COLOR GRADING OF COTTON-GRADING (PART II)

Luo Cheng, Kermit E. Duckett and Terezie Zapletalova
The University of Tennessee
Knoxville, TN
Michael Watson
Cotton Incorporated
Raleigh, NC
Hossein Ghorashi
Zellweger Uster
Knoxville, TN

Abstract

In this part of the series, two color grading systems were developed using expert system and neural networks. Both grading systems have two modes of operation, classification mode and training mode. In the training mode, the expert system can be trained by a statistical method based on Bayes' theorem or genetic algorithm. For neural network approach, the grading system can be trained by backpropagation algorithm or probabilistic neural network. Using 100 cotton samples from USDA, the agreement with classer can be improved from the original 50% from HVI grading to 86% - 100% depending on the training method and the training samples. The relative contributions of each measurement on color grading were also investigated using stepwise discriminate analysis.

Introduction

Color is vitally important to manufacturers and an accurate call on color is extremely important. As a result, instrumented color systems have been developed that attempt to make this call which cotton classers have made for many years. Assuming that the cotton classer can correctly evaluate color grade, a goal becomes one where High Volume Instrument (HVI) color grade matching with classer grade is predicted in all cases. Unfortunately, this is not the situation and much effort has been put forward to identify instrument system factors that enhance predictability of classer grade.

In Part I [1], color measurement was discussed using both a spectrometer CIE-based average color measurement and a color uniformity measurement using image analysis. The purpose of these efforts is to improve the accuracy of the measurement and to provide additional measurement for color grading. In this paper, we discuss the development of an expert system and a neural network based grading system that use these additional measurements to improve the agreement between cotton classer and instrument grading.

Reprinted from the *Proceedings of the Beltwide Cotton Conference*
Volume 1:649-653 (1999)
National Cotton Council, Memphis TN

Current Color Grading System

Classers Grade is composed of trash, preparation and color. The color grade reflects the varying amount of yellow color. In the process of grading, the classer visually compares the cotton samples with a set of standard cottons under standard illumination. The classer not only observes the color shade, but also variation in color, trash, yellow spot distribution and other visual effects in the process of grading. The HVI grades cotton based on a color/grade translation chart (Nickerson-Hunter color diagram). The color/grade translation chart is a 2-D graph of Rd versus +b, on which is overlaid the cotton classer's color grade (white, light spotted, spotted, tinged and yellow stained) and scale preparation (GM, SM, M, SLM, LM, SGO and GO). The HVI system currently provides Rd and +b parameters obtained from an average response of reflected and color-filtered light that is translated into a number representative of color grade. Figure 1 illustrates the HVI color chart used in color grading. For example, the number 43-3 indicates SLM for grade on higher side and spotted for color on lower side.

Expert System and Neural Networks for Color Grading

Expert System

An expert system is a computer-based system that uses knowledge and reasoning techniques to solve problems that normally require the abilities of human experts. The knowledge the expert system uses is made up either of rules or experience information about the behavior of the elements of a particular subject domain. Rules generally give descriptions of a condition, followed by implications of that condition. Experience information is a collection of experiences on a particular subject. The reasoning technique often used by expert systems is inferencing. An answer to a question posed to the expert system is inferred from the knowledge and facts presented to the expert system. Expert systems have been structured in many ways. The various expert system architectures include different components. However, certain components are common to most expert systems: a knowledge base and an inference mechanism.

HVI systems have been used in cotton grading for over 20 years, without necessarily replacing the cotton classer. During this period, experience has told us that if there is a disagreement between the HVI colorimeter and cotton classer, the cotton classer usually will assign the sample to a lower color grade than was assigned by the colorimeter. This is attributed to the fact that the human eye can see more detail than can the colorimeter. Therefore, we choose to develop an expert system that will use our experience to grade samples. Figure 2 illustrates the block diagram of the grading system. The inputs are HVI color grade, spectrometer redness (a*), variance and contrast of color, and percent area of yellow spots and trash. The expert system has two modes of operation - classification mode

and training mode. In the classification mode, based on HVI color grade, the expert system will evaluate a set of rules to make a decision whether it should be downgraded or not based on additional information. The decision boundaries are determined by a training module in training mode using two techniques, Bayesian and genetic algorithms (GA). The first decision-making technique used in this research is a statistical approach. A set of characteristic measurements, denoted features, is extracted from the input data and is used to assign each feature vector to one of the classes. The goal of any classifier is to minimize the number of errors made, or to maximize the likelihood of making a correct class assignment. This can be done using Bayesian theory. Alternatively, decision boundaries can be determined by genetic algorithms. A genetic algorithm (GA) is a directed random search technique, invented by Holland. Natural selection is a process by which nature causes those chromosomes that encode better characteristics to reproduce more often than those that encode poorer characteristics. Natural selection is the process that causes genetic algorithms to produce near-optimal solutions when the selected chromosome is decoded. This process involves creation of many chromosomes by reproduction, crossover, mutation and the survival of the chromosomes with the better characteristics. Successive generations of chromosomes improve in quality. The fitness evaluation function acts as an interface between the GA and the optimization problem. Where a mathematical equation cannot be formulated for this task, a rule-based procedure can be constructed for use as a fitness function or, in some cases, both can be combined. The fitness function in this case uses the same rules in classification mode. In each evaluation process, the fitness function executes the rules and returns the error of classification. After the selection and reproduction process, the classification error is reduced in successive iterations. The evolution process will be terminated if there is no significant improvement in a pre-specified number of iterations.

Neural Networks

Historically, the two major approaches to pattern recognition are the statistical and syntactic approaches. Recently, the emerging technology of neural networks (NN) has provided a third approach for pattern recognition algorithms. To some extent, the NN approach is a non-algorithmic, black-box strategy, which is trainable. We hope to 'train' the neural black box to learn the correct response or output for each of the training samples. This strategy is attractive to the classifier designer, since the required amount of *a priori* knowledge and detailed knowledge of the internal system operation is minimal. Furthermore, after training we hope that the internal structure of the artificial implementation will self-organize to enable extrapolation when faced with new, yet similar, patterns on the basis of 'experience' with the training set.

Multi-layer Perceptron (MLP) is perhaps the best known type of feedforward networks. Neurons in the input layer only act as buffers for distributing the input signal to neurons in the hidden layer. Each neuron in the hidden layer sums up its input signals after weighting them with the strength of the respective connections from the input layer and computes its output using a sigmoid function of the sum of the weighted inputs. The backpropagation algorithm is the most commonly adopted MLP training algorithm. Iteratively, beginning with the output layer, the error term is computed for neurons in all layers and weight updates determined for all connections. The weight updating process can take place after the presentation of each training pattern (pattern-based training) or after the presentation of the whole set of training patterns (batch training). In either case, a training epoch is said to have been completed when all training patterns have been presented once to the MLP. The probabilistic neural network provides a special neural network for solving pattern classification problems. The PNN is a neural network implementation of Bayesian classifiers. It has been shown that the probabilistic neural network (PNN) approaches the Baysian classifier if the training set is large. PNN is fairly specialized, not having the wide applicability of the MLP. PNN is intrinsically a classifier and the advantage of this network model is that a byproduct of its computations is Bayesian posterior probabilities.

Experimental

Descriptive Statistics of the Samples

One hundred cotton samples were provided by USDA for which HVI color grade agrees and disagrees with classer grade, respectively. Among the 100 samples, 50 samples show agreement between HVI colorimeter and the classer, whereas the remaining 50 samples show disagreement. The cottons do not cover the whole range of classer grades, but are restricted to White, Light Spotted and Spotted. The discrepancies are wholly between the first two classifications and, hence, very localized for overall color/grade translation.

Table 1 lists the categorized means and standard deviations for these 100 samples. Generally, mean values of L* decrease, whereas mean values of a* and b* increase as color grade goes down from white to spotted. These mean values are well separated in different color grades. For color variations, both L* variance and b* variance increase with the decrease of the grades as expected, but the mean value separation in L* is small in light spotted and spotted grade. Similarly, b* contrast shows good separation, but not L* contrast. Trash area and yellow spot area increases with the decrease in color grade. The mean values are well separated, but the standard deviations are high also. It should be noted that the variance of the sample is very important. Specifically, one can ask whether or not two or more groups are significantly different from each other with respect to the mean of a particular variable. If the means for a variable are significantly different in different groups, then we can say that this variable discriminantes between the groups.

Table 2 lists the correlation matrix of the variables. The highest correlation among the variables is between a* and b* with an R-value of 0.80. Thus when a sample has higher yellowness value, it usually also has higher redness value. Therefore, if more precise description of the sample is needed, a* should not be ignored. Correlation between a* and L* comes next with an R-value of –0.63. The negative R-value means that the lower the lightness, the higher the redness.

Procedure

An optical spectrometer, manufactured by *Ocean Optics*, was used to obtain CIE L*, a* and b* means. Because the spectrometer measures the entire visible spectral distribution, the means for L*, a* and b* are superior to the scanner approach. A color scanner has been used to obtain color images from which yellowness variation [variation in b* in CIE standardization] and yellow spot area can be obtained. The color scanner image is converted pixel-by-pixel in an algorithmic translation of R, G and B parameters to CIE color space parameters L*, a* and b*. Color uniformity, contrast, and trash/yellow spot content were obtained by software we developed and described in Part I of the series.

The inputs to expert system are HVI color grade and the additional measurements described in Part I to make adjustment on the initial HVI color grade. The neural network based grading system does not use the color translation chart. Instead, the raw measurements such as L* and b* are used together with additional measurements.

A flow chart of the process of designing a learning machine for classification is illustrated in Figure 3. The first step is the selection of architecture of the classifier. Once the architecture is determined, the next step is the feature selection process. The collected experimental data will be divided into two sets, one for training and the other for testing. In each of these two sets are 25 samples for which there is HVI and classer agreement and 25 samples for which there is disagreement between HVI and classer. If the testing result is good, the system is said to be trained and ready for use. Otherwise, the architecture must be modified or the features need to be refined.

Results and Discussion

Feature Selection

The objective of feature selection is to reduce the number of measurements we require. What one would like to do is to select the best m variables where 'best' implies lowest error rate. This is similar to the variable selection problem in regression analysis where one attempts to find the best m variables to predict the dependent variable. The most common methods are stepwise forward and stepwise backward discriminate analysis. Usually, the measures used in stepwise methods are partial Wilks' lambda, F- ratio and tolerance. In general, Wilks' lambda is the standard statistic that is used to denote the statistical significance of the discriminatory power of the model. Its value will range from 1.0 (no discriminatory power) to 0.0 (perfect discriminatory power). A Partial Wilk's lambda is the Wilks' lambda for the unique contribution of the respective variable to the discrimination between groups. One can look at this value as the equivalent to the partial correlation coefficients reported in Multiple Regression Analysis. Because a lambda of 0.0 denotes perfect discriminatory power, the lower the value, the greater is the unique discriminatory power of the respective variable. The conditional F-ratio is the univariate F-ratio, associated with a particular variable, testing the difference between the group means conditional on the variables already entered in the equation. It is a measure of how much a given variable contributes to the group differences given the variables already included. The F-ratio is proportional to the ratio of the between groups sums of squares and the within groups sums of squares and, hence, it is easy to see that large values of F correspond to well separated groups. The tolerance value is defined as 1 minus R-square of the respective variable with all other variables in the model, and this value gives an indication of the redundancy of the respective variable. It ranges from 0.0 to 1.0 with 0.0 representing 100% redundancy.

The results of stepwise forward analysis are presented in Table 3. Among the 10 variables, 8 of them were selected by stepwise analysis with the criteria of an F value greater than 1.0 to enter the model. Redness (a*) was selected as the most significant variable in contributing to the discrimination with a 'Partial Lambda' value of 0.659 (significantly lower than the rest) and an F value of 23.25 (significantly higher than the rest). Yellowness (b*) comes next with a 'Partial Lambda' value of 0.831 and an F value of 9.12. Trash area, b* variance, and Yellow spots area take the third, forth and fifth places in contributing discrimination power with closer 'Partial Lambda' values and F values. The remaining variables - a* variance, L* variance and L* contrast - contribute only a little in discrimination with higher "Partial Lambda' values and lower F-values. It is a little surprising that lightness (L*) and yellowness contrast were not selected during the stepwise analysis because of low F values (< 1.0). From the tolerance value, the results show that there are no redundant variables.

It is important to realize that, although the stepwise forward and stepwise backward methods are very similar in their appearance, they can produce different results even when using the same measure of goodness. The result of stepwise backward analysis is presented in Table 4. The results are very similar to stepwise forward analysis. The contribution in discrimination are in the order of a*, b*, trash area, yellow spots area, b* variance, a* variance and L* variance. L* contrast was not selected in stepwise backward when comparing with the stepwise forward. Yellowness (b*) variance ranked third place in stepwise forward, but it was ranked fifth in stepwise backward. Similar to stepwise

forward, L* and b* contrast were not selected because of the low F values.

Combining the results from both stepwise forward and backward, a*, b*, b* variance, yellow spots area, trash area, a* variance and L* variance were selected as the inputs to the neural networks. The reason that L* was not selected by the stepwise analysis may come from the narrow regions of the samples. With a wider range in samples, L* might be expected to be an important parameter and, therefore, it is included in the input vector to the neural networks.

Expert System Training and Classification

The training set was trained separately by both statistical and genetic algorithms. After training, another set was tested on the expert system and compared with their opposing group. The results are presented in Table 5. When we used set #1 for training and set #2 for testing, a 100% agreement between HVI and classer was obtained for the testing set, for the expert system trained using the statistical approach. If set #2 was used for training, a 98% agreement was obtained for testing set #1.

For the expert system trained using the genetic algorithm, where we trained the expert system using set #1, a 92% agreement was obtained between HVI and classer for set #2. If set #2 was trained, an 86% agreement was obtained for set #1. This result shows that the Bayes approach performs better than the genetic algorithms approach, for the samples we investigated. This is not surprising because the Bayes classifier is an optimized classifier for a sample distribution following a normal distribution. The training sample size is small and that also limited the performance of the genetic algorithms approach. This study is an exploration of the feasibility of the two approaches. Additional samples should be tested before drawing final conclusions.

It is also found that the training sample is very important. Training samples should be representative and sample size should be large enough to ensure unbiased classification. The most encouraging result is that this study demonstrates that the agreement between HVI grading and that of the cotton classer can be significantly improved using additional measurements discussed in Part I of the series, provided the classifier is properly trained.

Neural Network Training and Classification

Data set #1 was trained separately by both MLP and PNN. After training, set #2 was tested and compared with their opposing group. In training the MLP network, some parameters need to be identified which are important to the performance. These parameters include the number of hidden layers, the number of neurons in the hidden layer, learning rate and momentum coefficients. The input layer has eight neurons that are connected to the eight measurements selected by the feature selection process. The output layer has five neurons representing each of the five color grades. It has been proven that a network with only one hidden layer can compute any arbitrary function of its inputs; therefore, one hidden layer will be used. The number of neurons in the hidden layer will be determined by experiment. With too few hidden neurons, the network may be unable to create adequately complex decision boundaries. However, if there is a large number of hidden neurons, it is more difficult for the trained network to create a generalized mapping using the training data. The network will perform well on training data, but fail on the unseen data set. Learning rate and momentum are also important parameters in training a MLP network. We use a low learning rate with a high momentum for the starting point considering the noise in the data set.

The classification result on the testing set is illustrated in Figure 4. The lowest error rate was found when the hidden layer has six neurons. The error rate increased when the number of neurons were either increased or decreased.

PNN incorporates a parameter, σ, which is a smoothing factor that applies to the probability density estimate. There is no universal, mathematically rigorous method for choosing the best value of σ, yet it is crucial to competent performance. The optimal value of σ for every problem is different. It may be on the order of 1, or it may be on the order of 0.001. Figure 5 illustrates the classification error at three σ values from the testing set. A choice of σ that is either too low or too high will result in higher classification error. The lowest error rate was found for a smoothing factor of 0.1. Compared with the lowest error rate from MLP, it is higher (12% vs 6%).

It should be noticed that the training sample set is small. Therefore, it is not easy to draw a firm conclusion for the broadest application across color grading at this time. But this research has explored the feasibility of MLP and PNN to color classification schemes. In order to draw solid conclusions, more samples should be tested. In general, both approaches improved agreement between cotton classer and instrument grading. The agreement was raised from 50% to 88% or 94% depending on the neural network structure.

Conclusions

The feasibility of color grading using expert system and neural networks have been explored. Based on additional measurements and advanced training algorithms, the grading system can grade cotton color very close to that of the experienced human classer. Training sample selection and sample size is very important for classifier performance. The training samples should cover a wide range of features and should be representative of the population. In addition, the sample size should be sufficiently large.

Although this study is limited by the small amount of samples, its success is impetus for the pursuit of further studies to achieve higher classification rates. The neural

network training process is very important. Parameters such as the number of neurons in the hidden layer for MLP and the smoothing factor for PNN should be carefully determined to achieve optimum performance.

From the stepwise discriminate analysis for feature selection, redness (a*) was found to be the most important factor in color grading and yellowness (b*) to be next. Yellowness variance, yellow spot area and trash area also contribute in a smaller way to color grading.

Acknowledgements

We gratefully acknowledge Quality Section of USDA for providing cotton samples for this research.

Literature Cited

1. Duckett, K. E., Cheng, L., Zapletalova, T., Watson M., and Ghorashi, H., Color Grading of Cotton-Measurement (Part I). Beltwide Cotton Conference, Orlando, Jan. 5-8, 1999.

Table 1. Categorized Mean and Standard Deviation of the Samples.

	Classer Color Grade					
	White		Light Spotted		Spotted	
Measurement	Mean	Std.	Mean	Std.	Mean	Std.
L*	95.25	2.09	92.62	1.97	89.89	2.57
a*	1.80	0.24	2.49	0.24	3.08	0.28
b*	8.70	0.83	9.97	0.76	12.10	1.16
L* Variance	10.26	3.82	11.53	3.77	11.80	3.90
a* Variance	0.92	0.49	0.80	0.35	0.64	0.28
b* Variance	0.72	0.23	1.07	0.40	1.46	0.34
L* Contrast	12.30	5.53	13.96	4.84	13.24	6.47
b* Contrast	52.57	19.88	57.37	20.96	61.86	38.49
Trash Area %	0.076	0.11	0.10	0.05	0.199	0.11
Yellow Spots %	0.005	0.02	0.029	0.05	0.131	0.11

Table 2.1 Correlation Matrix.

	L*	a*	b*	L* _v	a*_v
L*	1.00				
a*	-0.63	1.00			
b*	-0.32	0.80	1.00		
L*_v	-0.29	0.08	-0.1	1.00	
a*_v	0.61	-0.12	0.11	-0.25	1.00
b*_v	-0.27	0.43	0.33	0.19	-0.04
L*_c	-0.01	0.08	0.11	-0.10	0.13
b*_c	-0.04	0.06	0.03	-0.03	0.19
T%	-0.46	0.21	0.03	0.55	-0.33
Y%	-0.31	0.45	0.47	0.24	0.01

Table 2.2 Correlation Matrix.

	b*_v	L*_c	b*_c	T%	Y%
L*					
a*					
b*					
L*_v					
a*_v					
b*_v	1.00				
L*_c	-0.05	1.00			
b*_c	0.22	0.45	1.00		
T%	0.24	-0.07	0.10	1.00	
Y%	0.41	0.01	0.20	0.18	1.00

L*_v --- L* Variance
a*_v --- a* Variance
b*_v --- b* Variance
L*_c --- L* Contrast
b*_c --- b* Contrast
T% --- Trash Area%
Y% --- Yellow Spot Area%

Table 3. Results from Stepwise Forward Analysis, F Value to Enter: 1.00.

Variable	Partial Lambda	F value	p-level	Tolerance
a*	0.659	23.25	0.0000	0.72
b*	0.831	9.12	0.0002	0.55
L* Variance	0.948	2.48	0.0893	0.60
a* Variance	0.921	3.86	0.0247	0.76
b* Variance	0.858	7.45	0.0010	0.89
L* Contrast	0.961	1.82	0.1678	0.94
Yellow Spots	0.870	6.75	0.0019	0.84
Trash Area	0.865	7.04	0.0014	0.62

Table 4. Results from Stepwise Backward Analysis, F Value to Remove: 1.00.

Variable	Partial Lambda	F value	p-level	Tolerance
a*	0.659	23.52	0.0000	0.74
b*	0.831	9.24	0.0002	0.55
L* Variance	0.950	2.41	0.0951	0.60
a* Variance	0.928	3.54	0.0331	0.78
b* Variance	0.870	6.81	0.0017	0.91
Yellow Spots	0.870	6.82	0.0017	0.84
Trash Area	0.865	7.11	0.0013	0.62

Table 5. Result from expert system.

Training Method	Bayes		GA	
	Training	Testing	Training	Testing
Set #1	96	100	94	92
Set #2	100	98	100	86

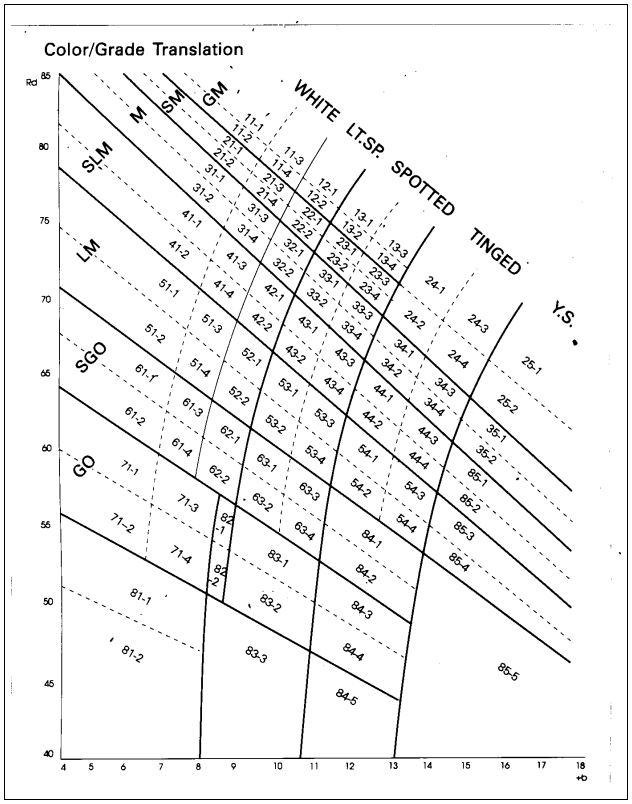

Figure 1. Color / grade translation chart

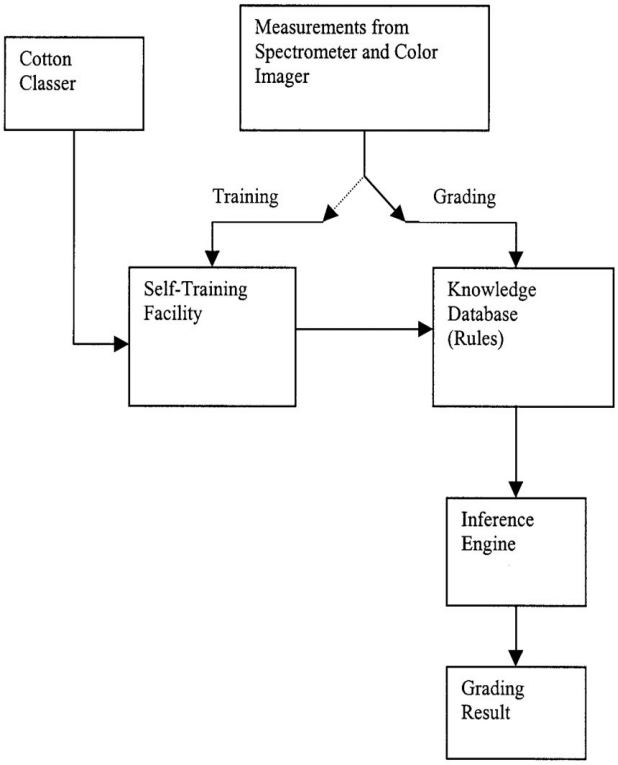

Figure 2. Block diagram of the expert system for color grading.

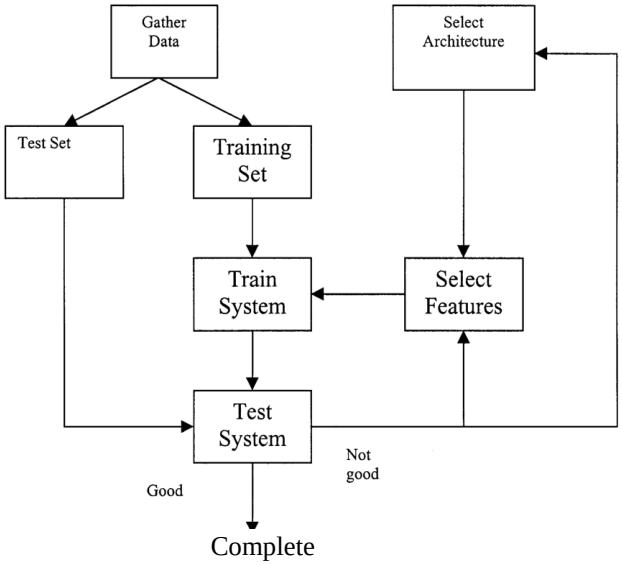

Figure 3. Steps involved in the design of a typical pattern recognition system.

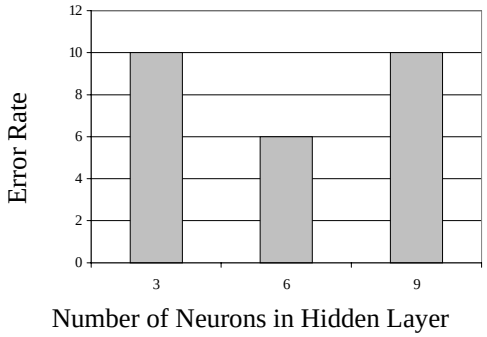

Figure 4. Error rate on testing set using MLP.

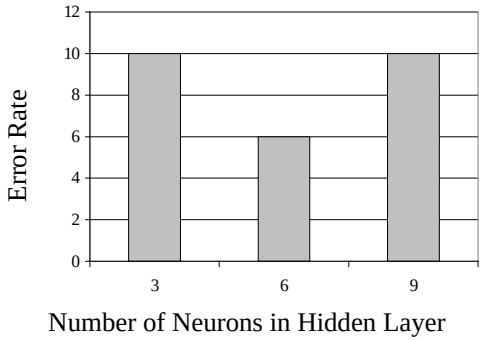

Figure 5. Error rate on testing set using PNN.