❒ 460

Prediction of hypertention drug therapy response using K-NN imputation and SVM algorithm

Lailil Muflikhah, Nurul Hidayat, Dimas Joko Hariyanto
Faculty of Computer Science, Brawijaya University, Indonesia

Article Info	ABSTRACT
  	Hypertention is a degenerative disease but its healing takes a long time by consuming hypertention drugs until patient's lifetime. The research is conducted to predict response of drug therapy using bioinformatics approach which is a blend of biological and informatics engineering methods. It is used medical record data of hypertensive patient in drug therapy which has an impact on genetic characteristics. The data is constructed as modelling for learning process. Then, it is implemented as a prediction whether the blood presure is under control or not. However, the amount data have no values, then they are required to be applied preprocessing data. Therefore, this research is proposed K-Nearest Neighbor (K-NN) Imputation algorithm for refining data. After that, it is implemented using Support Vector Machine (SVM) algorithm for prediction. The experiment result is achieved the highest accuracy rate of 90% at the best parameter value $\lambda = 0.9$, $\Sigma = 2$, $C = 0.1$, $\varepsilon = 0.001$ in ten times iterations.

Corresponding Author:

Lailil Muflikhah,
Faculty of Computer Science,
Brawijaya University, Indonesia.
Email: lailil@ub.ac.id

1. INTRODUCTION

Hypertention is a degenerative disease but it is a serious problem in the world. Approximately, there are 65 million patients that are diagnosed hypertention in ratio 1:3 adults and it is 28% American's prehypertention [1]. However, it is only 31% that the targeted blood pressure is achieved which the systolic blood pressure is less than 140 mmHg and or the diastolic blood pressure is less than 90 mmHg. In Indonesia, the prevalence of hypertention reaches 31.7%. Based on the previous research, it was found that 65.8% of patients in Harapan Kita polyclinic were detected having hypertention and only 39.3% reached the blood pressure target [2]. The 66% of hypertensive patients consume medicine on regular, and 60.7% of those have not reached the targeted blood pressure. As the previous research at the heart polyclinic of RSU dr. Saiful Anwar Malang shows that it is only 20.3% of hypertensive patients achieve the target of blood pressure [3]. These results show a high rate of uncontrolled hypertention.

Uncontrolled hypertention is a risk factor for cardiovascular events which is caused of the coronary heart disease and the cerebrovascular disease [4]. Based on the World Health Organization (WHO), uncontrolled hypertention is effect to 7 million deaths in each productive age and 64 million disabilities [5]. Therefore, the efforts to decrease blood pressure is addressed to achieve the targets by adequate therapy. It is a very important to reduce mortality and morbidity of the related hypertention diseases. The most causes of hypertention are multi-factorial, including the related activity of the renin angiotensin system, increasing sympathetic system, obesity, stress, excessive salt consumption and genetics. Angiotensin receptor blockers (ARBs) are powerful vasilidators for inhibitor of Renin Angiotensin Aldosterone System (RAAS) as well as Angiotensin Converting Enzyme inhibitors (ACEi). Therapy using ACEi and ARB has proven many clinical benefits and it is widely used in clinical practice [6]. Uncontrolled blood pressure may be related to

adherence, choice of drug combinations, or genetic variant molecules which are involved in RAAS. Research in Japan showed that the renin C-5312T variant, which nucleotide substitution from C to T in the nucleotide sequence-5312, was an independent predictor of resistance for ARB users in Japan [7].

The other hand, many researchers have developed tools for diagnosis some diseases using bioinformatics approach including machine learning algorithm. The previous research has been conducted on pairwise DNA sequence alignment between hepatitis B virus (HBV) and hepatocarcinoma (HCC) using modified dynamic programming which improved performance of computation space and time [8]. Also, the related research have been conducted for classification of breast cancer using logistic regression, diabetic analylitics using data mining approach. Beside that, classification of brain tumor image segmentation is implemented using hybrid strategy for clustering and segmentation method [9-11]. Furthermore, Support vector Machine (SVM) is one of machine learning algorithms which has high accuracy in medical research. This algorithm is applied to detect hypertention based on radial pulse wave and some risk factors obesity, stress, systolic and diastolic blood pressure, physical exercises, cigaret consumption and diet lifestyle [12-13]. However, the involved data of this research is incomplete. Several features have null values. Therefore, this study is purposed the K-NN Imputation and SVM algorithm to implement a prediction system based on the characteristics of genetic variation in hypertensive patients against drug therapy. The system is developed using patient data that treats hypertention drugs including polymorphime from the angiotensin and renin genes which have an important role in the cardiovascular system. The first step is preprocessing data to solve the missing value using KNN-Imputation, then it will be constructed the model in SVM method in order to predict the drug therapy respond for hypertensive patiens.

2. HYPERTENTION PATHOLOGY

Hypertention is a complex pathophysiological disease. Systemic blood pressure regulation is multifactorial. It is basically the end result of cardiac autoregulation and peripheral vascular resistance (Figure 2.1) [14]. The renin system angiotensin aldosterone plays an important role in the regulation of blood pressure, electrolyte balance, and the pathogenesis of atherosclerosis [15-17]. In the initial phase is increasing RAAS activity which occurs by increasing production of angiotension and or expression or activity of Renin (REN). The Renin catalyzes the breakdown of angiotensinogen (AGT) into angiotensin I. Then, it is catalyzed to angiotensin II with angiotensin converting enzyme (ACE). Furthermore, Angiotensin II increases blood pressure through strong vasoconstriction and sodium retention [15-16]. As an illustration it can be shown in Figure 1.
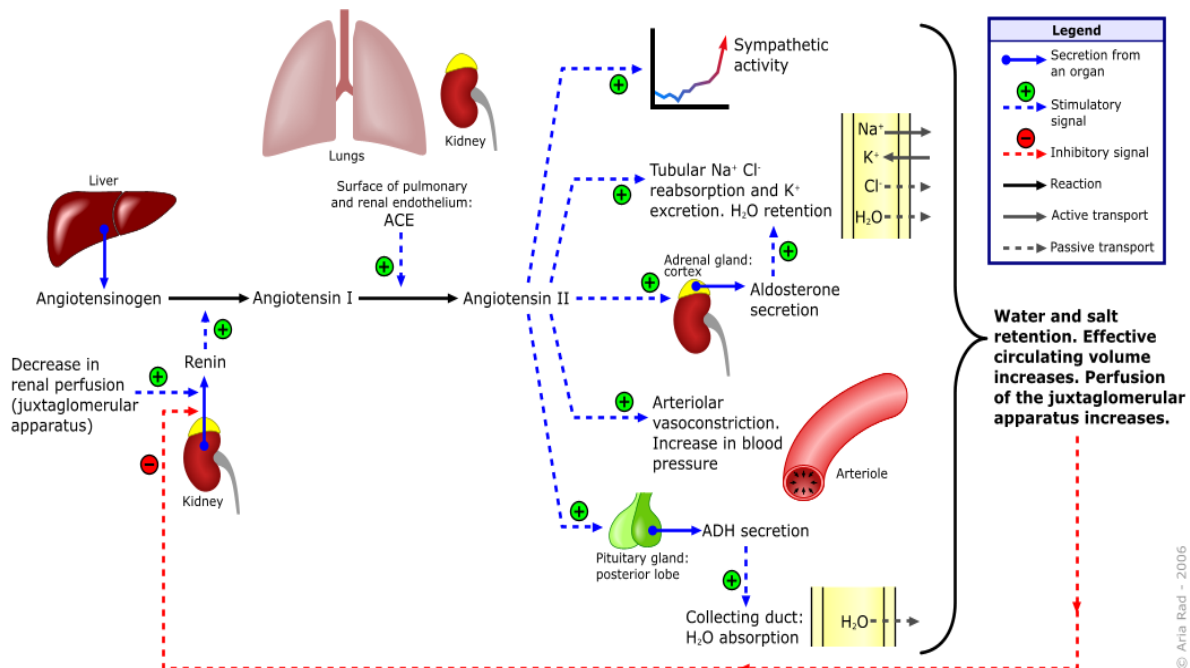

Figure 1. The renin system of angiotensin aldosterone (RAAS) in regulating blood pressure

Furthermore, based on the pathophysiological complexity of hypotension there are various anti-hypertensive therapies which aim to inhibit pathophysiology. There are many classes of antihypertensive drugs such as RAAS inhibitors consisting of ACEi and ARB, calcium channel blockers (CCB), diuretics, nervous system inhibitors sympathetic such as beta blockers, alpha blockers, and the details are shown in Table 1 [18].

Table 1. Anti-hypertention Drug and Dosage

Type of Drug	Name of Drug	Dosage
ACE inhibitors	Captopril	12.5-50 mg twice daily
	Enlapril	5-40 mg once daily or in two equally divided doses
	Fosinopril	10-40 mg once daily
	Lisinopril	5-40 mg once daily
	Perindopril erbumine	4-8 mg once daily
	Perindopril arginine	5-10 mg once daily
	Quinapril	5-40 mg once daily or in two equally divided doses
	Ramipril	2.5-10 mg once daily or in two equally divided doses
	Trandolapril	1-4 mg once daily
Calcium channel blockers-dihydropyridine	Amlodipine	2.5-10 mg once daily
	Felodipine	5-20 mg once daily (controlled release)
	Lercanidipine	10-20 mg once daily
	Nifedipine	10-40 mg once daily (conventional)
		20-120 mg once daily (controlled release)
Calcium channel blockers-nondihydropyridine	Diltiazem	180-360 mg once daily (controlled release)
	Verapamil	120-240 mg once daily (controlled release)
Angiotensin II receptor antagonists	Candesartan	8-16 mg once daily
	Eprosartan	600-800 mg once daily
	Irbesartan	150-300 mg once daily
	Losartan	50-100 mg once daily
	Telmisartan	20-80 mg once daily
	Olmesartan	20-40 mg once daily
Thiazide diuretics	Chlorthalodone	12.5-25 mg once daily
	Hydrochlorothiazide	12.5-25 mg once daily
	Indapamide	12.5-2.5 mg once daily
Beta-blockers	Bisoprolol	1.25-10 mg once daily
	Atenolol	25-100 mg once daily
	Carvedilol	12.5-50 mg once daily
	Labetalol	100-400 mg twice daily
	Metoprolol tartrate	50-100 mg twice daily
	Metoprolol succine (controlled release)	12-190 mg daily
	Oxprenolol	40-160 mg twice daily
Other	Clonidine	50-300 µg twice daily
	Hydralazine	12.5-100 mg twice daily
	Methyldopa	125-500 mg twice daily
	Moxonidine	200-600 µg daily
	Prazosin	0.5-10 mg twice daily

3. RESEARCH METHOD

The system consists of two main stages, are preprocessing data using K-NN Imputation algorithm and prediction using SVM method as shown in Figure 2.
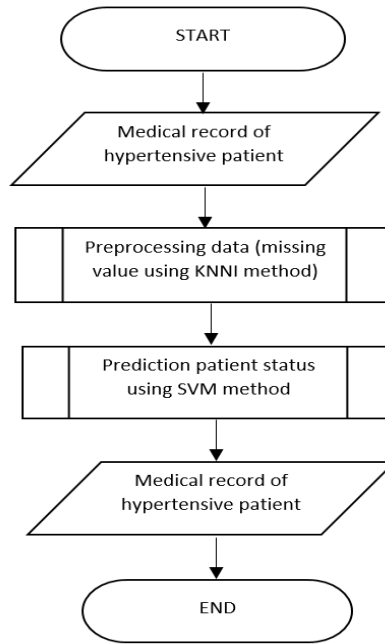
Figure 2. Flowchart of prediction system design for hypertention drug therapy response

3.1. K-Nearest Neighbor (K-NN) Imputation

K-Nearest Neighbor (KNN) Imputation is a method to get the attribute value based on the similarity between new cases and old cases at the appropriate features. According to Olivas [19], it is technique of Machine Learning to handle missing values data by imputation under considering the most similar record values.

At the first stage, input data is taken from medical record of hypertensive patients. However, the number of ignorance data is required to fill values using this method which is based on the appropriate data similarity. By separately, the data is selected which has complete value and has incomplete values. The next process is calculating the eucledian distance of data value to be shorted the similarity. Finally, it is selected to the most frequency values.

3.2. SUPPORT VECTOR MACHINE (SVM)

The Support Vector Machine (SVM) method is a linear classification method by finding the best hyperplane that functions as a separator of two classes in input space. The basic concept of Support Vector Machine is linear classifiers, and then it is developed into non-linear classifiers by incorporating kernel tricks in high-dimensional space as in Figure 3.

In Figure 3, it is illustrated the SVM method. The thick black line in the middle is a hyperplane that separates data +1 and data -1, in this study the data to be used is positive review data and negative review data. The closest point to the hyperplane is called Support Vector. The distance between a support vector and a hyperplane is called a margin. A support vector is a point that intersects a small black line.

Basically, SVM method is a linear classifier that can only be used for linear data. Therefore, it is developed by adding a kernel trick in order to classify non-linear data. The classified data must be transformed to the vector space in high dimension. The kernel trick functions that can be used in non-linear SVM classifications are Polynomial, Gaussian (RBF) and Sigmoid. Each label is denoted $y¬i \in \{-1, +1\}$ for i = 1, 2, ..., n, where n is the amount of data. It is assumed that +1 and -1 classes can be completely separated from the hyperplane, which is defined:

$$w.x + b = 0 \tag{1}$$

Data xi is included into -1 and it is as stated in the (2).

$$w.x_i + b \le -1 \tag{2}$$

Data xi is included into +1 and it is as stated in the (3)

$$w.x_i + b \geq +1 \tag{3}$$

The largest margin is calculated by maximizing the distance between the nearest point and hyperplane.

$$\frac{1}{\|w\|} \tag{4}$$

In general, therefore the real problem is a non-linearly separable form, then two classes cannot be separated by a hyperplane completely. Therefore, SVM modification is needed by entering kernel functions. The non-linear SVM concept is to change the data x that is mapped by the function Φ (x) to a higher dimensional vector space. This mapping aims to represent data in the new vector space.

The learning process of SVM is finding support vectors by dot product data that has been transformed into the new space. The dot product value can be calculated without knowing the data transformation process Φ. The kernel function provides convenience in the SVM learning process to determine support vectors in non-linear data [20]. The kernel function can be formulated as (5)

$$K(x_i.x_j) = \Phi(x_i).\Phi(x_j) \tag{5}$$

In the SVM method, Radial Basis Function (RBF) is a kernel to be applied in this research as in (6)

$$K(X_i \cdot X_j) = \exp\left(-\left(\frac{\|X_i - X_j\|^2}{2\sigma^2}\right)\right) \tag{6}$$

The next step is to make predictions by implementing the Sequential Support Vector Machine method including: calculation of the Hessian matrix, iteration to reach the maximum in least error rate or Max ($|\delta\alpha|$) $<\varepsilon$. After that, the bias and similarities between the testing data and training data are calculated. As a result, it will be obtained the positive or negative classes as it shows in Figure 4. It is a flow diagram of the Support Vector Machine sequential process.

4.   RESULT AND DISCUSSION

The data is taken from Syaiful Anwar Malang Hospital, at heart polyclinic. The feature details are as in Table 2 with balanced classes, which have the same total of data in each class. However, before testing with different data, validation tests are achieved with accuracy rate of 100%. This shows the system that is built is reliable.

Table 2. The Feature of Data Set for Hypertensive Patient in Drug Therapy

Feature	Remark
Code	Identify of patient
Gender	Male/ Female
Age	Birth date
Ethnic	Javanese or others
Waist C	Waist circumference
Hip C	Hip Circumference
Weight	Weight of patient
Height	Height of patient
Smoking	Active/ passive
Menopause	Stop of menstruation
Hypertention	Historical Hypertention
Ur	Ureum
Cr	Creatine
HDL	High Density Lipoprotein
LDL	Low Density Lipoprotein
TG	Trigliserida
Cholesterol	Level of cholesterol total
Glycemia	Level of blood sugar
AGT pre	Level of Angiotensinogen at pre intervension (drug)
AGT post	Level of Angiotensinogen at post intervension (drug)
DeltaAGT	Difference between AGTPre and AGTPost
Difference AGT	up/ down
Quartile AGT	Quartile of Angiotensinogen level
Q3AGTPost	The last quartile of Angiotensinogen

|	
AGT217_1	Polymorfisme of *AGT-217*
AGT20	Polymorfisme of *AGT-20*
AGT6	Polymorfisme of *AGT-6*
REN5312_1	Polymorfisme of *renin-5312*
SBP_ABPM1	Blood pressure systolic before intervention (drug)
SBP_ABPM_2	Blood pressure systolic after intervention (drug)
Undercontrol	Under control in normal constraint before
Alteration_in_24	intervension
DBP_ABPM1	Alteration of blood pressure in 24 hours (before/after
DBP_ABPM2	intervension)
DM	Blood pressure diastolic before intervension
	Blood pressure diastolic after intervension
	Diabetes Militus
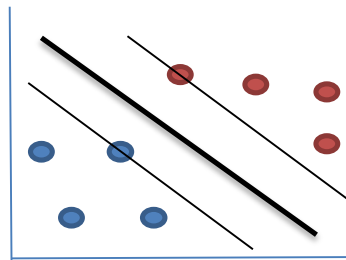
Figure 3. Support vector machine
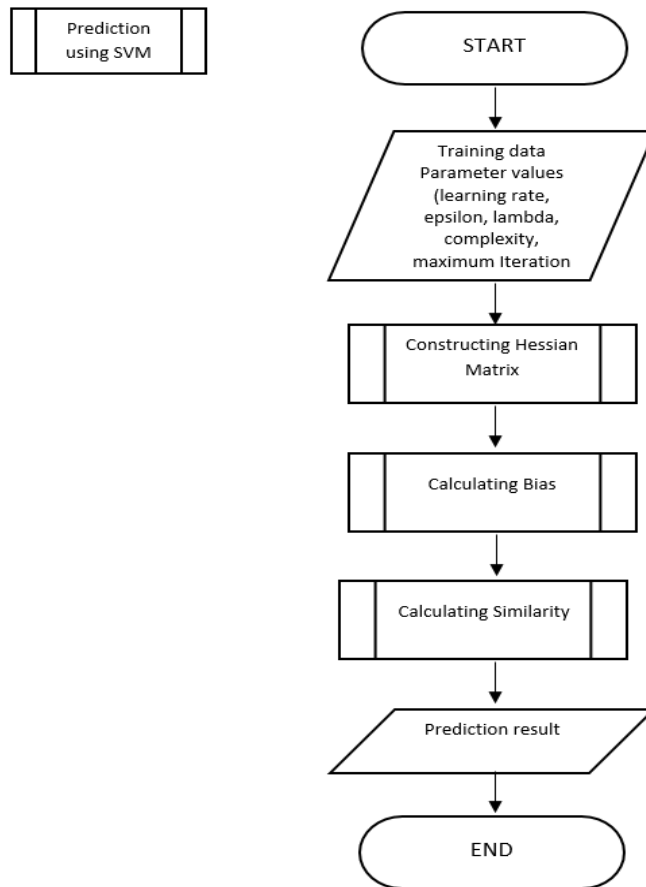
Figure 4. Flowchart of prediction using sequential support vector machine

The interface of prediction system is shown at Figure 5. There are some input parameter for training and testing data including lambda, gamma, C, epsilon and the number of iteration.

Figure 5. Implementation interface

Then, based on the experiment result that the accuracy rate is achieved at 90%. It is applied at the best parameter value as in Table 3.

Table 3. The Optimal Parameter Value for Testing Using SVM Algorithm

Parameter	Value
lambda	0.9
sigma	2
C	0.1
Epsilon	0.001
Number of iteration	10

The accuration rate of testing result is effect to many factors including preprocessing data of missing value using KNNI algorithm which have accuracy rate of 87%. Therefore, the next process for prediction is achieved of 90%.

5. CONCLUSSION

Identification system of drug therapy response for hepatitis patients has been applied using a combination of K-Nearest Neighbor (K-NN) Imputation and Support Vector Machine (SVM) algorithm. The amount missing value data have been covered using K-NN Imputation based on the similarity measure of the attribute value. It was applied at the initial stage before implemented SVM algorithm for prediction. Overall, the accuracy result is achieved of 90%.

REFERENCE

[1] American Heart Association. 2006. *American Heart Association Heart Disease and Stroke statistic 2006 Update*. www.circahajournal.org,

[2] Mohammad Saifur Rohman, Nani Hersunarti. *Achievement of Blood Pressure Target in Out Patient Clinic*. The 3rd scientific meeting on hypertention of Indonesian society of hypertention. 2009:74

[3] Mifetika Lukitasari, Mohammad Saifur Rohman, Dadang Hendrawan. *Achievement of blood pressure target with angiotensin blockage based therapy in outpatient clinic*. 5th scientific meeting of Indonesian society of hypertention abstract book. 2011.

[4]   Whitworth JA, World Health Organization, International Society of Hypertention Writing Group. 2003. *World Health Organization (WHO)/ International Society of Hypertention (ISH) statement on management of Hypertention.* J Hypertens. 2003;21:1983 – 1992

[5]   World Health Organization. 2005. Updated projects of global mortality and burden of disease, 2002–2030: Data sources, methods, and results. Evidence and Information for Policy World Health Organization. Retrieved October 15, 2007, from http://www.who.int/healthinfo/statistics/bodprojections2030/ en/index.html

[6]   Mancia G, De Backer G, Dominiczak A, Cifkova R, Fagard R, Germano G, et al. 2007 Guidelines for the management of arterial hypertention: The Task Force for the Management of Arterial Hypertention of the European Society of Hypertention (ESH) and of the European Society of Cardiology (ESC). J Hypertens. 2007; 25:1105–1187.

[7]   Konoshita T, Kato N, Fuchs S, Mizuno S, Aoyama C, Motomura M, Makino Y, Wakahara S, Inoki I, Miyamori I, Pinet F.. Genetic variant of the Renin-Angiotensin system and diabetes influences blood pressure response to Angiotensin receptor blockers. Diabetes Care. Genomic Disease Outcome Consortium (G-DOC) Study Investigators 2009; 32:1485-1490.

[8]   Lailil Muflikhah and Edy Santoso. Pairwise Sequence Alignment between HBV and HCC Using Modified Needleman-Wunsch Algorithm. *TELKOMNIKA*. 2017; 15(4); pp. 1785-1793

[9]   Mohammed Abdulrazaq Kahya. Classification enhancement of breast cancer histopathological image using penalized logistic regression. *Indonesian Journal of Electrical Engineering and Computer Science*. 2019; 13(1): pp. 405-410

[10] Sinan Adnan Diwan Alalwan. Diabetic analytics: proposed conceptual data mining approaches in type 2 diabetes dataset. *Indonesian Journal of Electrical Engineering and Computer Science*. 2019; 14(1): pp. 92-99

[11] Thrivikram Bathini, Baswaraj Gadgay. Hybrid enhanced ICA & KSVM based brain tumor image. *Indonesian Journal of Electrical Engineering and Computer Science.* 2019; 14(1): pp. 478-489

[12] Zhi-yu Luo, Ji Cui, Xiao-juan Hu, Li-ping Tu,Hai-dan Liu, Wen Jiao, Ling-zhi Zeng, Cong-cong Jing, Li-jie Qiao, Xu-xiang Ma, Yu Wang,Jue Wang,Ching-Hsuan Pai, Zhen Qi, Zhi-feng Zhang, and Jia-tuo Xu. A Study of Machine-Learning Classifiers for Hypertention Basedon Radial Pulse Wave. *Hindawi. BioMed Research International*. 2018; 12 pages

[13] Zainab Assaghir, Ali Janbain, Sara Makki, Mazen Kurdi, Rita Karam. Support Vector Machine to Detect Hypertention. *International Journal on Recent and Innovation Trends in Computing and Communication*. 2017; 5 (2): pp.6-9.

[14] Kaplan NM. Clinical Hypertention 7th ed. 2002; 63

[15] Oparil S, Haber E. The renin-angiotensin system (first of two parts). *N Engl J Med.* 1974; 291: 389–401

[16] Oparil S, Haber E. The renin-angiotensin system (second of two parts). *N Engl J Med.* 1974; 291: 446–457.

[17] Dzau VJ. Theodore Cooper Lecture: tissue angiotensin and pathobiology of vascular disease: a unifying hypothesis. *hypertention.* 2001; 37: 1047–1052.

[18] National Heart Foundation of Australia.Guide to management of hypertention 2008 Assessing and managing raised blood pressure in adults Updated August 2009.

[19] Olivas, Emilio Soria, et.al. Application and Trends: Algorithms, Methods, and Techniques. Handbook of Research on Machine Learning. IGI Global, United States of America. 2010

[20] Vijayakumar, S., Wu, S. *Sequential Support Vector Classification and Regression* International Conference on Soft Computing. 1999.