R.M. receives research support from the Epilepsy Foundation of America, the National  
402 Institute of Neurological Disorders and Stroke, and the National Institute on Aging and reports  
403 no conflict of interest.

404 **Supplemental Materials**

405 Link to GitHub Code to replicability

406 Figures S1-S4

407 Tables S1-9

408

## 409 REFERENCES

- 410 1. Tsao CW, Aday AW, Almarzooq ZI, et al. Heart Disease and Stroke Statistics—2023  
411 Update: A Report From the American Heart Association. *Circulation*.  
412 2023;147(8):e93-e621. doi:10.1161/CIR.0000000000001123
- 413 2. CDC. Stroke Facts. *Stroke*. October 24, 2024. Accessed November 22, 2024.  
414 <https://www.cdc.gov/stroke/data-research/facts-stats/index.html>
- 415 3. Ziaeeian B, Xu H, Matsouaka RA, et al. US Surveillance of Acute Ischemic Stroke  
416 Patient Characteristics, Care Quality, and Outcomes for 2019. *Stroke*.  
417 2022;53(11):3386-3393. doi:10.1161/STROKEAHA.122.039098
- 418 4. Boehme AK, Esenwa C, Elkind MSV. Stroke Risk Factors, Genetics, and Prevention.  
419 *Circ Res*. 2017;120(3):472-495. doi:10.1161/CIRCRESAHA.116.308398
- 420 5. Zhang MY, Mlynash M, Sainani KL, Albers GW, Lansberg MG. Ordinal Prediction  
421 Model of 90-Day Modified Rankin Scale in Ischemic Stroke. *Front Neurol*. 2021;12.  
422 doi:10.3389/fneur.2021.727171
- 423 6. Fonarow GC, Kapral MK, Schwamm LH. Future of Quality and Outcomes Research  
424 in Stroke. *Circ Cardiovasc Qual Outcomes*. 2015;8(6\_suppl\_3):S66-S68.  
425 doi:10.1161/CIRCOUTCOMES.115.002309
- 426 7. Fralick M, Colak E, Mamdani M. Machine Learning in Medicine. *N Engl J Med*.  
427 2019;380(26):2588-2590. doi:10.1056/NEJMc1906060
- 428 8. Esra Zihni MSC, Bryony McGarry PHD, John Kelleher PHD. Moving Toward  
429 Explainable Decisions of Artificial Intelligence Models for the Prediction of Functional  
430 Outcomes of Ischemic Stroke Patients. *Exon Publ*. Published online April 29,  
431 2022:73-90. doi:10.36255/exon-publications-digital-health-explainable-decisions
- 432 9. Su PY, Wei YC, Luo H, et al. Machine Learning Models for Predicting Influential  
433 Factors of Early Outcomes in Acute Ischemic Stroke: Registry-Based Study. *JMIR*  
434 *Med Inform*. 2022;10(3):e32508. doi:10.2196/32508
- 435 10. Li X, Pan X, Jiang C, et al. Predicting 6-Month Unfavorable Outcome of Acute  
436 Ischemic Stroke Using Machine Learning. *Front Neurol*. 2020;11.  
437 doi:10.3389/fneur.2020.539509
- 438 11. Liu Y, Yu Y, Ouyang J, et al. Functional Outcome Prediction in Acute Ischemic  
439 Stroke Using a Fused Imaging and Clinical Deep Learning Model. *Stroke*.  
440 2023;54(9):2316-2327. doi:10.1161/STROKEAHA.123.044072
- 441 12. Banks JL, Marotta CA. Outcomes Validity and Reliability of the Modified Rankin  
442 Scale: Implications for Stroke Clinical Trials. *Stroke*. 2007;38(3):1091-1096.  
443 doi:10.1161/01.STR.0000258355.23810.c6

- 444 13. Lee J, Park KM, Park S. Interpretable machine learning for prediction of clinical  
445 outcomes in acute ischemic stroke. *Front Neurol.* 2023;14:1234046.  
446 doi:10.3389/fneur.2023.1234046
- 447 14. ElHabr AK, Katz JM, Wang J, et al. Predicting 90-day modified Rankin Scale score  
448 with discharge information in acute ischaemic stroke patients following treatment.  
449 *BMJ Neurol Open.* 2021;3(1). doi:10.1136/bmjno-2021-000177
- 450 15. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of  
451 Observational Studies in Epidemiology (STROBE) statement: guidelines for  
452 reporting observational studies. *Lancet Lond Engl.* 2007;370(9596):1453-1457.  
453 doi:10.1016/S0140-6736(07)61602-X
- 454 16. Debray TPA, Collins GS, Riley RD, et al. Transparent reporting of multivariable  
455 prediction models developed or validated using clustered data (TRIPOD-Cluster):  
456 explanation and elaboration. *BMJ.* 2023;380:e071058. doi:10.1136/bmj-2022-  
457 071058
- 458 17. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance  
459 for reporting clinical prediction models that use regression or machine learning  
460 methods. *BMJ.* 2024;385:e078378. doi:10.1136/bmj-2023-078378
- 461 18. Taha M, Habib M, Lomachinsky V, et al. Evaluating the concordance between  
462 International Classification of Diseases, Tenth Revision Code and stroke severity as  
463 measured by the National Institutes of Health Stroke Scale. *BMJ Neurol Open.*  
464 2024;6(2). doi:10.1136/bmjno-2024-000831
- 465 19. Centers for Medicare and Medicaid Services. Centers for Medicare and Medicaid  
466 Services. Accessed January 7, 2025. <https://www.cms.gov>
- 467 20. Medicare Provider Analysis and Review (MedPAR). ResDAC (Research Data  
468 Assistance Center). Accessed January 7, 2025. [https://resdac.org/cms-  
469 data/files/medpar](https://resdac.org/cms-data/files/medpar)
- 470 21. Patorno E, Schneeweiss S, George MG, et al. Linking the Paul Coverdell National  
471 Acute Stroke Program to commercial claims to establish a framework for real-world  
472 longitudinal stroke research. *Stroke Vasc Neurol.* 2022;7(2):114-123.  
473 doi:10.1136/svn-2021-001134
- 474 22. Chronic Conditions Data Warehouse. Chronic Conditions Data Warehouse. July 1,  
475 2025. Accessed January 7, 2025. <https://www2.ccwdata.org/web/guest/home/>
- 476 23. Ganesh A, Luengo-Fernandez R, Wharton RM, Rothwell PM, on behalf of the  
477 Oxford Vascular Study. Ordinal vs dichotomous analyses of modified Rankin Scale,  
478 5-year outcome, and cost of stroke. *Neurology.* 2018;91(21):e1951-e1960.  
479 doi:10.1212/WNL.0000000000006554

- 480 24. Steinar CM, Smith MC, Byblow WD. Prediction Tools for Stroke Rehabilitation.  
481 *Stroke*. 2019;50(11):3314-3322. doi:10.1161/STROKEAHA.119.025696
- 482 25. Brott T, Adams HP, Olinger CP, et al. Measurements of acute cerebral infarction: a  
483 clinical examination scale. *Stroke*. 1989;20(7):864-870.  
484 doi:10.1161/01.STR.20.7.864
- 485 26. Rankin J. Cerebral Vascular Accidents in Patients over the Age of 60: II. Prognosis.  
486 *Scott Med J*. 1957;2(5):200-215. doi:10.1177/003693305700200504
- 487 27. de Haan R, Limburg M, Bossuyt P, van der Meulen J, Aaronson N. The Clinical  
488 Meaning of Rankin 'Handicap' Grades After Stroke. *Stroke*. 1995;26(11):2027-2030.  
489 doi:10.1161/01.STR.26.11.2027
- 490 28. Heo J, Yoon JG, Park H, Kim YD, Nam HS, Heo JH. Machine Learning–Based  
491 Model for Prediction of Outcomes in Acute Stroke. *Stroke*. 2019;50(5):1263-1265.  
492 doi:10.1161/STROKEAHA.118.024293
- 493 29. Lotan E. Emerging Artificial Intelligence Imaging Applications for Stroke  
494 Interventions. *Am J Neuroradiol*. 2021;42(2):255-256. doi:10.3174/ajnr.A6902
- 495 30. Mouridsen K, Thurner P, Zaharchuk G. Artificial Intelligence Applications in Stroke.  
496 *Stroke*. 2020;51(8):2573-2579. doi:10.1161/STROKEAHA.119.027479
- 497 31. Daidone M, Ferrantelli S, Tuttolomondo A. Machine learning applications in stroke  
498 medicine: advancements, challenges, and future perspectives. *Neural Regen Res*.  
499 2024;19(4):769. doi:10.4103/1673-5374.382228
- 500 32. Risselada R, Lingsma HF, Molyneux AJ, et al. Prediction of two month modified  
501 Rankin Scale with an ordinal prediction model in patients with aneurysmal  
502 subarachnoid haemorrhage. *BMC Med Res Methodol*. 2010;10(1):86.  
503 doi:10.1186/1471-2288-10-86
- 504 33. Roozenbeek B, Lingsma HF, Perel P, et al. The added value of ordinal analysis in  
505 clinical trials: an example in traumatic brain injury. *Crit Care*. 2011;15(3):R127.  
506 doi:10.1186/cc10240
- 507 34. Ovbiagele B, Goldstein LB, Higashida RT, et al. Forecasting the future of stroke in  
508 the United States: a policy statement from the American Heart Association and  
509 American Stroke Association. *Stroke*. 2013;44(8):2361-2375.  
510 doi:10.1161/STR.0b013e31829734f2
- 511 35. Meyers DJ, Mor V, Rahman M, Trivedi AN. Growth In Medicare Advantage Greatest  
512 Among Black And Hispanic Enrollees. *Health Aff (Millwood)*. 2021;40(6):945-950.  
513 doi:10.1377/hlthaff.2021.00118

- 514 36. Fernandes M, Cardall A, Jing J, et al. Identification of patients with epilepsy using  
515 automated electronic health records phenotyping. *Epilepsia*. 2023;64(6):1472-1481.  
516 doi:10.1111/epi.17589
- 517 37. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for  
518 epidemiologic research on therapeutics. *J Clin Epidemiol*. 2005;58(4):323-337.  
519 doi:10.1016/j.jclinepi.2004.10.012
- 520 38. MacKay EJ, Stubna MD, Chivers C, et al. Application of machine learning  
521 approaches to administrative claims data to predict clinical outcomes in medical  
522 and surgical patient populations. *PLoS ONE*. 2021;16(6):e0252585.  
523 doi:10.1371/journal.pone.0252585
- 524

525 **TABLES AND FIGURES**

526 **Table 1. Demographic Characteristics**

<b>Characteristics</b>	<b>Full Sample (N = 68,636)</b>	<b>Training / Validation (n = 40,650)</b>	<b>Test (n = 27,986)</b>
<b>Age, mean (SD)</b>	79.53 (8.67)	79.38 (8.63)	79.76 (8.71)
<b>Gender (%)</b>			
Female	37,439 (54.54)	22,045 (54.23)	15,394 (55.00)
Male	31,197 (45.45)	18,605 (45.76)	12,592 (45.00)
<b>Race (%)</b>			
White	53,192 (77.49)	31,794 (78.21)	21,398 (76.45)
Black	9,629 (14.02)	5,394 (13.26)	4,235 (15.13)
Asian	1,821 (2.65)	1,146 (2.81)	675 (2.41)
Hispanic	1,361 (1.98)	753 (1.85)	608 (2.17)
Other	1,483 (2.16)	876 (2.15)	607 (2.16)
Unknown	997 (1.45)	593 (1.45)	404 (1.44)
North American Native	153 (0.22)	94 (0.23)	59 (0.21)
<b>Admission Type (%)</b>			
Emergency	62,639 (91.26)	36,657 (90.18)	25,982 (92.83)
Urgently	4,911 (7.15)	3,375 (8.30)	1,536 (5.48)
Trauma Center	559 (0.81)	326 (0.80)	233 (0.83)
<b>Intensive Care Unit (ICU) Type (%)</b>			
Intermediate IOCU	13,325 (19.41)	8,379 (20.61)	4,946 (17.6)
General	11,569 (16.85)	6,786 (16.69)	4,783 (17.09)
Medical	3,033 (4.41)	1,599 (3.93)	1,434 (5.12)

Surgical	1,501 (2.18)	1,073 (2.63)	428 (1.52)
Trauma	153 (0.22)	117 (0.28)	36 (0.12)
Other	144 (0.20)	63 (0.15)	81 (0.28)
<b>Discharge Disposition (%)*</b>			
Home/Self-care	18,931 (27.58)	11,233 (27.63)	7,698 (27.50)
Skilled Nursing Facility	15,426 (22.47)	9,056 (22.27)	6,370 (22.76)
Inpatient Rehabilitation Facility	12,856 (18.73)	7,213 (18.67)	5,266 (18.81)
<b>Interventions (%)</b>			
Tissue Plasminogen Activator	9,001 (13.11)	5,579 (13.72)	3,422 (12.22)
Endovascular Intervention	3,089 (4.50)	1,780 (4.37)	1,309 (4.67)
<b>Comorbidities (%)</b>			
Acute Myocardial Infarction	4,290 (6.25)	2,544 (6.26)	1,746 (6.24)
Atrial Fibrillation	13,304 (19.38)	7,700 (18.94)	5,604 (20.02)
Diabetes	26,708 (38.91)	15,581 (38.33)	11,127 (39.76)
Congestive Heart Failure	19,766 (28.80)	11,555 (28.43)	8,211 (29.34)
Hypertension	48,451 (70.59)	28,418 (69.91)	20,033 (71.58)

527 **Legend:** Baseline demographics, admission type, Intensive Care Unit (ICU) Type, and

528 comorbidities stratified by sample, training, and test groups.

529 \*We did not include all discharge disposition variable in the table, as there are over 100 existing

530 items. We reported the most relevant ones to this table.

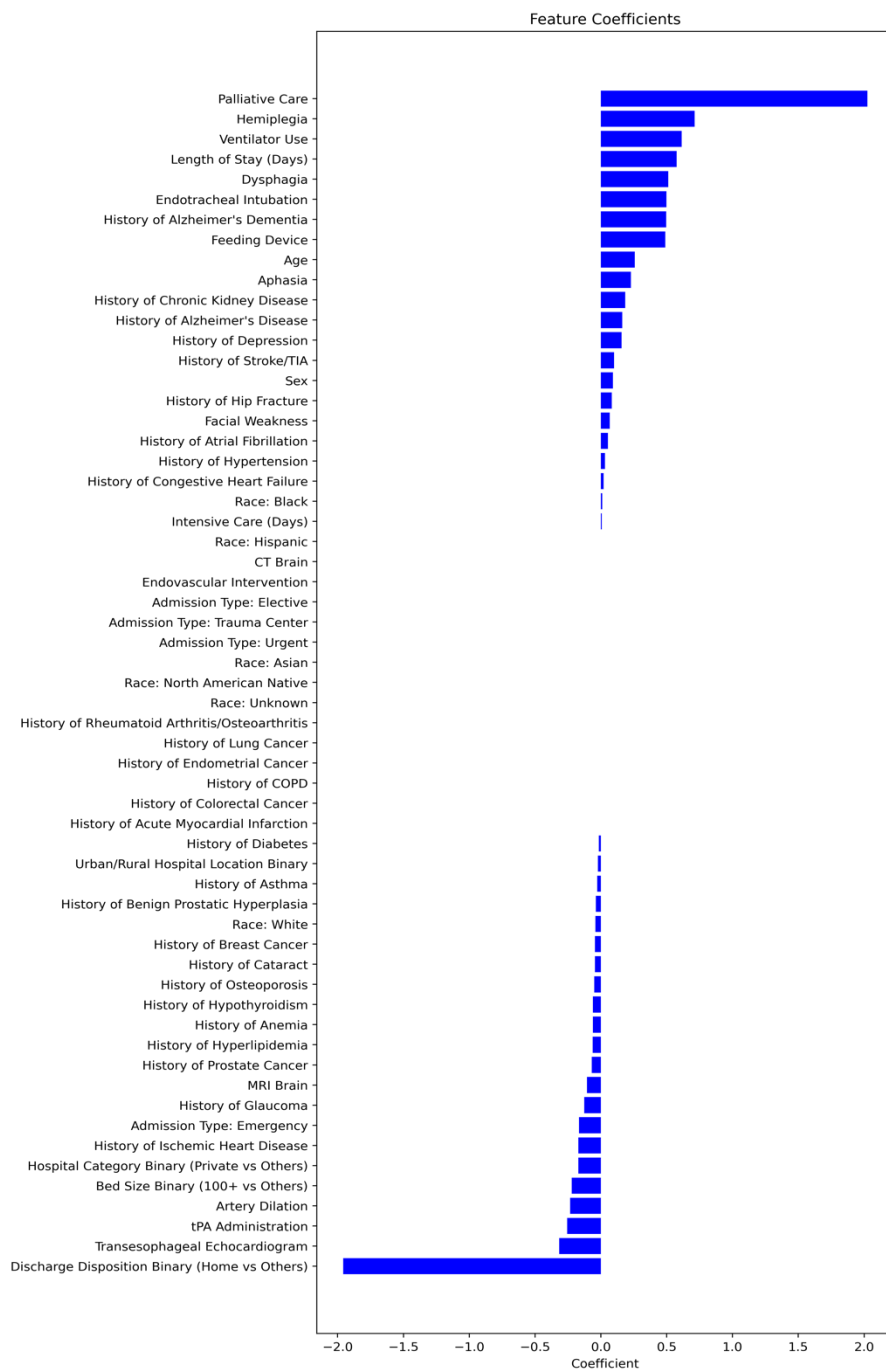
531 **Table 2. Full-Scale Ordinal Model Performance**

Metric	Score [CI]
ROC AUC	0.81 [0.80 – 0.81]
Precision-Recall AUC	0.39 [0.37 – 0.39]
Sensitivity	0.42 [0.41-0.2]
Specificity	0.89 [0.88 - 0.89]

532 **Legend:** Performance Metrics from Full-Scale Ordinal. We report micro-average ROC

533 AUC and Precision-Recall AUC.

534 **Figure 1. Model Features**



535

536 **Legend:** The full list of the classifier's features and their coefficient values. COPD: Chronic

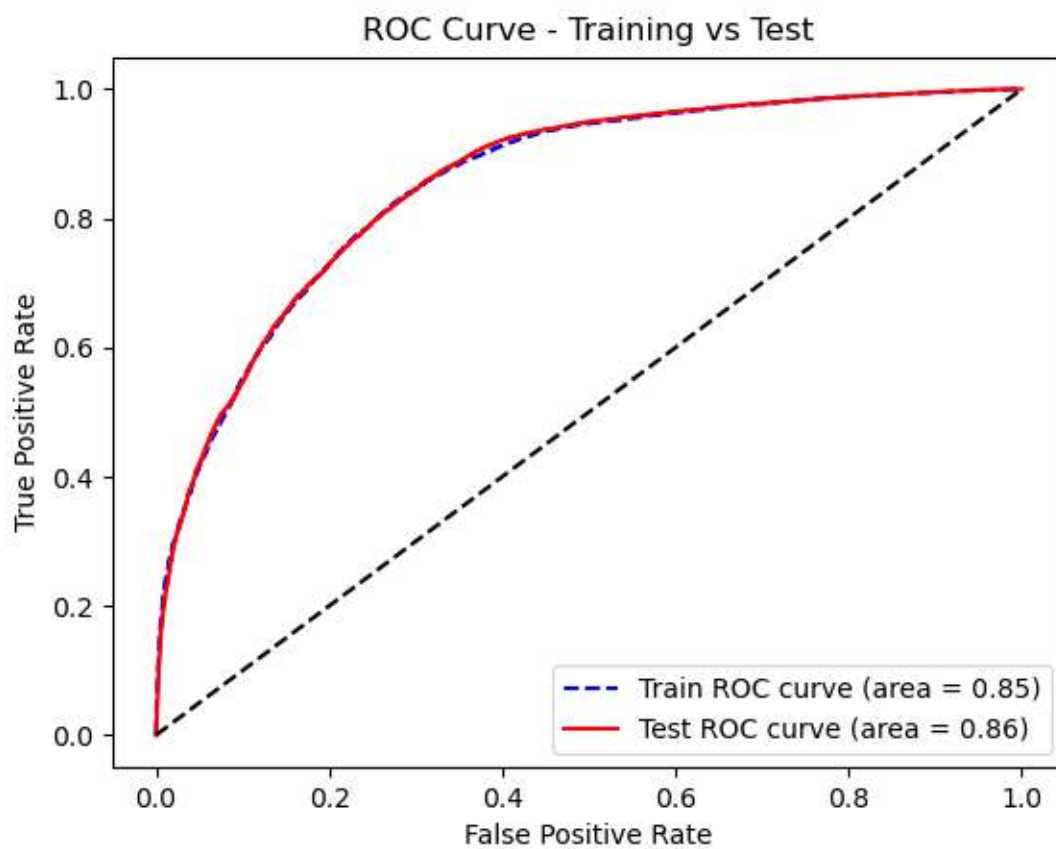
537 obstructive pulmonary disease; ICU: Intensive Care Unit; tPA, tissue plasminogen activator; CT,

538 computed tomography; MRI, Magnetic resonance imaging.

539

24

540 **Figure 2. ROC (Receiver Operating Characteristic) Curve**

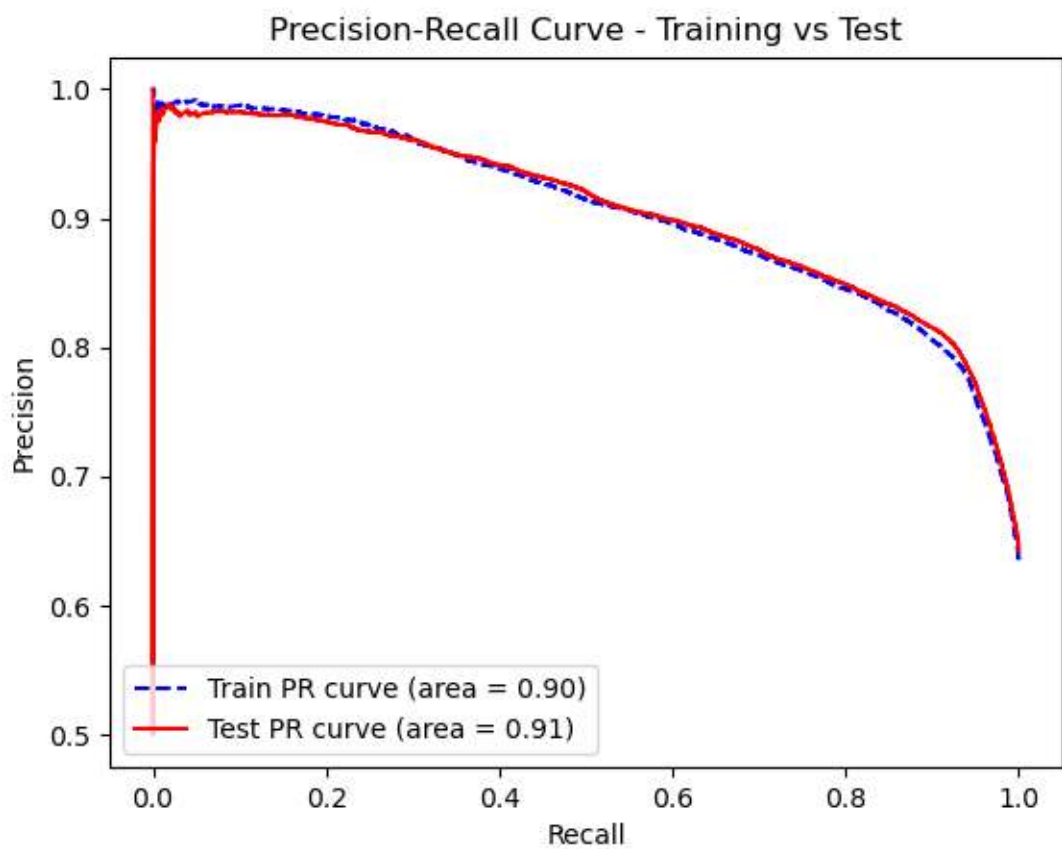


541

542 **Legend:** Comparison of the ROC in both the training and test sets of the classifier.

543

544 **Figure 3. Precision-Recall Area Under the Curve for Binary Classifier**



545

546 **Legend:** Comparison of Precision-Recall Curve of the classifier in the training and test sets.

547