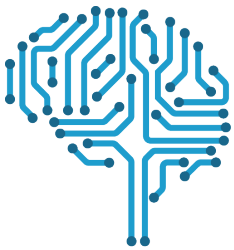

EasyANN

Predictions made easy

User Guide
v1.102

Gabriel Vilella[1], Pedro F. Arce[1], José O. Valderrama[2]

[1] Engineering School of Lorena, University of São Paulo, Brazil
[2] University of La Serena, Chile

[1]Department of Chemical Engineering, Engineering School of Lorena, University of São Paulo, Brazil.
[2]Department of Mechanical Engineering, University of La Serena, Chile.

Summary

1. Introduction 4

2. How to Use 5

2.1. ANN Topology 5

2.2. Training Function 6

2.3. Maximum Number of Iterations 7

2.4. Maximum Global Error 7

2.5. Number of Training and Predictions 7

2.6. Inserting Data 8

2.6.1. Read_Var_Training 9

2.6.2. Read_Depend_Var_Training 9

2.6.3. Read_Var_Testing 10

2.6.4. Read_Depend_Var_Testing 11

2.6.5. Read_Var_Prediction 11

2.7. Visualizing Results 12

2.7.1. Results_Training 12

2.7.2. Results_Testing 13

2.7.3. Data_Analysis 13

2.7.4. Results_Prediction 15

2.8. Prediction from a File 15

2.9. Before closing the application 16

3. Updating EasyANN 17

4. Uninstall 17

1. Introduction

EasyANN is a software aimed to the prediction of thermodynamic properties based on the training of artificial neural networks (ANN). The program reads data of a given desired property and data of the variables that have influence or determining the value of the desired property. For instance, it could read 100 critical pressure data (P_c) of ionic liquids and the same amount of data of the variables mass of the cation (M+), mass of the anion (M-), normal boiling temperature (T_b), critical temperature (T_c), and acentric factor (ω). The data is organized in an Excel file in different sheets (one for P_c and one for the variables M+, M-, T_b, T_c, and ω). The network finds a relation between the desired property P_c and the variables (M+, M-, T_b, T_c, and ω), by iteratively assigning to each neuron in the network a weight and a bias (adjustable parameters) until the calculated value of P_c is equal to the given value of the property (within a given margin of error). When this is reached, it is said that the network is trained.

After the network is trained, the program can predict the critical pressure P_c of other ionic liquids by knowing the properties (M+, M-, T_b, T_c, and ω). This manual explains how to use the program, for the estimation of P_c or any other property.

2. How to Use

The Figure 1 shows the user interface of the EasyANN software. On this screen, some variables and parameters of the network can be modified. Also, the user can choose between the optimization methods available for the process of finding the optimum weights and biases.

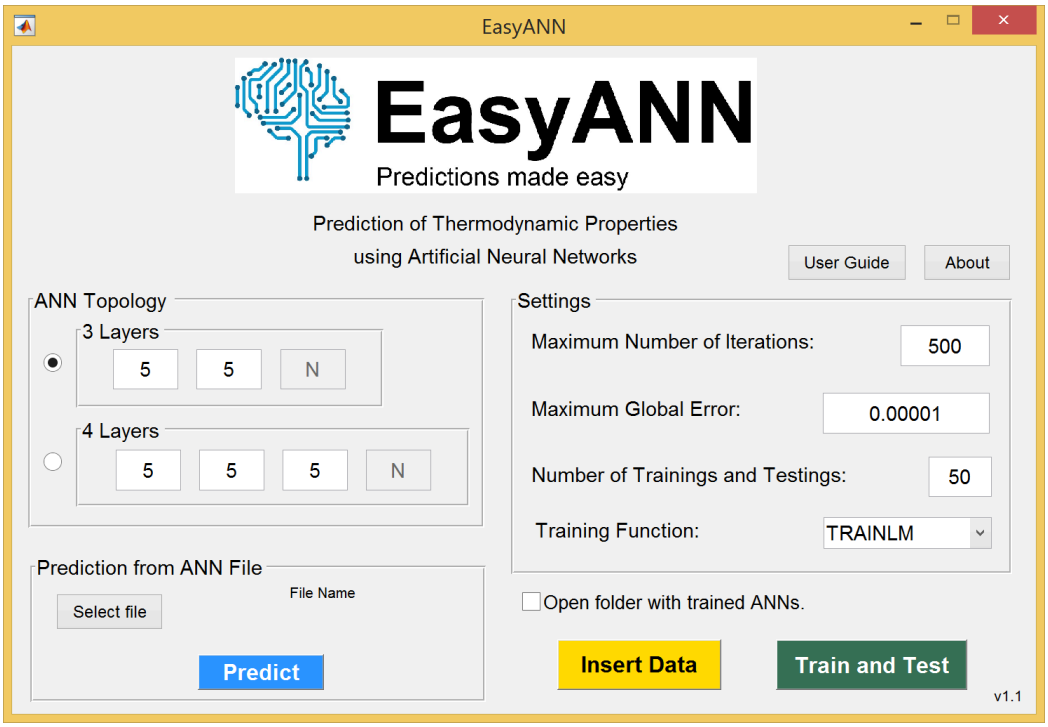

Figure 1 – EasyANN User interface.

2.1. ANN Topology

Here it is possible to set the topology of the ANN. You can choose between three or four layers and the number of neurons you want in each of the layers. The maximum is 50. The output layer that is represented by an N is the number of dependent variables and it will be filled automatically by the program based on the data it reads from Excel.

To know the best configuration for the set of data you have is to test various topologies and see which of them produces results with the minimum error.

2.2. Training Function

There are nine options available here for the training function. You just should select the one you want. The TRAILM that is marked by default is the one that is commonly used in fluid property estimation and provides good results. It is very difficult to know which training algorithm will be the fastest for a given problem and the ANN may not even converge any values with one function. The Table 1 shows all the available algorithms.

Table 1 – Training Function available for selection.

Acronym	Algorithm	Description
BFG	TRAINBFG	BFGS Quasi-Newton
CGB	TRAINCGB	Conjugate Gradient with Powell/Beale Restarts
CGF	TRAINCGF	Fletcher-Powell Conjugate Gradient
CGP	TRAINCGP	Polak-Ribiére Conjugate Gradient
GDX	TRAINGDX	Variable Learning Rate Backpropagation
LM	TRAINLM	Levenberg-Marquardt
OSS	TRAINOSS	One Step Secant
RP	TRAINRP	Resilient Backpropagation
SCG	TRAINSCG	Scaled Conjugate Gradient

2.3. Maximum Number of Iterations

Maximum Number of Iterations are also known as Epochs. This number represents the maximum number of times the testing will be repeated to get the results based on the Maximum Global Error. If the ANN cannot converge to any value, the maximum number will be reached.

2.4. Maximum Global Error

This is the error that one expects that the programs find between the value calculated in one iteration and the next one. The program will stop when this error is reached, or when the maximum numbers of iterations (500 by default), whichever comes first. This is not the error between the values of the training/testing dependent variables and the results being generated by the ANN. This is called deviation (% Dev) and can be seen on the Results_ tabs on the Excel file.

2.5. Number of Training and Predictions

Here you can specify how many times the software will repeat the training and prediction process. For each repetition, it calculates what was the minimum of the maximums errors among the substances during the training. Based on that, you know which iteration was the best. Each iteration also generates a file with all the prediction data. These files are unique because they represent the way the ANN learned and it can be used later to predict other set of data. It is stored on a folder that can be accessed if the checkbox *"Open folder with trained ANNs."* was checked.

Or, you can also navigate to:

C:\Users\[your_username]\Documents\Matlab\EasyANN\ANN_save

The maximum number you can enter for the repetition of the training and prediction process is 100.

ATTENTION: Do NOT modify or remove any of the files under the *\EasyANN* folder.

2.6. Inserting Data

Clicking on the *Insert Data* button the Excel file *EasyANN_Training_Prediction.xlsx* will open. There will be 9 sheets:

- Read_Var_Training;
- Read_Depend_Var_Training;
- Read_Var_Testing;
- Read_Depend_Var_Testing;
- Results_Training;
- Results_Testing;
- Data_Analysis;
- Read_Var_Prediction;
- Results_Prediction.

Do NOT change any of these names or the program will not run.

You can also access this file going to:

C:\Users\[your_username]\Documents\Matlab\EasyANN

ATTENTION: Do NOT modify or remove any of the files under the *\EasyANN* folder.

Check the EasyANN website to download an example of the Excel file, from where the screen shots here were taken.

2.6.1. Read_Var_Training

This is where you should insert the data relative to the substances you want to be used to trained the ANN. All the substances must be on the rows and the properties of each one of them must be on the columns. The Figure 2 shows an example.

	A	B	C	D	E	F	G
1	Cation	Anion	MW(+)	MW(-)	Tb (K)	Tc (K)	w
2	[DPEHIM	[PF6]	317.391	144.962	784.640	957.572	1.083
3	[omim]	[Cl]	195.330	35.452	649.550	869.410	0.657
4	[C2F3mim]	[TfO]	165.138	149.071	656.540	961.457	0.335
5	[omim]	[PF6]	195.330	144.962	646.100	810.848	0.938
6	[emim]	[TfO]	111.168	149.071	661.960	992.346	0.326

Figure 2 – Read_Var_Training example.

The program only reads numbers. All cells containing text are ignored and are used just for visual organization.

2.6.2. Read_Depend_Var_Training

Now you should insert the data relative to the property you are looking for. This is the property the ANN will learn to predict based on the other known properties. Figure 3 shows an example with one dependent variable (one desired property). You can have one to four dependent variables.

	A	B	C
1	Cation	Anion	Pc (bar)
2	[DPEHIM	[PF6]	11.65
3	[omim]	[Cl]	20.32
4	[C2F3mim]	[TfO]	28.53
5	[omim]	[PF6]	14.05
6	[emim]	[TfO]	35.84

Figure 3 – Read_Depend_Var_Training example.

It is important to notice that the order of the substances on the rows must be the same. And this must be maintained throughout the sheets. Also, the order of the columns (the properties) must be the same as well.

2.6.3. Read_Var_Testing

In this sheet the values of the independent variables (M+, M-, T_b, T_c, and ω) for those substances that you want to be used as tests to determine whether the ANN had a great training or not. Figure 4 shows an example. The five ionic liquids shown in the table were not used for training. You should note that order of the known properties on the columns is the same as it was on the first sheet: (M+, M-, T_b, T_c, and ω).

	A	B	C	D	E	F	G
1	Cation	Anion	MW(+)	MW(-)	Tb (K)	Tc (K)	w
2	[bmim]	[NfO]	139.222	299.092	762.29	1028.85	0.5150
3	[Me2NO2im	[NO3]	142.139	62.007	783.04	1084.70	0.7033
4	[BrC3mim]	[PF6]	204.091	144.962	597.86	782.99	0.7615
5	[Bemim]	[TfO]	173.239	149.071	803.04	1157.97	0.4118
6	[(C4H9OCH	[PF6]	241.356	144.962	713.82	886.61	0.9965

Figure 4 – Read_Var_Prediction example.

2.6.4. Read_Depend_Var_Testing

To know if the prediction was good or not, you should insert the data relative the substances you have chosen to be used to predict the desired property. The program will calculate the relative error for the training and prediction. Figure 5 shows an example.

	A	B	C
1	Cation	Anion	Pc (bar)
2	[bmim]	[NfO]	17.28
3	[Me2NO2im	[NO3]	34.20
4	[BrC3mim]	[PF6]	18.60
5	[Bemim]	[TfO]	29.01
6	[(C4H9OCH	[PF6]	13.26

Figure 5 – Read_Depend_Var_Prediction example.

2.6.5. Read_Var_Prediction

Here you are going to repeat the same process as described on the Read_Var_Training and Read_Var_Prediction. However, now you are going to insert

the parameters for the substances you want to predict a value. You should note that order of the known properties on the columns is the same as it was on the first sheet: (M+, M-, T_b, T_c, and ω). Figure 6 shows an example.

	A	B	C	D	E	F	G
1	Cation	Anion	MW(+)	MW(-)	Tb (K)	Tc (K)	w
2	[mmim]	[Br]	97.141	79.904	518.12	776.81	0.3797
3	[CH3CH(OH	[Cl]	183.275	35.452	695.53	914.61	0.9006
4	[emim]	[BF4]	111.168	86.803	449.46	596.23	0.8087
5	[DPEOIM	[BF4]	345.445	86.803	771.04	936.30	1.1675
6	[mim]	[TA]	83.114	113.016	511.6	716.73	0.5824

Figure 6 – Read_Var_Prediction example.

2.7. Visualizing Results

All the following sheets will be filled automatically by the program: Results_Training, Results_Testing, Data_Analysis and Results_Prediction. The only part that the user can insert content is the first two columns that are reserved to a text that describes the substances. The program will insert data starting on the C2 cell on every result sheet.

2.7.1. Results_Training

In this sheet, you can see the results and %deviations for each time the training was repeated and the error for each substance. Figure 7 shows an example for the first two training runs.

	A	B	C	D	E	F	G	H	I
1	Cation	Anion	Pc (bar)	R Calc_1	% Dev R_1	% Dev Abs R_1	R Calc_2	% Dev R_2	% Dev Abs R_2
2	[DPEHIM	[PF6]	11.65	11.8594	1.7823	1.7823	11.7448	0.7988	0.7988
3	[omim]	[Cl]	20.32	20.3813	0.3212	0.3212	20.7369	2.0714	2.0714
4	[C2F3mim	[TfO]	28.53	28.4490	-0.2674	0.2674	28.0977	-1.4990	1.4990
5	[omim]	[PF6]	14.05	14.2400	1.3529	1.3529	13.6842	-2.6034	2.6034
6	[emim]	[TfO]	35.84	36.0563	0.5925	0.5925	35.9007	0.1584	0.1584

Figure 7 – Results_Training example.

Column D shows the result for the first training for each substance. Column E is the relative error and the following column is the absolute value of the relative error. This will be used to determine which was the best iteration.

2.7.2. Results_Testing

Now you can check the results for the substances you have chosen to have the desired property predicted on the testing phase and see the errors for each substance. Figure 8 shows an example.

	A	B	C	D	E	F	G	H	I
1	cation	anion	Pc (bar)	R Calc_1	% Dev R_1	% Dev Abs R_1	R Calc_2	% Dev R_2	% Dev Abs R_2
2	[bmim]	[NfO]	17.28	17.4651	1.0713	1.0713	17.3191	0.2263	0.2263
3	[Me2NO2i	[NO3]	34.20	33.8308	-1.0797	1.0797	34.1058	-0.2754	0.2754
4	[BrC3mim]	[PF6]	18.60	18.6646	0.3475	0.3475	18.9918	2.1063	2.1063
5	[Bemim]	[TfO]	29.01	28.8140	-0.6757	0.6757	29.1195	0.3776	0.3776
6	[(C4H9OC	[PF6]	13.26	13.3871	0.9585	0.9585	13.3390	0.5960	0.5960

Figure 8 – Results_Training example.

2.7.3. Data_Analysis

The last sheet is the data analysis of all the results obtained by the ANN. Here you can see what were the errors for each run during training and prediction. Figure 9 shows an example for the first 10 runs.

	A	B	C	D	E	F	G	H	I	J
1		Training						Testing		
2	Iteration	Average	Average Abs	Max Abs	Epochs		Iteration	Average	Average Abs	Max Abs
3	1	21.125411	43.63174412	348.51051	3		1	19.532063	43.95386109	129.48075
4	2	0.1410348	1.076934081	8.4968953	55		2	0.3041335	1.34567532	5.5310944
5	3	0.5551234	13.43429084	61.512235	16		3	-5.5933255	10.32141217	34.682061
6	4	0.1285857	0.586533936	20.613864	33		4	-0.1926418	0.429759834	2.4373846
7	5	0.0897782	0.488489895	9.6351451	77		5	-0.0253073	0.910969031	2.7011518
8	6	0.1713677	0.498482846	5.2433029	72		6	0.2636652	0.651538101	2.0357627
9	7	2.4920803	10.03589152	67.71153	9		7	-2.2102575	7.751392755	17.620371
10	8	-0.0177256	0.698434601	14.171214	63		8	0.0403722	0.602554558	1.8667003
11	9	-0.0867388	1.252689851	12.008442	29		9	-0.1004539	1.300780368	3.9159301
12	10	0.6902134	7.527970533	70.653644	17		10	-1.8063288	6.124980229	14.676084

Figure 9 – Data_Analysis example of the errors.

Based on the maximum value of the absolute relative error (Max Abs) it is possible to determine the best iteration, which will be the one with the lowest value of Max Abs, that at the same time low absolute deviation. Figure 10 shows an example.

	L	M	N	O	P	Q	R	S	T	U
1	Values for the minimun Max Abs									
2	Training						Testing			
3	Iteration	Average	Average Abs	Max Abs	Epochs		Iteration	Average	Average Abs	Max Abs
4	6	0.17137	0.498482846	5.243303	72		8	0.04037	0.602554558	1.8667
5										
6										
7										
8										
9	Settings									
10	ANN Topology	5	15	15	1					
11	Max Global Error	0.00001								
12	Max Iterations (Epochs)	500								
13	Training Function	TRAINLM								
14										
15	Avg Epochs Needed	37.4								

Figure 10 – Data_Analysis example of the best iteration.

You can see that the best training results were in iteration 6 and for the prediction it was on iteration 8. The best iterations will only be automatically calculated only if you are working with one or two dependent variables. If you have three or four, you can use the Excel function that finds the minimum value. The program also writes the settings you have entered to run the ANN. Also, for four dependent variables, the results for each iteration will not be available due to the huge amount of data.

2.7.4. Results_Prediction

Here you can check the results for the substances you have chosen to have the desired property predicted. Figure 11 shows an example.

	A	B	C
1	Cation	Anion	Pc (bar)
2	[mmim]	[Br]	42.1225121
3	[CH3CH(OH)	[Cl]	24.6787101
4	[emim]	[BF4]	23.4065556
5	[DPEOIM	[BF4]	11.8796871
6	[mim]	[TA]	30.4355751

Figure 11 – Results_Prediction example.

2.8. Prediction from a File

After you have run EasyANN and determined a great trained ANN, you should save the file generated on that iteration (e.g. *Iteration3.eann*). The files with the trained ANN can be accessed if the checkbox *"Open folder with trained ANNs."* was checked or navigating to:

C:\Users\[your_username]\Documents\Matlab\EasyANN

ATTENTION: Do NOT modify or remove any of the files under the *\EasyANN* folder.

Then, you can load this file from anywhere on your computer and do as many predictions as you want. All the prediction results will have that same range of %deviation present on the testing phase. Before you click the Predict button, you should update the Read_Var_Prediction sheet on Excel with the substances you want to predict a value. You must have the same independent variables, on the same order, as you had when you trained the ANN, but now for the substances that you want to predict a value. The results will be written on the Results_Prediction sheet and all the other sheets will remain untouched.

2.9. Before closing the application

Now that you have run EasyANN and checked every result you need, you can copy all data you need from the Excel file to another one you have easy access. After that you should close the Excel file and run EasyANN again.

Remember: each time the program runs all the Results_ and Data Analysis sheets will be overwritten with the new content, as well as the Iteration files on the folder *ANN_save*. **It is recommended that you clear all contents of the Results_Training, Results_Testing and Data_Analysis to avoid confusion, especially if you change the Number of Trainings and Testings or the number of layers.** The Read_ sheets will remain unchanged, with the same data you have inserted before.

3. Updating EasyANN

Go to the EasyANN website to download the latest version. EasyANN checks every day you open it if there is an update available.

http://easyann.gvilella.me/download/

Download the Update Version and install it.

4. Uninstall

If you which to remove EasyANN from your computer, please go to Control Panel > Programs > Uninstall a program. Search for EasyANN and click on Uninstall.