Design of a mobile headache detection application with Naïve Bayes classifier method

by Hanik Badriyah Hidayati

Design of a mobile headache detection application with Naïve Bayes classifier method

Cite as: AIP Conference Proceedings 2314, 040005 (2020); https://doi.org/10.1063/5.0035197
Published Online: 09 December 2020

Riries Rulaningtyas, Hanik Badriyah Hidayati, and Ni Putu Desya Esprillia

Design of A Mobile Headache Detection Application with Naïve Bayes Classifier Method

Riries Rulaningtyas[1, a)], Hanik Badriyah Hidayati[2], and Ni Putu Desya Esprillia[1]

[1]*Biomedical Engineering, Department of Physics, Faculty of Science and Technology, Universitas Airlangga, Surabaya, Indonesia*
[2]*Department of Neurology, Faculty of Medicine, Universitas Airlangga, Surabaya, Indonesia*

[a)] Corresponding author: riries-r@fst.unair.ac.id

Abstract. Recently, many people still ignore the dangers of headaches and have not received yet the effective health care. This condition happens because the communities' awareness are still low and lack of knowledge about the type of headache experienced. This study aims to detect type of headache early with the Naive Bayes Classifier on Android. The Naive Bayes Classifier method includes probabilities' calculations in each class of all data (prior), probabilities' features calculations (likelihood) and multiplying of those two probabilities. The highest multiplications values would become the result of detection. The features which were used in headache detection were classified into two, namely red flags and primary headache. The red flags feature would be detected in the first detection, and the primary headache would be detected in the second detection. In the testing process gave accuracy, sensitivity, and specificity at first detection all with 100% values. Whereas the second detection produced 96.67% accuracy, sensitivity of migraine class was 100%, sensitivity of cluster class was 80%, sensitivity of Tension-Type Headache (TTH) class was 100%, specificity of migraine class was 92.86%, specificity of cluster class was 100% and specificity of TTH class was 100%. The results of accuracy, sensitivity, and specificity in this study were proven that the application had a good performance.

INTRODUCTION

Headache is the most commonly experienced complaint in daily life and have a high prevalence [1]. Data from the World Health Organization in 2011 showed that 75% adults in the world (age 18-65 years) have the experience of headache at least once a year [8]. Most of the headaches can be treated, but still many have not been identified or diagnosed and have not yet received optimal handling. At the present, there are still many people with disabilities due to ignore the dangers of headaches and not yet getting effective health care. Based on the survey from the Atlas of Headache in WHO 2011 shows that 50% of headache sufferers in Africa and Southeast Asia do their own treatment or do not check to the doctor, only 10% have been examined by neurologists [2]. This condition now still cannot be handled and still becomes a serious problem in the society. Beside that this condition happens because the awareness of the communities are still low and lack of knowledge about the type of headache [1]. In the initial step in the diagnosis of headache requires red flags for categorizing the type of headache that classified as headache primary or secondary headache [3]. Several previous studies have focused only on studying of primary headaches. The research in [4] did classification of migraine by using several classification methods on the data mining namely K-NN, SVM, Random Forest, and Naive Bayes by using 144 patients data. The Naive Bayes Classifier method on that research was only limited to the classification of migraine. Another Research used the expert system as a tool for early diagnosis of primary headache with the method Naive Bayes Classifier in Python programming language and still was not used in mobile application [5]. In this research tried to use Naïve Bayes Classifier to classify the type of headache by using android programming for mobile application. Mobile applications are becoming a trend in industrial information systems because they have features and design interfaces that can be developed as users' need.

The 2nd International Conference on Physical Instrumentation and Advanced Materials 2019
AIP Conf. Proc. 2314, 040005-1–040005-8; https://doi.org/10.1063/5.0035197
Published by AIP Publishing. 978-0-7354-4056-2/$30.00

METHOD

This study uses the Naive Bayes Classifier method to detect headaches. The steps are as follows.

Data Collection

Data are collected in the data form with the red flags criteria to detect whether the headache is classified as secondary or primary headache. Then, the criteria of primary headache is needed to detect further about the type of primary headache like migraine, cluster or tension-type headache (TTH).

Data Pre-processing

The pre-processing data has the function for data selection. Inappropriate data will be deleted while the complete data will be processed in the calculation. Data which is not feasible is the data which does not have complete information, the data will then be deleted. Decent data is the data which has complete information, which means all features are met.

Interface Design

The interface is a link between user and system interactions. On the interface, there is a Home Dashboard which consists of Splash screen, Detection, and Help. Splash screen and Home Dashboard will be displayed in Fig. 1, First Detection and Second Detection in Fig. 2. The detection is designed with questions according to the features mentioned in Table 1. This question will help the calculation process of Naive Bayes Classifier.

Calculation of Naive Bayes Classifier

The calculation of the Naive Bayes Classifier begins by calculating the prior probability. The prior probability (P (Gi)) is the number of features (Gi) divided by the entire training data (Gs), which is represented using the following equation 1.

$$P(Gi) = \frac{Gi}{Gs} \tag{1}$$

The next step in this calculation is to calculate likelihood. For likelihood the type of categorical data is calculated by obtaining the possibility of the appearance of each feature (Xi) at the appearance of each class (Gk), is represented by equation 2.

$$P(Xi|Gk) = \frac{Xi}{Gk} \tag{2}$$

While the likelihood type of numerical data is calculated by obtaining an average value (μ_{ij}) as well as standard deviations (σ_{ij}) of each class in training data. The equation 3 is the numeric data type as follows.

$$P(Xi|Gk) = \frac{1}{\sqrt{2\pi}\sigma_{ij}} exp^{\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}{}^2}} \tag{3}$$

After getting the prior probability and likelihood values, the overall calculation of Naive Bayes will be calculated with the posterior probability, following the equation 4 and 5.

$$Posterior = \frac{likehood.\,class\ prior}{predictor\ prior} \tag{4}$$

$$P(Xi|Gk) = \frac{P(Xi|Gk)P(Gi)}{P(Gk)} \tag{5}$$

Where $P(Gk)$ is assumed to be constant so that the above equation becomes:

$$P(Xi|Gk) = P(Xi|Gk)P(Gi) \tag{6}$$

Furthermore, to determine the detected results, posterior probability is obtained from the highest ratio of the two posterior probability values. for example, comparing the probability of secondary headaches $(G1)$ and primary headache $(G2)$is represented by equation 7.

$$P(Xi|G1).P(G1) > P(Xi|G2).P(G2) \tag{7}$$

Performance Measurement

This step calculates the accuracy, sensitivity, and specificity of the results of the calculations have been done.

Data Collection

The collection of medical record data was carried out in two places, namely the Section of the Marketing Center and Medical Record of Dr. Soetomo Hospital and the Information and Communication Technology (ICT) of Dr. Soetomo Hospital. The data that are taken consists of secondary headache data and primary headache data in year 2016-2018. The number of secondary headache data which were taken as many as 100 data, the data coms from several diseases that cause secondary headaches, such as bacterial encephalitis, meningioma, Subarachnoid haemorrhage (SAH), Intracerebral Hematoma (ICH), cerebral neoplasm, meningioma encephalitis, cerebral tuberculoma, and cerebral abscess. The primary headache data consist of 100 training data divided into migraine, clusters, and tension-type headache data. Migraine data which are used as training data are 34 data, cluster data are 13 data and TTH data are 53 data.

Data Pre-processing

Pre-processing data are processed the data into categorical data and numerical data. The result of pre-processed data will be presented in Table 1.

TABLE 1. Categorical and Numerical Data

X1	Feature	Data Type	Note
X1	Age > 50 years old	Categorical	Yes No
X2	Body temperature > 37.5 °C	Categorical	Yes No
X3	Neurological deficit (decreased consciousness, thick feeling or weakness in half body part	Categorical	Yes No
X4	Headache (the longer it increases)	Categorical	Yes No
X5	Seizures	Categorical	Yes No
X6	Edemapapil	Categorical	Yes No
X7	Sudden onset	Categorical	Yes No
X8	Medical records (HIV, Ca, TB, tumors, stroke)	Categorical	Yes No
X9	Precipitating factors (headache arises when coughing and straining)	Categorical	Yes No
X10	Gender	Categorical	Female Male
X11	Headache location	Categorical	Unilateral Bilateral
X12	Head pain scale	Categorical	Mild Moderate Severe

X1	Feature	Data Type	Note
X13	Duration	Numerical	
X14	Red/ watery eyes and runny nose	Categorical	Yes No

Calculation of The Naïve Bayes Classifier

The program reads features on each training data, then input from the user will be matched with the features that are already available. When the user input the data, the probabilities will be calculated. The probabilities are consisting of prior probability, likelihood, and posterior probability. Posterior probability is obtained by multiplying prior probabilities and likelihood values. Posterior probability can be found by calculate the prior probability and likelihood value first. The prior probability of first detection and second detection are represented in Table 2.

TABLE 2. Prior Probability of First Detection

	Prior Probability
Secondary Headache	0.5
Primary Headache	0.5

TABLE 3. Prior Probability of Second Detection

	Prior Probability
Migraine	0.34
Cluster	0.13
Tension-type Headache	0.53

The respective values of first detection likelihood for 100 secondary headache data and 100 primary headache data will be presented at Table 4.

TABLE 4. Likelihood of First Detection

Feature	Information	Primary Headache		Secondary Headache	
		Amount (Xi)	Likelihood (Xi/ G1)	Amount (Xi)	Likelihood (Xi/ G2)
Age > 50 years old (X1)	Yes	16	0.16	71	0.71
	No	84	0.84	29	0.29
Body temperature > 37.5 ºC (X2)	Yes	0	0	28	0.28
	No	100	1	72	0.72
Neurological deficit (decreased consciousness, thick feeling or weakness in half body part (X3)	Yes	0	0	99	0.99
	No	100	1	1	0.01
Headache (the longer it increases) (X4)	Yes	0	0	37	0.37
	No	100	1	63	0.63
Seizures (X5)	Yes	0	0	42	0.42
	No	100	1	58	0.58
Edemapapil (X6)	Yes	0	0	8	0.08
	No	100	1	92	0.92
Sudden onset (X7)	Yes	0	0	31	0.31
	No	100	1	69	0.69
Medical records (HIV, Ca, TB, tumors, stroke) (X8)	Yes	0	0	20	0.2
	No	100	1	80	0.8
	Yes	0	0	15	0.15

Feature	Information	Primary Headache		Secondary Headache	
		Amount (Xi)	Likelihood (Xi/ G1)	Amount (Xi)	Likelihood (Xi/ G2)
Precipitating factors (headache arises when coughing and straining) (X9)	No	100	1	85	0.85

The calculation of second detection likelihood are consist of 34 migraine data, 13 cluster data and 53 TTH data are presented in the following Table 5.

TABLE 5. Likelihood of Second Detection

Feature	Information	Migraine		Cluster		TTH	
		Amount	Probability	Amount	Probability	Amount	Probability
Gender (X1)	Woman	24	0.705882	5	0.384615	37	0.698113
	Man	20	0.294118	8	0.615385	16	0.301887
Location (X2)	Unilateral	34	1	13	1	4	0.075472
	Bilateral	0	0	0	0	49	0.924528
Characteristics (X3)	Throbbing	29	0.85	1	0.076923	5	0.09434
	Drilled	4	0.117647	11	0.846154	1	0.018868
	Pressed	1	0.029412	1	0.076923	47	0.886792
Head pain scale (X4)	Mild	1	0.029412	0	0	4	0.075472
	Moderate	12	0.352941	0	0	37	0.698113
	Severe	21	0.617647	13	1	12	0.226415
Red/ watery eyes and runny nose (X5)	Yes	0	0	12	0.923077	1	0.018868
	No	34	1	1	0.076923	52	0.981132

Whereas in numerical features, the average and standard deviation will be calculated in Table 6.

TABLE 6. Average and Standard Deviation of Migraine, Cluster and Tension-type Headache

	Migraine (G1)	Cluster (G2)	Tension-type Headache (G3)
x_i	240	240	240
μ_{ij}	990	165	2743.869
σ_{ij}	529.133	386.296	2612.729

Then the value of prior probability and likelihood will be multiplied to get the value of the posterior probability. The probability of each class is then compared to find the highest value. The highest value will become the result of detection.

Interface Design

This detection is designed in an Android smartphone and presented on each menu as Splash screen and Home Dashboard in Fig. 1, First Detection and Second Detection Fig. 2.

FIGURE 1. Splash screen and Home dashboard

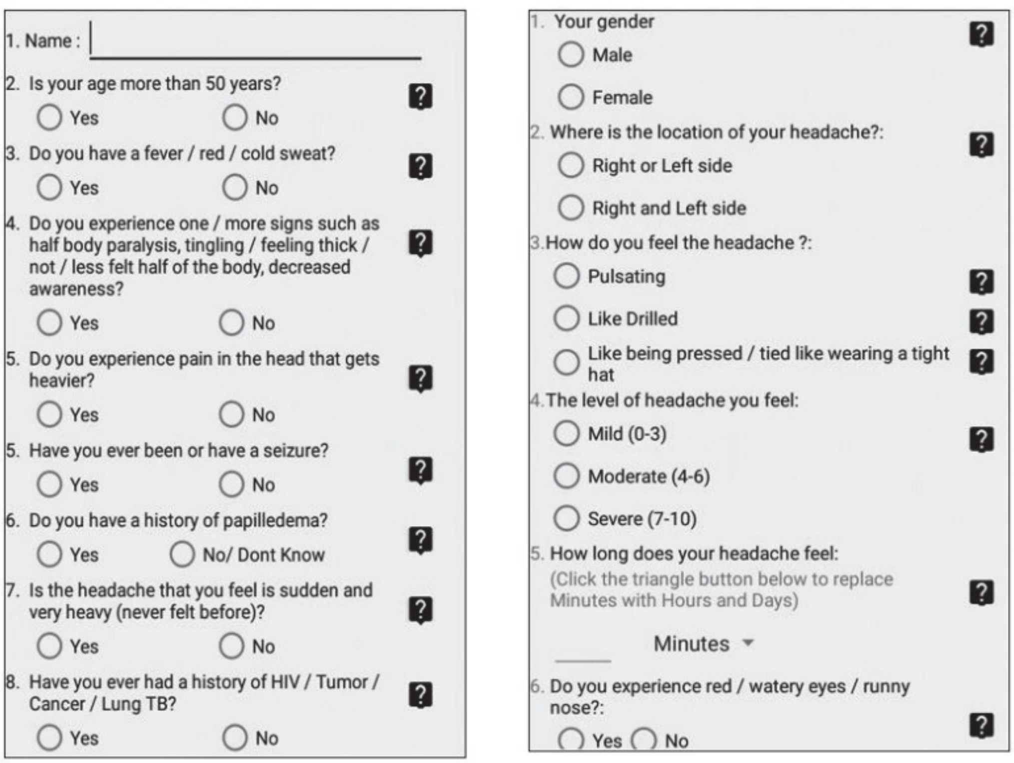

FIGURE 2. First detection and Second Detection

RESULT AND DISCUSSION

Performance Measurement

The program is tested into 50 secondary headache data and 30 primary headache data. Then confusion matrix of the data testing is available in Table 7 and Table 8.

TABLE 7. Matrix Confusion of First Detection

	Classified of Secondary Headache	Classified of Primary Headache
Secondary Headache	100	0
Primary Headache	0	100

TABLE 8. Matric Confusion of Second Detection

	Classified of Migraine	Classified of Cluster	Classified of TTH
Actual of Migraine	13	0	0
Actual of Cluster	1	4	0
Actual of TTH	0	0	12

The results of the calculation of the accuracy, sensitivity and specificity of each data represented in Table 9 and Table 10.

TABLE 9. Accuracy

Type of Headache	Accuracy (%)
Secondary Headache	100%
Primary Headache (Migraine, Cluster and TTH)	96.67%

TABLE 10. Sensitivity and Specificity

Type of Headache	Sensitivity (%)	Specificity (%)
Secondary Headache	100%	100%
Migraine	100%	94.11%
Cluster	80%	100%
TTH	100%	100%

The results of the performance measure at the first detection are 100% accuracy, 100% sensitivity, and 100% specificity by using 50 test data from Dr. Soetomo General Hospital. Whereas in the second detection yield the specificity 94.11% because there is one detection result which should be classified as cluster class but classified as a migraine class. The sensitivity of cluster class is 80% when detecting because the number of cluster data in Indonesia is very small amount [5].

CONCLUSION

The highest accuracy training namely 97.1429% obtained by the texture features SRE, LRE, GLN, RLN, and RPC. The most optimal parameter variations Artificial Neural Network Backpropagation training with accuracy using the number of hidden layer 3, learning rate 1, and maximum MSE of 0.00767216 hence highest level accuracy of retinal fundus image classification system obtained to detect DR using Gray Level Run Length Matrices (GLRLM) feature extraction and Backpropagation is 92.5%. The DR detection program has a sensitivity value of 93.33% and specificity value is 90%.

REFERENCES

1. Vaughan D. Oftalmologi Umum. 7th ed. Jakarta: Penerbit Buku Kedokteran; 2008.
2. Tripathi S, Singh KK, Singh BK, Mehrotra. Automatic Detection of Exudates in Retinal Fundus Images using Differential Morphological Profile. International Journal of Engineering and Technology 2013; 5(3):2024–2029.
3. Putri RAM, Apsari R, Purwanti E. Deteksi DR Melalui Citra Fundus Retina Berbasis Fitur Gray Level Co-occurence Matrices (GLCM) dan Jaringan Syaraf Tiruan [Bachelor thesis]. Universitas Airlangga; 2016. [cited 2018 January 10] Available from: http://repository.unair.ac.id/50600.
4. Saaddine JB, Honeycutt AA, Narayan KMV, Zhang X, Klein R, Boyle JP. Projection of DR and Other Major Eye Diseases Among People With Diabetes Mellitus. American Medical Association 2008; 126(12):1740–1747.
5. International Diabetes Federation (IDF). IDF Diabetes Atlas. 6th ed. Brussels, Belgium: IDF; 2013.

Design of a mobile headache detection application with Naïve Bayes classifier method

ORIGINALITY REPORT

5%	4%	3%	0%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	seap.taylors.edu.my Internet Source	2%
2	Sofis Midori, Dyah Hikmawati, Adri Supardi. "Molecular dynamics study of mechanical properties of Zn – xMg alloy for metal biomaterial", AIP Publishing, 2020 Publication	1%
3	doaj.org Internet Source	1%
4	D Silahudin, Henderi, A Holidin. "Model Expert System for Diagnosis of Covid-19 Using Naïve Bayes Classifier", IOP Conference Series: Materials Science and Engineering, 2020 Publication	1%
5	www.ijitee.org Internet Source	<1%

Exclude quotes	Off	Exclude matches	< 10 words
Exclude bibliography	On		