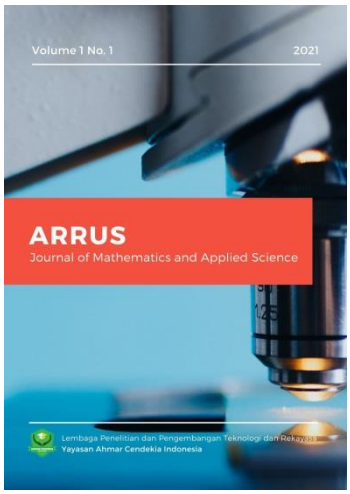
https://doi.org/10.35877/mathscience3626

OPEN ACCESS

ISSN : 2776-7922 (Print) / 2807-3037 (Online)



RESEARCH ARTICLE

Metabolite-Group Selection on *Temu Ireng (Curcuma Aeruginosa)* Contains Related to Toxicity Activity by Using Group Lasso Regression

Rahmat H. S.[1*], Farit Mochamad Afendi[2], Hari Wijayanto[2]

[1]Statistics Study Program, Faculty of Mathematics and Natural Sciences, State University of Makassar, Makassar, South Sulawesi.
[2]Department of Statistics, Faculty of Mathematics and Natural Sciences, IPB University, Bogor, West Java.

Abstract: Metabolites are expressed in mass-to-charge ratio (m/z) on mass spectrometry experiments. They can be identified more than once. Some of m/z representing same metabolites can be considered as a group of metabolites. Evaluation of metabolite effects can be considered based on the groups. Group least absolute shrinkage and selection operator (group lasso) regression can be used to evaluate these groups. It shrinks some coefficients of regression exactly to be zero by adding intermediate penalty on ordinary least square (OLS) objective function. The purposes of this study were to estimate groups of metabolite contains of *Curcuma aeruginosa* (*Temu ireng*) affecting toxicity activity by using group lasso regression and to compare it to partial least square regression (PLSR). The data used were toxicity activity and metabolite contain, obtained from LC-MS, of *temu ireng* from three areas in Java. The groups of metabolites which affected toxicity activity, of group lasso regression by using dedicated software of R with gglasso package, were groups of m/z 238.150, 250.165, 262.128, 264.144, 312.275, and 456.183. The estimates of metabolites that affected of group lasso regression and PLSR had similarities. Based on the goodness of fit, group lasso regression was better than PLSR to estimate the affecting groups.

Keywords: *Curcuma aeruginosa, group-variable selection, group lasso regression, toxicity activity.*

*Corresponding author: Rahmat H. S., Statistics Study Program, Faculty of Mathematics and Natural Sciences, State University of Makassar, Makassar, Indonesia

E-mail: rahmaths@unm.ac id

1. INTRODUCTION

Zingiberaceae is a family of plants used as sources of herbal medicine or jamu. One of the family is *Curcuma aeruginosa* (*Temu ireng*). It can be used as drugs of anti-rheumatic, skin diseases such as scabies, itching, and blood cleansing after giving birth (Djauharia & Sufiani, 2007) colic, obesity and rheumatism, asthma and coughing, and mental disorders (Waras et al., 2015). The chemical active compound or metabolite contains are important to be identified for further utilization of *temu ireng*. Secondary metabolites of plants are used as medicinal and food ingredients (Khafid et al., 2023). By recognizing its metabolite contains, it can be related to other characteristics such as bioactivity or biological activity. Bioactivity describes the advantages or disadvantages of a kind of drugs toward living organisms. One of bioactivities is toxicity activity, the extent to which chemicals or mixtures of certain substances can harm organisms (Shiel, 2008), which is one of the considerations for the safety of sample usage as a mixture of drugs or food.

https://doi.org/10.35877/mathscience3626

ISSN : 2776-7922 (Print) / 2807-3037 (Online)

Liquid chromatography, a chemical technique, combined to mass spectrometry (LC-MS) has been widely used to identify metabolite contains of samples (De Vos et al., 2007; Grata et al., 2009; Rakibe et al., 2018; Roux et al., 2011). Besides being easy to use, LC-MS has good accuracy in detecting metabolites that have been separated from sample solutions (Lacorte & Fernandez-Alba, 2006; Roux et al., 2011; Yoshida et al., 2009). The metabolites that are successfully identified by LC-MS are expressed in mass-to-charge ratio (m/z) (Ardrey, 2003). The metabolite of an m/z cannot be justified immediately because some metabolites have same molecular mass. (Ravindran et al., 2007) compiled estimates of plant metabolites on genus of *Curcuma* based on m/z.

The output data structure of LC-MS can consist of large quantities of m/z which is far exceeding the number of samples used. Because an m/z can be identified more than once. The number of predictors far exceeding the number of samples ($p >> n$) is recognized as high dimensional data. This circumstance produces singular matrix of $\boldsymbol{X}^T\boldsymbol{X}$ because input matrix $\boldsymbol{X}$ which is not full rank. Matrix $\boldsymbol{X}$ which is not full rank can also occur because columns which are not linearly independent. It is due to collinearity i.e. a predictor is a function of another predictor. So that, they have a perfect correlation. It is known as ill-conditioned. Estimation using ordinary least squares (OLS) method will not be unique (Frank & Friedman, 1993; Tibshirani, 1996; Wold et al., 2001).

(Septaningsih et al., 2018) estimated metabolite contains of *temu ireng* effecting toxicity activity by using partial least square regression (PLSR). It stated that metabolites contain of *temu ireng* which contributed greatly to antioxidant and toxicity activities were metabolites with m/z of 248.15 which was estimated as 9-Oxoneoprocurcumenol, 7α, 11α, -Epoxy- 5β-hydroxy-9-guaiaen-8-one, Curcumenolactone A, or Curcumenolactone B and 312.28 which was estimated as tetrahydro-bisdemethoxycurcumin. That study considered that m/z represents same metabolites as a different of m/z. On the other hand, some of m/z that represent the same metabolites can be considered as a group of metabolites. Therefore, evaluation of metabolite effect is considered based on the groups.

Analytical methods that can be used are group nonnegative garrote (Breiman, 1995), group least angle regression (group LARS): modification of least absolute shrinkage and selection operator (lasso) (Efron et al., 2004) and group lasso regression (Bakin, 1999; Yuan & Lin, 2006). (Yuan & Lin, 2006) said that a drawback of group nonnegative garrote is that its estimation performance is not optimal on high dimensional data where group LARS and group lasso regression can do so. Group LARS is a modification of least angle regression as an approach to group lasso regression. Group LARS estimates are slightly different with group lasso regression. Therefore, this study used group lasso regression.

Group lasso regression is a development of lasso regression introduced by (Tibshirani, 1996). Group lasso regression introduced by (Bakin, 1999) and then developed by (Yuan & Lin, 2006). Group lasso regression considers predictor variables consisting of different groups of predictor variables. Group lasso regression adds intermediate penalty, ℓ_1 for lasso regression and penalty ℓ_2 for ridge regression, on OLS objective function yielding some coefficients of regression are shrunken exactly to be zero so that variable selection performed. Some studies applied group lasso regression on gene expression data such as (Ma et al., 2007; Meier et al., 2006; Silver et al., 2012). The purposes of this study were to estimate the affected groups of metabolite contains of *temu ireng* to toxicity activity using group lasso regression and compare it to PLSR as (Septaningsih et al., 2018).

2. Research Method and Materials

This study used data of bioactivity test and a spectrometry experiment conducted by (Septaningsih et al., 2018) at the Analytical Chemistry Laboratory, Department of Chemistry and the Tropical Medical Study Center of Bogor Agricultural University. The data was toxicity activity and metabolite contain of *temu ireng* rhizome extracts. *Temu ireng* samples captured from three areas in Java, namely Cikabayan, Nagrak and Tawangmangu.

https://doi.org/10.35877/mathscience3626

ISSN : 2776-7922 (Print) / 2807-3037 (Online)

Toxicity activity data, as a respond variable, were obtained from toxicity test using brine shrimp lethality test (BSLT) in (LC_{50}) ppm (Krishnaraju et al., 2005). Five samples of each area were then selected based on their antioxidant and toxicity classifications.

Identification of *temu ireng* contains obtained from LC-MS Waters Xevo G2-S QTOF consisted of retention time, m/z and peak intensity. The identification results of all samples were merged and then MZmine was applied for eliminating effects of the baseline shifts and noises on the chromatograms. After that, the dimension of retention time was removed, so that the remaining dimensions were m/z and peak intensity. Furthermore, the amount of an m/z were obtained from ratio of a peak intensity of an m/z to the total peak intensity referred to as % relative area (Murray et al., 2013) as predictor variables.

The number of m/z identified by LC-MS output, of the 15 samples used, was 39 m/z consisting of 175 estimated metabolites. The dataset consisted of 308 m/z. Grouping was based on m/z hence the groups of predictors used was consisted of 39 groups. Those groups represented four groups of secondary metabolites on genus of *Curcuma* namely diphenylheptanoids, phenylphropenederivates, terpenoids, and miscellaneous.

Data analysis procedures were conducted on this study as follows:

(1). Exploring data
   (a). Identifying compositions of secondary metabolite contains of *temu ireng* rhizomes used to identify the dominant secondary metabolites.
   (b). Identifying groups of metabolites which highly correlated to toxicity activity with threshold $-1 < r \leq -0.5$ and $0.5 \leq r < 1$ (Paul et al., 2008). It aimed at identifying high correlation of groups of metabolites to toxicity activity.

(2). Standardizing predictor variables because estimates of coefficients cannot be compared if the predictors are not be standardized, not equivariant (Hastie et al., 2009):

$$x_{ij}^{*} = \frac{x_{ij} - \bar{x}_{ij}}{s_{x_{ij}}}$$

where x^* is predictor variables after standardization. Respectively, x_{ij}, $\bar{x}_{ij}$ and $s_{x_{ij}}$ are predictor variables, the mean of predictor variables and standard deviance of predictor variables before standardization.

(3). Applying group lasso regression for minimizing:

$$\beta^{glasso} = \arg\min_{\beta} \frac{1}{2}\left\| y - \sum_{l=1}^{m} X^{(l)}\beta^{(l)} \right\|_2^2 + \lambda \sum_{l=1}^{m} \sqrt{p_l}\left\|\beta^{(l)}\right\|_2, \qquad \lambda > 0$$

where $\boldsymbol{X}_{(l)}$ is submatrix of $\boldsymbol{X}$ with columns corresponding to l groups of predictors, $\beta^{(l)}$ is coefficient vectors of groups of predictor variables, p_l is length of vector $\beta^{(l)}$, $\|\beta^{(l)}\|_2$ is Euclidian norm and λ is tuning parameter. Estimation of group lasso regression can be approached using groupwise-majorization-descent (GMD) algorithm introduced by (Yang & Zou, 2015) as follows:
   (a). For $k = 1, \ldots, K$, computing γ_k i.e. the largest eigenvalue of $\boldsymbol{H}_{(k)}$.
   (b). Initialing $\tilde{\beta}$
   (c). For $k = 1, \ldots, K$:
      - Calculating $U(\tilde{\beta}) = -\nabla L(\tilde{\beta}|D)$
      - Calculating $\tilde{\beta}^{(k)}(new) = \frac{1}{\gamma k}\left(U^{(k)} + \gamma_k \tilde{\beta}^{(k)}\right)\left(1 - \frac{\lambda_{wk}}{\left\|U^{(k)} + \gamma_k \tilde{\beta}^{(k)}\right\|_2}\right)_+$
      - Setting $\tilde{\beta}^{(k)} = \tilde{\beta}^{(k)}(new)$
   (d). Updating groupwise cycle until convergence.

The GMD algorithm was packed in R package i.e. gglasso:

(a). Determining quantity of λ using cross-validation. The magnitude of λ chosen was it can yield interpretable estimated to compare as point 4.
   where $\hat{f}^{-\kappa(i)}(x_i, \lambda)$ is used function of training data with the k^{th} part removed and λ as the tuning parameter. Function of L is the objective function defined as

https://doi.org/10.35877/mathscience3626

ISSN : 2776-7922 (Print) / 2807-3037 (Online)

$L(y_i, \hat{f}^{-\kappa(i)}(x_i, \lambda))$. CV($\hat{f}$,$\lambda$) yields estimated prediction error curve of testing data and determines estimated tuning parameter λ minimizing prediction error. The type of cross-validation used was leave-one-out cross-validation (LOOCV), k=n, because sample size was small (Hastie et al., 2009).

(b). Calculating root mean square prediction (RMSEP)

$$RMSE = RMSEP = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}.$$

(c). Estimating coefficients of regression and calculating RMSE.

(d). Determining the tendency of groups of metabolites to toxicity activity by voting if the coefficient signs of a group were imbalanced and picking up the largest coefficient sign of a group member if coefficient signs of regression were balanced.

(e). Determining the most influential group of metabolites by using plot of group-norm path coefficients respect to log(λ).

(4). Comparing between estimates of group lasso regression and PLSR by using RMSE, the goodness of fit used in (Septaningsih et al., 2018).

3. Results and Discussion

3.1. Data exploration

The compositions of metabolite groups contain of *temu ireng* were dominated by groups of metabolites with m/z of 228.123, 230.138, 234.170, 216.155, 246.133, and 232.154 as the frequency of occurrence. These groups were classified as secondary metabolites of terpenoids as depicted on Figure 1. This was in line with (Afzal et al., 2013) stated that terpenoids were one of the largest groups of natural compounds identified in plants.

Summary of the groups of metabolite contains of *temu ireng* which had strong relationship to toxicity activity was depicted on Figure 2. There were 27 groups of metabolites high correlated to toxicity activity. Twenty groups of metabolites had relatively high correlations to toxicity activity.

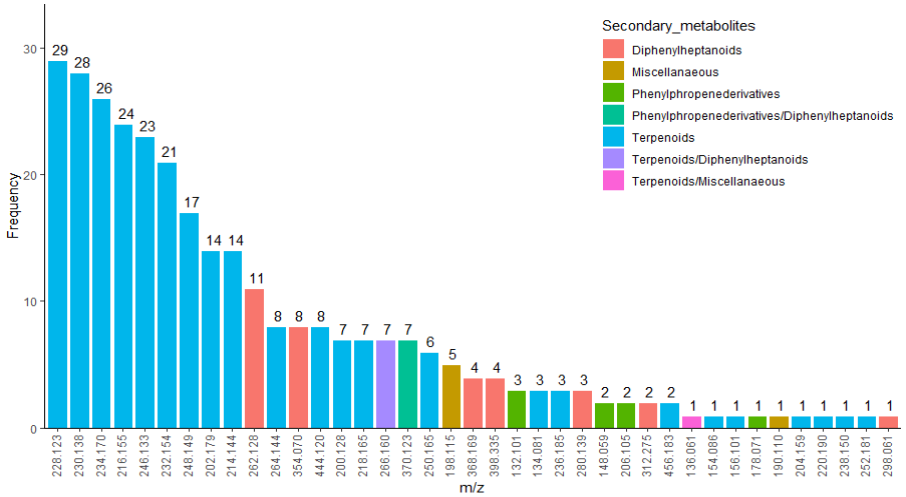


Figure 1: Composition of groups of metabolite contains of *temu ireng*. Groups of metabolites (bar), frequency of occurrence (number labels) and secondary metabolites (colors)

It was indicated by the absolute coefficients of correlation greater than or equal to 0.70. The groups of metabolites with a directly proportional relationship were the groups of metabolites with m/z of 132.101, 198.115, 202.179, 204.159, 214.144, 216.155, 228.123, 232.154, 234.170, 246.133, 248.149, 250.165, 262.128, 264.144, and 266.160. The higher the amount of metabolite groups, the higher the toxicity activity. Instead, groups of metabolites with an inverse relationship were groups of metabolites with m/z of 200.128, 218.165, 230.138, 236.185 and 368.169. The higher the number of metabolite groups, the lower

https://doi.org/10.35877/mathscience3626

ISSN : 2776-7922 (Print) / 2807-3037 (Online)

toxicity activity. It was estimated that the relationship occurred because metabolite contains of *temu ireng* was working synergistically. It was in line with (Liang et al., 2009). The groups of metabolites whose strong correlation to toxicity were dominated by secondary metabolites of terpenoids.

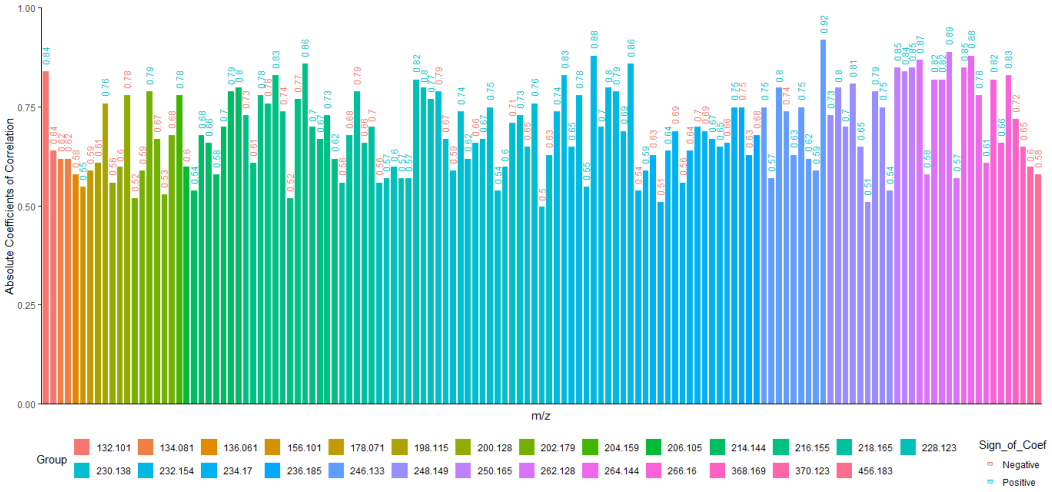


Figure 2: Groups of metabolites whose strong correlation to toxicity activity. Groups of metabolites (bar), absolute coefficient of correlation (number labels), correlation signs (red negative, blue positive), and groups of metabolites (colors)

3.2. Metabolite-group selection

The first step in metabolite-group selection using group lasso regression was determination of quantity of λ as the tuning parameter as shrinkage control of the coefficients of regression. Magnitude of λ were determined using LOOCV technique. Figure 3 depicted 100 estimated magnitudes of λ could be used, with the prediction error for each λ. There were two recommended magnitudes of λ namely lambda minimum which produced more complex model with minimum of RMSEP and lambda one standard error produced the simplest model with comparable RMSEP with RMSEP minimum. The differences of mean prediction error as in the cross-validation curve were not so large after lambda minimum, which was shown as the second dash vertical line (read from right to left). Therefore, the magnitude of λ could be selected at that interval by considering the prediction error was close and interpretable to compare. So that, the magnitude of selected λ, shown as solid line, was 4.313 or log(λ) was 1.462 with RMSEP 31.421.

The results of metabolite-group selection which affected toxicity activity using group lasso regression were depicted on Figure 4. The greater m/z value, the smaller coefficients regression tent to be. The affected groups of metabolites were metabolites with m/z of 238.150, 250.165, 262.128, 264.144, 312.275, and 456.183. Groups of metabolites with m/z of 250.164, 262.128, 264.143, and 312.273 tent affected positively to toxicity activity. It meant that the groups of metabolites could improve the toxicity of *temu ireng*, so that the *temu ireng* rhizomes was better to be utilized. Groups of metabolites with m/z of 238.150 and 456.183 tent affected negatively toxicity. The RMSE value of groups of metabolites used was 15.446. The estimates of groups of metabolites that affected toxicity activity were shown in Table 1. The most affected group of metabolites was the group of metabolites with m/z of 262.128 because it was the first entered to the model or nonzero coefficient from the largest to the smallest λ value as depicted on Figure 5.

https://doi.org/10.35877/mathscience3626

ISSN : 2776-7922 (Print) / 2807-3037 (Online)

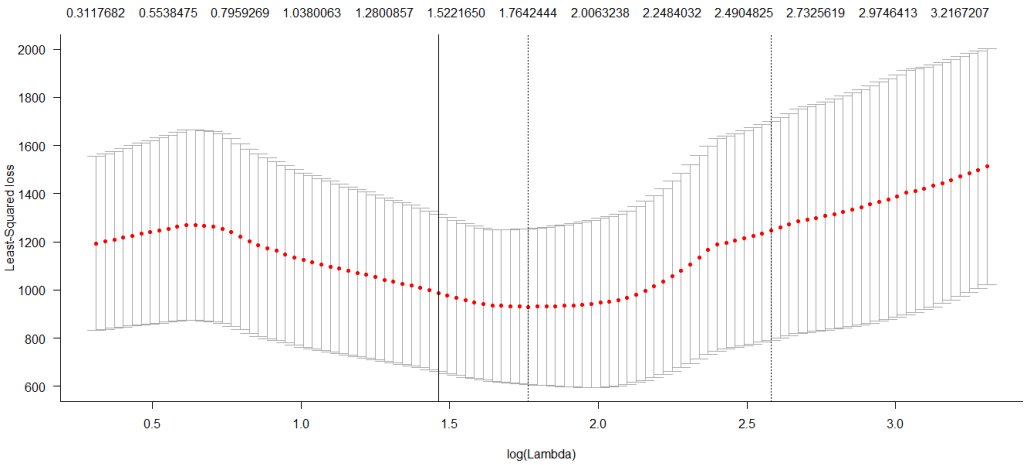


Figure 3: Prediction error curve of each λ. Mean of prediction error (•), prediction error deviance (I), λ that can be used (⋮), and λ was used (|)

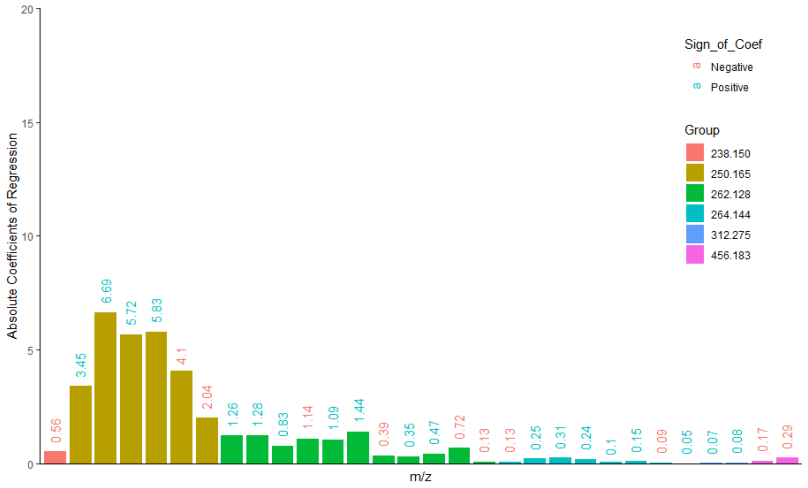


Figure 4: Absolute coefficients of regression. Groups of metabolites (bar), absolute coefficient of regression (number labels), coefficient signs (red negative, blue positive), and groups of metabolites (colors)

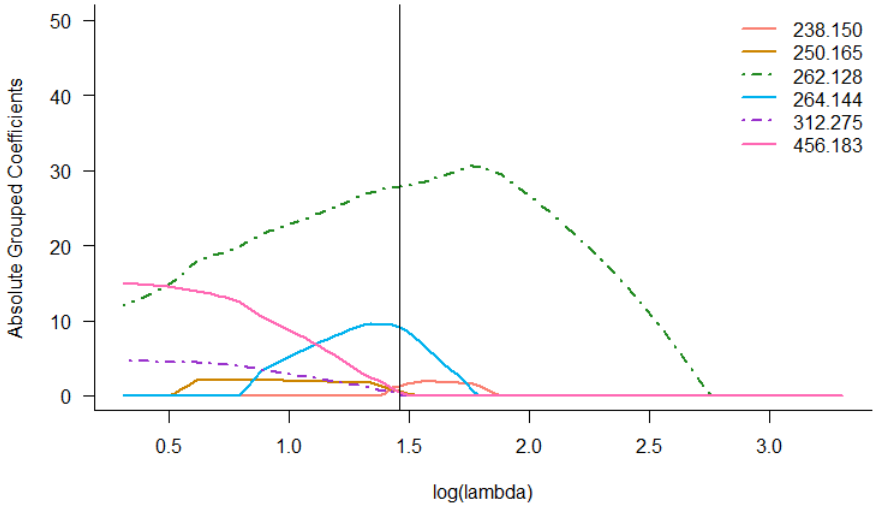


Figure 5: Solution path group norm of for each λ. Groups of metabolites effected (colors), terpenoids (—) and diphenylheptanoids (-·-)

https://doi.org/10.35877/mathscience3626

ISSN : 2776-7922 (Print) / 2807-3037 (Online)

Table 1: Estimates of metabolites

m/z	Estimates of metabolites
238.150	1,10-Bisaboladiene-3,4-diol, 2,10-Bisaboladiene-1,4-diol, dihidroxybisabola-3,10-diene, dihidroxybisabola-2,10-diene, Alismoxide, Curcumadiol
250.165	13-Hydroxydehydro-curdione, (1α,4β,5α,10β) 1,10: 4,5- diepoxy-7(11)-germacren-8-one, (1S,10S), (4S,5S)-germacrone-1(10), 4- diepoxide, 4,5-Epoxy-12-hydroxy-1(10),7(11)-germacredien-8-one, Aerugidiol, Oxycurcumenol, Procurcumadiol
262.128	(1E, 3E)-1,7-Diphenyl-1,3-heptadien-5-one
264.144	(E)1,7-diphenyl-1-hepten-5-one, (1E, 3E)-1,7-Diphenyl-1,3-heptadien-5-ol
312.275	Tetrahydro-bisdemethoxycurcumin
456.183	Parviflorene B, Parviflorene C, Parviflorene D, Parviflorene E

Groups of metabolites affected toxicity activity dominated by groups of metabolites classified as secondary metabolite of terpenoids. It was in line with (Dar et al., 2011) who reported that some derivatives of secondary metabolite of terpenoids had cytotoxic to human cancer cell lines. (Paduch et al., 2016) also reported that due to cytotoxic activity of terpenes caused substances or essential oils (EOs) contained terpenes as natural products may positively affect human health, exerting slight or no side effects.

The previous research, (Septaningsih et al., 2018) considered the singles of m/z represented same metabolites as a single different m/z using PLSR, concluded that metabolites that affected toxicity activity were metabolites with m/z of 206.105, 250.164, 262.128, 312.273, 368.169, and 368.170. The RMSE yielded was 17.434. The estimates of groups of metabolites that affected toxicity activities using group lasso regression had similarities to the estimates of PLSR. Three groups of metabolites of group lasso regression i.e. m/z of 250.165, 262,128 and 312.273 included to m/z which affected toxicity activity using PLSR.

The goodness of fit of estimation method on metabolites that affected toxicity activity was viewed based on the RMSE. The best analysis method was a method with a small RMSE. Therefore, it was sufficient evidence to draw conclusion that the estimation of *temu ireng* metabolite contains that affected toxicity activity using group lasso regression was better than using PLSR.

4. Conclusion

This study used group lasso regression to estimate the groups of metabolite contains of *temu ireng* affected toxicity activity. The groups of metabolite contents of *temu ireng* affected toxicity activity were groups of metabolites with m/z of 238.150, 250.165, 262.128, 264.144, 312.275, and 456.183. The estimates of group lasso regression had similarities to estimates of PLSR. Those were 250.165, 262.128 and 312.275. Estimation of metabolites that affected toxicity activity using group lasso regression was better than using PLSR based on RMSE.

References

Afzal, A., Oriqat, G., Akram Khan, M., Jose, J., & Afzal, M. (2013). Chemistry and biochemistry of terpenoids from Curcuma and related species. *Journal of Biologically Active Products from Nature*, *3*(1), 1–55.

Ardrey, R. E. (2003). *Liquid chromatography-mass spectrometry: an introduction* (Vol. 2). John Wiley & Sons.

Bakin, S. (1999). Adaptive regression and model selection in data mining problems. *PhD Thesis Mathematical Science Institute, The Australian National University*.

Breiman, L. (1995). Better Subset Regression Using the Nonnegative Garrote. *Technometrics*, *37*(4), 373. https://doi.org/10.2307/1269730

Dar, M. Y., Shah, W. A., Rather, M. A., Qurishi, Y., Hamid, A., & Qurishi, M. A. (2011). Chemical composition, in vitro cytotoxic and antioxidant activities of the essential oil and major

ISSN : 2776-7922 (Print) / 2807-3037 (Online)

constituents of Cymbopogon jawarancusa (Kashmir). *Food Chemistry*, *129*(4), 1606–1611.

De Vos, R. C. H., Moco, S., Lommen, A., Keurentjes, J. J. B., Bino, R. J., & Hall, R. D. (2007). Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nature Protocols*, *2*(4), 778–791. https://doi.org/10.1038/nprot.2007.95

Djauharia, E., & Sufiani, S. (2007). Observasi keragaan tanaman temu hitam (Curcuma aeruginosa Roxb.) pada berbagai jarak tanam. *Warta Tumbuhan Obat Indonesia*, *7*(1), 21–23.

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). *Least angle regression*.

Frank, Ll. E., & Friedman, J. H. (1993). A statistical view of some chemometrics regression tools. *Technometrics*, *35*(2), 109–135.

Grata, E., Guillarme, D., Glauser, G., Boccard, J., Carrupt, P. A., Veuthey, J. L., Rudaz, S., & Wolfender, J. L. (2009). Metabolite profiling of plant extracts by ultra-high-pressure liquid chromatography at elevated temperature coupled to time-of-flight mass spectrometry. *Journal of Chromatography A*, *1216*(30), 5660–5668. https://doi.org/10.1016/j.chroma.2009.05.069

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2). Springer.

Khafid, A., Wiraputra, M. D., Putra, A. C., Khoirunnisa, N., Putri, A. A. K., Suedy, S. W. A., & Nurchayati, Y. (2023). Uji Kualitatif Metabolit Sekunder pada Beberapa Tanaman yang Berkhasiat sebagai Obat Tradisional. *Buletin Anatomi Dan Fisiologi*, *8*(1), 61–70. https://doi.org/10.14710/baf.8.1.2023.61-70

Krishnaraju, A. V, Rao, T. V. N., Sundararaju, D., Vanisree, M., Tsay, H.-S., & Subbaraju, G. V. (2005). Assessment of bioactivity of Indian medicinal plants using brine shrimp (Artemia salina) lethality assay. *International Journal of Applied Science and Engineering*, *3*(2), 125–134.

Lacorte, S., & Fernandez-Alba, A. R. (2006). Time of flight mass spectrometry applied to the liquid chromatographic analysis of pesticides in water and food. *Mass Spectrometry Reviews*, *25*(6), 866–880.

Liang, X., Jin, Y., Wang, Y., Jin, G., Fu, Q., & Xiao, Y. (2009). Qualitative and quantitative analysis in quality control of traditional Chinese medicines. *Journal of Chromatography A*, *1216*(11), 2033–2044.

Ma, S., Song, X., & Huang, J. (2007). Supervised group Lasso with applications to microarray data analysis. *BMC Bioinformatics*, *8*, 1–17. https://doi.org/10.1186/1471-2105-8-60

Meier, L., Van De Geer, S., & Zou, H. (2006). The group lasso for logistic regression - Meier - 2008 - Journal of the Royal Statistical Society: Series B (Statistical Methodology) - Wiley Online Library. *Journal of the American Statistical Association*, *101*(476), 1418–1429. http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9868.2007.00627.x/full%5Cnpapers2://publication/uuid/969A5A32-A179-4503-84BE-C1B33F057B06

Murray, K. K., Boyd, R. K., Eberlin, M. N., John Langley, G., Li, L., & Naito, Y. (2013). Definitions of terms relating to mass spectrometry (IUPAC Recommendations 2013). *Pure and Applied Chemistry*, *85*(7), 1515–1609. https://doi.org/10.1351/PAC-REC-06-04-06

Paduch, R., Trytek, M., Król, S. K., Kud, J., Frant, M., Kandefer-Szerszeń, M., & Fiedurek, J. (2016). Biological activity of terpene compounds produced by biotechnological methods. *Pharmaceutical Biology*, *54*(6), 1096–1107. https://doi.org/10.3109/13880209.2015.1103753

Paul, D., Bair, E., Hastie, T., & Tibshirani, R. (2008). *"Preconditioning" for feature selection and regression in high-dimensional problems*.

Rakibe, U., Tiwari, R., Mahajan, A., Rane, V., & Wakte, P. (2018). LC and LC–MS/MS studies for the identification and characterization of degradation products of acebutolol. *Journal of Pharmaceutical Analysis*, *8*(6), 357–365. https://doi.org/10.1016/j.jpha.2018.03.001

Ravindran, P. N., Babu, K. N., & Sivaraman, K. (2007). *Turmeric: the genus Curcuma*. CRC press.

Roux, A., Lison, D., Junot, C., & Heilier, J. F. (2011). Applications of liquid chromatography coupled to mass spectrometry-based metabolomics in clinical chemistry and toxicology: A review. *Clinical Biochemistry*, *44*(1), 119–135. https://doi.org/10.1016/j.clinbiochem.2010.08.016

Septaningsih, D. A., Darusman, L. K., Afendi, F. M., & Heryanto, R. (2018). Liquid chromatography mass spectrometry (LC-MS) fingerprint combined with chemometrics for identification of metabolites content and biological activities of Curcuma Aeruginosa. *Indonesian Journal of Chemistry*, *18*(1), 43–52. https://doi.org/10.22146/ijc.25456

https://doi.org/10.35877/mathscience3626

ISSN : 2776-7922 (Print) / 2807-3037 (Online)

Shiel, W. (2008). *Webster's new world medical dictionary*. Wiley Publishing.

Silver, M., Montana, G., & Alzheimer's Disease Neuroimaging Initiative, and. (2012). Fast identification of biological pathways associated with a quantitative trait using group lasso with overlaps. *Statistical Applications in Genetics and Molecular Biology*, *11*(1), 000010220215446115 1755.

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society. Series B: Methodological*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Waras, N., Nurul, K., Muhamad, S., Maria, B., & Ardyani, I. D. A. A. C. (2015). Phytochemical screening, antioxidant and cytotoxic activities in extracts of different rhizome parts from Curcuma aeruginosa RoxB. *International Journal of Research in Ayurveda and Pharmacy*, *6*(5), 634–637. https://doi.org/10.7897/2277-4343.065118

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, *58*(2), 109–130.

Yang, Y., & Zou, H. (2015). A fast unified algorithm for solving group-lasso penalize learning problems. *Statistics and Computing*, *25*, 1129–1141.

Yoshida, H., Yamazaki, J., Ozawa, S., Mizukoshi, T., & Miyano, H. (2009). Advantage of LC-MS metabolomis methodology targeting hydrophillic compounds in the studies of fermented food samples. *Journal of Agricultural and Food Chemistry*, *57*(4), 1119–1126. https://doi.org/10.1021/jf803235m

Yuan, M., & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *68*(1), 49–67. https://doi.org/10.1111/j.1467-9868.2005.00532.x