Machine Learning Approach for Prediction of Crimp in Cotton Woven Fabrics

Muhammad Zohaib FAZAL, Sharifullah KHAN*, Muhammad Azeem ABBAS, Yasir NAWAB, Shahzad YOUNIS

Abstract: The interlacements of yarns in woven fabrics cause the yarn to follow a wavy path that produces crimp. Off-loom width of the fabric is determined by the percentage of the induced crimp. Therefore, the final width of the fabric will be less or surplus than required if crimp percentage is not precisely measured. Both excessive or recessive fabric width is unwanted and leads to huge loss of cost (profit), manufacturing time, energy (electricity) and ultimately loss of competition. Crimp percentage in yarns is determined by physically measuring the extra yarn length or by predicting it based on fabric structural parameters. Existing methods are mainly post-production, time and resource intensive that require specialized skills and tangible fabric samples. The proposed framework applies supervised machine learning for crimp prediction to cater for the limitations of the existing techniques. The framework has been cross-validated and has prediction accuracy (R^2) of 0.86 and 0.79 for warp and weft yarn crimp respectively. It has prediction accuracy (R^2) for warp and weft yarns crimp of 0.99 and 0.81 respectively for the unseen industrial dataset. The proposed prediction model shows better performance when compared with an existing standard system.

Keywords: cotton woven fabric; crimp prediction; machine learning; pre-production prediction

1 INTRODUCTION

Woven fabrics are produced by the interlacements of warp and weft yarns. These interlacements of yarns cause the yarns to follow a wavy path in fabric. Extra yarn length is required to overcome these undulations or waviness. The waviness or contraction in the yarns is called crimp [1] which leads to fabric shrinkage when it is removed from the loom after weaving (Off-loom fabric).

Off-loom width of the fabric is determined by the percentage of crimp present in the weft yarn. Therefore, the final width of the fabric will be less or surplus than required if crimp percentage is not precisely measured. Both excessive or recessive fabric width is unwanted and leads to huge loss of cost (profit), manufacturing time, energy (electricity) and ultimately loss of competition [2]. In addition to fabric width, several other mechanical properties like weight, thickness, cover factor, drape, flexibility, stiffness, bending, air permeability, impact loading, shear and hygral expansion are influenced by the yarn crimp. At the point of contacts of yarns, crimp interchange is affected due to interlacements or weave design. The difference between the crimp percentage values of warp and weft yarns influences the strength, elongation and energy absorption levels of woven fabrics differently in both warp and weft directions [3]. All these properties are important while designing fabrics for high end technical applications like heat and radiation protected firefighter clothing, molten metal protection for welders, stab protection and bulletproof vests, and spacesuits [4].

The yarn crimp is highly dependent on the variations in structural parameters of the fabric such as warp and weft yarn linear densities, weave design, reed count and several others. Among structural parameters the weave design that causes crimp is characterized by its float length i.e. the number of yarns passing without interlacement over the other set of yarns. Moreover, the selection of raw material (i.e. fibre) and the yarn spinning techniques (carded or combed) directs the physical and mechanical properties of the fabric [5]. In addition, several factors related with weaving machine i.e. warp yarn tension, heald-crossing time, the position of backrest, cloth support, and shed geometry also induce their effects on the fabric [6]. For example, a high warp yarn tension or warp linear density will result in a low crimp percentage in the warp but on the other hand, causes a higher crimp percentage in the weft.

At the present, the textile industry uses standardized measurement method to measure the crimp percentage from the weaved fabric. ASTM Standard test method D3883 [7, 8] and its variations are the most widely used methods in the textile industry for measuring crimp from the woven fabric. These measurement methods are mainly post-production, time and resource intensive that require specialized skills and tangible fabric samples. However, researchers have proposed methods for pre-production prediction of the yarn crimp using recorded data of the fabric structural parameters [9]. Prediction is made through statistical methods [2, 10] or using neural networks [11, 12]. An existing prediction model [2] based on linear regression is unable to handle fabric structural parameters that are inter-dependent and multi-correlated. For predicting the yarn crimp percentage, existing techniques have used small sized samples that do not reflect the true variations of the fabric parameters.

The present work proposes a technique for the prediction of yarn crimp in cotton fabric using a machine learning approach. In machine learning, classifier algorithms make predictions using a precisely built mathematical model of sample data called the training set [13-16]. Accuracy of the prediction is highly dependent on the quality of the training set. The present work presents modelling of the training set using data extracted from thousands of quality sheets of real fabric samples produced by a leading textile industry of Pakistan. Every quality sheet contains fabric parameters of the produced fabric and its crimp percentage (manual and standardized measurement). For the proposed training set, the fabric parameters (carefully selected) and the crimp percentage from the quality sheets are modelled as input vector (predictors) and the target (or the class label) respectively [17].

The rest of the paper is organized as follows. Section 2 provides a review of existing crimp prediction methods. Section 3 presents the methodology of the proposed crimp prediction method followed by the validation of the proposal using standard and comparative techniques. The last section concludes the present work.

2 RELATED WORK

The earlier attempts [20] were made to predict crimp from the geometry of yarns in the fabric. Geometrically, crimp percentage in the woven fabric is the percentage of extra yarn length over the fabric length. Mathematical models are incapable of explaining the exact geometrical forms of fibre, yarns, and fabrics. Due to several physical changes, yarns and fabrics change their behaviour and results deviate from the proposed mathematical model [20].

With the advancements in prediction tools, statistical (regression) and neural networks were used to predict the crimp percentage with reasonable high accuracy but on a small sized dataset. The dataset used by Lin [18] contain only 26 samples of cotton yarn. All the samples are with float length 1 and single yarn linear density only. Half of the samples were used as training and the rest for the test. The artificial neural network (ANN) model of the Lin [18] was constructed with an input layer having three nodes (Cover factor of fabric, warp and weft), the hidden layers having sixteen nodes and an output layer with two nodes for warp and weft crimp. Their model has an R^2 value of 85.09% for warp crimp and 87.23% for the weft. Training a neural network is a time-consuming job. Authors have [11] attempted to predict the crimp percentages in warp and weft yarns of polyester fabric. This model was based on 56 fabric samples. Yarn Linear density, PPI, width, float length, shed closing time and loom speed were used as the input predictors to train the model. ASTM Standard D3883 was used to validate the results.

Table 1 Summary and comparison of previous methods and techniques of crimp measurement and prediction

Related work	Measurement (M) / Prediction (P)	Technique	Sample Size	Raw Material	Float Length	Validation Method
ASTM Standard D3883 [7]	M	Manual Way	-	-	-	-
Peirce (1937)	P	Geometry of yarn in fabric	-	-	-	-
Lin [18]	P	Neural Networks	26	Cotton	1	Holdout Method (13 Traning - 13 test)
Maqsood et al. [2]	P	Linear Regression	68	Cotton	1, 2, 3, 4	Holdout Method (60 Traning - 8 test)
Shahabi et al. [19]	M	Tensile Strength Testing Machine (Instron) with Image Analysis	9	45:55 Wool-polyester blend	2, 3	ASTM Standard D3883
Mertova et al. [20]	P	Tensile Strength Testing Machine (Instron)	30	Polyester	1	ASTM Standard D3883, Image Analysis
Kaplan et al. [21]	M	Image Analysis	12	Cotton	4	ASTM Standard D3883
Malik et al. [11]	P	Neural Networks	52	Polyester	1, 2	ASTM Standard D3883

A statistical model was developed based on 68 samples that were manufactured in a controlled environment [2]. Among those, sixty samples were used for training purpose and eight samples for validation/testing. Cotton yarn with float length varying from 1-4 was used. A linear regression equation was developed for both warp and weft yarn crimps separately with R^2 values of 71.81% and 67.05% respectively. MINITAB statistical software was used for the analysis. Significant factors identified for warp crimp were weft count, PPI, width, float length and reed count while for weft crimp they were warp count, weft count, float length, reed count, and reed denting.

A summary and comparison of previous prediction models and measurement techniques are shown in Tab. 1. It has been observed from the resultant comparison that fabric swatch is physically required for crimp measurement. The physical measurement methods are laborious, time-consuming and involve costly equipment (e.g., tensile strength testing machine, Instron & Charged Coupled Device (CCD) cameras) [17]. The calculated results deviate from the proposed mathematical model due to variations in the geometrical structure. Prediction inaccuracy prevails in results because existing prediction models are trained and tested on a small and controlled dataset. The linear regression model has been used in crimp prediction while the predictors are multiple correlated dependent variables. Neural network models are trained on a very small dataset while neural networks perform well when they have been trained on a large dataset. Due to that, these models are unable to handle all the variations that arise in real time. These models perform very well for the dataset on which they are trained but perform below par for the unseen data.

3 MATERIALS AND METHODS

The following section presents the details of fabric data used and the predictor selection method for the construction of a training model that predicts the crimp percentage. Accuracy of the prediction is highly dependent on the selection of appropriate predictors. The present work follows a systematic approach for the selection of predictors that has yielded a high accuracy. Fig. 1 shows the sequence of steps (represented as rectangular blocks) that were carried out for the predictor selection process. At each step, several classifiers were executed to determine the significance of the underlying process towards predicted accuracy. For the present case, the predicted accuracy is the coefficient of determination (R^2) i.e. the percentage of variance in the dependent variable (crimp percentage) that is predictable from the independent variable (predictors). Solid arrows in Fig. 1 represent that the modified predictor or a set of predictors after applying a process will be selected to the next process only if the applied process results in improved R^2 value. Otherwise, the original pre-processed data will be used (dashed line arrow in Fig. 1). The same procedure has been followed for both wrap and weft crimp prediction. Brief description of each step is presented in the following subsections. Later, the present section discusses the effect of individual processes in terms of crimp prediction.
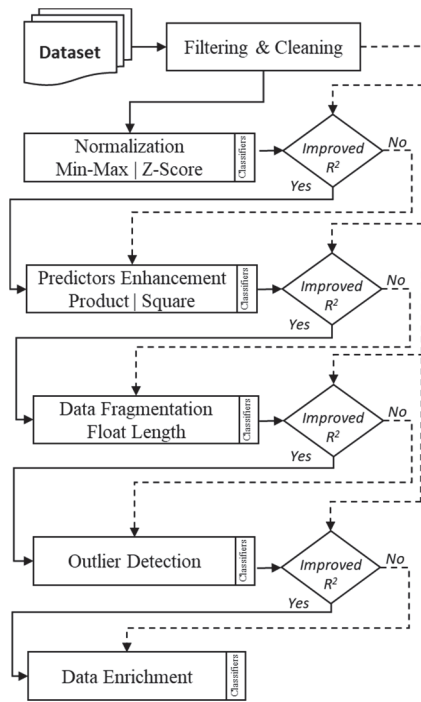
Figure 1 A systematic predictor selection method

3.1 Dataset and Pre-processing

Fabric data has been collected from a textile factory located in Faisalabad, Pakistan. The factory is heavily involved in export and domestic quality fabric weaving. All fabrics are produced as per customer requirements and fabric domain of use. A quality sheet containing fabric parameters i.e. yarn linear densities, yarn densities, grams per square meter (GSM) etc. is maintained for record keeping with every fabric produced by the factory. Traditionally, the textile industry uses these quality sheets as an instrument for cost estimations of future orders. Predicting accurate crimp percentage is an important aspect of the profitable cost estimation. The collected quality sheets contain data of 5428 samples of different fabrics produced by the factory. The parameters of the quality sheet are shown in Tab. 2. These parameters are either categorial type or continuous having defined categories and ranges of values respectively.

Table 2 Fabric parameters recoded in the quality sheets

Sr.	Parameter	Type of Data	Values
1	Warp Yarn linear density	Categorical	7, 10, 12, 14, 16, 20, 24, 30
2	Weft Yarn linear density	Categorical	7, 10, 12, 14, 16, 18, 20, 22, 24, 30
3	Warp density (Ends per inch, EPI)	Continuous	38-144
4	Weft density (Picks per inch, PPI)	Continuous	25-102
5	Weave Design (Float Length)	Categorical	1, 2, 3, 4
6	GSM	Continuous	81-386
7	Reed Count	Continuous	16.25-49.5
8	Reed Denting (Ends per dent)	Categorical	1, 2, 3, 4, 5

The present study focuses on fabric made up of carded cotton yarn on air-jet loom. The whole dataset contains 1390 cotton fabric samples. This raw data has been cleaned for missing values like empty spaces, alphabetical characters, anomalies etc. After exclusion of these records, the final sample size contains 1273 fabric samples.

3.2 Normalization

Normalization prepares data for the prediction model by bringing values of the predictors at a uniform scale or range. With normalized data, classifiers give equal weight to all predictors while constructing the prediction model [22]. However, not all predictors require normalization. It is applied to only those predictors that have different ranges of values. Mix-Max and Z-score normalization techniques were used separately for the normalization of selected data. Min-Max Normalization retains the relationship among original data whereas Z-score normalization performs normalization based on the mean and standard deviation of the predictors. After each normalization the normalized data is tested for its crimp prediction accuracy.

3.3 Predictors Enhancement

The prediction accuracy can be enhanced by increasing the number of predictors. Original predictors can be enhanced by creating new predictors using different mathematical techniques. The enhancement can be performed either through experts using domain knowledge or by mathematical expressions. For the present case, two mathematical expressions i.e. product and square are used separately to develop new predictors. The number of predictors has increased from eight to thirty-six by using product techniques whereas the number of predictors was doubled through square enhancement technique.

3.4 Data Fragmentation

Our dataset comprises fabric samples with four different float lengths. Range of crimp percentage induced in the yarns is different for each float length. To reduce the variations within the dataset, Data is divided into four fragments based on float length values as shown in Tab. 3.

Table 3 Data Fragmentation and number of instances in each fragment

Data Fragments	Float length	Initial No. of Instances	After Outliers Removal	After Data Enrichment
Fragment 1	1	390	377	658
Fragment 2	2	355	324	685
Fragment 3	3	472	461	652
Fragment 4	4	56	37	203

Fragmenting the data will improve the prediction accuracy of the model, as data within the fragments is more coherent and homogeneous. Traditionally, woven fabrics having the same float length possess the crimp percentage of warp and weft yarns in the same range.

3.5 Outliers Detection

Outliers are the anomalies in the dataset. Removal of the outliers results in improved prediction accuracy [23]. Basic data descriptions (measures of central tendency and

measures of dispersion) and statistical visualization (histograms and scatter plots) are outlier detection methods that provide valuable insight into the overall behaviour of data. The outlier detection methods (data description and visualization) used in the present work do not account the multivariate nature of the data as all predictors were independently examined. Moreover, another technique used for the detection of outliers is the use of a linear regression model. The linear regression model takes into account the multivariate nature of the dataset. Responses are dependent on predictors and a linear regression model is built on this assumption. Observations that show a large deviation from the fitted model were considered as outliers.

3.6 Data Enrichment

The number of samples in each fragment-set got reduced with the removal of noise and outliers. Prediction models considerably work well with a large number of instances in the dataset. To further enhance the dataset, a statistical technique considering mean and standard deviation of the continuous variables (predictors) was used. More instances are generated with the provided mean and standard deviation. A certain noise is added to avoid any potential duplication of instances.

Table 4 Warp and weft yarn crimp percentages (response variables)

Sr.	Response	Type of Data	Range
1	Warp crimp percentage	Continuous	6-19.80
2	Weft crimp percentage	Continuous	1-11.5

3.7 Prediction Models

The response variables are the warp and weft yarn crimp percentages that are of the continuous variable type. The range of both variables for cotton fabric is shown in Tab. 4. For the present case, the prediction model is a regression predictive model that performs the task of approximating a mapping function from input variables (predictors) to a continuous output variable (response or the class label). Every prediction model is provided with a set of N training instances of the form $[(f_1, y_1), …, (f_n, y_n)]$ where f_i is the predictor vector of the ith training instance and y_i is its labelled crimp percentage. The prediction model constructs a function $Y(f_i)$ that best fits the data. The loss function evaluates how well any prediction models have fitted the given training data. The loss function is minimized to increase the accuracy of the prediction model.

Present work is focused to use state of the art machine learning algorithms like Ridge regression, Kernel tricks (Polynomial and Radial Based Function (RBF)), K-nearest neighbours (KNN), Decision tree (Random Forest), Ensemble method (AdaBoost) and Deep learning (Neural Networks). Anaconda 64-bit windows python 3.6.3 data science platform is used to apply machine learning techniques on our data to build models. PyCharm community edition which is a freeware integrated development environment (IDE) is used for the application of selected techniques. Scikit-learn, numpy, and pandas libraries of python are used.

4 VALIDATION

The validation of the crimp prediction through constructed training model was performed using i) k-fold cross-validation technique, ii) ASTM Standard D3883 method and iii) comparative analysis with state-of-the-art crimp prediction technique.

4.1 K-fold Cross-Validation

Evaluation of prediction made through limited data is an on-going research area in machine learning and data mining. Many different techniques were proposed (holdout method, random subsampling, cross-validation, bootstrapping, etc.), among those the cross-validation is gaining popularity and is considered by the researchers as a valid evaluation method for situations with small or medium scale data size. The k-fold cross-validation is a regression method in which training data is partitioned into a fixed number of splits or folds. For every iteration, one-fold of the data is taken as a test while the rest is used for the training. The procedure continues for the total number of folds so that finally every instance has been tested at least one time. The debate is present in literature about setting the right number of folds i.e. the value of K. However, 10-fold cross-validation is the most widely used method for evaluating the performance of the prediction model. In the 10-fold cross-validation, the training data is split into ten equal parts. Each part is held out in an iteration as a test and the remaining nine-tenths are used for training. The training/learning procedure continues for ten times and finally, the performance measures are averaged to get the result.

The 10-fold cross-validation results are used to measure the significance of a process that has been applied to the data during the construction of the training model. The 10-fold cross-validation results (R^2 value) for wrap and weft crimp prediction are shown in Tabs. 5 and 6 respectively for every process. As discussed, the predictor or a set of predictors are selected for the next process only if they show improved R^2 value.

Wrap crimp prediction: The prediction accuracy (R^2 value) of the wrap yarn crimp made by classifier algorithms after every process being applied to the data is shown in Tab. 5. Initially, after pre-processing (filtering and cleaning) of the data - among classifiers - Kernel ridge (RBF) has resulted in the 0.70 R^2 value. However, random forest (RF) and neural network (NN) resulted in the same 0.69. The negative R^2 value is produced by the Kernel ridge (poly) classifier i.e. −0.5. After filtering and cleaning, the data was normalized using min-max and z-score methods. The results of 10-fold cross-validation for both normalization techniques shows no improvement in the R^2 value for all classifiers. As proposed by the research framework, normalized data will be carried for further processing only if it helps in improving the prediction accuracy of the models. However, normalization shows no significant improvement and as a result filtered and cleaned data is used for the predictor's enhancement process. Even with the predictor enhancement, none of the classifiers has shown improvement in R^2 value. Later, during the data fragmentation process, the whole data is split based on float length value (F-1, F-2, F-3 and F-4). A significant

improvement has been shown by all classifiers for only the fragment with float length value equal to 3 (F-3). Since improvement is observed after applying the data

fragmentation process, the fragmented data has been carried to the next process i.e. outlier detection.

Table 5 Summary of prediction accuracies of R^2 values for warp crimp prediction models

| Model | Filtering & Cleaning | Normalization | | Predictors Enhancement | | Data Fragmentation | | | | Outlier Detection | | | | Data Enrichment | | | | | |
		Min-Max	Z-score	Product	Square	F1	F2	F3	F4	F-1	F-2	F-3	F-4	F1FE	F2OE	F3OE	F4FE	Combine Dataset without FL	Combine Dataset with FL
Linear Regression	0.63	0.64	0.64	0.65	0.63	0.58	0.59	0.75	0.21	0.56	0.64	0.76	−0.1	0.69	0.71	0.82	0.74	0.39	0.68
Ridge Regression	0.64	0.64	0.64	0.66	0.63	0.58	0.6	0.75	0.22	0.56	0.64	0.76	−0	0.69	0.71	0.82	0.7	0.38	0.68
KNN	0.66	0.66	0.66	0.66	0.67	0.64	0.61	0.69	0.16	0.57	0.67	0.7	0.2	0.73	0.8	0.9	0.67	0.8	0.85
Kernel Ridge (poly)	−0.5	−0.5	−0.5	0.68	0.67	0.62	−1.4	0.75	0.36	0.58	0.63	0.76	0.24	0.77	0.79	0.88	0.71	0.54	0.78
Kernel Ridge (RBF)	0.7	0.7	0.7	0.67	0.64	0.65	0.55	0.76	0.12	0.62	0.66	0.76	−0.1	0.8	0.72	0.86	0.7	0.4	0.72
Random Forest	0.69	0.69	0.69	0.7	0.7	0.67	0.61	0.72	0.17	0.61	0.65	0.72	0.35	0.74	0.67	0.88	0.71	0.69	0.9
AdaBoost	0.65	0.65	0.65	0.68	0.66	0.64	0.6	0.71	−0.1	0.54	0.62	0.71	0.36	0.75	0.78	0.84	0.67	0.69	0.82
Neural Network	0.69	0.7	0.7	0.69	0.69	0.68	0.64	0.7	−0.2	0.63	0.67	0.66	−0.7	0.73	0.81	0.9	0.29	0.7	0.9

Table 6 Summary of prediction accuracies of R^2 values for weft crimp prediction models

| Model | Filtering & Cleaning | Normalization | | Predictors Enhancement | | Data Fragmentation | | | | Outlier Detection | | | | Data Enrichment | | | | | |
		Min-Max	Z-score	Product	Square	F1	F2	F3	F4	F-1	F-2	F-3	F-4	F1FE	F2FE	F3FE	F4OE	Combine Dataset without FL	Combine Dataset with FL
Linear Regression	−0.42	−0.42	−0.42	0.59	0.61	0.61	0.55	0.18	0.36	0.54	0.49	0.13	0.72	0.69	0.49	0.35	0.88	0.58	0.58
Ridge Regression	−0.36	−0.36	−0.36	0.6	0.6	0.61	0.55	0.18	0.37	0.54	0.49	0.13	0.72	0.69	0.48	0.35	0.88	0.58	0.58
KNN	−15.29	−15.29	−15.29	0.57	0.57	0.64	0.57	0.21	0.38	0.6	0.54	0.12	0.7	0.79	0.71	0.71	0.89	0.78	0.79
Kernel Ridge (poly)	−0.06	−0.06	−0.06	0.58	0.6	0.65	0.58	0.16	0.36	0.56	0.56	0.1	0.73	0.77	0.61	0.34	0.9	0.65	0.65
Kernel Ridge (RBF)	0.11	0.11	0.11	0.59	0.59	0.64	0.52	0.17	0.25	0.58	−0.04	0.13	0.75	0.81	0.61	0.24	0.88	0.61	0.61
Random Forest	−8.97	−9.39	−8.97	0.59	0.62	0.69	0.63	0.18	0.51	0.63	0.53	0.14	0.66	0.78	0.75	0.63	0.9	0.78	0.78
AdaBoost	0.51	0.51	0.51	0.6	0.63	0.68	0.59	0.24	0.21	0.62	0.49	0.15	0.5	0.78	0.69	0.5	0.87	0.74	0.75
Neural Network	−0.93	−0.93	−0.93	0.62	0.63	0.7	0.61	0.13	0.01	0.61	0.49	0.12	0.52	0.8	0.58	0.62	0.87	0.7	0.69

After removal of potential anomalies within the fragments, the prediction accuracies for F-2 and F-3 were improved with a maximum R^2 value of 0.67 and 0.76 respectively. However, prediction accuracies for F-1 and F-4 have decreased to 0.63 and 0.35 respectively. The reduction in prediction accuracies indicates that eliminated instances were not truly outliers for F1 and F4 fragments and they were kept for the data enrichment process. The data enrichment process takes two fragments (F2 and F3) with excluded outliers, the other two fragments (F1 and F4) having original instances. After applying data enrichment over individual fragments, the highest predicted accuracy (0.78) for fragment (F-1) was resulted by Kernel Ridge (RBF) classifier. Similarly, for F-2 and F-3 fragments, KNN has resulted in the highest R^2 values as 0.83 and 0.89 respectively. In the case of fragment F4, the Ridge

Regression and Kernel Ridge (RBF) has resulted in the same highest R^2 value i.e. 0.74. Individual fragments had shown improved results however, these fragments were combined together for evaluation. The evaluation of combined fragments with and without float length predictor was carried out. For all combined fragments but without the float length predictor, the KNN classifier resulted in an accuracy of 0.75. The accuracy has improved further when combining fragments and including the float length predictor. The highest accuracy for combined fragments with float length resulted by Random Forest and Neural Network classifiers as 0.75 and 0.86 respectively. The predictor selection process presented by the present work has shown a clear improvement in the prediction accuracy of wrap crimp from initially 0.69 to finally 0.9 R^2 value.

Weft crimp prediction: The training model for weft crimp prediction has been trained similarly as wrap crimp prediction model. The initial accuracy results after filtering and cleaning of the data for weft crimp prediction is 0.51 R^2 value by the Adaptive Boost (AdaBoost) classifier. The data is then passed through the data normalization process and was evaluated for prediction accuracy. The highest prediction accuracy after normalization is 0.51 by the AdaBoost classifier. The accuracy value improved from 0.51 to 0.63, after applying the predictor enhancement process. Later, fragmentation of the data based on float length values showed improvement in prediction accuracy. The Neural Network classifier resulted in the highest accuracy of 0.70 for the fragment F-1. Random forest classifier outperformed all others for F2 and F4 with R^2 values of 0.63 and 0.51 respectively. The fragment F3 had a very low prediction accuracy among all fragments with a 0.24 R^2 value, resulted by the AdaBoost classifier. Outlier

detection and removal process had shown decreased prediction accuracies for fragments F-1, F-2 and F-3 with the R^2 values 0.63, 0.56 and 0.15 respectively. However, fragment F4 gained an improvement in R^2 value from 0.51 to 0.75 for Kernel Ridge (RBF) classifier after passing through outlier detection and removal process. As per results, the fragments F1, F2 and F3 with all instances and fragment F4 with excluded outliers were carried further to data enrichment process. The highest prediction accuracies of the individual fragment after applying data enrichment were 0.81, 0.75, 0.71 and 0.90 for fragment F1, F2, F3 and F4 respectively. Finally, all fragments were combined together for collective weft crimp prediction. The perdition accuracy of combined fragments with float length was 0.78 and 0.79 for combined fragments without float length. The predictor selection process presented by the present work has shown a clear improvement in the prediction accuracy of weft crimp from initially 0.51 to finally 0.79 R^2 value.

Table 7 Parameters and crimp percentage of test fabrics

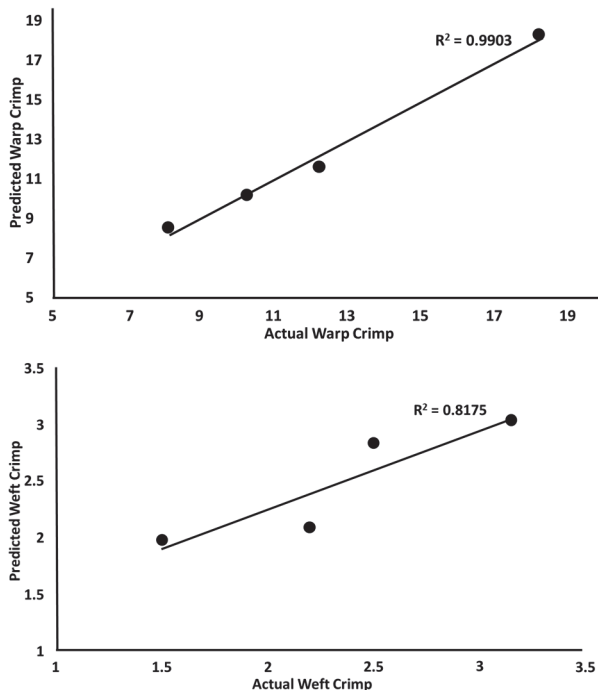| Test Fabric	Wrap Linear Density	Weft Linear Density	EPI	PPI	Float Length	GSM	Reed Count	Reed Denting	Actual (D3883)		Predicted	
									Wrap Crimp	Weft Crimp	Wrap Crimp	Weft Crimp
1	14	16	94	62	1	320	23.28	4	18.25	1.5	18.61	1.98
2	30	30	130	69	2	175	42	3	10.30	2.5	10.34	2.84
3	21	21	108	56	3	210	26.32	4	12.25	3.15	11.79	3.04
4	40	40	140	70	4	140	34	4	8.15	2.2	8.66	2.09

Figure 2 Validation result with ASTM standard D3883 for a) warp crimp; b) weft crimp

4.2 ASTM Standard D3883 Method

In this subsection, the proposed prediction model is validated by comparing the predicted crimp values with actual crimp values i.e. measured through ASTM standard D3883 crimp measurement method [7]. Any crimp prediction technique is prone to error as the weaving of fabric is highly associated with its weaving methods and raw materials used that induces noise. This error can be measured by comparing the predicted crimp values with the actual crimp values. Moreover, a small difference

between these two values provides validation of the prediction technique. For this purpose, four fabric samples were made in weaving lab at National Textile University. Fabrics were weaved from cotton carded yarn on the air-jet loom and fabric parameters were recorded as shown in Tab. 7. The predicted crimp values for each sample are computed from the proposed model whereas the actual crimp percentage of warp and weft yarns is measured through ASTM standard D3883 method.

The Pearson correlation analysis is performed to determine the difference between predicted and actual crimp values. Fig. 2a-b shows fitted line plots of wrap and weft crimp differences between predicted and actual crimp values. The strong, positive correlation for both warp and weft was observed with a p-value of 0.000. The correlation values $R^2 = 0.9903$ for wrap and $R^2 = 0.8175$ for weft show high accuracy of the proposed prediction model.

4.3 Comparative Analysis with State-of-the-art

The comparative analysis between state-of-the-art [2] and the proposed prediction technique provides further validation of the proposed technique. Maqsood et al. [2] had predicted wrap and weft crimp percentage in woven cotton fabric using linear regression, a statistical model. The prediction accuracy was assessed by correlation analysis of the training and test sets using holdout evaluation method. Their dataset contains sixty instances of cotton woven fabric i.e. labelled with wrap and weft crimp percentages. Fabric parameters used by Maqsood [2] as predictors were yarn linear densities, yarn densities, the width of fabric, reed count and reed denting. These predictors are the same as the predictors used in present work except for fabric width where the present work used the GSM. The reason for choosing Maqsood et al. [2] is because their dataset has the same specifications as the

proposed dataset. Similarly, the proposed model can easily be applied to Maqsood dataset and vice versa. Tab. 8 shows the comparison analysis between Maqsood et al. and the present model. An implementation of Maqsood's model was made through MINITAB statistical software. Both datasets were loaded in the software. Maqsood's model performs well on their own dataset with 0.98 R^2 value for both wrap and weft crimp prediction. However, their model with the present dataset has resulted in a very low prediction accuracy i.e. 0.50 and 0.17 for wrap and weft crimp respectively. The reason for this low prediction accuracy is that their model is incapable of handling large size dataset. On the other hand, a high prediction accuracy (0.97) for both wrap and weft crimp resulted in the proposed model when applied on the Maqsood dataset. The comparison in Tab. 8 shows the fact that previous models trained on very small and controlled datasets are unable to predict the unseen data with the same accuracy as reported. However, the proposed model has achieved an increased prediction accuracy for both small and medium size datasets.

5 CONCLUSIONS

The aim of the present work is to automate the decision making of crimp prediction in pre-production cotton fabric. The proposed prediction model eliminates the requirement of expert involvement from the fabric manufacturing process. Moreover, the proposed approach ensures accurate, cost and time effective pre-production crimp prediction that improves the productivity of the textile industry. Prediction models for prediction of warp and weft yarn crimp percentages were developed using a machine learning technique.

Table 8 Comparison between proposed model and state-of-the-art

	Warp Crimp (R^2)		Weft Crimp (R^2)	
	Maqsood 2014 Model	Present Work Model	Maqsood 2014 Model	Present Work Model
Maqsood 2014 Dataset	0.98	0.97	0.98	0.97
Present Work Dataset	0.50	0.99	0.17	0.81

Previously only neural networks and linear regression models were used to predict the crimp percentage in yarn. The major difference of present work from previous studies is the size and the nature of the sample dataset and predictor engineering for training data. Our dataset is quite large as compared to previous studies i.e. 1273 fabric samples. Previously the largest dataset used by Maqsood (2014) was of 68 samples. Moreover, fabric samples used previously were controlled samples manufactured at labs with predefined fabric parameters. The present dataset is the real data taken from the industry containing all the possible variations. Our dataset being large also takes care of float lengths 1, 2, 3 and 4, while most of the previous works are based on single float length value. Our proposed prediction model predicts unseen data with high accuracy of R^2 value above 0.95. Among all classifiers, the prediction models which outperformed others were KNN, Kernel ridge (RBF), Random Forest and Deep Neural Network. The proposed model was validated using three different methods i.e. K-fold cross-validation, ASTM standard D3883 and comparative analysis with state-of-the-art technique.

The present system is trained on the fabric samples that are produced from cotton yarns manufactured with ring spinning (carded) on air-jet looms. However, further studies are needed for other types of fabrics. The scope of the present system can be further enhanced by including fabrics made up from other raw materials and yarn made from open-end spinning (combed yarn) and air vortex techniques on looms other than air-jet.

6 REFERENCES

[1] Adanur, S. (2000). *Handbook of weaving*. CRC press.
[2] Maqsood, M., Hussain, T., Nawab, Y., Shaker, K., & Umair, M. (2015). Prediction of warp and weft yarn crimp in cotton woven fabrics. *The Journal of the Textile Institute, 106*(11), 1180-1189. https://doi.org/10.1080/00405000.2014.981041
[3] Afroz, F. & Siddika, A. (2014). Effect of warp yarn tension on crimp % in woven fabric. *European Scientific Journal, ESJ, 10*(24).
[4] Nazir, M. U., Shaker, K., Nawab, Y., Fazal, M. Z., Khan, M. I., & Umair, M. (2017). Investigating the effect of material and weave design on comfort properties of bilayer-woven fabrics. *The Journal of the Textile Institute*, 108(8), 1319-1326. https://doi.org/10.1080/00405000.2016.1247616
[5] Topalbekiroğlu, M. & Kaynak, H. K. (2008). The effect of weave type on dimensional stability of woven fabrics. *International Journal of Clothing Science and Technology, 20*(5), 281-288. https://doi.org/10.1108/09556220810898890
[6] Kumpikaitė, E. & Milašius, V. (2003). Analysis of interrelation between fabric structure factors and beat-up parameters. *Materials Science, 9*(2), 228-232.
[7] ASTM International (2016). ASTM D3883-04(2016) Standard Test Method for Yarn Crimp and Yarn Take-up in Woven Fabrics.
[8] Banerjee, P. K., Mishra, S., & Ramkumar, T. (2010). Effect of Sett and Construction on Uniaxial Tensile Properties of Woven Fabrics. *Journal of Engineered Fibers and Fabrics, 5*(2), 8-21. https://doi.org/10.1177/155892501000500202
[9] Majumdar, A., Mitra, A., Banerjee, D., & Majumdar, P. K. (2010). Soft computing applications in fabrics and clothing: a comprehensive review. *Research Journal of Textile and Apparel, 14*(1), 1-17. https://doi.org/10.1108/RJTA-14-01-2010-B001
[10] Shaker, K. et al. (2015). A statistical approach for obtaining the controlled woven fabric width. *Autex Research Journal, 15*(4), 275-279. https://doi.org/10.1515/aut-2015-0008
[11] Malik, S. A., Gereke, T., Farooq, A., Aibibu, D., & Cherif, C. (2018). Prediction of yarn crimp in PES multifilament woven barrier fabrics using artificial neural network. *The Journal of the Textile Institute*, 109(7), 942-951. https://doi.org/10.1080/00405000.2017.1393786
[12] Lin, J.-J. (2003). A genetic algorithm for searching weaving parameters for woven fabrics. *Textile Research Journal, 73*(2), 105-112. https://doi.org/10.1177/004051750307300203
[13] Li, Z, Wu, D., & Yu, T. (2019). Prediction of Material Removal Rate for Chemical Mechanical Planarization Using Decision Tree-Based Ensemble Learning. *Journal of Manufacturing Science and Engineering, 141*(3). https://doi.org/10.1115/1.4042051
[14] Li, Z., Zhang, Z., Shi, J., & Wu, D. (2019). Prediction of surface roughness in extrusion-based additive manufacturing with machine learning. *Robotics and Computer-Integrated Manufacturing, 57*, 488-495. https://doi.org/10.1016/j.rcim.2019.01.004
[15] Gao, H. N., Shen, D. H., Yu, L., & Zhang, W. C. (2020).

Identification of Cutting Chatter through Deep Learning and Classification. *International Journal of Simulation Modelling*, *19*(4), 667-677. https://doi.org/10.2507/IJSIMM19-4-CO16

[16] Sousa Junior, W. T. de, Montevechi, J. A. B., Miranda, R. de C., Rocha, F., & Vilela, F. F. (2019). Economic Lot-Size Using Machine Learning, Parallelism, Metaheuristic and Simulation. *International Journal of Simulation Modelling*, *18*(2), 205-216. https://doi.org/10.2507/IJSIMM18(2)461

[17] Li, Z., Goebel, K., & Wu, D. (2018). Degradation Modeling and Remaining Useful Life Prediction of Aircraft Engines Using Ensemble Learning. *Journal of Engineering for Gas Turbines and Power Print, 141*(4). https://doi.org/10.1115/1.4041674

[18] Lin, J.-J. (2007). Prediction of yarn shrinkage using neural nets. *Textile Research Journal, 77*(5), 336-342. https://doi.org/10.1177/0040517507078787

[19] Shahabi, N. E., Mousazadegan, F., Varkiyani, S. M. H., & Saharkhiz, S. (2014). Crimp analysis of worsted fabrics in the terms of fabric extension behaviour. *Fibers Polymers, 15*(6), 1211-1220. https://doi.org/10.1007/s12221-014-1211-y

[20] Mertova, I., Neckar, B., & Ishtiaque, S. M. (2016). New method to measure yarn crimp in woven fabric. *Textile Research Journal, 86*(10), 1084-1096. https://doi.org/10.1177/0040517514551464

[21] Kaplan, V., Varan, N. Y., Dayik, M., Turhan, Y., & Durur, G. (2016). Detection of Warp Elongation in Satin Woven Cotton Fabrics Using Image Processing. *Fibres Textiles in Eastern Europe*. https://doi.org/10.5604/12303666.1183202

[22] Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. 3rd ed. Morgan Kaufmann.

[23] Han, J., Kamber, M., & Pei, J. (2012). *12 - Outlier Detection*. Data Mining (Third Edition), 543-584. https://doi.org/10.1016/B978-0-12-381479-1.00012-5

Contact information:

Muhammad Zohaib FAZAL
School of Electrical Engineering and Computer Science (SEECS),
National University of Sciences and Technology (NUST),
Islamabad, Pakistan
E-mail: 14msitmfazal@seecs.edu.pk

Sharifullah KHAN
(Corresponding author)
School of Electrical Engineering and Computer Science (SEECS),
National University of Sciences and Technology (NUST),
Islamabad, Pakistan
E-mail: sharifullah.khan@seecs.edu.pk

Shahzad YOUNIS
School of Electrical Engineering and Computer Science (SEECS),
National University of Sciences and Technology (NUST),
Islamabad, Pakistan
E-mail: muhammad.shahzad@seecs.edu.pk

Muhammad Azeem ABBAS
University Institute of InformationTechnology,
Pir Mehr Ali Shah Arid Agriculture University Rawalpindi,
Rawalpindi, Pakistan
E-mail: azeem.abbas@uaar.edu.pk

Yasir NAWAB
National Textile University,
Faisalabad, Pakistan
E-mail: yasir.nawab@yahoo.com