Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 211−218
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3625

Performance Analysis of LVQ 1 Using Feature Selection Gain Ratio for Sex Classification in Forensic Anthropology

Yulia Harni, Iis Afrianty*, Suwanto Sanjaya, Rahmad Abdillah, Febi Yanto, Fadhilah Syafria

Faculty of Sains and Teknologi, Informatics Engineering, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia
Email: [1]11950121752@students.uin-suska.ac.id, [2,*]iis.afrianty@uin-suska.ac.id, [3]suwantosanjaya@uin-suska.ac.id, [4]rahmad.abdillah@uin-suska.ac.id, [5]febi.yanto@uin-suska.ac.id, [5]fadhilah.syafria@uin-suska.ac.id
Correspondence Author Email: iis.afrianty@uin-suska.ac.id
Submitted: **12/06/2023**; Accepted: **28/06/2023**; Published**: 29/06/2023**

Abstract−One approach to handling large of data dimensions is feature selection. Effective feature selection techniques produce the essential features and can improve classification algorithms. The accuracy performance results can measure the accuracy of the method used in the classification process. This research uses the Learning Vector Quantization (LVQ) 1 method combined with Gain Ratio feature selection. The data used is male and female skull bone measurement data totaling 2524. The highest accuracy results are obtained by LVQ 1, which uses a Gain Ratio with a threshold of 0.01 with a learning rate = 0.1, which is 92.01%, and the default threshold weka(-1.7976931348623157E308) with a learning rate = 0.1, which is 92.19%. In comparison, previous research that did not use gain ratio or that did not use GR only had the best results of 91.39% with a learning rate = 0.1, 0.4, 0.7, 0.9. This shows that LVQ 1 using the Gain Ratio can be recommended to improve the performance of the Skull dataset compared to LVQ 1 without Gain Ratio.

Keywords: Accuracy; Gain Ratio; LVQ 1; Performance; Skull

1. INTRODUCTION

Forensic anthropology is reconstructing a deceased individual's biological profile, i.e., estimating sex, age at death, ancestry, and stature based on skeletal remains[1]. Forensic experts often use bones to determine a person's identity, such as race, sex, estimated age, estimated height, estimated cause of death, and estimated time of death in forensic cases, such as cases of bodies buried with only bone remaining, mutilation cases, and body parts due to natural disaster[2].

The most critical component of forensic anthropology of an individual is sex determination, which is the first step in identifying skeletal remains[3]. Knowledge of the sex of an unknown set of remains is essential for making more accurate age estimates[4]. The skull can be one of the skeleton parts used to determine sex and is the best bone after the pelvic bone[5].

The most common methods used to explore and classify sex are Discriminant Function Analysis (DFA), Logistic Regression (LR), and Support Vector Machine (SVM). Discriminant Function Analysis (DFA) is also an essential statistical method for determining data accuracy and is often used in forensic anthropology analysis[6]. Sex classification is also the most commonly used exploratory task in machine learning (ML), especially using Support Vector Machine (SVM) and artificial neural network (ANN) algorithms[7].

One of the techniques in sex classification is by utilizing one of the Artificial Neural Networks (ANN), namely learning vector quantization (LVQ)[8]. Learning vector quantization or LVQ, is one part of the neural network that performs supervised learning. LVQ is one of the supervised ANN classification algorithms based on the Kohonen mode. LVQ is also called a supervised version of Self Organizing Map networks (SOM) which is considered an unsupervised learning algorithm, LVQ uses a vector quantization architecture along with labeled vectors and supervised training[9]. In its development, the LVQ method is divided into several LVQ: LVQ 1, LVQ 2, LVQ 2.1, and LVQ 3[10].

In the previous research (Darmila et al., 2022), the implementation of LVQ 1, LVQ 2, and LVQ 3 methods on sex classification based on skull measurements got the best accuracy in LVQ 1 method with accuracy reaching 91.39% with learning rate (α) 0.1, 0.4, 0.7, 0.9, subtractor_alfa value 0.01, min_alfa.0.01. Therefore, the previous research became a benchmark to compare the accuracy results in this research with the same classification method, namely LVQ 1[11].

In addition to using LVQ 1 as a classification method, this research also uses feature selection, which is a method used to reduce the impact of dimensionality on data sets by finding feature subsets that efficiently define data and selecting essential and relevant features for mining tasks from input data and removing redundant and irrelevant features and helps detect a good feature subset that is suitable for a given problem[12]. Feature selection aims to build a new dataset from the merging attributes of the feature selection technique[13]. One of the feature selection methods is Gain Ratio, Gain ratio (GR) is a modification of Information Gain[14], [15].

In other research (Pasha & Mohamed, 2022; Safii et al., 2021), it has been stated that the gain ratio feature selection technique is used to rank the features from highest to lowest, which measures the performance based on the average merit of each feature and allocates a rank, where the high-ranked features are considered the most essential feature[16]. Feature selection gain ratio is used to reduce the dimensionality of the collected data, effectively shortening the process's time [17]. Selected features on Gain Ratio can affect accuracy performance[18].Therefore,

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 211−218
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3625

this research uses Gain Ratio (GR) as a feature selection to optimize the LVQ 1 algorithm and see the accuracy value obtained using and not using Gain Ratio (GR) feature selection.

2. RESEARCH METHODOLOGY

2.1 Research Workflow

The following is the flow of this research which is presented in Figure 1:

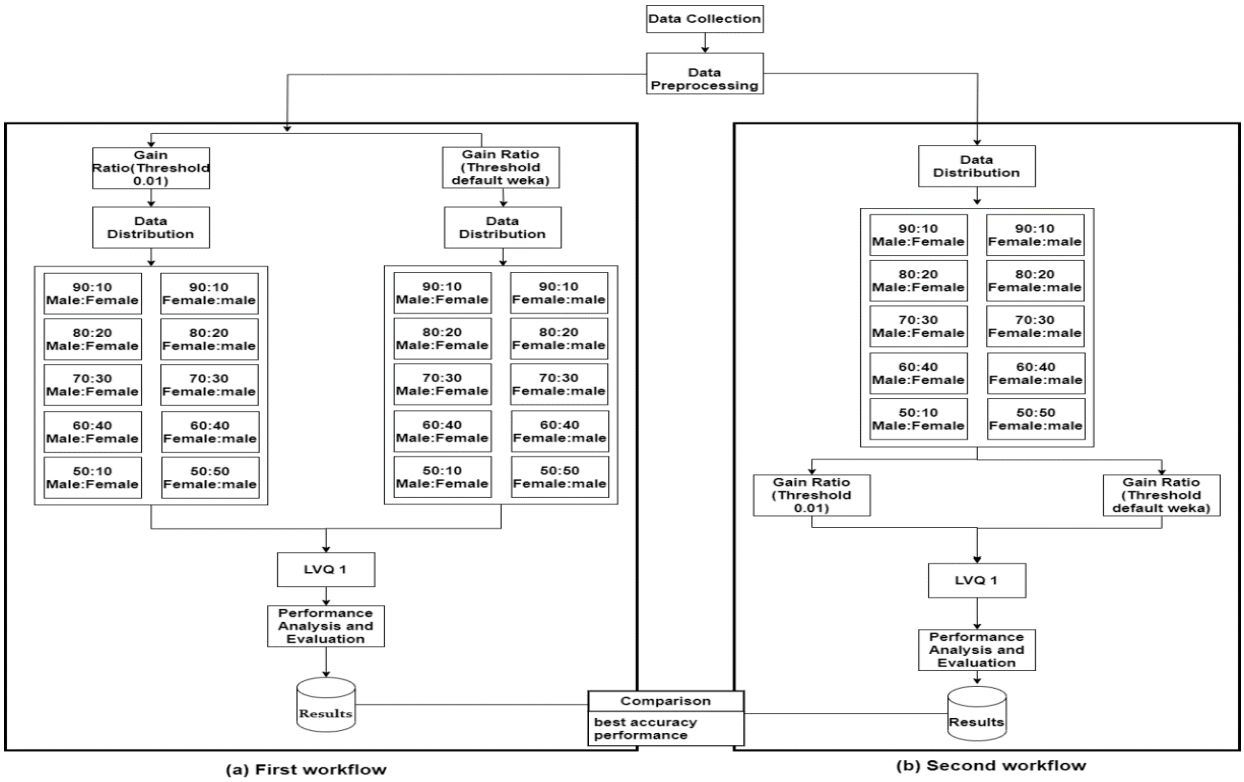

Figure 1. Research workflow

Explanation of the research workflow in Figure 1:

a. The first workflow from the dataset then enters the data preprocessing stage then enters the Gain Ratio feature selection stage with two different thresholds, then from each threshold will select features then from the selected data, features will be distributed into sampling data, then classified with LVQ1 and evaluated.
b. The second workflow from the dataset then enters the preprocessing stage then the data is distributed into sampling from each sampling, the data enters the Gain Ratio feature selection stage with two different thresholds, then from each threshold will select features then from the selected data features will be classified with LVQ1 and evaluated.

The two workflows will get the accuracy value and then compare their performance. Based on the research workflow in Figure 1, the research stages can be described as follows:

a. Data Collection

In this research, the skull dataset is sourced from the William W. Howells dataset, and secondary data from the website https://web.utk.edu/~auerbach/HOWL.htm. Skull bones can determine the sex of humans. In a dataset of 2524 consisting of 1368 male skull data and 1156 female skull data, there are 82 skull bone measurement features.

Table 1. Dataset

Sex	…	ASB	BPL	NPH	NLH	JUB	…	PAS	OCC	OCS	OCF	FOL
M	…	112	96	66	50	118	…	26	98	30	51	34
M	…	113	108	64	48	118	…	24	93	27	39	34
M	…	112	102	67	53	112	…	23	88	30	45	41
M	…	113	95	76	53	114	…	24	94	34	50	38
M	…	111	90	67	51	115	…	26	94	32	40	34
…	…	…	…	…	…	…	…	…	…	…	…	…
F	…	106	103	68	48	110	…	106	103	68	48	110
F	…	101	90	65	48	110	…	101	90	65	48	110
F	…	108	96	59	45	110	…	108	96	59	45	110

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 211−218
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3625

F	…	106	86	57	45	106	…	106	86	57	45	106
F	…	107	95	63	47	113	…	107	95	63	47	113

Table 1 shows the skull bone measurement dataset and Table 2 below offers 12 features out of 82 skull bone measurement features with their respective codes.

Table 2. Skull bone measurement features

Code	Features of Skull Bone Measurements	Code	Features of Skull Bone Measurements
ASB	Biasterionic breath	PAS	Bregma-lamda subtense(parietal subtense)
BPL	Basion-prosthion lengthz	OCC	Lamnda-opisthion chord (Occipital chord)
NPH	Nasion-prosthion height	OCS	Lamnda-opisthion Subtense (Occipital subtense)
NLH	Nasal height	OCF	Occipital-frontal circumference
JUB	Bujugal breath	FOL	Foramen magnum length
NLB	Nasal breadth	OBH	Orbit Height, left

b. Data Preprocessing

1. Data Cleaning
   Removed some features on this dataset such as Popnum and Population. These features are removed because they contain population names and population numbering descriptions.
2. Data Transformation
   Data transformation that converts categorical data to numerical data[19].
3. Data Normalization
   Data normalization can be done with several varied approaches. In this research applying the Min-Max normalization method. Normalization is done by mapping into numbers between 0 and 1 [11], [20].

$$v' = \frac{v - min_A}{max_A - min_A}(new\_max_A - new\_max_A) + new\_min_A \quad (1)$$

c. Feature Selection Gain Ratio

Normalization Feature Selection can help to find the ranking results of each attribute, useful for the process of creating learning models and improving their accuracy [21]. In this research we use Gain Ratio (GR) feature selection, a modification of Information Gain that reduces bias on entropy values. The Gain Ratio determination is as follows[22], [23]:

Calculate the Entropy value for each attribute

$$Entropy\ (S) = \sum_{i=1}^{n} -pi * log_2 pi \quad (2)$$

Calculate the information gain value for each attribute

$$InformationGain(S, A) = Entropy(S) - \sum_{i=1}^{n} \frac{|Si|}{|S|}\ x\ Entropy(Si) \quad (3)$$

Calculate the Split Information value for each attribute

$$SplitInfo\ (D) = -\sum_{j=1}^{v} \frac{Dj}{D}\ x\ log2\ \frac{Dj}{D} \quad (4)$$

Calculate the Gain Ratio value for each attribute

$$GainRatio(A) = \frac{Gain(A)}{SplitInfo(A)} \quad (5)$$

Researchers used the WEKA tool in this research to perform GR feature selection[24]. the output of WEKA is used as the basis for feature ranking[25]. Furthermore, there are 2 thresholds used, namely the default threshold of the Weka application and with a threshold of 0.01[26]. The Threshold selected in the Gain Ratio will select highly relevant characteristics to the data class.

d. Data Distribuition

This step will be divided into 9 samples from each data class. This step will be divided into 9 samples from each data class .The distribution of data used during testing, sampling data after using Gain Ratio and not using Gain RatiIn this dataset with both classes, namely male skull bones and female skull bones divided into each sample (SI - S IX).

e. Classification with LVQ 1

In this step, the required data is modeled according to the steps of LVQ algorithm 1. In the classification process, training and testing processes are carried out using data that has known object classes. LVQ 1 algorithm is the initial basis of the LVQ algorithm, where only the reference vector that is closest to the class distribution will be refreshed.

In learning on LVQ 1 the parameters used are[27]:

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 211−218
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3625

1. Determine the value of the X vector, the target value (T), the weight vector value (Wj), the class value (Cj), the learning rate value (α) with the conditions $0 < \alpha < 1$, the value of reducing or decreasing the learning rate (α), and the minimum learning rate value (min α).
2. Update the weight value if it matches the condition:
   if T=Cj, then it is solved by Eq.

$$Wj(\text{new}) = Wj(\text{old}) + \alpha[Xi - Wj(\text{old})] \quad (6)$$

   if T≠Cj, then it is solved by Eq.

$$Wj(\text{new}) = Wj(\text{old}) - \alpha[Xi - Wj(\text{old})] \quad (7)$$

f. Performance Analysis and Evaluation

This stage aims to evaluate the accuracy performance using LVQ1 by doing Gain Ratio and without doing Gain Ratio.

3. RESULT AND DISCUSSION

3.1 Data Preprocessing

a. Data cleaning

Only removed 3 attributes from the data set, namely population names and population numbering descriptions.

Table 3. Before Cleaning

Sex	PopNum	Population	…	ASB	BPL
M	1	NORSE	…	112	96
M	1	NORSE	…	113	108
M	1	NORSE	…	112	102
F	1	NORSE	…	106	103
F	1	NORSE	…	101	90
F	1	NORSE	…	108	96

Table 4. After Celaning

Sex	…	ASB	BPL
M	…	112	96
M	…	113	108
M	…	112	102
F	…	106	103
F	…	101	90
F	…	108	96

b. Data Transformation

This step changes the sex feature of male to 1 and female to 0.

Table 5. Data Transformation

Sex	…	ASB	BPL
1	…	112	96
1	…	113	108
1	…	112	102
0	…	106	103
0	…	101	90
0	…	108	96

c. Data normalization

Data normalization in this research is done before the data enters the selection feature stage and is also done in the Weka application.

3.2 Feature Selection Gain Ratio

The following is the total features selected from both thresholds from the first workflow in Table 6:

Table 6. Total selected features in the first workflow

Threshold	**Total**
0.01	53
Default Weka (-1.7976931348623157E308)	81

Table 6 shows the total of features selected according to the threshold that has been determined. The features selected at the 0.01 threshold are 53 features and 81 features for the default threshold. Furthermore, the selected features from each threshold will give birth to a new dataset according to the total of features, which will then be distributed into 9 sampling data as in Table 8.

The following features are selected from the 2 thresholds of the second workflow in Table 7.

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 211−218
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3625

Table 7. Total selected features in the second workflow

Threshold	Sample	Total Selected Features	Threshold	Sample	Total Selected Features	Total selection of the same feature
0.01	I	37	Default weka	I	82	37
	II	47	(-	II	82	47
	III	58	1.7976931348623157E308)	III	82	58
	IV	67		IV	82	67
	V	67		V	82	67
	VI	59		VI	82	59
	VII	57		VII	82	57
	VIII	61		VIII	82	61
	IX	59		IX	82	59

In Table 7, the total features by the 0.01 threshold obtained for each sample vary, some are the same and some are different, but the default threshold in Weka gets the same total features in each sample, because no features are selected with this default threshold. The sampling distribution process in the flow is described in Table 8.

3.3 Data Distribution

In the sample I as much as 90% male 10% female. In sample II as much as 10% male 90% female, in sample III as much as 80% male 20% female, in sample IV as much as 20% male 80% female, in sample V as much as 70% male 30% female, in sample VI as much as 30% male 70% female, in sample VII as much as 60% male 40% female, in sample VIII as much as 40% male 60% female, in sample XI as much as 50% male 50% female.
The following Table contains the amount of data based on data distribution:

Table 8. Data Distribution

Class	Sample								
	I	II	III	IV	V	VI	VII	VIII	IX
male	1231	137	1094	274	958	410	821	547	684
Female	116	1040	231	925	347	809	462	694	578

3.4 Classification with LVQ1

William's dataset was divided into training and testing data and then divided into 9 samples. The data is tested using k-fold cross-validation testing using the learning rate (α) value parameters 0.00001, 0.0001, 0.001, 0.01, 0.1, 0.4, 0.7, and 0.9, minimum α 0.01, subtractor α 0.01, and the epoch value used is 1000.

3.5 Performance Analysis and Evaluation of LVQ 1

a. Previous Research Results (LVQ 1 without Gain Ratio)

In LVQ 1 without Gain ratio get the highest accuracy achievement in sampling I which is 91.39% with the value of α = 0.1, 0.4, 0.7, 0.9.

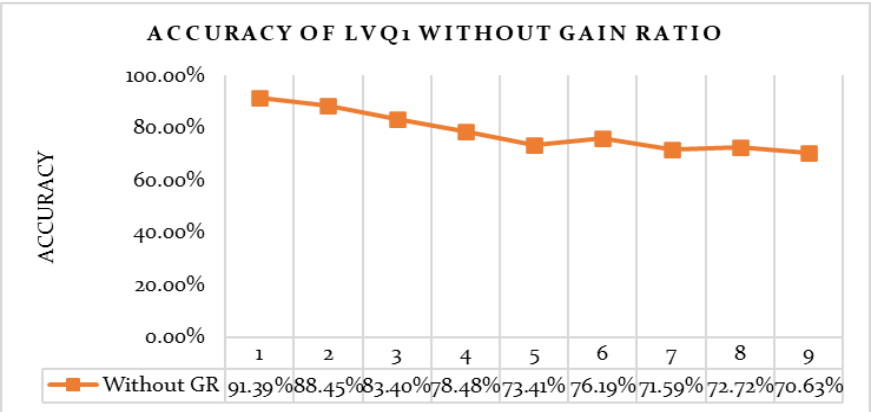

Figure 2. Accurasy Result of LVQ 1 without Gain Ratio

In Figure 2, shows the best accuracy results of the 9 samples tested with LVQ method 1. In sampling II, III, and IV get the best accuracy at α = 0.1. In sampling V, the best accuracy is produced at α = 0.4, 0.7, 0.9. While in sampling VI, VII, VIII, IX achieved the best accuracy at α = 0.01.

b. Best Accuracy Results of LVQ 1 with Gain Ratio with Threshold 0.01 and without Gain Ratio

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 211−218
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3625

The best accuracy results obtained by LVQ1 using Gain Ratio with threshold 0.01 is with the second workflow. The following results compare LVQ 1 accuracy using a feature selection Gain Ratio with threshold 0.01 and those that do not use feature selection in previous research.

In Figure 3, the best accuracy comparison results are obtained in sample II using GR with a value of α = 0.1, which is 92.01%, while in previous research or without GR only has the best result of 88.45% at α = 0.1. In sample I, the best accuracy results using GR was obtained with a value of α = 0.4, 0.7, and 0.9, while in previous research, the best results in sample I were obtained with a value of α = 0.1, 0.4, 0.7, 0.9. In sample III, the accuracy results using GR were obtained at a value of α = 0.4, while in previous research the best accuracy results were obtained at a value of α = 0.1. In sample IV, the accuracy results using GR and the accuracy results in the previous research, the best accuracy results were both obtained at a value of α = 0.1. In sample V, the best accuracy results using GR were obtained at a value of α = 0.001, while in previous research, the best accuracy results were obtained at a value of α = 0.4, 0.7, 0.9. In sample VI the best accuracy results using GR were obtained at a value of α = 0.001, while in previous research, the best accuracy results were obtained at a value of α = 0.01. In samples VII, VIII, IX the best accuracy results using GR and the best accuracy results in previous research were both obtained at a value of α = 0.01.

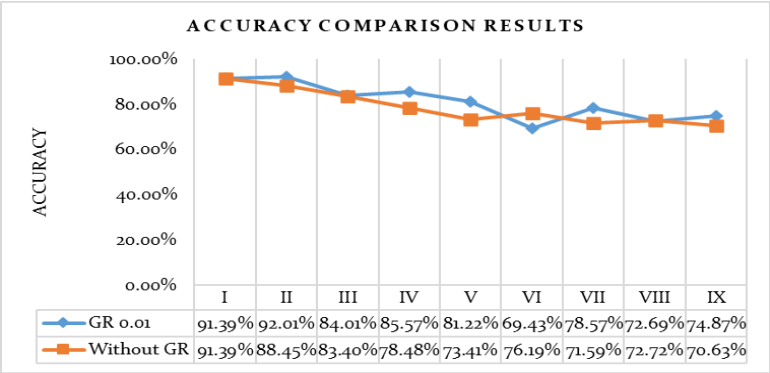

Figure 3. The results of the comparison of the two highest accuracy results in samples I - IX with GR 0.01 and without GR

c. Best Accuracy Results of LVQ 1 with Gain Ratio with Default Threshold and Without Gain Rasio

The best accuracy result obtained by LVQ1 using Gain Ratio with default threshold Weka(-1.7976931348623157E308) is with the first workflow. The following is the comparison result of LVQ 1 accuracy using Gain Ratio feature selection with threshold 0.01, which does not use feature selection in previous research[11]. In Figure 4, the best accuracy comparison results are obtained in sample II using GR with a value of α = 0.1 which is 92.19%, while in previous research or without Gain Ratio only has the best result of 88.45% at α = 0.1. In sample I, the best accuracy results using GR were obtained with α = 0.4, 0.7, and 0.9, while in previous research the best results in sample I were obtained with α = 0.1, 0.4, 0.7, and 0.9. In sample III, the accuracy results using GR were obtained at α = 0.4 and 0.7, while in previous research, accuracy results were obtained at α = 0.1. In sample IV, the accuracy results using GR and the accuracy results in the previous research, the best accuracy results were both obtained at a value of α = 0.1. In sample V, the best accuracy results using GR were obtained at a value of α = 0.01, while in previous research,the best accuracy results were obtained at a value of α = 0.4, 0.7, 0.9. In samples VI and VII, the best accuracy results using GR were obtained at a value of α = 0.001, while in previous research, the best accuracy results were obtained at a value of α = 0.01. In samples VIII and IX the best accuracy results using GR and the best accuracy results in previous research were obtained at a value of α = 0.01.

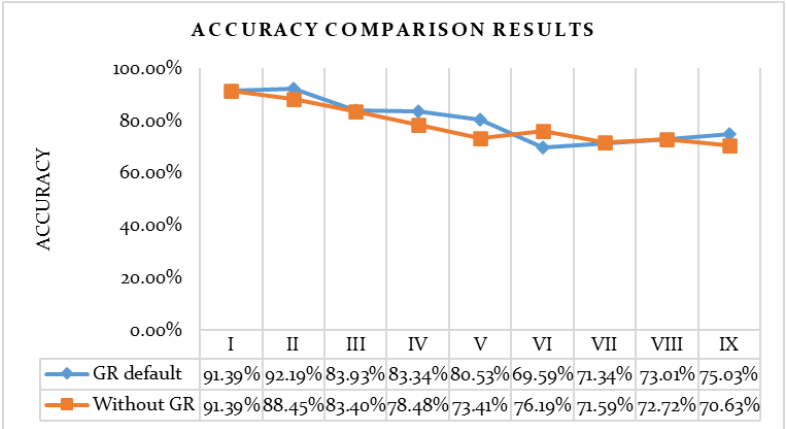

Figure 4. The results of the comparison of the two highest accuracy results in samples I - IX with GR default Weka and without GR

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 211−218
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3625

d. Highest Accuracy Difference Results in Comparison of LVQ 1 with gain ratio with LVQ 1 and without gain ratio

Table 9 shows the difference in the comparison of results between LVQ 1, which uses gain ratio and LVQ 1 which does without gain ratio. It can be seen that the difference in accuracy results in the yellow-colored column shows the accuracy results of LVQ1 using a gain ratio higher than the accuracy results with LVQ1 without gain ratio. In comparison, the white column is the result of LVQ 1 accuracy using gain ratio lower than LVQ 1 that does without gain ratio.

Table 9. The difference result of accuracy comparison

Difference result of accuracy comparison LVQ1 GR threshold 0.01 - LVQ1									
Testing	sample								
	I	II	III	IV	V	VI	VII	VIII	IX
LVQ1 GR threshold 0.01	91.39%	92.01%	84.01%	85.57%	81.22%	69.43%	78.57%	72.69%	74.87%
LVQ 1	91.39%	88.45%	83.40%	78.48%	73.41%	76.19%	71.59%	72.72%	70.63%
Result Accuracy margin	**0.00%**	**3.56%**	**0.61%**	**7.09%**	**7.81%**	**-6.76%**	**6.98%**	**-0.03%**	**4.24%**
Difference result of LVQ1 GR threshold default accuracy comparison - LVQ1									
LVQ1 GR threshold Default weka (-1.7976931348623157E308)	91.39%	92.19%	83.93%	83.34%	80.53%	69.59%	71.34%	73.01%	75.03%
LVQ1	91.39%	88.45%	83.40%	78.48%	73.41%	76.19%	71.59%	72.72%	70.63%
Result Accuracy margin	**0.00%**	**3.74%**	**0.53%**	**4.86%**	**7.12%**	**-6.60%**	**-0.25%**	**0.29%**	**4.40%**

The result of the difference between LVQ 1 using a gain ratio with a threshold of 0.01 and LVQ 1 that does not use a gain ratio, it can be seen that more yellow columns indicate that the accuracy results by LVQ 1 using a gain ratio with a threshold of 0.01 is better than LVQ 1 without a gain ratio with the highest accuracy difference results is in sample V which is 7.81%. The same is the case with LVQ 1, with a default threshold getting the highest accuracy difference than LVQ 1 without Gain Ratio, with the highest accuracy difference in sample V of 7.12%.

4. CONCLUSION

In this research, it can be concluded that the accuracy results obtained by the LVQ 1 algorithm are strongly influenced by the gain ratio feature selection implemented on the data before classification with LVQ 1. As explained earlier, the gain ratio feature selection mechanism is used in two research workflows. Different gain ratio thresholds (0.01 and default weka(-1.7976931348623157E308)) will select features that are relevant to the data so that each threshold will produce a different total of data features, then the new dataset is classified and tested so that the resulting accuracy value is better than the accuracy value obtained by LVQ1 without gain ratio feature selection. Thus, the implementation of gain ratio feature selection is proven to improve the accuracy of LVQ1 algorithm, so gain ratio feature selection can be recommended to handle numerical data with many features. For future research, we can use different thresholds from this research.

REFERENCES

[1] E. Nikita and P. Nikitas, “On the use of machine learning algorithms in forensic anthropology,” Leg Med, vol. 47, Nov. 2020, doi: 10.1016/j.legalmed.2020.101771.

[2] I. Afrianty, D. Nasien, and H. Haron, “Performance Analysis of Support Vector Machine in Sex Classification of The Sacrum Bone in Forensic Anthropology,” JURNAL TEKNIK INFORMATIKA, vol. 15, no. 1, pp. 63–72, Jun. 2022, doi: 10.15408/jti.v15i1.25254.

[3] J. Bewes, A. Low, A. Morphett, F. D. Pate, and M. Henneberg, “Artificial intelligence for sex determination of skeletal remains: Application of a deep learning artificial neural network to human skulls,” J Forensic Leg Med, vol. 62, pp. 40–43, Feb. 2019, doi: 10.1016/j.jflm.2019.01.004.

[4] D. H. Ubelaker and H. Khosrowshahi, “Estimation of age in forensic anthropology: historical perspective and recent methodological advances,” Forensic Sciences Research, vol. 4, no. 1. Taylor and Francis Ltd., pp. 1–9, Jan. 02, 2019. doi: 10.1080/20961790.2018.1549711.

[5] R. G. Arthy and others, “Determination of Sex by Osteometry of Third Metatarsal,” Indian Journal of Forensic Medicine & Toxicology, vol. 14, no. 3, pp. 1–6, 2020.

[6] M. Bozdag et al., “Sex estimation in a modern Turkish population using the clavicle: a computed tomography study,” Australian Journal of Forensic Sciences, vol. 54, no. 2, pp. 187–198, 2022, doi: 10.1080/00450618.2020.1781255.

[7] D. Toneva, S. Nikolova, G. Agre, D. Zlatareva, V. Hadjidekov, and N. Lazarov, “Machine learning approaches for sex estimation using cranial measurements,” Int J Legal Med, vol. 135, no. 3, pp. 951–966, May 2021, doi: 10.1007/s00414-020-02460-4.

[8] T. S. Fatayer and M. N. Azara, “IoT secure communication using ANN classification algorithms,” in Proceedings - 2019 International Conference on Promising Electronic Technologies, ICPET 2019, Institute of Electrical and Electronics Engineers Inc., Oct. 2019, pp. 142–146. doi: 10.1109/ICPET.2019.00033.

[9] A. Dutt and M. A. Ismail, “Can we predict student learning performance from LMS data? A classification approach,” in 3rd International Conference on Current Issues in Education (ICCIE 2018), 2019, pp. 24–29.

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 211−218
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3625

[10] K.-L. Du and M. N. S. Swamy, Neural Networks and Statistical Learning. London: Springer London, 2014. doi: 10.1007/978-1-4471-5571-3.
[11] Darmila, Afrianty Iis, Sanjaya Suwanto, Abdillah Rahmat, Iskandar Iwan, and Syafria Fadhillah, "Evaluasi Perbandingan Performansi LVQ 1, LVQ 2, DAN LVQ 3 Dalam Klasifikasi Jenis Kelamin Menggunakan Tulang Tengkorak," Jurnal Instek, vol. 7, no. 2, 2022, doi: https://doi.org/10.24252/instek.v7i2.32659.
[12] R. Zebari, A. Abdulazeez, D. Zeebaree, D. Zebari, and J. Saeed, "A Comprehensive Review of Dimensionality Reduction Techniques for Feature Selection and Feature Extraction," Journal of Applied Science and Technology Trends, vol. 1, no. 2, pp. 56–70, May 2020, doi: 10.38094/jastt1224.
[13] Ü. Çavuşoğlu, "A new hybrid approach for intrusion detection using machine learning methods," Applied Intelligence, vol. 49, no. 7, pp. 2735–2761, Jul. 2019, doi: 10.1007/s10489-018-01408-x.
[14] B. Pes, "Ensemble feature selection for high-dimensional data: a stability analysis across multiple domains," Neural Comput Appl, vol. 32, no. 10, pp. 5951–5973, May 2020, doi: 10.1007/s00521-019-04082-3.
[15] L. Yu, Y. Cao, C. Zhou, Y. Wang, and Z. Huo, "Landslide susceptibility mapping combining information gain ratio and support vector machines: A case study fromWushan Segment in the Three Gorges Reservoir Area, China," Applied Sciences (Switzerland), vol. 9, no. 22, Nov. 2019, doi: 10.3390/app9224756.
[16] S. J. Pasha and E. S. Mohamed, "Advanced hybrid ensemble gain ratio feature selection model using machine learning for enhanced disease risk prediction," Inform Med Unlocked, vol. 32, Jan. 2022, doi: 10.1016/j.imu.2022.101064.
[17] R.-H. Dong, H.-H. Yan, and Q.-Y. Zhang, "An Intrusion Detection Model for Wireless Sensor Network Based on Information Gain Ratio and Bagging Algorithm," International Journal of Network Security, vol. 22, no. 2, pp. 218–230, 2020, doi: 10.6633/IJNS.202003.
[18] B. Prasetiyo, Alamsyah, M. A. Muslim, and N. Baroroh, "Evaluation of feature selection using information gain and gain ratio on bank marketing classification using naïve bayes," in Journal of Physics: Conference Series, IOP Publishing Ltd, Jun. 2021. doi: 10.1088/1742-6596/1918/4/042153.
[19] C. Fan, M. Chen, X. Wang, J. Wang, and B. Huang, "A Review on Data Preprocessing Techniques Toward Efficient and Reliable Knowledge Discovery From Building Operational Data," Frontiers in Energy Research, vol. 9. Frontiers Media S.A., Mar. 29, 2021. doi: 10.3389/fenrg.2021.652801.
[20] R. Hidayawanti and Y. Latief, "RAW MATERIAL OPTIMIZATION WITH NEURAL NETWORK METHOD IN CONCRETE PRODUCTION ON PRECAST INDUSTRY," International Journal of GEOMATE, vol. 24, no. 102, pp. 10–17, 2023, doi: 10.21660/2023.102.g12146.
[21] I. Safii, M. Kamisutara, T. M. Faahrudin, and S. Data, "Heart Disease Classification using Gain Ratio Feature Selection with Hidden Layer Modification in Extreme Learning Machine," 2021.
[22] Information Technology and Career Education. Hong Kong: CRC Press, 2015. doi: 10.1201/b18416.
[23] N. A. Siagian, S. Wage, and Sawaluddin, "Dataset Weighting Features Using Gain Ratio to Improve Method Accuracy Naïve Bayesian Classification," in IOP Conference Series: Earth and Environmental Science, IOP Publishing Ltd, May 2021. doi: 10.1088/1755-1315/748/1/012034.
[24] B. Geyik, K. Erensoy, and E. Kocyigit, "Detection of Phishing Websites from URLs by using Classification Techniques on WEKA," in Proceedings of the 6th International Conference on Inventive Computation Technologies, ICICT 2021, Institute of Electrical and Electronics Engineers Inc., Jan. 2021, pp. 120–125. doi: 10.1109/ICICT50816.2021.9358642.
[25] D. Stiawan et al., "An Approach for Optimizing Ensemble Intrusion Detection Systems," IEEE Access, vol. 9, pp. 6930–6947, 2021, doi: 10.1109/ACCESS.2020.3046246.
[26] S. Magdy, Y. Abouelseoud, and M. Mikhail, "Efficient spam and phishing emails filtering based on deep learning," Computer Networks, vol. 206, p. 108826, 2022, doi: https://doi.org/10.1016/j.comnet.2022.108826.
[27] X. Lei, T. Shangqin, W. Zhenglei, X. Yongbo, and W. Xiaofei, "UCAV situation assessment method based on C-LSHADE-means and SAE-LVQ," Journal of Systems Engineering and Electronics, pp. 1–17, 2023, doi: 10.23919/JSEE.2023.000062.