



Energy Sources, Part A: Recovery, Utilization, and **Environmental Effects** 

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ueso20

# Toward estimating standard enthalpy of combustion of pure chemical compounds: extreme learning machine approach

Roy Setiawan & Samira Mohammadinia

To cite this article: Roy Setiawan & Samira Mohammadinia (2021): Toward estimating standard enthalpy of combustion of pure chemical compounds: extreme learning machine approach, Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, DOI: 10.1080/15567036.2021.1917730

To link to this article: <u>https://doi.org/10.1080/15567036.2021.1917730</u>



Published online: 25 Apr 2021.

|--|

Submit your article to this journal 🖸





View related articles



View Crossmark data 🗹

## **RESEARCH PAPER**



Check for updates

# Toward estimating standard enthalpy of combustion of pure chemical compounds: extreme learning machine approach

Roy Setiawan<sup>a</sup> and Samira Mohammadinia<sup>b</sup>

<sup>a</sup>Universitas Kristen Petra, Surabaya, Indonesia; <sup>b</sup>Chemical Engineering Department, Islamic Azad University, Mahshahr, Iran

#### ABSTRACT

One of the effective thermochemical properties in the determination of heat process efficiency is the combustion enthalpy changes during complete combustion of the compounds. According to the importance of this property in different processes, the main aim of this work is selected as the development of extreme learning machine (ELM) approach to predict the combustion enthalpy in terms of functional groups. To achieve this goal, a comprehensive data set containing 4,590 experimental enthalpy points is used for the preparation and validation of ELM. To investigate the accuracy of the ELM approach in the estimation of the enthalpy, various visual and statistical comparisons are used. These comparisons lead into  $R^2$  value of one and low error values for overall phase. The standard deviation, root mean squared error, and mean relative error for overall phase are determined to be 11.18, 14.92, and 0.28, respectively. The relative deviations between the estimated and actual enthalpy points are below 8%. According to the statistical and graphical results, ELM algorithm has great potential in the prediction of enthalpy of combustion for pure chemical materials.

#### **ARTICLE HISTORY**

Received 23 July 2020 Revised 25 March 2021 Accepted 26 March 2021

#### **KEYWORDS**

ELM; predicting model; enthalpy; combustion; heating value

# Introduction

The combustion enthalpy ( $\Delta_c H^\circ$ ) for chemical material is expressed for the amount of changes in enthalpy when the material undergoes oxidation to combustion products in standard condition at the pressure of 1 atm and temperature of 298.15 K. The combustion products include N<sub>2</sub>(g), CO<sub>2</sub>(g), Cl<sub>2</sub> (g), F<sub>2</sub>(g), SO<sub>2</sub>(g), H<sub>2</sub>O(g), Br<sub>2</sub>(g), H<sub>3</sub>PO<sub>4</sub>(s), I<sub>2</sub>(g), and SiO<sub>2</sub>(crystobalite) (International 2009).

The knowledge about the heat of reactions, and this parameter is highly important when researchers and engineers are dealing with the calculation of thermal efficiency of the process to produce heat and power. The enthalpy of combustion gives knowledge for the value of available heat energy from the combustion of a fuel to perform efficient work. Thus, the importance of accurate and reliable heat of combustion values becomes highlighted when designing plant because of time consumption and hardships in the measurement of this value for many materials including toxic, explosive, volatile, and highly reactive materials. Due to these facts, the interest to develop a high ability approach for prediction of this property in terms of the material chemical structure has been increased (Suleymani and Bemani 2018).

In the literature, there are several approaches for forecasting enthalpy of combustion. Cardozo suggested a group contribution approach based on equivalent normal alkane chain length. This approach was constructed by 1,168 different pure organic compounds; however, the details of the accuracy of this method were not mentioned by Cardozo (Cardozo 1986). After that, Seaton et al. proposed an approach derived from Benson's approaches that had been used for prediction of

#### 2 👄 R. SETIAWAN AND S. MOHAMMADINIA

formation enthalpy of pure materials. Unfortunately, they did not discuss the degree of precision for their model. Moreover, they did not report any details about databank, which they utilized to present their method (Seaton and Harrison 1990). Then, two empirical models were developed by Hshieh to estimate combustion enthalpy of polymers and organosilicon compounds. These models have limitation in their applicability to a definite number of chemical group compounds(Hshieh, Hirsch, and Beeson 2003). Besides these approaches, Gharagheizi suggested a quantitative structure-property relation to predict enthalpy of combustion. Gharagheizi used 1,714 pure compounds for the development of the model with  $R^2$  of 0.996 (Gharagheizi 2008). In another research, Pan developed two methods based on artificial neural network and quantitative structure-property relation. The first one utilizes atom-type electro topological state index to predict the enthalpy. This method concluded to  $R^2$ of 0.991. The other method had  $R^2$  of 0.995 for 1,650 pure compounds (Pan et al. 2011). Then, an artificial neural network that used group contributions for the estimation of the enthalpy of combustion with an absolute deviation of 0.16% is presented by Gharagheizi (Gharagheizi, Mirkhani, and Tofangchi Mahyari 2011). The experimental investigations require lots of cost (Li et al. 2019) and time (Shi et al. 2018); therefore, the computational approaches are preferred in different issues (Xu et al. 2020). The artificial intelligence methods are known as the most reliable methods (Chen et al. 2020).

The artificial intelligence methods have extensive applications in other issues such as determination of higher heating values based on proximate (Keybondorian et al. 2017) and ultimate systems (Darvishan, BakhDarvishan et al. 2018), asphaltene adsorption (Mazloom et al. 2020), hydrocarbon and carbon dioxide interfacial tension (Suleymani and Bemani 2018), and sulfur deposition (Bemani, Baghban, and Mohammadi 2020). Due to extensive applications of machine learning methods in chemical engineering (Ershadnia et al. 2020), in the present work, a novel very accurate machine learning method is proposed for the estimation of combustion enthalpy of pure chemical compounds based on group contributions. The extreme learning machine (ELM) algorithm has low running time in comparison with other methods. Additionally, user who has some knowledge of artificial intelligence methods can use this algorithm. The simplicity and low cost of ELM highlight the importance of this method in different issues.

## Methodology

The process of the present work can be described as follows: first, a comprehensive experimental databank has been collected, then it is divided into two main groups known as testing and training data sets. After that, the parameters of the ELM model are obtained by using the training data set and next the accuracy of the model is evaluated based on graphical and statistical approaches. Additionally, the outlier detection has been done on the experimental data set to ensure the utilized data set.

# Experimental data set collection

The reliability and accuracy of models for prediction of properties, especially those dealing with a considerable amount of measured data, are highly functional of the comprehensiveness and quality of the applied databank for its training. The mentioned characteristics of such a method are both a large number of pure compounds exist in databank and diversity in the studied chemical groups. In this paper, the databank prepared by Yaws is used, which is known as one of the popular sources of physical properties data, especially enthalpy of combustion for pure compounds. Thus, the enthalpy values for 4,590 materials have been gathered in this source and used for training of ELM algorithm (Yaws and Gabbula 2003). They are divided into 3,443 and 1,147 points for training and testing phases randomly.

#### Extreme learning machine

The complex relationship of multiple variables with a special target (Zhang et al. 2020) such as prediction of a chemical engineering process is really difficult to be determined for an extensive combination of data by applying common approaches (Abdi-Khanghah et al. 2018). To handle this type of issues, many alternative methods have been proposed (Keybondorian, Soltani Soulgani, and Bemani 2018); artificial intelligence modeling is presented as a prediction tool that can be applied to identify and express complicated relationships between input variables and consequent outputs (Zarei and Baghban 2017).

Huang et al. proposed ELM for the first time as a fast machine learning strategy with impressive generalization performance that fundamentally utilized single-hidden layer feedforward neural networks (SLFNNs) with random generation of hidden nodes (Huang, Zhu, and Siew 2006). All the utilized parameters that are tuned in the common feedforward neural networks cause a dependency between different parameters layers. The significant point of ELM approach is that there is no need for the hidden layer to be tuned so training speed is more than FFNNs; furthermore, it has better generalization performance (Huang, Zhu, and Siew 2004)

The smallest norm of output weights, the smallest training error, and best generalization performance are known as the main objectives of the ELM algorithm. This strategy has wide application in dynamic action recognition, terms series estimation, mobile object index (Wang et al. 2012), multimedia recognition (Lan et al. 2013), and bioinformatics (Wang, Zhao, and Wang 2008). The ELM is known as an an efficient strategy with several advantages such as limited human intervention, simplicity of development, and fast training speed so it can be used as a pleasant alternative that acts as a learning base for large-scale feature mapping and computations (Bengio 2009). Figure 1 illustrates the architecture of the ELM algorithm, which has a layered structure. The training ( $x_i$ ) and target ( $y_i$ ) values are assumed in space of {( $x_i, y_i$ ), i = 1, ..., N} in  $\mathbb{R}^n \times \mathbb{R}^m$  so that m and n are dimensions of inputs and outputs, respectively. The ELM algorithm for L nodes can be written as follows (Torabi et al. 2010)



Figure 1. Schematic diagram of ELM algorithm.

#### 4 👄 R. SETIAWAN AND S. MOHAMMADINIA

$$f_L(x) = \sum_{i=1}^L \beta_i h_i(x) = h(x)\beta = \tilde{y}$$
(1)

where weight vector and nonlinear feature mapping are shown by  $\beta = [\beta_1, ..., \beta_L]$  and  $h(x) = [h_1(x), ..., h_L(x)]$ , respectively. In this algorithm, different types of function including cosine, multi quadric, and hyperbolic tangent can be used.  $h_i(x)$  in real situations can be written as Rao and Mitra 2010):

$$h_i(x) = G(a_i, b_i, x) \tag{2}$$

Here,  $b_i isinR$ ,  $a_i isinR^d$ , and nonlinear differential function is shown by *G* as a function of *a* and *b*, which are hidden node parameters. The random feature mapping and solving linear parameter are two main steps of the ELM algorithm. First, it implements inputs weight and hidden biases for estimation of the hidden layer matrix. After that,  $\beta$  is used to connect the hidden layer weights to outputs. These values are calculated based on minimizing squared error (Rao and Mitra 2010):

$$\min_{\beta \in \mathbb{R}^{L \times m}} ||y - u||_2 \tag{3}$$

where *u* is the matrix output for the hidden layer:

$$u = \begin{bmatrix} u_1(x_1) & \cdots & u_L(x_1) \\ \vdots & \ddots & \vdots \\ u_1(x_N) & \cdots & u_L(x_N) \end{bmatrix}$$
(4)

y is a training target matrix, which is defined as follows:

$$y = \begin{bmatrix} y_{11} & \cdots & y_{1m} \\ \vdots & \ddots & \vdots \\ y_1 & \cdots & y_{Nm} \end{bmatrix}$$
(5)

To solve Eq. (3), the below formulation is expressed:

$$\beta^* = \left(u^T u\right)^{-1} u^T y \tag{6}$$

Here,  $u^{\mathrm{T}}$  denotes the transposed u.

# **Results and discussion**

To develop the ELM algorithm for prediction of the combustion enthalpy, the algorithm parameters should be set. To this end, the sigmoid function has been used as an activation function. Moreover, the cross-validation method is used to determine the number of the hidden layer neurons, which is 50 for the heat of combustion estimation model. Table 1 gives information about the estimation results of this algorithm based on the following parameters: mean relative error (MRE), root mean square error (RMSE), standard deviation (STD), mean squared error (MSE), and  $R^2$ . As shown,  $R^2 = 1$ , STD = 11.18, RMSE = 14.92, MSE = 222.62, and MRE = 0.28 are estimated for the overall phase. Furthermore, the predicted and experimental values of the combustion enthalpy are presented in Figure 2 simultaneously. The high degree of agreement is achieved between predicted and experimental enthalpy points. Additionally, the real enthalpy value is illustrated versus ELM enthalpy in

Table 1. Statistic parameters for ELM algorithm in prediction of enthalpy.

······································					
	R <sup>2</sup>	MRE (%)	MSE	RMSE	STD
Train	1.000	0.28	224.95	15.00	11.23
Test	1.000	0.28	215.64	14.68	11.04
Total	1.000	0.28	222.62	14.92	11.18



Figure 2. Simultaneous comparison of actual and predicted combustion enthalpy values.

Figure 3. Two fitting lines are implemented on training and testing data set so that the obtained lines have similar formulations to y = x approximately. These lines show that the ELM algorithm has great accuracy in the prediction of the enthalpy. As shown in Figure 4, the relative deviations between the calculated and real combustion enthalpy values vary in the range of -4 and 8. These values have high compaction alongside *x*-axis, which expresses the quality of accuracy.

One of the important parts of the development of a model is the detection of inaccurate or suspected data points (Torabi et al. 2012). The actual data have straight effect on the performance of models (Cao et al. 2020). To this end, the leverage method should be employed (Pirdashti, Taheri and Dragoi et al. 2020). There are some necessary parts such as Hat matrix in this method (Razavi et al. 2019b):

$$H = X \left( X^T X \right)^{-1} X^T \tag{7}$$

in which *H* represents Hat matrix and *X* points to the  $q \times p$  matrix. *p* and *q* are the model parameters number and number of samples, respectively (Razavi et al. 2019a). The major step of this method is the illustration of William's plot. In this depiction, the reliable portion of data set is located between standard residual indexes (-3 and 3) and a leverage limit, which is calculated as :

$$H^* = 3(p+1)/q$$
(8)

The leverage limit shows the upper limit for reliable data. If the Hat value of a sample is more than leverage limit, the sample point is suspected data. William's plot for the combustion enthalpy data set is shown in Figure 5. Due to this plot, most numbers of the enthalpy data points in this data set are accurate and reliable.



Figure 3. Cross plot of real and forcasted enthalpy.



**○**Train ○Test

Figure 4. Relative deviation of predicted and actual enthalpy values.



Figure 5. William's plot for enthalpy databank.

# Conclusion

The main aim of this paper is the suggestion of a group contribution approach based on ELM algorithm for the calculation of the standard combustion enthalpy for pure chemical materials by using a data set of 4,590 pure compounds. The accuracy of the collected data set has been judged to detect the suspected data for improving ELM prediction. The needed inputs of this approach are the number of available molecular groups among 142 groups in the investigated structure. In this study, the leverage method is used to detect correct data for preparation of the algorithm. TheSTD, RMSE, and MRE for the overall phase are determined to be 11.18, 14.92, and 0.28 for the proposed model. The comparison of experimental data and forecasted enthalpy points has concluded to the fact that our proposed algorithm has great potential in the prediction of combustion enthalpy. The most significant outcomes of this study are the development of a powerful tool to determine combustion enthalpy. Additionally, a mathematical analysis of the accuracy of data set in this topic is introduced. Due to the accuracy of the model and the investigated analysis, the present study can be an interesting study for engineers dealing with combustion process.

# **Notes on contributors**

*Dr. Roy Setiawan* is currently working as Assistant professor at Department of Management, Petra Christian University, Indonasia. He holds his PhD in Management from Universitas Airlangga, Surabaya, Jawa Timur, Indonaisa. He has published several research papers in the journals of international repute in the areas of Leadership, Human Resource Management, Innovative Work Behavior, Green Campus Competitiveness, Organizational Behavior, and Change Management.

*Mrs Samira Mohammadinia* studied chemical engineering at islamic Azad University of Mahshahr. Her research interests are machine learning , thermodynamic modeling, and surface reaction applications.

#### References

- Abdi-Khanghah, M., A. Bemani, Z. Naserzadeh, and Z. Zhang. 2018. Prediction of solubility of N-alkanes in supercritical CO2 using RBF-ANN and MLP-ANN. *Journal of CO2 Utilization* 25:108–19. doi:10.1016/j.jcou.2018.03.008.
- Bemani, A., A. Baghban, and A. H. Mohammadi. 2020. An insight into the modeling of sulfur content of sour gases in supercritical region. *Journal of Petroleum Science and Engineering* 184:106459. doi:10.1016/j. petrol.2019.106459.
- Bengio, Y. 2009. Learning deep architectures for AI. Foundations and Trends<sup>®</sup> in Machine Learning 2 (1):1–127. doi:10.1561/2200000006.
- Cao, B., J. Zhao, Z. Lv, Y. Gu, P. Yang, and S. K. Halgamuge. 2020. Multiobjective Evolution of Fuzzy Rough Neural Network via Distributed Parallelism for Stock Prediction. *IEEE Transactions on Fuzzy Systems* 28 (5):939–52. doi:10.1109/TFUZZ.2020.2972207.
- Cardozo, R. 1986. Prediction of the enthalpy of combustion of organic compounds. AIChE Journal 32 (5):844-48. doi:10.1002/aic.690320514.
- Chen, H., A. Chen, L. Xu, H. Xie, H. Qiao, Q. Lin, and K. Cai. 2020. A deep learning CNN architecture applied in smart near-infrared analysis of water pollution for agricultural irrigation resources. *Agricultural Water Management* 240:106303. doi:10.1016/j.agwat.2020.106303.
- Darvishan, A., H. Bakhshi, M. Madadkhani, M. Mir, and A. Bemani. 2018. Application of MLP-ANN as a novel predictive method for prediction of the higher heating value of biomass in terms of ultimate analysis. *Energy Sources*, *Part A: Recovery, Utilization, and Environmental Effects* 40 (24):2960–66. doi:10.1080/15567036.2018.1514437.
- Ershadnia, R., M. A. Amooie, R. Shams, S. Hajirezaie, Y. Liu, S. Jamshidi, and M. R. Soltanian. 2020. Non-Newtonian fluid flow dynamics in rotating annular media: Physics-based and data-driven modeling. *Journal of Petroleum Science* and Engineering 185:106641. doi:10.1016/j.petrol.2019.106641.
- Gharagheizi, F. 2008. A simple equation for prediction of net heat of combustion of pure chemicals. *Chemometrics and Intelligent Laboratory Systems* 91 (2):177–80. doi:10.1016/j.chemolab.2007.11.003.
- Gharagheizi, F., S. A. Mirkhani, and A.-R. Tofangchi Mahyari. 2011. Prediction of standard enthalpy of combustion of pure compounds using a very accurate group-contribution-based method. *Energy & Fuels* 25 (6):2651–54. doi:10.1021/ef200081a.
- Hshieh, F. Y., D. B. Hirsch, and H. D. Beeson. 2003. Predicting heats of combustion of polymers using an empirical approach. *Fire and Materials* 27 (1):9–17. doi:10.1002/fam.815.
- Huang, G.-B., Q.-Y. Zhu, and C.-K. Siew. 2004. Extreme learning machine: A new learning scheme of feedforward neural networks. Neural Networks 2:985–90.
- Huang, G.-B., Q.-Y. Zhu, and C.-K. Siew. 2006. Extreme learning machine: Theory and applications. *Neurocomputing* 70 (1–3):489–501. doi:10.1016/j.neucom.2005.12.126.
- International, A. 2009. Standard Test Method for Heat of Combustion of Liquid Hydrocarbon Fuels by Bomb Calorimeter. ASTM International.
- Keybondorian, E., B. Soltani Soulgani, and A. Bemani. 2018. Application of ANFIS-GA algorithm for forecasting oil flocculated asphaltene weight percentage in different operation conditions. *Petroleum Science and Technology* 36 (12):862–68. doi:10.1080/10916466.2018.1447960.
- Keybondorian, E., H. Zanbouri, A. Bemani, and T. Hamule. 2017. Estimation of the higher heating value of biomass using proximate analysis. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects* 39 (20):2025–30. doi:10.1080/15567036.2017.1400609.
- Lan, Y., Z. Hu, Y. C. Soh, and G.-B. Huang. 2013. An extreme learning machine approach for speaker recognition. *Neural Computing and Applications* 22 (3–4):417–25. doi:10.1007/s00521-012-0946-x.
- Li, T., M. Xu, C. Zhu, R. Yang, Z. Wang, and Z. Guan. 2019. A deep learning approach for multi-frame in-loop filter of HEVC. *IEEE Transactions on Image Processing* 28 (11):5663–78. doi:10.1109/TIP.2019.2921877.
- Mazloom, M. S., F. Rezaei, A. Hemmati-Sarapardeh, M. M. Husein, S. Zendehboudi, and A. Bemani. 2020. Artificial Intelligence Based Methods for Asphaltenes Adsorption by Nanocomposites: Application of Group Method of Data Handling, Least Squares Support Vector Machine, and Artificial Neural Networks. *Nanomaterials* 10 (5):890. doi:10.3390/nano10050890.
- Pan, Y., J. Jiang, R. Wang, and J. Jiang. 2011. Predicting the net heat of combustion of organic compounds from molecular structures based on ant colony optimization. *Journal of Loss Prevention in the Process Industries* 24 (1):85–89. doi:10.1016/j.jlp.2010.11.001.
- Pirdashti, M., Taheri, M., Dragoi, E.N. and Curteanu, S., 2020. Machine Learning Approaches for Prediction of Phase Equilibria in Poly (Ethylene Glycol)+ Sodium Phosphate Aqueous Two-Phase Systems. *Iranian Journal of Chemistry* and Chemical Engineering (IJCCE),39(6), pp.185–197.
- Razavi, R., A. Bemani, A. Baghban, A. H. Mohammadi, and S. Habibzadeh. 2019a. An insight into the estimation of fatty acid methyl ester based biodiesel properties using a LSSVM model. *Fuel* 243:133–41. doi:10.1016/j.fuel.2019.01.077.
- Razavi, R., A. Sabaghmoghadam, A. Bemani, A. Baghban, K.-W. Chau, and E. Salwana. 2019b. Application of ANFIS and LSSVM strategies for estimating thermal conductivity enhancement of metal and metal oxide based

nanofluids. Engineering Applications of Computational Fluid Mechanics 13 (1):560-78. doi:10.1080/19942060.2019.1620130.

- Seaton, W. H., and B. K. Harrison. 1990. A new general method for estimation of heats of combustion for hazard evaluation. Journal of Loss Prevention in the Process Industries 3 (3):311–20. doi:10.1016/0950-4230(90)80025-6.
- Shi, K., Y. Tang, S. Zhong, C. Yin, X. Huang, and W. Wang. 2018. Nonfragile asynchronous control for uncertain chaotic Lurie network systems with Bernoulli stochastic process. *International Journal of Robust and Nonlinear Control* 28 (5):1693–714. doi:10.1002/rnc.3980.
- Suleymani, M., and A. Bemani. 2018. Prediction of the interfacial tension between hydrocarbons and carbon dioxide. *Petroleum Science and Technology* 36 (3):227–31. doi:10.1080/10916466.2017.1416629.
- Torabi, D. Hesam, and N. A. Masoudi. 2010. *Mining biological repetitive sequences using support vector machines and fuzzy SVM*. 1–17.
- TORABI, D. HESAM, N. A. MASOUDI, and F. Zare. 2012. Finding Exact and Solo LTR-Retrotransposons in Biological Sequences Using SVM. 111–116.
- Wang, B., G. Wang, J. Li, and B. Wang. 2012. Update strategy based on region classification using ELM for mobile object index. Soft Computing 16 (9):1607–15. doi:10.1007/s00500-012-0821-9.
- Wang, G., Y. Zhao, and D. Wang. 2008. A protein secondary structure prediction framework based on the extreme learning machine. *Neurocomputing* 72 (1–3):262–68. doi:10.1016/j.neucom.2008.01.016.
- Xu, W., S. Qu, L. Zhao, and H. Zhang. 2020. An improved adaptive sliding mode observer for middle-and high-speed rotor tracking. *IEEE Transactions on Power Electronics* 36 (1):1043–53. doi:10.1109/TPEL.2020.3000785.
- Yaws, C. L., and C. Gabbula. 2003. Yaws" Handbook of Thermodynamic and Physical Properties of Chemical Compounds. Netherlands: Knovel.
- Zarei, F., and A. Baghban. 2017. Phase behavior modelling of asphaltene precipitation utilizing MLP-ANN approach. *Petroleum Science and Technology* 35 (20):2009–15. doi:10.1080/10916466.2017.1377233.
- Zhang, H., S. Qu, H. Li, J. Luo, and W. Xu. 2020. A Moving Shadow Elimination Method Based on Fusion of Multi-Feature. IEEE Access 8:63971–82. doi:10.1109/ACCESS.2020.2984680.