

# Comparing artificial neural network algorithms for prediction of higher heating value for different types of biomass

Olga M. Jakšić (✉ [olga@nanosys.ihtm.bg.ac.rs](mailto:olga@nanosys.ihtm.bg.ac.rs))

University of Belgrade: Univerzitet u Beogradu <https://orcid.org/0000-0002-0937-3677>

Zoran Jakšić

University of Belgrade: Univerzitet u Beogradu

Koushik Guha

National Institute of Technology Silchar

Ana G. Silva

Universidade Nova de Lisboa Faculdade de Ciências e Tecnologia: Universidade Nova de Lisboa  
Faculdade de Ciências e Tecnologia

---

## Research Article

**Keywords:** Artificial neural networks, Biomass, Higher heating value, Proximate analysis

**Posted Date:** March 3rd, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-267108/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

## Manuscript title page

**Manuscript title:** Comparing artificial neural network algorithms for prediction of higher heating value for different types of biomass

### Author information

- <sup>1</sup>\*Olga Jakšić • <sup>1</sup>Zoran Jakšić • <sup>2</sup>Koushik Guha • <sup>3</sup>Ana G. Silva
- <sup>1</sup>University of Belgrade (Institute of Chemistry, Technology and Metallurgy – ICTM, Centre of Microelectronic Technologies), Belgrade, Serbia; <sup>2</sup>National MEMS Design Centre, Department of Electronics and Communication Engineering, National Institute of Technology, Silchar, Assam, India; <sup>3</sup>CeFiTec, Faculty of Science and Technology, New University of Lisbon, FCT-NOVA, Lisbon, Portugal
- \*e-mail address of the corresponding author: olga@nanosys.ihtm.bg.ac.rs
- Olga Jakšić (0000-0002-0937-3677), Zoran Jakšić (0000-0003-4952-7607), Koushik Guha (0000-0002-5046-2786), Ana G. Silva (0000-0001-7590-3042)

**Abstract** – The paper presents the comparative qualitative and quantitative analysis of twelve algorithms for training artificial neural networks (ANN) which predict the higher heating value (HHV) of biomass based on the proximate analysis (fixed carbon, volatile matter, and ash percentage). The twelve networks, with the same structure but different training algorithm, were fed with 318 experimental data triplets from literature for different biomass species and trained with 318 corresponding HHV values used for the supervised learning. Our comparative analysis showed that several algorithms resulted in ANNs generating outputs well correlated with the true measured values of the biomass HHV. Of those, Levenberg-Marquardt algorithm gives the best results in terms of mean squared error calculated on the training set data, while Bayesian regularization gives the best results in terms of regression. When applied to new datasets, unknown to the ANNs trained here, the highest accuracy of the HHV prediction was obtained by Conjugate Gradient, Powell/Beale Restarts training function. It ensured prediction based on the unknown datasets better than Levenberg-Marquardt algorithm. The described approach can be used for predicting the calorific values of different biomass species, including newly proposed ones, as well as for optimizing the HHV for both pure biomass and biomass mixtures.

**Keywords:** Artificial neural networks · Biomass · Higher heating value · Proximate analysis.

### Declarations

**Funding** This work was financially supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia, Grant No. 451-03-68/2020-14/200026.

**Conflicts of interest/Competing interests** The authors declare no conflict of interest.

**Availability of data and material** All data are freely available on Mendeley Data online repository.

**Code availability** All custom code is freely available on Mendeley Data online repository.

**Authors' contributions** (optional: please review the submission guidelines from the journal whether statements are mandatory)

**Ethics approval** N.A.

**Consent to participate** All authors expressed their explicit consent to participate in this work.

**Consent for publication** All authors expressed their explicit consent with the publication of this joint manuscript in Soft Computing.

# Comparing artificial neural network algorithms for prediction of higher heating value for different types of biomass

<sup>1</sup>Olga Jakšić • <sup>1</sup>Zoran Jakšić • <sup>2</sup>Koushik Guha • <sup>3</sup>Ana G. Silva

Received: (date)

## Abstract

The paper presents the comparative qualitative and quantitative analysis of twelve algorithms for training artificial neural networks (ANN) which predict the higher heating value (HHV) of biomass based on the proximate analysis (fixed carbon, volatile matter, and ash percentage). The twelve networks, with the same structure but different training algorithm, were fed with 318 experimental data triplets from literature for different biomass species and trained with 318 corresponding HHV values used for the supervised learning. Our comparative analysis showed that several algorithms resulted in ANNs generating outputs well correlated with the true measured values of the biomass HHV. Of those, Levenberg-Marquardt algorithm gives the best results in terms of mean squared error calculated on the training set data, while Bayesian regularization gives the best results in terms of regression. When applied to new datasets, unknown to the ANNs trained here, the highest accuracy of the HHV prediction was obtained by Conjugate Gradient, Powell/Beale Restarts training function. It ensured prediction based on the unknown datasets better than Levenberg-Marquardt algorithm. The described approach can be used for predicting the calorific values of different biomass species, including newly proposed ones, as well as for optimizing the HHV for both pure biomass and biomass mixtures.

**Keywords** Artificial neural networks · Biomass · Higher heating value · Proximate analysis

## Introduction

The biomass is usually agricultural waste or waste in the food and wood industry. In the paradigm of circular economy, it is of high importance for the environmental wellbeing to explore the possibilities to use the biomass as an alternative, non-toxic, renewable energy source. The main benefit from the transition from fossil fuels to renewable energy sources is the reduction of greenhouse gas emission, the most important goal in strategies for sustainable development (Klass 1998). The number of potential plant species proposed for the use as a biomass-based energy source is growing every day and there is a need for the estimation of calorific values of various types of biomass (Channiwala et al. 2002), (Demirbaş et al. 2004), or for the optimal design of the system with renewable energy sources (Krishnamoorthy et al. 2020).

Among biomass sources belong plant pits, shells, seeds, cobs, prunings, stalks, leaves, husks, grass... Common biomass resources considered to produce briquettes and pellets like corn, wheat, sunflower, barley and oats, are accompanied by tobacco (Mijailović et al. 2014), (Liang et al. 2019). Most of these resources can also be used to produce gaseous fuels like methane (Wang et al. 2019). Besides methane production, biomass is also employed in the production of bio-adsorbents, composts, insecticides, land recovery of nicotine, chlorogenic acid and various agents, as well as in bioremediation (Alagić et al. 2015). The use of agricultural waste is versatile. Besides the mentioned use as an alternative fuel it may be used for other alternative energy sources also – for instance, cashew apple juice and many other kinds of agricultural waste may be used as a substrate for microbial fuel cells (Priya et al. 2019). However, the use of biomass as an alternative energy source has its disadvantages. Beside the lack of standardization, the main drawback of alternative energy sources based on the biomass is their low calorific value.

The main figure of merit for the calorific value of the biomass is its higher heating value (HHV), defined as the amount of heat released by a given quantity of fuel. In general, regarding the production of any fuel on an industrial scale, for system design and analysis, for an efficient generation of heat and power, the knowledge on the HHV value is crucial. Experimental determination of the HHV is most reliable, however measuring of HHV is not always an option and there are many efforts in

---

✉ Olga Jakšić  
olga@nanosys.ihtm.bg.ac.rs

<sup>1</sup> University of Belgrade, Institute of Chemistry, Technology and Metallurgy (ICTM), Centre of Microelectronic Technologies, Belgrade, Serbia

<sup>2</sup> National MEMS Design Centre, Department of Electronics and Communication Engineering, National Institute of Technology, Silchar, Assam, India

<sup>3</sup> CeFiTec, Faculty of Science and Technology, New University of Lisbon, FCT-NOVA, Lisbon, Portugal

determining HHV otherwise. There is a correlation between HHV and the composition of the fuel. For instance, it has been shown that the fraction of oxygen can be a predictor of HHV of gaseous, liquid, and solid fuels (Merckel et al. 2017). Due to the fact that HHV is correlated with the composition of the raw material used as fuel, there are numerous mathematical models for the estimation of calorific values of fuels based on ultimate analysis, proximate analysis, physical composition, chemical composition or structural analysis (Krishnan et al. 2018), (Cordero et al. 2001), (Nhuchhen et al. 2012)... In these papers, as well as in the references cited therein, it is shown that HHV can be determined based on the proximate analysis, i.e. using the known percentages of the fixed carbon (FC), volatile materials (VM) and ash (ASH). Naturally these percentages, residues from the combustion of the raw material must sum up to 1 (FC+VM+ASH=1). Since, the proximate analysis can be performed by relatively simple lab equipment, in comparison with ultimate elemental analysis, the determination of HHV based on proximate analysis is important.

Apart from the analytical approach to the determination of HHV based on the proximate analysis, the characterization of the biomass in terms of its HHV using artificial neural networks is an active field of research (Olatunji et al. 2020), (Abdulsalam et al. 2020), (Ghugare et al. 2014)... A nonlinear correlation between the higher heating value and the proximate, ultimate analysis has been proven (Aydinli et al. 2017), (Dashti et al. 2019). The ANNs trained on a relatively small set of data sets (25, referring to rice husks) outperformed empirical equations when compared to experimental HHV data (Yu et al. 2014). Compared to other soft computing techniques (Multilinear regression and Gene Expression Programming), ANN showed better results in predicting HHV of hydrothermally carbonized biomass (Abdulsalam et al. 2020). While HHV of biochar predictions in the quoted reference were based on hydrothermal carbonization temperature, biomass residence time in the reactor, and the composition of biomass as inputs, the majority of empirical and software-assisted correlations for the estimation of biomass fuel HHV reported in literature were based on their proximate, ultimate, and chemical analysis (Boumanchar et al. 2019).

Since it has been shown that software-assisted solutions, particularly ANNs, often outperform empirical correlations in terms of the *coefficient of determination*,  $R$ , with regard to experimentally measured HHV values, here we explore 12 different algorithms for training the ANNs aimed for the prediction of HHV based on known results of the proximate analysis.

## Methodology

This section introduces briefly ANN based modelling, the paradigm of learning, types of training algorithms and figures of merit of created ANNs that will be used for exemplary data in subsequent sections.

In the context of the Tom Mitchell's definition of machine learning ("A computer program is said to learn from experience  $E$  with the respect to some class of tasks

$T$  and performance measure  $P$ , if its performance at tasks in  $T$ , as measured by  $P$ , improves with experience  $E$ "), this method is based on the following interpretation  $E$  – experience of repeating determination of HHV based on given percentages of FC, VM and ASH, as per the proximate analysis of the biomass.

$T$  – task of computing one HHV value based on one triplet of input data (FC, VM and ASH).

$P$  – statistical performance metrics based on minimization of error of the computed value compared to the true value proved by the experiment.

Typical ANN learning tasks are function approximation, selection and clustering. The ANN here aims to solve nonlinear regression problem.

What we do is create a multiple-input single output (MISO), forward feed backward propagation multilayer shallow artificial neural networks by using algorithms with supervised learning (contrary to deep neural networks predominantly used for image classification). Supervised learning was realized by data sets based on experimentally obtained data from the literature, used as examples of proper network behaviour (proper network behaviour being the combination of proximate analysis results at the input of ANN, and corresponding measured HHV value at the output of ANN). In all cases the network was fed (supervised) with the same sets of input/output data.

Forward feed refers to the direction of data processing: input data feed the first hidden layer where each neuron is calculated in a way that a sum of weighted input data and a bias is subjected to an activation function. Every subsequent hidden layer is fed with the neuron outputs of the previous one and the output layer is fed with the neuron outputs of the last hidden layer. The obtained output of the ANN is evaluated with regard to the target values in the datasets used for supervising. Backward propagation refers to error computation throughout the network. Errors are calculated with respect to weights and then used for optimization of network performances which iteratively affects the adjustment of weights throughout the network. The weights are updated after all the inputs in the training set are applied – the batch training mode was chosen (opposed to incremental training mode where weights are updated after each of three inputs from the input data set is applied).

The quality of results was evaluated through MSE and regression values. MSE - Mean Squared Error is the average squared difference between outputs generated by the MATLAB function and the targets (true, measured HHV data that correspond to inputs given to the MATLAB functions). Lower values are obviously better. The algorithm with the minimal MSE was considered the most appropriate. Equation for the calculation of MSE is

$$MSE = \frac{1}{N} \sum_{i=1}^N (e_i)^2 = \frac{1}{N} \sum_{i=1}^N (t_i - a_i)^2 \quad (1)$$

$N$  is the number of samples (input-output pairs) used for training the network,  $t$ , is the target value, here it is the measured value for HHV of biomass,  $a$  is the value

calculated by the ANN,  $e$  is the error, *i.e.* the difference between the target and the calculated value.

The regression values  $R$  measures the correlation between the obtained ANN outputs and the targets. An  $R$  value of 1 means a close relationship, 0 is a random relationship. The expression related to the calculation of  $R$  is

$$R^2 = 1 - \frac{\sum_{i=1}^N (t_i - a_i)^2}{\sum_{i=1}^N (\bar{t} - t_i)^2} \quad (2)$$

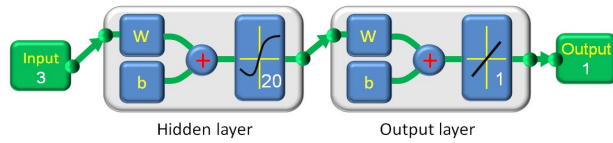
The notation is the same as for the calculation of MSE and  $\bar{t}$  is the arithmetic mean of the target values.

After training multiple ANNs and performing the comparative analysis, the algorithm with the highest correlation between the ANN output and the target values was considered to be the most appropriate for the estimation of HHV based on the known results of the proximate analysis of the biomass. All networks are cross-validated with the set of new data, unknown prior to the training and learning process.

### The structure of Artificial Neural Networks

For the comparative analysis of different algorithms for training the ANNs the same network structure is used but with different training functions.

The structure of all twelve artificial networks is the same. It is shown in Fig. 1.



**Fig. 1** The simplified presentation of the ANN structure used in all our algorithms for the determination of the HHV based on the proximate analysis of the biomass.

The ANNs use backpropagation as a learning algorithm. There are three inputs for all networks. They are the same, the results of the proximate analysis: the percentages of the fixed carbon (FC), volatile materials (VM) and ash (ASH). The networks have one hidden layer with 20 neurons, one output layer with 1 neuron and one output – the predicted HHV of the biomass. The number of neurons in the ANN is a parameter of interest in the process of choosing the network structure. If it is too small, the generated outputs do not converge well to the desired targets. If it is too high, the network predicts well in the framework of the given training set of data but may be not sufficiently good in predicting the output when fed with new, unknown data. The number of neurons in the hidden layer here was chosen as one that gave the best results considering outputs from ANN fed with known and unknown data.

The neuron values in both the hidden and the output layer are calculated in a similar manner: a sum of the

weighted values provided by the previous layer and some bias term is passed through an activation function. Some common activation functions are purely linear, and two sigmoid functions (logistic function and hyperbolic tangent, defined by expressions (3), (4) and (5), respectively.

$$y = x \quad (3)$$

$$y = \frac{1}{1 + e^{-x}} \quad (4)$$

$$y = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (5)$$

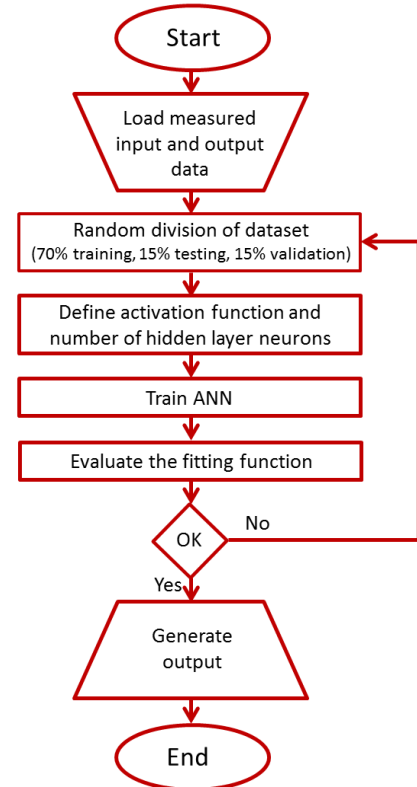
In the algorithms investigated in this paper the following sigmoid function (Sigmoid Symmetric Transfer Function) was used

$$y = \frac{2}{1 + e^{-2x}} - 1 \quad (6)$$

The sigmoid activation function is used in the hidden layer and linear activation function in the output layer.

### Training algorithms

Figure 2 shows the diagram of the training process that includes loading the collected data, creating and configuring the network, initializing the weights and biases, training and validating the network, storing outputs of interest for future usage of the network. The step ‘Train ANN’ refers to training by using one of twelve training algorithms available in MathWorks MATLAB or Octave environment. Short names and the acronyms of the training functions are given in Table 1.



**Fig. 2** The workflow diagram for training our ANNs.

There is no simple rule to choose the best training algorithm. Some are better/faster in solving regression problems (finding optimal function approximation) and some show better results in pattern recognition. The speed of a training algorithm depends on the complexity of the network structure, on the complexity of the problem represented by datasets, on the amount of data in datasets, etc. The same workflow presented in Fig 2. produces different results if different training functions are employed. Moreover, due to the nature of the training process itself, the same workflow presented in Fig. 2 produces different results if employed with the same training function repeatedly. Multiple training can be performed in search for the best possible outcome of each separate algorithm. Here, shallow learning is performed, multiple training is performed manually.

**Table 1** Short names and the acronyms of training functions

#	Training function	Acronym
1	Levenberg-Marquardt	LM
2	Bayesian Regularization	BR
3	Scaled Conjugate Gradient	SCG
4	Resilient Backpropagation	RP
5	One Step Secant	OSS
6	Variable Learning Rate Gradient Descent	GDX
7	Gradient Descent Momentum	GDM
8	Gradient Descent	GD
9	Polak-Ribière Conjugate Gradient	CGP
10	Fletcher-Powell Conjugate Gradient	CGF
11	Conjugate Gradient, Powell/Beale Restarts	CBG
12	BFGS Quasi Newton	BFG

A special attention was given here to generating the output after training the network. In this work, the output block in Fig. 2 refers to more than the results related to network performances. It also refers to the functions that can be used for a future ANN usage on new datasets and also to scripts for recreating ANN functions which may lead to their new versions with better performances.

### Input data for training ANNs

Data sets were formed on the basis of literature data on measured HHV of biomass characterized by the proximate analysis. The final set of data consisted of 318 training input records and corresponding 318 outputs. All input data, used in the process of training, testing and validation of ANNs, are taken from the references (Demirbaş et al. 2004), (Krishnan et al. 2018) and

(Nhuchhen et al. 2012) denoted in tables with i, ii and iii, respectively. The whole data set along with the custom programmed standalone MATLAB functions is published at Mendeley Data repository with open access (Jakšić 2021). As for the exemplary information in this paper, shortened lists of data are given in Tables 2–5.

**Table 2** Results of proximate analysis of biomass, measured values for species – straws, references are given in the main text

Raw materials straws [Reference]	Proximate analysis			HHV (MJ/kg)
	FC	VM	ASH	
Corn straw [iii]	19,19	73,15	7,65	17,68
Jawar straw [iii]	15,15	75,97	8,88	17,95
Millet straw [iii]	16,45	78,28	5,27	18,05
Paddy straw [iii]	11,8	72,7	15,5	14,52
Paddy straw [ii]	11,8	72,7	15,5	14,552
Rape straw [iii]	17,81	76,54	4,65	18,34
Rice straw [iii]	13,91	65,7	20,38	14,85
Sugar cane straw [iii]	14,6	76,2	9,2	17,19
Wheat straw [iii]	19,8	71,3	8,9	17,51
Wheat straw [i]	19,8	71,3	8,9	17,5
Wheat straw [ii]	23,5	63	13,5	17

**Table 3** results of proximate analysis of biomass, measured values for species – wood, references are given in the main text

Raw materials wood [Reference]	Proximate analysis			HHV (MJ/kg)
	FC	VM	ASH	
Alabama oak waste [iii]	21,9	74,7	3,3	19,23
Alabama oak waste [ii]	21,9	74,7	3,3	19,228
Beech wood [iii]	24,6	74	0,4	19,2
Beech wood [i]	24,6	74	0,7	19,9
Block wood [iii]	14,59	83,32	2,09	18,26
Block wood [ii]	14,59	83,32	2,09	18,261
Eucalyptus wood [iii]	15,91	78,48	4,11	16,42
Eucalyptus wood [ii]	21,3	75,35	3,35	18,64
Red wood [iii]	16,1	83,5	0,4	21,03
Red wood [ii]	19,92	79,72	0,36	20,72
Spruce wood [iii]	28,3	70,2	1,5	20,1
Spruce wood [i]	28,3	70,2	0,5	20,3
Subabul wood [iii]	18,52	81,02	1,2	19,78
Subabul wood [i]	18,52	81,02	1,2	19,777



**Table 4** Results of proximate analysis of biomass, measured values for species – husks, references are given in the main text

Raw materials husks [Reference]	Proximate analysis			HHV (MJ/kg)
	FC	VM	ASH	
Olive husk [i]	26,1	70,3	3,6	19
Paddy husk [ii]	15,1	62	16,5	17,141
Rice husk [iii]	16,95	61,81	21,24	14,69
Sal seed husk [iii]	28,06	62,54	9,4	20,6
Sal seed husk [ii]	28,06	62,54	9,4	20,06
Olive husk [iii]	26,1	70,3	3,6	19
Coconut husk [ii]	18	71	6	15,526
Broad bean [iii]	8,68	85,44	5,88	16,07
Rice husk [iii]	16,3	61,2	22,5	16,47

**Table 5** Results of proximate analysis of biomass, measured values for species – leaves, references are given in the main text

Raw materials leaves [Reference]	Proximate analysis			HHV (MJ/kg)
	FC	VM	ASH	
Casuarina equisetifolia [iii]	16,46	73,5	3,93	18,48
Lantana camara leaf [iii]	11,83	70,46	7,26	18,5
Moringa oleifera [iii]	10,7	67,8	21,5	14,23
Sena leaves [iii]	25,5	57,2	17,3	18,13
Sena leaves [ii]	25,5	57,2	17,3	18,125
Sugar cane leaves [iii]	14,9	77,4	7,7	17,41
Sugar cane leaves [ii]	14,9	77,4	7,7	17,41
Tobacco leaf [iii]	11,2	72,6	17,2	15

## Results and discussion

The results related to network performances allowed for the quantitative comparative analysis of the utilized training algorithms. All our scripts designed for training the networks, all of the generated ANN functions, all of the performances of the created networks and all of the data used for training ANNs in this paper are stored in the Mendeley Data, an online data repository with open access (Jakšić 2021).

Table 6 presents the results related to the mean squared error, MSE. Table 7 presents the results related to the coefficient of determination, R.

It is often convenient to normalize the input data to the  $(-1, 1)$  interval in order to avoid over fitting and to ensure dimensional uniformity. Since input data in our case are of the same order of magnitude, and the network contains only one hidden layer with 20 neurons, the normalization of the data was not performed. That is the reason that in

results gathered in Table 6 MSE for some training functions goes as high as 278,7862 (even higher values were obtained during the training process). Normalizing input data only for training ANNs with training functions that showed bad results (high MSE, low R) was not an option either.

**Table 6** Mean squared error of ANNs for various training functions and for training, validation and testing datasets.

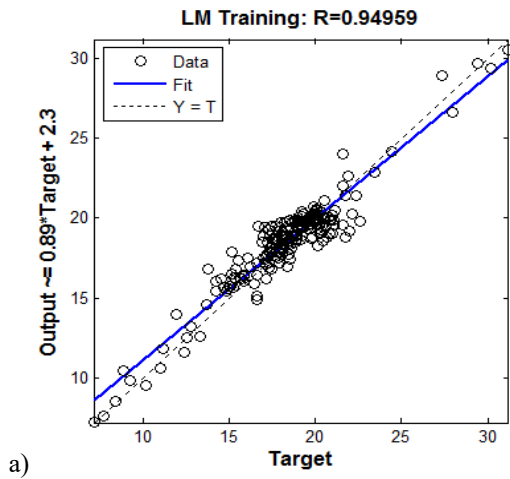
Training function	Training	Validation	Testing
Levenberg-Marquardt	0,896034	0,155095	0,969185
Bayesian Regularization	1,025	0	0,946005
Scaled Conjugate Gradient	1,1951	1,44898	1,09353
Resilient Backpropagation	2,1193	2,2943	1,6687
One Step Secant	1,4843	1,1892	1,4422
Variable Learning Rate Gradient Descent	2,4725	2,0742	5,8071
Gradient Descent Momentum	202,3968	278,7862	221,7867
Gradient Descent	181,3289	205,3820	192,3658
Polak-Ribière Conjugate Gradient	1,9354	1,7645	1,0998
Fletcher-Powell Conjugate Gradient	3,7774	1,7271	5,9445
Conjugate Gradient, Powell/Beale Restarts	0,9472	3,0801	1,9221
BFGS Quasi Newton	1,1067	1,0216	1,1801

**Table 7** The coefficient of determination of ANNs for various training functions and datasets (training, validation and testing)

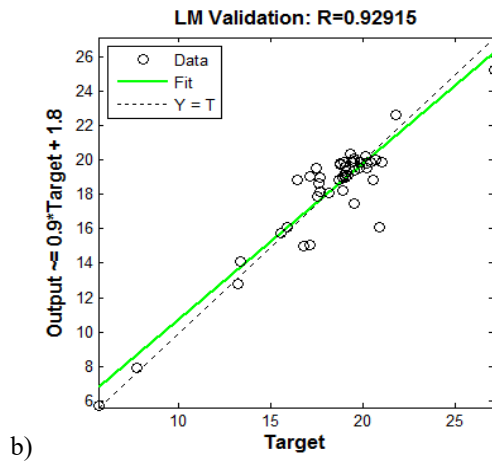
Training function \ R	Training	Validation	Testing
Levenberg-Marquardt	0,936368	0,919951	0,943208
Bayesian Regularization	0,944748	1	0,956280
Scaled Conjugate Gradient	0,93724	0,898488	0,94857
Resilient Backpropagation	0,8513	0,8378	0,8850
One Step Secant	0,8984	0,9195	0,9015
Variable Learning Rate Gradient Descent	0,8239	0,8547	0,4955
Gradient Descent Momentum	0 + 5,03i	0 + 5,93i	0 + 5,27i
Gradient Descent	0 + 4,74i	0 + 5,07i	0 + 4,9i
Polak-Ribière Conjugate Gradient	0,8652	0,8779	0,9258
Fletcher-Powell Conjugate Gradient	0,7136	0,8807	0,4772
Conjugate Gradient, Powell/Beale Restarts	0,9365	0,7745	0,8662
BFGS Quasi Newton	0,9253	0,9313	0,9202

The quantitative comparative analysis of all training functions was made so that every step in the workflow shown in Figure 1 is the same for all the training processes (loading the same set of collected data, creating and configuring the network with the same structure for all, etc.).

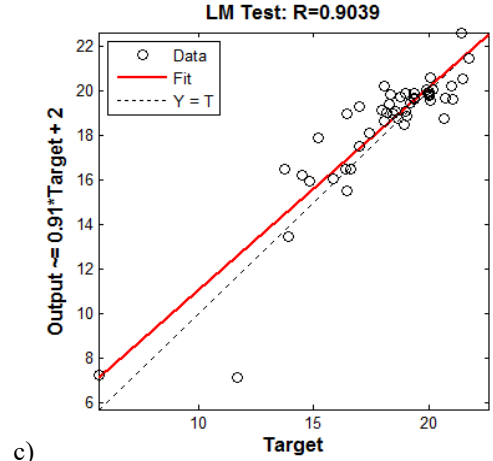
In terms of MSE calculated on a set of training data, the best performance (minimal MSE), had the Levenberg-Marquardt (LM) training function. For MSE calculated both for sets of validation and testing data, LM ranked second place. Since the coefficient of determination ( $R$ ) is related to MSE, as shown in Eq. (1) and Eq. (2), the ranking of LM training function is also high when  $R$  is used as the parameter for the comparison of training functions performances. For  $R$  calculated on a set of training data, the LM training function is ranked fourth, after BR, SCG and CBG training functions. For  $R$  calculated on a set of validation data, LM training function is ranked third, after BR and BFG training functions and for  $R$  calculated on a set of testing data, LM training function is also ranked third, after BR and SCG training functions. Figure 3 shows the regression of LM training function.



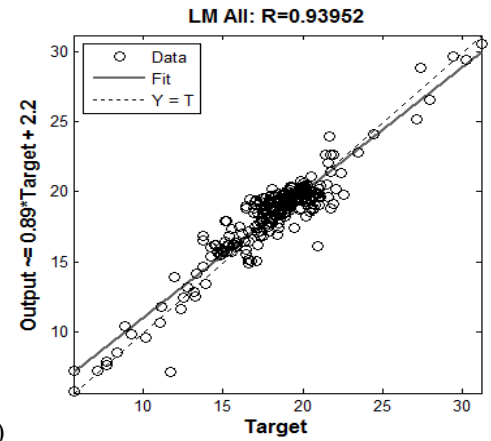
a)



b)



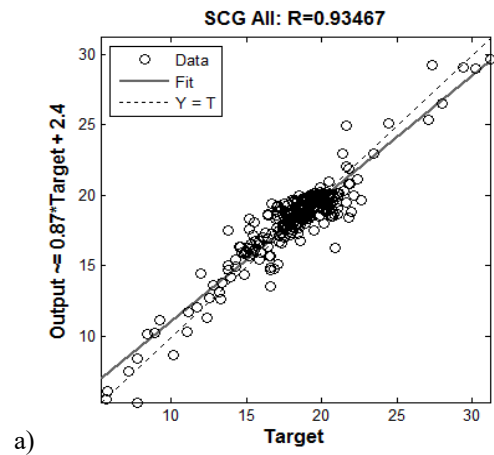
c)



d)

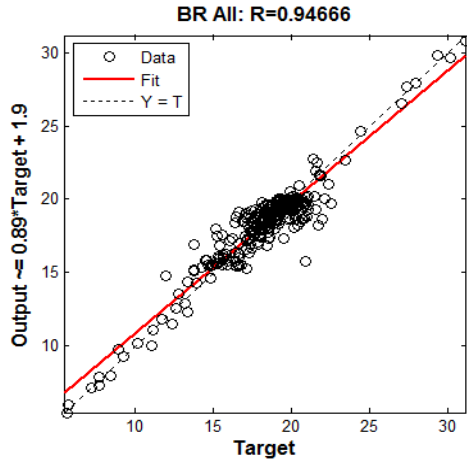
**Fig. 3** The regression of LM algorithm for different sets of data

Graphical representations of regressions for all twelve training algorithms along with the expressions for regression lines in the form  $\text{Output} = a\text{Target} + b$ , ( $a$  and  $b$  are coefficients), for training data, validation data, test data and all data, are not given here, but they are freely available for download from the open access Mendeley Data repository [20], along with all input, target data, designed scripts and generated ANN functions. Besides the results presented in Tables 6 and 7, graphical representations of all the training functions are given in Fig. 4, which shows the regression lines calculated for the whole set of input data.

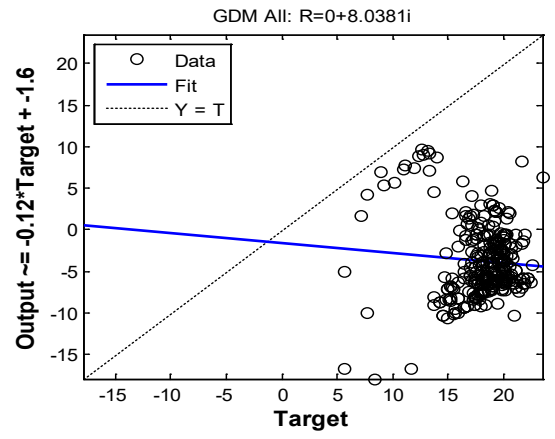


a)

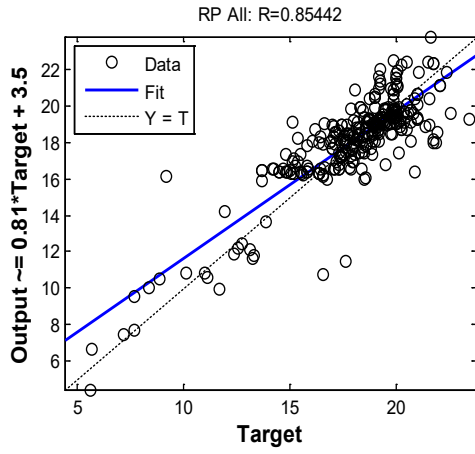




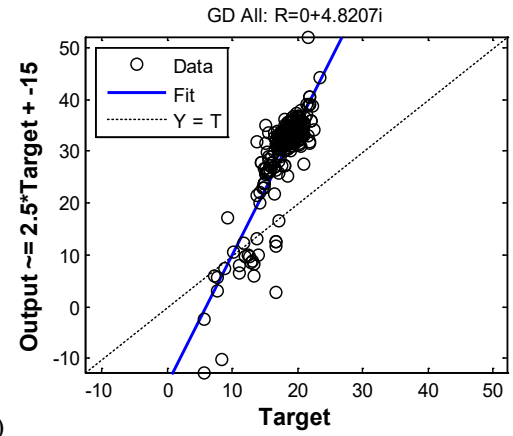
b)



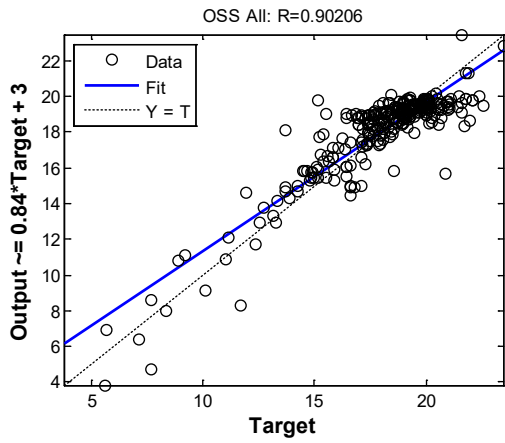
f)



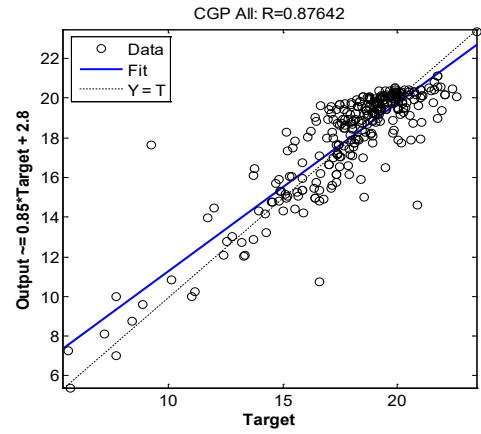
c)



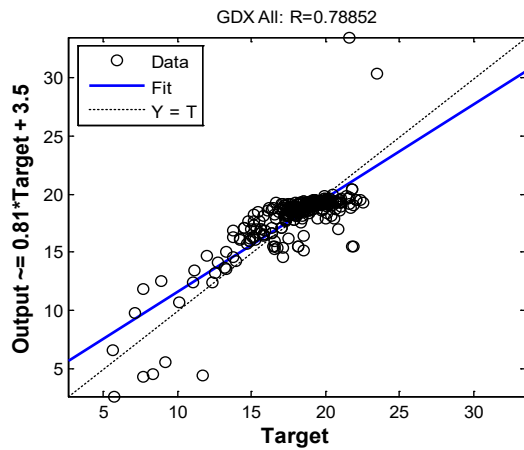
g)



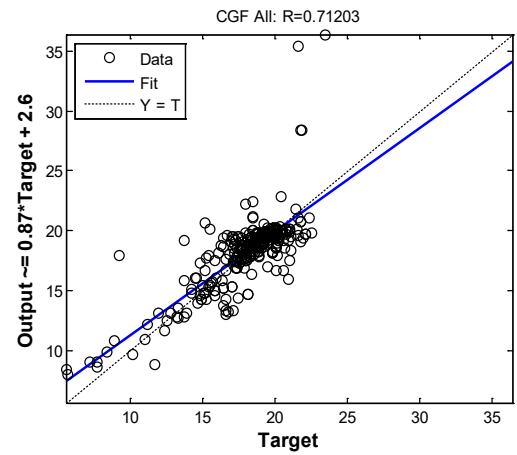
d)



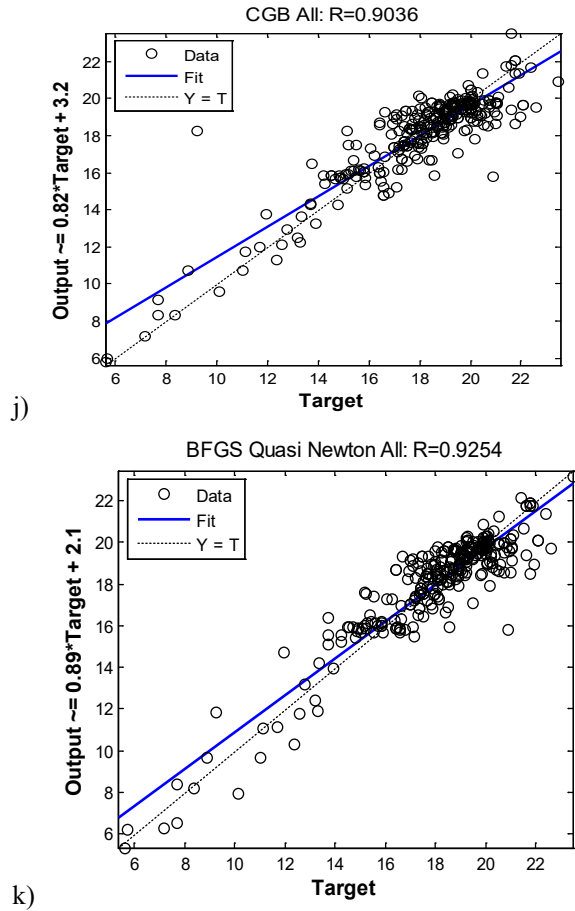
h)



e)



i)



**Fig. 4** The coefficient of determination of ANNs calculated for complete data sets (training, validation and testing) for: a) Scaled Conjugate Gradient, b) Bayesian Regularization, c) Resilient Backpropagation, d) One Step Secant, e) Variable Learning Rate Gradient Descent, f) Gradient Descent Momentum, g) Gradient Descent, h) Polak-Ribière Conjugate Gradient, i) Fletcher-Powell Conjugate Gradient, j) Conjugate Gradient with Powell/Beale Restarts and k) BFGS Quasi Newton training function.

Based on the results in Tables 6 and 7, Gradient Descent (GD) and Gradient Descent Momentum (DGM) training algorithms are disqualified from comparisons as unsuitable for training an ANN that solves this regression problem, because the training process did not converge to a useful ANN. In Tables 6 and 7 corresponding MSEs are too high, corresponding R values are complex and in Figure 4 e) and 4 f) the regression line deviates greatly from the ideal one.

The comparative analysis of different algorithms for training an ANN aimed for time series forecasting implemented in the MathWorks MATLAB neural network toolbox, performed in (Aggarwal et al. 2015) resulted with the conclusion that among all factors that affect the ANN's performance (amount of input data, the network complexity, activation function, number of hidden layers, number of weights and biases, etc), the training function is the most important factor for the accuracy of the network. Similarly, for predicting HHV of the biomass, all training functions exhibited reasonable accuracy, while the best performance has the one based on the Levenberg-Marquardt algorithm.

(Torrecilla et al. 2008) gave the comparative analysis of fourteen training algorithms with advantages and disadvantages while solving the prediction of the moisture value in a drying process of olive oil mill waste. The recommendations in previous works, related to the algorithm speed and the memory usage, favour LM training function for small datasets emphasizing that its performances are relatively poor on pattern recognition problems which can be solved faster by using Resilient Backpropagation (RP) algorithm.

In this work, in terms of MSE calculated on a set of training data, the Resilient Backpropagation (RP) training function ranked eighth out of twelve. For MSE calculated for sets of validation data RP ranked ninth and for MSE calculated for sets of testing data, RP ranked seventh. For R calculated on a set of training data, RP training function ranked eighth; for R calculated on a set of validation data, RP training function is ranked ninth and for R calculated on a set of testing data, RP training function is ranked seventh. This result is in accord with the fact that the determination of HHV based on the proximate analysis of biomass is a regression problem.

The training function recommended as similar to LM regarding the speed of the algorithm and less storage demanding, is BFGS Quasi Newton training function (BFG). It should be remarked, however, that its implementation gets slower with an increasing network complexity. In solving the problem of HHV prediction based on the proximate analysis of biomass, BFG performance was behind the performance of LM training function regarding the MSE calculated on a set of training data: the BFG training function ranked fourth, preceded by LM, CGB and BR training functions. For MSE calculated for sets of validation data BFG ranked right after the LM while BR was the best and for MSE calculated for sets of testing data, BFG was the fifth, preceded by BR, LM, SCG CGP training functions. For R calculated on a set of training data, BFG training function is right behind LM; for R calculated on a set of validation data, BFG training function is right before LM (it is slightly better than LM) and for R calculated on a set of testing data, BFG training function is slightly weaker than LM alongside with CGP training function.

Functions CGP, BFG, RP, OSS, GDX, GDM, GD, CGP, CGF and CGB are less common in literature on the prediction of HHV based on the proximate analysis of biomass. Probably this is at least partially due to the fact that they are not part of the built-in application with the graphical user interface in MATLAB application for training ANNs. In this paper it is emphasized that the network performances may differ depending on many factors and that it is advisable to explore the training functions alternatives to the three most commonly met in practical situations: LM, BR and SCG.

As for these three most commonly used training functions, LM, BR and SCG, general recommendations are that LM is the first choice for fast prediction of function fitting problems (nonlinear regression) on small datasets and that SCG is favourable in pattern recognition problems because SCG algorithm is the least memory demanding. Training with SCG automatically stops when generalization stops improving, as indicated by an

increase in the mean square error of the validation samples. BR algorithm is more time consuming, but can result in good generalization for difficult, small, or noisy datasets. Training stops according to adaptive weight minimization (regularization).

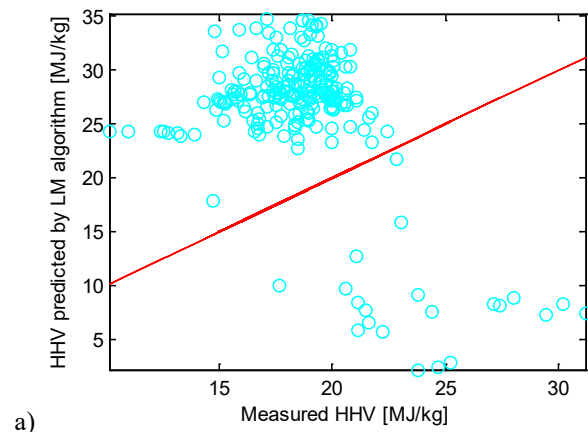
The quantitative comparative analysis of performances of the twelve training algorithms with all aspects identical except the training function, performed in this work, resulted in following conclusions: For the dataset used in the process of training all twelve networks, the best performances were shown by Bayesian Regularization training function in terms of R in general (calculated for the training, validation and testing data) and in terms of MSE calculated for the validation and testing data. LM training function had the smallest MSE regarding the training set of data.

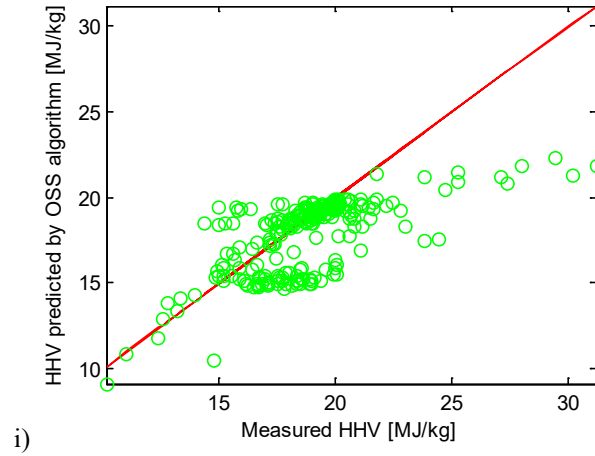
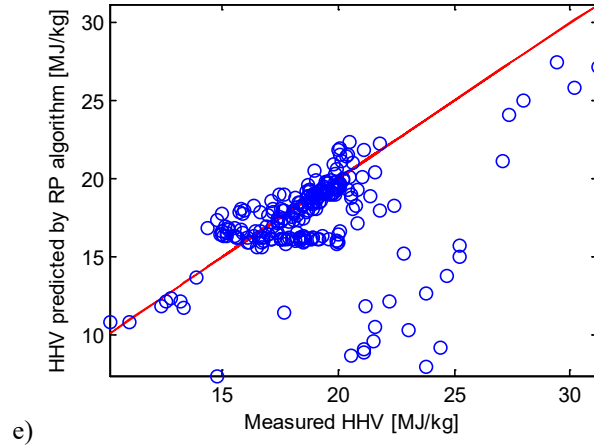
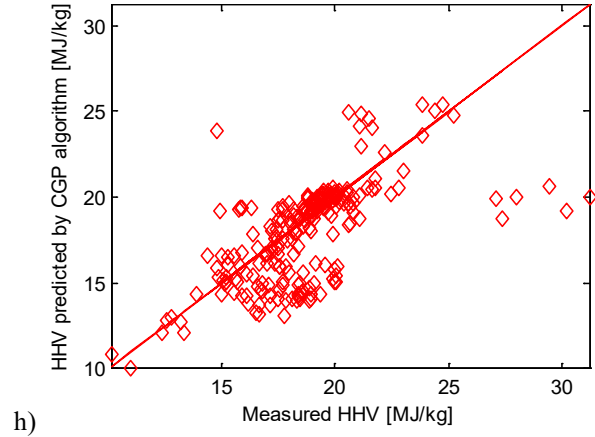
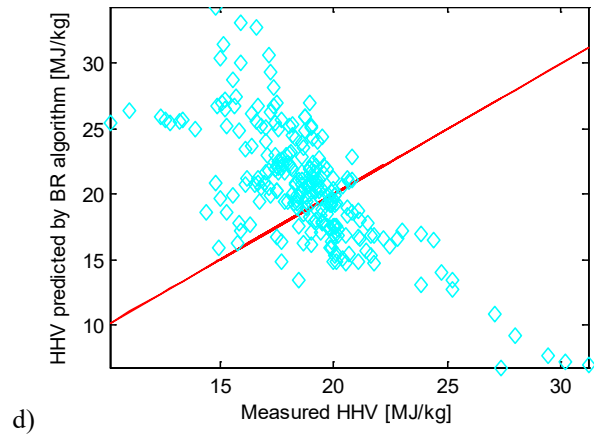
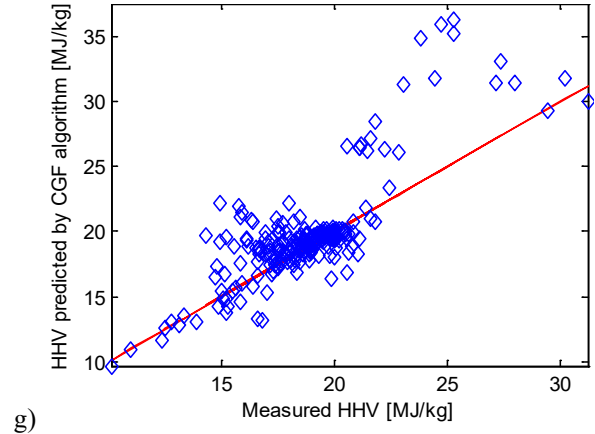
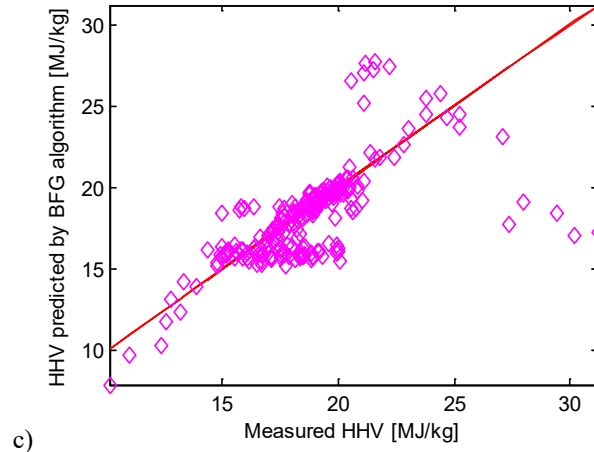
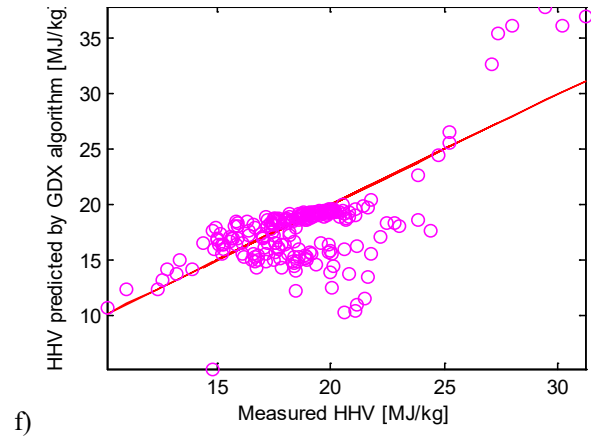
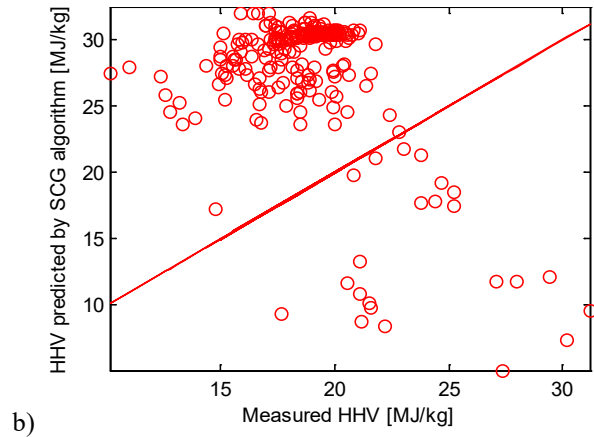
A very important figure of merit of any ANN is its ability to generalize the knowledge acquired in training process to new, completely unknown data. Since Gradient Descent and Gradient Descent Momentum algorithms were abandoned because they did not pass performance evaluation, ten ANN functions generated as output in the training process were fed with a completely new set of input data in order to estimate the ability of obtained networks to generalize the acquired knowledge. Complete set of new data, used for this estimation, consisted of 237 examples gathered from the literature on experimentally determined values for FC, VM, ASH and HHV. The literature that is used comprised five papers: (Cordero et al. 2001), (Alaba et al. 2020), (Conag et al. 2019), (Lakovic et al. 2021) and (Pattanayak et al. 2020), denoted with I, II, III, IV and V, respectively, in the Table 8 where the excerpt of the full data set is given. The complete dataset is available from (Jakšić 2021).

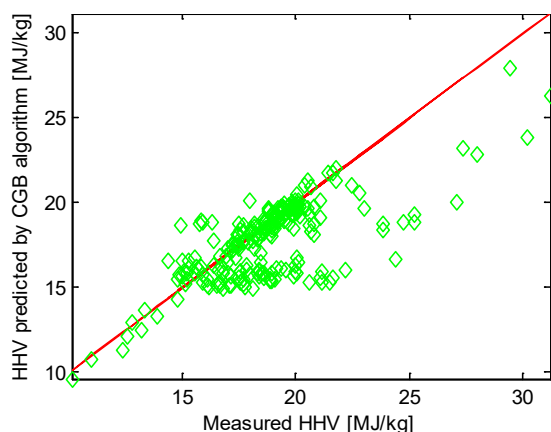
The results presented in Fig. 5 show the predicted HHV values obtained by ten ANNs (for 237 input triplets with FC, VM and ASH values) over the corresponding measured values of HHV related to input triplets. The solid red line represents the idealistic response of ANN when the predicted and measured values coincide. Symbols represent the predicted data. The quality of the prediction of ANNs was estimated by calculating the mean squared error as defined by Equation (1), the mean of squared differences between the predicted and the measured HHV values. Contrary to MSE ranking based on calculations over data used for training validation and testing of different algorithms, MSE ranking based on calculations over data completely new to ANNs, showed that the Levenberg-Marquardt training function and the Bayesian regularization training function were not superior to other training functions. The algorithm that generated data with the greatest correlation with the corresponding experimentally obtained values unknown to the ANN was the one based on the Conjugate Gradient, Powell/Beale Restarts training function (CGB). In terms of MSE calculated over the data unknown to ANNs, the ranking from the best to worst is as follows: CGB, OSS, CGP, CGF, GDX, RP, BR, BFG, SCG and LM. In this work CGB training function exhibited the best generalization properties – the smallest MSE of predicted HHV based on previously unseen triplets of input data.

Table 8 Excerpt of the dataset used as input for the estimation of the ANN ability to generalize the acquired knowledge when introduced to previously unknown data. The dataset is available from (Jakšić 2021) and I, II, III, IV and V are defined in the main text.

Biomass [Reference]	FC (%)	VM (%)	ASH(%)	HHV(M J/kg)
Sample 128 [IV]	36.1	60.5	3.30	21.4000
SCB-300-030 [III]	35.1 800	57.98 00	4.140 0	22.4200
Sample 129 [IV]	34.6 000	62.10 00	2.800 0	21.6000
SCL-300-030 [III]	31.9 500	51.23 00	14.78 00	20.8000
SCL-250-075 [III]	28.8 900	56.76 00	13.57 00	20.5700
SCB-350-015 [III]	31.6 900	62.11 00	4.710 0	20.6500
Sample 131 [IV]	33.9 000	54.70 00	0.400 0	21.7800
Bambusa-mizorameana [V]	5.19 00	65.45 00	20.06 00	18.4600
SCL-250-060 [III]	27.2 500	58.72 00	13.06 00	19.8700
SCB-250-075 [III]	31.3 000	63.22 00	3.670 0	20.3800
SCB-250-060 [III]	31.2 000	63.34 00	3.640 0	20.3900
Dendrocalamus-longispathus-Kurz [V]	4.35 00	66.85 00	18.80 00	17.8600
Sample 130 [IV]	33.9 000	65.70 00	0.400 0	21.7700
Dendrocalamus-asper [V]	5.19 00	62.65 00	15.16 00	18.4600
Bambusa-nutans [V]	4.85 00	67.95 00	16.20 00	18.3600
Bambusa-vulgaris-var.-vittata [V]	7.70 00	65.48 00	17.22 00	16.7100
Schizostachyum-polymorphum [V]	4.85 00	68.57 00	16.58 00	18.3600
Dendrocalamus-giganteus-Munro [V]	5.70 00	68.48 00	16.22 00	18.5100
Dendrocalamus-manipureanus [V]	6.40 00	67.00 00	17.60 00	19.3300
Lignite char [I]	72,6	13,3	14	27,371
Rice husk [II]	13,8	62,16	24,04	15,21







j)  
**Fig. 5** The ANNs output over the true measured value for the HHV for every training algorithm: a) Levenberg-Marquardt; b) Scaled Conjugate Gradient; c) BFGS Quasi Newton; d) Bayesian regularization; e) Resilient Backpropagation; f) Variable Learning Rate Gradient Descent; g) Fletcher-Powell Conjugate Gradient; h) Polak-Ribière Conjugate Gradient; i) One Step Secant; j) Conjugate Gradient, Powell/Beale Restarts

Graphical presentation of predicted HHV generated as outputs of ten selected ANNs showed that the cost function (the quality measure of ANN, here it is MSE) may not be informative enough when it comes to specifics related to data. Graphs in Figs. 5a) – 5j) show that the prediction of all ANNs was poor for input triplets whose measured HHV was high. The analysis of the datasets implies that the reason for such poor prediction is that ANNs were not taught to predict them. In the dataset that consisted the examples used for training, validation and testing of ANNs there were no example quadruplets (FC, VM, ASH, HHV) with HHV higher than 22.58 MJ/kg (soplillo biomass example in [12]). Another reason may be improper use of the obtained ANNs. All example quadruplets used for predictive HHV modelling of HHV for raw and torrefied sugarcane residues in [23] were used here for testing the generalization capability of observed ANNs without teaching them to treat such examples (some samples of torrefied biomass have higher HHV than raw biomass). The examples with high biomass HHV, used in [11] for investigation of predicting heating values of lignocellulosic and carbonaceous materials from proximate analysis, are used here for the estimation of the generalization capability of observed ANNs only and not in the process of training, validation and testing ANNs. Better performances of ANNs in the period of their usage and exploitation are the consequence of better training. The training is better if it is performed by using the examples of the data valid for the same type of the biomass or data obtained under the same laboratory conditions, etc.

## Conclusion

The outlined research gives the comparative qualitative and quantitative analysis of twelve algorithms for training artificial neural networks which predict the higher heating value (HHV) of the biomass characterized by the proximate analysis, i.e. by the percentages of the fixed carbon, volatile matter and ash. Besides the analysis of

performances of the investigated ANNs, the comparative qualitative and quantitative analysis of their usage, the prediction results of investigated ANNs when fed with the unknown data, are also given. All datasets with the examples, as well as the scripts used in this research for training of ANNs are published with open access, aimed for future training of new ANNs. ANNs in the form of functions that predict HHV of biomass based on the input triplets of data (percentages of the fixed carbon, volatile matter and ash in biomass) are also published with open access for all algorithms.

The algorithms investigated here were based on the training functions known as Levenberg-Marquardt, Scaled Conjugate Gradient, BFGS Quasi Newton, Bayesian regularization, Resilient Backpropagation, Variable Learning Rate Gradient Descent, Fletcher-Powell Conjugate Gradient, Polak-Ribière Conjugate Gradient, One Step Secant, Conjugate Gradient, Powell/Beale Restarts, Gradient Descent and Gradient Descent Momentum.

Besides of rankings of algorithms in terms of mean squared error and in terms of the coefficient of determination, the paper deals with the proper usage of the data used for the training, validation and testing of ANNs. The paradigm “the bigger the data set for the training, validation and testing of ANN, the better its prediction capabilities” is broadened here with the conclusion: a big dataset with the examples used for the training, validation and testing of ANN cannot guarantee its generalization if ANN’s usage goes beyond the dataset’s domain.

The presented results are a step toward our ultimate goal is to design a software tool for the calorific characterization of the biomass, easy to use and available for download from an open access data repository.

## Acknowledgement

This work was financially supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia (Grant No. 451-03-68/2020-14/200026).

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

- Abdulsalam J, Lawal A. I, Setsepu R. L, Onifade M, Bada S. (2020) Application of gene expression programming, artificial neural network and multilinear regression in predicting hydrochar physicochemical properties. *Bioresources and Bioprocessing* 7:1-22, <https://doi.org/10.1186/s40643-020-00350-6>.



- Aggarwal, R., & Kumar, R. (2015). Effect of training functions of artificial neural networks (ANN) on time series forecasting. *International Journal of Computer Applications*, 109(3), <https://doi.org/10.5120/19168-0634>.
- Alaba P. A, Popoola S. I, Abnisa F, Lee C. S, Ohunakin O. S, Adetiba E, Akanle M. B, Abdul Patah M. F, Atayero A. A. A, Wan Daud W. M. A. (2020) Thermal decomposition of rice husk: a comprehensive artificial intelligence predictive model. *Journal of Thermal Analysis and Calorimetry* 140:1811-1823, <https://doi.org/10.1007/s10973-019-08915-0>.
- Alagić S. Č, Maluckov B. S, Radojičić V. B. (2015) How can plants manage polycyclic aromatic hydrocarbons? May these effects represent a useful tool for an effective soil remediation? A review. *Clean Technologies and Environmental Policy* 17:597-614, <https://doi.org/10.1007/s10098-014-0840-6>.
- Aydinli B, Caglar A, Pekol S, Karaci A. (2017) The prediction of potential energy and matter production from biomass pyrolysis with artificial neural network. *Energy Exploration & Exploitation*. 35:698-712, <https://doi.org/10.1177/0144598717716282>.
- Boumanchar, I., Charafeddine, K., Chhiti, Y., Alaoui, F.E.M.H., Sahibed-Dine, A., Bentiss, F., Jama, C. and Bensitel, M. (2019) Biomass higher heating value prediction from ultimate analysis using multiple regression and genetic programming. *Biomass Conversion and Biorefinery*, 9(3):499-509, <https://doi.org/10.1007/s13399-019-00386-5>.
- Channiwal S, Parikh P. (2002) A unified correlation for estimating HHV of solid, liquid and gaseous fuels. *Fuel* 81:1051-1063 [https://doi.org/10.1016/S0016-2361\(01\)00131-4](https://doi.org/10.1016/S0016-2361(01)00131-4).
- Conag A. T, Villahermosa J. E. R, Cabatingan L. K, Go A. W. (2019) Predictive HHV Model for Raw and Torrefied Sugarcane Residues. *Waste and Biomass Valorization* 10:1929-1943, <https://doi.org/10.1007/s12649-018-0204-2>.
- Cordero T, Marquez F, Rodriguez-Mirasol J, Rodriguez J. (2001) Predicting heating values of lignocellulosics and carbonaceous materials from proximate analysis. *Fuel*. 80:1567-1571, [https://doi.org/10.1016/S0016-2361\(01\)00034-5](https://doi.org/10.1016/S0016-2361(01)00034-5).
- Dashti A, Noushabadi A. S, Raji M, Razmi A, Ceylan S, Mohammadi A. H. (2019) Estimation of biomass higher heating value (HHV) based on the proximate analysis: Smart modeling and correlation. *Fuel*. 257:115931, <https://doi.org/10.1016/j.fuel.2019.115931>.
- Demirbaş A, Demirbaş A. H. (2004) Estimating the calorific values of lignocellulosic fuels. *Energy exploration & exploitation*. 22:135-143, <https://doi.org/10.1260/0144598041475198>.
- Ghugare S, Tiwary S, Elangovan V, Tambe S. (2014) Prediction of higher heating value of solid biomass fuels using artificial intelligence formalisms. *BioEnergy Research*. 7:681-692, <https://doi.org/10.1007/s12155-013-9393-5>.
- Jakšić O. (2021) HHVannPRO: ANN trained to predict the biomass HHV based on the results of the proximate analysis Mendeley Data. <http://dx.doi.org/10.17632/dx8kbs5yxc.2>
- Klass D. L. (1998) Biomass for renewable energy, fuels, and chemicals; Academic Press: San Diego. <https://doi.org/10.1016/B978-0-12-410950-6.X5000-4>.
- Krishnamoorthy, M. and Raj, P.A.D.V., (2020) Optimum design and analysis of HRES for rural electrification: a case study of Korkadu district. *Soft Computing*, 24(17), p 13051-13068, <https://doi.org/10.1007/s00500-020-04724-y>.
- Krishnan R, Hauchhum L, Gupta R, Pattanayak S. (2018) Prediction of Equations for Higher Heating Values of Biomass Using Proximate and Ultimate Analysis. In *Proc. 2018 2nd International Conference on Power, Energy and Environment: Towards Smart Technology (ICEPE) Shillong, India*. p 1-5, <https://doi.org/10.1109/EPETSG.2018.8658984>.
- Lakovic N, Khan A, Petković B, Petkovic D, Kuzman B, Resic S, Jermisittiparsert K, Azam S. (2021) Management of higher heating value sensitivity of biomass by hybrid learning technique. *Biomass Conversion and Biorefinery* <https://doi.org/10.1007/s13399-020-01223-w>.
- Liang M, Lu W, Lei P, Wang L, Wang B, Li B, Shen Y, Zhang K. (2019) Physical and Combustion Properties of Binder-Assisted Hydrochar Pellets from Hydrothermal Carbonization of Tobacco Stem. *Waste and Biomass Valorization* 1-14, <https://doi.org/10.1007/s12649-019-00848-x>.
- Merckel R. D, Heydenrych M. (2017) Mass-fraction of oxygen as a predictor of HHV of gaseous, liquid and solid fuels. *Energy Procedia* 142:4124-4130 <https://doi.org/10.1016/j.egypro.2017.12.335>.
- Mijailovic I, Radojicic V, Ecim-Djuric O, Stefanovic G, Kulic G. (2014) Energy potential of tobacco stalks in briquettes and pellets production. *Journal of Environmental Protection and Ecology* 15:1034-1041.
- Nhuchhen D. R, Salam P. A. (2012) Estimation of higher heating value of biomass from proximate analysis: A new approach. *Fuel*. 99, 55-63, <https://doi.org/10.1016/j.fuel.2012.04.015>.
- Olatunji O, Akinlabi S, Madushele N. (2020) Application of Artificial Intelligence in the Prediction of Thermal Properties of Biomass; In *Valorization of Biomass to Value-Added Commodities; Daramola, M., Ayeni, A., Eds, Springer* p. 59-91, [https://doi.org/10.1007/978-3-030-38032-8\\_4](https://doi.org/10.1007/978-3-030-38032-8_4).
- Pattanayak S, Loha C, Hauchhum L, Sailo L. (2020) Application of MLP-ANN models for estimating the higher heating value of bamboo biomass. *Biomass Conversion and Biorefinery* <https://doi.org/10.1007/s13399-020-00685-2>.
- Priya A. D, Setty Y. P. (2019) Cashew apple juice as substrate for microbial fuel cell. *Fuel* 246:75-78 <https://doi.org/10.1016/j.fuel.2019.02.100>.



- Torrecilla, J. S., Aragón, J. M., & Palancar, M. C. (2008). Optimization of an artificial neural network by selecting the training function. Application to olive oil mills waste. *Industrial & engineering chemistry research*, 47(18), 7072-7080, <https://doi.org/10.1021/ie8001205>.
- Wang C, Li L, Chen R, Ma X, Lu M, Ma W, Peng H. (2019) Thermal conversion of tobacco stem into gaseous products. *Journal of Thermal Analysis and Calorimetry* 137:811-823, <https://doi.org/10.1007/s10973-019-08010-4>.
- Yu W, Chen C. (2014) Predicting the Heating Value of Rice Husk with Neural Network; In *Practical Applications of Intelligent Systems*; Springer: Berlin, Heidelberg. p. 877-885, [https://doi.org/10.1007/978-3-642-54927-4\\_84](https://doi.org/10.1007/978-3-642-54927-4_84).

# Figures

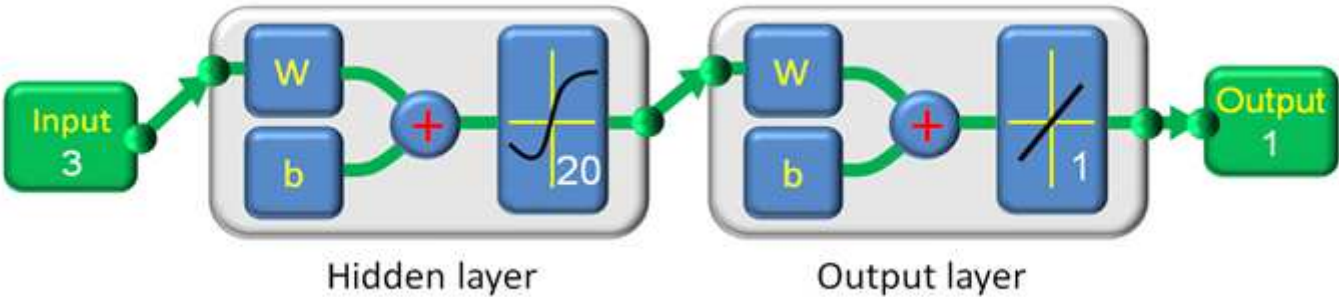


Figure 1

The simplified presentation of the ANN structure used in all our algorithms for the determination of the HHV based on the proximate analysis of the biomass.

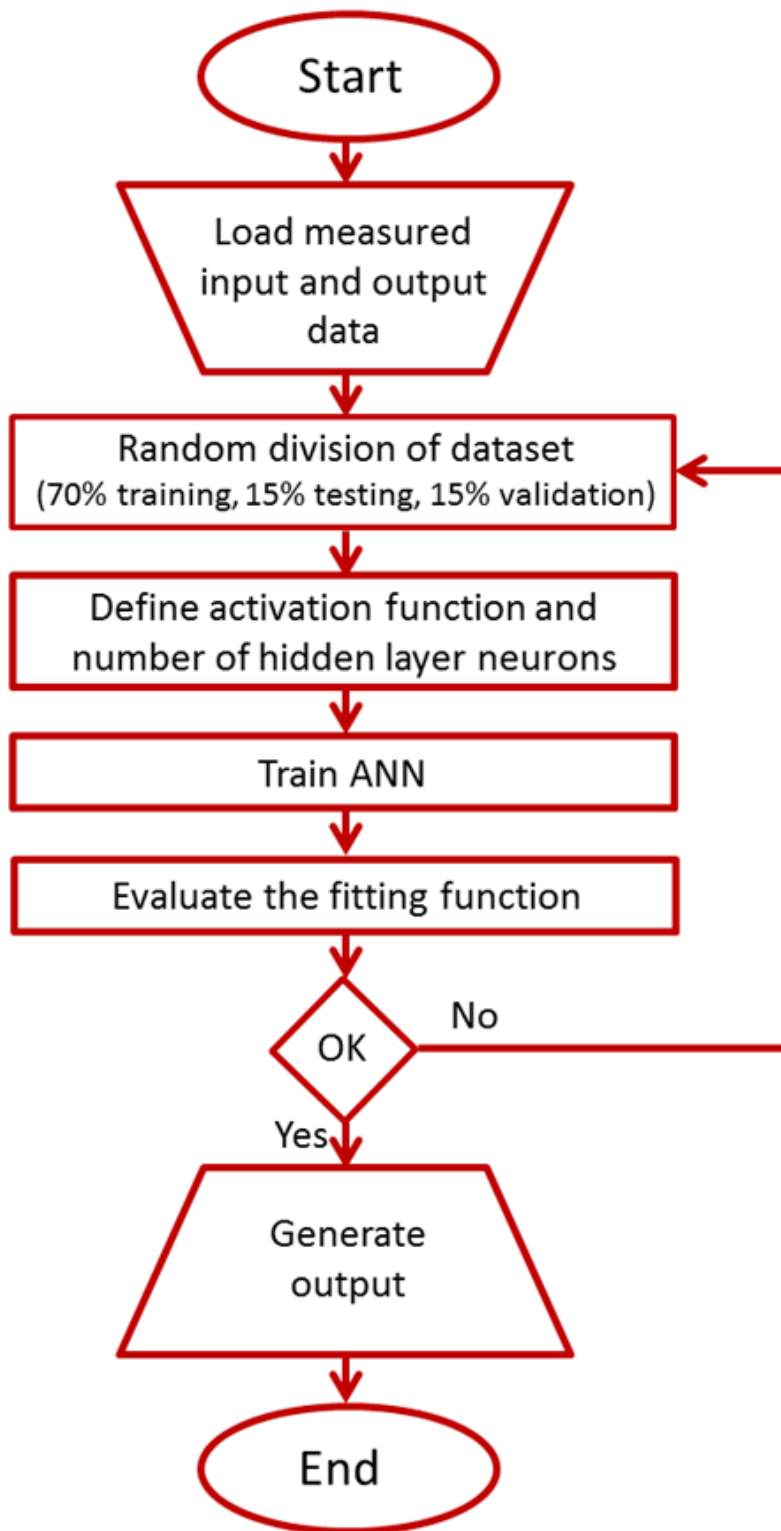
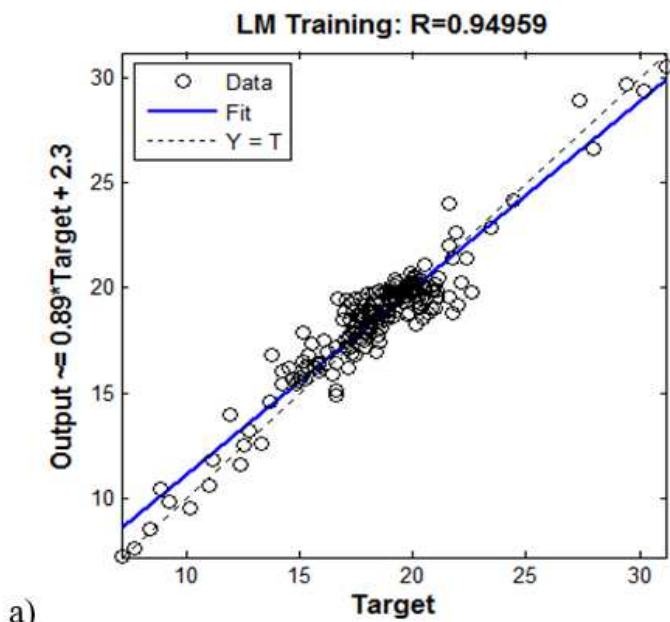
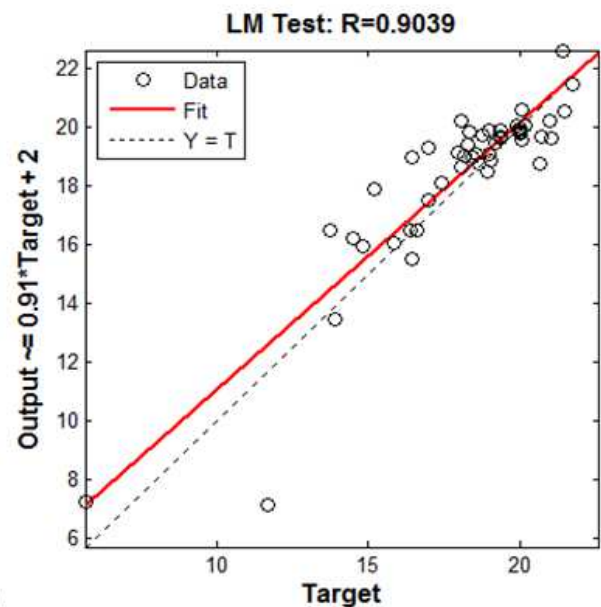


Figure 2

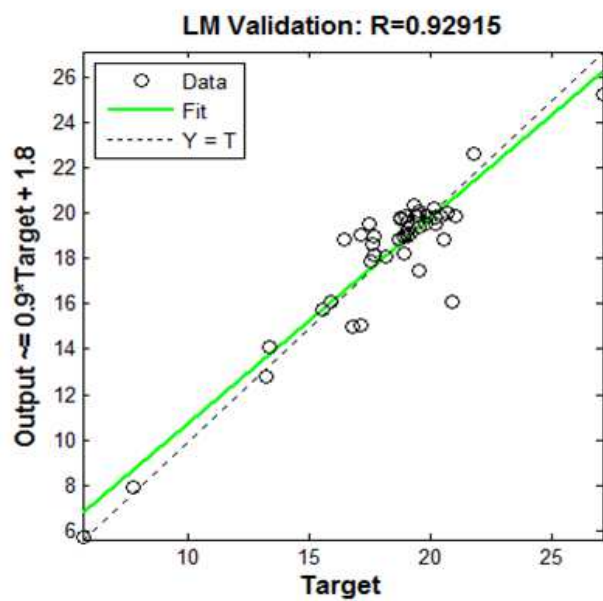
The workflow diagram for training our ANNs.



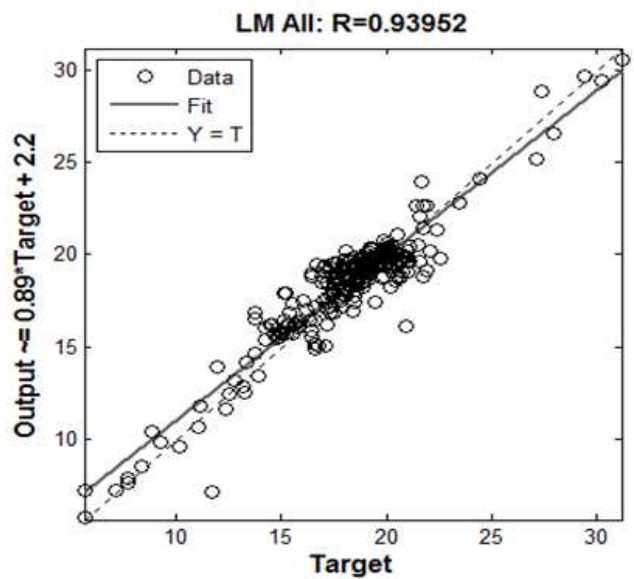
a)



c)



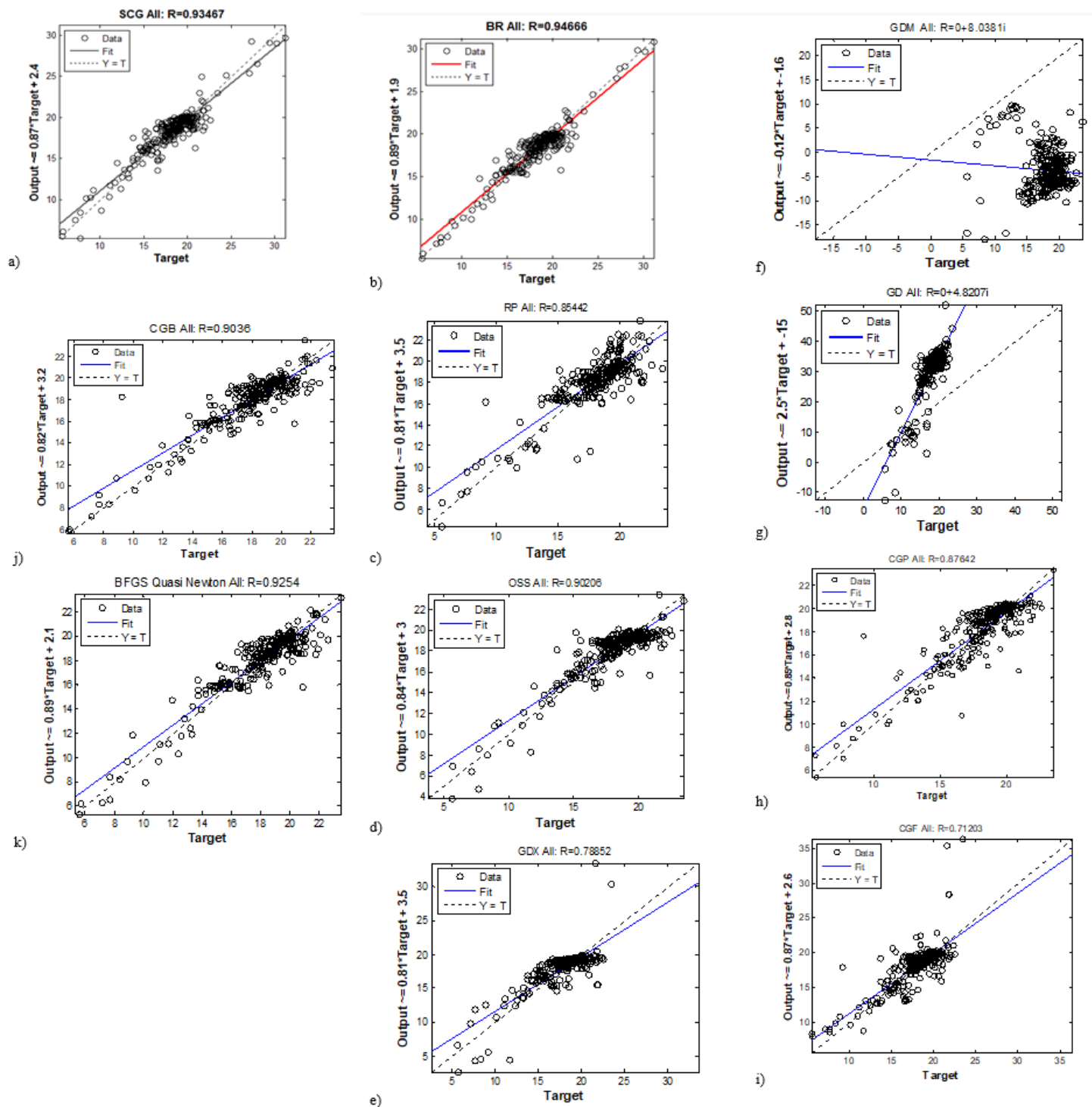
b)



d)

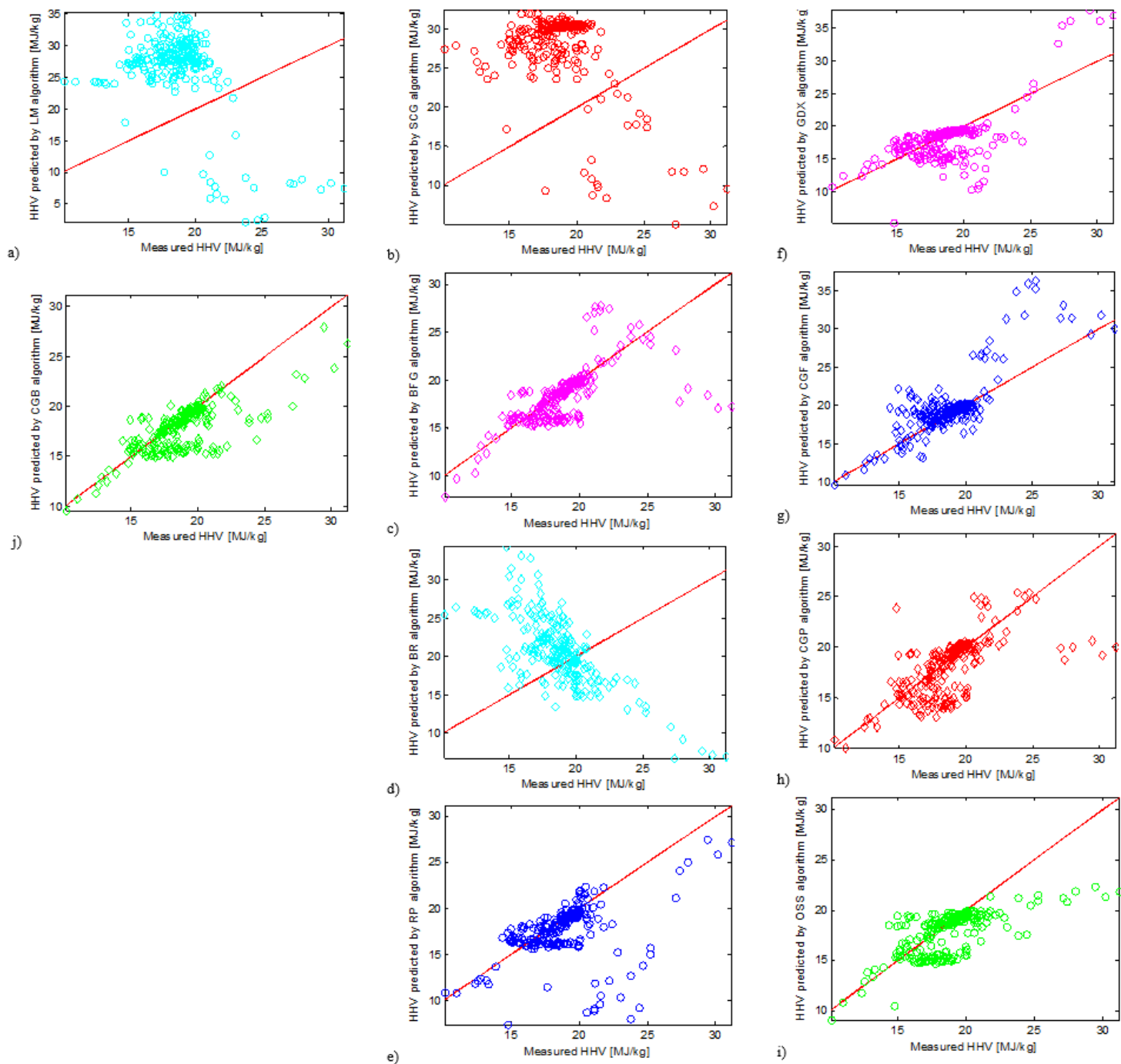
**Figure 3**

The regression of LM algorithm for different sets of data



**Figure 4**

The coefficient of determination of ANNs calculated for complete data sets (training, validation and testing) for: a) Scaled Conjugate Gradient, b) Bayesian Regularization, c) Resilient Backpropagation, d) One Step Secant, e) Variable Learning Rate Gradient Descent, f) Gradient Descent Momentum, g) Gradient Descent, h) Polak-Ribière Conjugate Gradient, i) Fletcher-Powell Conjugate Gradient, j) Conjugate Gradient with Powell/Beale Restarts and k) BFGS Quasi Newton training function.



**Figure 5**

The ANNs output over the true measured value for the HHV for every training algorithm: a) Levenberg-Marquardt; b) Scaled Conjugate Gradient; c) BFGS Quasi Newton; d) Bayesian regularization; e) Resilient Backpropagation; f) Variable Learning Rate Gradient Descent; g) Fletcher-Powell Conjugate Gradient; h) Polak-Ribière Conjugate Gradient; i) One Step Secant; j) Conjugate Gradient, Powell/Beale Restarts