Neural network design and feature selection using principal component analysis and Taguchi method for identifying wood veneer defects

Yuce, B. , Mastronique, E. , Packianather, M. S. , Pham, D. , Lambiase, A. and Fruggiero, F.

CURVE is the Institutional Repository for Coventry University

http://curve.coventry.ac.uk/open

Neural network design and feature selection using principal component analysis and Taguchi method for identifying wood veneer defects

Baris Yuce, Ernesto Mastrocinque, Michael Sylvester Packianather, Duc Pham, Alfredo Lambiase & Fabio Fruggiero

Neural network design and feature selection using principal component analysis and Taguchi method for identifying wood veneer defects

Baris Yuce[a]*, Ernesto Mastrocinque[b], Michael Sylvester Packianather[c], Duc Pham[d], Alfredo Lambiase[b] and Fabio Fruggiero[e]

[a]Institute of Sustainable Engineering, Cardiff University, Newport Road, The Parade, Queen's Building, Cardiff CF24 3AA, UK; [b]Department of Industrial Engineering, University of Salerno, Via Ponte don Melillo 1, Fisciano 80046, Italy; [c]Institute of Mechanical and Manufacturing Engineering, Cardiff University, Queen's Buildings, The Parade, Cardiff CF24 3AA, UK; [d]School of Mechanical Engineering, University of Birmingham, Birmingham B15 2TT, UK; [e]Department of Engineering & Environmental Physics, University of Basilicata, Potenza, Italy

Nowadays, ensuring high quality can be considered the main strength for a company's success. Especially, in a period of economic recession, quality control is crucial from the operational and strategic point of view. There are different quality control methods and it has been proven that on the whole companies using a continuous improvement approach, eliminating waste and maximizing productive flow, are more efficient and produce more with lower costs. This paper presents a method to optimize the quality control stage for a wood manufacturing firm. The method is based on the employment of the principal component analysis in order to reduce the number of critical variables to be given as input for an artificial neural network (ANN) to identify wood veneer defects. The proposed method allows the ANN classifier to identify defects in real time and increase the response speed during the quality control stage so that veneers with defects do not pass through the whole production cycle but are rejected at the beginning.

Keywords: ANN; PCA; feature extraction; Taguchi analysis

1. Introduction

In recent years, quality concept has become very crucial not just for the products themselves, but as a competitiveness factor for the companies. Nowadays, the concept of quality for companies is synonymous with efficiency. The companies which reach high levels of quality are more efficient because they produce better products with lower costs. Companies whose organization or processes do not exhibit acceptable levels of quality are characterized by a number of internal errors such as dead times, poor coordination for numerous activities which are overlapping and disjointed, waste of resources, and a lack of tools and procedures for collecting feedback to a process of continuous improvement.

Quality control may generally be defined as a system that maintains a desired level of quality, through feedback on product/service characteristics and implementation of

---

remedial actions in case of a deviation of such characteristics from a specified standard (Mitra, 2012). Several tools stemming from statistics, computer science, and other similar fields are used to perform and improve the process of quality control. For instance, artificial neural networks (ANNs) (McCulloch & Pitts, 1943) have been used in many real-world applications related to quality control and in particular an ANN classifier has been proven to give the best results to correctly recognize wood veneer defects (D'Addona & Teti, 2013; Pham & Liu, 1995). The ANN takes as input 17 statistical features extracted from wood veneer images captured using a charge-coupled device (CCD) matrix camera. These images were converted into gray-level histograms after applying segmentation and image-processing algorithms in order to extract the features. Using these features an ANN could be trained to distinguish between 12 veneer defects together with clear wood giving 13 classes. During training the ANN takes the 17 features as input and assigns it to one of the 13 classes, and during recall it receives the 17 features and indicates the class to which it belongs to.

In order to improve the quality control and reduce the processing time, it is necessary to identify the critical variables and eliminate the redundant and the noisy features. A useful tool for this is the principal component analysis (PCA) (Pearson, 1901); this multivariate statistical technique, employed in many fields uses orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components (PCs). In this way, it is possible to reduce an original space R^p by representing the variables in a new space R^k with $k < p$.

The aim of this paper is to improve the quality control for the identification of wood veneers defects through the integration of the PCA and ANN. The procedure consists of identifying the PCs in the 17 features in order to reduce the number of inputs to be given to the ANN necessary to detect the defects in real time with minimal error, whereby reducing the whole quality control process time.

This paper is organized as follows: Section 2 surveys the literature on the identification of wood veneer defects; the experimental procedure is given in Section 3; the PCA pre-processing is given in Section 4; neural network design is given in Section 5; the results are given in Section 6; and finally conclusions and future research directions are given in Section 7.

2.   Literature review

This section gives a survey of the papers regarding the identification of wood veneer defects, highlighting the pros and cons of the methods used.

In order to detect wood veneer defects, an automatic visual inspection system was developed by Pham and Alcock (1996, 1999a) to segment the images of birch wood boards. Monochrome images of the veneer are pre-processed by automated algorithms that locate defect areas (Pham & Alcock, 1999b) where a set of numerical descriptors is extracted for further analysis. Seventeen statistical attributes of the local gray-level distribution were identified as relevant for defect identification (Lappalainen, Alcock, & Wani, 1994; Pham & Alcock, 1999c). Twelve possible veneer defects were distinguished in contrast to clear wood giving 13 classes. For each data sample, a classifier takes the 17-dimensional vector of image feature and decides to which of the 13 classes the pattern belongs to.

Among several algorithms, ANN has given the best results regarding the ability to recognize wood veneers defects. A three-layered multi-layer perceptron (MLP) gave 85% identification rates (Packianather, 1997; Packianather, Drake, & Rowlands, 2000;

Pham & Liu, 1995). Learning vector quantization networks have been applied to perform the classification of wood defects with high accuracy (Pham & Sagiroglu, 2000). In order to improve the classification accuracy of a single network, a decision tree of smaller and more specialized modular neural networks were introduced to achieve classification by successive refinements (Packianather & Drake, 2000).

In Packianather and Drake (2004), Response surface methodology was used to design a MLP network for classifying surface defects on wood veneer. The results showed that although the performance of the neural network could be improved by this method extrapolation outside the tested parameter range should be avoided.

A comparison between the minimum distance classifier (MDC) and neural networks to identify wood veneer defects was performed showing that the MDC does not perform as well as a neural network (Packianather & Drake, 2005).

Further, the Bees Algorithm (BA) was employed in place of the Back Propagation (BP) algorithm to optimize the weights of neural network for identification of wood defects (Pham, Ghanbarzadeh, Koc, Otri, & Packianather, 2006). Both the algorithms showed the same accuracy and in addition the BA proved to be considerably faster.

Lastly, the evolutionary ANN Generation and Training algorithm was used in the design and training of MLP classifier for identification of wood veneer defects (Castellani & Rowlands, 2009). The algorithm enabled the neural network topology and the weights to evolve over time. Compared to the approach based on the Taguchi method for the manual optimization of the MLP structure and the control parameters of the BP rule for tuning the connection weights, this algorithm performed equally well to the ANN solutions but using considerably smaller NN architectures. This paper expands previous works on the classification of wood veneer defects using statistical features extracted from wood veneer defect images; in particular it proposes a method to optimize the identification processes using the integration of ANN and PCA. The PCA method has been widely used as a means of reducing the dimensionality of the input space (Ahmadzadeh & Lundberg, 2013; Charytoniuk & Chen, 2000; Mohamed-Saleh & Hoyle, 2008; Rajput, Das, Mishra, Singh, & Dwivedi, 2010; Sratthaphut, Jamrus Woothianusorn, & Toyama, 2013; Tabe, Simons, Savery, West, & Williams, 1999). In this paper, PCA is used to reduce the number of input features for the wood veneer defects identification problem.

2.1. Case study: wood veneer defects identification problem

Plywood is made of thin layers of wood, called veneers, joined together using an adhesive. The quality of a board is determined by the types and number of defects in the constituent sheets. In any case, high-quality boards should be made up of sheets containing as few defects as possible or preferably no defects. To ensure this, careful inspection of the sheets is required as part of the quality control process. Defects of the veneer are identified by human inspectors as the sheets are transported to an assembly on a conveyor. This task is extremely stressful and demanding and a short disturbance or loss of attention results in misclassifications. A study conducted on human inspectors on wood mills reported that they could only obtain up to 55% accuracy in wood sheet inspection (Pölzleitner & Schwingshakl, 1992). Hence, an automatic visual inspection system was developed in order to increase the accuracy in wood sheet inspection (Drake & Packianather, 1998; Pham & Alcock, 1996, 1998, 1999c). Wood sheets were presented to a CCD camera which captured their images. These were segmented to separate clear wood and defective areas. Features were then extracted from the segmented images. The

feature vectors obtained were finally presented to the defect classification module that performed the task of grouping them into one of 12 types. Seventeen statistical features (Table 1) of the local gray-level distribution were identified as relevant for defect identification (Lappalainen et al., 1994; Pham & Alcock 1999c). Twelve possible defects of the

Table 1.   Features extracted.

| Feature number	Description of feature
1	Mean grey level
2	Median grey level
3	Mode grey level
4	Standard deviation of the grey levels σ
5	Skewness
6	Kurtosis
7	Number of pixels with a grey level of less than or equal to 80
8	Number of pixels with a grey level of greater than or equal to 220
9	Grey level p for which there are 20 pixels below
10	Grey level s for which there are 20 pixels above
11	Histogram tail length on the dark side $(q - p)$
12	Histogram tail length on the bright side $(s - r)$
13	Number of edge pixels after thresholding a segmented window at mean value
14	Number of pixels after thresholding at $\mu - 2\sigma$
15	Calculate the number of edge pixels for feature 14
16	Number of pixels after thresholding at $\mu + 2\sigma$
17	Calculate the number of edge pixels for feature 16
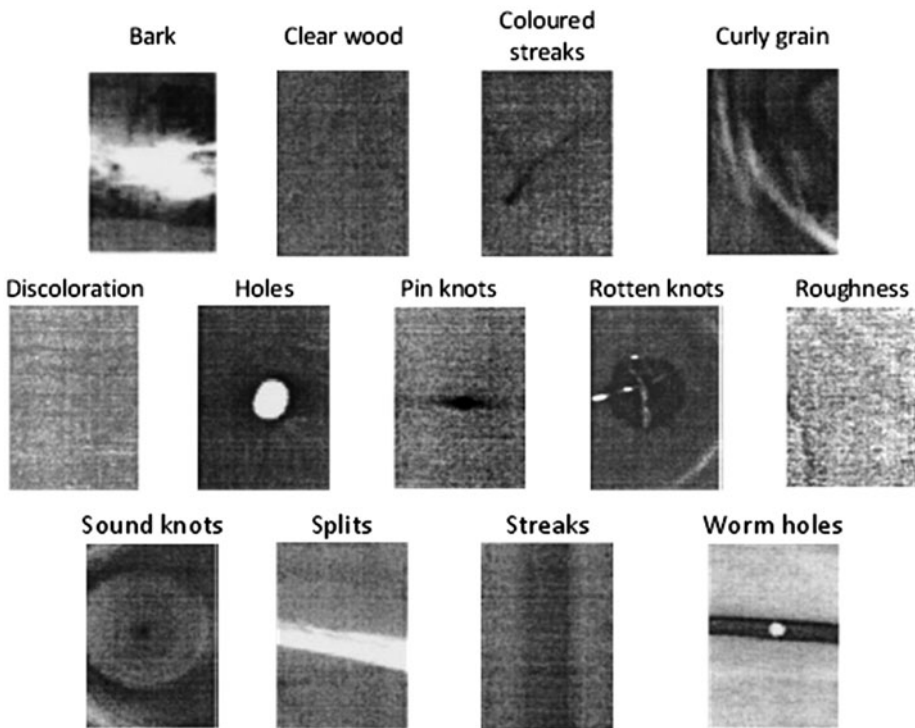
Figure 1.   Wood veneer defects classes and clear wood.

veneer can be distinguished in contrast to clear wood giving 13 possible classes shown in Figure 1. For each data sample, a classifier takes the 17-dimensional vector of image features and decides to which of the 13 classes the pattern belongs. The experimental procedure for training the ANN classifier is described in the following section.

3. Experimental procedure

For each defect shown in Figure 1, a 20×17 matrix was available except for the defects of curly grain, holes, and worm holes where their matrix dimensions were lower. In order to avoid any imbalance in the training data these three defects were excluded in this study. The number of rows in the matrix dimension indicates the number of exemplars available for each class and the number of columns indicates the total number of features extracted as given in Table 1. An initial examination of features 7 and 8 showed that a considerable number of feature values were zero and for this reason they were excluded from this study.

The PCA-based preprocessing was then applied on 15 features in order to find the principal variables to be given as input for the ANN. A feed-forward ANN with a Back-Propagation learning algorithm has been chosen using 75% of the data-set for training and 25% for testing. The results were compared with those obtained using all the features as input for the ANN.

Subsequently, several experiments have been conducted in order to find the best ANN configuration for the identification of wood veneer defects problem. The experiments have been carried out using different numbers of hidden layers, different numbers of neurons in the hidden layers, different values of the Pearson coefficient, and different numbers of input for the network. Finally, Taguchi analysis (Roy, 1990) has been performed in order to analyze the results and identify the best configuration of the ANN.

4. The PCA based feature selection

The PCA is a multivariate statistical tool that uses orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called PCs. The procedure starts calculating the covariance matrix of the variables if they were expressed in the same unit. In our case the original variables show different units and orders of magnitude. For this reason, the original variables had to be expressed in terms of standardized deviations using the correlation matrix instead of the covariance one.

Each PC is expressed as linear combination of the standardized deviations of the p variables. The first PC is:

$$y_1 = Za_1$$

The score for the i-th statistical unit is:

$$y_{i1} = a_{11}z_{i1} + \cdots + a_{1s}z_{is} + \cdots + a_{1p}z_{1p} \quad \text{for } i = 1, \ldots, n$$

where a_{1s} is the coefficient of the first PC and s-th variable. The sign of this coefficient reveals the relationship between the first PC and the s-th variable and its value shows how much this variable contributes to the scores of the first PC. In general, considering the first k PCs, the $n \times k$ score matrix is:

$$Y = ZA$$

Using the normalization condition,

$$a_v{'}a_v = 1$$

The scores become standardized with mean equal to zero and variance equal to the first eigenvalue. In fact, considering the first PC:

$$Var(Y_1) = \frac{y_1{'}y_1}{n} = a_1{'}\frac{ZZ'}{n}a_1 = a_1{'}Ra_1 = a_1{'}\lambda_1 a_1 = \lambda_1 a_1{'}a_1 = \lambda_1$$

Alternatively, it is possible to obtain the scores of the PCs with mean equal to zero but variance equal to one, dividing each score by the root of the respective eigenvalue.

Another way is by considering a_{vs} satisfying the following expression:

$$a_{vs}{'}a_{vs} = \lambda_v$$

The a_{vs} coefficients are the correlation coefficients between components of each of the p variables:

$$r(Y_v, X_s) = a_{vs} = a_{vs}\sqrt{\lambda_v}$$

4.1. Number of PCs

There are several criteria to choose the right number k of the PCs. Considering a PCA using the correlation matrix, the first criterion suggests choosing a number of PCs which contain a high percentage (at least 80%) of the total variance. This criterion can be modified considering a percentage threshold changing as a function of the starting number of variables.

The second criterion suggests considering all the PCs having eigenvalue higher than one. In this way, each PC explains a percentage of the total variance higher than that of a single variable.

The third criterion suggests making a graphic called scree plot of the eigenvalues λ_v as a function of the number v of PCs ($v = 1, 2 \dots p$). Since the eigenvalues are obtained in a decreasing order, this plot is descending; the criterion suggests choosing k as the number of PCs if there is a marked variation between kth and $k+$1th eigenvalues.

Usually, all the three criteria are used in order to choose the optimum number k of PCs.

4.2. PCA application

The PCA has been applied on the ten 20×15 matrices, one for each of the wood defects and clear wood considered, composing the $n \times p$ data matrix where n is the number of exemplars and goes from 1 to 200, while p is the number of features from 1 to 15. In Table 2 the mean, the standard deviation, and the number of measurements available for each variable have been reported, while in Table 3 the correlation matrix between the variables is reported.

The Bartlett's test of sphericity has been applied on the correlation matrix, and the null hypothesis of absence of correlation between the variables indicates the safe use of PCA on this data. Table 4 shows the percentages of variance related with the extracted PCs. This indicates that all the variables except 11th, are well explained by the extracted PCs having percentages of variance higher than 50%, between 58.8 and 98.4%. Table 5

Table 2.   Statistical measures calculated for each variable.

	Mean	Std. deviation	Total no. of exemplars N
var1	140.5200	34.4987	200
var2	140.7950	34.8882	200
var3	152.7650	47.7006	200
var4	16.5550	13.5653	200
var5	72.7250	204.6147	200
var6	2833.8050	4538.5314	200
var9	104.9650	40.3886	200
var10	188.5550	48.7751	200
var11	11.7600	11.7756	200
var12	15.3750	27.6862	200
var13	712.9050	366.6647	200
var14	82.7200	91.1395	200
var15	46.8800	48.2954	200
var16	151.7300	111.8529	200
var17	95.2250	60.1066	200

contains the amount of total variance contributed by each component and the percentage of cumulative variance. Considering just the first four components, the percentage of cumulative variance was equal to 81.876%.

The scree plot in Figure 2 shows a change of the slope between the considered four or five components. Therefore, according to the scree plot there are four PCs.

Since the first four components are responsible for 81.876% of the total variance, which is clearly higher than the 46.329% suggested for $p = 15$, and have eigenvalues higher than 1, and the scree plot shows a pronounced bend between 4 and 5, the three criteria agree to choose 4 as the optimum number of PCs.

In Table 6 the correlation coefficients between each PC and each variable are reported. The sign of the coefficients indicates if the relation between the coefficient and the variable is direct or inverse, while the numeric value indicates the correlation strength between the variables.

The first PC shows a strong direct relation with the variables number 4, 5, and 6; the second PC shows a strong direct relation with the variables 1, 2, 3, and 9; the third PC has an inverse relation with the variable 17; finally the fourth PC shows a strong relation with the variable 12. For all these cases of strong relation, a Pearson coefficient $\alpha \geq .7$ has been considered. The Pearson coefficient indicates the number of variables sufficient to represent a particular PC. Table 7 shows that when $\alpha \geq .6$, indicating a moderate relation between the variables, 13 variables instead of the initial 15 should be considered, while when $\alpha \geq .7$, indicating a strong relation between the variables, just 9 variables should be considered.

The above results have found four PCs. Based on the results in Table 7 and $\alpha \geq .6$, features 4, 5, 6, 15, and 16 are merged under one component which means they have something in common. These features are standard deviation, skewness, kurtosis, number of edge pixels for the pixels after thresholding at $\mu - 2\sigma$, and number of edge pixel for the pixels after thresholding at $\mu + 2\sigma$, respectively. These features are statistical features which illustrate the characteristics of the image data distributions, according to the Gaussian distribution. Therefore, these features merged under one PC represent the Gaussian characteristics of the image.

Table 3.   Correlation matrix between the variables.

	var1	var2	var3	var4	var5	var6	var9	var10	var11	var12	var13	var14	var15	var16	var17
var1	1	.987	.648	-.103	-.218	-.221	.833	.498	.205	-.175	.352	.018	.057	-.147	-.014
var2	.987	1	.573	-.188	-.327	-.297	.819	.435	.234	-.143	.384	.086	.119	-.225	-.053
var3	.648	.573	1	.536	.315	.315	.353	.684	.091	-.243	-.029	-.143	-.127	.151	-.034
var4	-.103	-.188	.536	1	.698	.800	-.442	.642	-.059	.108	-.610	-.173	-.269	.201	-.170
var5	-.218	-.327	.315	.698	1	.903	-.218	.576	-.278	.157	-.376	-.439	-.396	.577	.263
var6	-.221	-.297	.315	.800	.903	1	-.407	.656	-.022	.314	-.454	-.211	-.277	.323	.014
var9	.833	.819	.353	-.442	-.218	-.407	1	.202	-.152	-.243	.540	-.208	-.063	.072	.305
var10	.498	.435	.684	.642	.576	.656	.202	1	.057	.365	-.220	-.205	-.218	.185	.014
var11	.205	.234	.091	-.059	-.278	-.022	-.152	.057	1	.052	.036	.458	.329	-.454	-.478
var12	-.175	-.143	-.243	.108	.157	.314	-.243	.365	.052	1	-.173	.009	-.006	-.100	-.060
var13	.352	.384	-.029	-.610	-.376	-.454	.540	-.220	.036	-.173	1	.080	.265	-.027	.387
var14	.018	.086	-.143	-.173	-.439	-.211	-.208	-.205	.458	.009	.080	1	.836	-.564	-.523
var15	.057	.119	-.127	-.269	-.396	-.277	-.063	-.218	.329	-.006	.265	.836	1	-.493	-.382
var16	-.147	-.225	.151	.201	.577	.323	.072	.185	-.454	-.100	-.027	-.564	-.493	1	.778
var17	-.014	-.053	-.034	-.170	.263	.014	.305	.014	-.478	-.060	.387	-.523	-.382	.778	1

Table 4.   Amount of variance of each variable related with the extracted PCs.

	Initial	Extraction
var1	1.000	.984
var2	1.000	.972
var3	1.000	.879
var4	1.000	.915
var5	1.000	.847
var6	1.000	.885
var9	1.000	.918
var10	1.000	.977
var11	1.000	.491
var12	1.000	.892
var13	1.000	.648
var14	1.000	.715
var15	1.000	.588
var16	1.000	.750
var17	1.000	.820

Table 5.   Eigenvalues and percentage of variance.

Component	Initial eigenvalues			Extraction sums of squared loadings		
	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %
1	5	31	30.509	4.576	30.509	30.509
2	4	24	54.814	3.646	24.304	54.814
3	3	19	74.27	2.918	19.456	74.27
4	1	8	81.876	1.141	7.607	81.876
5	.99	7	88.491			
6	.67	4	92.93			
7	.38	3	95.462			
8	.29	2	97.416			
9	.16	1	98.453			
10	.11	.734	99.187			
11	.07	.47	99.657			
12	.03	.2	99.857			
13	.02	.109	99.965			
14	0	.027	99.992			
15	0	.008	100			

The features 1, 2, 3, 9, and 10 are mean gray level, median gray level, mode gray level, gray level p for which there are 20 pixels below, and gray level s for which there are 20 pixels above, respectively. What is common in these features is that they are calculated from the average gray level.

The third PC contains features 11 and 17, where feature 11 is related with the histogram tail length on the dark side and feature 17 is the number of the edge pixels after thresholding at $\mu + 2\sigma$. There is a link between the edge pixels and histogram of both dark and bright side, where it is expected to have brighter image on the edge after threshold which changes the histogram of both bright and dark pixels. However, PC has only found the direct relationship between dark side and thresholded edges. Therefore, this PC is sensitive to the histogram changes at the edge.

The last PC only contains the feature 12 which measures the histogram tail length on the bright side.
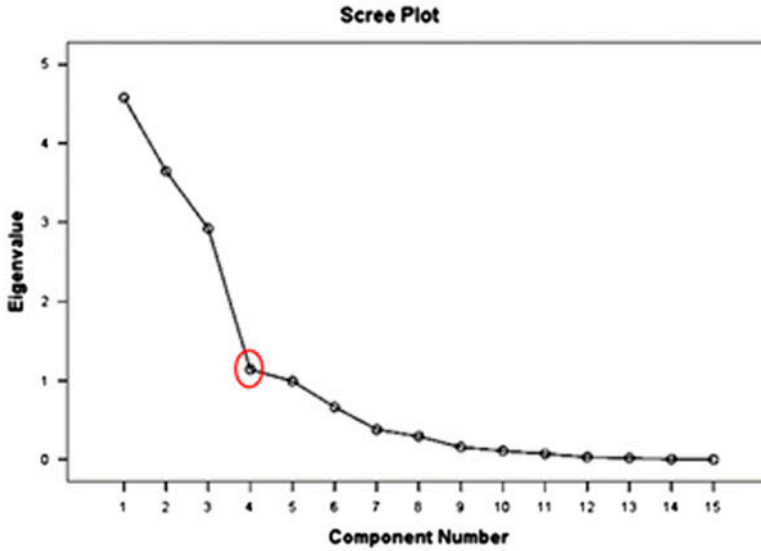
Figure 2.   Scree plot of eigenvalues.

Table 6.   Correlation coefficient – *moderate correlation in bold and high correlation in italics* – between the variables and the four extracted PCs.

	Component			
	1	2	3	4
var1	−.353	*.893*	.249	.011
var2	−.447	*.840*	.252	.053
var3	.244	*.749*	.411	−.300
var4	*.781*	.059	.503	−.217
var5	*.909*	.118	.062	.056
var6	*.854*	.005	.371	.140
var9	−.388	*.834*	−.257	.074
var10	.519	**.613**	.500	.287
var11	−.349	−.066	**.602**	.044
var12	.236	−.196	.234	*.862*
var13	−.560	.342	−.382	.270
var14	−.561	−.337	.532	.052
var15	**−.598**	−.245	.391	.136
var16	**.611**	.244	−.563	−.010
var17	.268	.316	−.772	.226

Table 7.   Critical variables related with Pearson coefficient.

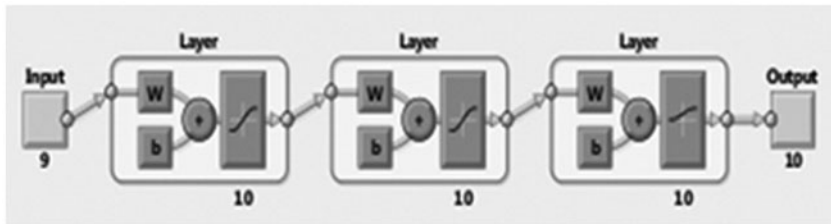| Pearson coefficient	PC 1	PC 2	PC 3	PC 4
$\alpha \geq .6$ (moderate correlation)	4, 5, 6, 15, 16	1, 2, 3, 9, 10	11, 17	12
$\alpha \geq .7$ (high correlation)	4, 5, 6	1, 2, 3, 9	17	12

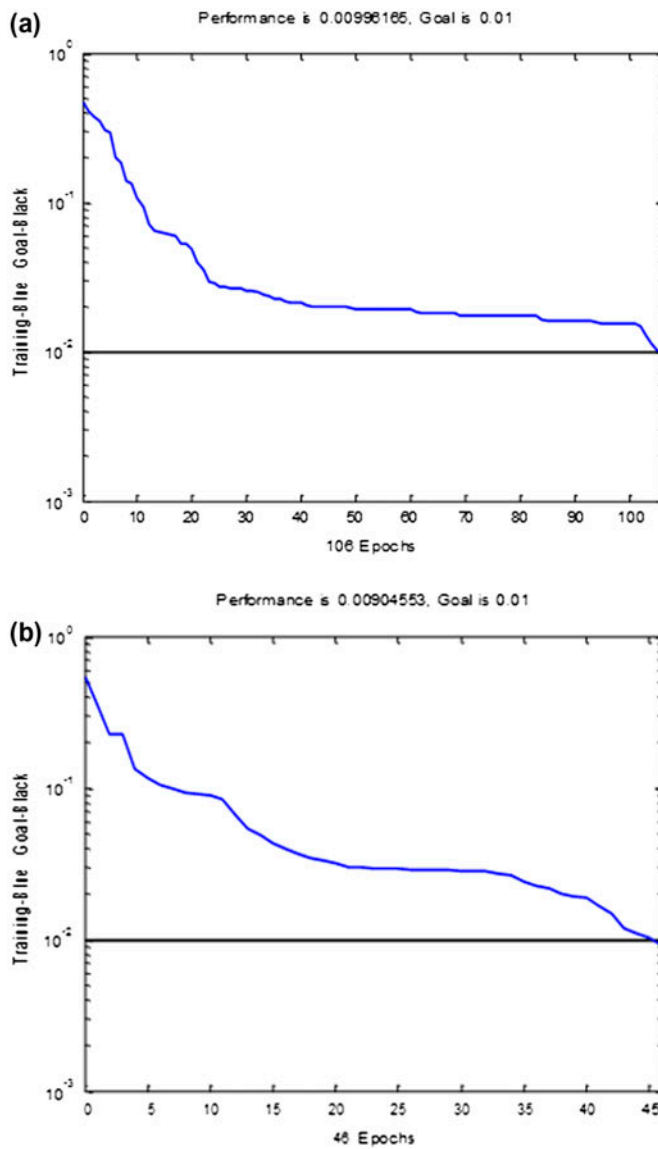Figure 3.   A feed-forward artificial neural network 9-10-10-10.

Figure 4.   (a) Training curve without feature reduction; (b) training curve with reduction feature set.

The reduced feature sets using PCA have been used to train the ANN classifier as described in the following section.

5.   ANN design

In this section the ANN, based on the results of the PCA, was designed in order to identify wood veneer defects.

A feed-forward ANN with a Back-Propagation learning algorithm has been chosen using 75% of the data-set for training and 25% for testing. The hidden layer has a *tan-sigmoid* transfer function because the data are normalized between −1 and +1, while the output layer has a *logsigmoid* transfer function because the output of the network should be 0 (for defect free) or 1 (for defect).

Figure 3 shows a network with an input layer of nine neurons representing the critical features considered, two hidden layers with 10 and 10 neurons, respectively, and finally the output layer with 10 neurons, one for each defect and clear wood. Figure 4 shows the classification results in terms of training curves in the case of reduced number of features according to the PCA and using all the features. It is clear how the performance of the network improves when the PCA is applied. The average results in terms of training time improved by 56.6%.

6.   Results and Taguchi analysis

Several network configurations have been tested in order to improve the performance. Moreover, Taguchi method has been applied on the training and test results in order to evaluate the effects of the three factors on the performance of the network. The Taguchi orthogonal array (design set) used is given in Table 8. The signal-to-noise ratio according to the criteria 'Larger is better' is expressed by the following expression

$$\frac{S}{N} = -10 \log_{10} \left[\frac{1}{N} \sum \frac{1}{Y^2} \right]$$

Table 8.   Taguchi L9 orthogonal array-design set.

Taguchi design – experiment number	Design set			Design level		
	Number of hidden layers	Number of neurons	Alpha Pearson	Number of layer	Number of neurons	Alpha Pearson
1	1	15-5-10	All features	1	1	1
2	1	13-10-10	$\alpha \geq .6$	1	2	2
3	1	9-15-10	$\alpha \geq .7$	1	3	3
4	2	13-5-5-10	$\alpha \geq .6$	2	1	2
5	2	9-10-10-10	$\alpha \geq .7$	2	2	3
6	2	15-15-15-10	All features	2	3	1
7	1	9-5-10	$\alpha \geq .7$	1[*]	1	3
8	1	15-10-10	All features	1[*]	2	1
9	1	13-15-10	$\alpha \geq .6$	1[*]	3	2

Note: 1[*] is the dummy array, set to 1.

Table 9. Results of the ANN training and test using one hidden layer.

Number of neurons	1 Train (%)	1 Test (%)	2 Train (%)	2 Test (%)	3 Train (%)	3 Test (%)	4 Train (%)	4 Test (%)	5 Train (%)	5 Test (%)	Mean Train (%)	Mean Test (%)
Pearson coefficient $\alpha \geq .7$												
9-5-10	93	40	93	48	91	68	91	44	97	60	93	52
9-10-10	91	48	93	42	95	84	91	50	97	60	93	57
9-15-10	90	50	87	56	96	56	86	70	93	78	91	62
Pearson coefficient $\alpha \geq .6$												
13-5-10	94	62	90	58	92	44	95	54	91	60	92	56
13-10-10	90	94	89	66	91	52	93	60	85	62	90	67
13-15-10	97	72	93	70	92	72	91	62	91	62	93	68
All features												
15-5-10	95	64	95	44	94	50	91	36	93	58	94	50
15-10-10	93	70	91	62	91	70	91	60	96	78	93	68
15-15-10	95	60	95	64	91	58	92	72	94	64	94	64

Table 10. Results of the ANN training and test using two hidden layer.

Number of neurons	1		2		3		4		5		Mean	
	Train (%)	Test (%)	Train (%)	Test (%)	Train (%)	Test (%)	Train (%)	Test (%)	Train (%)	Test (%)	Train (%)	Test (%)
Pearson coefficient $\alpha \geq .7$												
9-9-10-10	87	62	85	30	92	60	66	48	95	38	85	48
9-10-5-10	32	30	91	42	61	36	90	52	81	68	71	46
9-15-9-10	91	54	89	56	89	66	91	60	91	56	90	58
Pearson coefficient $\alpha \geq .6$												
13-12-11-10	89	86	93	70	90	82	100	84	91	78	92	80
13-13-10-10	95	84	89	80	89	84	90	56	86	82	90	77
13-10-5-10	90	58	92	60	62	40	89	72	92	74	85	61
All features												
15-10-5-10	92	40	76	36	93	46	92	34	92	66	89	44
15-15-10-10	93	74	73	34	70	54	87	56	89	46	82	53
15-13-12-10	90	80	93	58	94	82	93	62	89	70	92	70

Each experiment has been repeated five times and the average results have been taken into account. The main goal of these experiments is to investigate the outputs of the ANN in order to produce near-optimal expected results.

The first factor is the number of hidden layers for which two levels were considered, low and high. The low level denotes one hidden layer and high level denotes more than one hidden layer (in this case two hidden layers). In order to use an L9 Orthogonal Array a dummy level was used for level 3 which was set to level 1. The second factor was the number of neurons in the hidden layer and three levels were chosen, low, medium, and high. For single hidden layer, these levels varied from 5 to 15 neurons, whereas in the case of two hidden layers, levels were presented as lower than 10, 10, and higher than 10 neurons. Finally, the third factor considered was Pearson coefficient and three levels were chosen: complete correlation (level 1), moderate correlated features ($a \geq .6$, $a \leq -.6$) (level 2), and highly correlated features ($a \geq .7$, $a \leq -.7$) (level 3), respectively.
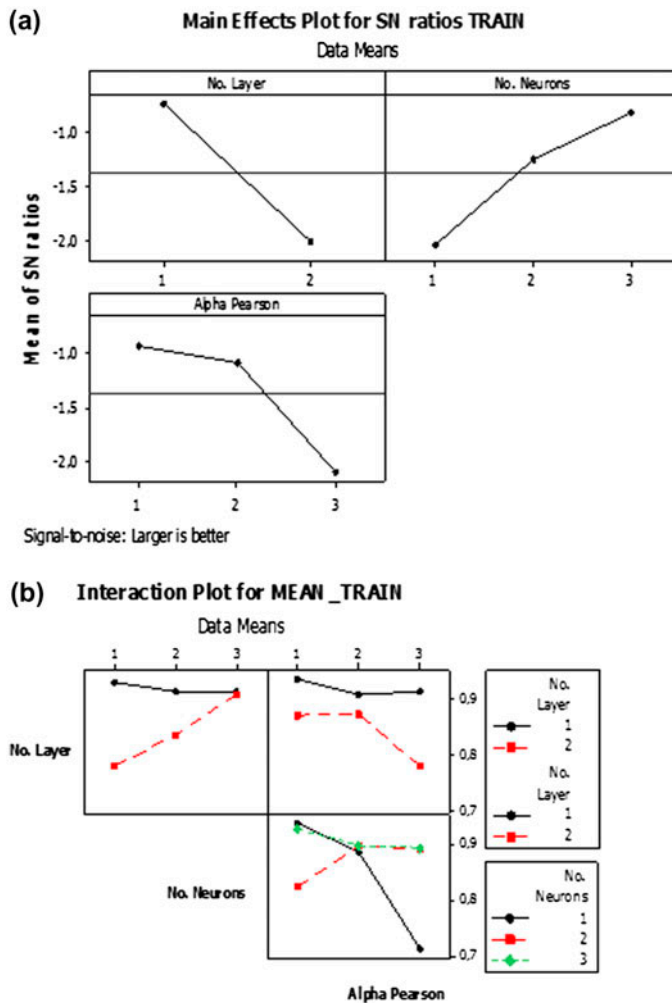

Figure 5.   (a) Taguchi analysis results during the training; (b) Interaction between factors during the training.

Table 9 shows the results using one hidden layer, changing the number of neurons in the hidden layer from 5 to 15 using three levels of the Pearson coefficient.

Table 10 shows the results when two hidden layers are used, changing the number of neurons in a way that their values are always included between the number of input and output neurons.

ANN training using one hidden layer, from 5 to 15 neurons per layer and the complete set of features in the issue maximizes the mean performance and minimizes standard deviation (i.e. *S/N* larger is better) (Figure 5(a)). The training of ANN is then influenced by the interaction between variables. Reducing the number of features to highly correlated variables (alpha person in level 3) has optimal effect in mean training performance if number of hidden layers is limited to 1. Increasing the number of neurons from 10 to 15 (level 3 in Figure 5(b)) has no interaction with layer complexity. When training ANNs with all features available (more input to the network), then the number of neurons does not influence the mean performance. Moderate neurons per layer (10, i.e. level 2) reduce training variance and time simultaneously maintaining
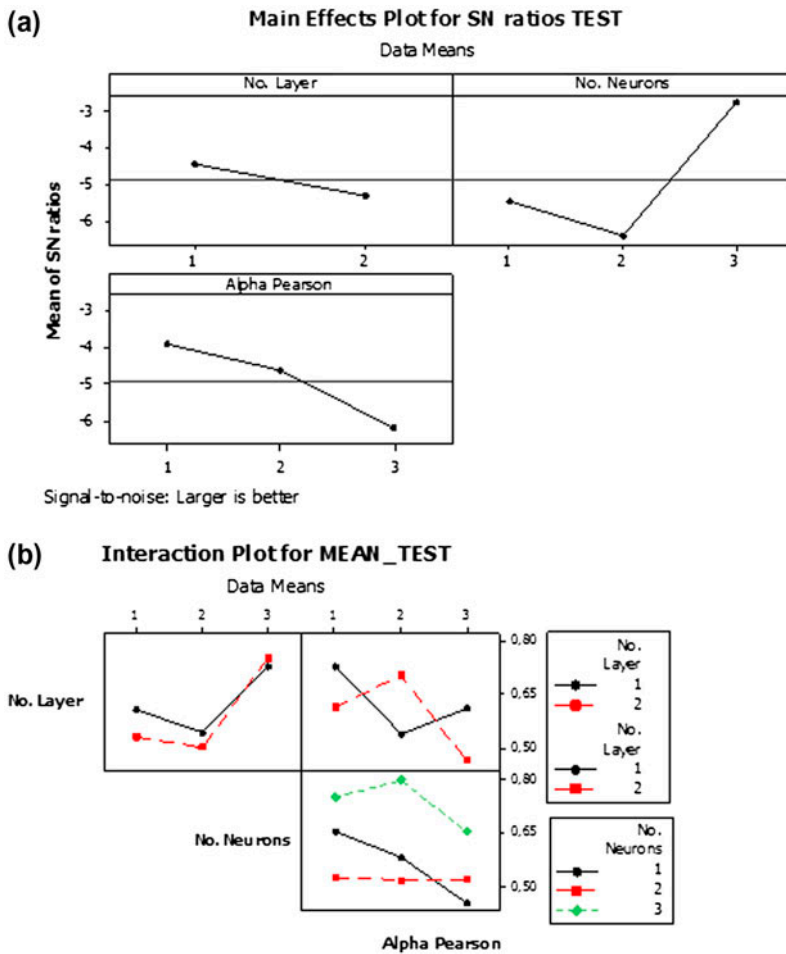

Figure 6.   (a) Taguchi analysis results during the test; (b) Interaction between factors during the test.

percentage of training success around 92%. The fact that the performance during the training does not change mean and S/N value much indicates that it is better to use lesser inputs to the network in order to reduce the computational time.

Figure 6(a) shows the signal-to-noise ratios and the results, according to the criteria 'Larger is better.' When the number of neurons increases, the variance minimizes. Moreover, the minimum number of hidden layers and more input in the network maximize the mean and minimize the variance. The optimal testing performance is achieved with 15 neurons. Figure 6(b) shows interaction between correlated variables and number of hidden layers of the mean performance of the ANN testing stage. The performance of ANN with the minimum number of hidden layers is found to be 73% with the reduced number of features. The performance of ANN testing stage with moderate correlated variables is 80%, which reduces testing process time by around 40%. This performance has been found with more than one hidden layer and 10 neurons in each layer. The worst performance is found with less number of inputs to the network and less number of neurons. The best performance is found with highly correlated features of the minimum level of hidden layers with more than 10 neurons per layer.

In summary, the results of Taguchi analysis show that the best performance of the ANN is reached using one hidden layer, a higher number of neurons in the hidden layer and a lower value of Pearson coefficient.

7. Conclusion

This paper has introduced ANN-based intelligent quality control for the detection of wood veneer defects with lower inspection criteria. To reduce the training time and increase the testing performance, a principal component analysis (PCA) based dimension reduction stage has been proposed. The proposed PCA method is based on the determination of the critical features for the inspection process. The proposed method has been applied on a case study for identifying defects on wood veneer. The reduced feature set has been used as inputs to train the ANN classifier which successfully identified the defects and clear wood. The best performance with one hidden layer was found to be with more than 10 neurons. The reduction of features reduces the number of epochs. According to the best performance, a reduction of 61 epochs increased the quality of outputs in testing stage by 18%. Different ANN designs have been studied by running some experiments with three control factors and carrying out Taguchi analysis on the results to determine the best-performing ANN topology.

References

Ahmadzadeh, F., & Lundberg, J. (2013). Remaining useful life prediction of grinding mill liners using an artificial neural network. *Minerals Engineering, 53,* 1–8.

Castellani, M., & Rowlands, H. (2009). Evolutionary artificial neural network design and training for wood veneer classification. *Engineering Applications of Artificial Intelligence, 22,* 732–741.

Charytoniuk, W., & Chen., M. S. (2000). *Neural network design for short-term load forecasting.* Proceedings of DRPT 2000, International Conference on IEEE, Electric Utility Deregulation and Restructuring and Power Technologies, London, UK.

D'Addona, D. M., & Teti, R. (2013). Image data processing via neural networks for tool wear prediction. *Procedia CIRP, 12,* 252–267.

Drake, P. R., & Packianather, M. S. (1998). A decision tree of neural networks for classifying images of wood veneer. *The International Journal of Advanced Manufacturing Technology, 14,* 280–285.

Lappalainen, T., Alcock, R. J., & Wani, M. A. (1994). *Plywood feature definition and extraction* (Report 3.1.2, QUAINT, BRITE/EURAM project 5560). Cardiff: Intelligent Systems Laboratory, School of Engineering, University of Wales.

McCulloch, W., & Pitts, W. (1943). A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics, 5*, 115–133. doi:10.1007/BF02478259

Mitra, A. (2012). *Fundamentals of quality control and improvement*. New Jersey, NJ: John Wiley and Sons.

Packianather, M. S. 1997. *Design and optimisation of neural network classifiers for automatic visual inspection of wood* (PhD thesis). School of Engineering, University of Wales, Cardiff.

Packianather, M. S., & Drake, P. R. (2000). Neural networks for classifying images of wood veneer. Part 2. *International journal of advanced manufacturing, 16*, 424–433.

Packianather, M. S., & Drake, P. R. (2004). Modelling neural network performance through response surface methodology for classifying wood veneer defects. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 218*, 459–466.

Packianather, M. S., & Drake, P. R. (2005). Comparison of neural and minimum distance classifiers in wood veneer defect identification. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 219*, 831–841.

Packianather, M. S., Drake, P. R., & Rowlands, H. (2000). Optimizing the parameters of multilayered feedforward neural networks through Taguchi design of experiments. *Quality and Reliability Engineering International, 16*, 461–473.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space (PDF). *Philosophical Magazine, 2*, 559–572.

Pham, D., & Alcock, R. J. (1996). Automatic detection of defects on birch wood boards. *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering, 210*, 45–52.

Pham, D. T., & Alcock, R. J. (1998). Artificial intelligence based techniques for processing segmented images of wood boards. *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering, 212*, 119–129.

Pham, D. T., & Alcock, R. J. (1999a). Recent developments in automated visual inspection of wood boards. In S. G. Tzafestas (Ed.), *Advances in manufacturing-decision control and information technology* (pp. 79–88). London: Springer.

Pham, D. T., & Alcock, R. J. (1999b). Plywood image segmentation using hardware-based image processing functions. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 213*, 431–434.

Pham, D. T., & Alcock, R. J. (1999c). Automated visual inspection of wood boards: Selection of features for defect classification by a neural network. *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering, 213*, 231–245.

Pham, D. T., Ghanbarzadeh, A., Koc, E., Otri, S., & Packianather, M. (2006). *Optimising neural networks for identification of wood defects using the Bees Algorithm*. Proceedings of the IEEE 2006 International Conference on Industrial Informatics, Singapore.

Pham, D. T., & Liu, X. (1995). *Neural networks for identification, prediction and control*. London: Springer.

Pham, D. T., & Sagiroglu, S. (2000). Neural network classification of defects in veneer boards. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 214*, 255–258.

Pölzleitner, W., & Schwingshakl, G. (1992). Real-time surface grading of profiled wooden boards. *Industrial Metrology, 2*, 283–298.

Rajput, N. S., Das, R. R., Mishra, V. N., Singh, K. P., & Dwivedi, R. (2010). A neural net implementation of SPCA pre-processor for gas/odor classification using the responses of thick film gas sensor array. *Sensors and Actuators B: Chemical, 148*, 550–558.

Roy, R. (1990). *A primer on Taguchi method*. New York, NY: Van Nostrand Reinhold.

Sratthaphut, L., Jamrus, S., Woothianusorn, S., & Toyama, O. (2013). Principal component analysis coupled with artificial neural networks for therapeutic indication prediction of thai herbal formulae. *Silpakorn University Science and Technology Journal, 7*, 41–48.

Tabe, H., Simons, S. J. R., Savery, J., West, R. M., & Williams, R. A. 1999. *Modelling of multiphase processes using tomographic data for optimisation and control*. Proceedings of 1st World Congress on Industrial Process Tomography, Buxton, UK.