Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2023, 2023:15
https://doi.org/10.28919/cmbn/7843
ISSN: 2052-2541

APPLICATION OF LASSO FOR IDENTIFICATION OF FUNCTIONAL GROUPS WITH SIGNIFICANT CONTRIBUTIONS TO ANTIOXIDANT ACTIVITIES OF *CENTELLA ASIATICA*

C. WIRDIASTUTI[1], U. D. SYAFITRI[1,2,*], I M. SUMERTAJAYA[1], E. ROHAETI[2,3], M. RAFI[2,3]

[1]Department of Statistics, IPB University, Bogor, Indonesia

[2]Tropical Biopharmaca Research Center, IPB University, Bogor, Indonesia

[3]Department of Chemistry, IPB University, Bogor, Indonesia

Abstract: High-dimensional data has more variables than observations (p>>n). In this case, modeling with regression analysis becomes ineffective because it will violate the multicollinearity assumption. The least absolute shrinkage and selection operator (LASSO) can handle high-dimensional data and multicollinearity because LASSO works by reducing the parameters of variables with significant effects and selecting variables with minor effects. In its application, several variables have the same characteristics. Reducing and selecting variables in the form of groups can solve the problem so that the group LASSO can be used as a solution. This study used data on antioxidant activity in *C. asiatica*. It is a plant that contains antioxidants. The spectroscopic technique can find important information about antioxidants, namely the Fourier transformed infrared spectrophotometer (FTIR). FTIR is a spectroscopic technique based on molecular vibrations subjected to infrared so that it can characterize molecules with functional groups. FTIR data has large dimensions and multicollinearity. This study has 1866 explanatory

---

[*]Corresponding author
E-mail address: utamids@apps.ipb.ac.id
Received November 28, 2022

variables (p) and 15 predictor variables (n). So, this study aimed to implement LASSO and the group LASSO to identify functional groups that major affect the antioxidant. This study concluded that group LASSO was better than the LASSO with modification of the LAR algorithm in identifying functional groups that had a major contribution to antioxidant activity. The results showed that the functional groups that had a major effect on the antioxidant activity of *C. asiatica* were –NH –OH, and C –O. In general, the functional groups that had a major effect on the antioxidant activity of *C. asiatica* came from phenolic compounds.

Keywords: antioxidant; FTIR spectra; high dimensional data; LASSO; *Centella asiatica*.

2020 AMS Subject Classification: 92B05.

1. INTRODUCTION

High-dimensional data has more explanatory variables than observations (p>>n). In handling this data, regression analysis modeling becomes ineffective because the assumption of multicollinearity will be violated [1]. Multicollinearity in regression analysis causes the estimation of model parameters to be biased and difficult to interpret with many analyzed variables. [2] developed LASSO by adding the norm l_1 penalty function to the regression model to overcome this. LASSO works by reducing the parameters of a variable with a significant effect and selecting a variable with a small impact.

LASSO can solve high-dimensional data problems. It is because LASSO can reduce the regression coefficient of variables that have a high correlation to error, with the aim that the regression coefficient is close to zero or equal to zero so that LASSO acts as a variable selection as well overcoming multicollinearity [3]. LASSO does not have an explicit solution for determining its estimated coefficients, so computational programming is needed to solve it [4]. One of the algorithms that are effective in helping to solve the LASSO regression solution computationally is the Least Angle Regression Selection (LARS) algorithm.

This study used data on antioxidant activity in *C. asiatica*. It is one of the plants that can be used as a natural antioxidant. [5] concluded that the methanol extract of the *C. asiatica* herb has 393 antioxidant effects and protects and reduces DNA damage. The antioxidant activity of *C.*

asiatica can be influenced by differences in the area of origin, climate (temperature, humidity, light, and wind), and geographical conditions [6]. Find out crucial information about the antioxidant activity of the *C. asiatica* plant and it can be done by using spectroscopic techniques. Fourier-transformed infrared spectrophotometer (FTIR) is a spectroscopy based on molecular vibrations subjected to infrared to characterize molecules with functional groups.

In general, spectroscopic data has large dimensions, and there are multicollinearity problems because spectroscopic data has more independent variables than the number of observations ($p>>n$) [7]. In this study, there were 1866 wave numbers as explanatory variables and 15 observational data, and there was a multicollinearity problem, so statistical techniques were needed to deal with this problem. Several previous studies have discussed this, such as a study that applied the Support Vector Machine method for classifying six *Zingiberaceae* plants using selected variables from the genetic algorithm [8]. [9] applies the partial least square (PLS) for finding functional groups having a major contribution to the antioxidant activity of *Syzygium polyanthum*. Previous research on the correlation between metabolites profiles and bioactivities of *Curcuma aeruginosa* has applied principle component analysis and PLS [10]. A study to determine the performance of PLS and LASSO regression on microarray data concluded that LASSO performance was better than PLS regression [11].

In this study, we used the LASSO method to predict the functional group of the antioxidant compound using FTIR spectra of *C. asiatica* because it follows high-dimensional data. In addition, LASSO could also reduce the regression coefficient of variables with high correlation error, with the aim that the regression coefficient is close to zero or equal to zero so that the LASSO method acts as a variable selection as well to overcomes multicollinearity [3]. LASSO works by adding constraints to the least-squares method. [12] applied the LASSO method to high-dimensional data. They found that the regression coefficients generated from the LASSO method were more capable of selecting explanatory variables than the multiple regression method. The LASSO method does not have a straightforward solution for determining the estimated coefficient, so computational programming is needed to solve it [4]. One of the algorithms that are effective in helping to solve the LASSO regression solution computationally is the LARS algorithm.

In addition, the advanced method of LASSO, namely group LASSO, is also used in this study because group LASSO is usually used for grouped data. FTIR data tend to be clustered, this is based on the theory that two compounds provide an infrared absorption peak at the exact location, so it can be said that the two compounds are identical [13]. Research conducted by [14] applied the group LASSO to high-dimensional FTIR data with 1798 wave numbers as explanatory variables and 280 observational data, concluding that the group LASSO provides a high model accuracy in handling high-dimensional data and contains multicollinearity. Therefore, in this study, LASSO and group LASSO are used to identify functional groups that contribute majorly to the antioxidant activity of *C. asiatica*.

2. LITERATURE REVIEW

2.1 *Centella Asiatica*

C. asiatica is one of the plants that are easy to grow in tropical and subtropical areas [15]. *C. asiatica* is a plant easily found in Indonesia and used as an herbal plant. Based on previous research, *C. asiatica* has many benefits and properties related to antimicrobial, antioxidant, wound healing, anti-inflammatory, and anticancer activities [16]. *C. asiatica* contains polyphenols, flavonoids, carotene, tannin, vitamin C, and triterpenoids (asiaticoside) which have antioxidant activity [17]. Based on previous research, methanol extract and water extract from *C. asiatica* can protect DNA damage [5].

2.2 Antioxidant Activity

Antioxidants play an essential role in protecting cells from reactions caused by free radicals because antioxidants will capture free radicals [18]. Free radicals are compounds that contain unpaired electrons in their orbits so that they are very reactive with surrounding molecules, and free radicals will cause damage to lipids, proteins, and nucleic acids. If the number of free radicals is more than antioxidants in the body, it can cause oxidative stress. Oxidative stress plays an essential role in several diseases, such as atherosclerosis, chronic kidney failure, diabetes mellitus, cancer, premature aging, cardiovascular disease, and neurological diseases [19].

Antioxidants are compounds that can slow the oxidation process of biological molecules or suppress free radicals [20]. Compounds that have antioxidant effects are phenols and polyphenols, and the most common are flavonoids (flavonols, isoflavones, flavones, catechins, and flavanones), cinnamic acid derivatives, coumarins, tocopherols, and polyfunctional organic acids [21].

2.3 Fourier transformed infrared spectrophotometer (FTIR)

The FTIR spectrophotometer allows for simultaneously measuring infrared absorption at various wavelengths [14]. Infrared absorption in a specific wavenumber region can be used to determine the functional groups formed. An organic compound can be identified from the functional groups contained in it. Different combinations provide different functional groups and infrared absorption forms [14]. If two compounds provide infrared absorption peaks at the exact location, they can be said to be identical [13]. The spectral data produced by the FTIR is quantitative data which generally has large dimensions and contains multicollinearity because the spectroscopic data have several independent variables (p) more significant than the number of observations ($p>>n$) [7].

2.4 LASSO

Tibshirani introduced the least absolute shrinkage and selection operator (LASSO) method in 1996. The coefficient estimator in the LASSO method is obtained by minimizing the following Eq. (1) [2].

$$\sum_{j=1}^{N}\left(y_i-\beta_0-\sum_{i=1}^{p}X_{ij}\beta_i\right)^2+\lambda\sum_{i=1}^{p}|\beta_i| \tag{1}$$

with penalty $\sum_{i=1}^{p}|\beta_i|\leq t$. The value of t is the quantity that controls the shrinkage of the coefficient estimator with $t\geq 0$, p is the number of explanatory variables, and N is the number of observations. The LASSO estimator is obtained by specifying a standardized limit, namely $s = t/\sum_{i=1}^{p}\left|\hat{\beta}_i^0\right|$ with $t=\sum_{i=1}^{p}\left|\beta_i^0\right|$ while $\hat{\beta}_i^0$ is the solution of the least squares estimator. If $t <$

$t_0 = \sum_{i=1}^{p} |\hat{\beta}_i^0|$ it will cause $\hat{\beta}_{LASSO}$ to shrink to zero or exactly zero. If $t \geq t_0 = \sum_{i=1}^{p} |\hat{\beta}_i^0|$ then the LASSO estimator is the same result as the least squares estimator ($\beta_{LASSO} = \hat{\beta}_i^0$). This causes LASSO to form an efficient model by maintaining the variables that affect the model. The solution of the LASSO method does not have an explicit solution because the constraint function of the LASSO method is in the form of an absolute function that cannot be derived at the inflection point [4]. The coefficient estimator cannot be obtained in closed form but must use quadratic programming [2].

After its first publication in 1996, this LASSO paper did not receive attention until 2002, after Hastie developed the Least Angle Regression (LAR) algorithm [23]. Modification of LAR for LASSO results in algorithm efficiency in estimating the LASSO coefficient estimator solution with faster computations than quadratic programming. The original LAR algorithm is as follows [24] :

1. Standardize the independent variable to have a median value of zero and a variance of one. Start with residual $r = y - \bar{y},\ \beta_1, \beta_2, \dots\ , \beta_p = 0$.
2. Find the independent variable x_j that is most correlated with r.
3. Change the value of β_j from 0, moving towards the least squares coefficient $\langle x_j, r\rangle$, until another competitor x_k has a correlation x_j equal to the correlation with the current remainder.
4. Change the values of β_j and β_k moving in the direction defined by the joint least squares coefficient of the current residual in (x_j, x_k) until another competitor (x_l) correlates with the current residual of the same magnitude.
5. Continue this way until all p independent variables have been entered. After the min(*N-1*,*p*) steps, the model solution for OLS is obtained.

Modification of the LAR algorithm to get the LASSO solution is to modify the 4th step, namely by:

4a. If the non-zero coefficient reaches zero, remove the variable from the active variables and recalculate the OLS direction together.

2.5 Group LASSO

The Group Least Absolute Shrinkage and Selection Operator (Group LASSO) method is a regression technique that applies LASSO selection to the explanatory variables in groups. This grouping aims to facilitate selecting explanatory variables with similar characteristics. Group LASSO allows variables to group in large numbers with uneven numbers [25]. The coefficient estimator in the group LASSO is obtained by minimizing the following sum of squares of the remainder as Eq. (2) [26]:

$$\frac{1}{2}\left\|\boldsymbol{Y}-\sum_{j=1}^{k}\boldsymbol{X_j}\boldsymbol{\beta_j}\right\|^2+\lambda\sum_{j=1}^{k}\sqrt{p_j}\left\|\boldsymbol{\beta_j}\right\| \quad (2)$$

where k is the number of groups, $\boldsymbol{X_j}$ is the j$^{\text{th}}$ explanatory variable, $\boldsymbol{\beta_j}$ is the j$^{\text{th}}$ regression coefficient, and p_j is the number of variables in the j$^{\text{th}}$ group. While $\lambda \geq 0$ is a controller of the amount of depreciation. The model will be in standard form when $\lambda = 0$. If the λ value is getting bigger, then the coefficient's estimated value will be smaller towards zero for going to infinity.

3. RESEARCH METHODOLOGY

3.1 Data Sources

The data used is the result of FTIR measurements to see the antioxidant content of the *C. asiatica*. This data results from an experiment by the Biopharmaceutical Study Center Team in collaboration with the Statistics Department of IPB in research conducted by Putri [27]. The observations used were obtained by extracting the *C. asiatica* with five extract materials, namely water, 30% ethanol, 50% ethanol, 70% ethanol, and p.a ethanol for each extract material. Experiments were carried out with three replications to produce 15 objects of observation. Furthermore, measurements were carried out using an FTIR spectrophotometer, resulting in 1866 absorption points. The explanatory variable (X) used is wave number. While the response variable (Y) used is the antioxidant activity of the *C. asiatica* extract.

3.2 Procedure Of Analysis

Data processing begins with exploring the data by standardizing the value of the explanatory variable and the response variable to have a mean of zero and a variance of 1. Then calculate the correlation value between the explanatory variables. The second step is to apply the modified LARS algorithm for LASSO to the data used. This step begins with determining each iteration's LASSO regression coefficient estimator, selecting the best model using a leave one out cross validation (LOOCV), and identifying functional groups that affect antioxidant activity with LASSO. The third step is to apply the group LASSO method. At this step, it is necessary to classify the variables based on chemical properties. Next, determine the best model by looking at the lambda that provides the minimum residual cross-validation (CVE) value and identifies the functional groups that affect antioxidant activity with the group LASSO. The fourth step is to identify the functional groups that contribute to the antioxidant activity of *C. asiatica*. The last step is to determine the size of the goodness of the LASSO analysis model and the group LASSO.

4. RESULTS

4.1 Data exploration

FTIR spectroscopy is important in fingerprint analysis because it can display different spectral patterns for each active compound [28]. This study used spectral data from FTIR measurements on *C. asiatica* plants. Each extract material was experimented with three times so that each color in Figure 1 represents the FTIR spectrum for each extract material in each repetition.

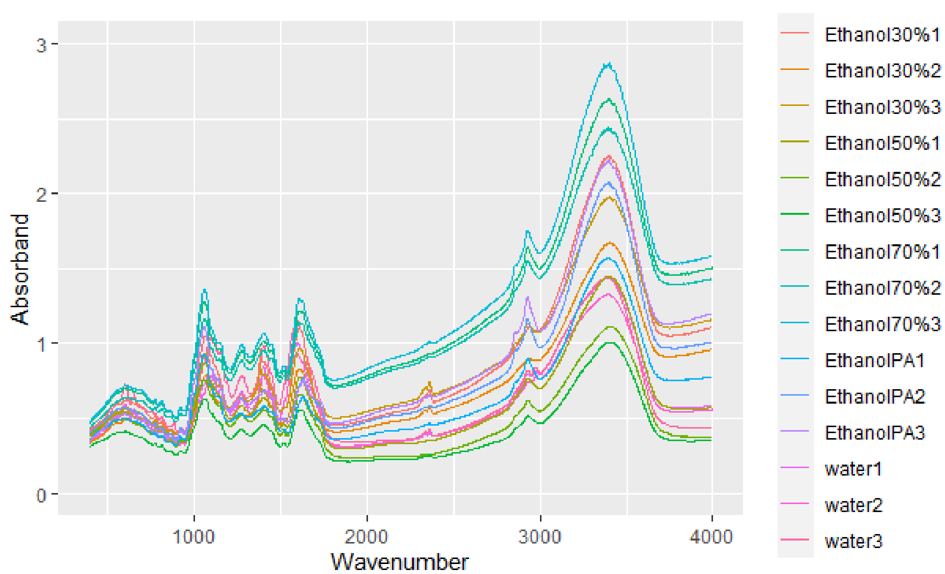

Figure 1. FTIR Spectrum of *C. asiatica* extract before standardization

The resulting FTIR spectrum is in the range of wave numbers between 399.2373 cm^{-1} to 3996.232 cm^{-1}. Figure 1 and 2 shows the infrared absorption intensity, which is not much difference between each spectrum of each extract material for each repetition. The infrared absorption causes a relatively similar pattern for the peaks or valleys formed in each spectrum at a specific interval.

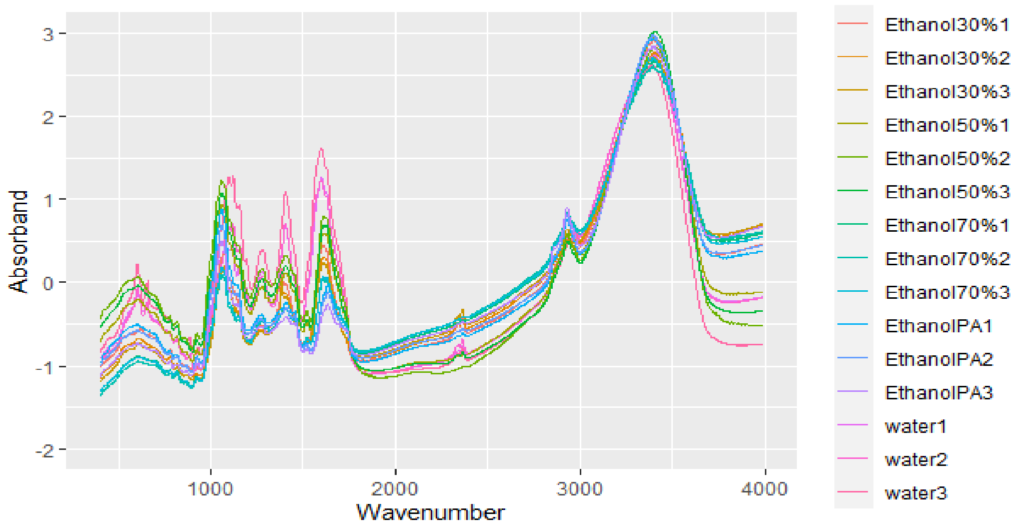

Figure 2. FTIR Spectrum of *C. asiatica* extract after standardization

Data exploration is carried out by standardizing the values of the explanatory variables and the response variables to have a mean of zero and a variance of 1 like in Figure 2. The next step is calculating the correlation between the explanatory variables of the formed wave numbers. The next step is to calculate the correlation between the explanatory variables. In Figure 3, the higher the correlation between independent variables, the stronger the color formed. Figure 3 is dominated by dark blue, meaning there is a multicollinearity problem in the data used.

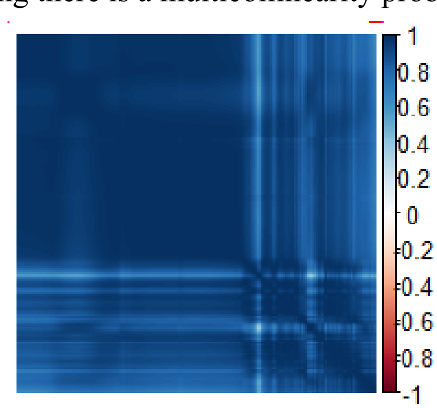

Figure 3. Correlation between explanatory variables

4.2 LASSO Analysis

The LASSO coefficient estimator is obtained through computation by modifying the LAR algorithm, commonly called LARS, to produce an algorithm that is more efficient than quadratic programming. Estimating the LASSO coefficient is carried out in iteration by setting all initial coefficients to be zero. Then, the independent variable that is most correlated with error is entered into the model. The plot of the stages of the independent variables that enter the model can be seen in Figure 4.

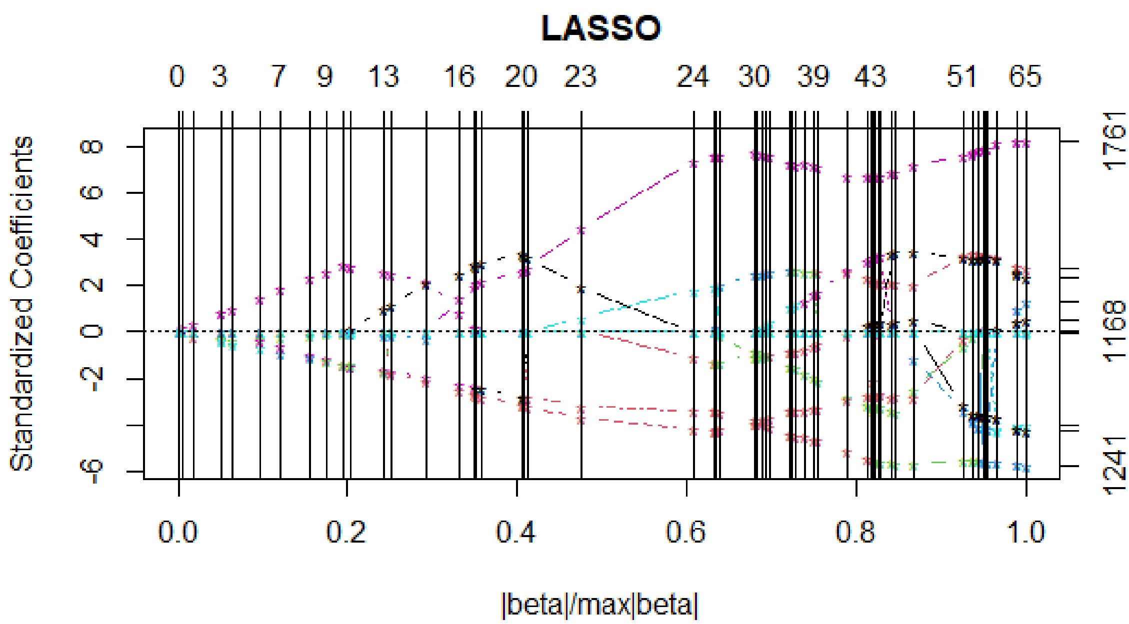

Figure 4. Plot the estimated LASSO regression coefficient with the LARS algorithm at each iteration

The plot of the estimated LASSO regression coefficients with the LARS algorithm represents the LASSO regression coefficients at each step of the LARS algorithm plotted at $s = \frac{t}{\sum\left|\hat{\beta}_j{}^{OLS}\right|}$ where $t = \sum\left|\hat{\beta}_j{}^{LASSO}\right|$. After all the independent variables are selected, the next step is to choose the best model. In Figure 4, the variable X_{1490} is included in the model in the first iteration, meaning that the variable X_{1490} has the highest correlation with the remainder compared to the other variables, with s around 0.005. Furthermore, in the second iteration, X_{1145} is entered into the model. In the second iteration, the model has two variables, namely X_{1490} and X_{1145}, with s around 0.0183. And so on until the variable X_{1168} enters the model at stage 65.

The selection of the best model is made by using cross-validation. In the package, the LARS algorithm uses fraction mode to see at what iteration it produces the best LASSO regression model. The optimal value of s can show this. The minimum cross-validation value indicates the optimal s value. The value of cross-validation can be seen in Figure 5.

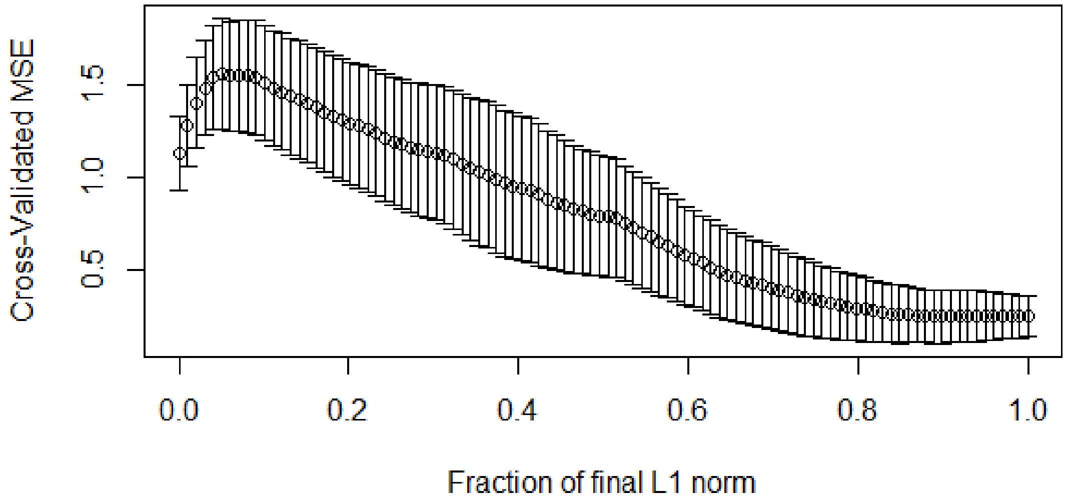

Figure 5. Cross-validation value using fraction mode

Figure 5 shows that the optimal value of s is around 0,8699. The value of s is in the 50th iteration, as shown in Figure 4. Therefore, the best LASSO model selected in this data is the model in the 50th iteration. At this iteration, eight variables enter the model, namely X_{1863}, X_{1761}, X_{1408}, X_{313}, X_1, X_{1240}, X_{1274}, and X_{1462}. This variable is the wave number 405.0234; 601.7493; 1282.574; 3394.483; 3996.232; 1606.593; 1541.018; 1178.425 cm^{-1}

After knowing the best model, the next step is to find the functional groups that affect the antioxidant activity by paying attention to Table 1 [29]. Based on Table 1, the functional groups that affect the antioxidant activity of *C. asiatica* are -NH, -OH, and C-O.

Table 1. FTIR table

Wavenumber (cm^{-1})	Functional groups
3100-2800	$-CH$
3600-3200 and 1500-1300	$-OH$
1200-1000	$C-O$
3500-3100 and 1640-1550	$-NH$

4.3 Group LASSO Analysis

Group LASSO analysis was applied to data with grouped explanatory variables. Therefore, it is necessary to group the explanatory variables in the data to be used based on their characteristic equations. This grouping is done by looking at the peaks of the FTIR spectrum. The explanatory variables formed in one wave's crest are considered in the same group. This is based on the theory that if two compounds provide an infrared absorption peak at the same location, it can be said that the two compounds are identical [13]. The grouping of explanatory variables can be seen in Figure 6.

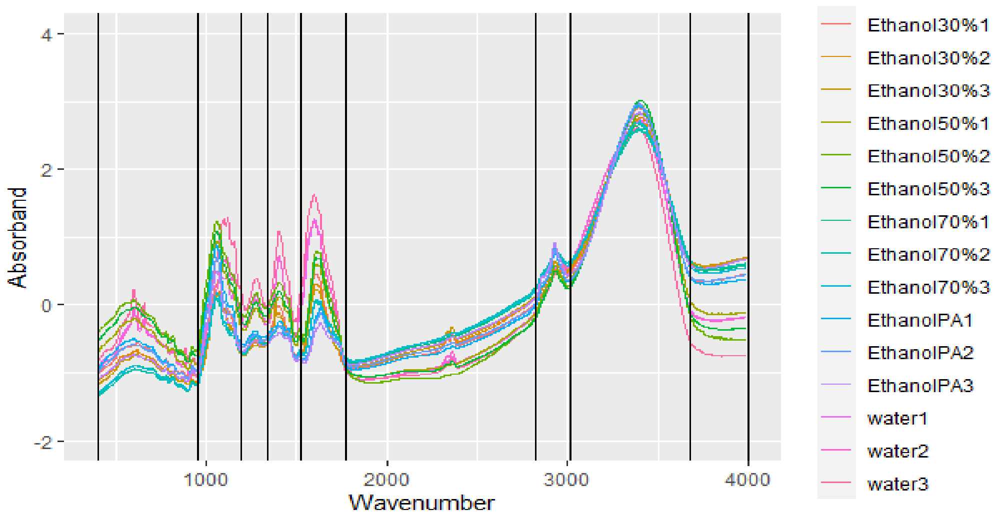

Figure 6. Classification of explanatory variables based on chemical properties

Based on Figure 6, data has a relatively similar pattern for the peaks or valleys formed by each spectrum at a specific interval. Thus, the explanatory variables can be grouped in the same way of grouping, as shown in Table 2.

Table 2. Classification of explanatory variables based on chemical properties

Group	Wavenumber (cm^{-1})
1	399 - 955
2	955 - 1192
3	1192 - 1343
4	1343 - 1524
5	1524 - 1773
6	1773 - 2818
7	2818 - 3011
8	3011 - 3600
9	3600 - 4000

In the group LASSO, the explanatory variable group with a non-zero coefficient is a variable that can explain the effect of *C. asiatica*'s antioxidant activity like LASSO. Based on this, the groups that affect the antioxidant activity of *C. asiatica* are shown in Figure 7.

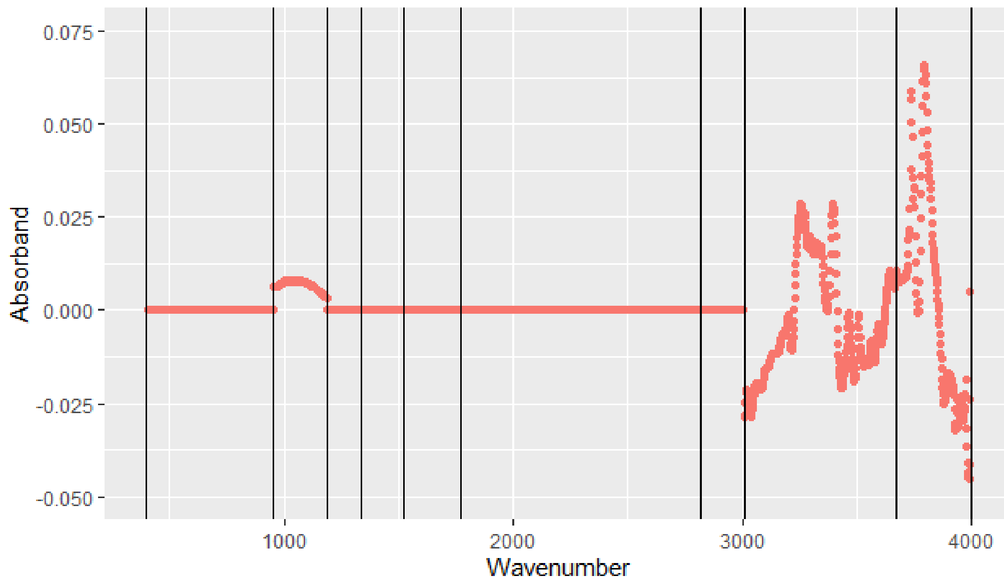

Figure 7. Group LASSO coefficient plots

Figure 7 shows that the groups of variables that affect the antioxidant activity of *C. asiatica* are groups 2, 8, and 9, with wave number intervals of 955 – 1192 cm^{-1}, 3011 – 3600cm^{-1}, and 3600 – 4000 cm^{-1}. Therefore, Table 1 shows that the functional groups that affect antioxidant activity are –NH, –OH, C–O, and –CH**.**

4.4 Identification of Functional Groups Contribute to the Antioxidant Activity

The functional groups that made a major contribution to this research were those identified using LASSO and the group LASSO. Based on Table 3, the functional groups that have a major effect on the antioxidant activity of *C. asiatica* are –NH and –OH, and C – O.

Table 3. The functional groups that affect the antioxidant activity of *C. asiatica*

Analysis Used	Functional groups
LASSO	–NH, –OH, C – O
Group LASSO	–NH, –OH, C – O, –CH

Table 3 shows that fewer functional groups are identified using LASSO than group LASSO. This difference is caused by differences in the model selection process carried out on LASSO and the group LASSO. In the LASSO analysis, the variables included in the model are individual variables, while the variables included in the group LASSO model are in the form of groups. In the group LASSO, if one or more variables in a group have a high correlation to the error, then all the variables in that group are included in the model so that the functional groups identified using the group LASSO are more than the group LASSO.

In general, it can be seen that metabolites with functional groups –NH, –OH, and C – O are thought to be metabolites that contribute to the major antioxidant activity of *C. asiatica* extract. The functional groups identified are the cumulative result of several compounds that act as antioxidants. Based on these results, it can be assumed that the functional groups identified are derived from phenolic compounds. This result follows the research conducted by [30] that phenolic compounds are the most significant contributor to antioxidant activity.

4.5 The Measure Of Goodness

The measure of goodness used is Mean Squared Error (MSE). MSE is the average value of the square of the error. MSE was used to compare the ability between the LASSO analysis and the group LASSO for high-dimensional data and the explanatory variables formed groups. MSE indicates the size of the error, so the smaller the MSE value, the better. Measures of the goodness of these two models are in Table 4. Based on MSE, the group LASSO is better than the LASSO with modification of the LAR algorithm in modeling functional groups that have a major contribution to antioxidant activity.

Table 4. Measures of the goodness

Analysis Used	MSE
LASSO	0.9373
Group LASSO	0.2703

5. CONCLUSION

This study concluded that group LASSO was better than the LASSO with modification of the LAR algorithm in identifying functional groups that had a major contribution to antioxidant activity. The metabolites having –NH and –OH and C –O functional groups were suspected as metabolites that contributed to the major antioxidant activity of *C. asiatica* extract. The functional groups identified are derived from phenolic compounds, the most significant contributor to antioxidant activity.

ACKNOWLEDGEMENTS

The authors gratefully acknowledged the Ministry of Education, Culture, Research and Technology of the Republic of Indonesia through the National Competitive Basic Research Grant for funding and supporting this research (No: 3753/IT3.L1/PT.01.03/P/B/2022).

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] M. Yunus, A. Saefudin, A.M. Soleh, Characteristics of group LASSO in handling high correlated data, Appl. Math. Sci. 11 (2017), 953–961. https://doi.org/10.12988/ams.2017.7276.

[2] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc.: Ser. B (Methodol.) 58 (1996), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

[3] I.T. Jolliffe, Principal component analysis, Second Edition, Springer-Verlag, New York, (2002).

[4] M. Robbani, F. Agustiani, N. Herrhyanto, Regresi least absolute shrinkage and selection operator (lasso) pada kasus inflasi di indonesia tahun 2014-2017, Jurnal EurekaMatika 7 (2019), 1-16.

[5] T. Anand, N. Mahadeva, K.G. Phani, Antioxidant and DNA damage preventive properties of Centella asiatica (L) Urb, Pharmacognosy J. 2 (2010), 53-58.

[6] M. Rafi, Quality control methods for some Zingiberaceous plants from Indonesia using liquid chromatography combined with chemometrics, Thesis, Gifu University, (2013).

[7] A. Risqa, Pemodelan regresi komponen utama fungsional pada data spektroskopi, Thesis, Institut Pertanian Bogor, (2016).

[8] T. Oktaria, U.D. Syafitri, M. Rafi, et al. Metode support vector machine (SVM) untuk klasifikasi enam tumbuhan zingiberaceae menggunakan variabel terpilih hasil algoritma genetika, Thesis, Institut Pertanian Bogor, (2020).

[9] E. Rohaeti, F. Karunina, M. Rafi, FTIR-based fingerprinting and chemometrics for rapid investigation of antioxidant activity from syzygium polyanthum extracts, Indones. J. Chem, 21 (2021), 128–136. https://doi.org/10.22146/ijc.54577.

[10] D.A. Septaningsih, L.K. Darusman, F.M. Afendi, R. Heryanto, Liquid chromatography mass spectrometry (LC-MS) fingerprint combined with chemometrics for identification of metabolites content and biological activities of curcuma aeruginosa, Indones. J. Chem. 18 (2018), 43–52.

[11] K.D. Farmani, P.E.N. Kencana, K.G. Sukarsa, Perbandingan analisis least absolute shrinkage and selection operator (LASSO) dan partial least squares (PLS) (Studi Kasus: Data Microarray), e-Jurnal Matematika. 1 (2012), 75-80.

[12] F.K.H. Prabowo, Y. Wilandari, A. Rusgiyono, Pemodelan pertumbuhan ekonomi jawa tengah menggunakan pendekatan least absolute shrinkage and selection operator (LASSO), Jurnal Gaussian, 4 (2015), 855–864.

[13] E.W. Wibowo, Metode eksploratif untuk menguji kesamaan spektrum FTIR Temulawak, Thesis, Institut Pertanian Bogor, (2009).

[14] H. Mardhotillah, B. Sartono, Y. Angraini, Penerapan group least absolute shrinkage and selection operator sebagai metode alternatif dalam menangani data berdimensi tinggi, Thesis, Institut Pertanian Bogor, (2015).

[15] K. Kesornbuakao, S. Amornraksa, M. Sriariyanun, et al. The antibacterial and antioxidant activity of Centella asiatica chloroform extract-loaded gelatin nanoparticles, MATEC Web Conf. 187 (2018), 02002. https://doi.org/10.1051/matecconf/201818702002.

[16] P. Yasurin, S. Malinee, P. Theerawut, Review: The bioavailibility activity of Centella asiatica, KMUTNB Int. J. Appl. Sci. Technol. 9 (2016), 1-9.

[17] N.A. Zainol, Profiling of Centella asiatica (L.) urban extract, Malays. J. Anal. Sci. 12 (2008), 322-327.

[18] I.S. Young, Antioxidants in health and disease, J. Clinic. Pathol. 54 (2001), 176–186. https://doi.org/10.1136/jcp.54.3.176.

[19] F. Pitella, Antioxidant and cytotoxic activities of Centella asiatica (L) Urb, Int. J. Mol. Sci. 10(2009), 3713-3721. https://doi.org/10.3390/ijms10093713.

[20] M. Valko, C.J. Rhodes, J. Moncol, et al. Free radicals, metals and antioxidants in oxidative stress-induced cancer, Chem.-Biol. Interactions. 160 (2006), 1–40. https://doi.org/10.1016/j.cbi.2005.12.009.

[21] D.E. Pratt, B.J.F. Hudson, Natural antioxidants not exploited commercially, in: B.J.F. Hudson (Ed.), Food Antioxidants, Springer, Dordrecht, 1990: pp. 171–191. https://doi.org/10.1007/978-94-009-0753-9_5.

[22] G.D. Christian, Analytical chemistry, Wiley, New York, (1986).

[23] A.M. Soleh, Aunuddin, LASSO: solusi alternatif seleksi peubah dan penyusutan koefisien model regresi linier, Forum Statistika dan Komputasi: Indonesian Journal of Statistics. 18 (2013), 21-27.

[24] T. Hastie, R. Tibshirani, J. Friedman, The elements of statistical learning, Springer, New York, (2008).

[25] J. Huang, T. Zhang, The benefit of group sparsity, Ann. Stat. 38 (2010), 1978-2004. https://doi.org/10.1214/09-aos778.

[26] M. Yuan, Y. Lin, Model selection and estimation in regression with grouped variables, J. R. Stat. Soc B. 68 (2006), 49–67. https://doi.org/10.1111/j.1467-9868.2005.00532.x.

[27] R.T.E. Putri, Korelasi antara spektrum ftir dan aktivitas antioksidan dari pegagan (*Centella Asiatica*) menggunakan kemometrik, Thesis, Institut Pertanin Bogor, 2018.

[28] S. Easmin, Md.Z.I. Sarker, K. Ghafoor, et al. Rapid investigation of α-glucosidase inhibitory activity of Phaleria macrocarpa extracts using FTIR-ATR based fingerprinting, J. Food Drug Anal. 25 (2017), 306–315. https://doi.org/10.1016/j.jfda.2016.09.007.

[29] R.L. Wehling, Infrared Spectroscopy, in: S.S. Nielsen (Ed.), Food Analysis, Springer, Boston, 2010: pp. 407–420. https://doi.org/10.1007/978-1-4419-1478-1_23.

[30] K.K. Chew, S.Y. Ng, Y.Y. Thoo, et al. Effect of ethanol concentration, extraction time and extraction temperature on the recovery of phenolic compounds and antioxidant capacity of *Centella asiatica* extracts, Int. Food Res. J. 18 (2011), 571-578.