Variance: Journal of Statistics and Its Applications
P-ISSN: 2685-8738 E-ISSN: 2685-872X
Variance: Journal of Statistics and Its Applications
Volume 5 Nomor 2 | October 2023
10.30598/variancevol5iss2page185-192

CLASSIFICATION OF THE GEOCHEMICAL COMPOSITION OF METEORITE OF PUNGGUR (ASTOMULYO) BY *k*-NEAREST NEIGHBOR ALGORITHM

Triyana Muliawati[1]*
[1] Department of Mathematics, Institut Teknologi Sumatera, Way Hui, 35352, Lampung, Indonesia

E-mail Corresponding Author: *triyana.muliawati@ma.itera.ac.id*

Abstract: The fall of a meteorite in Astomulyo Village, Punggur, Lampung Province in early 2021 is an interesting topic for further study. This rare object has been suggested to have a unique geochemical composition and a special connection with other meteorites. We aimed to trace its classification by comparing it with other well-known meteorites studied previously. We approach the classification process using the *k*-nearest neighbor algorithm. The database used 211 represents the geochemical data for each known meteorite group from chemical analyses of meteorites. As a result, we identified that with a *k*-value = 5 and the proportion of test data 5/95 (in %), the geochemical composition of this meteorite is relatively close to that of the H-type chondrite group with a value accuracy of 91.67%. These results are consistent with the fact that the meteorite of Punggur has a high total iron and metallic composition.

Keywords: Astomulyo, Geochemistry, *K*-nearest Neighbor, Meteorite

1. INTRODUCTION

Meteorite is a natural outer space object that is attracted by the Earth's gravity and can be found on the Earth's surface in solid form. This object can originate from extraterrestrial objects, such as planetesimal objects, asteroids, or other planetary materials ejected from its orbit. In several historical records, the fall of space objects to the Earth's surface can be destructive, including the mass extinction that occurred 65 million years ago. Researchers are now starting to review these space objects and group them for future use, both for mitigating the dangers of celestial bodies and for exploring outer space minerals, such as asteroids. In this case, it is very important to understand its geochemical composition because researchers can then identify the processes and origins of these celestial bodies, along with their grouping.

A meteorite fell in Indonesia in early 2021. This meteorite was confirmed to have fallen on January 28, 2021, around 09.31 PM and produced a loud bang. This meteorite fell into Astomulyo Village, Kec. Punggur, Lampung Province, Indonesia [1] [2]. The composition of this meteorite has been reviewed and classified as a stony-iron meteorite, with details of 27% SiO_2. This geochemical data has not been compared with geochemical data from other meteorites that have been encountered. In this study, we used the *k*-nearest neighbor algorithm. Because *k*-nearest neighbor is a simple algorithm for classification and the *K* value (number of nearest neighbors used) can be adjusted according to needs.

The *k*-nearest neighbor algorithm is a supervised learning algorithm, which is an algorithm that studies labeled data, which consists of features and labels. The ultimate goal of supervised learning is to identify new data labels with features in new data. The *k*-nearest neighbor algorithm classifies new objects based on their (k) nearest neighbors [3]. The condition for the value of k is that it cannot be greater than the number of training sets, and k must be odd and greater than one. The closest neighbor is seen from the distance or distance of the closest training set to the object to be classified and calculated using the Euclidean Distance method.

In the process of grouping, this algorithm has the advantage of being simple and easy to implement to solve classification problems, and there is no need to build a model and several parameters or make additional assumptions. Many researchers have used this method to determine the quality of rock masses [4], batusatam [5], hulusimpang formation [6], petrified wood [7], classification of igneous rock types [8] [9], classification of volcano status [10] [11], and human age range through tooth fossils [12]. Previous research classified Astomulyo (Punggur) meteorite based on major and minor geochemical compositions. We aim to explore the types of

 Submitted: 2023-09-29 *Accepted:2023-10-22*

Astomulyo (Punggur) meteorites based on their geochemical composition classification by comparing them with other well-known meteorites using the k-nearest neighbor algorithm.

2. METODOLOGY

2.1. Data

The data were obtained from a compilation of sorted stony meteorite geochemical data by Jerosewich [13]. Meteorites marked as weathered were not included in this dataset. In addition, we assumed FeO, Fe_2O_3, FeS, and Fe-metal, together with total Fe or Fe(t). The geochemical compositions involved in this calculation were SiO_2, MgO, Fe(t), TiO_2, Al_2O_3, Cr_2O_3, MnO, CaO, P_2O_5, Ni, and Co. 221 data have been grouped based on the type of chondrite, namely Carbonaceous (C), H Chondrite (H), L Chondrite (L), LL Chondrite (LL), Enstatite Chondrite (E), Achondrite (AC), Anomalous Meteorite (AN), and Mesosiderite Chondrite (Ms) which can be reviewed in detail in [13]. The data were used as a test for the Astomulyo (Punggur) Meteorite, with details in Table 1.

Table 1. Chemical composition of astomulyo (punggur) meteorite [1]

ID	Meteorite	Type	Composition (w/w) %										
			SiO_2	MgO	Fe(t)	TiO_2	Al_2O_3	Cr_2O_3	MnO	CaO	P_2O_5	Ni	Co
0	Astomulyo (Punggur)	Unidentified	38.76	22.60	29.15	0.06	3.70	0.46	0.50	2.51	0.20	0.48	0.09

2.2. *K*-nearest Neighbor Algorithm

The k-nearest neighbor (KNN) algorithm is used for classifying objects based on learning data that are closest to the object. In KNN, several distance rules can be used [14]. The classification search approach used in this study was based on the KNN algorithm, which was performed in three stages [15]. The first stage determines the parameter k or the number of nearest neighbors used as a reference. The value of k should be odd so that decisions can be made better. The second stage involves calculating the squared Euclidean distance [16] for each object with the given object data, following Equation (1). The third stage was performed by sorting the objects in the group with the smallest Euclidean distance.

$$d(x_s, x_p) = \sqrt{\sum_{i=1}^{n} \left(x_{s,i} - x_{p,i}\right)^2} \quad (1)$$

where $d(x_s, x_p)$ is the distance between variables x_s and x_p, x_s is the test data, x_p is the training data, i is the data variable n is the numbers of data.

2.3. Model Evaluation

An evaluation was conducted to measure the formation of the classification model that has been formed. The evaluation used in this study was a confusion matrix [17]. The confusion matrix is a table consisting of the number of rows of actual data (testing set) of the classification results predicted to be correct or incorrect by the classification model. The confusion matrix for classes with 2 × 2 dimensions is presented in Table 2.

Table 2. Confusion matrix

		Predict	
		+	-
Truth	+	∑ TP	∑ FN
	-	∑ FP	∑ TN

where True Positive (TP) is that a lot of data in the actual class is positive, and in the prediction class, it is also positive. False Negative (FN) is a lot of data in the actual class is positive, while the predicted class is negative. False Positive (FP) is a lot of data in the actual class that is a negative class, while the predicted class is positive. TN (True Negative) is a large amount of data in the actual class is a negative class, and the predicted class is also negative.

For detailed testing, the accuracy value was sought. Accuracy evaluation is good enough to help choose the best model by comparing the accuracy levels of various models. Accuracy is a test method based on the degree of closeness between the predicted and actual values as a whole. The accuracy formula is shown in Equation (2).

$$Accuracy = \frac{\sum TP + \sum TN}{\sum TP + \sum FN + \sum FP + \sum TN} \quad (2)$$

2.4. Model Evaluation

Correlation was used to measure the relationship between the results of an observation [18]. Correlation provides information about whether there is a relationship between two variables, as well as measuring how strong and direction the relationship is. The correlation used in this study is Pearson Correlation. In the correlation distance, the points are considered as a series of values, that is, the distance between the vector values x_s and x_p, using Equations (3) and (4).

$$d(x_s, x_p) = 1 - \frac{(x_s - \overline{x_s})(x_p - \overline{x_p})}{\sqrt{(x_s - \overline{x_s})(x_s - \overline{x_s})}\sqrt{(x_p - \overline{x_p})(x_p - \overline{x_p})}} \quad (3)$$

wheres

$$\overline{x_s} = \frac{1}{n}\sum_j x_{s,j} \text{ and } \overline{x_p} = \frac{1}{n}\sum_j x_{p,j} \quad (4)$$

This process was performed using Google Collaboratory (Python programming language facility from Google).

3. RESULTS AND DISCUSSION

The KNN algorithm was applied to the compilation data for meteorite geochemical composition ([13] and Table 1) using Equation (1). Then, the confusion matrix (Equation (2)) is calculated to see how correct the type class is with the classification that has been used. From the results of the confusion matrix (Table 3), the accuracy value (Equation (3)) was calculated and used as a comparison to determine the optimal *k* for this study. In Table 3 it can be seen that the accuracy value is 0.92 or 92%.

Table 3. Result of confusion matrix and accuracy value

Mt	Ac	C	H	L	LL	Ms	An	En
Ac	1	0	0	0	0	0		
C	0	4	0	0	0	0		
H	1	0	3	0	0	0		
L	1	0	0	1	0	0		
LL	0	0	0	0	1	0		
Ms	0	0	0	0	0	1		
An								
En								

⇨

Mt	Pr	Rc	f 1	Sp
Ac	0.50	1.00	0.67	1
C	1.00	1.00	1.00	4
H	1.00	0.75	0.86	4
L	1.00	1.00	1.00	1
LL	1.00	1.00	1.00	1
Ms	1.00	1.00	1.00	1
An				
En				
Accuracy			**0.92**	12
Macro avg	0.92	0.96	0.92	12
Weighted avg	0.96	0.92	0.92	12

Note: Mt: Meteorite type; C: C-Carbonaceus; H: H-Chondrite; L: L-Chondrite; LL: LL-Chondrite; E: Enstatite; An: Anomalous Chondrite; Ac: Achondrite; Ms: Mesosiderite. Pr: Precision; Rc: Recall; f 1: f 1-score; Sp: Support.

In this study, the k values in the KNN algorithm were used, namely 3, 5, 7, 9, 11. The learning model of the KNN algorithm was carried out 10 times for each k = 3, 5, 7, 9, 11. Comparison of testing data with training data with ratios of 5%/ 95%, 15%/85%, 25%/75%, and 35%/65%. Then, the accuracy is calculated, and the average value is taken from 10 iterations of the learning model. The average results are shown in Figure 1.

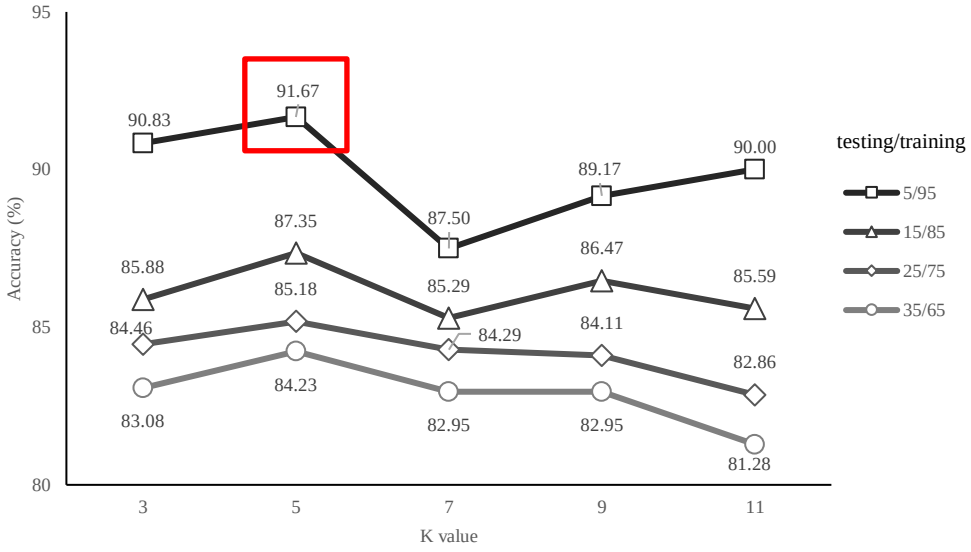

Figure 1. Graph of optimum *K* value vs % accuracy, based on testing/training data ratio

From Figure 1, the highest accuracy results were obtained, namely 91.67% at k = 5, with a ratio of testing data to training data of 5%/95%. Because k is chosen as 5, the five closest neighboring meteorites to the Astomulyo (Punggur) meteorite are calculated based on their closest distance. The distance results are sorted from largest to smallest; the smallest distance means that the closest distance is from the Astomulyo (Punggur) meteorite, whereas the largest distance is the farthest distance from the Astomulyo (Punggur) meteorite. The order of distances is shown in Figure 2.

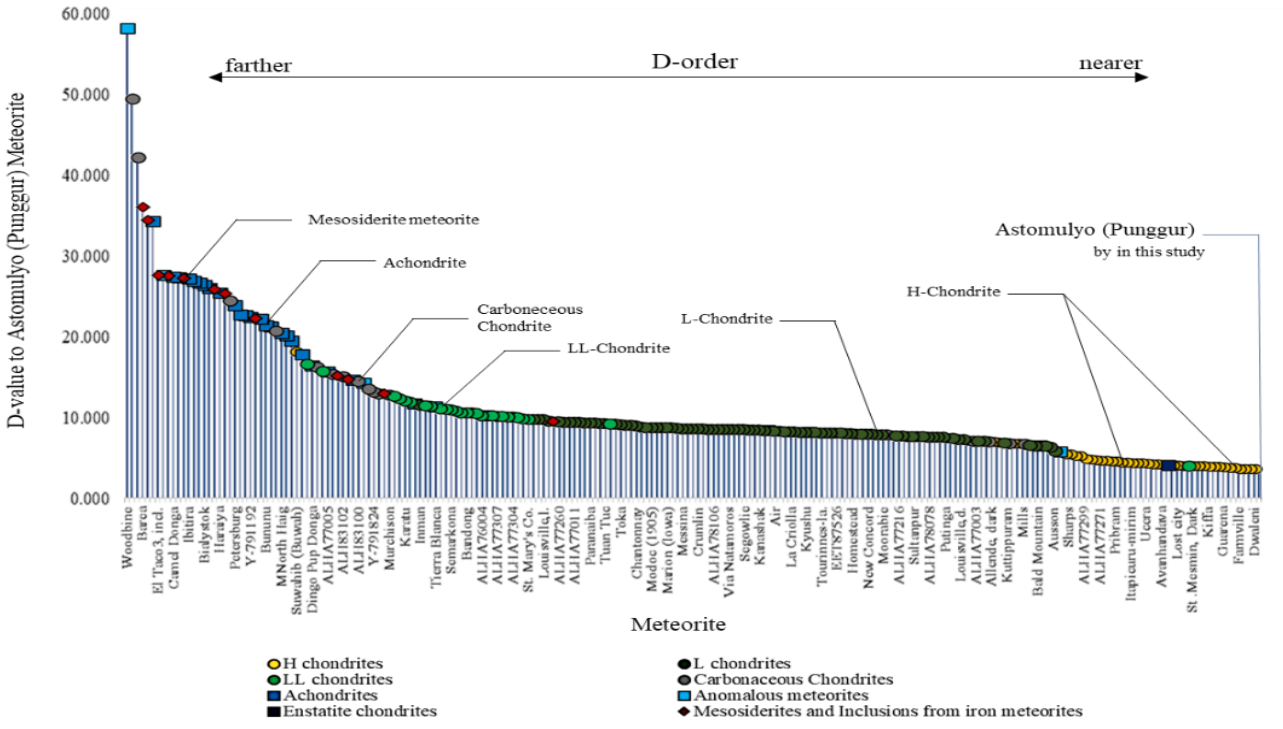

Figure 2. D-value and D-order of the other meteorite to Astomulyo (Punggur) meteorite

From Figure 2, the five meteorites closest to the Astomulyo (Punggur) meteorite are the Marilia, Dwaleni, Forest Vale, Allegan, and Farmville meteorites. Of the five meteorites, all five were meteorites with type H-Chondrites. Therefore, we can classify the astomulyo meteorite as a meteorite of the same type, namely, type H-Chondrites.

Correlation matrices were constructed to examine the characteristics of the level and direction of the relationship between the geochemical composition of each meteorite. The interval between the correlation values ranged from 0 to 1. The closer the correlation value is to 1, the stronger the level of relationship between the variables; the closer the correlation value is to 0, the weaker is the level of relationship between the variables. The direction of the relationship between each geochemical composition can be seen as positive (+) or negative (-). A positive sign (+) means that if one variable increases, the other variable also increases, or vice versa. The negative sign (-) indicates that if one variable decreases, the other variable increases, and vice versa. The results of the meteorite geochemical composition correlation matrix are listed in Table 4.

Table 4. Correlation value

Element	SiO_2	MgO	Fe(t)	TiO_2	Al_2O_3	Cr_2O_3	MnO	CaO	P_2O_5	Ni	Co
SiO_2	1.00	0.11	-0.68	0.33	0.23	0.38	0.73	0.25	-0.09	-0.58	-0.41
MgO	0.11	1.00	-0.24	-0.54	-0.62	0.17	-0.12	-0.59	-0.15	-0.18	0.03
Fe(t)	-0.68	-0.24	1.00	-0.38	-0.38	-0.19	-0.33	-0.41	0.16	**0.88**	0.71
TiO_2	0.33	-0.54	-0.38	1.00	0.78	-0.06	0.27	**0.84**	0.08	-0.33	-0.40
Al2O_3	0.23	-0.62	-0.38	0.78	1.00	-0.16	0.09	**0.98**	-0.07	-0.27	-0.38
Cr_2O_3	0.38	0.17	-0.19	-0.06	-0.16	1.00	0.42	-0.11	0.01	-0.23	-0.16
MnO	0.73	-0.12	-0.33	0.27	0.09	0.42	1.00	0.12	0.03	-0.33	-0.26
CaO	0.25	-0.59	-0.41	0.84	0.98	-0.11	0.12	1.00	-0.03	-0.31	-0.42
P_2O_5	-0.09	-0.15	0.16	0.08	-0.07	0.01	0.03	-0.03	1.00	0.06	0.05
Ni	-0.58	-0.18	0.88	-0.33	-0.27	-0.23	-0.33	-0.31	0.06	1.00	**0.81**
Co	-0.41	0.03	0.71	-0.40	-0.38	-0.16	-0.26	-0.42	0.05	0.81	1.00

From Table 4, there are four values with the highest correlation that have a very strong relationship: the correlation between Al_2O_3 and CaO is 0.98, Fe(t) and Ni is 0.88, TiO_2 and CaO is 0.84, and Ni and Co is 0.81. The direction of the relationship between the geochemical compositions of Al_2O_3 and CaO, Fe(t) with Ni, TiO2 with CaO, and Ni with Co is positive (+), which means that if one of the geochemical compounds increases, the other geochemical compounds increase or if one of the geochemical compounds decreases, the other geochemical compounds decrease. This process is illustrated in Figure 3.

Meteorite classification is the basic framework of meteorites, and cosmochemistry experts work and communicate well [19]. This process was designed to gather new ideas regarding the relationship between meteorites. Meteorite classification was performed to group the meteorites based on their constituent components. Grouping was performed to group the meteorites based on the level of similarity in each meteorite according to the meteorite constituent components. Each meteorite had a different geochemical composition. Based on the various meteorites found, it turns out that some do not match the existing meteorite groups.

Meteorites are classified based on their type, class, and sub-class. This proves the existence of diversity in the solar systems. Each chondrite is distinguished by the body from which the meteorite originated, which is an important issue for meteorites. In this study, it can be seen that the chemical composition of the Astomulyo (Punggur) meteorite based on the *k*-nearest Neighbor algorithm has a close/similarity to the H-Chondrite meteorite. This is reviewed based on the similarity of the composition of SiO_2, MgO, Fe(t), TiO_2, Al_2O_3, Cr_2O_3, MnO, CaO, P_2O_5, Ni, and Co.

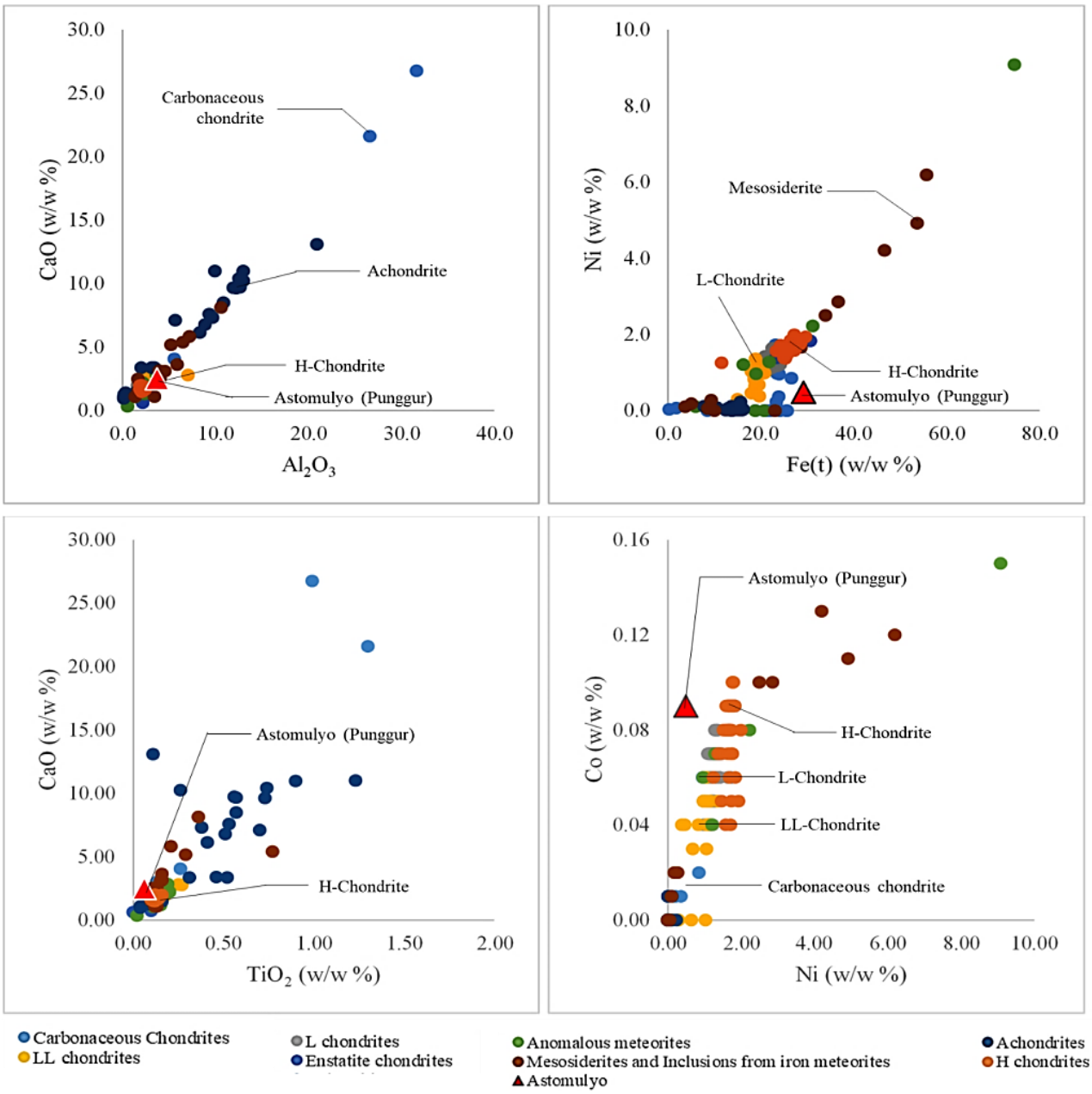

Figure 3. Scatter plot of the highest correlation and the position of Astomulyo (Punggur) meteorite.

4. CONCLUSIONS

It was found that with a value of $k = 5$ and a proportion of test data of 5%/95%, the geochemical composition of this meteorite is relatively close to this group, with a value of 91.67%. The results show that the back meteorite is of the H-Chondrites type. Type H-chondrite has a characteristic geochemical composition of Al_2O_3, CaO, Fe(t), Ni, TiO_2, and Co.

ACKNOWLEDGMENTS

The author is very grateful to everyone for their helpful comments, corrections, and suggestions that helped improve the presentation and quality of this work.

REFERENCES

[1] D. G. Harbowo, R. Muztaba, H. L. Malasan, S. Sumardi, L. K. Agustina, T. Julian, J. H. Sitorus, A. D. A. Denhi, D. J. P. Sihombing, M. P. Mahayu and D. Setyawan, "Meteorite from Astomulyo Village, Central Lampung, Indonesia: Investigation of its chemical properties," in *IOP Conference Series: Earth and Environmental Science*, 2021.

[2] D. G. Harbowo, J. H. Sitorus, L. K. Agustina, R. Muztaba, T. Julian and H. L. Malasan, "3D Modelling of Meteorite from Astomulyo Village, Lampung, Indonesia by Close Range Photogrammetry (CRP) Methods," in *IOP Conference Series: Earth and Environmental Science*, 2022.

[3] F. Gurunescu, Data Mining: Concepts, Models and Techniques, Berlin: Springer, 2011.

[4] S. Rahmadanti, Implementation of the K-nearest Neighbor Method in Determining Rock Mass Quality, Riau, 2020.

[5] D. G. Harbowo, M. Afdareva, V. Inggrid and S. Sumardi, "Batusatam physical and chemical properties review: A Billitonite tektite in Southeastern Belitung Island, Indonesia," in *IOP Conference Series: Earth and Environmental Science*, 2021.

[6] D. G. Harbowo, B. Priadi, T. Julian, R. N. Amelia, D. . J. P. Sihombing and F. S. Kencana, "A preliminary study on the element abundance in the Hulusimpang Formation, Way Kalianda, Pesawaran, Lampung, Indonesia," in *IOP Conference Series: Earth and Environmental Science*, 2021.

[7] T. Muliawati, D. G. Harbowo, A. M. F. Lubis, J. D. Turnip, E. R. Irda, A. Azahra and Y. Marito, "k-Means Clustering to Enhance the Petrified Wood Composition Data Analyses and Its Interpretation," *Indonesian Journal of Applied Mathematics*, vol. 3, no. 1, pp. 26-33, 2023.

[8] M. I. Sultoni, B. Hidayat and A. S. Subandrio, "Classification of igneous rock types through color imagery using local binary pattern and k-nearest neighbor methods," *imagery using local binary pattern and k-nearest neighbor methods Telecommunications, Control, Computers, Electrical, and Electronics,* vol. 4, no. 1, pp. 10-15, 2019.

[9] S. N. Wibowo, B. Hidayat and J. Arif, "Identification of Igneous Rock Types Looking at Rock Texture Using the Discrete Wavelet Transform (dwt) and K-nearest Neighbor (knn) Methods," in *eProceedings of Engineering*, 2017.

[10] F. C. Anggian, N. Hidayat and M. T. Furqon, "Technology Development and Computer Science," *Journal of Information Technology Development and Computer Science Classification of Volcano Status,* vol. 3, no. 12, pp. 11027-11033, 2020.

[11] F. Tempola, M. Muhammad and A. Khairan, "Comparison of Classification Between KNN and Naive Bayes in Determining Volcano Status with K-Fold Cross Validation," *Journal of Information Technology and Computer Science,* vol. 5, no. 5, pp. 577-584, 2018.

[12] H. J. Kuen, B. Hidayat, and J. Arif, "Identification of Human Age Range Through Molar Tooth Fossils Based on Digital Image Processing Using the Discrete Wavelet Transform and K-nearest Neighbor (k-nn) Methods," in *eProceedings of Engineering*, 2018.

[13] E. Jarosewich, "Chemical analyses of meteorites: A compilation of stony and iron meteorite analyses," *Meteoritics,* vol. 25, no. 4, pp. 323-337, 1990.

[14] P. Cunningham and S. J. Delany, "k-Nearest Neighbour Classifiers: 2nd Edition (with Python examples)," 2020.

[15] M. Kuhkan, "A method to improve the accuracy of k-nearest neighbor algorithm," *International Journal of Computer Engineering and Information Technology,* vol. 8, no. 6, p. 90, 2016.

[16] K. Yu, G. D. Guo, J. Li, and S. Lin, "Quantum algorithms for similarity measurement based on Euclidean distance," *International Journal of Theoretical Physics,* vol. 59, pp. 3134-3144, 2020.

[17] J. T. Townsend, "Theoretical analysis of an alphabetic confusion matrix," *Perception & Psychophysics,* vol. 9, pp. 40-50, 1971.

[18] K. R. Godfrey, "Correlation methods," *Automatica,* vol. 16, no. 5, pp. 527-534, 1980.

[19] M. K. Weisberg, T. J. McCoy, and A. N. Krot, "Systematics and evaluation of meteorite classification," *Meteorites and the early solar system II,* vol. 19, pp. 19-52, 2006.