MCW Biostatistics Technical Report 67: Predicting Left Ventricular Hypertrophy (LVH) with Bayesian Additive Regression Trees (BART)

Rodney Sparapani

October 22, 2018

In this technical report, we provide R users with the ability to predict Left Ventricular Hypertrophy (LVH) using ECG and subject data with Bayesian Additive Regression Trees (BART). You will find these files in a ZIP archive online at `https://www.mcw.edu/-/media/MCW/Departments/Biostatistics/tr067.zip`. You will also need the BART R package which you can find online at `https://cran.r-project.org/package=BART`. The R program is `tr067.R`. Two prediction models are presented: BART-LVH and an ECG-only sub-model, BART-LVH ECG-only. BART-LVH is preferable since it has slightly better performance than BART-LVH ECG-only in R^2 and Sensitivity. From here on, we denote these models by the number of covariates used to predict the outcome. BART-LVH is BART26 utilizing 26 covariates with ECG data, gender, age, height and SBP/DBP; and BART-LVH ECG-only is BART22 using 22 covariates from ECG data and gender but excludes age, height and SBP/DBP.

The BART26 and BART22 models predict Normalized Left Ventricular Mass (LVM) expressed mathematically as $y_i \sim \mathrm{N}\big(f(\boldsymbol{x}_i),\ w_i^2\sigma^2\big)$ with predictors $\boldsymbol{x}_i$ where $f \overset{\text{prior}}{\sim}$ BART. BART is a Bayesian nonparametric machine learning method. The BART function, f, is an ensemble sum of trees. These BART models were fitted and the trees from these fits are encoded in R objects: `bart26.rds` contains the details of the BART26 fit and `bart22.rds`, the BART22 fit. As you will see exemplified in the R program `tr067.R`, the R `predict` function can be used to generate predictions at new data points which are provided by an R matrix or an R data frame, e.g., `yhat=predict(bart26, x)`.

Normalized LVM has a heavier right tail than left, therefore, we have fitted BART with two weight values, w_i: 0.905 for values < 1.037 and 1.105 for values ≥ 1.037. Based on body size, a mean value of the LVM was predicted using the following allometric formula previously derived from a MESA subpopulation of 822 men and women without LVH risk factors (`typewriter` font represents variable names found in the R data frames):

$$\texttt{exp\_lvm} = c(\text{height in m})^{0.561}(\text{weight in kg})^{0.608} \text{ where } c = \begin{cases} 6.82 & \text{females} \\ 8.17 & \text{males} \end{cases}.$$

We define Normalized LVM as: `lvh_lvm = olvedm1/exp_lvm` where `olvedm1` is the LVM in grams as determined by cardiac magnetic resonance imaging (cMRI) which is estimated via BART as $\hat{y}_i = \hat{f}(\boldsymbol{x}_i)$. The 95% quantile for `lvh_lvm` is 1.31 so values ≥ 1.31 are indicative of LVH. However, this cutoff is at the prediction level, y_i, rather than at the mean level, $\mathrm{E}[y_i] = f(\boldsymbol{x}_i)$; hence, 1.31 is too extreme. An appropriate cutoff for LVH by $f(\boldsymbol{x}_i)$ was determined to be 1.19 corresponding to the 93% quantile.

Table 1 is a list of the predictors and their units. To demonstrate the prediction process, we have supplied synthesized data based on the MESA study including the training, `mesa.train.rds`, and the validation, `mesa.test.rds`. So that these subjects can not be identified, a 5% error was added to the continuous variables and 5% of the discrete variables were altered. This process is presented as a prescription that the user can follow with their own data to make LVM and LVH predictions with the BART models.

The following results were generated by running the R program `tr067.R` which loads the data and computes statistical summaries and graphics files. In Table 2, we summarize R^2 between cMRI LVM and the BART predictions. In Table 3, we present contingency tables between cMRI LVH and the BART predictions. Furthermore, the generated Figures are presented as well where the red dots are females and the blue dots are males.

Table 1: Table of predictors used to estimate Normalized Left Ventricular Mass. Predictors denoted by an asterisk (*) are included with BART26 but excluded from BART22.

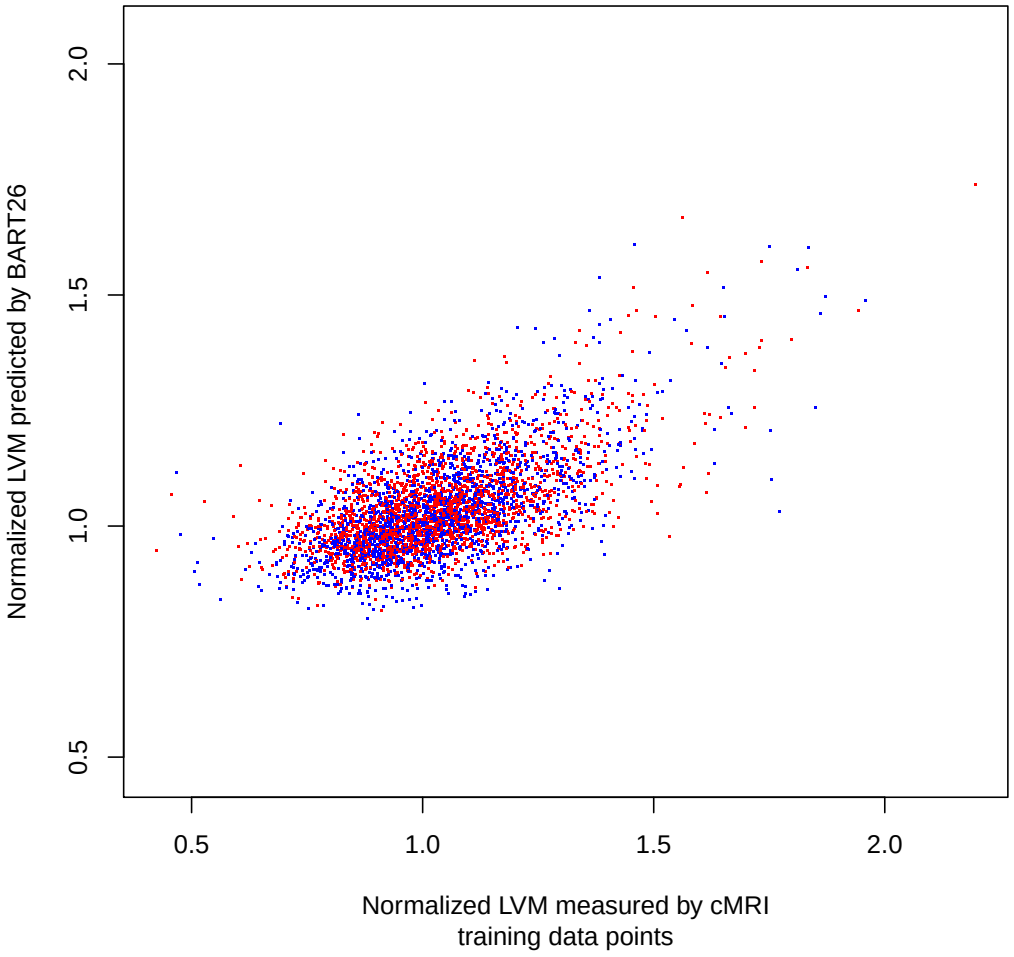
Predictors	Description
`samv21`	S amplitude lead V_2, μV
`heartrt1`	heart rate, bpm
`sbp1c`*	systolic blood pressure, mmHg
`ppdrv21`	P' duration lead V_2, ms
`tamv61`	T amplitude lead V_6, μV
`ramv41`	R amplitude lead V_4, μV
`pdurv41`	P duration lead V_2, ms
`qrsdur1`	QRS duration, ms
`stjv41`	STJ amplitude lead V_4, μV
`ramv11`	R amplitude lead V_1, μV
`age1c`*	age, years
`samv11`	S amplitude lead V_1, μV
`ramavl1`	R amplitude lead aVL, μV
`briv61`	intrinsicoid deflection R wave lead V_6
`ramv51`	R amplitude lead V_5, μV
`tamv11`	T amplitude lead V_1, μV
`gender1`	sex: 0=female, 1=male
`htcm1`*	height, cm
`stjv21`	STJ amplitude lead V_2, μV
`samv31`	S amplitude lead V_3, μV
`ramv31`	R amplitude lead V_3, μV
`tamv41`	T amplitude lead V_4, μV
`dbp1c`*	diastolic blood pressure, mmHg
`tami1`	T amplitude lead I, μV
`ramavf1`	R amplitude lead aVF, μV
`stjv11`	STJ amplitude lead V_1, μV

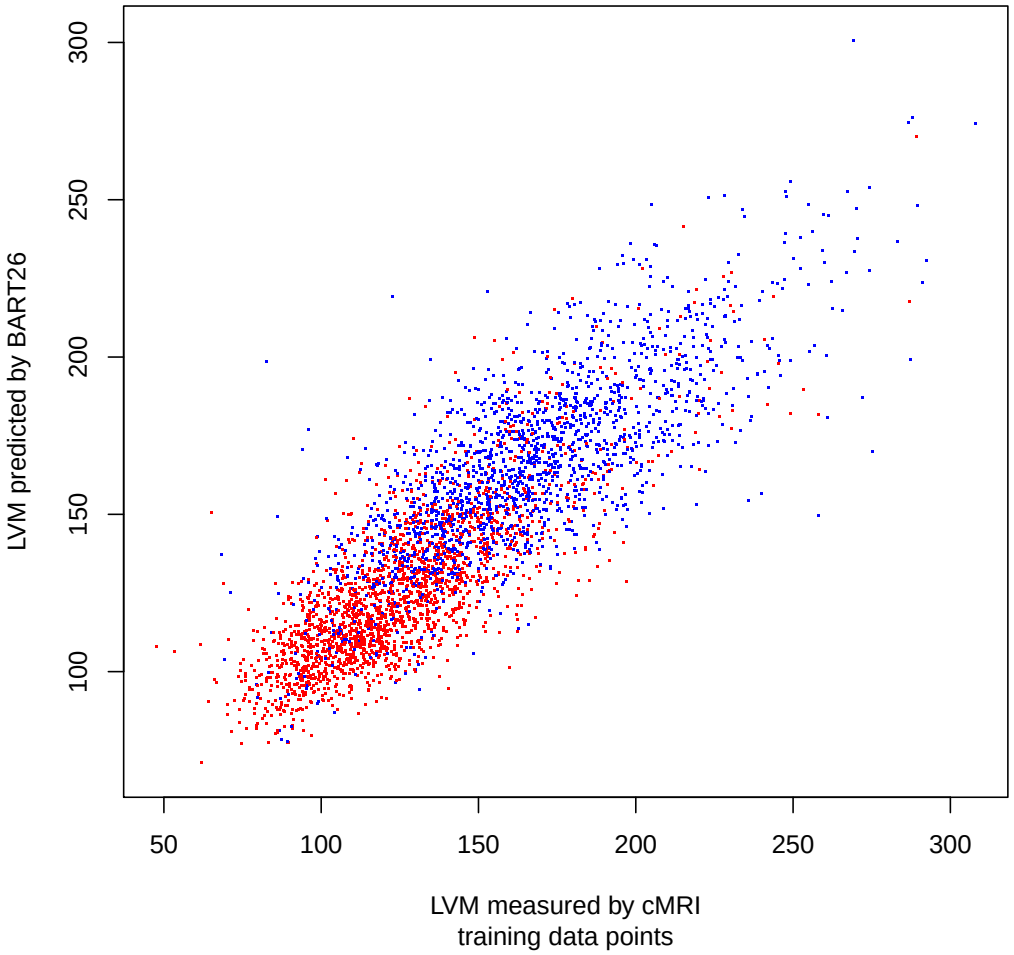
Table 2: R^2 between cMRI LVM and the BART predictions

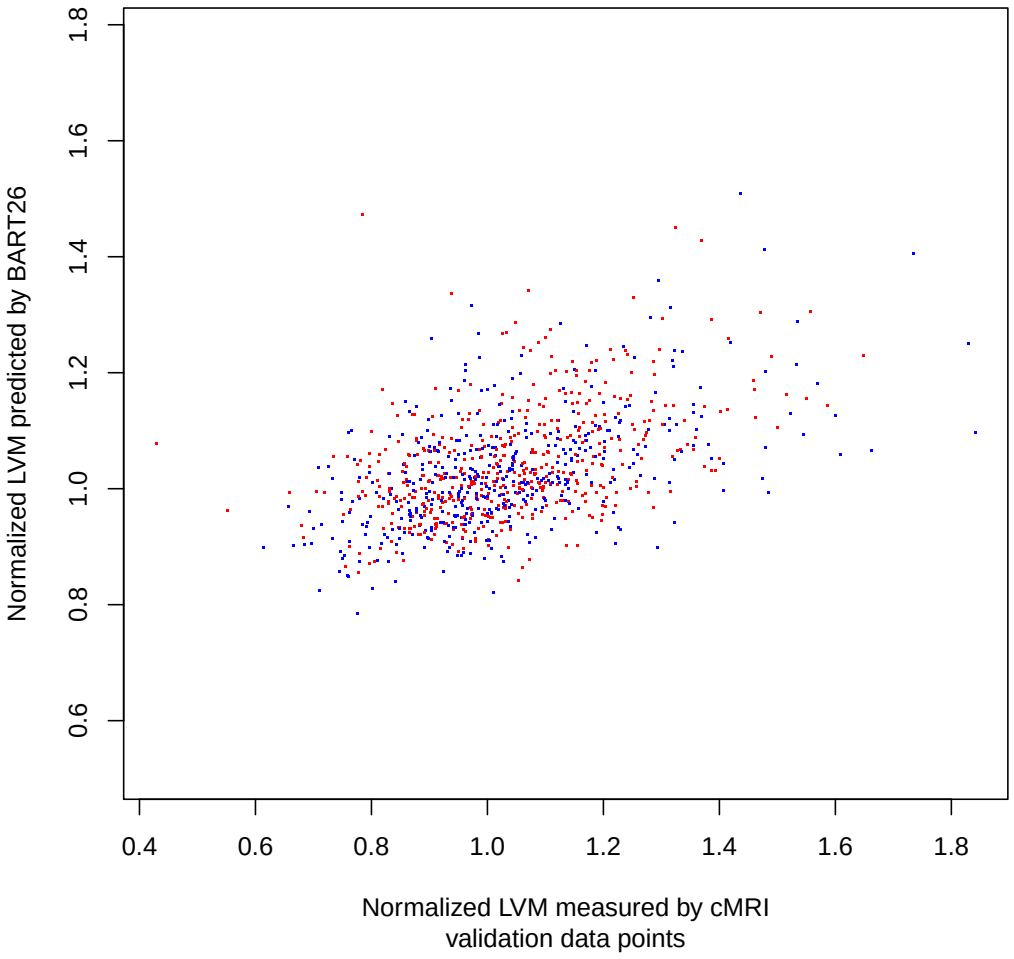
	BART26		BART22	
Set	Normalized LVM	LVM	Normalized LVM	LVM
Training	40.2%	72.9%	35.2%	70.9%
Validation	24.8%	65.3%	20.0%	62.9%

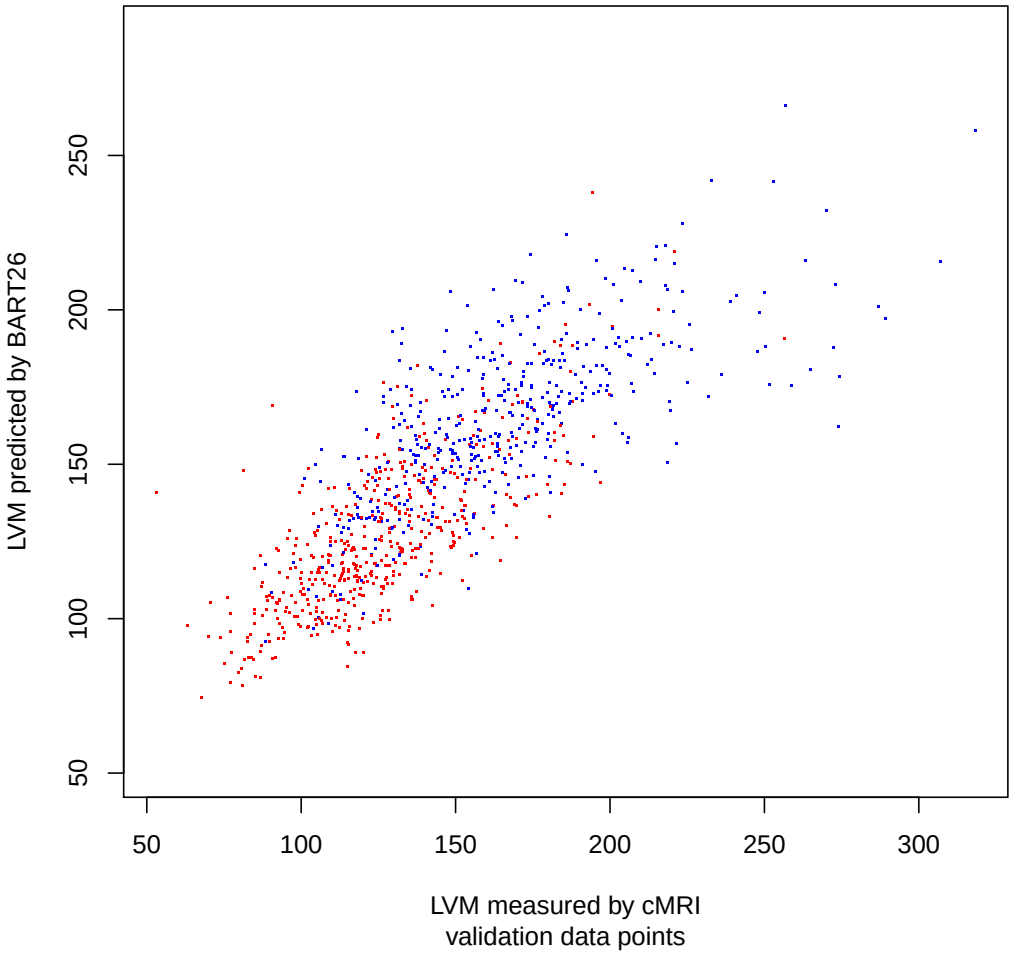
Table 3: Contingency table of cMRI LVH and the BART predictions

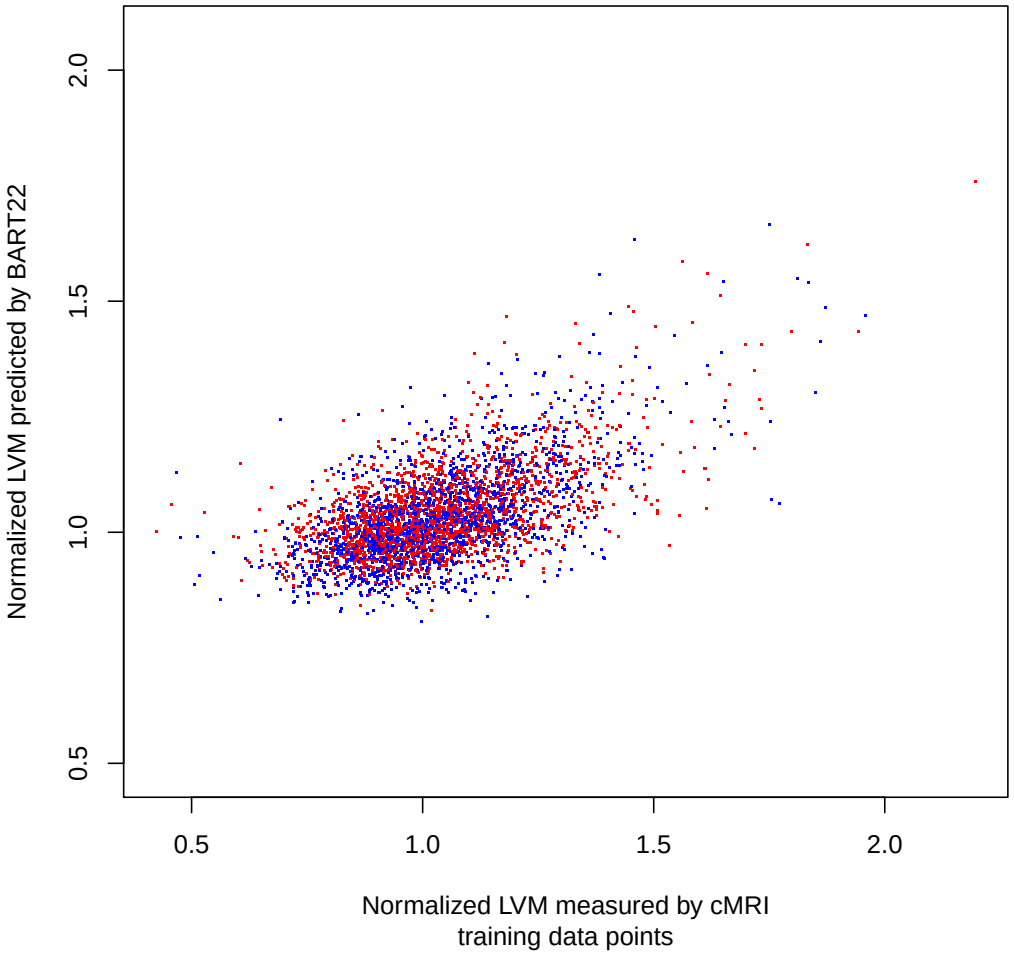
		BART26			BART22		
Set	cMRI	No LVH	LVH		No LVH	LVH	
Training	No LVH	3348	156	Spec.=95.5%	3363	141	Spec.=96.0%
	LVH	140	130	Sens.=48.1%	157	113	Sens.=41.9%
Validation	No LVH	821	48	Spec.=94.5%	837	32	Spec.=96.5%
	LVH	50	21	Sens.=29.6%	54	17	Sens.=23.9%

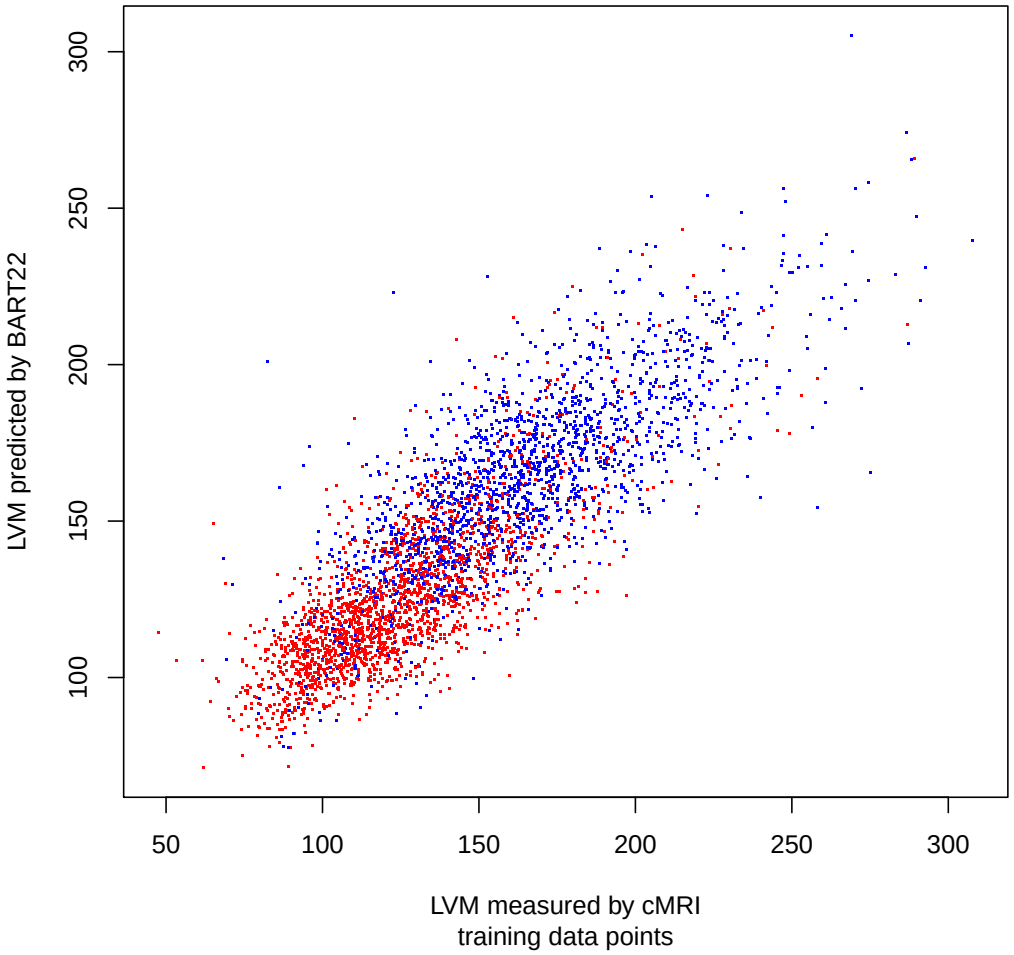

Normalized LVM predicted by BART26
0.5
1.0
1.5
2.0
0.5
1.0
1.5
2.0
Normalized LVM measured by cMRI
training data points

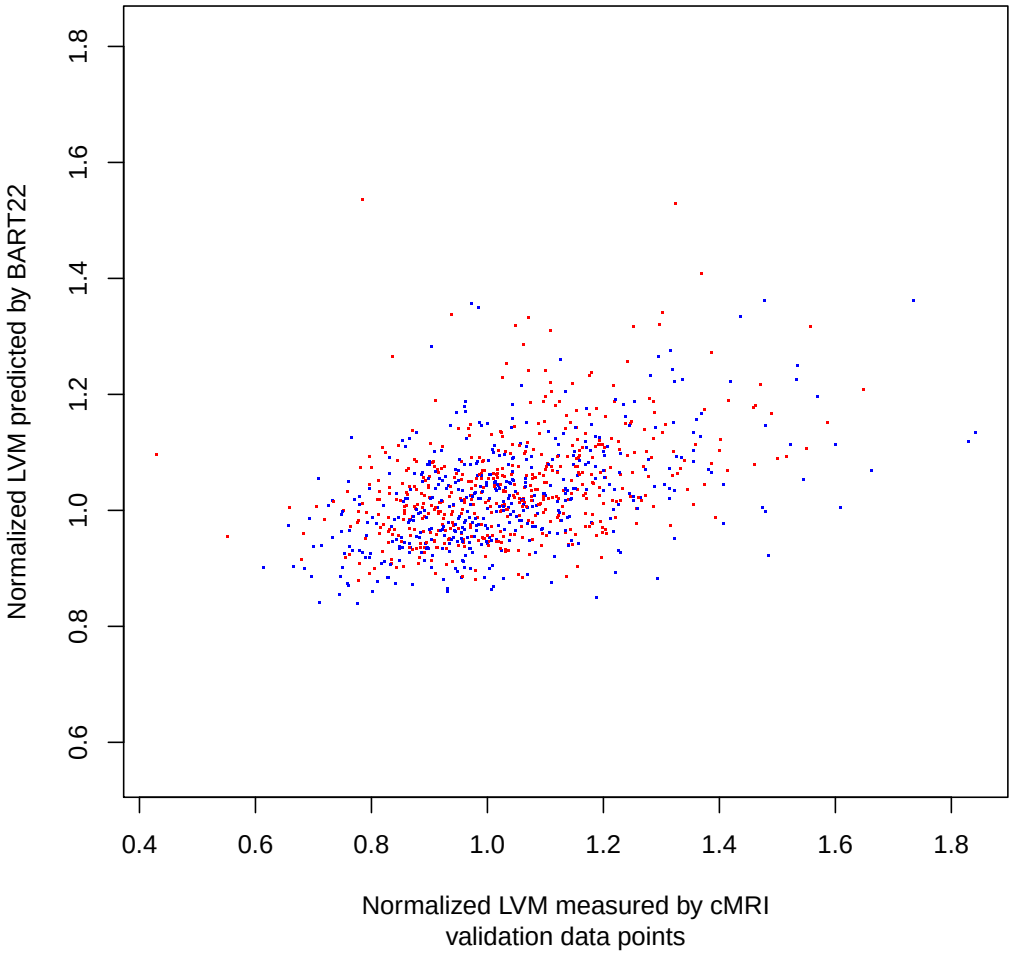

LVM predicted by BART26
100
150
200
250
300
50
100
150
200
250
300
LVM measured by cMRI
training data points

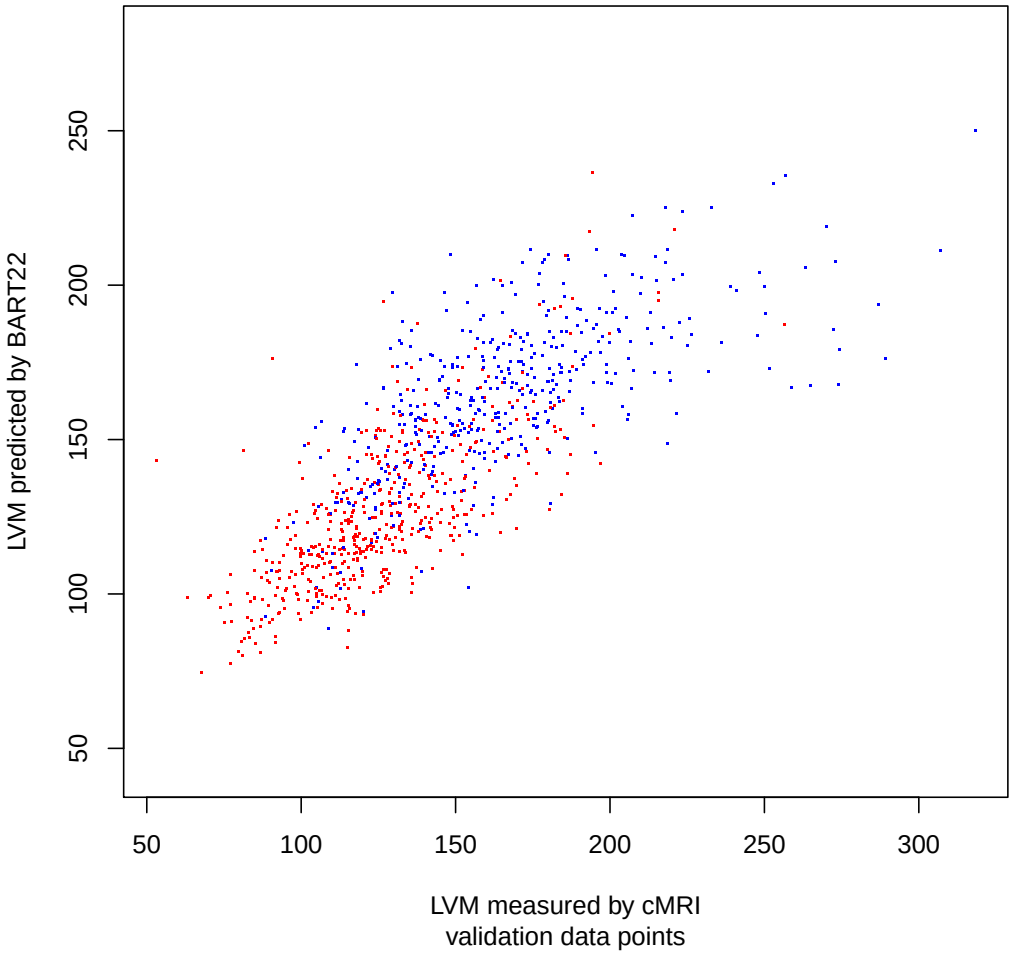

Normalized LVM predicted by BART26
0.6
0.8
1.0
1.2
1.4
1.6
1.8
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
Normalized LVM measured by cMRI
validation data points


LVM predicted by BART26
50
100
150
200
250
50
100
150
200
250
300
LVM measured by cMRI
validation data points


Normalized LVM predicted by BART22
0.5
1.0
1.5
2.0
0.5
1.0
1.5
2.0
Normalized LVM measured by cMRI
training data points


LVM predicted by BART22
100
150
200
250
300
50
100
150
200
250
300
LVM measured by cMRI
training data points


Normalized LVM predicted by BART22
0.6
0.8
1.0
1.2
1.4
1.6
1.8
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
Normalized LVM measured by cMRI
validation data points


LVM predicted by BART22
50
100
150
200
250
50
100
150
200
250
300
LVM measured by cMRI
validation data points