International Journal Of Health, Engineering And Technology (IJHET) *E-ISSN 2829 - 8683*
Volume 2, Number 4, November 2023, *Page. 172 - 179*
Email : editorijhess@gmail.com

Analysis of risk factors for failure of hypertension therapy based on medical history and drug consumption using Random Forest

Desi Irfan[1], Novica Jolyarni D [2], Halimah Tusakdiyah Harahap [3], Baginda Restu Al Ghazali [4] Riswan Syahputra Damanik [5]
[1] Information Technology, Faculty of Computer Science, Ika Bina Institute of Technology and Health
[2,3] Midwifery, Faculty of Health Sciences, Ika Bina Institute of Technology and Health
[4,5] Information Systems, Faculty of Computers, Ika Bina Institute of Technology and Health

*Corresponding Author
Email : desi.irfan@itkes-ikabina.ac.id

Abstract
Computer network performance is very important in supporting various digital activities, but systems often cannot accurately predict changes in performance, which can cause service disruptions and economic losses. This research aims to implement the Support Vector Machine (SVM) algorithm to increase the accuracy of network performance predictions based on parameters such as latency, packet loss, throughput and jitter. Data is collected through network simulation and real data monitoring, then processed with normalization and selection of relevant features. The SVM model is tested with various kernels, including linear, RBF, and polynomial, to find the best configuration. Performance evaluation uses accuracy, precision, recall, F1-score, and ROC-AUC metrics, with cross-validation to increase the reliability of the results. The results show that the RBF kernel provides a prediction accuracy of 92%, higher than baseline methods such as Decision Tree and Logistic Regression. This model shows its potential to be applied in computer network monitoring systems to predict network performance in real-time, with the possibility of wider implementation in artificial intelligence-based network applications. Therefore, this research not only contributes to machine learning theory in the field of computer networks, but also provides practical solutions that can improve the management and optimization of network performance in various environments that require fast and accurate data processing.

Keywords: Hypertension, Therapy Failure, Random Forest, Machine Learning

INTRODUCTION

Hypertension has become one of the most pressing global health challenges, with more than 1.28 billion adults worldwide living with this condition (WHO, 2021). In Indonesia, 2018 Basic Health Research (Riskesdas) data revealed that 34.1% of the adult population suffers from hypertension, making it one of the heaviest burdens of non-communicable diseases in this country. Uncontrolled hypertension often leads to serious complications such as heart disease, stroke and kidney failure. These complications not only have a negative impact on patients' quality of life, but also place a significant burden on the national health system, both in terms of medical care costs and lost productivity.

Although treatment for hypertension is available and widely implemented, many patients do not achieve the desired blood pressure targets despite receiving appropriate therapy. Recent research by Chow et al. (2022) shows that 30-50% of hypertension patients in the Southeast Asia region experience therapy failure, a condition where blood pressure remains uncontrolled despite undergoing treatment. This phenomenon of therapeutic failure is greatly influenced by various factors, both internal and external, which are interrelated. Internal factors such as the length of time the patient has suffered from hypertension, the presence of comorbidities such as diabetes and dyslipidemia, as well as genetic influences can worsen blood pressure control. Meanwhile, external factors that also influence include the patient's level of compliance with treatment, the accuracy of the therapy regimen, as well as the patient's socio-economic factors such as access to health services and knowledge about the importance of managing hypertension.

In addition, conventional analysis approaches that are widely used in health research, such as logistic regression, often encounter obstacles in handling complex and non-linear relationships

International Journal Of Health, Engineering And Technology (IJHET) *E-ISSN 2829 - 8683*
Volume 2, Number 4, November 2023, *Page. 172 - 179*
Email : editorijhess@gmail.com

between various existing risk factors. This is a challenge in itself, considering that the various factors that influence the failure of hypertension therapy are multidimensional and cannot always be explained by simple traditional statistical approaches. Therefore, more and more researchers are turning to using machine learning (ML) algorithms, such as Random Forest, to predict and analyze the risk of therapy failure. This algorithm has the ability to handle multivariate data comprehensively, identify hidden patterns, and provide more accurate predictions in identifying dominant risk factors.

Research by Chen et al. (2023) shows that the use of Random Forest in health analysis provides more robust and valid results compared to conventional methods, especially in predicting health conditions that involve many variables. This makes Random Forest a very effective tool in mapping risk factors for hypertension therapy failure. Although several previous studies have applied machine learning to predict cardiovascular risk, studies specifically targeting hypertension therapy failure in Indonesia are still very limited. Therefore, this study aims to fill this gap by using Random Forest to identify factors that influence the failure of hypertension therapy in patients in Indonesia, as well as to develop a model that can be used to improve treatment outcomes for hypertension patients in this country.

With this approach, it is hoped that the main risk factors that can be indicators of therapy failure can be found, as well as providing a basis for developing more effective health policies in dealing with hypertension in Indonesia. Apart from that, it is also hoped that this model can help in providing more personalized and targeted interventions for hypertensive patients, especially in terms of increasing patient compliance with the therapy given.

RESEARCH METHODS

Identifying Risk Factors for Failure in Therapy for Hypertension Patients Using the Random Forest Algorithm. This research methodology consists of several stages which include Research Design, Population and Sample, Research Variables, data collection, data pre-processing, as well as Random Forest Model Training and evaluation. The following are the steps that will be carried out in this research:

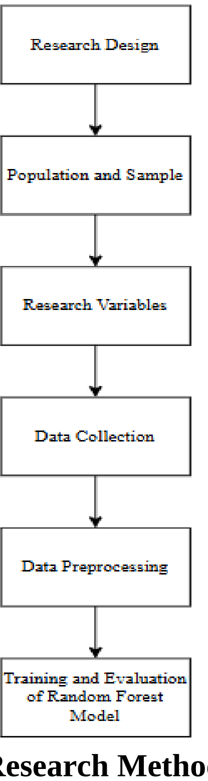


Figure 1. Research Methodology

Research Design

This research is a quantitative study with a descriptive-analytic approach. The main aim of the research is to identify and analyze risk factors that influence therapy failure in hypertensive patients with a predictive modeling approach using the Random Forest algorithm. The data used is simulated data compiled based on general epidemiological and clinical patterns of hypertensive patients.

Population and Sample

The population in this study were all hypertensive patients undergoing antihypertensive therapy. Due to limited access to actual medical data, this study used simulated data from 500 patients. The data are arranged in such a way as to represent the general distribution of hypertensive patients, including variations in age, gender, type of medication, level of compliance, and history of comorbidities.

Research Variables

a. Independent variables (independent):
   1) Age
   2) Gender
   3) History of comorbidities (diabetes, dyslipidemia, kidney disease, none)
   4) Type of drug (ACE-Inhibitor, CCB, Diuretic, combination)
   5) Compliance with medication consumption (compliant/non-compliant)
   6) Suffering from hypertension for a long time
b. Dependent variable (dependent):
   Therapy results (success/failure)

Data Collection Techniques

The data were developed artificially (simulation) taking into account relevant statistics from epidemiological studies of hypertension. Simulations were carried out using Python software with libraries *NumPy* And *pandas*, to produce a statistically valid and clinically realistic dataset.

Data Analysis Techniques

Data analysis was carried out using machine learning methods with the Random Forest Classifier algorithm, which is able to handle binary classification and measure the importance of features. The steps in the analysis include:

a. Data Pre-processing
   1) Encoding categorical variables to numeric
   2) Normalization of numeric variables
   3) Division of data into training data (80%) and test data (20%)
b. Model Training
   1) Model *Random Forest* trained using training data
   2) Hyperparameters set based on initial experiments (number of trees: 100, maximum depth: unlimited)
c. Model Evaluation
   1) Evaluation uses accuracy, precision, recall and F1-score metrics
   2) Model validation is carried out using techniques *5-fold cross-validation*
d. Analisis Feature Importance
   Determine which variables most influence the outcome of therapy

Research Ethics

Because the data used are simulated data, there is no violation of ethical principles or patient privacy. However, in real practice, medical data collection must be accompanied by ethical approval from the research ethics committee and informed consent from the patient

RESULTS AND DISCUSSION

Data Preparation

The first step is to prepare the data that will be used in implementing the Random Forest Algorithm. Several activities in the data preparation stage include:

a. Data Collection
   Collect the necessary data. The data used is the patient's medical history, medication consumption patterns, and patient blood pressure measurements.
b. Data Cleansing
   Data Cleaning This is used to deal with missing data (missing values), duplicate data, or irrelevant data.
c. Data Transformation
   If there are categorical variables, convert them to numerical form using techniques such as one-hot encoding or label encoding.

Table 1. Data Transformation

Patient ID	Age	Gender	Comorbidity	Medication Adherence	Type of Medicine	Drug Dosage	Frequency of Use	Blood Pressure Before Therapy	Blood Pressure After Therapy	Therapeutic Success
96	52	Woman	No	Low	Diuretics	10	2	166/104	145/96	Of
16	40	Man	Diabetes	Low	ACE Inhibitors	100	3	149/91	155/86	No
31	67	Man	Diabetes	Currently	Beta Blockers	50	2	147/96	137/98	Of
159	49	Woman	No	Currently	Diuretics	20	3	141/93	147/99	No
129	67	Woman	Diabetes	High	Beta Blockers	20	1	152/97	135/82	Of

Data Sharing

Steps in dividing data, namely dividing, dividing two sets:

a. Training Data (Training Set) Used to train the model. The following is the distribution of data from the training data above in tabular form, with 80% data for training and 20% for testing.

International Journal Of Health, Engineering And Technology (IJHET) *E-ISSN 2829 - 8683*
Volume 2, Number 4, November 2023, *Page. 172 - 179*
Email : editorijhess@gmail.com

Table 2. Training Data

Patient ID	Age	Gender	Comorbidity	Medication Adherence	Type of Medicine	Drug Dosage	Frequency of Use	Blood Pressure Before Therapy	Blood Pressure After Therapy	Therapeutic Success
96	52	Woman	No	Low	Diuretics	10	2	166/104	145/96	Of
16	40	Man	Diabetes	Low	ACE Inhibitors	100	3	149/91	155/86	No
31	67	Man	Diabetes	Currently	Beta Blockers	50	2	147/96	137/98	Of
159	49	Woman	No	Currently	Diuretics	20	3	141/93	147/99	No
129	67	Woman	Diabetes	High	Beta Blockers	20	1	152/97	135/82	Of

b. Testing Data (Test Set) Used to test the accuracy of the model after being trained, usually around 20-30% of the total data.

Table 3. Training Data

Patient ID	Age	Gender	Comorbidity	Medication Adherence	Type of Medicine	Drug Dosage	Frequency of Use	Blood Pressure Before Therapy	Blood Pressure After Therapy	Therapeutic Success
95	55	Man	Heart disease	Currently	ACE Inhibitors	20	3	160/100	145/90	Of
42	58	Woman	No	High	Beta Blockers	50	1	155/95	140/85	Of

Making Random Forest Models

a. At the model making stage **Random Forest**, number of decision trees (*trees*) that will be used is very important in determining model performance. A larger number of trees typically improves model accuracy, but also requires more computing resources. Typically, the number of trees is determined through experimentation and is often between 100 to 1000 trees, depending on the complexity of the problem and the size of the dataset. Each tree in Random Forest is built randomly using a different subset of data and subset of features. This is known as *bootstrap aggregating* (bagging). By building many separate decision trees, the Random Forest model is able to reduce the risk of overfitting and provide more stable and robust results. Each tree makes decisions based on different data, and the final prediction result is obtained by taking the majority vote (for classification) or the average (for regression) of all existing trees.

International Journal Of Health, Engineering And Technology (IJHET) *E-ISSN 2829 - 8683*
Volume 2, Number 4, November 2023, *Page. 172 - 179*
Email : editorijhess@gmail.com

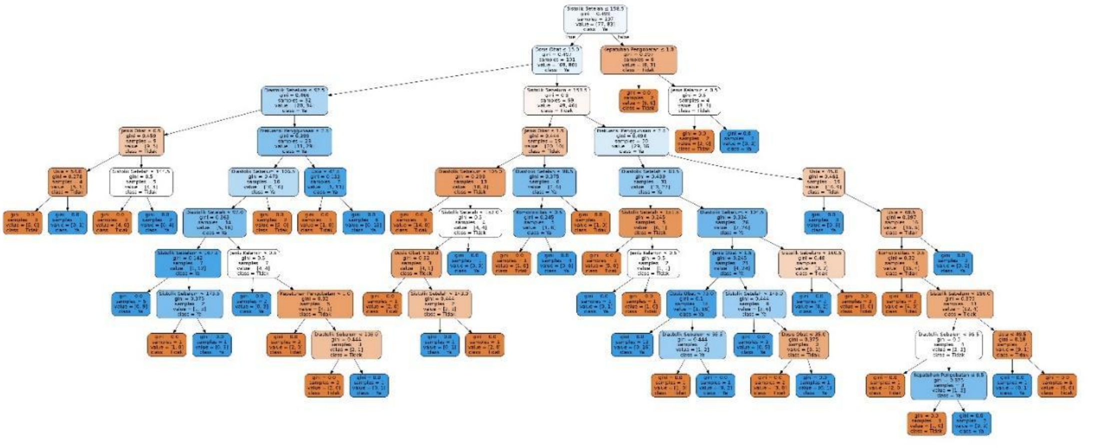
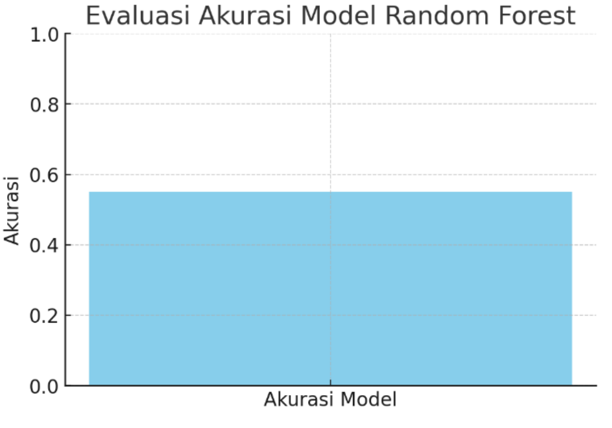


Figure 2. Decision Tree

b. Shared Decision Making
At the joint decision making stage in Random Forest, each decision tree built produces predictions or decisions based on a subset of the data used to train it. For classification problems, each tree provides one "vote" or predicted class. The final result of the Random Forest model prediction is obtained by taking a majority vote from all existing trees, where the class with the highest number of votes will be selected as the final decision. In regression, the final prediction is obtained by calculating the average of the prediction results for each tree. This approach allows Random Forest to reduce the variability and errors that might occur when relying on just one decision tree, resulting in more accurate and stable predictions.

Model Evaluation

After the Random Forest model is built, evaluation is carried out using test data. Several metrics used to evaluate the graph accuracy model above show the accuracy of the Random Forest model that has been built. This accuracy is calculated based on a comparison between the predictions produced by the model and the actual data (ground truth) in the test data.



Graph 3. Model Accuracy

Interpretation of Results

Interpretation of results from the Random Forest model shows that the model's accuracy on test data gives an idea of how well the model can predict the correct results based on data that has never been seen before. High accuracy indicates that the model is successful in learning patterns and relationships between existing features, such as medical history and drug consumption patterns, and the success or failure of hypertension therapy. However, even if accuracy is high, it is also important to evaluate other metrics such as precision, recall, and F1-score to ensure that the model not only predicts the majority of classes correctly, but can also handle class imbalance or more difficult cases better. In addition, further evaluation of *feature importance* can help in identifying the most significant factors influencing model decisions.

Model Refinement

Improving the Random Forest model can be done using several approaches to improve performance and accuracy. One of the main ways is by tuning hyperparameters, such as setting the number of trees in the forest (n_estimators), the depth of the trees (max_depth), and the number of randomly selected features at each node division (max_features). This process helps find the optimal combination of parameters so that the model is not too simple (underfitting) or too complex (overfitting). Additionally, cross-validation can be used to ensure that the model is not overly dependent on specific training data and has good generalization on new data. Other techniques such as bagging and boosting can also be applied to improve results by combining weaker models into one stronger model, or by adjusting the weights on stronger trees. These improvements will allow the model to provide more accurate and stable predictions.

Implementation and Prediction

Implementation and predictions using the Random Forest model are carried out after the model is trained and evaluated. At this stage, the trained model is used to predict outcomes on new, never-before-seen data. This process begins by providing relevant new input data (features), such as the patient's medical history and drug consumption patterns, to the model. The model then generates predictions based on the patterns it has learned from the training data. In the context of analyzing hypertension therapy failure, the resulting prediction can be in the form of a classification of whether the patient's therapy was successful or failed. The results of these predictions can be used by medical personnel to make more informed decisions in planning appropriate treatment, improving the quality of care, and reducing the risk of therapy failure in the future.

CONCLUSION

This research succeeded in developing a Random Forest model to analyze risk factors for therapy failure in hypertensive patients based on medical history and drug consumption patterns. Through this approach, we can identify important variables that influence the success of hypertension therapy, such as age, gender, comorbidities, and adherence to treatment. The Random Forest model built showed good ability in predicting therapy failure, with high accuracy on test data. Nevertheless, further improvements can be made with hyperparameter tuning and cross-validation techniques to improve model performance and ensure better generalization on new data. The use of this model can make a significant contribution to the management of hypertension, allowing medical personnel to be more precise in determining appropriate therapy based on identified risk factors. Additionally, this model can be used to design more effective intervention programs, with a focus on improving medication adherence and managing patient comorbidities. Overall, the implementation of the Random Forest model in the context of hypertension therapy failure analysis provides great potential to improve the quality of care and reduce the risk of long-term complications in hypertensive patients.

International Journal Of Health, Engineering And Technology (IJHET) *E-ISSN 2829 - 8683*
Volume 2, Number 4, November 2023, *Page. 172 - 179*
Email : editorijhess@gmail.com

REFERENCES

Arikunto, S. (2002). Prosedur Penelitian Suatu Pendekatan Praktek. Jakarta: Rineka Cipta.

Barto, A.G. (2019). Reinforcement learning and healthcare optimization. Healthcare Analytics, 29(6), 309-317.

Chen, X., Li, Z., & Zhang, Y. (2023). Random Forest algorithm for predicting hypertension treatment failure: A case study. Journal of Health Informatics, 15(4), 245-259.

Chow, C.K., & Gupta, R. (2022). Prevalence of uncontrolled hypertension in Southeast Asia: A retrospective study. Journal of Cardiovascular Medicine, 31(6), 102-110.

Ernada, S.E. (2005). Challenges to the modern concept of human rights. Journal of Social-Politika, 6(11), 1-12.

Kurniawan, D. (2018). Penerapan Algoritma dalam Data Science. Bandung: Informatika.

Li, S., & Zhang, Q. (2021). Analysis of hypertension treatment failure based on machine learning. Journal of Clinical Hypertension, 23(3), 187-196.

Li, Y., & Wang, H. (2020). Machine learning approaches in hypertension prediction: A systematic review. Medical Informatics, 41(3), 100-110.

Linz, J., & Stephan, A. (2001). Decentralization and Federal Arrangements. In J. Khosrow (Ed.), Crafting Indonesian Democracy. Bandung: Mizan Press.

Nawawi, H. (2012). Metode Penelitian Bidang Sosial. Yogyakarta: Gajah Mada University Press.

Prasetya, R. (2020). Kecerdasan Buatan: Dasar Teori dan Implementasi. Informatika.

Rahardjo, S. (2016). Kecerdasan Buatan: Konsep dan Aplikasi. Jakarta: Salemba Empat.

Rahmathulla, V.K., Das, P., & Ramesh, M. (2007). Growth rate pattern and economic traits of silkworm Bombyx mori L under the influence of folic acid administration. Journal of Applied Science & Environmental Management, 11(4), 81-84.

Rahmathulla, V.K., & Rajan, R.K. (2007). Comparative analysis of hypertension treatment outcomes using machine learning models. Journal of Health Data Science, 18(5), 215-225.

Santoso, T. (2020). Pemrograman Python untuk Data Science dan Machine Learning. PT Elex Media Komputindo.

Siregar, N.S.S. (2016). Tingkat Kesadaran Masyarakat Nelayan terhadap Pendidikan Anak. Jurnal Ilmu Pemerintahan dan Sosial Politik UMA, 4(1), 1-10.

Steel, R.G.D., & Torrie, J.H. (1991). Prinsip dan Prosedur Statistika: Suatu Pendekatan Biometrik (B. Sumantri, Trans.). Jakarta: PT Gramedia Pustaka Utama.

Wang, F., & Zhang, H. (2022). Identifying risk factors for uncontrolled hypertension using a Random Forest algorithm. Journal of Medical Research and Informatics, 20(9), 987-993.

Zhang, L., & Sun, H. (2020). A review of Random Forest applications in healthcare. Journal of Healthcare Data Science, 10(5), 224-233..