Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2020, 2020:22
https://doi.org/10.28919/cmbn/4584
ISSN: 2052-2541

FEATURE IMPORTANCE OF THE AORTIC ANATOMY ON ENDOVASCULAR ANEURYSM REPAIR (EVAR) USING BORUTA AND BAYESIAN MCMC

REZZY EKO CARAKA[1,2,5], NYITYASMONO TRI NUGROHO[3,4], SHAO-KUO TAI[1,*], RUNG-CHING CHEN[1,*], TONI TOHARUDIN[5], BENS PARDAMEAN[2,6]

[1]Department of Information Management, College of Informatics, Chaoyang University of Technology, Taichung City 41349, Taiwan (R.O.C.)

[2]Bioinformatics and Data Science Research Center, Bina Nusantara University, Jakarta, Indonesia

[3]Vascular and Endovascular Surgery Division, Department of Surgery, Cipto Mangunkusumo National Hospital, Faculty of Medicine University of Indonesia, Jakarta, Indonesia

[4]Department of Vascular and Endovascular Surgery, University Hospital of Muenster, Muenster, Germany

[5]Department of Statistics, Faculty of Mathematics and Natural Science, Padjadjaran University, Indonesia

[6]Computer Science Department, BINUS Graduate Program Master of Computer Science Bina Nusantara University, Jakarta, Indonesia

Abstract: A retrospective study of the abdominal aortic aneurysm (AAA) with EVAR treated patients. The third-party collected the data from twelve vascular centres in Indonesia during 2012-2017. Patient demographics and computed tomography data were evaluated with Osirix MD Software. During five years, we had 148 EVAR cases done using Endurant stent graft (Medtronic). In this paper, we perform Bayesian modelling and selection of feature selection by Boruta. Before performing the models, we will determine the selection of dependent variables start with the Age, Class, and Sex. It will get what is important to be dependent and independent. The difference between Bayesian and the

*Corresponding authors
E-mail addresses: sgdai@cyut.edu.tw (S. Tai), crching@cyut.edu.tw (R. Chen)
Received March 17, 2020

classical method is the introduction of prior information in the form of probability distributions. In addition, to determine the parameters using the Bayesian method obtained from the probability statement. Parameter estimation in Bayesian is no longer a point estimate but, on the contrary, is a statistical distribution. In other words, Bayesian states that a parameter is a variable that has a distribution. Bayesian has become a popular method in modern statistical analysis. Bayesian is applied to a broad spectrum in the scientific and research fields. Bayesian data analysis involves learning from data that uses probability models for many observations and some information to be studied. In other words, analysing statistical models are by combining prior knowledge about the model or parameters of the model. In a nutshell, the simulation results obtained modelling with Bayesian-ZIP-MCMC R^2 87.52 and Bayesian-Boruta R^2 88.28%.

Keywords: aorta; Boruta; feature selection; Bayesian; MCMC.

2010 AMS Subject Classification: 62C10, 62F15, 65C05, 97M60.

1. INTRODUCTION

Abdominal aortic aneurysm screening or broadly known as AAA screening is an examination procedure that aims to check for the abnormal widening of the aorta [1]. The examination is recommended since the beginning of the abdominal aortic aneurysm occurs because if it is too late, the size of the aorta can increasingly widen and burst. Moreover, the AAA case is predominantly an asymptomatic case, that is the reason an AAA screening is recommended on a high-risk patient with age >65 years and has a familial history of AAA [2]. Abdominal aortic aneurysm (AAA) is a condition of widening of the aorta in the abdomen, abnormally. The aorta is the main artery blood vessel that comes out of the heart to supply blood throughout the body.

The exact cause of an abdominal aortic aneurysm is not yet known, but several factors are thought to cause this condition. Among smoking habits, high blood pressure (hypertension)[3], trauma due to accidents[4], hereditary diseases [5], as well as infections and thickening of the arteries (atherosclerosis). The characteristics of the aorta in each individual are different, especially for each country [6] conduct studies on the characteristics of abdominal aortic aneurysm (AAA) in Korean patients. Moreover, [7] an epidemiology study for AAA patients is performed in the South-East Asian state of Sarawak in Borneo Island. This study shows that AAA in this Asian population is not uncommon, and the incidence is comparable to the Western world. It is important to analyse the characteristics of AAA in Indonesia [8], [9],[10].

Abdominal aortic aneurysm (AAA) are defined as pathologic dilatation of the abdominal aortic segment and due to this dilatation the aorta prone to expand and rupture that leads to a higher mortality rate [11]. Usually diameter of the abdominal aorta >30 mm is diagnosed with AAA[12].

The prevalence rates of AAA is 1.3 to 3.9% in male gender and 1.0-2.2% in the female gender, and AAA is predominantly in the age >65 years [13]. The risk factors for AAA development are male gender, hypertensive disorder, dyslipidemia, smoking history, familial history of aneurysm, greater height, coronary artery disease, cerebrovascular disease, atherosclerosis [14]. AAAs are degenerative disease and has tendency due to loss of the elastin and collagen in the aortic wall through matrix metalloproteinases (MMPs), macrophages, proteolytic enzymes, and many other inflammatory cells [15]. The total prevalence rates [16] per 100 000 persons/year in the age of 45-54 was 0, age of 55-64 was 6 (95% CI 2-12), age of 65-74 was 32 (95% CI 22-47), age of 75-84 was 67 (95% CI 47-92),), age ≥85 was 151 (95% CI 102-216).

The progressivity of AAAs is different between male and female gender. It is hypothesized that estrogen has a role in the modulation of the immune system and resulting the protective issue in women via reducing the macrophage MMP production. Besides, the expansion and rupture rate are greater in the female gender, presumably due to the beginning of aneurysm process is begun in smaller diameter of aorta in women. The AAA events is also higher in the patients with smoking history and reaches 222 (95% CI 129-355 per 100 000 persons/year). If indication of AAA treatment exists, endovascular treatment is the gold standard of the treatment, despite there is an option for open aortic repair (OAR). Endovascular treatment belongs to EVAR (endovascular aneurysm repair), TEVAR (thoracic endovascular aortic repair), ch-EVAR (Chimney EVAR), f-EVAR (fenestrated EVAR), and b-EVAR (branched EVAR) depends on the aortic segment which involves in aneurysmal sac.

The surveillance of successfulness of endovascular treatment can be described in the re-intervention rates, 30-days post-op mortality rates, and follow up mortality rates in years. In the EVAR procedure, from the meta-analysis of four studies (EVAR-1, DREAM, OVER, and ACE), the re-intervention rates are 5.1-8.5 in 100 000 persons/year, the 30-days post-op mortality rates are 0.2-1.8%, and mortality rates are 3.1 (2.1-3.4) to 6.0 (3.9-7.3) years. In the TEVAR procedure vs. open surgical repair for thoracic aorta aneurysm, mortality rates at 180 days are 10.2% (95% CI 9.0-11.4) vs. 23.8% (95% CI 21.4-26.1), respectively. The mortality rates at 5 years for TEVAR is 42.2% (95% CI 40.2-44.1) and in open surgical repair is 46.7% (95% CI 43.9-49.4) [17].

Feature selection is one of the most important techniques and is often used in data mining pre-processing [18], especially for knowledge discovery and discovery science. This technique reduces the number of features involved in determining a target class's value, reducing irrelevant, redundant,and data features that cause misunderstanding of the target class. The selection of the best features also provides high accuracy and a parsimony model will be obtained, which is easy to apply because not all variables need to be involved for the construction of the model. In addition, discussions with some experts in the field of health will also be done to improve insight so that it can produce a solution that is really appropriate [19].

The main task in feature selection is to determine which features are selected and used in the context of forecasting attribute attributes. Features are considered relevant if their value varies systematically with category membership. Feature selection is also useful for reducing data dimensions.

In addition, Boruta widely used in the field that requires superior features selection algorithms which is very important to create a ranking evaluation framework is based on the similarity between classifiers used as the core feature selection algorithm. Basically, the subset that will be obtained from the feature selection is a ranking order of all features selected by each algorithm.

In this paper, we analyse the proposed feature selection using Boruta and then Bayesian regression for the formation of models for the characteristics of the aortic for endovascular aneurysm repair in the Indonesian population in five years experiences.

Regression analysis is used to determine the effect of one or more independent variables on the independent variables[20],[21]. Currently, the use of regression analysis mostly uses a classic approach that does not include prior information [22],[23]. In this case, Bayes regression fills these weaknesses[24],[25]. The Bayes approach allows researchers to combine prior information and information obtained from a sample and then use it together to estimate posterior parameters. Moreover, this method is also useful to further understand the importance of stability in the feature selection algorithm so that it can then be used as a basis for decision making, which algorithm is suitable for one problem domain. The remainder of the paper is organized as follows. Section II provides the methodology. Section III provides five years of study experience. The feature selection and Bayesian regression are presented in Section IV. Finally, conclusions and future research directions are indicated in Section V.

2. METHODS

In this section, we introduce the feature selection Boruta and Bayesian Regression.

2.1 Feature Selection

Feature selection is the process of identifying and removing irrelevant and redundant variables. Feature selection is considered relevant if their values vary systematically with category membership. This process is essential in machine learning [26]. Many machine learning algorithms experience a decrease in accuracy when the number of variables is significant but not optimal [27],[28]. In addition, the number of variables slows down algorithm performance and takes up too many resources [28]. Boruta is one of the relevant feature selection algorithms, which can work with classification methods that measure the importance of variables. By default, Boruta uses Random Forest. The steps for Boruta's algorithm in finding all relevant variables are as follows.

1. Add shadow attributes, i.e., random data from copies of all attributes.
2. Run a random forest classifier and get *Z* scores.
3. Get the maximum *Z* scores among the shadow attributes (MZSA).
4. Mark the first attribute that has lower importance than MZSA as '*Rejected*' and deletes it from the system.
5. Mark the first attribute that has higher importance than MZSA as '*Confirmed*'.
6. Remove all shadow attributes.
7. Repeat the procedure until only the Confirmed attribute is left, or when reached the specified iteration limit.

2.2 Bayesian Regression

The posterior density function in the Bayes rule is proportional to the product of the prior distribution and the data distribution [24],[29],[30]. Following are the Bayes rules in establishing the posterior distribution of data:

$$p(\theta|\boldsymbol{y},\boldsymbol{X}) = \frac{p(\theta,\boldsymbol{y})}{p(\boldsymbol{y})} = \frac{p(\theta)p(\boldsymbol{y}|\theta)}{p(\boldsymbol{y})}$$

Where, $p(\boldsymbol{y}) = \sum_\theta p(\theta)p(y|\theta)$, for θ discrete and $p(\boldsymbol{y}) = \int p(\theta)p(y|\theta)\ d\theta$,for θ Continue. . The form equivalent to the above equation is obtained by issuing $p(\boldsymbol{y})$ because it does not depend on θ. Therefore, $p(\boldsymbol{y})$ considered constant. The form of the equation above becomes:

$$p(\theta|\boldsymbol{y},\boldsymbol{X}) \propto p(\theta)\, p(\boldsymbol{y}|\theta)$$

Which is a construction of$Posterior \propto Prior\ x\ Likelihood$. Furthermore, in Bayes regression, the Bayes estimator is assumed to use the expected value of the posterior distribution.

$$E(\theta|\boldsymbol{y},\boldsymbol{X}) = \int \theta\, f(\theta|\boldsymbol{y},\boldsymbol{X})\, d\theta = \int \frac{\theta\, f(\theta,\boldsymbol{y})}{f(\boldsymbol{y})}\, d\theta$$
$$= \frac{1}{f(\boldsymbol{y})}\int \theta\, f(\theta,\boldsymbol{y})\, d\theta$$
$$\propto \int \theta\, f(\theta,\boldsymbol{y})\, d\theta$$

The equation above is a form of $Posterior \propto Prior\ x\ Likelihood$. With the Bayes approach, regression parameters $\beta_0, \beta_1, \ldots, \beta_k$ considered a random variable [30]. Next, estimating these regression parameters takes into account prior information. There are two scenarios for priors, namely diffuse improper prior and prior informative conjugate provers for the regression parameter vector, $(\boldsymbol{\beta},\ \sigma^2)$. The regression model in this study assumes a normal distribution [31]. The prior noninformative distribution that is appropriate for normal heretical assumptions is uniform or uniform continuous on $(\boldsymbol{\beta}, \log \sigma)$. The prior is equivalent to diffuse improper prior.

$$\pi(\boldsymbol{\beta},\sigma^2) \propto \frac{1}{\sigma^2}$$

where the regression coefficient can take all real values, $-\infty < \beta_k < \infty$, for $k = 0,1, \ldots, K-1$, and variance $\sigma^2 > 0$. By carrying out the multiplication operation between likelihood and prior to produce a posterior of the model parameters as Posterior distribution of $\boldsymbol{\beta}$ if σ^2 known,

$$(\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{X},\sigma^2) \sim N(\widehat{\boldsymbol{\beta}}, (\boldsymbol{X}'\boldsymbol{X})^{-1}\sigma^2),$$

Where $\widehat{\boldsymbol{\beta}}$ is the OLS estimator, and $(\boldsymbol{X}'\boldsymbol{X})^{-1}\sigma^2$ is the covariance matrix of $\widehat{\boldsymbol{\beta}}$. Moreover, Posterior distribution of σ^2 can be written as follows:

Inverted χ^2:

$$(\sigma^2|\boldsymbol{y},\boldsymbol{X}) \sim Inv-\chi^2(n-K,\hat{\sigma}^2),$$

To find out the characteristics $\boldsymbol{\beta}$ that does not depend on σ^2 the marginal distribution must be obtained from $\boldsymbol{\beta}$. In practice, $\sigma^2$is an unknown parameter. This can be demonstrated by uniting the combined posterior distribution.

$$p(\boldsymbol{\beta},\sigma^2|\boldsymbol{y},\boldsymbol{X}) = p(\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{X},\sigma^2)\, p(\sigma^2|\boldsymbol{y},\boldsymbol{X})$$

by integrating the function against σ^2, obtained unconditionally posterior distribution of $\boldsymbol{\beta}$, and a multivariate or Student-t distribution $(\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{X}) \sim \boldsymbol{t}$. The multivariate Student-t distribution with degree of freedom, $v = n - K$, with the density function as follows:

$$p(\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{X}) = \frac{v^{\frac{v}{2}}\,\Gamma\left(\frac{v+n}{2}\right)\left|\frac{\boldsymbol{X}'\boldsymbol{X}}{\hat{\sigma}^2}\right|^{1/2}}{\pi^{\frac{n}{2}}\Gamma(\frac{v}{2})}(v + (\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}_{\boldsymbol{OLS}})'\frac{\boldsymbol{X}'\boldsymbol{X}}{\hat{\sigma}^2}(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}_{\boldsymbol{OLS}}))^{-(n+v)/2}$$

It can be seen that issued σ^2 makes distribution $\boldsymbol{\beta}$ *heavy-tailed* as it should reflect uncertainty σ^2. Although, the mean vector of $\boldsymbol{\beta}$ unchanged, the variance increases with its *v/(v-2):*

$$\sum\nolimits_{\boldsymbol{\beta}} = var(\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{X}) \sum\nolimits_{\boldsymbol{\beta}} = \hat{\sigma}^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\frac{v}{v-2},$$

Where $v = n - K$ is the degree of freedom of parameter distribution of the multivariate Student-*t* distribution. As a conclusion in the discussion of the posterior one regression coefficient, β_k, under the diffuse improper prior scenario, it can be shown that β_k *standardis*ed has a Student-*t* distribution with degrees of freedom $n - K$. That is because the marginal posterior distribution is as follows,

$$\frac{\beta_k - \widehat{\beta_k}}{(h_{k,k})^{1/2}}|y, X \sim t_{n-K}$$

Where, $h_{k,k}$ is a diagonal element to-*k* from $\sum_{\boldsymbol{\beta}}$ and $\hat{\beta}_k$ is the OLS estimator of β_k**.** Marginal posterior distribution of σ^2 can be written as follows:

$$p(\sigma^2|\boldsymbol{y},\boldsymbol{X}) = \frac{p(\boldsymbol{\beta},\sigma^2\,|\mathbf{y})}{p(\boldsymbol{\beta}|\sigma^2\,,\mathbf{y})}$$

This form is an Inverse form -χ^2

$$(\sigma^2|\boldsymbol{y}) \sim \text{Inv} - \chi^2(n-K,\hat{\sigma}^2),$$

Where

$$\hat{\sigma}^2 = \frac{1}{(n-K)}\left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\boldsymbol{OLS}}\right)'\left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\boldsymbol{OLS}}\right).$$

3. FIVE YEARS OF EXPERIENCES

Indonesia has had an emerging endovascular treatment during this last half-decade, especially for aortic. With only thirty-nine vascular surgeons, we have for 261.1 million populations; we had treated many cases. Despite some cases also were treated by other specialities. During these five years, we collected data from endovascular aneurysm repair (EVAR) treatment.

This is a retrospective study of the abdominal aortic aneurysm (AAA) patients that were treated by EVAR. The third-party collected the data from twelve vascular centres in Indonesia during 2012-2017. Patient demographics and computed tomography data were evaluated with Osirix MD software.

During 5 years, we had 148 EVAR cases done using Endurant stent graft (Medtronic). Due to

the lack of complete data, we excluded 54 cases. Figure 1 represents the evaluation of 94 patients revealed 76.6% was male with a mean age of 67.8 ± 9.9 years old, and 17% of them were treated both for an abdominal and thoracic aneurysm. All of them had a fusiform with 14.9% dissected, which more than half were infra-renal (72.3%) and had >50% intra-mural thrombus in the aneurysmal sac (57.4%).

Mean values of proximal neck length, diameter, and angle were 26.8 ± 16.7 mm, 21.0 ± 5.9 mm, 41.3 ± 19.9 degrees, respectively, with four patients had less than 10 mm in neck length. The maximum diameter of the aneurysmal sac was 57.0 ± 22.1 mm on average. Most aneurysmal sacs didn't have common iliac artery involvement; only 19.1% occurred in both arteries. Eleven percent of patients had external iliac artery circumferential thrombus >50%, which occurred either on the left or right side. Sealing zone in both CIA less than 15 mm was none. Figure 1 describes the aortic blood vessels of patients who have AAA (abdominal aortic aneurysm), or blood vessel dilation. Dilated blood vessels will be very susceptible to rupture and cause the patient to die. This AAA disease can be treated with surgery. There are two ways to operate, open surgery (dissected stomach) or endovascular. Endovascular means to stent-graft into the aorta through the thight

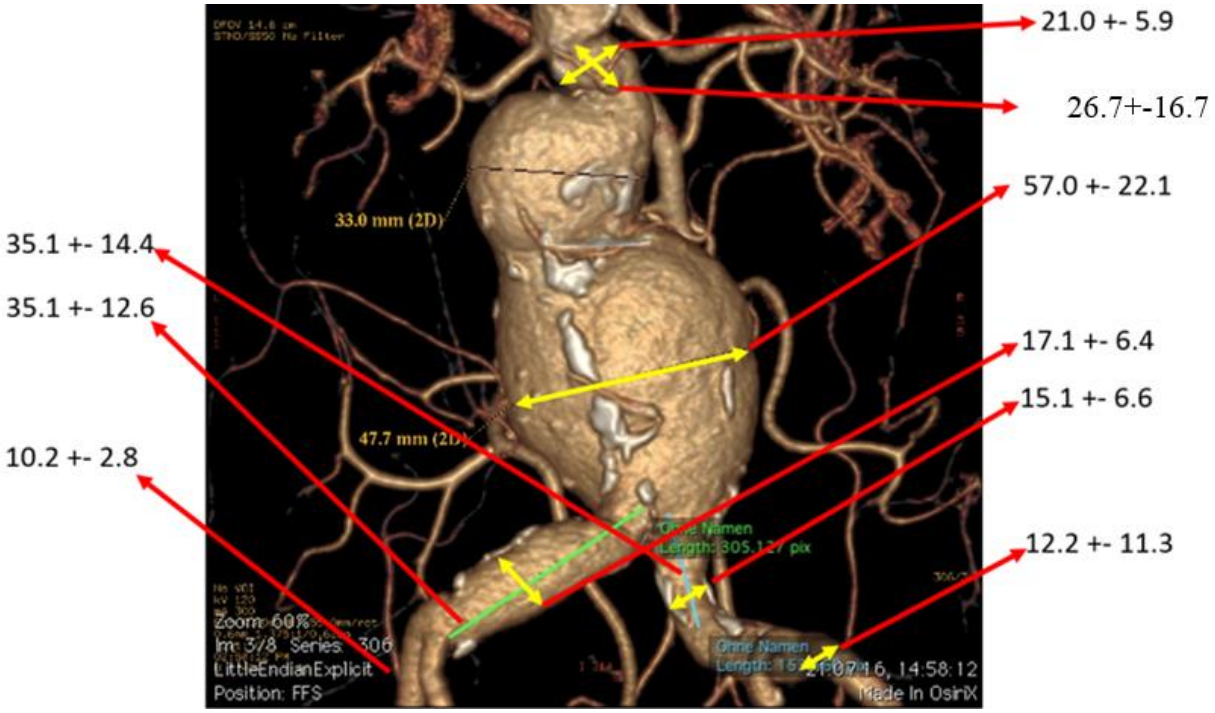

Figure 1. Anatomy of an abdominal aortic aneurysm (AAA) in a reconstructive anterior view AAA Anatomy. The numbers represent the measurement of diameter or length of its correspondence measure of anatomical site in millimeter (mean ± SD mm)

4. ANALYSIS

The dataset used in this paper can be seen in Table 1. In the selection feature, the first step is to find out how if we place 3 Dependent variables for feature selection. First simulation by choosing Class (1 Evar, 2 Tevar, and 3 Both). Then, Age and lastly Sex (1 Male, 2 Female).

Table 1. Dataset

Variables	
Age	Continuous
Sex	Male (1), Female (2)
Class	EVAR (1), TEVAR (2), BOTH (3)
Aortic Type	Fusiform (1), Saccular (2)
Renal Involvement	Supra Renal (1), Juxta Renal (2), Infra Renal (3)
Dis	Dissection (1), Non Dissection (2)
Proximal neck length	mm
Proximal neck diameter	mm
Proximal neck angle	degree
Intramural thrombus >50%	Yes (1), No (2)
Max. aneurysmal sac	diameter (mm)
Right CIA diameter	diameter (mm)
Left CIA diameter	diameter (mm)
Right EIA	diameter (mm)
Left EIA diameter	diameter (mm)
CIA involvement	Left (1), Right (2), Both (3), None (4)
CIA right sealing zone length	diameter (mm)
CIA left sealing zone length	diameter (mm)
CIA sealing zone length <15 mm in both side	Yes (1), No (2)
CIA thrombus >50%,	Left (1), Right (2), Both (3), None (4)
EIA thrombus >50%,	Left (1), Right (2), Both (3), None (4)

In the first simulation, Boruta performed 99 iterations in 2.633223 secs. With four attributes confirmed important: “Renal Inv,” Age, EIAthrombus, ProximalNA, and three tentative attributes left: Dissection, ProximalND, ProximalNL are fully explained in Figure 2.

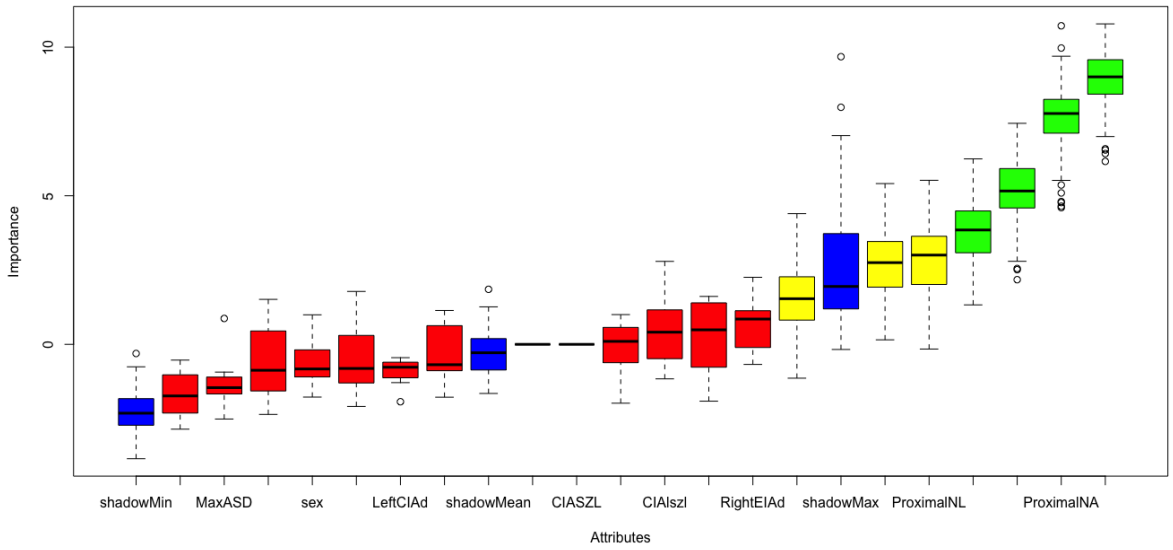

Figure 2. Feature Selection Based on Class

Then, a second simulation is performed to perform feature selection when considering age. Figure 3 explains that there is only one important attribute, which is renal inv and 18 attributes confirmed unimportant.

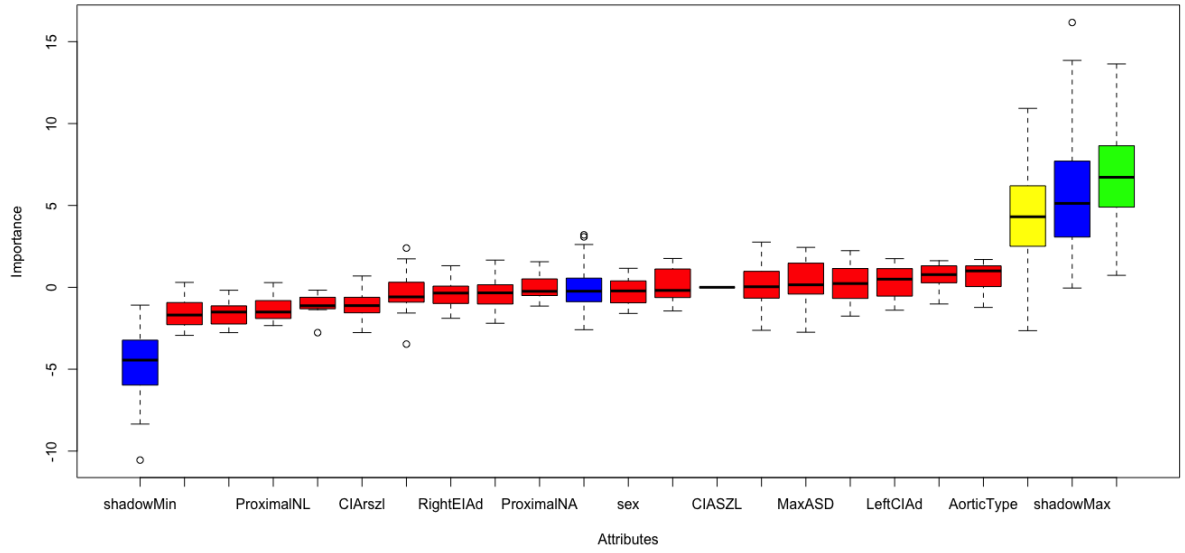

Figure 3. Feature Selection Based on Age

Finally, the selection of the best features of Boruta performed 99 iterations in 2.683058 secs using the Sex variable (1 Male, 2 Female). Figure 4 explains that 2 confirmed important attributes: LeftCIAd, LeftEIAd, and 17 confirmed unimportant attributes. We have then obtained one tentative attribute left: RightEIAd.

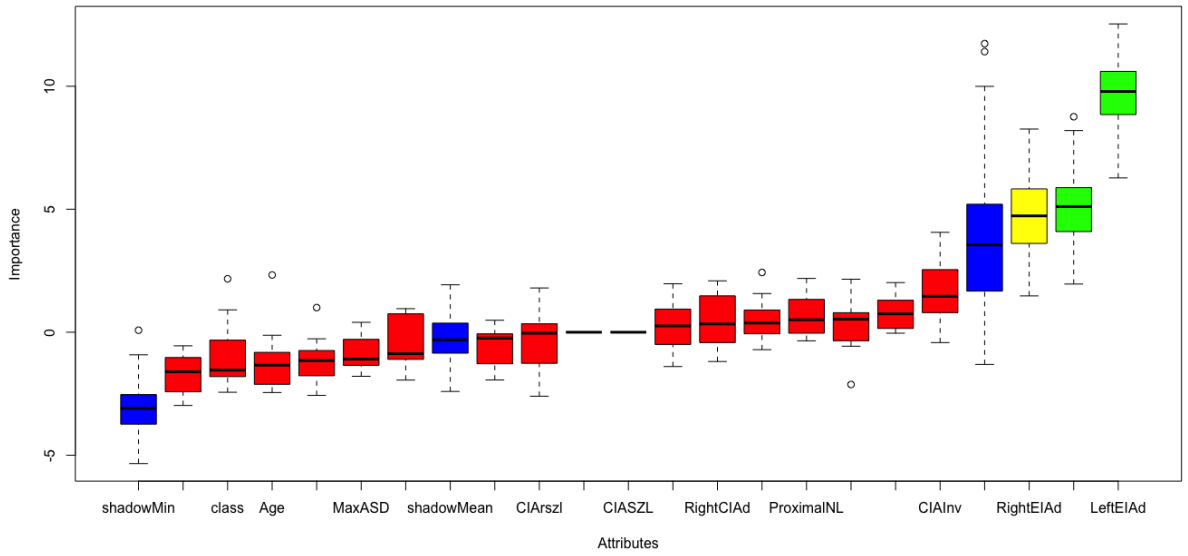

Figure 4. Feature Selection Based on Sex

Based on the three simulations using feature selection, it was found that the good use of variables is Class (1 Evar, 2 Tevar, and 3 Both) because this simulation has a lot of important variable information compared only using age and sex. Furthermore, Bayesian regression will be built with MCMC and prior ZS-Null.

Bayesian and MCMC methods are complicated to separate[32]. For the posterior model of a Bayesian that is very complicated, it requires a challenging integration process in determining the marginal posterior parameters. Then, we need an alternative solution with a numerical approach. Another way to perform MCMC is by Gibbs Sampler. Gibbs Sampler facilitates the numerical solution to generate random data $\theta = (\pi, \sigma^2, \beta)$. It contains as a random vector with a precise distribution and has an estimated value of $f(\overline{\Theta})$ of function is related in $f(\Theta)$ [33]. Since all parameters $= (\pi, \sigma^2, \beta)$in bayesian are treated as variables, then the inference in the estimation will be based on:

$$f(\overline{\Theta}) = E(f(\Theta)|y^n)$$

$$= \int f(\theta) f\mu(\theta)$$

$$= \int f(\theta) p(\theta|y^n) d\,(\theta)$$

Gibbs Sampler is a very efficient generator, so it is often used as a generating of random variables in data analysis using the MCMC method [34],[35]. Gibbs Sampler method can be defined as a simulation technique to generate random variables from a particular distribution indirectly, without having to calculate the density function of a data distribution assuming several random variables $\beta_0, \beta_1, \beta_2, \dots, \beta_p, \sigma_d$ in a mixture distribution.

$$f\beta_0, \beta_1, \beta_2, \dots, \beta_p, \sigma_d$$

The mixture of the function distribution considered known as bellows.

$$f(\beta_0|\beta_1, \beta_2, \dots, \sigma_d)$$
$$f(\beta_1|\beta_0, \beta_1, \dots, \sigma_d)$$
$$\vdots$$
$$f(\sigma_d|\beta_0, \beta_1, \dots, \sigma_p)$$

To get the distribution characteristics of *x* or often said to be the marginal distribution of $f(x)$, multiple process integrals must be carried out as many random variables are left in their combined distribution.

$$f(x) = \int \dots \int \int f(\beta_0, \beta_1, \beta_2, \dots, \beta_p, \sigma_d) d\beta_0 d\beta_1 d\beta_2 \dots d\sigma_p$$

Integration in the above equation will be very difficult or even impossible if the combined function is very complicated. To solve this problem, we can use the Gibbs Sampler method with the condition that the distribution of each variable contained therein is known. With this method, without having to calculate and know what the shape of the marginal function is, the characteristics of each marginal random variable in the combined function will be known. The way to approach the marginal distribution characteristics with Gibbs Sampler through numerical data generation with Monte Carlo simulations. Otherwise, the conditional distribution of random variables is being studied for all remaining random variables under the combined distribution function. Then to find out the characteristics of the marginal distribution $f(\beta_0)$, the Gibbs Sampler with the Monte Carlo method will generate some data *X* that has distribution $f(\beta_0|\beta_1, \beta_2, \dots, \sigma_d)$ and its marginal distribution characteristics. It is estimated based on data from the Monte Carlo simulation. This explains that with huge amounts of data, the values obtained based on these data will reflect the condition of a population. Data generation of each random variable in the combined density function using the Gibbs Sampler is as follows.

Table 2. Pseudocode Gibbs Sampler MCMC

1. Set the initial value $\theta^0 = \beta_0^0, \beta_1^0, \beta_2^0, \ldots, \sigma_d^0$

2. Generate the data

Step 1. β_0^1 from $f(\beta_0|\beta_1^0, \beta_2^0, \ldots, \sigma_d^0)$

Step 2. β_1^1 from $f(\beta_1|\beta_0^0, \beta_1^0, \ldots, \sigma_d^0)$

Step 3. β_2^1 from $f(\beta_2|\beta_0^0, \beta_1^0, \ldots, \sigma_d^0)$

$\vdots$

Step *d*. σ_d^1 from $f(\sigma_d|\beta_1^1, \beta_1^1, \ldots, \beta_p^1)$

The generation of data values obtained in the first iteration is

$$\theta^1 = \beta_0^1, \beta_1^1, \beta_2^1, \ldots, \sigma_d^1$$

3. Use the result of the generation in the first iteration as the initial value in the second iteration. Then generate the following stages:

Step 1. β_0^2 from $f(\beta_0|\beta_1^1, \beta_2^1, \ldots, \sigma_d^1)$

Step 2. β_1^0 from $f(\beta_1|\beta_1^1, \beta_2^1, \ldots, \sigma_d^1)$

Step 3. β_2^0 from $f(\beta_2|\beta_1^1, \beta_2^1, \ldots, \sigma_d^1)$

$\vdots$

Step *d*. σ_d^1 from $f(\sigma_d|\beta_1^1, \beta_2^1, \ldots, \beta_p^1)$

To obtain the data generation value in the second iteration,

$$\theta^2 = \beta_0^2, \beta_1^2, \beta_2^2, \ldots, \sigma_d^2$$

Then, do it until the *N* iteration. Use the result of generation in the *N-1* iteration as the initial value in the *N* iteration. To obtain the value of generating data in the *N* iteration.

$$\theta^N = \beta_0^N, \beta_1^N, \beta_2^N, \ldots, \sigma_d^N$$

4. Based on *N* data that has been generated, then estimated the characteristics of each marginal density function

$$f(\beta_1), f(\beta_2), \ldots, f(\sigma_d)$$

As mentioned in the previous section, we can generate random data in each iteration. Other words, the data generated is very dependent on the random value that has been generated by one iteration before. Moreover,the data generated does not depend on random values in the second, third, until *N* iteration previously.The model that can be produced explained below:

As mentioned in the previous section, we can generate random data in each iteration. Other words, the data generated is very dependent on the random value that has been generated by one iteration

before. Moreover, the data generated does not depend on random values in the second, third, until *N* iteration previously. The model that can be produced, is explained below:

$$y_i = \beta_0 + \beta_1\left(x_{1,i} - \overline{x_1}\right) + \beta_1\left(x_{2,i} - \overline{x_2}\right) + \cdots + \beta_{19}\left(x_{19,i} - \overline{x_{19}}\right) + \in_i, 1 \leq i \leq n$$

Then, we compared the accuracy of the models by performing all the variables we had and also compared to the feature selection with Boruta. Table 3 explains that the best model used is model 3, with an R^2 87.52% compared to other models.

Table 3. Summary Model All Variable

	P(B != 0 | Y)	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept	1	1	1	1	1	1
Age	0.3642689	0	0	1	0	0
Sex	0.2055305	0	0	0	0	0
AorticType	0.2135014	0	0	0	0	0
`Renal Inv`	0.9957249	1	1	1	1	1
Dissection	0.6874533	1	0	1	1	1
ProximalNL	0.2022366	0	0	0	0	0
ProximalND	0.2273428	0	0	0	0	0
ProximalNA	0.7469482	1	1	1	1	1
IntramuralThrombus	0.3422043	0	0	0	0	0
MaxASD	0.2162348	0	0	0	0	0
RightCIAd	0.3181459	0	0	0	0	1
LeftCIAd	0.2496117	0	0	0	0	0
RightEIAd	0.3057865	0	0	0	0	0
LeftEIAd	0.2168787	0	0	0	0	0
CIAInv	0.2322077	0	0	0	0	0
CIArszl	0.2346521	0	0	0	0	0
CIAlszl	0.2914526	0	0	0	1	0
CIAthrombus	0.2119417	0	0	0	0	0
EIAthrombus	0.9997458	1	1	1	1	1
BF	-	1	0.4867886	0.3602048	0.3577225	0.316961
PostProbs	-	0.011	0.0056	0.0041	0.004	0.0036
R^2	-	0.8652	0.8238	**0.8752**	0.8751	0.8732
dim	-	5	4	6	6	6
logmarg	-	15.59144	14.8715114	14.570354	14.5634389	14.44246

We are likewise presenting a shortened beta-binomial dissemination on the model size. This condition allows all models with more than $\frac{n}{2}$ coefficients a zero likelihood from the earlier which confines the number of potential anomalies. Then the Bayesian inference model selection is used with the Bayesian Model Averaging (BMA); at this stage, we can have predictions and a combination of estimates from the resulting model. In Figure 4, it can also be seen that the linear line (red line) can fit the residual (dot), which can be assumed that our model is good enough to be used as a prediction.

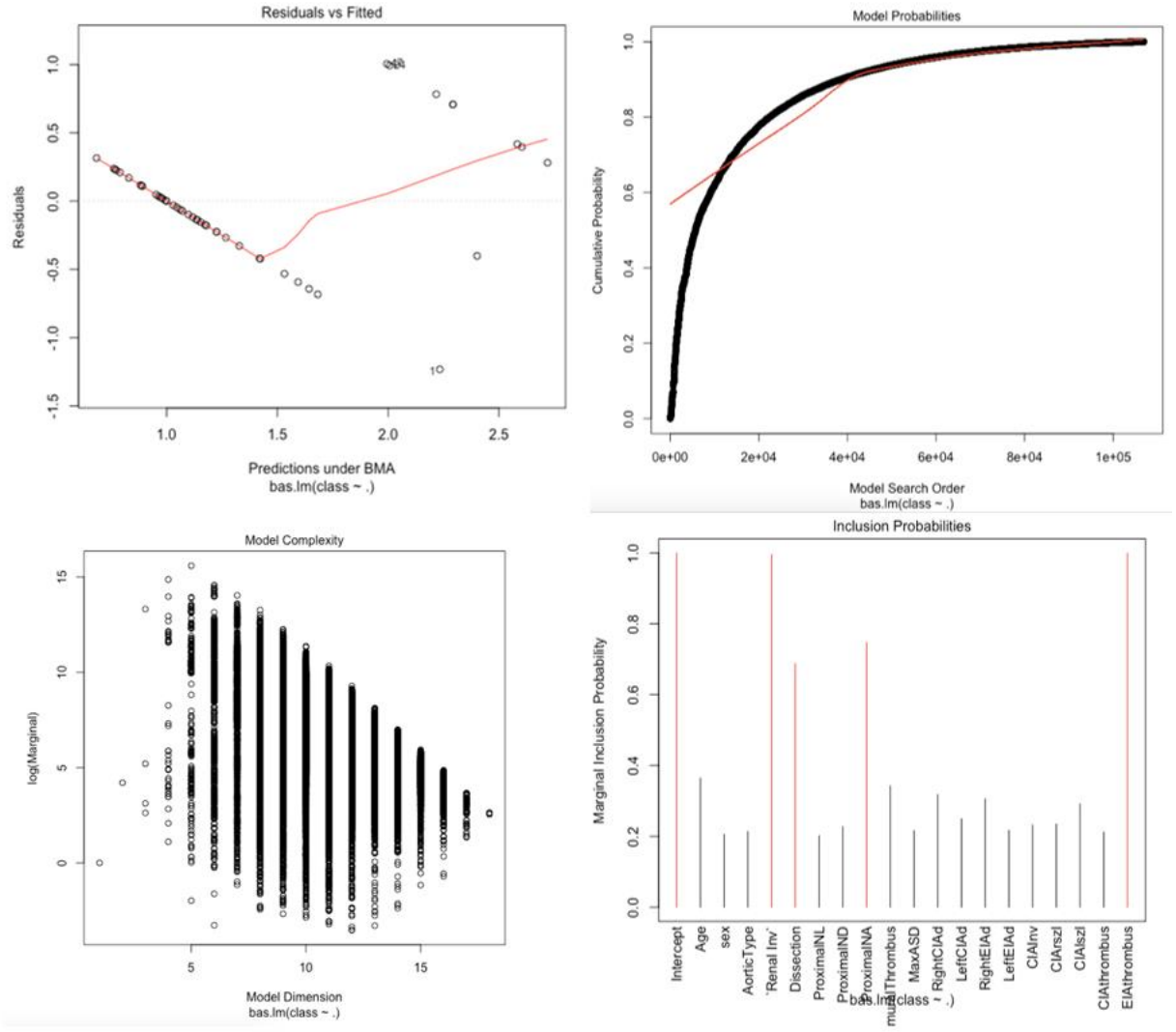

Figure 5. Model Summary Bayesian All Variable

Table 5. Summary Model Bayes-Boruta Selection

	P(B != 0 | Y)	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept	1	1	1	1	1	1
Age	0.2875	0	1	0	1	1
`Renal Inv`	0.984375	1	1	1	1	0
ProximalNA	0.815625	1	1	0	0	0
EIAthrombus	0.9875	1	1	1	1	1
BF	-	1	0.4642043	0.2093472	0.1305416	0.00010808
PostProbs	-	0.5701	0.2368	0.134	0.0312	0.0093
R^2	-	0.8238	0.8373	0.8601	0.8828	0.8547
dim	-	4	5	3	4	3
logmarg	-	14.59773	13.8303036	13.033973	12.5616705	5.46505303

Bayesian Model Averaging (BMA) is an application of Bayesian inference to the problems of model selection, combined estimation, and the prediction produces a straightforward model choice criteria and less risky predictions. The results obtained when using the Boruta selection feature are better with an R^2 value of 88.28% for model 4. It should be noted that when using all variables and also, the variable reduction does not provide significant results to improve the accuracy of the model. Visually can be seen in Figure 6, and Figure 7, respectively.

The ensemble predict results tend to be under dispersive or overdispersion (the estimated value of the aortic measure is centred on a value with low or high variance). This condition can cause the forecast results to be less reliable because under dispersive indicates relatively narrow forecast intervals, while overdispersion indicates relatively more full forecast intervals. These problems can be handled by the calibration of ensemble forecasts to produce a more reliable and sharper predictive density function (PDF).

If the calibrated ensemble interval forecasts have a standard deviation of proportional size, the results of the forecast reach an equidispersive state. This method is based on the concept of dynamic models. The advantage of this method lies in the ability to integrate several models of ensemble forecasts and correct forecast biases so that the mean and variance are closer to reality so that the calibrated ensemble forecasts can approach the actual observational values. By utilizing information from several ensemble models, the forecasts are produced does not only depend on one model. In addition, the BMA parameter estimator is not stagnant and always changes following dynamic weather and atmosphere information. BMA predictive robust function (PDF) is a linear combination of several forecast models where each model has a different contribution or weight to the predictive PDF formation. In BMA, weight determination is based on the average concept

(averaging model). The magnitude of the weight depends on the ability of the model to forecast the weather component where the weight will be higher if the forecast results get closer to the real observation value. From a Bayesian point of view, these weights are the posterior probabilities of each member/ensemble model.

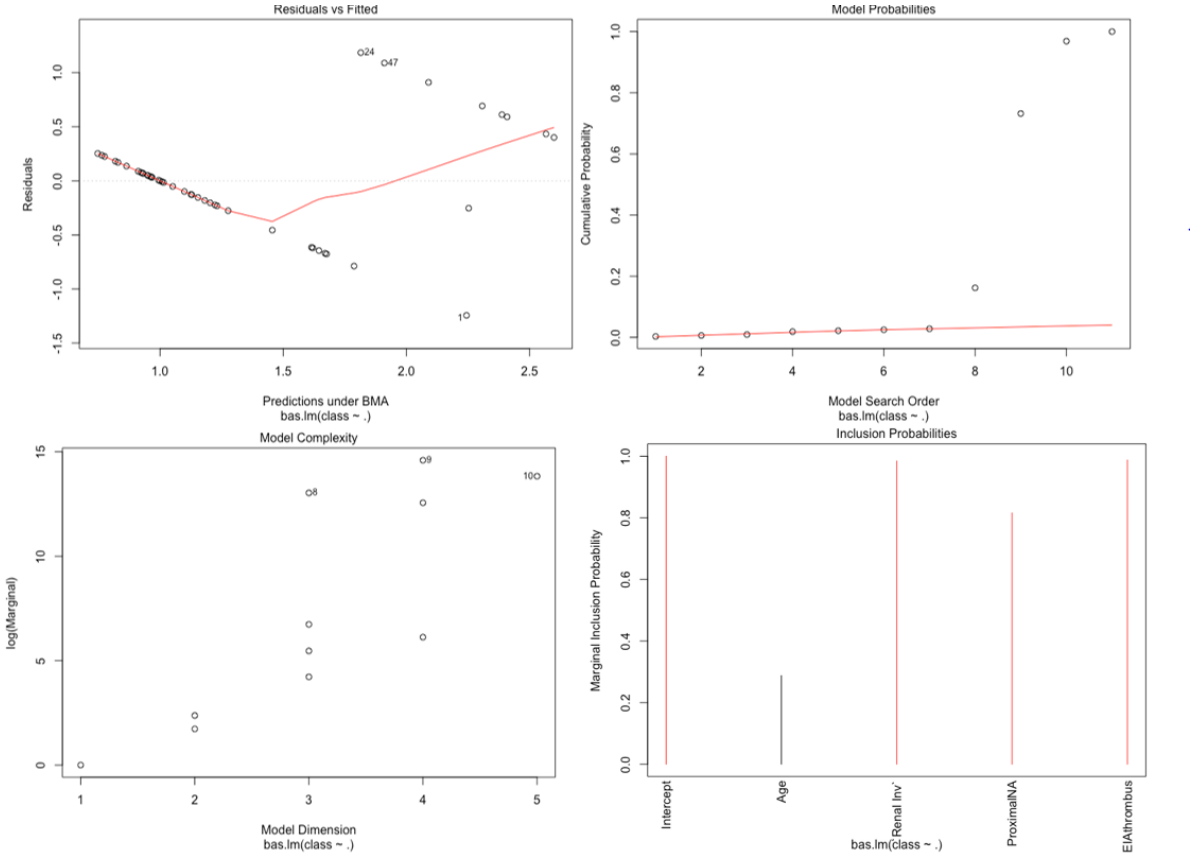

Figure 6. Model Summary Bayesian Boruta Feature Selection

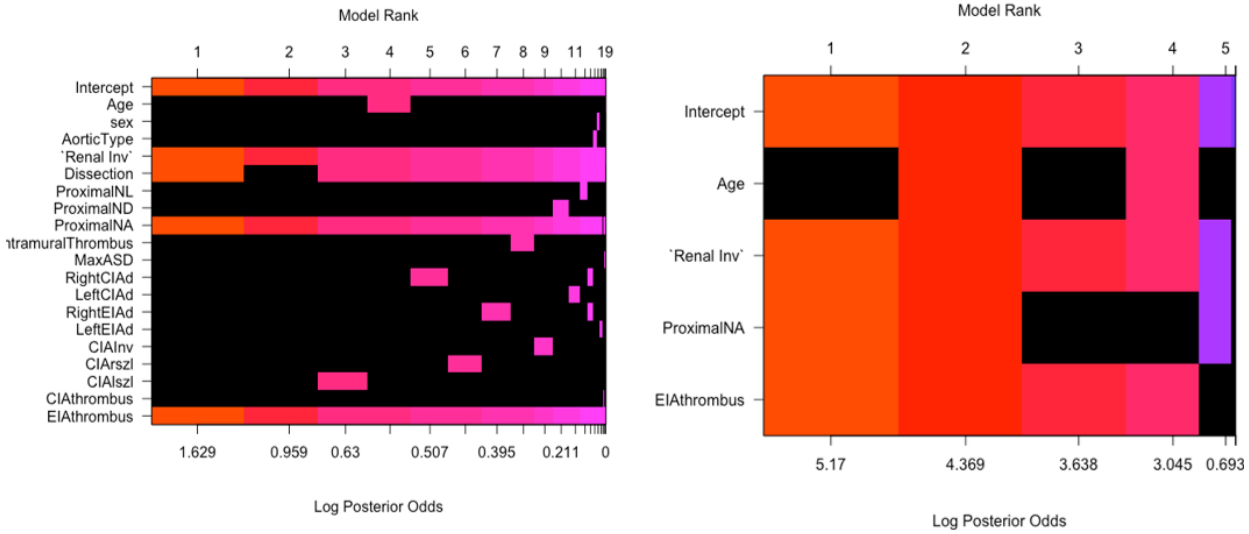

Figure 7. Model Rank All Variables (left), Model Rank Feature Selection (Right)

5. CONCLUSION

Eleven percent of patients had external iliac artery circumferential thrombus > 50%, which occurred either on the left or right side. Sealing zone in both CIA less than 15 mm was none. EVAR in Indonesia started in 2012 with only 4 cases. Despite being relatively new, many cases of EVAR have been managed in our centres in Indonesia during the last five years. The treatment was increasing almost doubled every following year. Most cases were done by vascular surgeons with the referral system to particular university hospitals. Aortic characteristic for AAA in Indonesia has common features; nevertheless, those were relatively long proximal neck and the huge size of the aneurysmal sac in particularly young age. The health care system and insurance in Indonesia could be the factor for late diagnosis. In this paper, we already conduct the feature selection in Bayesian regression as well as MCMC. To begin with, Bayesian modelling using all variables that can reach the R^2 is 87.52%. Moreover, when we were performing feature selection, our model can reach the R^2 88.28%. Generally speaking, feature selection in this simulation does not provide significant differences in accuracy. Overall, it may be said Boruta gives the main results that aortic class must be considered in Renal Inv, Age, EIAthrombus, and ProximalNA. Then if we want to pay attention to age, the important variable is renal inv, and finally, if we pay attention to the aortic class, the important variable is LeftCIAd, LeftEIAd. It means that the older the patient will give a clear difference in renal iv. At the same time, there are significant differences in the aortic class when justified on LeftCIAd, LeftEIAd. Statistically descriptive the size of the diameter of the aorta as an adult (after 18 years) is always the same. If the sex depends on the posture. Indeed, women are usually smaller. AAA is prevalently in men over 65 years. In frequency modelling or classical methods to find out the effect of parameters on the model, then hypothesis testing is performed. The purpose of testing is to find out the predictors that affect the response. Like the classical method, Bayesian also needs to test the parameters obtained. The difference with Bayesian people with frequency is that Bayesian people use credible intervals to find out the parameters that affect the model, while frequency use tests that are full of distribution assumptions. Credible intervals or Bayesian intervals are almost similar to the frequency confidence interval. Future research should be perform with the H-GLM [36],[37],[38], Mixture distribution [39]-[40].

AUTHOR CONTRIBUTIONS

Rezzy Eko Caraka, lead this research and do the Conceptualization, Methodology, Software, Investigation, Data Curation, Formal Analysis, writing-original draft preparation, writing-review

and editing. Nyitiasmono Tri Nugroho, do the Data Curation, Formal Analysis, writing-original draft preparation, writing-review and editing. Shao-Kuo Tai do the supervision, project administration, and funding acquisition. Rung-Ching Chen do the supervision, project administration, and funding acquisition. Toni Toharudin do the project administration, writing-review and editing, and Bens Pardamean do the project administration, funding acquisition, writing-review and editing.

ACKNOWLEDGEMENT AND FUNDING

This paper is supported by the Ministry of Science and Technology, Taiwan. The Nos are MOST-107-2221-E-324-018-MY2 and MOST-106-2218-E-324-002, Taiwan. This paper is also partially sponsored by Chaoyang University of Technology (CYUT) and the Higher Education Sprout Project, Ministry of Education (MOE), Taiwan, under the project name: "The R&D and the cultivation of talent for health-enhancement products".

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

SUPPLEMENTARY MATERIALS

The dataset is available by request to corresponding author.

APPENDIX

At point π, the objective thickness or the full contingent of θ_j in the Gibbs sampler is of a non-standard structure, can be utilized to bring about a Gibbs sampler having a lot of effectively tested standard full conditionals. It increases the objective thickness with a positive dormant variable u developing the joint thickness of u and θ. Along these lines, the minimal thickness for θ is given by π and the Gibbs sampler is reached out to incorporate an additional full contingent for u. Assume that we wish to test from a thickness π given by

$$\pi(\theta) \propto q(\theta)\, f(\theta)$$

where q is a thickness of known structure and f is a non–negative invertible capacity. With the presentation of a latent $u: \Omega \longrightarrow R^{+}$ with the joint thickness can be composed as follows:

$$\pi(\theta, u) \propto q(\theta) I(u < f(\theta))$$

Then, we can minimize u and we may get $\pi(\theta)$ along these lines, the growth is substantial. The prior $p(\theta)$ distribution which has parameter a and comes from the family of distribution $\mathbb{D}$ is

said to be conjugate to $p(y|\theta)$, if the posterior distribution $p(\theta|y)$, which is obtained is also a family of distribution $\mathbb{D}$. As $y_s \sim poisson\ (\theta),\ v = 1,2,\dots,V$ with the likelihood function.

$$p(y|\theta) = \prod_{v=1}^{V} \frac{\exp(-\theta) \times \theta^{\frac{y}{v}}}{y_v!} = \frac{\exp(-V\theta) \times \theta^{(\sum_{v=1}^{V} y_v)}}{\prod_{i=1}^{n} y_v!}$$

It is assumed that prior $p(\theta)$ has gamma distribution with parameter $\boldsymbol{c} = (a, b)^T$ or $\theta \sim gamma\ (a, b)$ so that it can be written as follows,

$$p(\theta) = \frac{b^a}{\Gamma(\mathrm{a})} \theta^{a-1} \exp(-b\theta)$$

Then the posterior $p(\theta|y)$ is written as follows

$$p(\theta|y) = \frac{p(y|\theta) \times p(\theta)}{\int_{\Omega\theta} p(y|\theta) \times p(\theta) d\theta} \propto p(y|\theta) \times p(\theta)$$

with $\Omega\,\theta$ is the domain of θ

$$p(\theta|y) \propto \frac{\exp(-V\theta) \times \theta^{\sum_{v=1}^{V} y_v}}{\prod_{i=1}^{n} y_v!} \times \frac{b^a}{\Gamma(\mathrm{a})} \theta^{a-1} \exp(-b\theta)$$

$$\propto \exp[-\theta(V + b)] \times \frac{\theta^{\overline{V_y}}\, b^a \theta^{a-1}}{\Gamma(\mathrm{a})}$$

$$\propto \exp[-\theta(V + b)] \times \frac{\theta^{(\overline{V_y}+a)-1}}{\Gamma(\mathrm{a})}$$

These results can be said that posterior $p(\theta|y)$ has gamma distribution or can be written $\theta|y \sim gamma\ \left(\overline{V_y} + a, V + b\right)$.

REFERENCES

[1] C. B. Ernst, Abdominal aortic aneurysm, N. Engl. J. Med. 328 (16) (1993), 1167–1172.

[2] T. F. O'Donnell, B. E. Landon, and M. L. Schermerhorn, The case for expanding abdominal aortic aneurysm screening, J. Vasc. Surg. Available, In Press, https://doi.org/10.1016/j.jvs.2019.10.024.

[3] A. Virdis, C. Giannarelli, M. Fritsch Neves, S. Taddei, and L. Ghiadoni, Cigarette Smoking and Hypertension, Curr. Pharm. Des. 16 (23) (2010), 2518–2525.

[4] M. M. Benjamin and W. C. Roberts, Fatal Aortic Rupture from Nonpenetrating Chest Trauma, Baylor Univ. Med. Cent. Proc. 25 (2) (2012), 121–123.

[5] K. McDonald and B. J. Maurer, Familial aortic valve disease: evidence for a genetic influence?, Eur. Heart J. 10

(7) (1989), 676–667.

[6] J. M. Lee et al., Candidates and major determinants for endovascular repair of abdominal aortic aneurysms in Korean patients, Heart Vessels, 28 (2013), 215–221.

[7] M. K. Yii, Epidemiology of abdominal aortic aneurysm in an Asian population, ANZ J. Surg. 73 (6) (2003), 393–395.

[8] N. T. Nugroho and R. Suhartono, Level of Patency for a One Year Period of Infra-inguinal Arterial Bypass in Patients with Peripheral Artery Disease, J. Indones. Soc. Vasc. Endovasc. Surg. 1 (1) (2020), 26-29.

[9] E. A. Idelevich et al., Comparative in vitro activity of bacteriophage endolysin HY-133 against Staphylococcus aureus attached to vascular graft surface, Med. Microbiol. Immunol. 209 (1) (2020), 51-57.

[10] N. T. Nugroho, M. Herten, N. Osada, S. Sielker, and G. B. Torsello, Correlation Among Six Single Nucleotide Polymorphisms Related to Cell Survival, Inflammation and Lipoprotein Regulation for Abdominal Aortic Aneurysm Risk Factor, J. Vasc. Surg. 67 (6) (2018), e232–e233.

[11] A. Mathur, V. Mohan, D. Ameta, G. Bhardwaj, and P. Haranahalli, Aortic aneurysm, J. Transl. Intern. Med. 4 (1) (2016), 35–41.

[12] D. T. Bradley, S. A. Badger, M. McFarland, and A. E. Hughes, Abdominal Aortic Aneurysm Genetic Associations: Mostly False? A Systematic Review and Meta-analysis, Eur. J. Vasc. Endovasc. Surg. 51 (1) (2016), 64-75.

[13] F. A. Lederle, The Aneurysm Detection and Management Study Screening Program, Arch. Internal Med. 160 (10) (2000), 1425-1430.

[14] T. Li et al., TLR4 and MMP2 polymorphisms and their associations with cardiovascular risk factors in susceptibility to aortic aneurysmal diseases, Biosci. Rep. 39 (1) (2019), BSR20181591.

[15] R. C. Lo and M. L. Schermerhorn, Abdominal aortic aneurysms in women, J. Vasc. Surg. 63 (3) (2016), 839-844.

[16] D. P. J. Howard, A. Banerjee, J. F. Fairhead, A. Handa, L. E. Silver, and P. M. Rothwell, Age-specific incidence, risk factors and outcome of acute abdominal aortic aneurysms in a defined population, Br. J. Surg. 102 (8) (2015), 907-915.

[17] P. Chiu et al., Endovascular Versus Open Repair of Intact Descending Thoracic Aortic Aneurysms, J. Amer. Coll. Cardiol. 73 (6) (2019), 643-651.

[18] T. W. Cenggoro, B. Mahesworo, A. Budiarto, J. Baurley, T. Suparyanto, and B. Pardamean, Features Importance in Classification Models for Colorectal Cancer Cases Phenotype in Indonesia, Proc. Comput. Sci. 157 (2019), 313–320.

[19] A. Budiarto, T. Febriana, T. Suparyanto, R.E. Caraka, B. Pardamean, Health Assistant Wearable-Based Data Science System Model: A Pilot Study, in: 2018 International Conference on Information Management and Technology (ICIMTech), IEEE, Jakarta, 2018: pp. 438–442.

[20] L. Rokach, Decision forest: Twenty years of research, Inf. Fusion, 27 (2016), 111–125.

[21] R. E. Caraka and H. Yasin, Geographically Weighted Regression (GWR) Sebuah Pendekatan Regresi Geografis, 1st ed. Yogyakarta: MOBIUS GRAHA ILMU, 2017.

[22] R. Herliansyah and I. Fitria, Latent variable models for multi-species counts modeling in ecology, Biodiversitas J. Biol. Diversity, 19 (5) (2018), 1871-1876.

[23] R. E. Caraka et al., Ecological Show Cave and Wild Cave: Negative Binomial Gllvm's Arthropod Community Modelling, Proc. Comput. Sci. 135 (2018), 377–384.

[24] X. Wang, Y. Yue, J.J. Faraway, Bayesian regression modeling with INLA, CRC Press, Boca Raton, 2018.

[25] T. Murniati, N. Iriawan, D.D. Prastyo, Bayesian Mixture Poisson Regression for Modeling Spatial Point Pattern of Primary Health Centers in Surabaya, MATEMATIKA (Malaysian J. Ind. Appl. Math.), 36 (2020) 51–67.

[26] C. Dewi and R. C. Chen, Random Forest and Support Vector Machine on Features Selection for Regression Analysis, Int. J. Innov. Comput. Inf. Control, 15 (6) (2019), 2027–2037.

[27] R. E. Caraka, R. C. Chen, S. A. Bakar, M. Tahmid, T. Toharudin, and B. Pardamean, Employing Best Input SVR Robust Lost Function with Nature-Inspired Metaheuristics in Wind Speed Energy Forecasting, IAENG Int. J. Comput. Sci. In Press.

[28] M. B. Kursa and W. R. Rudnicki, Feature selection with the boruta package, J. Stat. Softw. 36 (11) (2010), 1-13.

[29] H. Zheng, L. V. Hampson, and S. Wandel, A robust Bayesian meta-analytic approach to incorporate animal data into phase I oncology trials, Stat. Methods Med. Res. 29 (1) (2019), 94–110.

[30] C. Joyner, C. McMahan, J. Baurley, and B. Pardamean, A two-phase Bayesian methodology for the analysis of binary phenotypes in genome-wide association studies, Biometric. J. 62 (1) (2019), 191–201.

[31] R. E. Caraka et al., Variational Approximation Multivariate Generalized Linear Latent Variable Model in Diversity Termites, Sylwan, 164 (1) (2020), 161–177.

[32] K. R. Koch, Monte Carlo methods, GEM - Int. J. Geomath. 9 (2018), 117-143..

[33] M. Carlo, Markov Chain Monte Carlo and Gibbs Sampling, Lecture Notes for EEB 581, version 26 April 2004.

[34] C.P. Robert, G. Casella, Introducing Monte Carlo methods with R, Springer, New York, 2010.

[35] P.U. Gio, E. Rosmaini, The Robustness of Two Independent Samples t Test Using Monte Carlo Simulation with R, IOP Conf. Ser.: Mater. Sci. Eng. 300 (2018), 012030.

[36] Y. Lee and M. Noh, Modelling random effect variance with double hierarchical generalized linear models, Stat. Model. 12 (6) (2012), 487–502.

[37] Y. Lee, L. Rönnegård, M. Noh, Data analysis using hierarchical generalized linear models with R, CRC Press, Taylor & Francis Group, Boca Raton, 2017.

[38] M. Noh, Y. Lee, J. H. . Oud, and Toharudin, Hierarchical likelihood approach to non-Gaussian factor analysis, J.

Stat. Comput. Simul. 89 (3) (2019), 1555–1573.

[39] R. R. M. Tajuddin, N. Ismail, and K. Ibrahim, “Several two-component mixture distributions for count data,” Commun. Stat. Simul. Comput., pp. 1–12, 2020. https://doi.org/10.1080/03610918.2020.1722834

[40] B. Moravveji, Z. Khodadadi, and M. Maleki, A Bayesian Analysis of Two-Piece Distributions Based on the Scale Mixtures of Normal Family, Iran. J. Sci. Technol. Trans. A Sci. 43 (3) (2019), 991–1001.