J. Cotton Res. Dev. **37**(1), 134-141 (January, 2023) **ISSN No. 0972-8619**

Developing AI model using auto machine learning platform for highest spinnable count index prediction from cotton fibre properties

N. SHANMUGAM*,GTV PRABHU, P. JAGAJANANTHA AND H. CHAURASIA
ICAR-Central Institute for Research on Cotton Technology, Mumbai 400 019
**Email : dr.shanmugam@gmail.com*

Abstract : Single integrated indexes were developed for quick evaluation and ranking of cottons. Central Institute for Research on Cotton Technology (CIRCOT) has developed Highest Spinnable count (HSC) index to rank cottons grown in India. HSC index is calculated after spinning cotton to 2 counts namely under spun and over spun based on CIRCOT CSP standards and evaluating its lea strength. Cotton fibre parameters namely UHML, Strength, micronaire, elongation and uniformity were used in the development of models for predicting highest spinnable count. Models were developed using regression method and AI based machine learning algorithms. Automl GBM algorithm was found to give lease root mean square error of 3.0 and highest coefficient of determination (R2) of 0.97 when compared with generalized linear regression and deep neural network. The percentage of error reduction varied from 62 to 85 per cent in comparison with multiple regression model depending up on error parameter used for evaluation of model. Important variables for determining the highest spinnable count of yarn is found to be UHML and fibre strength in both the models namely automl and regression.

Keyword: Cotton, high volume instrument (HVI), machine learning, model, upper half mean length (UHML)

Cotton fibre properties plays a key role in deciding the yarn quality produced. One instrument that is widely used throughout the world for measuring cotton fibre properties is High Volume Instrument (HVI). Cotton fibre properties that are reported by HVI are fourteen in number. Even though all the HVI properties have its application and relevance in mechanical processing of cotton but only limited number of properties are used in trading of cotton. The measurement and supply of too many parameters increases the complexity in making procurement decisions. In order to mitigate or reduce the complexity, single integrated indexes were developed for quick evaluation and ranking of cottons. Two popular indices, Fibre quality index (FQI) and Spinning Consistency Index (SCI) were mainly used in the bale management for choosing the bales for blending and mixing of cottons .FQI and SCI are calculated from fibre properties and does not convey any meaning except that higher the value of index better the cotton. In the FQI formula, Length, strength, fineness and maturity are given equal weightage. But incase of SCI, multiple regression is used with different coefficients for every fibre parameter. SCI uses colour values namely Reflectance (Rd) and yellowness (+b) apart from fibre length, strength, uniformity and micronaire. The coefficients of Rd and +b are positive coefficients used in the equation indicates that increase in these will increase the SCI value. However, in practice, one can see that when Rd value increases, there is a drop in +b. Apart from that higher +b, in majority of cases are associated with higher non-lint content like trash. CIRCOT has developed Highest Spinnable count index to rank cottons grown in India. HSC index is calculated after spinning cotton to 2 counts namely under spun and over spun based on CIRCOT Count Strength Product (CSP) standards and evaluating its lea strength. HSC index is used to rank cotton. HSC index gives a direct meaning unlike SCI or FQI, for example if one cotton is having a HSC of 83, it means that particular cotton is spinnable to 83s Ne and it will have a CIRCOT standard CSP value of 2603.

Statistical modelling of cotton fibre properties and yarn strength has been studied using discriminant analysis, measurement of alternatives and ranking according to compromise solution (MARCOS) method, multiple linear regression , stepwise regression and response surface regression. Apart from using statistical tools for modelling, textile researchers had applied Artificial Intelligence (AI) based models in predicting yarn properties from fibre properties. Artificial intelligence models that were applied in textile are: Fuzzy logic model, hierarchical deep learning neural network, Genetic Algorithm– , hybrid neural network, and artificial neural network . Highest standard count index was estimated using back propagation algorithm based neural network and this model is based on ICC mode data like 2.5 per cent span length. Presently HVI mode cotton fibre assessment has been implemented at CIRCOT from 2016 and textile mills were also switched over HVI mode data. Hence, there exists a need to develop models relating HVI mode cotton fibre properties and HSC.

Machine learning is the key technology in artificial intelligence. Today, machine learning is able to pick up knowledge from examples and it is able to code implicit knowledge. Simplifying training and tuning of machine learning models frees the data scientist to focus on data-preprocessing, feature engineering and model deployment. The success of machine learning in a broad range of applications has led to an ever-growing demand for machine learning systems. Auto machine learning systems available in the commercial and opensource platforms i nclude BigML.com, Wise.io, H_2O.ai, feedzai.com, RapidMiner.com, Prediction.io, DataRobot.com, Microsoft's Azure Machine Learning, Google's Cloud Machine Learning Engine, and Amazon Machine Learning. Among these machine learning platforms, H_2O AutoML produces high quality models that are suitable for deployment in an enterprise environment. H_2O AutoML supports supervised training of regression, binary classification and multi class classification models on tabular datasets. In this study H_2O automl plalform selected as it is an open source platform and this platform was used to develop models linking fibre properties and highest spinnable count index.

MATERIALS AND METHODS

ICAR-CIRCOT conducts fibrequality evaluation of cotton samples received from breeders and also performs spinning of agronomy experiment samples which were in the final stage of release for commercial cultivation. Data on the results of quality evaluation are presented annually in AICRP workshop and publishedas cotton technological report. Cotton fibre quality was assessed using HVI and data on Upper Half Mean Length (UHML), Uniformity Index (UI), Micronaire (Mic), Bundle strength (Str) and elongation (E). Agronomy trial samples were spun to two yarn counts one under spun and other over spun with reference to CIRCOT CSP standards in mechanical processing division of CIRCOT. Yarn count and lea strength were tested as perIS 1671: 1971 and were presented in AICRP reports. In this study, data for the period 2016-2017, 2017-2018, 2018-2019, 2019-2020, 2020-2021, 2021-2022 pertaining to agronomy trials were collected from AICRP cotton technological reports and used for the development of models. Highest spinnable count was calculated using a developed algorithm as detailed in an earlier research paper. Multiple regression equation was developed using back ward regression in the data analysis mode of SPSS software. In the regression, variable HSC was used as the independent and UHML, UI, Mic, Str, and E were used as dependent variables. Similarly, machine learning models were developed using Automl of H_2O platform. In the machine learning, fibre parameters are kept as input variable and HSC was used as response variables.

RESULTS AND DISCUSSION

Summary of data of fibre and yarn properties are given at Table 1. Total number of data used in the study are 244. UHML is the average length by number of the longer half (50 per cent) of the fibres distributed by weight. This study included a minimum UHML of 20.8 mm. It was observed that fibre length below this level was found to be un spinnable. Fibre length determines the settings of spinning machines. Longer fibres can be spun at higher processing speeds and allow for lower twist levels and increased yarn strength. Uniformity ratio is the ratio between the mean length and the UHML expressed as a percentage. Variations in fibre length can lead to an increase in waste, deterioration in processing performance and ultimately yarn quality. As per the SITRA norms , which was adopted from cotton incorporated, USA, UI value less than 77 is given a rating of very low and above 85 as very high. UI of cotton used in this modelling covered the entire spectrum of UI, that is very low to very high.

HVI micronairetest measures the resistance offered by a weighed plug of fibres in a chamber of fixed volume to a metered airflow. The unit of micronaire is µg/inchIf the micronaire is coarse, the number of fibres in the yarn cross section will be less. This always results in lower strength and lower elongation for a given yarn count assuming maturity and strength of fibre being equal. It can be seen at Table 1 that a maximum count of 120s Ne was spun from a cotton having a minimum micronaire of 2.3.

The strength of cotton fibres is usually defined as the breaking force required for a bundle of fibres of a given weight and fineness. The ability of cotton to withstand tensile force is fundamentally important in spinning. It is well known that yarn and fabric strength correlates with fibre strength. Cotton fibre strength used in this study were in the range of 20.3 g/tex to 41.2 g/tex. It is to note here that fibre parameters used in this study were measured in HVI mode. Fibre strength evaluated using HVI mode will have a value of 1.25 times fibre strength that was measured using ICC mode. In addition to using fibre strength as input variable, two additional parameters *viz*. SL ratio and SLSTR. Integration

Table 1. Summary of cotton fibre and yarn properties

Parameters	Minimum	Maximum	**Mean**
Upper half mean length (mm)	20.8	36.9	**27.9**
Uniformity index	77.0	89.0	**83.2**
Micronaire (µg/inch)	2.7	5.9	**4.2**
Bundle strength (g/tex)	20.3	41.2	**28.0**
Elongation (%)	3.7	6.9	**5.7**
Yarn count_1 (Ne)	12.0	95.7	**35.5**
Count strength product_1 (CSP)	1529	2915	**2260**
Yarn count_2 (Ne)	16.0	120.0	**46.6**
Coutn strength product_2 (CSP)	1400	2608	**2040**
Highest spinnable count index (Ne)	2.3	91.1	**38.8**

Table 2. Correlation between cotton fibre parameters and HSC

Parameter	UHML	UI	Mic	Str	E	HSC
UHML	1	.712**	-.393**	.753**	.199**	.815**
UI	.712**	1	-.087	.648**	-.004	.592**
Mic	-.393**	-.087	1	-.169**	-.307**	-.534**
Str	.753**	.648**	-.169**	1	.050	.718**
E	.199**	-.004	-.307**	.050	1	.230**
HSC	.815**	.592**	-.534**	.718**	.230**	1

**. Correlation is significant at the 0.01 level (2-tailed).

Table 3. Regression models summary

Model	R	R Square	Adjusted Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.876a	.767	.762	7.8793	.767	156.904	5	238	.000
2	.875b	.766	.762	7.8891	-.002	1.599	1	238	.207
3	.874c	.763	.760	7.9141	-.002	2.519	1	239	.114

a. Predictors: (Constant), E, UI, Mic, Str, UHML, b. Predictors: (Constant), UI, Mic, Str, UHML, c. Predictors: (Constant), Mic, Str, UHML, d. Dependent Variable: HSC

parameter of strength and length were calculated by dividing fibre length by strength and called it as SLratio. Similarly another parameter namely SLSTR is calculated by multiplying SL and STR. When using ICC mode data, it was advised to cotton breeders to achieve a SLratio of 0.8. But however, in HVI mode, this ratio becomes 1.0 as strength is 1.25 times that of ICC mode strength. It can be seen at Table 1 that strength value starts from a minimum value of 0.8 and goes up to 1.2. Hence when using HVI mode data for breeding purposes, one should aim for a minimum slratio of 1.0.

HVI measure the elongation-to-break while performing the bundle-strength testing. The increase in length or deformation of the fibre before it breaks as a result of stretching is called elongation. Expressed as a per cent increase over its original length. Cotton normally hasa elongation per cent in the range of 6-9 per cent. In this study elongation found to vary from 3.7 to 6.9. One issue with fibre elongation measurement in HVI is its calibration for elongation and its reproducibility . It was reported that HVI values on elongation are inconsistent.

HSC is expected to lie in between C1 and C2 as C1 is under spun and C2 is over spun. In some cases, during spinning, particularly in short staple cottons, the minimum count spun was 12s Ne even though it did not able to meet the standard CSP norms of CIRCOT. In those cases, the HSC will be lower than the under spun count. Similarly in case of overspun, if second count is not spun above the standard csp of CIRCOT, HSC will go beyond C2.Minimum underspun count (C1) was found to be 12.0 whereas HSC is found to be 2.3. This reveals the information that some cotton that are spun to higher count did not meet the required CSP standards.

Correlation coefficient matrix is given at Table 2. HSC is having signification correlation with all the input variables namely UHML, UI, Mic, str and E. Highest significant correlation coefficient of0.815 is observed for UHML indicating that it is the most important variable in the prediction of HSC. Higher fibre length is associated with finer count. Except micronaire, all the other fibre parameters are positively correlated with HSC. Micronaire being indication of fibre linear density, higher micronaire will result in less number of fibres in yarn cross section and hence higher micronaire cotton may not offer required number of fibres per yarn cross section for optimum spinning. UHML is found to have significant correlation with all the variables. Elongation found to be unrelated with UI and strength showing non-significant correlation coefficient.

Backward regression performed on the data resulted in 3 models (Table 3). Model 1 includes all five input variables, model 2with four input variables except elongation and model 3 with three variables except elongation and UI. A backward elimination rule starts with all possible explanatory variables and then discards the least statistically significant variables, one by one . [Smith 2018] The discarding stops when each variable remaining in the equation is statistically significant. Even though all the models are significant, the model with all five variables (Model 1) are selected for studying the effect of individual variables on HSC.

$$HSC = (2.095 \times UHML) + (0.67 \times UI) + (-7.686 \times Mic) + (1.777 \times str) + (1.175 \times E) - 99.829 \quad \text{-----Model 1}$$

Percentage relative contribution of each variable towards HSC was studied using the formula given as equation 1: Where, Bi is the standardized coefficient of ith value (i = 1,2,3……k) and R2 is the coefficient of determination.

$$100\left(\frac{B_i}{\sum_1^k B_i}\right)R^2 \text{---------------------Eq.1}$$

It is observed that the contribution of UHML is the maximum (33.9%) followed by Str at 28 per cent and Mic at 27 per cent (Fig. 1). The least influencing variables are UI and elongation. The regression model is able to explain 76 per cent of the variation of in HSC.

In order to compare with effect of interaction parameters of SL and str, 2 integrated properties namely SLstr ratio and SLStr multiplication factor also studied using them as input variables. It was observed that use of SL ratio dropped the R square value by 22 per cent from 0.76 to 0.59. On the contrary, use SLSTR integration parameter gave the coefficient of determination value (R2) equivalent to that of that of all input variable model that is 0.76.

H2O AutoML was used to develop models by keeping all the five cotton fibre parameters as input and HSC as output. Automltrains and cross-validates the algorithms namely three pre-specified XGBoost Gradient Boosting Machine (GBM) models, a fixed grid of Generalized Linear Machine learning (GLM), a default Random Forest (DRF), five pre-specified H_2O GBMs, a near-default Deep Neural Net, an Extremely Randomized Forest (XRT), a random grid of XGBoost GBMs, a random grid of H_2O GBMs, and a random grid of Deep Neural Nets [H_2O Automl 2022]. AutoML performs a hyperparameter search for all these algorithms in order to deliver the best model. The trained models are ranked by a default sort metric of Room Mean Square Error (RMSE). Among the models developed, the model ID: GBM_5_AutoML is found to have the least RMSE. This model is based on forward learning ensemble method. The developed model was used to predict HSC values of all data of 244 Nos. Error analysis was carried out to judge the performance of model. Here all the data were used for both training and validation. Normally one uses 75:25 or 80:20 ratio of train/validation to get generalization in the model. However, in this case, it was not done as H_2O automl cross validates the prediction using sample group of data while training. In addition to that as regression was carried out using complete data, it was decided to use complete data for automl training to enable for comparison.

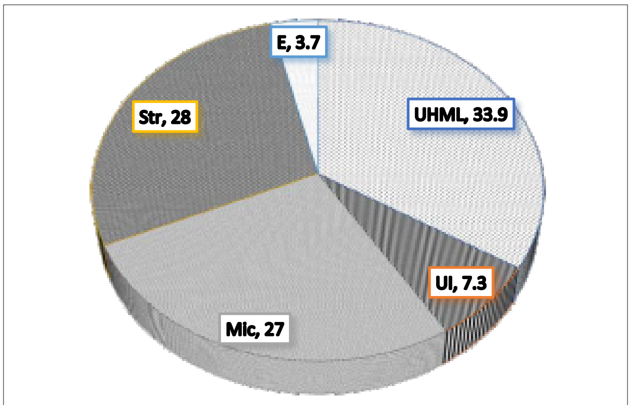

Fig. 1. Importance of difference fibre properties on HSC

Error analysis of regression and automl GBM model results are given at Table 7. The R2 value represents the degree that the predicted value and the actual value move in unison. The R2 value varies between 0 and 1 where 0 represents no correlation between the predicted and actual value and 1 represents complete correlation. Compared to regression model, R squared has improved from 0.77 to 0.97 between predicted and actual HSC values.

The Mean Absolute Error (MAE) is an average of the absolute errors. The MAE units are the same as the predicted target, which is useful for understanding whether the size of the error is of concern or not. The smaller the MAE the better the model's performance. In

Table 4. Error analysis of regression and Auto Ml model

Parameter	Model		Reduction (%)
	Regression	AutoMl GBM	
Coefficient of determination	0.77	0.97	26
Mean Absolute Error	6.1	2.3	62
Mean Square Error	61.6	9.0	85
Root Mean Square Error	7.9	3.0	62
Mean Absolute Percentage Error	20.0	7.5	62

autoML GBM model, MAE was reduced by 62 per cent while predicting HSC.

The Mean Square Error (MSE) metric measures the average of the squares of the errors or deviations. MSE takes the distances from the points to the regression line and squaring them to remove any negative signs. MSE incorporates both the variance and the bias of the predictor. Among the model metrics studied, largest drop is occurred in MSE *i.e.* 85 per cent.

The RMSE metric evaluates how well a model can predict a continuous value. The RMSE units are the same as the predicted target, which is useful for understanding if the size of the error is of concern or not. The smaller the RMSE, the better the model's performance. RMSE has come down by 4.9 units while prediction was done using auotml model.

One additional goodness of fit measure is the Mean Absolute Percentage Error (MAPE). Absolute percentage error is a measure of how much a target series varies from its model predicted level, expressed as a percentage value. MAPE of automl GBM model is observed to have a value of 7.5 which is 62 per cent lower than that of regression model.

Prediction versus actual values of HSC are depicted at Fig 2 and 3. It is observed that in both the models the values are almost distributed equally in the 2 sides of the line. Fig 3 it can be seen that the predicted values are closer to line, even for fine counts. Automl GBM model determines the fibre parameter importance by calculating the relative influence of each variable. Importance of variable is measured, based on how much squared error decreased due to the inclusion of that particular variable. According to Automl GBM model, the strength is the most important parameter followed by UHML, mic, E and UI. The ranking is similar to regression where in first 2 important variables are UHML and str. The least important variables are E and UI identical to regression model.

CONCLUSION

Cotton fibre parameters namely UHML, Strength, Micronaire, elongation and uniformity

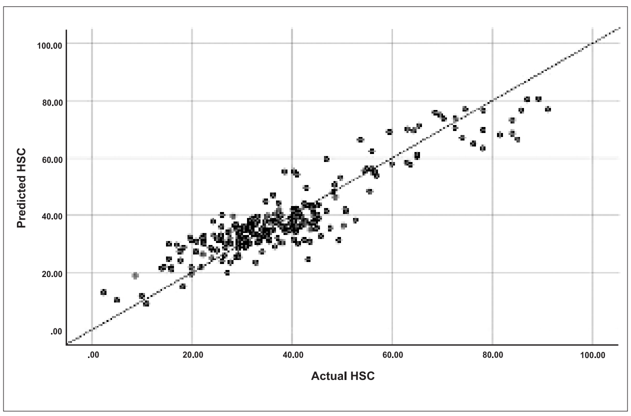

Fig. 3. AutoMl GBM model – Actual vs predicted HSC

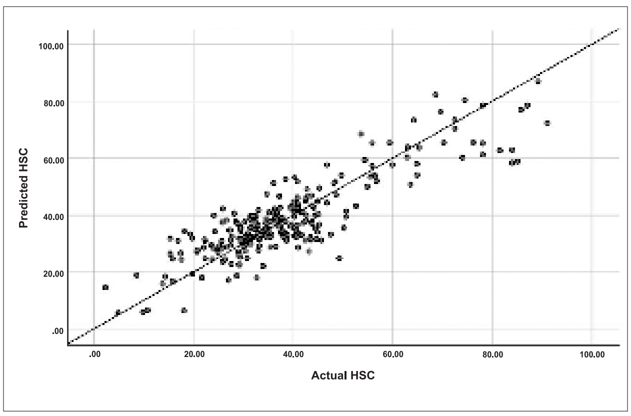

Fig. 2. Regression model – Actual vs predicted HSC

were used in the development of models for predicting highest spinnable count. Models were developed using regression method and AI based machine learning algorithms. Automl GBM algorithm found to give lease root mean square error of 3.0 and highest coefficient of determination (R2) of 0.97 when compared with generalized linear regression and deep neural network. The percentage of error reduction varied from 62 to 85 per cent depending up on error parameter used for evaluation of model. Important variables for determining the highest spinnable count of yarn is found to be UHML and strength in both the automl GBM and regression model. The least important variable in the prediction of HSC is cotton fibre elongation and uniformity index in both the models.

REFERENCES

Ahmad, G.G. 2016. Using artificial neural networks with graphical user interface to predict the strength of carded cotton yarns. *J. Text. Inst.*, **107**: 386–94

Ghanmi, H., Ghith, A. and Benameur, T. 2015. Response Surface Regression Models for Prediction of Ring Spun Yarn Properties. *Res. J. Text. Appar.*, **19**: 1–10

Ghanmi, H., Ghith, A. and Benameur, T. 2021. Ring spun yarn quality prediction using hybrid neural networks. *J. Text. Inst.*, 1-9 https://doi.org/10.1080/00405000.2021.2022826

H2O, Automl. 2022. AutoML: Automatic Machine Learning — H2O 3.38.0.1 documentation. https://docs.h2o.ai/h2o/latest-stable/h2o-docs/automl.html (Accessed September 22, 2022)

Hassan, A.A. and Ibrahim, A.E. 2018. Using some fiber properties to predict strength and regularity of carded yarn for some Egyptian cotton varieties. *Middle East J. Appl. Sci.*, **8**: 1256-63

Hutter, F., Kotthoff, L. and Vanschoren, J. (eds). 2019. Automated Machine Learning: Methods, Systems, Challenges. Springer International Publishing:Cham http://link.springer.com/10.1007/978-3-030-05318-5 (Accessed September 17, 2022)

ICAC, ITMF. 2020. Interpretation and Use of Instrument Measured Cotton Characteristics. https://csitc.org/sitecontent//instrument_testing/public_documents_it/0_2020-ITMF_ICCTM%20&%20ICAC_CSITC_Task_Force-INTERPRETATION_Guide-V1.0-English.pdf

LeDell, E. and Poirier, S. 2020. H2O AutoML: Scalable Automatic Machine Learning. In H2O AutoML: Scalable Automatic Machine Learning,. 1–16

Mitra, A. 2022. Cotton fibre selection based on quality value using measurement of alternatives and ranking according to compromise solution (MARCOS) method. *Res. J. Text. Appar.*, ahead-of-print. https://doi.org/10.1108/RJTA-03-2022-0030 (Accessed August 12, 2022)

Öner, E. Topcuoglu, S. and Kutlu, O. 2018. The effect of cotton fibre characteristic on yarn properties. *IOP Conf. Ser. Mater. Sci. Eng.*, **459**: 012057

Paknejad, S.H., Vadood, M., Soltani, P. and Ghane, M. 2021. Modeling the sound absorption behavior of carpets using artificial intelligence. *J. Text. Inst.*, **112**: 1763–71

Saha, S. 2021. Hierarchical Deep Learning Neural Network (HiDeNN): An artificial intelligence (AI) framework for computational science and engineering. *Comput. Methods Appl. Mech. Eng.*, **373**: 113452

Sarker, B. and Chakraborty, S. 2022. Discriminant analysis-based modeling of cotton fiber and yarn properties. *Res. J. Text. Appar.*, **26**: 18–40

Shanmugam, N. and Doke, S. 2005. Highest standard count estimation from fibre parameters using neural network techniques. *Indian J. Fibre Text. Res.*, **30**: 302–08

Smith G. 2018. Step away from stepwise. *J. Big Data*, **5**: 32

Vasudevan, P., Jaraman, D., Sreenivasan, J. and Pasupathy, R. 2019. *SITRA norms for spinning mills*. 9th ed. The South Indian Textile Research Association: Comibatore

Visalakshi, V., Yogalakshmi, T. and Castillo, O. 2022. Analysis of Cotton Yarn Count by Fuzzy Logic Model. In Proceedings of International Conference on Computational Intelligence, Tiwari R, Mishra A, Yadav N, Pavone M. (eds). Springer Singapore: Singapore; 349–361 https://link.springer.com/10.1007/978-981-16-3802-2_29 (Accessed April 16, 2022)

Received for publication : December 12, 2022
Accepted for publication : December 26, 2022