

Research on Pattern Recognition Method for Honey Nectar Detection by Electronic Nose

Ningjing Liu¹, Bolin Shi^{2,*}, Lei Zhao², Zhaoshen Qing¹,
Baopin Ji¹, and Feng Zhou¹

¹ College of Food Science and Nutritional Engineering, China Agricultural University,
Beijing 100083, P.R. China

² Food and Agriculture Standardization Institute,
China National Institute of Standardization, Beijing 100191, P.R. China

Abstract. Electronic nose (e-nose) utilizes the gas sensors array to absorb the volatile organic compounds (VOCs) of samples to classify them into different clusters, and it is noted by the sensitive of the sensors. However, limited by the methodologies of the pattern recognition, this kind of advantage had not been exploited fully. The research studied on different types of pattern recognition method, and selected the optimum method for the detection of samples with little nuance, exemplified by honey nectar detection for including rape honey, linden honey and acacia honey. It was found that support vector machine (SVM, non-linear prediction model) showed better performance than linear discriminate analysis (LDA, linear prediction model) for the classification of tiny different samples, especially when multi-group detection was involved. After the optimized method was selected, the key points of the SVM model were analyzed, and two key parameters were displayed, which were kernel parameter and penalty parameter. Three algorithms, including grid searching (GS), particle swarm optimization (PSO) and genetic algorithm (GA) were applied to find the appropriate parameter values. The results showed parameters optimized by genetic algorithm (kernel parameter and penalty parameter is 0.11 and 14.38 respectively) led to the optimal model, whose training accuracy was 98.78% and prediction accuracy was 97.5%. The results suggested that in the method of SVM with parameters selected by GA, e-nose could handle well the discrimination of similar samples like honey nectar detection.

Keywords: electronic nose, nectar detection, pattern recognition, support vector machine.

1 Introduction

Electronic nose is a kind of a newly developing technology detection system. It utilizes gas sensors array, signal processing and pattern recognition to imitate the olfactory system of human beings. Gas sensors array, usually constituted by metal-oxide semiconductor sensors, absorbs the volatile organic compounds (VOCs) of samples, and the absorption reaction of the VOCs and sensors leads to the change of

* Corresponding author.

surface electron intensity [1]. This kind of current change will be transformed by signal processing to digital signals. After the digital signals obtained, they are analyzed by the pattern recognition system, and make the final discriminant results.

Unlike other detection methods, such as gas chromatography and high efficiency liquid chromatography, e-nose detection is based on the overall responses from all sensors, combining with the characterization from various angle, not only analyzing by one or two feature ingredients, nor one or two characteristic signals. In this case, focusing on one or two sensors results tends to be meaningless. So pattern recognition is brought into the system to deal with the overall detection [2]. Pattern recognition run by using training samples to cultivate the system and foster its ability of discriminate the special items like a person [3]. It has been applied into lots of research areas, like military, finance, medicine, industry and agriculture. Combining with different chemo metrics algorithms, it has shown the great potential in formulation optimization, production control and discriminate detection.

Pattern recognition plays an important role in the e-nose system, and affects the detection result directly, which means the performance of e-nose system largely depends on the method chosen for the pattern recognition. For the e-nose detection, the advantage of sensitive is presented when similarities are analyzed, which brings an advanced requirement to the pattern recognition [4]. However, some pattern recognition methods which were being used were not accurate enough for this level of sensitiveness. Linear discriminate models, like linear discriminate analysis (LDA), discriminate partial least-square (DPLS) etc., had been widely used in e-nose analysis, and turned to be more and more mature. But this kind of models might show weakness in nuance since in this case, it was not guaranteed that all the samples could be divided accurate by the linear classifier. Focusing on this, the research study the appliance of non-linear discriminate model, support vector machine (SVM). By mapping the signal data into higher dimension space, data points in the space tended to be more disperse, which made it easier for the model to classify the samples into different clusters. Besides, SVM is based on the principle of structural risk minimization, which will enhance the robust of model significantly [5]. However the complex of SVM not only brought the steady and robust of the model, but also the difficulty of parameter selecting. Under the different value of parameters, the model showed various degree of performance [6]. The study chose different chemo metrics algorithms to optimize the penalty parameter and kernel parameter of SVM so that the best classifier could be obtained.

2 Experiments and Methods

2.1 Experimental Samples

Three different kinds of botanical origin honey were chosen, including rape honey (76 units), linden honey (55 units) and acacia honey (113 units). To ensure the authenticity of the samples, all the honey were collected from beekeepers directly by the members of the group. Considering the florescence of different nectar honey was

not the same, samples collected in different time were stored in refrigerator in -18°C until the collection had been completed.

2.2 Instrument

The e-nose system was FOX 4000, made by Alpha MOS in France. It was constituted by 18 metal-oxide semiconductor gas sensors, distributing in 3 chambers. The e-nose was installed with HS100 head space auto-sampler, which contained 2 pallets with 64 head space bottles (10ml).

2.3 Methodology

To avoid the disturb of crystals which was formed during storage, water bath was employed to heat the honey before the test under the temperature of 40°C for 15 minutes [7]. The parameters of the detection were shown in Table 1, which had been optimized by the orthogonal test [8].

Table 1. Parameters of e-nose detection

Parameters	Value
Head Space Time(S)	300
Head Space temperature (°C)	40
Inject Volume (ul)	500
Inject Speed(μL/s)	500
Inject Temperature(°C)	45
Acquisition Time(S)	240
Delay Time(S)	1080

Each sample obtained a 18*120 matrix, 18 lines for 18 sensors, and 120 columns for 120s detection time. Independent Component Analysis (ICA) was used to extract the feature information in the data matrix, which generated a new matrix of 8*120 for each sample. Afterwards, genetic algorithm was utilized to select 20 characteristic points for each unit, which represented the characterization of samples. According to the ratio of 2:1, all the samples were divided into 2 parts, training data (164 samples) and prediction data (86 samples), to build and validate the model.

2.4 Chemometrics Algorithms

2.4.1 Prediction Model

Two different kinds of methods were selected to make a comparison, non-linear model and linear model. Non-linear model exemplified by support vector machine (SVM). SVM was after the principle of structural risk minimization, which was great helpful for the model to avoid over training of the model and especially suitable for the small margin data detection [9]. It utilized the non-linear kernel function to map the data into a higher dimension space, and built a classified plane to distinguish the

data sets into different clusters. To guarantee the minimum structural risk, the plane should be kept as far as possible from the data of both clusters. The linear model was built by the linear discriminate analysis (LDA), which made the classification through mapping the data to a specific direction where all the data could be separated from each other as disperse as possible [10]. Fisher function was normally applied to find this specific direction.

2.4.2 Parameter Optimization

The approaches for parameter optimization included grid searching, genetic algorithm (GA) and particle swarm optimization (PSO).

Grid searching was a kind of exhaustive searching method. By presetting the searching range and searching step length (usually the logarithm of 2), the algorithm could find the optimum solution as long as the range was suitable and step length was precise [11]. In this study, the search range was from 2-4 to 210, and the length of step was log2.

GA, simulating the process of natural evolutionary, could search multi-direction with retaining a population of candidate solution. The algorithm coded the candidates by binary system and run the iteration including choosing, crossing and variation until the optimal item obtained [12]. In this study, the accuracy of the model was set as the fitness function, and the number of population was 20 with the max iteration being 100.

PSO was similar to the GA, but it replaced the crossing and variation by calculating the difference between the fitness value of items and the fitness value of optimum in the population. Compared with GA, PSO was only affected by the optimal item, not the whole population, which would bring a faster convergence rate [13]. Same with GA, the number of population was 20 with the max iteration being 100.

3 Results and Discussion

3.1 Comparing Different Kinds Model

The research had comparing the performance of two different kinds of discriminate model, LDA and SVM, standing for linear model and non-linear model respectively. The accuracy of the models was shown in table 2.

Table 2. Accuracy of different discriminate models

Model Methods	Training Set				Prediction Set			
	Accuracy %	Clusters			Accuracy %	Clusters		
		Rape	Linden	Acacia		Rape	Linden	Acacia
LDA	89.6	45/51	27/38	75/75	83.8	17/25	12/17	38/38
SVM	95.7	50/51	35/38	72/75	92.5	24/25	14/17	36/38

From Table 2, it could be found that linear discriminate classifier showed excellent ability in acacia honey detection, the accuracy of training set and prediction set could reach 100% at the same time. However, this kind of performance was absent upon the detection of other two kinds of honey, especially for linden honey. This phenomenon illustrated that it was difficult for linear model to find out an appropriate classifier to discriminate the rape honey and linden honey, and these two kinds of honey confused badly in linear space. However, this matter did not occur in SVM model. Compared with LDA, SVM could distinguish three kinds honey well. The main reason of this was that SVM utilized the kernel function to map the data into higher dimension space, where the data sets got more disperse and it was great helpful for the classifier to find the difference among three clusters. Simultaneously, SVM was after the principle of structural risk minimization, which demanded the separating hyperplane to keep as far as possible from the data sets, and this ensured great probability of generalization, reflected in the higher value prediction accuracy.

For the further analysis, three bipartition model were built by these two methods, referring to Table 3 to Table 5.

Table 3. Discriminate model of Rape honey and Linden honey

Methods	Training Set			Prediction Set		
	Accuracy%	Clusters		Accuracy%	Clusters	
		Rape	Linden		Rape	Linden
LDA	97.6	51/51	36/38	78.6	17/25	16/17
SVM	92.1	50/51	32/38	86.0	21/25	15/17

Table 4. Discriminate model of Rape honey and Acacia honey

Methods	Training Set			Prediction Set		
	Accuracy%	Clusters		Accuracy%	Clusters	
		Rape	Acacia		Rape	Acacia
LDA	100.0	51/51	75/75	100.0	25/25	38/38
SVM	100.0	51/51	75/75	100.0	25/25	38/38

Table 5. Discriminate model of Linden honey and Acacia honey

Methods	Training Set			Prediction Set		
	Accuracy%	Clusters		Accuracy%	Clusters	
		Linden	Acacia		Linden	Acacia
LDA	98.2	38/38	73/75	98.2	17/17	37/38
SVM	100.0	38/38	725/75	100.0	17/17	38/38

The above three tables illustrated that compared with tripartition model, LDA tended to adapt to bipartition classification models, especially for the rape-acacia discrimination and linden-acacia discrimination. In these two models, there was no capability performance difference between LDA method and SVM method. Furthermore, in the case of rape-linden models, training accuracy of LDA was higher than that of SVM, despite of poor performance in prediction accuracy of LDA. However, neither LDA nor SVM could make efficient division of rape and linden

honey, which meant these two kinds of honey had similar properties, but the prediction model of SVM still kept in higher level than LDA. It was because that when dealing with the similar samples, SVM model did not pursue training accuracy blindly, but minimized the risk of mistaken as low as possible to maintain the generalization ability of the model [14].

Meanwhile, by comparing the tripartition model and bipartition models, it could be proved that SVM had great advantages on multi-classification. It was mainly because that SVM utilized the kernel function to map the data into higher dimension where the data got more separate than in the lower dimension, which was great helpful for generating the classifier. This kind of improvement showed significantly in multi-classification. However, as to the linear model, especially when the micro-difference samples were detected, it was hard for the model to search a well-fit classify plane, which led to the phenomenon that when this case occurred, neither training accuracy nor the prediction accuracy tended to be acceptable. The principle of SVM, structural risk minimization, allowed SVM model obtain the least diversity between the accuracy of training and the accuracy of prediction. However, LDA model, in some cases, demanded the high value of training accuracy too excessively to ignoring the generalized ability of the model, leading to the over training of the model.

The results above showed that when dealing with the similar samples of multi-clusters, SVM model showed the better performance, while as to the samples with large difference, and for the bipartition classifier, LDA model was also a good choice, considering its less complex model structure.

3.2 Effects of SVM Parameters

Although SVM model received better discriminate results, it was limited by the choice of different parameter values. Among all the parameters, the most three important factors were kernel function, kernel parameter value and penalty value. Since the task of kernel function was mapping the data into higher dimension space, as long as the function met the basic requirements of the kernel function, it did not make any significant difference [15]. While the other two factors, values of kernel parameter and penalty parameter were the key points of an excellent discriminate model. Figure 1 showed the accuracy of training sets under different parameter values by grid searching. It could be found that different value groups led to the different performance. Specifically, the figure did not illustrate an upward trend for the accuracy with the increase or decrease of the value, but tended to be a wave curve, which could be illuminated clearly by analysis the risk of SVM principle. Models with the kernel parameter (r) of 2^0 to 2^{-5} and penalty parameter (c) of 2^3 to 2^6 showed higher accuracy, and when $r=0.11$ and $c=16$, it reached the peak.

Where R was real risk, R_{emp} was empirical risk, n was the number of the samples, h was VC dimension, η was confidence level, and Φ was confidence interval.

In SVM, the higher dimension was the key point of detection, which was decided by the kernel parameter. In non-linear mapping, the kernel function was just like a kind of tool which could influence how the data mapping into the space, while the kernel parameter could confirm which space the data mapping into. In the high

dimension space, the dimension determined the Vapnik-Chervonenkis (VC) Dimension of the space, namely the capability of the plane classification [16]. The relationship between the real risk and VC dimension was shown in formulation 1 (h stands for the VC dimension).

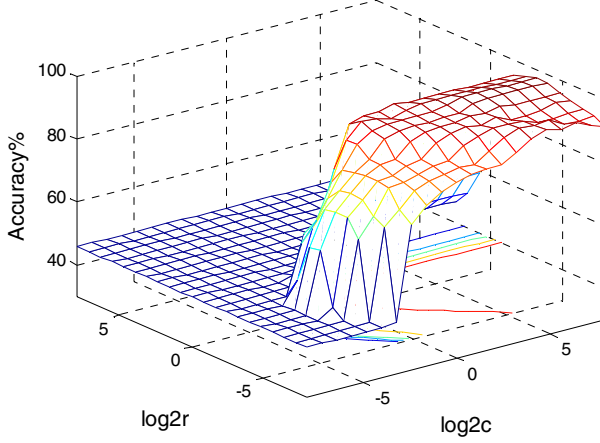


Fig. 1. Grid search graph for the optimization for the parameters of SVM

$$R(W) < R_{emp}(W) + \sqrt{\frac{h \left[\ln\left(\frac{2n}{h}\right) + 1 \right] - \ln\left(\frac{\eta}{4}\right)}{n}} \quad (1)$$

$$R(W) < R_{emp}(W) + \Phi \quad (2)$$

When the dimension got higher, the structure of the space tended to be complex, and the data sets became more diffuse, leading to the increase of VC dimension, enhancing the ability of the classifier, meaning the lower empirical risk. Along with that, confidence interval got wider, which brought larger disparity between the real risk and empirical risk. It could be shown by the phenomenon of higher training accuracy and lower prediction accuracy. In contrast, when the dimension got lower, the sets tended to be more concentrate. Although the classification would be complicate, the gap between the effectively risk and empirical risk would not be two wider, resulting in the little difference between training accuracy and prediction accuracy. For the detection requirements, the empirical risk and confidence interval needed to be minimized at the same time if it was possible, so that the final performance of the model could be acceptable.

Apart from the kernel parameter, penalty parameter also played a great importance role in model classification. Limited by the characteristics of sets, even in higher space, there still be the possibility of indivisibility. When this occurred, the overall classification should be considered in order the entire results obstructed. This could be obtained by setting the penalty parameters to ignore the individual samples to

maintain the total classification. By changing the weight of penalty, the demanding proportion of real risk and empirical risk could be confirmed. When the penalty got larger, the model pursued lower empirical risk, but brought wider confidence interval, while the penalty got smaller, it turned reverse. Hence the penalty parameter was also an important role.

The analysis above proved that great emphasis should be put upon these parameters selection.

3.3 SVM Parameters Selection

Besides the grid searching (in 3.2), the study utilized the genetic algorithm and particle swarm optimization to select the most suitable parameter value. The optimization results of these two algorithms were shown in figure 2 and figure 3.

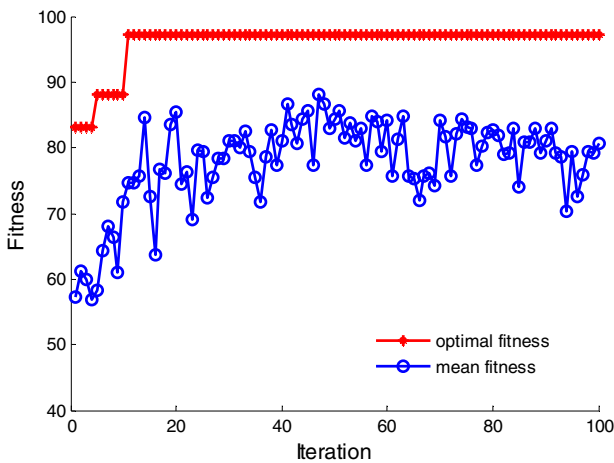


Fig. 2. PSO for the optimization for the parameters of SVM

Figure 2 illustrated the process of PSO. From the figure, it could be found that PSO got faster convergence rate. Fitness function became plane and stable in 18th generation. As to the mean fitness, the value increased wavy and in th 40th it reaches at approximate 80%. However, when it generated to 65th iteration, the mean value dropped a little, but it regained into 80% soon. It was because some bad items were generated but these items were weed out quickly to keep the main trend in normal direction. It proved that PSO had great ability of resistance for the accidental fault to avoid influence upon the final results. Finally, the optimization results was $r=20.02$, $c=0.09$.

Compared with PSO, the GA had a slower convergence rate. From the figure 3, it could be found that GA got stable in 44th generation. The difference in convergence rate was mainly led by the difference of algorithm theories. PSO put more energy on excellent individual, while GA needed an overall evaluation of all the items in one generation. Finally, when $c=14.38$, $r=0.11$ the model received the best results.

After that, parameters optimization results by three methods are shown in table 6.

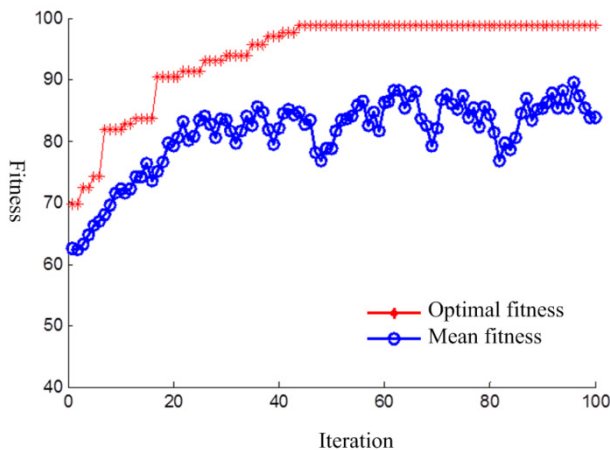


Fig. 3. GA for the optimization for the parameters of SVM

Table 6. Predict accuracy of SVMs optimized by different methods

Optimization Methods	c	R	Training Set				Prediction Set			
			Accuracy %	Clusters			Accuracy %	Clusters		
				Rape	Linden	Acacia		Rape	Linden	Acacia
Grid Searching	16.00	0.11	97.56	50/51	37/38	73/75	95%	24/25	16/17	36/38
PSO	20.02	0.09	96.95	50/51	36/38	73/75	93.75%	24/25	15/17	36/38
GA	14.38	0.11	98.78	51/51	37/38	74/75	97.5%	25/25	16/17	37/38

The results referring to Table 6 showed that there was little different between the kernel parameter selected by different methods, while the value of penalty parameter seemed more significant. It was mainly because the model largely relied on the structure of higher space, which was determined by the kernel parameter, so the optimization results of different methods got close to each other; while penalty only presented the tolerance of the error, not as critical as the kernel parameter, which led to fluctuate in different methods.

As to the capability of the model built by different parameter values, GA got the highest scores. Grid searching limited by the step length which was setting artificially, may not correspond with best searching length, being easily to jump the optimum if the length was too long; while for the PSO, considering the fast convergence rate and little improvement in the later stage, it could be doubted that the PSO had been fallen into local minimum, leading to the ignorance of globally optimal solution. If the algorithm fell into this trap, the process would be limited by searching the solution around local minimum repeatedly, and could not jump out of this [17]. In the study, the fast convergence rate and the poor optimization of PSO was identical to it. Ultimately, GA was applied to select the optimized values, and in the case of $r=0.11$ and $c=14.38$, the SVM got the best performance with the training accuracy of 98.78% and prediction accuracy of 97.5%.

4 Conclusions

The study focused on building a suitable prediction model for the e-nose. The results show that when dealing with similar samples like the honey of different nectar, non-linear model SVM had more energy to classify them into different cluster accurately, while linear model was relatively poor. Later the impact of kernel parameter and penalty parameter was discussed and utilized three different methods to select the best value, including grid searching, PSO and GA. Ultimately, GA was confirmed as the most suitable methods, bringing the kernel parameter (r) with the value of 0.11 and 14.38 for the penalty parameter (c), under which case, the performance of the model got the peak of 98.78% training accuracy and 97.5% prediction accuracy. The study provided a methodology for e-nose detection for similar samples, which could widely broaden the areas of e-nose application.

Acknowledgment. This study had been supported by National Natural Science Foundation of China (31101292).

References

1. Mirasoli, M., Gotti, R., Di Fusco, M., Leoni, A., Colliva, C., Roda, A.: Electronic nose and chiral-capillary electrophoresis in evaluation of the quality changes in commercial green tea leaves during a long-term storage. *Talanta* 129(1), 32–38 (2014)
2. Singh, H., Bhasker Raj, V., Kumar, J., Mittal, U., Mishra, M., Nimal, A.T., Sharma, M.U., Gupta, V.: Metal oxide SAW E-nose employing PCA and ANN for the identification of binary mixture of DMMP and methanol. *Sensors and Actuators B: Chemical* 200, 147–156 (2014)
3. Meyer-Baese, A., Schmid, V.: *Statistical and Syntactic Pattern Recognition*, pp. 151–196. Academic Press, Oxford (2014)
4. Rodriguez-Mendez, M.L., Apetrei, C., Gay, M., Medina-Plaza, C., de Saja, J.A., Vidal, S., Aagaard, O., Ugliano, M., Wirth, J., Cheynier, V.: Evaluation of oxygen exposure levels and polyphenolic content of red wines using an electronic panel formed by an electronic nose and an electronic tongue. *Food Chemistry* 155(15), 91–97 (2014)
5. Karaçalı, B., Ramanath, R., Snyder, W.E.: A comparative analysis of structural risk minimization by support vector machines and nearest neighbor rule. *Pattern Recognition Letters* 25(1), 63–71 (2004)
6. Lin, S.-W., Lee, Z.-J., Chen, S.-C., Tseng, T.-Y.: Parameter determination of support vector machine and feature selection using simulated annealing approach. *Applied Soft Computing* 8(4), 1505–1512 (2008)
7. Ajlouni, S., Sujirapinyokul, P.: Hydroxymethylfurfuraldehyde and amylase contents in Australian honey. *Food Chemistry* 119(3), 1000–1005 (2010)
8. Liu, N., Shi, B., Zhao, L., Qing, Z., Baopin, J., Zhou, F., Tian, W.: Optimizing an Electronic Nose for Analysis Honey from Different Nectar Sources. *Sensor Letters* 11, 1145–1148 (2013)
9. Ahmad, A.S., Hassan, M.Y., Abdullah, M.P., Rahman, H.A., Hussin, F., Abdullah, H., Saidur, R.: A review on applications of ANN and SVM for building electrical energy consumption forecasting. *Renewable and Sustainable Energy Reviews* 33, 102–109 (2014)

10. Sharma, A., Paliwal, K.K., Onwubolu, G.C.: Class-dependent PCA, MDC and LDA: A combined classifier for pattern classification. *Pattern Recognition* 39(7), 1215–1229 (2006)
11. Liu, R., Liu, E., Yang, J., Li, M., Wang, F.: Optimizing the Hyper-parameters for SVM by Combining Evolution Strategies with a Grid Search. *Ruiming* 344, 712–721 (2006)
12. Tam, S.M., Cheung, K.C.: Genetic algorithm based defect identification system. *Expert Systems with Applications* 18(1), 17–25 (2000)
13. Subasi, A.: Classification of EMG signals using PSO optimized SVM for diagnosis of neuromuscular disorders. *Computers in Biology and Medicine* 43(5), 576–586 (2013)
14. Corani, G., Gatto, M.: VC-dimension and structural risk minimization for the analysis of nonlinear ecological models. *Applied Mathematics and Computation* 176(1), 166–176 (2006)
15. Vapnik, V.: The Nature of Statistical Learning Theory. *Data Mining and Knowledge Discovery* 6, 1–47 (1999)
16. Koiran, P., Sontag, E.D.: Vapnik-Chervonenkis dimension of recurrent neural networks. *Discrete Applied Mathematics* 86(1), 63–79 (1998)
17. Nitta, T.: Local minima in hierarchical structures of complex-valued neural networks. *Neural Networks* 43, 1–7 (2013)