

<https://doi.org/10.1038/s43856-024-00603-x>

Comparison of machine learning models for the prediction of hypertension in transgender patients undergoing gynecologic surgery

Check for updates

Reetam Ganguli^{1,2}, Jordan Franklin³, Xiaotian Yu⁴, Alice Lin⁵, Aditi Vichare⁶✉ & Stephen Wagner⁶

Abstract

Background Transgender patients face a higher burden of cardiovascular morbidity due to structural and biological stressors, particularly in low-resource settings. No studies exist comparing machine learning model development strategies for this unique patient cohort and limited literature exists comparing data/outcomes between transgender and cisgender populations.

Methods We compare machine learning models trained solely on transgender patients against models developed on a size-matched and ratio-matched cohort of cisgender patients and a 300-fold larger, ratio-matched cohort of cisgender patients undergoing obstetric/gynecologic procedures in the National Surgical Quality Improvement Program from January 1, 2005 through December 31, 2019. All models were developed to predict the outcome of hypertension. Statistical significance between models was calculated using 5-by-2 fold cross validation hypothesis testing.

Results Among 626,102 patients having an obstetric/gynecologic surgery, there are 1959 transgender patients of which 85,405 (13.7%) have hypertension requiring medication. Saliiently, the logistic regression machine learning models trained selectively on the transgender cohort have an AUC of 0.865 (95% CI: 0.83–0.90), with an accuracy of 85% (95% CI: 0.80–0.87) compared to ($p < 0.05$) the logistic regression model trained on the 300-fold larger combined cohort which has an AUC of 0.861 (95% CI: 0.82–0.90), with an accuracy of 83% (95% CI: 0.80–0.87).

Conclusion Machine learning models can be trained on smaller, selectively transgender populations and may perform similarly or better to predict cardiovascular outcomes in transgender patients, than models developed on predominantly cisgender patients; this can be useful in lower-resource settings with smaller-volume transgender patients.

Plain language summary

Transgender patients face a higher burden of cardiovascular disease. Statistical models that predict cardiovascular disease-related outcomes, such as high blood pressure (hypertension), may be useful to clinicians to guide treatment, but existing models are mainly developed in cisgender populations. Here, we developed models to predict hypertension in patients undergoing surgery, and compared models developed using data from cisgender patients, transgender patients, or mixed populations to see if this affected how well these models could predict hypertension in the transgender population. We ultimately found that one of our models trained on a much smaller cohort of solely transgender patients outperformed the same model trained on a 300-times larger population of mixed cisgender and transgender patients. These findings might help to guide future efforts to develop statistical approaches to accurately predict health outcomes in transgender patients.

Predicting hypertension in transgender patients, remains a challenge, and the efficacy of machine learning (ML) models developed solely on transgender cohorts to predict cardiovascular outcomes compared to traditional ML models trained on predominantly cisgender cohorts has been undocumented in the medical literature. Transgender individuals have a current gender identity or expression that differs

from their assigned sex at birth. Though the prevalence of transgender individuals as reported in medical literature is rising, there currently exists no medical literature examining the efficacy of ML in predicting surgical outcomes for transgender patients¹. Current ML models are trained on large cohorts of patients, the majority of which are cisgender². Cisgender patient cohorts do not accurately reflect the

¹Brown University, Providence, RI, USA. ²University of California Los Angeles, Los Angeles, CA, USA. ³Department of Computer Sciences, Georgia Institute of Technology, Atlanta, GA, USA. ⁴Department of Mathematics, University of Virginia, Charlottesville, VA, USA. ⁵Warren Alpert Medical School, Providence, RI, USA. ⁶Department of Obstetrics and Gynecology, Harvard Medical School, Boston, MA, USA. ✉e-mail: avichare@g.ucla.edu

unique sociologic, demographic, and clinical features of transgender patient cohorts^{3,4}.

Current literature indicates that in addition to notable structural health disparities, transgender individuals also have differing complex clinical features in comparison to cisgender individuals^{5,6}. Transgender patients often face violence, discrimination, and barriers in access to health care disproportionately more often than cisgender patients which can serve as a risk factor for heart disease^{5,7}. Multiple studies have shown an increased prevalence in cardiovascular disease among transgender patients^{8,9}. This high hypertensive disease burden in transgender populations, compared to their cisgender counterparts, compared to any other comorbidity is why hypertension was selected as the outcome variable for this study. Hypertensive comorbidity is associated with more complex clinical management, increased health care costs, and worse health outcomes accounting for more cardiovascular disease deaths than any other modifiable risk factor¹⁰. Given that interpreting health data with reference levels and statistics specific to transgender patients can improve patient outcomes, further study of hypertension in transgender patient populations can help inform a physician’s decision to prescribe anti-hypertensive medications to reduce the risk of cardiovascular disease in transgender patients¹¹. The development of automated prediction models trained solely on transgender cohorts, to prognose cardiovascular outcomes for transgender patient populations have not been studied to date.

It has been well documented that ML models are able to have better prognostic potential when compared to clinical judgment, highlighting the evolving importance of their role in managing patient care¹². The prediction power of ML models is largely dependent on the composition of the training data; current ML models do not fit transgender patients well due to the predominantly cisgender cohort used to train models². We hypothesized that ML models developed on selectively transgender cohorts would be more accurate than general ML models trained on mixed cisgender and transgender cohorts, specifically for the prediction of hypertension severe enough to require medication in transgender patients undergoing a gynecologic or obstetric operation of any kind. Ultimately, we found that ML models trained on transgender compared to ML models trained on much larger cisgender cohorts were generally statistically similar with the sole exception of logistic regression models. Logistic regression models developed on solely transgender cohorts were able to statistically outperform logistic regression models trained on primarily cisgender cohorts that were over 300-fold larger in prediction hypertension for transgender patients.

Methods

Data

The American College of Surgeons National Surgical Quality Improvement Program (ACSNSQIP) database is a national surgical registry utilized to measure risk-adjusted outcomes of multiple surgical procedures spanning multiple surgical specialties. Over 700 hospitals report over one million surgical cases a year in the NSQIP dataset. The data is audited for accuracy and prospective variables are collected by trained clinical reviewers.

Study population

Patients in the ACSNSQIP database who were coded as having a gynecologic or obstetric surgery within January 2005 through December 2019 and were coded as having a Male sex met the inclusion criteria for the transgender cohort for this study and patients coded as having a Female sex met the inclusion criteria for the cisgender cohort for this study. This study was exempt from IRB review pursuant to section 4ii of the of the IRB Exemption requirements and Brown University’s Institutional Guidelines and agreement with the ACSNSQIP data use agreement was required. The American College of Surgeons collects the ACSNSQIP data with informed consent and provides the data to medical researchers; therefore, there was no necessity to reobtain patient consent.

Cohort development

We assembled 3 cohorts of patients for 2 main experiments (Table 1). The first cohort consists of all transgender patients that met the inclusion criteria (we will refer to this as the transgender cohort). The second cohort consists of a volume-matched and class ratio-matched cisgender and transgender patients available in the data (which we will refer to as the cisgender cohort). The goal for the cisgender cohort was to create a smaller dataset that emulated the observed compositions of the transgender patients and cisgender patients in the NSQIP dataset. The third cohort consists of all the transgender patients and cisgender patients who met the inclusion criteria of being recorded in the ACS NSQIP within January 2005 through December 2019 and were coded as having an obstetric or gynecologic surgery (which we will refer to as the combined cohort).

The cisgender cohort and transgender cohort were derived from the combined cohort. Cisgender patients were selected at random from cisgender patients in the combined cohort and transgender patients were selected at random from the transgender cohort to create the volume and ratio matched cisgender cohort of predominantly cisgender patients.

In the cisgender cohort, the total number of patients selected were equal to the total number of patients in the transgender cohort, in order to have a consistent sample size during model development. This cisgender cohort was intended to be a microcosm of the combined cohort and was therefore volume matched to the lower sample size of the transgender cohort to get a fairer comparison between ML models developed on the 2 cohorts. The ratio of cisgender to transgender patients in this cisgender cohort were directly predicated on the ratio of observed cisgender to transgender patients in the combined cohort, to emulate the observed ratios of transgender and cisgender patients in a real medical database. Therefore, this cohort was predominantly cisgender, due to the lower representation of transgender patients in the ACSNSQIP, representative of the lower proportion of transgender patients documented in most medical databases.

Outcome

The primary outcome variable analyzed was a diagnosis of hypertension severe enough to require medication, which may impact the patient’s risk for cerebrovascular, renal and cardiac disease. To be documented as a positive, the patient’s hypertension must be recorded in their medical record and their hypertension must be severe enough that to warrant administration of antihypertensive medication (like calcium channel blockers, diuretics, beta

Table 1 | Cohort development breakdown

Cohort Development Table			
Cohort Name	Transgender Cohort	Cisgender Cohort (Volume-Matched)	Combined Cohort
Number of Transgender	1959 (100%)	6 (0.31%)	1959
Patients in cohort			(0.31%)
Number of Cisgender Patients in cohort	0 (0%)	1953 (99.7%)	624143 (99.7%)
Total Number of Patients in cohort	1959	1959	626,102

Cohort development breakdown table with number and proportion of transgender and cisgender patients per cohort.

blockers, and ACE inhibitors) within 30 days prior to their index surgery, or during the time the patient is being considered as a candidate for surgery. Furthermore, the patient must have been receiving or required (if non-compliant) long-term treatment of their chronic hypertension exceeding 2 weeks to be coded as a yes for this outcome. Although this dataset consisted of surgical patients because this variable was solely recorded preoperatively, it can be used to model and predict hypertension in nonsurgical candidates as well.

The class balance ratio for the hypertension outcome variable was kept consistent between the combined cohort and the volume-matched cisgender cohort. The ratio of cisgender patients that had hypertension to cisgender patients who did not have hypertension in the large, combined dataset were preserved in the development of the smaller, volume-matched cisgender cohort to emulate the real, observed distribution of hypertension cases in cisgender patients. For transgender patients in the volume-matched cisgender cohort, the same ratio of transgender patients with hypertension to transgender patients without hypertension were kept constant to the observed ratio in the transgender cohort.

Machine learning models

Any patients carrying blank/NULL values for the outcome variable column were removed to eliminate any uncertainty/inaccuracy from the training. These patients with missing values were omitted from the analysis to avoid any ascertainment bias in erroneously classifying a positive case as a negative case and vice versa. The recording of these values are audited by the NSQIP and quality checked to ensure that they are accurately documented. Then, blank data were handled by multivariate iterative imputation in order to reduce bias in the data. Binary values that were imputed through multivariate imputation were rounded to the nearest whole number (0 or 1) to maintain medical consistency and interpretability within the data. The outcome variable was removed from the data frame prior to this process and was appended back on after imputation to avoid introducing inaccuracies in model development.

The cohort was split at the patient level such that no training data could appear in the testing set. All variables studied in the analysis were included in the model to optimize the predictive potential of the model and preserve intervariable correlations to optimize model performance.

Selecting individuals was done randomly to assemble all cohorts. For each of the 3 cohorts, a 75–25% stratified train test split was performed to preserve the hypertension class ratio between the training set and test set. The test set for all models developed on all cohorts was a set of 25% of the patients in the transgender cohort, unique from the patients in the training set for transgender patients. This was done to ensure that the predictive potential of all models specifically in the prognosis on cardiovascular outcomes in transgender patients was being evaluated and compared. The scikit learn package's train-test-split function was used as a random assortment algorithm were used to segment cohorts into training and testing sets to reduce bias. Blinding was not possible due to need to develop ML models, but no patients were fully observed at the individual level, patient data in the NSQIP is de-identified, and aggregate patient data was stored in the form of variables to mitigate bias.

ML models were selected based on existing literature^{3,12} and narrowed to supervised models due to their higher accuracy rates and the presence of labeled data in the training set. ML models were hyperparameter optimized through a grid search and was validated through a 5-fold cross validation to obtain the optimal hyperparameters yielding the best results on the testing set.

Variable importance

Variable importance was determined based on the model. For the random forest model, variable importance (VI) is determined using the mean decrease in Gini index/impurity. High mean decrease in the Gini Index indicates more importance. For the logistic regression model, VI is found by taking the absolute value of coefficients of the ultimate model, ranking the coefficients by magnitude; a larger coefficient value indicates

higher importance. For the XGBoost model, VI is calculated for a single tree's importance by improving the node purity, and then summing the importance over each boosting iteration. The VI averages all importances across each variable for all decision trees to formulate a ranking. For this model, we used the gain of each tree to formulate the importance rankings, where a larger gain indicates higher importance¹².

Statistical analysis

Descriptive statistical analysis was utilized to assess differences in the mean clinical features for the cisgender and transgender cohort. Measurements were taken from distinct samples. Initial analysis was done by conducting an independent, one-way analysis of variance (ANOVA) test, equivalent to a 2-tail t-test when done for two independent groups, of every independent variable included in the models, segmented between the cisgender and transgender cohorts, to compare if these features were represented more in transgender vs cisgender cohorts.

After ML models were developed on the cisgender, transgender, and combined cohorts, they were assessed on the testing set of transgender patients, unique from model development, by calculating the area under the curve (AUC) of the model's receiver operating characteristic (ROC), which was obtained through bootstrapping. The threshold-independent nature of discrimination of the AUC makes it a strong metric for our analysis. A salient limitation of using AUC ROC for imbalanced datasets include sensitivity to changes in predictions for the minority class. For example, if there are a low number of patients for positive class, then the AUC score may vary widely depending on how the model predicts for the positive class, which may not be indicative of how the model would prospectively perform given the real distribution.

Furthermore, AUC scores in imbalanced data may be artificially inflated because false positive rates do not drop as drastically when the number of total true negatives is very large. This is why metrics like the F1 score that account for precision (which is highly sensitive to false positive rates irrespective of high true negative values) help to better contextualize model performance. Because AUC ROC metrics can be affected by class imbalance present within the data, the unweighted F1 score and Matthew's Correlation Coefficient (MCC) metrics were also obtained for each model, along with a 95% confidence interval for each metric across each model. The MCC is a statistical test evaluating model performance by calculating the total discrepancy between the model prediction and true value.

To compare the statistical significance between the performance of the ML models developed on the transgender, cisgender, and combined cohorts, 5 by 2 cross validation fold hypothesis testing was utilized between the ML models developed on the transgender and cisgender cohorts and between the ML models developed on the transgender and combined cohorts¹³. Only ML models of the same type, developed on the different cohorts, were compared against each other. This hypothesis testing framework was chosen over other frameworks like ten-fold cross validation due to its relatively lower Type I error, its ability to be modified to overcome lack of independence in the data, and its ability to obtain a strong estimate of generalization error and variance of the generalization error between the performance of the 2 compared models. In 5 by 2 cross validation, a paired t-test is conducted between the performance of the 2 models compared and a *p* value is generated under the null hypothesis that that both models perform equally well on their given dataset.

All analyses were conducted using the Sklearn version 0.24.2 package and pandas version 1.5.0 package in Python (Python Software Foundation) and R 4.1.0.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Results

Cohort characteristics

The combined cohort contained 1959 transgender patients and 624,143 cisgender patients, totaling 626,102 patients having an obstetric or gynecologic surgery logged in the ACS NSQIP database between the 2005 and 2019 calendar years. The transgender cohort contained solely the 1959 transgender patients from the combined cohort (and therefore 0 cisgender patients). The cisgender cohort contained 6 transgender patients and 1953 cisgender patients, the same total size as the transgender cohort. The transgender/cisgender ratio for the cisgender cohort was matched to be as close as possible (rounding up) to the transgender/cisgender ratio observed in the combined cohort. The 6 transgender patients in the cisgender cohort were pulled from the test set of transgender patients to ensure that the training set for the cisgender cohort and transgender cohort contained different transgender patients to keep patient cohorts unique for optimal comparison.

Out of 1959 patients in the transgender cohort, 418 had severe hypertension. Out of 1959 patients in the cisgender cohort, 533 had hypertension requiring medical therapy. Out of 624,143 cisgender and 1959 transgender patients in the combined cohort, 85,405 had hypertension requiring medication.

The input data was sectioned off into binary features which were represented as preoperative and clinical history features, found in Table 2, and postoperative parameters and clinical outcomes found in Table 3.

Model variable importance

52 statistically significant clinical parameters were included in the model. The model variable importance plots of all 3 models trained on the transgender, cisgender, and combined cohort can be found in Fig. 1. The top postoperative predictors of death, across models trained on all cohorts, were age and weight. Interestingly, ML models developed on the transgender cohort highlighted the presence of diabetes mellitus as a high-weight feature for the prediction of severe hypertension in transgender patients, which models trained on the cisgender and combined cohort failed to identify.

Model performance for models developed on transgender cohort

The extreme gradient boosting model (XGBoost) had an AUC of 0.84 (95% CI: 0.79–0.88), with an accuracy of 85% (95% CI: 0.79–0.86), and a weighted F1 score of 0.72 (95% CI: 0.67–0.77).

The random forest (RF) model had an AUC of 0.85 (95% CI: 0.81–0.89), with an accuracy of 86% (95% CI: 0.81–0.87), and a weighted F1 score of 0.74 (95% CI: 0.69–0.79).

The logistic regression (LR) model had an AUC of 0.86 (95% CI: 0.83–0.90), with an accuracy of 85% (95% CI: 0.80–0.87), and a weighted

F1 score of 0.73 (95% CI: 0.68–0.78). A comprehensive comparison of model performance metrics across models developed on the transgender, cisgender, and combined cohorts, respectively, is included in Table 4. A comparative AUC ROC curve plot for the logistic regression model developed across the transgender, cisgender, and combined cohorts can be observed in Fig. 2.

When hypothesis testing to compare the statistical significance between the performance of the RF model developed on the transgender cohorts compared to the RF model developed on the cisgender cohort, there was no statistical significance found between the performance of the models ($p > 0.05$). The performance of the LR models between the transgender and cisgender cohort was unable to be determined.

When comparing the performance of the RF models developed on the transgender cohorts against the RF models developed on the 300-fold larger combined cohort, there was also no statistical significance found between the performance of the models ($p > 0.05$). Lastly, the performance of the LR models between the transgender and combined cohort was found to be significantly different ($p < 0.05$).

Discussion

In our ultimate analysis, it was found that all 3 ML models had the highest or equal raw accuracy when selectively trained on the transgender patient cohort compared to the to-scale trans-cis cohort and the combined cohort. It is important to note that while the confidence intervals for the accuracies overlap across cohorts on all models, 95% confidence intervals from two subgroups may overlap substantially and still have a statistically significant difference¹⁴. To better ascertain whether the results between cohorts were statistically significant, we used a 5 by 2 cross validation fold hypothesis testing framework to ascertain statistical significance in the performance of the RF and LR models between the transgender and cisgender cohorts, as well as between the transgender and combined cohorts.

The lack of statistical significance found between the RF model developed on the transgender and cisgender cohorts indicates that for similar sized transgender and mixed, predominantly cisgender cohorts, forest-based ML algorithms can perform similarly for the prediction of severe hypertension in transgender populations. However, the presence of similarly sized transgender and cisgender cohorts are unlikely to be found in the vast majority of observed medical data, where the patients are predominantly cisgender. The lack of statistical significance found between our RF models developed on the transgender cohort and the RF models developed on the combined cohort is an important finding helping to address this; it demonstrates that ML models developed on solely transgender populations can perform statistically similar to ML models developed on predominantly cisgender cohorts over 300-fold larger in the prediction of hypertensive morbidity in transgender populations. Interestingly, the statistically significant difference in the performance of the LR models developed on the transgender cohort versus the LR models in the combined cohort indicate that simpler, logistic regression predicated classifiers have the potential to be comparatively better at capturing statistical relationships unique to transgender cohorts for the prediction of outcomes in transgender patients when compared against decision tree-based classifiers.

A challenge for the development of ML models on transgender patient populations is the lower representation of transgender patients in ML training data sets. This is since transgender patient populations have a smaller observed representation in existing medical databases; this comparatively smaller sample size can be a deterrent to developing ML models for transgender communities developed on cohorts who actually identify as transgender patients. By demonstrating the potential benefit or statistically similar performance of ML models developed on saliently smaller transgender cohorts, we hope to motivate further development of ML models on transgender patient populations helping to address the observed lack of ML models for transgender populations, who face a disproportionately higher disease and discrimination burden compared to their cisgender counterparts¹⁵.

Table 2 | Preoperative Variable Features

Characteristic	Transgender N = 1959	Cisgender N = 1959	p value	OR
Inpatient/ outpatient status	1125 (57%) 834 (44%)	1089 (56%) 869 (44%)	0.25	0.87
Age of Patients	45 ± 1	49 ± 1	>0.99	1.03
Principal anesthesia technique	1958 (100%) 1 (<0.1%)	1958 (100%) 1 (<0.1%)	<0.001	217,970.00
Disseminated Cancer	1930 (99%) 29 (1.5%)	1922 (98%) 36 (1.8%)	0.38	0.64
Open Wound (with or without infection)	1941 (99%) 18 (0.9%)	1952 (100%) 7 (0.3%)	0.014	0.22
ASA Classification	345 (18%) 1614 (82%)	229 (12%) 1730 (88%)	<0.001	1.15

Summary table comparing preoperative variable features between transgender and cisgender patients, where hypertension requiring medication is the outcome variable.

Table 3 | Postoperative Variable Summary Table

Characteristic	Transgender N = 1959	Cisgender N = 1959	p value	OR	95% CI
Length of total hospital stay	1.46 ± 0.13	1.53 ± 0.13	0.46	0.98	0.91, 1.06
Occurrences Superficial Surgical site			0.36		
No Complication	1940 (99%)	1933 (99%)			
Superficial Incisional SSI	19 (1.0%)	25 (1.3%)		1.19	0.60, 2.32
Open Wound/Wound Infection			0.070		
No	1952 (100%)	1957 (100%)			
Yes	7 (0.4%)	1 (<0.1%)		0.06	0.01, 0.71
Occurrences Organ Space SSI			0.60		
No	1941 (99%)	1943 (99%)			
Yes	18 (0.9%)	15 (0.8%)		1.05	0.48, 2.30
Occurrences Wound Disrupt			0.44		
No	1950 (100%)	1952 (100%)			
Yes	9 (0.5%)	6 (0.3%)		0.82	0.24, 2.75
Occurrences Pneumonia			0.32		
No	1949 (99%)	1952 (100%)			
Yes	10 (0.5%)	6 (0.3%)		0.22	0.05, 0.88
Occurrences Unplanned Intubation			>0.99		
No	1957 (100%)	1957 (100%)			
Yes	2 (0.1%)	1 (<0.1%)		0.16	0.01, 2.90
Occurrences Pulmonary Embolism			0.50		
No	1956 (99%)	1956 (99%)			
Yes	3 (0.2%)	2 (0.2%)		0.38	0.05, 2.83
Occurrences Ventilator > 48 H			>0.99		
No	1957 (100%)	1958 (100%)			
Yes	2 (0.1%)	0 (0%)		0.00	0.00, Inf
Occurrences Progressive Renal Insufficiency			>0.99		
No	1958 (100%)	1957 (100%)			
Yes	1 (<0.1%)	1 (<0.1%)		2.44	0.12, 51.7
Occurrences Acute Renal Fail			0.50		
No	1959 (100%)	1957 (100%)			
Yes	0 (0%)	1 (<0.1%)		1,138,109	0.00, Inf
Occurrences Urinary Tract Infection			0.91		
No	1922 (98%)	1920 (98%)			
Yes	37 (1.9%)	38 (1.9%)		0.86	0.52, 1.42
CVA/Stroke with neurological deficit			>0.99		
No	1958 (100%)	1958 (100%)			
Yes	1 (<0.1%)	0 (0%)		0.00	0.00, Inf
Occurrences Cardiac Arrest Requiring CPR			>0.99		
No	1958 (100%)	1957 (100%)			
Yes	1 (<0.1%)	1 (<0.1%)		8.00	0.30, 215
Occurrences Bleeding Transfusions			<0.001		
No Bleeding Complication	1912 (98%)	1873 (96%)			
Bleeding Complication	47 (2.4%)	85 (4.3%)		1.12	0.70, 1.78
Occurrences DVT/Thrombophlebitis			0.083		
No	1956 (100%)	1949 (100%)			
Yes	3 (0.2%)	9 (0.5%)		6.59	1.20, 36.2
Occurrences Sepsis			0.089		
No	1947 (99%)	1953 (100%)			
Yes	12 (0.6%)	5 (0.3%)		0.63	0.17, 2.26
Return to OR			0.038		
No	1925 (98%)	1939 (99%)			
Yes	34 (1.7%)	19 (1.0%)		0.54	0.28, 1.05
Days from operation to discharge	1.33 ± 0.11	1.41 ± 0.11	0.26	1.01	0.91, 1.11

Summary table comparing postoperative variable features between transgender and cisgender patients, where hypertension requiring medication is the outcome variable.

Model Variable Importance Plot for Random Forest Model

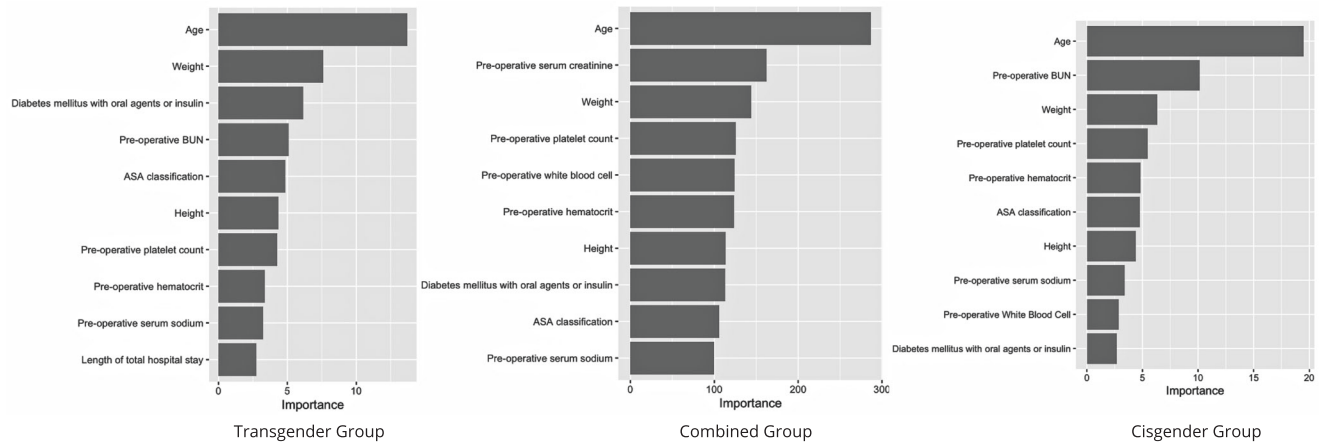


Fig. 1 | Model variable importance plots. Model variable importance plots for Random Forest Model trained on all 3 cohorts. BUN blood urea nitrogen, ASA American Society of Anesthesia.

Table 4 | Comparative model performance metrics

Comparative Model Metrics Between Cohorts				
Model	Accuracy (95% CI)	ROC AUC (95% CI)	F1 Score (95% CI)	MCC (95% CI)
ML Models Developed on Transgender Cohort				
Logistic Regression	85% (80–87%)	0.86 (0.83–0.90)	0.73 (0.68–0.78)	0.48 (0.38–0.57)
Random Forest	86% (81–87%)	0.85 (0.81–0.89)	0.74 (0.69–0.79)	0.49 (0.39–0.59)
XGBoost	85% (79–86%)	0.84 (0.79–0.88)	0.72 (0.67–0.77)	0.45 (0.35–0.55)
ML Models Developed on Cisgender Cohort				
Logistic Regression	80% (79–86%)	0.85 (0.81–0.89)	0.72 (0.67–0.77)	0.45 (0.35–0.55)
Random Forest	80% (79–86%)	0.84 (0.80–0.88)	0.70 (0.65–0.75)	0.42 (0.32–0.52)
XGBoost	80% (78–85%)	0.84 (0.79–0.88)	0.64 (0.59–0.70)	0.35 (0.25–0.46)
ML Models Developed on Combined Cohort				
Logistic Regression	83% (80–87%)	0.86 (0.82–0.90)	0.73 (0.68–0.78)	0.47 (0.37–0.57)
Random Forest	85% (82–88%)	0.89 (0.85–0.92)	0.76 (0.71–0.81)	0.53 (0.43–0.62)
XGBoost	85% (82–88%)	0.88 (0.85–0.91)	0.76 (0.71–0.81)	0.52 (0.43–0.62)

Comparative model performance metrics (with 95% CI) across 3 ML models independently developed on the transgender, cisgender, and combined cohorts (MCC = Matthew’s correlation coefficient). CI confidence interval, AUC ROC area under the receiver operating characteristic curve, MCC Matthew’s correlation co-efficient.

Even with statistically similar results, our analysis reveals that a cohort of just 1953 transgender patients can be used for the development of ML models that perform statistically similarly and possibly better than ML models trained on a combined cohort of 626,230 patients for the prediction of hypertension in transgender patients. This is an important implication, suggesting that healthcare providers can use models developed specifically on smaller transgender patient cohorts to achieve comparable performance to ML models trained on combined transgender and cisgender patient cohorts that are larger by over 300-fold. Several healthcare institutions lack

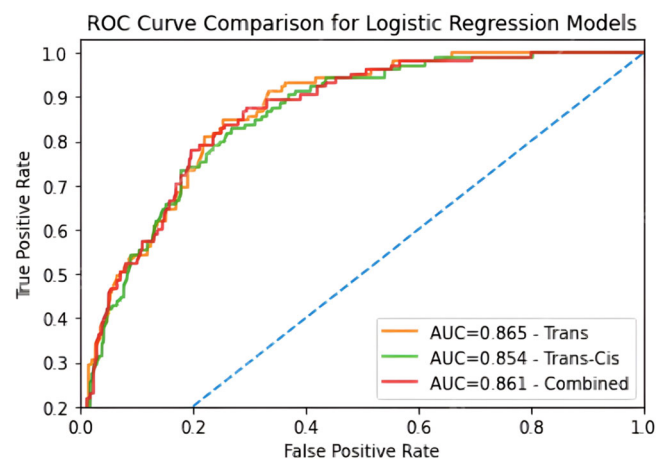


Fig. 2 | Comparative AUC ROC curves for Logistic Regression models. Comparative AUC ROC curves for Logistic Regression models developed on the transgender, cisgender, and combined cohorts for the prediction of severe hypertension in transgender patients. ROC receiver operating characteristic, AUC area under the curve.

the ability to develop robust ML models due to the lack of patient data available for training combined with the large number of patients needed to train an accurate model¹⁶. This challenge can be addressed for transgender patients by using low-volume transgender patient cohorts for training ML models in more healthcare settings to account for clinical differences found in transgender patients⁵. This can help accelerate the development of ML models and help to optimize clinical impact for transgender patients, who can receive accurate, automated prognoses sooner, in a larger number of locations.

Our results could be explained by the salient structural health disparities faced by transgender individuals distinct from their cisgender counterparts—specifically limited access to health care settings and adverse physiologic effects from discrimination⁵. There is a clear need for ML models with disadvantaged populations in mind¹⁷. Caceres et al. have posited that sociodemographic, clinical, and discriminatory stressors are correlated with poorer cardiovascular health outcomes and higher cardiovascular disease rates in Black and Latinx LGBTQ adults, particularly sexual minority women¹⁵. Additionally, cross-sex hormone therapy has been

shown to lead to adverse cardiovascular effects, such as increased risk of thromboembolisms, hypertension, and cardiovascular mortality^{18,19}. The use of gender affirming hormone therapy has been shown to increase visceral adipose tissue deposits which may mechanistically predispose transgender patients to thromboembolic and myocardial infarction events. The combination of these factors could explain the differences in transgender patient cohorts in comparison to cisgender patient cohorts. Given the importance of early intervention in cardiovascular disease, the use of ML models to provide early diagnoses should improve patient outcomes²⁰.

Future ML models for transgender patient care can be expanded for use in broader healthcare settings based on our findings. ML models can be improved with training on data collected from larger cohorts of transgender patients, such as the long-term cross-sex hormone therapy cohort studied by Asscheman et al.²¹. Being able to diagnose the presence of severe hypertension earlier leads to improved outcomes for patients, especially for patients at high-risk for cardiovascular disease complications residing in lower resource settings. This is why all our models were trained with basic clinical history data, preoperative lab tests, and postoperative outcomes²², which are all easily obtained as a standard of care in most healthcare settings^{22,23}. This allows our models to be able to make predictions on transgender patient populations residing in remote areas as a next step, where it is more likely for patients to have undiagnosed hypertension compared to their urban counterparts²⁴. In these lower resource settings, such as low- and middle-income countries, transgender individuals experience a high burden of adverse health and disease outcomes²⁵. Our models can be scaled into an eventual mobile app, where healthcare providers can manually enter in corresponding clinical and operative features for each patient in order to get an ML-driven hypertensive risk score for each patient, with models specifically tailored for their gender identity. In these settings, ML models could help physicians with less training and experience in treating transgender populations to develop earlier and more structured diagnoses. Integration of ML models into electronic health records and mobile health applications could contribute towards personalized, precision healthcare as a future direction for transgender individuals with significant barriers in access to healthcare^{26,27}.

Conclusion

We have been able to demonstrate that ML models trained on selectively transgender populations performs statistically similar (in the case of RF models) may perform statistically better or similar with certain datasets (in the case of LR models) to ML models developed on a combined cohort of predominantly cisgender patients over 300-fold larger for the prediction of hypertension in transgender populations. Overall, our findings serve as a starting point for future ML models made specifically for transgender patient populations, particularly in settings with lower volume patients to train ML models upon.

Data availability

The data that support the findings of this study are available from the ACS NSQIP but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request with written permission of the ACS NSQIP. Source Data File cannot be made available in accordance with the ACS NSQIP Data Use Agreement, as revealing source data to noncovered entities may render it possible, via deliberate technical engineering, to reveal identifying information of particular persons, in violation of the NSQIP data use agreement, rendering authors subject to penalties under statutes that apply to NSQIP data.

Code availability

Code file is publicly available²⁸.

Received: 21 August 2022; Accepted: 2 September 2024;

Published online: 30 September 2024

References

1. Arcelus, J. et al. Systematic review and meta-analysis of prevalence studies in transsexualism. *Eur. Psychiatry* **30**, 807–815 (2015).
2. Khera, R. et al. Use of machine learning models to predict death after acute myocardial infarction. *JAMA Cardiol.* **6**, 633–641 (2021).
3. Grant J. M., Motter L. A. & Tanis J. *Injustice at Every Turn: A Report of the National Transgender Discrimination Survey* (National Center for Transgender Equality, 2011).
4. Macapagal, K., Bhatia, R. & Greene, G. J. Differences in healthcare access, use, and experiences within a community sample of racially diverse lesbian, gay, bisexual, transgender, and questioning emerging adults. *LGBT Health* **3**, 434–442 (2016).
5. Bradford, J., Reisner, S. L., Honnold, J. A. & Xavier, J. Experiences of transgender-related discrimination and implications for health: results from the Virginia Transgender Health Initiative Study. *Am. J. Public Health* **103**, 1820–1829 (2013).
6. Tollinche, L. E., Van Rooyen, C., Afonso, A., Fischer, G. W. & Yeoh, C. B. Considerations for transgender patients perioperatively. *Anesthesiol. Clin.* **38**, 311–326 (2020).
7. Kenagy, G. P. Transgender health: findings from two needs assessment studies in Philadelphia. *Health Soc. Work* **30**, 19–26 (2005).
8. Alzahrani, T. et al. Cardiovascular disease risk factors and myocardial infarction in the transgender population. *Circ. Cardiovasc. Qual. Outcomes* **12**, e005597 (2019).
9. Wierckx, K. et al. Prevalence of cardiovascular disease and cancer during cross-sex hormone therapy in a large cohort of trans persons: a case-control study. *Eur. J. Endocrinol.* **169**, 471–478 (2013).
10. Carey, R. M. & Whelton, P. K. Prevention, detection, evaluation, and management of high blood pressure in adults: synopsis of the 2017 American College of Cardiology/American Heart Association Hypertension Guideline. *Ann. Intern. Med.* **168**, 351–358 (2018).
11. Roberts, T. K. et al. Interpreting laboratory results in transgender patients on hormone therapy. *Am. J. Med.* **127**, 159–162 (2014).
12. Ganguli, R. et al. Machine learning models to prognose 30-day mortality in postoperative disseminated cancer patients. *Surg. Oncol.* **44**, 101810 (2022).
13. Dietterich, T. G. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Comput.* **10**, 1895–1923 (1998).
14. Greenland, S. et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur. J. Epidemiol.* **31**, 337–350 (2016).
15. Caceres, B. A. et al. Assessing and addressing cardiovascular health in LGBTQ adults: a scientific statement from the American Heart Association. *Circulation* **142**, e321–e332 (2020).
16. Paranjape, K., Schinkel, M., Nannan Panday, R., Car, J. & Nanayakkara, P. Introducing artificial intelligence training in medical education. *JMIR Med. Educ.* **5**, e16048 (2019).
17. Ganguli, R., Lad, R., Lin, A. & Yu, X. Novel generative recurrent neural network framework to produce accurate, applicable, and deidentified synthetic medical data for patients with metastatic cancer. *JCO Clin. Cancer Inform.* **7**, e2200125 (2023).
18. Defreyne, J., Van de Bruaene, L. D. L., Rietzschel, E., Van Schuylenbergh, J. & T'Sjoen, G. G. R. Effects of gender-affirming hormones on lipid, metabolic, and cardiac surrogate blood markers in transgender persons. *Clin. Chem.* **65**, 119–134 (2019).
19. Streed, C. G. Jr et al. Cardiovascular disease among transgender adults receiving hormone therapy: a narrative review. *Ann. Intern. Med.* **167**, 256–267 (2017).
20. Hartley, C. J. et al. Cardiovascular health informatics: risk screening and intervention. *IEEE Trans. Inf. Technol. Biomed.* **16**, 791–794 (2012).

21. Asscheman, H. et al. A long-term follow-up study of mortality in transsexuals receiving treatment with cross-sex hormones. *Eur. J. Endocrinol.* **164**, 635–642 (2011).
22. Ganguli, R. & Wagner, S. M. 444 National and international validation of a prenatal machine learning algorithm to forecast cesarean section risk. *Am. J. Obstet. Gynecol.* **230**, S244–S245 (2024).
23. Ganguli, R. et al. Machine learning methods to predict presence of residual cancer following hysterectomy. *Sci. Rep.* **12**, 2738 (2022).
24. Zhou J. & Fang S. Association between undiagnosed hypertension and health factors among middle-aged and elderly Chinese population. *Int. J. Environ. Res. Public Health* **16**, 1214 (2019).
25. Reisner, S. L. et al. Global health burden and needs of transgender populations: a review. *Lancet* **388**, 412–436 (2016).
26. Shickel, B., Tighe, P. J., Bihorac, A. & Rashidi, P. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE J. Biomed. Health Inf.* **22**, 1589–1604 (2018).
27. Solomon, D. H. & Rudin, R. S. Digital health technologies: opportunities and challenges in rheumatology. *Nat. Rev. Rheumatol.* **16**, 525–535 (2020).
28. Ganguli, R. et al. Comparison of machine learning models for the prediction of hypertension in transgender patients undergoing gynecologic surgery. <https://doi.org/10.5281/zenodo.7844140> (2024).

Author contributions

R.G. developed study design, helped in design of machine learning models, and helped write all sections of manuscript. J.F. preprocessed data, conducted statistical testing between machine learning models, developed machine learning models and coding workflow, X.Y. helped generate figures and data tables and conducted statistical analysis, A.L. helped with statistical analysis and helped write introduction and discussion, S.W. added edits and comments throughout and informed study design, A.V. helped with study design, and aided table development.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s43856-024-00603-x>.

Correspondence and requests for materials should be addressed to Aditi Vichare.

Peer review information *Communications Medicine* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024