Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2022, 2022:125
https://doi.org/10.28919/cmbn/7752
ISSN: 2052-2541

CLASSIFICATION MODEL FOR TYPE OF STROKE USING KERNEL LOGISTIC REGRESSION

SUWARDI ANNAS[1], BOBBY POERWANTO[1,*], ASWI[1], MUHAMMAD ABDY[2], RISKA YANU FA'RIFAH[3]

[1]Department of Statistics, Universitas Negeri Makassar, Indonesia

[2]Department of Mathematics, Universitas Negeri Makassar, Indonesia

[3]Department of Information System, Telkom University, Indonesia

Abstract: Stroke is the second leading cause of death in the world and has a high contribution to disability. Many stroke sufferers do not recognize the symptoms of stroke or do not even have knowledge related to stroke. This causes many sufferers to be late to the hospital for first aid. This can lead to even greater risk. One effort that can be done is to find out the factors that influence stroke so that it can be prevented. This study aims to create a classification model that can be used to predict the type of stroke and to find out what factors have a significant effect on the type of stroke. The method used is Kernel Logistic Regression (KLR) which is the development of Logistic Regression (LR) by using a linear combination of regularized LR. In the modeling, two scenarios for the distribution of training and testing data were also carried out, namely, scenarios 7:3 and 8:2. The results of the accuracy of the two scenarios, are 75.97% for scenario 8:2 and 73.97% for scenario 7:3. The accuracy of the KLR is 92.12% which increased by 16.15% from the LR. From the modeling results for scenario 8:2, it was found that four predictors affected the type of stroke significantly, namely cholesterol level, temperature, length of stay, and disease history.

Keywords: logistic regression; kernel logistic regression; stroke.

2010 AMS Subject Classification: 62P10.

[*]Corresponding author

E-mail address: bobby_poerwanto@unm.ac.id

Received September 23, 2022

1. INTRODUCTION

Stroke is currently listed as the second leading cause of death in the world, and is also one of the diseases with a high contribution to the main cause of disability [1]. There are about 80.1 million people worldwide and 6.2 million of them die each year due to stroke according to the latest study from the Global Burden of Disease Study [2]–[4]. In Indonesia, the Ministry of Health conducted Basic Health Research in 2018. They found that the prevalence of stroke based on diagnosis in the population aged 15 years in Indonesia increased to 10.9 per mil where previously it was only 7 in 2013. In other words, there is an increase in the prevalence of stroke in all provinces in Indonesia [5].

Many of the impacts experienced by stroke survivors include neurological deficits that result in varied quality-of-life disturbances which are also a heavy burden for patients, nurses, and those around the patient. The incidence of recurrent stroke will be oriented toward higher socioeconomic impact, higher case mortality, and clinical examination with worse outcomes than the first stroke experience [6].

A study from the ASEAN Neurological Association (ASNA) stated that in Indonesia, the majority of 2065 stroke patients spread across 28 hospitals were about 6 hours late arriving at the hospital for first aid. The most common reasons are that patients were not aware of the symptoms of stroke, and the lack of knowledge about the symptoms of a stroke. From the data, it was found that 20% of them had repeated strokes [7]. If this continues, then the number of stroke deaths in Indonesia is likely to remain high.

In general, there are two types of stroke, namely ischemic stroke and hemorrhagic stroke. An ischemic stroke is a stroke that occurs when blood flow to the brain is blocked by a blood clot, while a hemorrhagic stroke occurs when a weak blood vessel bursts and bleeds into the brain. Most strokes that occur are ischemic strokes. Only about 15% of hemorrhagic strokes occur from the average stroke, but this type of stroke is more deadly [8].

One of the efforts that can be done to reduce the mortality rate due to stroke is by modeling the type of stroke classification which can also find out the significant factors that affect the type of stroke. The built model can be used as a predictive model for the classification of stroke, whether it is ischemic or hemorrhagic. Furthermore, knowing the factors that influence the type of stroke, can be used as an early warning so that preventive efforts can be made against these influencing factors. Disease modeling can be used as a basis for government policies or programs [9].

Several methods have been used in classifying strokes such as neural network [10], support vector machine and random forest [11], naïve Bayes [12], logistic regression [13], and other methods. However, many of the methods mentioned have not been able to determine the relationship between response and predictors. Logistics regression (LR) is a method that not only being able to perform classification but also able to see the relationship between factors and the response. However, the weakness of this method is that the accuracy produced when the data has large dimensions tends to be low [14]. One effort that can be done to improve the accuracy of logistic regression is to use a linear combination of regularized LR. The form of a linear combination of these parameters is called a kernel.

A kernel is a machine-learning method for classification that is formed based on the results of the log-likelihood transformation of regularized LR using kernel functions. This method is known as Kernel Logistic Regression (KLR). By using the maximum likelihood estimation (MLE) method, the parameter estimation results in KLR are not closed form. Minka (2003), used Newton Raphson's numerical method to obtain parameter estimates for KLR. However, this numerical method does not provide optimum estimation results because the dimensions of the resulting hessian matrix are high. Maalouf, Trafalis, and Adrianto [15] use the truncated newton numerical method to solve this problem.

Research conducted by Maalouf, Trafalis, and Adrianto [15] explains that the KLR method can provide a higher classification accuracy than the Support Vector Machine (SVM) and regularized LR methods. The results of the analysis show that at 95% confidence intervals for the classification of sonar data with a sample of 200, the KLR is (89 ± 4.3), the SVM method is (87 ± 4.5), and IRLS is (79.2 ± 5.5). For the larger sample size, namely the diabetes data (768 samples), it provides a classification accuracy that is not much different.

2. MATERIALS AND METHODS

2.1. Logistic Regression

LR is the most frequently used linear prediction method for binary data. This method performs as well as the machine learning (ML) model for predicting the risk of major chronic diseases with low incidence and simple clinical predictors [16]. The binary logistic regression model can be written as follows:

$$\pi(\mathbf{x}_i) = \frac{\exp(\mathbf{x}_i \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i \boldsymbol{\beta})}$$

The model consists of one parameter $\boldsymbol{\beta}$ and the predictor in the variable $\mathbf{x}_i$. The number of parameters consists of the number of predictors plus one intercept. The method that is often used in estimating parameters in LR is maximum likelihood estimation (MLE) with Newton Raphson's numerical iteration. The interpretation of the model obtained from the analysis results using the odds ratio $(e^{-\beta})$ and the goodness of the model obtained is sought with the level of accuracy in classification [17].

2.2. Kernel Logistic Regression

KLR is one of the machine-learning methods of LR development [18]. In its development, this KLR is used to overcome overfitting and non-optimized accuracy of LR modeling. Maalouf, Trafalis, and Adrianto [15] also mention that KLR is a powerful machine-learning method for classification. The KLR model is formed by adding kernel functions to the LR model. The KLR model can be written as follows:

$$\begin{aligned}\log[\pi(\mathbf{k}_i)] &= \sum_{l=1}^{n} \alpha_l \mathbf{K}((\mathbf{x}_i),(\mathbf{x}_l)) \\ &= \mathbf{k}_i \boldsymbol{\alpha}\end{aligned}$$

$\mathbf{k}_i$ is the kernel element in the i[th] row of the kernel matrix $\mathbf{K}((\mathbf{x}_i),(\mathbf{x}_l))$, with the kernel function commonly used is the radial basis function (RBF) kernel. The RBF kernel function is as follows:

$$\begin{aligned}\mathbf{K} &= e^{-\frac{1}{2\sigma}((\mathbf{x}_i-\mathbf{x}_l)^T(\mathbf{x}_i-\mathbf{x}_l))} \\ &= e^{-\gamma\|(\mathbf{x}_i-\mathbf{x}_l)\|^2}\end{aligned}$$

2.3. Classification Accuracy

The validation of the classification prediction model can be measured using accuracy, specificity, and sensitivity. The three indicators can be seen based on the confusion matrix given in Table 1 [19]:

Table 1. Confusion Matrix for Binary Data

Y	Predicted Class = 1	Predicted Class = 0
Actual Class = 1	TP	FN
Actual Class = 0	FP	TN

where

TP (true positive) : the number of predictions coded 1 that matches the observation

FP (false positive) : the number of predictions coded 1 that does not match the observation

FN (false negative) : the number of predictions coded 0 that does not match the observation

TN (true negative) : the number of predictions coded 0 that matches the observation

The accuracy, sensitivity, and specification of the classification can be seen in the following equation:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Spesificity = \frac{TN}{TN + FP}$$

To find out the results of a good classification prediction in LR, accuracy is the most frequently used.

2.4. Data Source

The data used are medical records of stroke patients at the Stroke Center of the DADI RSKD, South Sulawesi Province, Indonesia. The data are 262 data consisting of 161 ischemic stroke data, and 101 hemorrhagic stroke data for approximately the last 6 years. The variables used in this study consisted of two types, namely responses and predictors. The description of the variables used can be seen in Table 2.

Table 2. Description of Variables Used

Variable	Description	Data Scale
Stroke Type (Y)	Response	Nominal 0: ischemic 1: hemorrhagic
Cholesterol Level (X_1)	Predictor 1	Interval
Blood Sugar Level (X_2)	Predictor 2	Interval
Temperature (X_3)	Predictor 3	Interval
Length of stay (X_4)	Predictor 4	Interval
Pulse Frequency (X_5)	Predictor 5	Ratio
Gender (X_6)	Predictor 6	Nominal 0: Male 1: Female
History (X_7)	Predictor 7	Nominal 0: No history of hypertension/diabetes mellitus/stroke 1: has a history

2.5. Procedure to analyze Data

The analytical method used is LR and KLR. LR is used to predict factors that influence different categories of stroke patients. KLR is used to optimize the prediction results of LR modeling by involving a significant predictor. The stages of data analysis in this study are as follows.

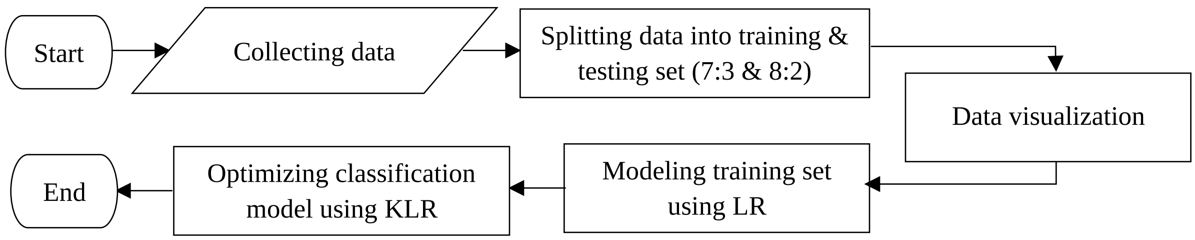

Figure 1. Research Stages

Based on Figure 1, broadly speaking, this research consists of 3 stages. The first stage is collecting and splitting data into training and testing with a ratio of 7:3 and 8:2. The second stage is data visualization which aims to provide an overview of the variables used in this study. The third stage is the data analysis section, which is to predict the classification with LR and interpret the formed model, evaluate/validate the model, and optimize the classification model formed using KLR. The data used in the classification model is training data. Testing data is used to evaluate/validate the classification model formed on the LR and the KLR is used for classification optimization by looking at the resulting increase in accuracy.

3. RESULTS AND DISCUSSION

3.1.Results

Before doing the modeling, the description of each variable used is first explained. The first is the response variable, namely the type of stroke where this variable is binary consisting of "0" for the ischemic stroke type, and "1" for the hemorrhagic stroke type. From a total of 262 data, each percentage is obtained as follows:

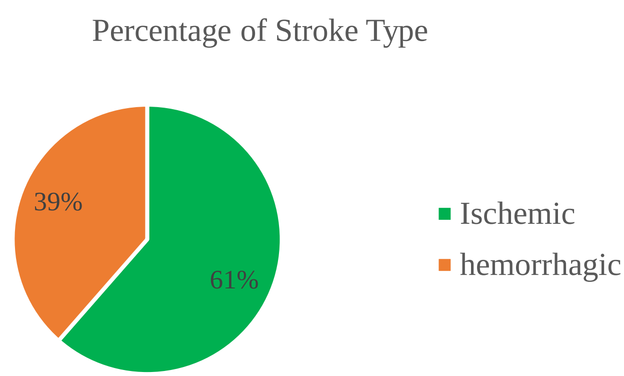

Figure 2. Pie Chart of Stroke Type

It can be seen in Figure 2 that from 262 data, ischemic stroke types have a percentage of 61%, and the rest are hemorrhagic strokes. This is also in accordance with research conducted by Joy [8] which states that ischemic stroke sufferers are the most common.

Furthermore, the variables for cholesterol level, blood sugar level, temperature, length of stay, and pulse frequency can be seen in Table 3.

Table 3. Description of Predictor with Interval Scale

Cholesterol Level (X_1)		Blood Sugar Level (X_2)		Temperature (X_3)		Length of stay (X_4)		Pulse Frequency (X_5)	
Mean	202.07	Mean	126.69	Mean	36.51	Mean	8.27	Mean	92.83
S.E	3.07	S.E	3.55	S.E	0.05	S.E	0.31	S.E	0.89
Median	202.07	Median	113.5	Median	36.5	Median	7	Median	88
Std. Deviation	49.74	Std. Deviation	57.49	Std. Deviation	0.76	Std. Deviation	5.07	Std. Deviation	14.48
Variance	2474.32	Variance	3304.58	Variance	0.57	Variance	25.70	Variance	209.71
Range	298	Range	430	Range	9.2	Range	34	Range	86
Min	89	Min	29	Min	31	Min	1	Min	60
Max	387	Max	459	Max	40.2	Max	35	Max	146
Sum	52941.84	Sum	33192.03	Sum	9564.98	Sum	2166	Sum	24320.97

In this study, two additional predictors are categorical data. The percentage of gender and disease history can be seen in Figure 3.

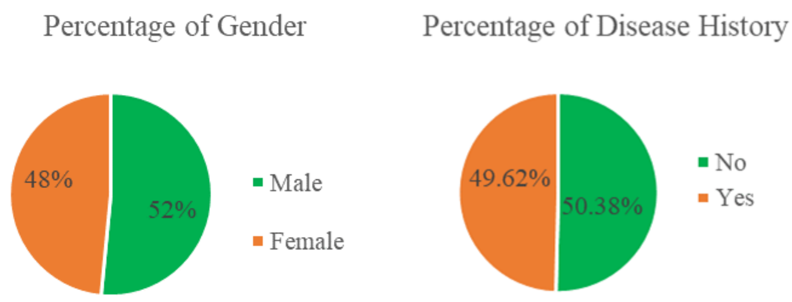

Figure 3. Pie Chart of Gender (a), and Disease History (b)

For the gender percentage, it can be seen that men suffer from stroke (52%), and the rest are women. For disease history, the difference is not too significant, namely patients who have a history of hypertension, diabetes mellitus, or stroke, which is 49.62% of the 262 data collected.

In modeling, this research begins with splitting the data from the dataset used. Splitting data is divided into two scenarios, namely the comparison of training and testing 7:3 and 8:2. Splitting of data is carried out by randomly using the help of statistical tools with results as shown in Table 4.

Table 4. Data Splitting Results

Scenario	Training	Testing
7:3	189	73
8:2	215	47

After splitting the data, the next step is classification modeling with training data for each scenario. The modeling of stroke patient data with LR as a reference is category 0. The results of the analysis with LR for prediction of classification of stroke patient data show that of the 7 predictors, there are 3 significant predictors seen from the p-value smaller than 0.05. These results are obtained for all scenarios. The details of the coefficient results from all predictors used to predict the classification of different types of stroke in scenarios 7:3 and 8:2 can be seen in Table 5.

Table 5. Coefficients of All Predictors of LR Analysis

Scenario 7:3			
Variables	Coef.	P-value	Exp (Coef.)
X1	-0.006	0.102	0.994
X2	-0.003	0.245	0.997
X3	-0.624	0.022	0.536
X4	0.177	0.00	1.194
X5	0.017	0.180	1.017
X6	0.506	0.147	1.658
X7	0.515	0.044	1.673
Constant	20.305	0.036	657996619.2
Scenario 8:2			
Variables	Coef.	P-value	Exp (Coef.)
X1	-0.009	0.009	0.991
X2	-0.002	0.524	0.998
X3	-0.354	0.042	0.702
X4	0.191	0.000	1.211
X5	0.016	0.171	1.016
X6	0.457	0.170	1.579
X7	0.708	0.031	2.030
Constant	10.849	0.235	51479.723

Based on Table 5, there are differences in the classification prediction results from the two scenarios used. These results are shown by significant predictors of the model. In the 7:3 scenario, three predictors (temperature, length of stay, and disease history) are significant. However, in the 8:2 scenario, four predictors are significant (temperature, length of stay, disease history, and cholesterol levels). The significance of this predictor can be seen from the p-value (less than 0.05) of the predictor. After getting a significant predictor in the prediction of stroke patient data classification, the next step is to form a model for each scenario. The model formed can be seen in the equations (1) dan (2):

The LR model for scenario 7 : 3 can be seen in equation (1)

$$\pi(\boldsymbol{x}) = \frac{exp(20.305 - 0.624X_3 + 0.177X_4 + 0.515X_7)}{1 + exp(20.305 - 0.624X_3 + 0.177X_4 + 0.515X_7)} \quad (1)$$

The LR model for scenario 8: 2 can be seen in equation (2)

$$\pi(\boldsymbol{x}) = \frac{exp(10.849 - 0.009X_1 - 0.354X_3 + 0.191X_4 + 0.708X_7)}{1 + exp(10.849 - 0.009X_1 - 0.354X_3 + 0.191X_4 + 0.708X_7)} \quad (2)$$

The model performance of the two scenarios formed can be seen by modeling the testing data based on the coefficients generated in the classification process with training data. The results of the accuracy of each scenario are given in Table 6.

Table 6. Result of Model Accuracy

Scenario	Training	Testing
7:3	76.20%	73.97%
8:2	76.30%	75.97%

The accuracy shown in Table 6 has a value above 70%. This shows that the classification model formed has a fairly good performance. However, there is a difference in results between scenarios 7:3 and 8:2. Accuracy in scenario 8:2 has a higher result than 7:3. Thus, from the series of analyzes, it can be concluded that scenario 8:2 provides better results in classifying stroke patient data.

From Table 5 and Table 6, the classification prediction used in this study is based on the best results and performance of LR modeling for stroke patients (scenario 8:2). Thus, cholesterol levels, temperature, length of stay, and disease history are predictors that can be interpreted as factors that are thought to affect the type of stroke suffered by the patient. The interpretation of the results of this analysis can be seen from the value of exp (coef.) in Table 5 (scenario 8:2). As an example of a disease history predictor (X_7), the resulting exp (coef.) is 2,030. This value means that if the patient has a history of hypertension/diabetes mellitus/stroke before, they will be twice more likely to have a hemorrhagic stroke than those without a history of the disease.

3.2. Kernel Logistic Regression (KLR)

The performance results using LR to classify the type of stroke suffered by the patient are in the fairly good category. This can be seen from the results of the best data testing accuracy in the 8:2 scenario, which is 75.97%. According to Martin-Baos [20], the performance generated from LR can be optimized using ML methods, namely KLR. KLR is a form of regularized LR by adding kernel functions to the model. This study uses the RBF kernel function.

This classification analysis with KLR is based on the results of LR by involving significant predictors in scenario 8:2, namely cholesterol levels, temperature, length of stay, and disease history. Therefore, the classification model with KLR is applied to scenario 8:2 to get optimum results. The optimization of this classification model is based on the accuracy value obtained from the testing data modeling as shown in Table 7.

Table 7. KLR Accuracy for scenario 8:2

Threshold Value	Accuracy
λ=0.05 and γ=0.05	83.67%
λ=0.01 and γ=0.075	87.12%
λ=0.005 and γ=0.1	89.97%
λ=0.001 and γ=0.5	90.24%
λ=0.0005 and γ=0.75	92.12%

Table 7 provides five examples of accurate results in modeling stroke patient data using KLR based on influencing factors, namely cholesterol levels, temperature, length of stay, and disease history. The highest accuracy results were obtained using λ and γ respectively 0.0005 and 0.75 with an accuracy value of 92.12%. This accuracy is the optimum result of the iteration process using λ between 0.0005 to 0.1 and γ between 0.01 and 1.

Based on the results in Table 7, it can be concluded that the use of KLR in optimizing the accuracy of classification modeling with LR on stroke patient data has been successful. This is indicated by looking at the accuracy value of KLR (92.12%) which is higher than LR (75.97%).

The accuracy of LR is categorized as good enough while KLR is an excellent classification category.

3.3. Discussion

The stroke-type classification model is based on seven predictors. Four of seven predictors are significant, namely cholesterol levels, temperature, length of stay, and disease history. This result is based on the prediction of the classification using the best LR, which is in the 8:2 scenario. The results of this LR performance are included in the fairly good category, with an accuracy of 75.97%. This LR classification model can be optimized using ML methods, namely KLR [14], [15]. Martin-baos et al. [20] stated that KLR was able to improve performance by up to 10% from its predecessor method. This is in line with this research where the performance obtained in the optimum model reaches 16.15%, with the accuracy results based on Table 7 being 92.12%.

Analysis of the classification of stroke types using LR is known that the factors that are thought to influence the different types of stroke are cholesterol levels, temperature, length of stay, and disease history. Of the four predictors that are most closely related to stroke are cholesterol levels and disease history. In stroke patients, cholesterol levels are one of the comorbidities that are considered because high cholesterol levels cause blockage of blood vessels and the occurrence of stroke [21]. Based on research conducted by Pascoe et al. [22], showed that 88% of stroke patients had high cholesterol levels. The results of this study are in line with two previous studies, namely an increase in cholesterol levels in the body that is not controlled will result in a person having a stroke. Aini et al. [23] explained that the average cholesterol level of hemorrhagic stroke patients was lower than ischemic. This is also in line with the results of the research conducted. The resulting coefficient is negative which means that hemorrhagic stroke patients have lower cholesterol levels than ischemic stroke patients.

A history of the disease such as hypertension, diabetes mellitus, or a history of a previous stroke is also something that should be considered for stroke sufferers. Lin et al. [24] stated that premature death of stroke patients was caused by inadequate treatment of hypertension in stroke patients. Of the 2.5 million stroke patients who are screened, one-third of patients have hypertension and are not treated. One-third of the patients had a stroke condition that was more severe than the others. Hypertension is one of the main factors for stroke in young people [25]. The research of Masuda et al. (2022) states that stroke control in young people can be done using strict blood pressure control for severe hypertension. In addition to hypertension, risk factors for diabetes mellitus must be treated early and aggressive management is carried out to reduce the burden of stroke [26].

Diabetes mellitus not only affects the arterial architecture of the blood-brain causing side effects of pathological neovascularization and vasoregression but also changes cerebrovascular function which results in impaired myogenic reactivity and endothelial dysfunction. Diabetes mellitus increases the risk and longevity of stroke and impairs physical and cognitive recovery after stroke [27]. Based on this description, this study can be used as a reference in preventing stroke by paying attention to factors that cause stroke severity, including high cholesterol levels, and a history of diseases such as hypertension and diabetes mellitus. It is necessary early screening of indicators suspected to be the cause of stroke so that it does not occur in severity.

4. CONCLUSIONS

Stroke is a non-communicable disease that is included in the dangerous category because it appears suddenly. This study predicts the classification of stroke types in patients treated at the Stroke Center of RSKD DADI, South Sulawesi Province using LR. LR's performance in classifying the type of stroke is categorized as quite good. The performance of this classification changes to the category of excellent classification when using KLR. The resulting accuracy of the KLR increased by 16.15% from the LR to 92.12%. Factors that are thought to influence the severity of the type of stroke suffered by the patient are cholesterol levels, temperature, length of stay, and disease history. To reduce the severity of a stroke, this study suggests providing intensive care and tight control of cholesterol levels and a history of diseases such as hypertension and diabetes mellitus in cholesterol sufferers. In addition, early screening of these factors is also very necessary for the early prevention of stroke.

ACKNOWLEDGMENTS

Thanks are given to the Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi (Kemdikbudristek) Republic of Indonesia for providing assistance through the PDUPT scheme so that this research can be carried out properly.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] P. Willeit, T. Toell, C. Boehme, et al. STROKE-CARD care to prevent cardiovascular events and improve quality of life after acute ischaemic stroke or TIA: A randomised clinical trial, EClinicalMedicine. 25 (2020), 100476. https://doi.org/10.1016/j.eclinm.2020.100476.

[2] V.L. Feigin, B.A. Stark, C.O. Johnson, et al. Global, regional, and national burden of stroke and its risk factors, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019, Lancet Neurol. 20 (2021), 795–820. https://doi.org/10.1016/s1474-4422(21)00252-0.

[3] M. Naghavi, A.A. Abajobir, C. Abbafati, et al. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016, Lancet. 390 (2017), 1151–1210. https://doi.org/10.1016/s0140-6736(17)32152-9.

[4] G.A. Roth, D. Abate, K.H. Abate, et al. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017, Lancet. 392 (2018), 1736–1788. https://doi.org/10.1016/s0140-6736(18)32203-7.

[5] K. K. R. Indonesia, Hasil Utama Riskesdas 2018, Jakarta, 2018. https://www.litbang.kemkes.go.id/hasil-utama-riskesdas-2018.

[6] G.P. Samsa, J. Bian, J. Lipscomb, et al. Epidemiology of recurrent cerebral infarction, Stroke. 30 (1999), 338–349. https://doi.org/10.1161/01.str.30.2.338.

[7] J. Misbach, W. Ali, Stroke in Indonesia: A first large prospective hospital-based study of acute stroke in 28

hospitals in Indonesia, J. Clinic. Neurosci. 8 (2001), 245–249. https://doi.org/10.1054/jocn.1999.0667.

[8] K. Joy, Ischemic vs. Hemorrhagic stroke: What's the difference?, Michigan Health, 2018. https://healthblog.uofmhealth.org/ischemic-vs-hemorrhagic-stroke-perfcon (accessed Sep. 20, 2022).

[9] R.Y. Fa'rifah, B. Poerwanto, Platelets and hematocrit in the survival model of dengue hemorrhagic fever (DHF) sufferers in palopo, Mater. Sci. Forum 967 (2019) 3–8. https://doi.org/10.4028/www.scientific.net/msf.967.3.

[10] H.G. Jeong, B.J. Kim, T. Kim, et al. Classification of cardioembolic stroke based on a deep neural network using chest radiographs, eBioMedicine. 69 (2021), 103466. https://doi.org/10.1016/j.ebiom.2021.103466.

[11] M.S. Sirsat, E. Fermé, J. Câmara, Machine learning for brain stroke: A review, J. Stroke Cerebrovascular Dis. 29 (2020), 105162. https://doi.org/10.1016/j.jstrokecerebrovasdis.2020.105162.

[12] J.C. Griffis, J.B. Allendorfer, J.P. Szaflarski, Voxel-based Gaussian naïve Bayes classification of ischemic stroke lesions in individual T1-weighted MRI scans, J. Neurosci. Meth. 257 (2016) 97–108. https://doi.org/10.1016/j.jneumeth.2015.09.019.

[13] S. Annas, A. Aswi, M. Abdy, et al. Stroke classification model using logistic regression, J. Phys.: Conf. Ser. 2123 (2021), 012016. https://doi.org/10.1088/1742-6596/2123/1/012016.

[14] A. Al-Aamri, K. Taha, M. Maalouf, et al. Inferring causation in yeast gene association networks with kernel logistic regression, Evol. Bioinform. Online. 16 (2020), 117693432092031. https://doi.org/10.1177/1176934320920310.

[15] M. Maalouf, T.B. Trafalis, I. Adrianto, Kernel logistic regression using truncated Newton method, Comput. Manag. Sci. 8 (2010), 415–428. https://doi.org/10.1007/s10287-010-0128-1.

[16] S. Nusinovici, Y.C. Tham, M.Y. Chak Yan, et al. Logistic regression was as good as machine learning for predicting major chronic diseases, J. Clinic. Epidemiol. 122 (2020), 56–69. https://doi.org/10.1016/j.jclinepi.2020.03.002.

[17] A. Agresti, An introduction to categorical data analysis, Second edition, John Wiley & Sons, Inc., New Jersey, 2007.

[18] M. Ekström, P.A. Esseen, B. Westerlund, et al. Logistic regression for clustered data from environmental monitoring programs, Ecol. Inform. 43 (2018), 165–173. https://doi.org/10.1016/j.ecoinf.2017.10.006.

[19] R. Ott, M.T. Longnecker, An introduction to statistical methods and data analys, 5th Ed. Thomson Learning, Inc. Duxbury, 2015.

[20] J.Á. Martín-Baos, R. García-Ródenas, L. Rodriguez-Benitez, A Python package for performing penalized maximum likelihood estimation of conditional logit models using Kernel Logistic Regression, Transport. Res.

Procedia. 58 (2021), 61–68. https://doi.org/10.1016/j.trpro.2021.11.009.

[21] E. Schwedhelm, S. Tiedt, S. Lezius, et al. Effective high-density lipoprotein cholesterol is associated with carotid intima-media thickness and vascular events after acute ischemic stroke, Atherosclerosis. 357 (2022), 9–13. https://doi.org/10.1016/j.atherosclerosis.2022.08.001.

[22] M.C. Pascoe, C.F. Ski, D.R. Thompson, et al. Serum cholesterol, body mass index and smoking status do not predict long-term cognitive impairment in elderly stroke patients, J. Neurol. Sci. 406 (2019), 116476. https://doi.org/10.1016/j.jns.2019.116476.

[23] A.Q. Aini, L.A. Pujarini, D.D. Nirlawati, The differences of total cholesterol levels between patients with ischemic and hemorragic stroke, Biomedika, 8 (2016), 1–5. https://doi.org/10.23917/biomedika.v8i2.2909.

[24] Q. Lin, T. Ye, P. Ye, et al. Hypertension in stroke survivors and associations with national premature stroke mortality: data for 2·5 million participants from multinational screening campaigns, Lancet Glob. Health. 10 (2022), e1141–e1149. https://doi.org/10.1016/s2214-109x(22)00238-8.

[25] H. Masuda, A. Singla, K. Dinh, et al. An unusual case of malignant hypertension causing stroke in a young adult, Ann. Vascular Surg. - Brief Rep. Innov. 2 (2022), 100052. https://doi.org/10.1016/j.avsurg.2022.100052.

[26] P. Mendys, L.R. Jackson II, E.Z. Solimon, et al. The atrial fibrillation paradox -connecting hypertension to atrial disease and stroke, Amer. J. Prevent. Cardiol. 8 (2021), 100284. https://doi.org/10.1016/j.ajpc.2021.100284..

[27] V. Wolf, Y. Abdul, W. Li, et al. Impact of diabetes and ischemic stroke on the cerebrovasculature: A female perspective, Neurobiol. Dis. 167 (2022), 105667. https://doi.org/10.1016/j.nbd.2022.105667.