COTTON COLOR GRADING
BY NEURAL NETWORK

B. Xu, J. Su and D.S. Dale
University of Texas at Austin
Austin, TX
M.D. Watson
Cotton Incorporated
Raleigh, NC

Abstract

It is well known that disagreements in cotton color grades between the high volume instrument and classer are substantial. The machine-classer disagreement deters the full acceptance for the use of machine grading of cotton color. This paper first provides a quantitative analysis on the distributions of the disagreements across all the color grades, the major and sub-color categories. The study proves that the disagreements can be both systematic and random, and further analyzes the possible sources for these two types of disagreements. The paper devotes its second part to the introduction of a novel design of a neural network classifier for cotton color classification. This classifier consists of multiple networks performing a two-step classification that identifies the major and sub-color categories separately. The classifier can be trained by any desirable data. In this research, it was trained by using a set of classers' grades, and exhibited good generalization for the new testing data. The classifier seems to have reduced the machine-classer disagreements to a minimal level, which is limited by the classer's sustainability.

According to the USDA universal standards for Upland cotton, cotton colors are classified into five major categories based on chromatic differences (1-white, 2-light spotted, 3-spotted, 4-tinged and 5-yellow stained), and three to eight subcategories in one major category based on differences in grayness (1-good middling, 2-strict middling, …8-below grade) [3]. A double-digit number that indicates both the major and sub-categories of the color is used to denote a color grade. For example, color grade 21 refers to a white, strict-middling cotton. The color grade of a sample is determined either by a classer who compares the sample with the universal standards; or by the colorimeter of a high volume instrument (HVI) that calculates the location of the color data of the sample in the cotton color diagram. Although the HVI is a unique instrument currently used for grading cotton colors in the cotton classing system, its output has not been accepted as official color grading by the industry because of substantial disagreement with grades provided by a classer. A classer, who is trained to visually grade cotton color and trash, has the right to correct the HVI's rating when a dispute occurs. Since visual grading has been the traditional and widely accepted method for cotton color grading, the machine-classer disagreement undermines the industrial acceptance for the machine grading. To investigate new methods that can reduce this disagreement, it is necessary to understand the possible reasons causing the disagreements. In this paper, we will first report a study on the HVI-classer disagreements in color grades, and then present the preliminary results of using a neural network classifier to reduce these disagreements.

Reprinted from the *Proceedings of the Beltwide Cotton Conference*
Volume 1:654-657 (1999)
National Cotton Council, Memphis TN

HVI-Classer Disagreements in Color Grades

The HVI and classer color grades of 2489 cotton samples were collected randomly from 1996 U.S. crops. The numbers were counted for those that have conflicting HVI and classer grades. It was found that the total rate of disagreements across all the grades available from the samples was as high as 54.08%. The disagreements in respective grades are displayed in Figure 1. Each bar in the figure indicates the percentage of the samples that have conflicting grades. More than 99% of the disagreements occur between adjacent grades. A number of prominent peaks rise in the figure, indicating more severe HVI-classer disagreements in those grades than in others. The four highest peaks appear in the following pairs of HVI-classer grades: 21-22, 31-32, 41-42, and 51-52, revealing that the disagreements primarily occur among two major color categories: white and light spotted. In addition, all the high peaks stand above the diagonal line in the HVI-classer plane (horizontal), which means that the disagreements are biased. Samples labeled "light spotted" by classer are often graded "white" by HVI. The opposite case almost does not exist in these samples.

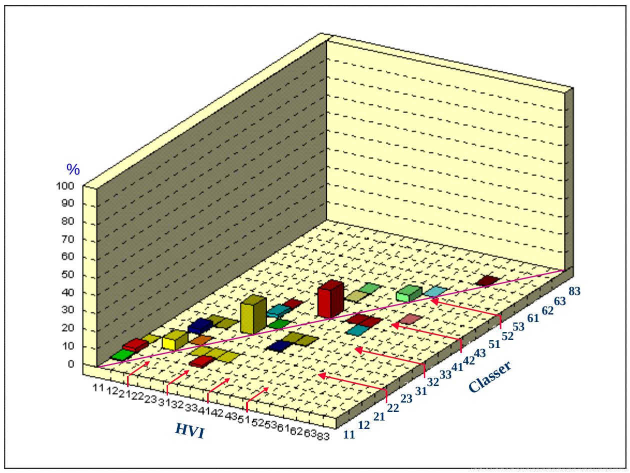

Figure 1 Distributions of HVI-Classer Disagreement in Color Grades

The disagreements can be examined from the perspectives of the major and sub-categories. Table I shows the distributions of the disagreements among the five major categories. A substantial amount (44.3%) of the samples were graded "white" by the HVI, but disputably graded "light spotted" by the classer. However, almost no "light spotted" samples graded by the HVI were graded "white" by classer. Hence, there is a biased trend in the disagreements between the white and light spotted categories. The disagreement in these two categories is a determinant in the total disagreement. Between the light spotted and spotted categories, the disagreements are nearly negligible and unbiased. The disagreements among other categories are not available from this sample set, but they are expected to be low. Table II shows the distributions of the disagreements among the seven subcategories. Overall, the disagreements in the subcategories are much lower, and more widely spread than in the major categories. The HVI has a slight tendency to give higher grades to the samples than the classer in the subcategories.

Table I Disagreement (%) in the Major Categories

Classer / HVI	White	Light Spotted	Spotted	Tinged	Yellow Stained
White		44.3	0		
Light Spotted	0		0.2		
Spotted	0	0.08			
Tinged					
Yellow Stained					

Table II Disagreement (%) in the Subcategories

Classer / HVI	GM	SM	M	SLM	LM	SGO	GO
GM		2.79					
SM	0		3.34				
M		1.07		2.19			
SLM			0.56		0.52		
LM				0.12		0	
SGO							
GO							

GM: good middling, SM: strict middling, M: middling, SLM: strict low middling, LM: low middling, SGO: strict good ordinary; GO: good ordinary.

From the above analysis, it is noted that the possible sources attributable to HVI-classer disagreements can be both systematic and random. Systematic disagreements mainly occur among the major color categories, particularly between "white" and "light spotted", and are the dominant component in the total disagreements. Following are the main reasons for systematic disagreements.

Inaccurate partition of the color space used in HVI

The color space, known as the Nickerson-Hunter color diagram, is a two dimensional (R_d~b) space that constitutes 30 blocks, representing 25 official color grades and five categories below grade color [3]. The partition of the color space was based on the experimental data of cotton crops in the 1950's. The boundaries between color blocks may not accurately represent color differences in today's cotton. Figure 2 shows the frequency of the 2489 samples at each grade with the boundary curves that separate the main color categories. The samples seem to gather naturally into two distinct groups, each having an approximately normal distribution (a bell-like shape). The two groups, which overlap extensively, were labeled "white" and "light spotted" by the classer. However, the white-light spotted boundary used by the HVI does not properly separate these two groups. The majority of the light spotted samples graded by classers were arranged into the "white" side of this boundary. This is why the HVI, which uses the boundary to classify cotton color, is more likely to grade a sample "white". The Cotton Division of USDA has decided to shift this boundary leftward to reduce HVI-classer discrepancies. This raises critical questions such as "what is the optimal partition of the color space?" and "are visible boundaries really needed to grade cotton colors?"

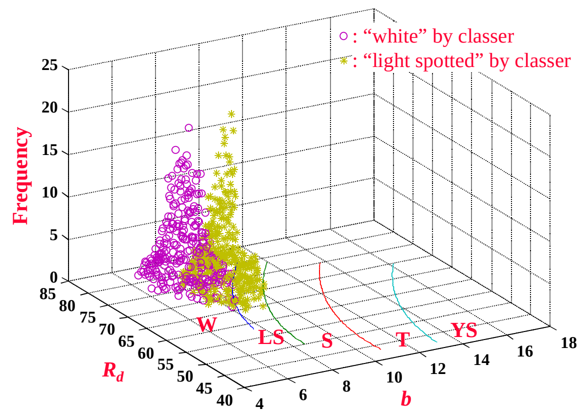

Figure 2. Sample Distributions and HVI Color Boundaries

Absence of physical standards for the light-spotted category

In the universal standards for Upland cotton, the standards for the light spotted category are descriptive. A sample between "white" and "light spotted" can be physically compared only with the white standards. If it does not match with one of the white standards, it will be assigned to "light spotted". But this sample might not *physically* match a light-spotted standard if it existed. Hence, a classer is more likely to label a "between" sample "light spotted." The lack of the physical standards for all the grades is also the major reason for low sustainability in classer's grades.

Ignorance of a color attribute—redness by HVI

An HVI colorimeter measures only two of the three color attributes of a sample: reflectance R_d and yellowness b. The third attribute, redness (or greenness) a, is ignored by the HVI, but influences the classer's color perception. A preliminary study has found that redness a of cotton makes up from 10% of chroma in the white category to 33% in the tinged category [6]. a may play an important role in color grading. Besides, the HVI colorimeter does not measure the variations of these color attributes, which are more or less utilized by classer.

Difference in viewing conditions

The classer can view a larger area of the sample than the HVI colorimeter can, and can also ignore the influence of trash and yellow spots that may otherwise affect the grade call. The HVI's rating is based on the average of the R_d, $+b$ data in a smaller area of the sample.

It is also noted that the disagreements among the subcategories are mainly random and insignificant. The factors that lead to random disagreements are:

Hard boundary. This refers to the arbitrary assignment of color grades made by the HVI when a color data point is on or close to the boundary of two grade blocks in the color diagram. When the color point is near one boundary, the assigned color grade will be very sensitive to slight changes in color.

Human subjectivity. Visual comparison of a sample with the universal standards is a subjective process. Human errors in grading are considered random.

Machine instability. The HVI colorimeter needs to be calibrated frequently to correct electronic drifting. The drifting changes the R_d and b readings.

The cotton industry is dissatisfied with both the HVI grading and the visual grading because of their high disagreements. The industry's consensus is that a new cotton color grading method should neither simply mimic classers nor solely rely on the current HVI color diagram; it should comply with the USDA universal standards. The new color measurement and data classification technologies provide better solutions for cotton color grading. In the previous papers [4.5,6], we reported the development of an imaging colorimeter that uses the image processing technology to measure the statistics of the color attributes of raw cotton. We plan to present a series of papers to discuss different classification methods applicable to color grading based on color measurement data. The following section of this paper deals with the application of the neural network to this problem.

Neural Network Classifier

From Figure 2, it can be seen that the severe overlap of the color data occurs between the white and light spotted categories. The boundary is not sharp or linear. A neural network (NN) is a computational system that can provide sophisticated mappings from a set of input variables to a set of output variables according to the relationships learned from the training data [1, 2]. An NN usually contains massive processing units (neurons) organized in successive layers. The neurons between two adjacent layers are connected with adjustable parameters governing the form of the input-output mapping. To perform an explicit mapping, the connections of neurons must be feed-forward. One of the most common feed-forward networks is a *multilayer perceptron* (MLP), which normally composes one input layer, one or more hidden layers and one output layer (Figure 3). To design an MLP for solving a specific classification problem, the developer needs to determine the inputs, outputs, number of hidden layers, number of neurons in each layer, and the training algorithm that are suited for the problem.

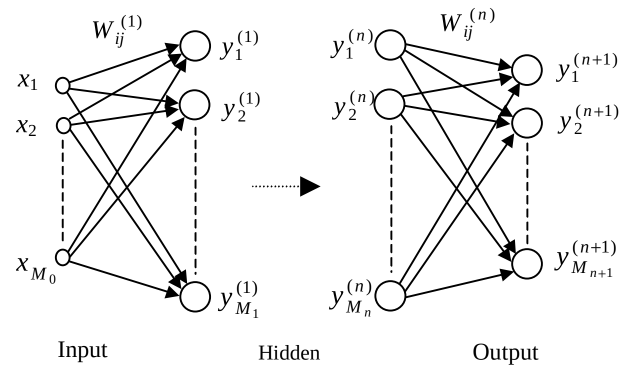

Figure 3. MLP Topology

To classify cotton color, the inputs of the MLP should utilize the statistic information, such as the means and standard deviations, of $R_d ab$ of samples. In this research, we were unable to collect enough samples that had been graded by the classer and HVI for the imaging colorimeter to measure these needed data. We had to use only the R_d and b means from the HVI as the inputs. The principle and procedure established by using these two inputs, however, are directly applicable to the one using more inputs.

Normally, the MLP uses one output neuron to represent one respective category, such as a color grade. Since there are 25 official color grades in the USDA universal standards, a neural network should have 25 output neurons to differentiate these grades. However, the color grades of U.S. cotton heavily concentrate in the white and light-spotted categories. The sample set randomly selected for this research cannot equally represent all the color grades. There would be negative biases over less represented grades if all the grades were judged simultaneously. Therefore, a two-step approach was adopted in developing a neural network based classifier. This classifier consists of multiple neural networks that perform the classifications of the major and sub-color categories separately (Figure 4). The color data of a sample are first classified by an MLP to determine the main color category, and then sent to a separate MLP to determine the sample's subcategory within the identified main category. The classifier has one main MLP that has two inputs, two hidden layers, and five output neurons corresponding to the five main categories (1-white, 2-light spotted, …). The two hidden layers have six and twelve neurons, respectively. For each main category, there is a sub-MLP that may have three to eight output neurons depending on how many subcategories (1-good middling, 2-strict middling, …) are available in this main category. A sub-MLP also uses a two hidden-layer structure, with six neurons being on the first layer and 15 neurons on the second layer. Each output of the main MLP is also used to control a switch that permits the color data to be sent to the corresponding sub-MLP when it is turned on. After two categories are identified, a color grade is generated by placing two digits together.

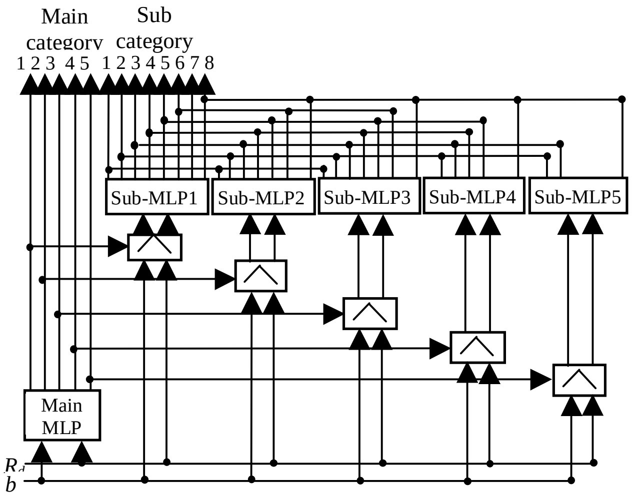

Figure 4. The Neural Network Classifier

The connecting weights w_{ij}^{a1} of two adjacent layers in each MLP were determined through a supervised training procedure called the *error back-propagation* algorithm [1, 2]. To make the classification results acceptable to the cotton industry, the training data used should be those obtained from the universal standards for Upland cotton. Unfortunately, the universal standards do not include the physical samples (biscuits) for all the color grades, whose colors can be measured by an instrument. We had to use the classer's color grades as targeted grades in the network training, since they are currently "official". The same sample set (2489 color data of 1996 crops) was used as the training set. 1385 more samples from 1997 U.S. corps were used as a test set to check the generalization performance of the classifier.

It was found that the NN classifier reduced the machine-classer disagreements from 54.08% to 16.35% for the training set. The NN-classer disagreement seems to have reached a minimal level (around 20%), because the classer's sustainability is generally 80%. Figure 5 presents the distributions of the NN-classer disagreements of the training samples. Compared with the distributions of HVI-classer disagreements in Figure 1, the disagreements are much smaller and more evenly spread across all the grades. That means that the NN classifier primarily reduces systematic disagreements with the classer. The NN-classer disagreements of the test samples decrease from 62.09% to 22.89%, which are consistent with those of the training samples. The result from the test set shows that the NN classifier provides a good generalization for new cotton color data.

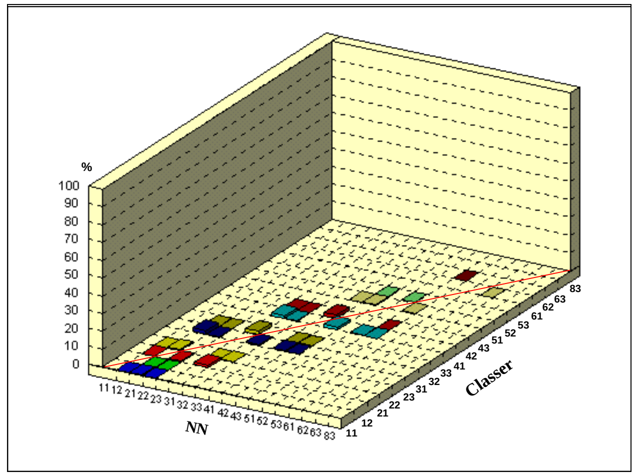

Figure 5 Distributions of NN-Classer Disagreement in Color Grades

Table III shows the NN-classer disagreements of both the training set and the test set (in parentheses) in the major color categories. The total disagreements in the major categories have decreased from 44.61% to 9.37% in the training set, and from 57.11% to 13.07% in the test set. As mentioned before, the HVI is more likely to grade light spotted cotton as "white" than to grade white cotton as "light spotted". The HVI-classer disagreements in these two opposite ways are seriously unbalanced. Although the NN-classer disagreements still occur mainly between "white" and "light spotted", the two-way disagreements are more comparable. For the test set, the NN classifier actually reversed the HVI's trend, putting more white samples (8.16%) into the light-spotted category than light-spotted samples (4.69%) into the white category. This may be due to over-training of the classifier. More testing data are needed to verify the generalization performance of the classifier.

Table III Disagreement (%) in the Major Categories

Classer / ANN	White	Light Spotted	Spotted	Tinged	Yellow Stained
White		5.4 (4.7)	0 (0)		
Light Spotted	3.9 (8.2)		0 (0.2)		
Spotted	0 (0)	0 (0)			
Tinged					
Yellow Stained					

The data in the parentheses are those of the test set.

Table IV Disagreement (%) in the Subcategories

Classer / ANN	GM	SM	M	SLM	LM	SGO
GM		0.2 (1.3)				
SM	0.4 (0.8)		2.8 (3.1)			
M		1.5 (1.5)		1.9 (2.5)		
SLM			1.0 (1.7)		0.3 (0.2)	
LM				0.2 (0.5)		0 (0)
SGO					0 (0)	

The data in the parentheses are those of the test set.
GM: good middling, SM: strict middling, M: middling, SLM: strict low middling, LM: low middling, SGO: strict good ordinary; GO: good ordinary.

Table IV shows the distributions of the disagreements in the subcategories. Compared with the HVI-classer disagreements in the subcategories (Table II), the NN-classer disagreements remain roughly at the same level (from 9.5% to 8.31% in the training set and from 11.13% to 11.68% in the test set). The classifier tends to give higher subcategory grades to the samples as does the HVI, but this tendency is less noticeable with the classifier than with the HVI. The disagreements primarily occur in the categories of strict middling, middling and strict low middling.

Conclusions

The HVI-classer disagreements in cotton color grades can be both systematic and random. Systematic disagreements mainly occur between the two major color categories, "white" and "light spotted", and take more 80% of the total disagreements. The disagreements among the subcategories are basically random and insignificant. To improve the acceptance of the machine grading for cotton color, efforts should be taken to reduce the systematic disagreements. The neural network classifier developed in this research proves very effective in improving machine-classer agreements in color grades, even if it inputs the same color information as the HVI does. The classifier uses separate neural networks to determine the major category and the subcategory of cotton color in a consecutive way. The classifier can be trained with any desirable data to gain the good generalization over new test data.

Acknowledgement

We are grateful for the financial support provided by Cotton Incorporated and Texas Technology Development and Transfer Program for this research. We would like to thank Mr. James Knowlton, AMS of USDA, Memphis, TN, for his valuable opinions on the research subject.

Literatures Cited

1. Bishop, C. M., "Neural Networks for Pattern Recognition", Clarendon Press, Oxford, U.K., 1995.
2. Lin, Chin-Teng and Lee, George C. S. "Neural Fuzzy Systems", Prentice Hall, NJ, 1996.
3. United States Department of Agriculture, "The Classification of Cotton," Agricultural Handbook 566, 1993.
4. Xu, B., Fang, C., Huang, R. and Watson, M.D., Chromatic Image Analysis of Cotton Trash and Color Measurements, *Textile Research Journal*, **67**, 881-890, 1997.
5. Xu, B., Fang, C., Huang, R. and Watson, M.D., Cotton Color Measurements by an Imaging Colorimeter, *Textile Research Journal*, **68**, 351-358, 1998.
6. Xu, B., Fang, C. and Watson, M.D., Investigating New Factors in Cotton Color Grading, *Textile Research Journal*, **68**, 779-787, 1998.