Identification of factor affecting atrial fibrillation in a patient with risk of obstructive sleep apnea at Rumah Sakit dr. Cipto Mangunkusumo using decision tree method

D N Vitase1,*, T Siswantining1 and T Kamelia2
1 Department of Mathematics, Universitas Indonesia
2 Division of Pulmonology, Department of Internal Medicine, FK Universitas Indonesia/dr. Cipto Mangunkusumo Hospital, Jl. Diponegoro No. 71, Central Jakarta 10430, Indonesia

*Corresponding author: titin@sci.ui.ac.id

Abstract. Obstructive Sleep Apnea (OSA) have a higher prevalence in Asia population. OSA have associated with Atrial Fibrillation (AF) from hypoxemia condition even though the association between OSA and AF could be contributed to risk factor AF among OSA patient. AF has four times the higher prevalence in a patient with OSA than without OSA (4.8% vs. 0.9%). Although, the study about factor affecting AF in OSA patient has not been widely practiced, especially in Indonesia. This study aims to determine the factors affecting AF in patients with risk of OSA. The data used in this study is primary data of patients with risk of OSA at RSCM based on interviews with the Berlin questionnaire. The sample size are 145 patients with risk of OSA based on the Berlin questionnaire using a technique of non-probability sampling. The number of patients diagnosed AF is 45 patients (31.03%). The decision tree used is CHAID and CART to determine which factors significantly affect AF and classify based on the factor. Dependent variables in this study are AF, while independent variables are age, sex, body mass index (BMI), waist circumference (WC), neck circumference (NC), blood pressure (BP), smoking history, history of consuming alcoholic, history of heart failure, history of coronary heart disease (CHD), and history of stroke. The results of this study, which factor significantly affect AF in a patient with risk of OSA using CHAID are age, BP, WC, and NC with the accuracy rate is 71.7%, although using CART are age, BP, WC, BMI, CHD, and smoking with the accuracy rate is 77.2%.

1. Introduction
Obstructive sleep apnea (OSA) is a disorder repetitive collapse of the upper airway at least five times per hour of sleep. OSA is defined by the occurrence of fragmented and snoring during sleep, daytime sleepiness [1]. A valid tool to do screening OSA for Indonesian population is the Berlin questionnaire [2]. In western population, OSA prevalence is estimated about 4% in men and 2% in women, while OSA prevalence in Asia ranged from 3.7% - 97.3% risk factor for OSA is gender, age, blood pressure, neck circumference, smoking, and body mass index (BMI) [16]. In Indonesia, research about OSA showed that significant factors of OSA were age ≥ 36 years, BMI ≥ 25 kg/m², snoring history, and neck circumference ≥ 40 cm, also waist circumference inpatient COPD [3,4]. The period of apnea in OSA causes a lack of oxygen in the blood called hypoxemia. The condition of hypoxemia triggers
imbalance between sympathetic and parasympathetic nerves affecting irregularity heartbeat cause heart disorder, that is Atrial Fibrillation [5].

Atrial Fibrillation (AF) is also cardiovascular diseases due to an irregular heartbeat that affecting blood clots, stroke, heart failure, and other heart complication [6]. AF occurred in 4.3% OSA patient with risk factor were age, sex, hypertension, coronary artery disease, heart failure, and smoking [7]. Research by the Sleep Heart Health Study in 2005 showed about 4.8% of OSA patients have AF, while factors such as age, sex, BMI, and coronary heart disease are associated with AF [8]. However, research about factor affecting AF in OSA patient in Indonesia has not been widely practiced. Therefore, it is necessary to lead research about predictions which factor affecting AF in a patient with risk of OSA. The purpose of this research was to determine which factors affecting AF in patients with risk of OSA at Rumah Sakit dr. Cipto Mangunkusumo (RSCM). Also, classify the characteristic of OSA patients based on the factors most significantly affecting AF in a patient with risk of OSA.

2. Methods
2.1. Data Collection and Variables
This research was conducted at RSCM from March 2018 to May 2018. This research has been approved and passed the code of ethics verification by the Health Research Ethics Committee, Universitas Indonesia-Cipto Mangunkusumo Hospital. The population and sample in this research are patient with risk of OSA at RSCM and patient with risk of OSA based on the Berlin questionnaire do treatment or examination of an electrocardiogram (ECG) at RSCM. The sample is chosen by using a technique of non-probability sampling with the purposive sampling method. Moreover, the sample size is 145 patients. The dependent variable is Atrial Fibrillation (AF), which is categories as Non-AF or AF based on the results of the examination of an ECG. The independent variable that involved are: Age (< 40 years, 40-60 years, > 60 years); Sex (Men, Women); Body Mass Index (Normal, Obesity 1, Obesity 2); Neck Circumference; Waist Circumference; Blood Pressure (Normal, Hypertension Grade 1, Hypertension Grade 2); Smoking (Yes, No); Alcohol Consumption (Yes, No); History of Heart Failure (Yes, No); History of Coronary Heart Disease (Yes, No); History of Stroke (Yes, No).

2.2. Chi-square Automatic Interaction Detection (CHAID)
CHAID method is the iterative technique that examines categorical variables used in the classification and constructs based on the significance level of chi-square to the categorical dependent variable [9,10]. The strength of CHAID is one of non-parametric statistic which not require the normal assumption, a type nominal or ordinal variables could be the independent variable, and the number of categories in each the independent variable can be more than 2 categories [11]. CHAID has three steps, which are merging, splitting, and stopping [12,13]. The CHAID method step was explained in Figure 1. In the CHAID method, chi-square test used to the equality of proportions test and independence test on merging step [9,14]. These are the steps of independency test:

a. *Hypothesis* $H_0$: Both variables are independent vs $H_1$: both variables are not independent
b. Determine the significance level $\alpha$
c. Statistic test. The statistic test used is $\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$. Then, find $\chi^2$ value in a $\chi^2$ table with freedom of degree $(k - 1)(c - 1)$.
d. Decision rule. Reject $H_0$ if $\chi^2 > \chi^2_{\alpha; (k-1)(c-1)}$ or p-value < $\alpha$.
e. Conclusion. If reject $H_0$, then both variables have significantly related while accepting $H_0$, then both variable not significant.
2.3. Classification and Regression Tree (CART)
CART is one of decision tree method to determine and classify the relationship between dependent and independent variable. The result of CART is a classification tree or regression tree depend on the type of independent variable [11]. CART called a binary tree and CART will binary splitting to the child node. The steps of the CART method illustrated in figure 2.

3. Results and Discussions
The result of CHAID and CART is a tree diagram, and the most significant variable will appear in the first node of the graph. Figure 3 illustrated tree diagram as a result of CHAID method, the tree diagram has 3 depths, and the significant factors to AF in patients with risk of OSA were age, waist circumference, blood pressure, and neck circumference. The tree diagram divided classification of OSA patient into 5 classifications, classification (1) is merged node 1,3, classification (2) is merged
node 1,4, classification (3) is merged node 2,5,7, classification (4) is merged node 2,5,8, classification (5) is merged node 2,6 and the classification used to the characteristics of patients with risk of OSA diagnosed AF is first classification (55.9%) with the characteristics are patients over 40 years of age, and having a waist circumference ≤ 80 cm for women or waist circumference ≤ 90 cm for men (normal waist circumference).

Meanwhile, Figure 4 shows the tree diagram a result of the CART method. Figure 4 has five depths, and the variable which significant effect AF in the patient risk of OSA is age, blood pressure, waist circumference, BMI, coronary heart disease, and smoking. Thus, tree diagram (Figure 4) divided the classification of a patient with the risk of OSA into 13 classifications by each classification has different characteristic and percentage. The classification which used to see the characteristic of a patient with risk of OSA is classification twelve (80%) merge node 2,6,12,20,23 with the characteristics are age patients 40-60 years, obesity, hypertension, and having a waist circumference 80 cm for women or waist circumference 90 cm for men (over WC). The significant variable to AF using CHAID and CART method has some different, which is coronary heart disease, BMI, smoking, and neck circumference. The same significant variable to AF using CHAID and CART is age, blood pressure, and waist circumference.

The accuracy classification using the CHAID and CART method is calculated from the classification table and show in Table 1. The CHAID method can predict AF patient, and Non-AF patient overall is 71.7%, while the CART method can predict AF patient and Non-AF patient overall is 77.2% greater than CHAID method. Based on table 1, the CHAID method can predict AF patient in patient who observed have AF correctly is 19 patients (42.2%) and predict Non-AF patient correctly is 85 patients (85%) meanwhile CART method can predict AF patient in patient who observed have AF correctly is 16 patients (35.2%) and predict Non AF patient correctly is 96 patients (96%). This show that the CHAID method has greater than CART method for predict AF patient correctly, although CART gives higher result in overall accuracy rate classification.

![Figure 3. Tree diagram using CHAID method](image-url)
Figure 4. Tree diagram using CART method

Table 1. The accuracy classification result

| Method | Observed | Predicted | Overall Percentage Correct |
|--------|----------|-----------|---------------------------|
|        |          | Non AF | AF | Total |                   |
| CHAID  | Non AF   | 85     | 15 | 100   | 69%   | 85.0% |
|        | AF       | 26     | 19 | 45    | 31%   | 42.2% |
|        | Total    | 111    | 34 | 145   | 100%  | 77.2% |
| CART   | Non AF   | 96     | 4  | 100   | 69%   | 96.0% |
|        | AF       | 29     | 16 | 45    | 31%   | 35.6% |
|        | Total    | 125    | 20 | 145   | 100%  | 77.2% |

4. Conclusion
The main factor which affecting AF in a patient with risk of OSA at RSCM is age. Also, there are other factors that affect AF in patients with risk of OSA at RSCM, CHAID has blood pressure, waist circumference, and neck circumference, while CART has blood pressure, waist circumference, body mass index, coronary heart disease, and smoking. By using CHAID method, we obtain five classifications with the accuracy is 71.7%. Meanwhile, by using the CART method, we obtain thirteen classifications with the accuracy classifications is 77.2%.
Acknowledgment
This research is supported by PITTA UI 2018 Research Grant.

References
[1] Epstein L J, Kristo D, Strollo P J, Friedman N, Malhotra A, Patil S P, Schwab R J, Weaver E M, Weinstein M D 2009 Clinical guideline for the evaluation, management and long-term care of obstructive sleep apnea in adults (USA: J Clin Sleep Med5)
[2] T Kamelia 2017 J. Sleep Med. 40 152
[3] Wiadnyana I P, Susanto A D, Amri Z and Antariksa B 2010 J. Respir Indonesia 30
[4] I R Anisa, T Siswantining, T Kamelia and D Sarwinda 2018 J. Phys.: Conf. Ser. 1108 012076
[5] Goyal S K and Sharma A 2013 World J. Cardiol. 5 157
[6] American Heart Association 2017, February 6 Retrieved from http://www.heart.org/HEARTORG/Conditions/Arrhythmia/AboutArrhythmia/What-is-AtrialFibrillation-AFib-or-AF_UCM_423748_Article.jsp#.WytnKacza00
[7] Gami A S, Hodge D, Herges R M, Olson E J, Nykodym J and Kara T 2007 J. Am. Coll. Cardiol. 49 565
[8] Mehra R, Benjamin E J, Shahar E, Gottlieb D J, Nawabit R, Kirchner H L and Redline S 2006 Am. J. Respir Crit. Care. Med. 173 910
[9] Gallagher C A, Monroe, H M and Fish J L 2000 J. Appl. Stat. 29 256
[10] Kass G 1980 Appl. Stat. 29 119
[11] Diaz-Perez F M and Bethencourt-Cejas M 2016 J. Destination Mark. Manag. 5 275
[12] Alkhasawneh M S, Ngah U K, Tay L T, Isa N A and Al-Batah M S 2014 J. Applied Math. 2014 929768
[13] Susanti Y, Zukhronah E, Respatiwulan and Handajani S S 2017 J.Phys.: Conf. Ser. 909 012041
[14] T Siswantining, Magfiroh, T Kamelia and D Sarwinda 2018 J. Phys.: Conf. Ser. 1108 012118
[15] Ritschard G 2013 CHAID and Earlier Supervised Tree Methods in J J McArdle & G Ritschard (eds), Contemporary Issue in Exploratory Data Mining in Behavioral Sciences, (New York : J J McArdle & G Ritschard )
[16] Mirrakhimov A E, Soornbaev T and Mirrakhimov E M 2013 BMC Pulm. Med. 13 10
[17] Eherler D and Lehmann T 2001 Data Min. Mark. Appl. 49
[18] Breiman L, Friedman J, Stone C J and Olshen R 1993 Classification and Regression Trees (New York: Chapman and Hall)