

Article

Intelligent Image Classification for Grading Egyptian Cotton Lint

Oliver J. Fisher¹, Ahmed Rady¹, Aly A. A. El-Banna², Nicholas J. Watson^{1*}, Haitham H. Emaish³

¹ Food Water Waste Research Group, Faculty of Engineering, University of Nottingham, Nottingham, UK; olivert.fisher@nottingham.ac.uk; ahmed.rady@nottingham.ac.uk

² Department of Plant Production, Faculty of Agriculture, Saba Basha, Alexandria University, Egypt; aly-elsawy@alexu.edu.eg

³ Department of Soils and Agricultural Chemistry, Faculty of Agriculture, Saba Basha, Alexandria University, Egypt; drhaitham1976@alexu.edu.eg

* nicholas.watson@nottingham.ac.uk

Abstract: Egyptian cotton is one of the most important commodities to the Egyptian economy and is renowned globally for its quality, which is currently graded by manual inspection. This has several drawbacks including significant labour requirement, low inspection efficiency, and influence from inspection conditions such as light and human subjectivity. This current work uses a low-cost colour vision system, combined with machine learning to predict the cotton lint grade of the cultivars Giza 86, 97, 90, 94 and 96. Unsupervised and supervised machine learning approaches were explored and compared. Three different supervised learning algorithms were evaluated: linear discriminant analysis, decision trees and ensemble modelling. The highest accuracy models (77.3-98.2%) used an ensemble modelling technique to classify samples within the Egyptian cotton grades: Fully Good, Good, Fully Good Fair, Good Fair and Fully Fair. The unsupervised learning technique k-means showed that human error is more likely to occur when classifying lint belonging to the higher quality grades and underlined the need for an intelligent system to replace manual inspection.

Keywords: Digital Manufacturing; Sensors; Machine Learning; Industry 4.0; Optical; Cotton Lint; Industrial Digital Technologies

1. Introduction

Cotton is an important textile crop, accounting for 90% of all-natural fibres used in the textile industry [1]. The textile industry plays a significant role in the Egyptian economy and wider society, contributing around 14% towards GDP [2] and employing 25.8% of the industrial workforce [3]. Among global cotton producers, Egyptian cotton has a reputation for producing the highest quality cotton [4]. The high quality of Egyptian cotton is primarily attributed to (A) the cotton lint being handpicked to avoid mixing yield of mature and immature plants and (B) the cotton fibre's exceptional length, fineness and brightness [5]. Since the mid-1980s, the production of Egyptian cotton has been declining [3], and exports have decreased from 164000 tonnes in 1980 to 71000 tonnes by 2019 [6]. Current challenges faced by the industry include high production costs and low productivity [3]. Domestic and international measures have been introduced to strengthen the Egyptian cotton industry. In 2017, the Egyptian Government introduced a 19-step plan to reverse the cotton industry's decline with the aim to increase the domestic cotton industry value, as the majority of cotton lint (85%) is exported as a raw product with no added value [7]. Additionally, in 2016 The United Nations Industrial Development Organisation (UNIDO) recognised the lack of innovation and scarce sustainable agricultural practices within the Egyptian cotton industry and initiated activities to improve the economic, social and environmental performance of cotton growers and cotton processors [8]. This work aims to address inefficiencies during the grading of Egyptian cotton lint by

developing an affordable intelligent sensor to grade cotton lint. This work complements the goal of UNIDO to support SMEs cotton growers and processors, as there is an opportunity to use this system to help drive value back towards SME Egyptian cotton processors, who have been identified as needing support in light of falling cotton prices [3]. Providing processors with the tools to grade their harvested cotton lint will strengthen their bargaining position to demand a fair price for their cotton lint, as higher graded cotton lint requires less downstream processing.

Egyptian cotton market value is determined first according to the cultivator (species) and then according to grade [9]. The grading is performed by manual inspection and is influenced by fibre colour and length, presence of “trash” (e.g., dried cotton leaves, seed-coats, barks, grass and dust) and maturity (age of plant harvested from) [10]. Manual inspection has several drawbacks including significant labour requirements, low inspection efficiency, and influence from inspection conditions such as light [11]. Previous work has explored developing decision support systems to improve, or replace, manual inspection. One example applied pair-wise correlation analysis to evaluate the relative importance of cotton characteristics, measured using the Cotton Classifying System Version-5 (CCS-V5) instrument, to determining fibre cotton grade quality [9]. This study identified that total trash (%), micronaire value and reflectance degree had the most influence on fibre cotton grades [9]. Other studies have looked into developing a Multi-Criteria Decision-Making (MCDM) support software to support fibre grading and selection problems using key cotton characteristics measured via a High Volume Instrument (HVI), which images the samples [12,13]. Previous work employed The Analytic Hierarchy Process method to determine weights of the cotton fibre properties, while Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) was used to compute the final quality values of cotton lint, which were subsequently graded and ranked [12]. Recent work has expanded on the MCDM approach to increase the number of fibre physical properties taken into consideration from six to thirteen [13]. Both these methods were validated on published cotton data and found to achieve a good correlation between the quality determined by the MCDM approach and manual inspection [12,13]. While these studies demonstrate the potential of removing human error from the fibre grading and selection process, they rely on cotton lint measurements from traditional fibre analytical equipment, such as the HVI, as input data. These analytical equipment have a high capital cost making them unattractive to SME Egyptian cotton lint producers that already suffer from high production costs [3]. However, combining affordable, easy to use, low-cost sensors with data-driven modelling will enable the development of intelligent sensors [14] that can replace high-cost analytical equipment for grading Egyptian fibre cotton.

Data-driven modelling techniques, such as Machine Learning (ML), may be used to analyse sensor measurements and generate actionable information. Machine learning methods develop models which learn from a training data set and are capable of fitting complex functions between input and output data. Machine learning models can accurately predict material properties from simple sensor measurements (e.g., optical [15], acoustic [16,17], pressure and temperature [18]) meaning they offer an affordable alternative to traditional high-cost analytical techniques.

Within the cotton industry, machine learning models have been developed to predict fibre properties from optical measurements [15], grade pilling on a fleece from optical measurements [19], recognise fibres from acoustic measurements [20] and detect pseudo-foreign fibres in cotton lint from optical measurements [10]. This work aims to develop an intelligent sensor to grade Egyptian cotton lint from image measurements recorded by a low-cost optical sensor. Previous work has already demonstrated potential in this area [11]. Lv *et al.* were able to develop a support vector machine ML model to classify the grade of cotton lint entering China based on the US upland grades (Good Middling, Strict Middling, Middling, Strict Low Middling, Low Middling, Strict Good Ordinary, Good Ordinary) [11]. However, this work did not provide details on how and by whom the cotton lint samples were labelled in order to create a data set to train and validate the model [11], which is important to assess the validity of the final ML model. If the data

used is not labelled correctly (e.g. by trained and experienced cotton grading professionals) the final ML model trustworthiness is questionable. This work expands on that of Lv *et al.* by first evaluating if it is possible to apply a similar methodology to classify Egyptian cotton lint grade which has unique properties compared to other cotton from other countries of origin [9] and is subject to more specific grading criteria (nine instead of seven used for the US cotton lint grading). Secondly, validating the methodology on a labelled dataset provided by the Cotton Arbitration & Testing General Organization (CATGO), Alexandria, Egypt. Thirdly, by expanding the number of features extracted from only the percentage of impurity within the cotton lint to include features that describe the colour and staining of the fibre lint, as previous work has identified a strong correlation of these features to Egyptian cotton quality [9].

Within this study, two applications of machine learning are implemented, which are supervised and unsupervised learning. Supervised machine learning refers to problems where the training data contains both input and labelled output data [21]. Whereas if the training data consist of a set of input data without any corresponding output data this is known as an unsupervised machine learning problem and is used analyse and cluster unlabelled data [21]. Unsupervised learning is used when there are concerns regarding the labelling of the data. The machine learning models developed for this study are examples of a classification models, in which the models aims to assign input data to one of a finite number of discrete categories or classes [21]. The complexity of manual inspection leads to inevitable human error in the labelling of cotton lint grades. Unsupervised learning was used to understand where human error is likely to occur by grouping together similar samples into clusters. This is important as, the cotton lint quality can vary between harvest seasons due to changes to environments, soil quality, biotic and abiotic stresses [22]. This means that the classification models will likely require retraining each season using new labelled training data to ensure the models are able to describe this variability and that the models' predictions remain accurate. The time associated with the collection and labelling of a training data set may make supervised machine learning unfeasible. Therefore, the unsupervised machine learning modelling route was also explored as a method to reduce the volume of labelled output data required.

2. Materials and Methods

2.1 Cotton lint samples

To develop the ML classification models, a key component of the intelligent sensor, a dataset of images of Egyptian cotton lint samples were collected and graded (labelled) by the Cotton Arbitration & Testing General Organization (CATGO), Alexandria, Egypt. The CATGO is responsible for providing the official certificates for authenticating cotton lint in terms of determining the quality attributes, and grade for Egyptian cotton cultivars for cotton ginning companies. The CATGO provided samples for five cultivars Giza 86, 87, 90, 94 and 96. Egyptian cotton lint is categorised into nine grades, Fully Good through to Fully Fair. Example images of the grades for Giza 96 are presented in Figure 1, to show the change in quality between grades. To aid communication the grades have been assigned a number from 1 for the highest quality grade (Fully Good) to 9 for the lowest quality grade (Fully Fair).

**(a)****(b)****(c)****(d)****(e)****(f)****(g)**

Figure 1 . Sample images of each grade (a) 1: fully good, (b) 3: good, (c) 4: fully good fair to good, (d) 5: fully good fair, (e) 6: good fair to fully good fair, (f) 7: good fair and (g) 9: fully fair to good fair for the Giza 96 cultivar, apart from grades 2 (good to fully good) and 8 (fully fair to good fair) that was not represented in the data.

A total of 3,447 samples were provided, but unfortunately, samples of all grades for each cultivar were not available due to an insufficient number of samples being provided for certain grades. The breakdown of the number of samples of each grade for each cultivar is reported in Table 1.

Table 1 The breakdown of the number of samples of each grade for each cultivar.

Grade	Giza 86	Giza 87	Giza 90	Giza 94	Giza 96
(1) Fully good	115	N/A*	N/A	102	103
(2) Good to fully good	118	116	100	115	N/A
(3) Good	113	108	131	100	109
(4) Fully good fair to good	119	110	116	108	118
(5) Fully good fair	150	115	124	109	97
(6) Good fair to fully good fair	115	N/A	131	99	102
(7) Good fair	115	N/A	101	N/A	120
(8) Fully fair to good fair	N/A	N/A	N/A	N/A	N/A
(9) Fully fair	N/A	N/A	N/A	104	64

*N/A: not available

2.2 Colour vision system

A schematic configuration of the colour vision system, utilised in this study, is shown in Figure 2. The system contained an 8.1 MP Fuji A850 digital camera (FUJIFILM Corporation, Minato-ku, Tokyo, Japan), which captured images in JPEG format with dimensions of 2248x3264 pixels. The camera was mounted 11 cm vertically above the surface of the samples. A triangle 10 W LED light source was mounted 11 cm vertically above the surface of the samples and used to ensure consistent illumination of the samples. The average light intensity on the sample surface was 4879 lux, calibrated by a Samsung Galaxy M31, 10 times, with the average value reported. To ensure a consistent illumination condition, the camera along with the light source was enclosed in a light aluminium box whose dimensions were 54x40x40 cm. The inside of the box was coloured black to minimise the surface reflection from the sides. Samples of 4 cm thickness were placed directly below the camera on the black surface. Each image was captured with no flash, and the output images were transferred to a PC for analysis.

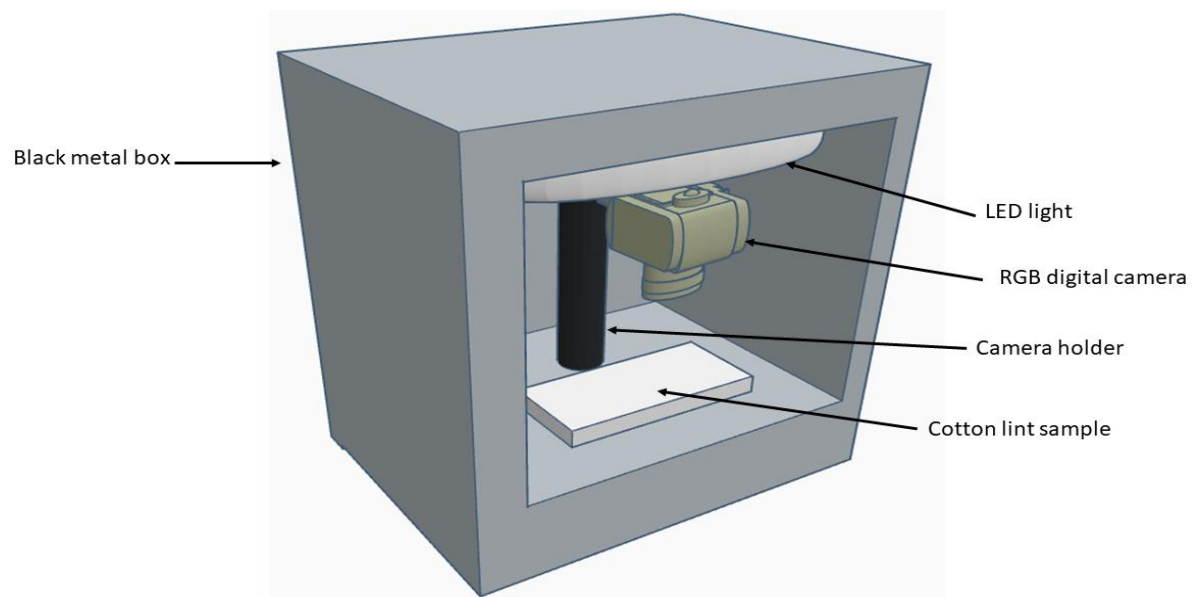


Figure 2 Schematic diagram of the colour vision system used to acquire colour images of the cotton lint samples.

2.3 Image processing procedures

Ten features were extracted from the cotton lint images and used as inputs for the classification models developed following the steps outlined in Figure 3. First, all the images were cropped from 2448x3264 to 2448x2965 pixels to eliminate the areas of the sample with less uniform light. To obtain the region of interest (ROI), colour segmentation using k-means clustering was applied. The RGB colour values were transformed into the International Commission on Illumination (CIE) $L^*a^*b^*$ colour space, where L^* , a^* , and b^* refer to lightness from black to white, green to red component, and blue-yellow component, respectively [23]. The $L^*a^*b^*$ colour coordinate are more appropriate for colour segmentation as it is device-independent, perceptually uniform, and more appropriate to show small colour differences compared with RGB colour coordinate [24]. In this study, each image was segmented into sub-images. The first sub-image includes the lightest areas in the image (S1), the second sub-image includes the darkest areas in the image (S2), and the third sub-image includes the areas in between the lightest and darkest areas (S3) as shown in Figure 3. The $L^*a^*b^*$ values for all sub-images were considered as features. Additionally, the impurity was calculated by applying the histogram thresholding on the grey image to obtain the binary image [24]. The ratio between the area of the white segmented areas in the binary image and the total image area was considered as the impurity. This process resulted in a total of 10 features extracted from each sample image.

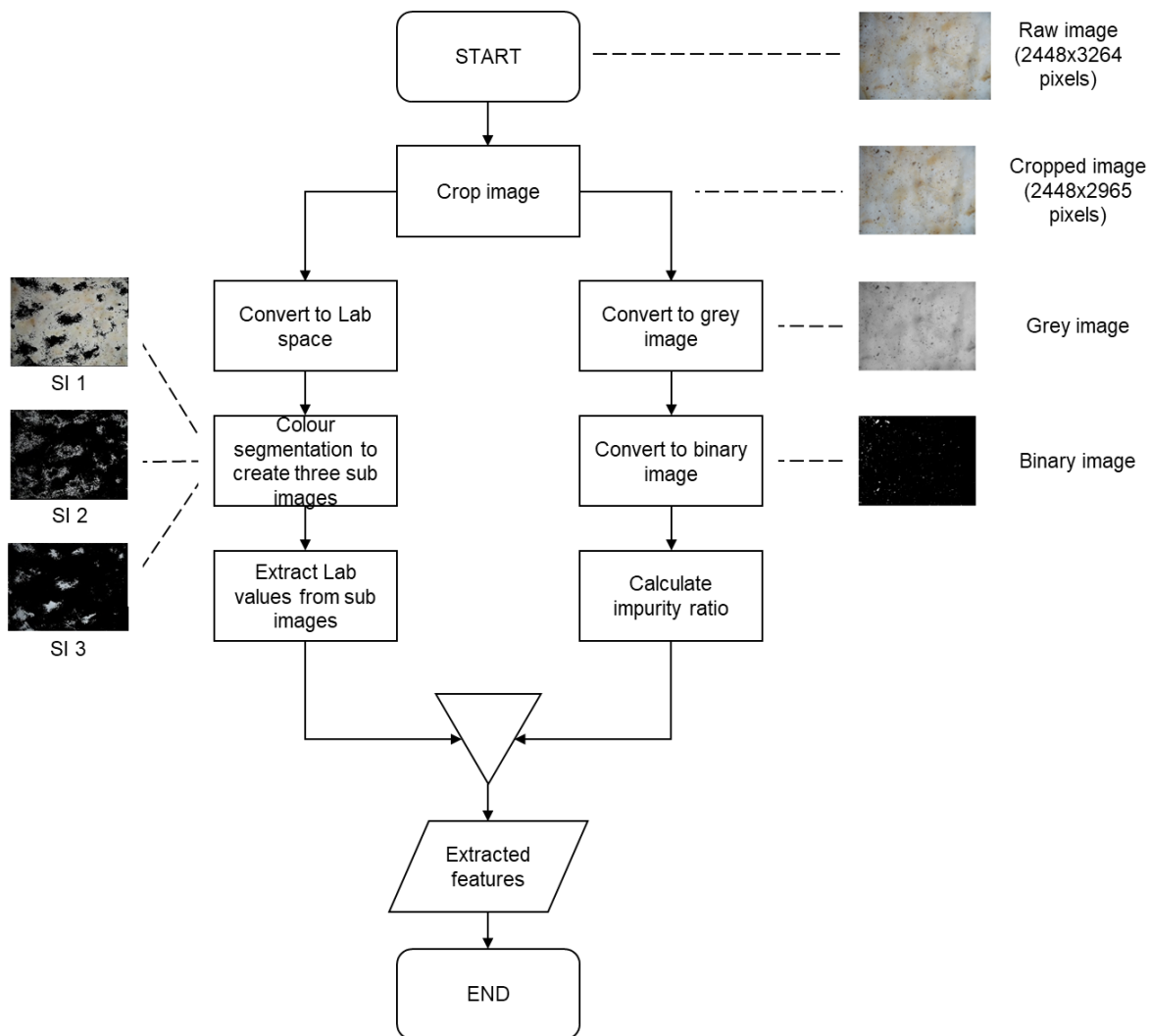


Figure 3 Image processing and feature extraction steps with example images of each step.

2.4 Feature analysis

The features extracted from the images were analysed to understand how feature values vary throughout the cotton grades and to determine the relative importance of each feature when classifying cotton lint samples. Random forest models were used as part of the analysis to determine the relative importance of each feature. A random forest model is a specific example of an ensemble modelling technique, outlined in Section 2.5, that is constructed from multiple weak decision tree classifiers, each trained on a different subset of the data [25]. The random forest models were built using the MATLAB R2021a software (MathWorks). Default parameter values were used except for the number of trees, which was increased to 500 to improve model accuracy with acceptable increases in computational time [26]. The model accuracy was evaluated using the Out-Of-Bag accuracy, which is the average accuracy of each tree assessed on the subset of data not used for training that specific tree. The relative importance of each feature was obtained with the mean decrease impurity method [27].

2.5 Machine learning classification models

For the supervised classification part of this work, two machine learning algorithms were investigated, which were Linear Discriminant Analysis (LDA) and Decision Trees (DT). The LDA classifier was chosen as it is simple to implement and relatively fast to train [21]. This classifier includes a linear combination of features and is based on the value of the discriminant function before the image is classified into the appropriate class [21]. Decision trees are based on sequential partitioning of the input feature space into several binary sub-regions based on model parameters. The sample is then classified into the class that has the minimum error [21]. Ensemble techniques are based on optimising the classification performance by weighing several individual classifiers and combining them to obtain a classifier that out-performs the single classifier [25]. Among several ensemble methods, the bagging technique was applied in this study. The idea of ensemble bagging is to form a better-performing classifier from a number of weaker performing classifiers, where each weak classifier is a result of training the samples with replacement [25].

The extracted features were normalised to ensure all variables were given equal weight by the classification algorithms. To normalise without any loss of information the minimax function was applied [28]. For each cultivar data set, the data was partitioned into training (80%) and testing (20%) data sets. Stratified random sampling was used to ensure that the training and testing data sets were balanced across the grades [29]. The training data was used to train the models following cross-validation (10-fold) procedures to optimise the models' hyperparameters. Hyperparameters are parameters (e.g., the maximum depth of a decision tree) that are set before training a model [30]. The developed models' ability to classify cotton grade was then evaluated using the unseen test data.

The unsupervised machine learning part of this work, a K-means clustering algorithm was evaluated as an alternative method to classify the cotton lint images. K-means clustering is a method that aims to partition observations into k clusters in which each observation belongs to the clusters with the nearest means [21]. The algorithms identify centroids of the k clusters, which can then be used to label new data [21]. For this work, k was set to the number of known grades contained within each cultivar dataset. These machine learning models were again developed using the MATLAB R2021a software.

Preliminary results from the supervised classification (e.g., accuracy between 63.94–65.38% for Giza 86 cultivar) indicated that the ML models were struggling to classify the cotton lint samples when the transitional grades were included (i.e., those containing the word "to" in Table 1, Section 2.). This was likely due to the boundaries between the grades not being sufficiently defined within the feature data (Section 3.1). Therefore, separate models were developed from data containing all the grades and data where the "to" grades were removed from the data. This was to assess whether if by reducing the granularity of the model (i.e., number of grades) the accuracy could be increased.

3. Results and discussion

3.1 Feature analysis

Following the image processing steps outline in Section 2.3, ten features were extracted from the cotton images and used as inputs for the classification models. The data is visualised in Figure 4 and Figures A.1-4 to better understand the effect of each feature on grading cotton quality. The feature relating to the percentage of impurity within the cotton lint images standouts as the feature that displays the most consistent impact on cotton quality, as demonstrated by variation in boxplot heights between grades (Figure 4 and Figures S.1-4). Figure 4(D) shows that for the Giza 86 cultivar, as the cotton grade deteriorates from grade 1 to 7 the percentage of impurities within the cotton lint image increases. This relationship is repeated in the other cultivars studied (Figures S.1-4) and supports previous evidence of a strong negative correlation between impurities and Egyptian cotton quality [9]. However, the distinction between grades is not always captured solely by the percentage of impurities. For example, Figure 4(D) shows the grades 5 and 6 have similar distributions for the feature percentage of impurities. This suggests that this feature alone may not be able to classify cotton lint grade to sufficient accuracies.

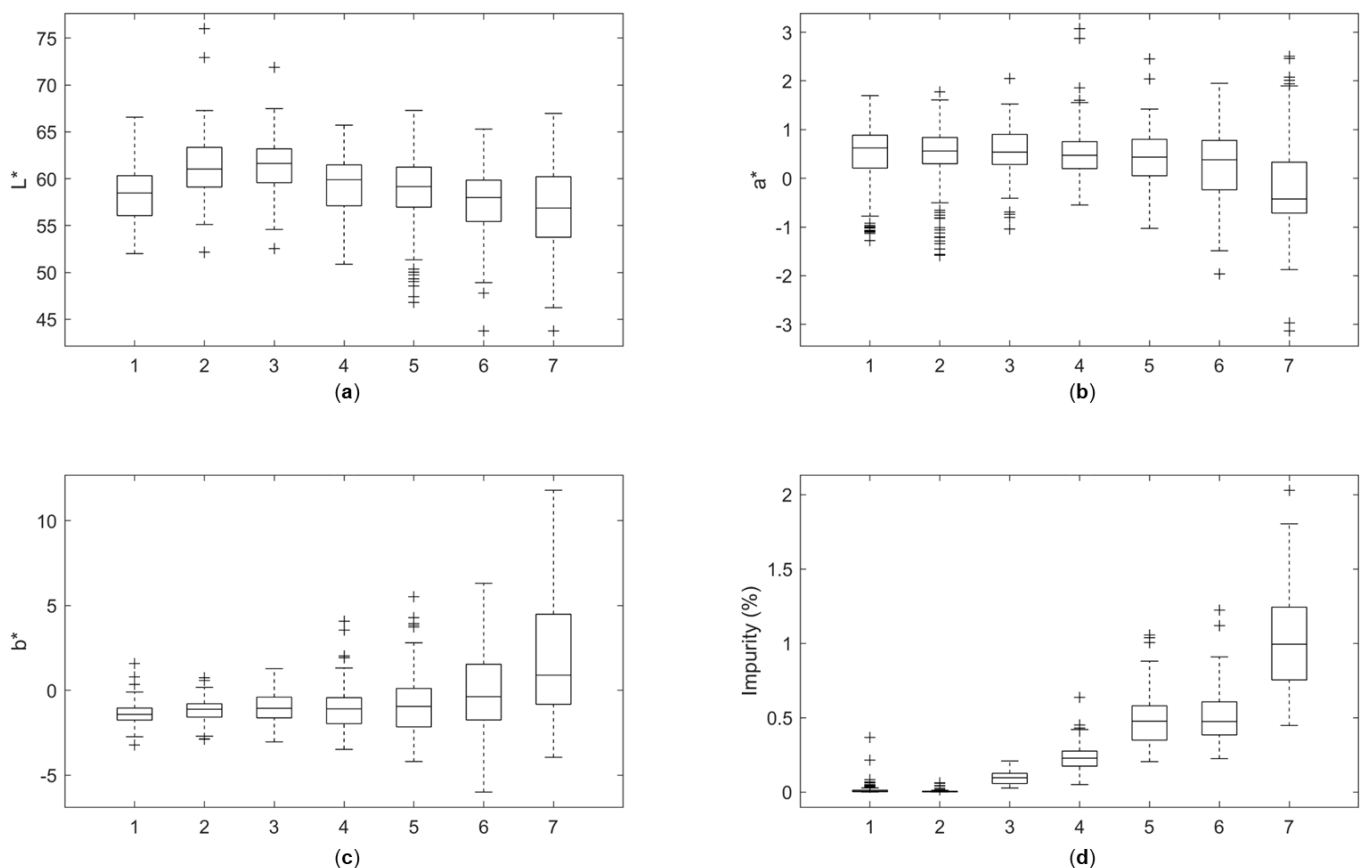


Figure 4 Distribution of the cultivar Giza 86 extracted features (a) L^* , (b) a^* , (c) b^* and (d) impurity data across the Egyptian cotton lint grades

The three remaining features contain information relating to the colour of the cotton lint, a key cotton property linked to quality [31]. Previous path correlation analysis has identified that colour measurements from a HVI machine show a strong negative correlation between the degree of yellowness (b^*) and Egyptian cotton quality, and a strong positive correlation between the degree of reflectance (L^*) and Egyptian cotton quality [9]. The first of these relationships is, to an extent, reflected in the cotton lint data, as all but Giza 87 report an increase in the b^* median from their respective best to their worst grades

(increase from between -1.42 and -0.96 to between 0.89 and 2.30). However, Figure 4(A) shows little evidence of a strong positive correlation between L^* and Egyptian cotton quality, as the median L^* between the best and worst grades, are similar (58.47 and 56.87 respectively) and fluctuates between these grades. There is some evidence of a positive relationship within the Giza 96 samples measured, as median L^* increases from 54.80 to 60.02 from the worst to best grade. Although, Figure A4(A) shows that this relationship is not linear for the grades in between. When using a HVI machine to measure the cotton quality, the colour measurements are limited to only b^* and L^* , yet previous studies have shown that degree of redness (a^*) to influence cotton quality [32]. The Figure 4(B) and Figures A1-4(B) shows that between the grades 1 and 6 the a^* feature has similar distributions. However, a^* measurements decrease when moving from grade 6 to 7, illustrated by the a^* medians decreasing from between 0.25 and 0.71 to between -0.60 and 0.20. This suggests there is a relatively large colour change occurring between these grades that may help the classification models distinguish between these grades.

While examining the features extracted provides some distinctions between the grades, Figure 4 and Figures A1-4 shows a large overlap between the distributions of features between adjacent grades. This is particularly true for the colour related features. The reason for the overlap in distributions is because cotton lint grade is not solely evaluated based on one attribute, but a combination of attributes (colour, fibre length, percentage of impurities, maturity and more) [9]. This, and its reliance on human intuition, causes challenges in previous attempts to model the grading of Egyptian cotton lint [9]. This is where machine learning classifications models may excel, due to their ability to detect hidden relationships within the data [21]. Subsequently, the relative importance of input variables to adsorption efficiency was calculated using a random forest model with the mean impurity decrease method.

As shown in Figure 5, the percentage of impurity is the most influential factor in the model decision-making process, which estimated to be a minimum of between 3.03 and 9.06 times more important than the colour features for classifying cotton lint grade, dependant on the cultivar. The difference was most pronounced in the Giza 87 cultivar where the impurity feature was more than 9 times more important than all the features. The Out-Of-Bag prediction accuracy for the random forest was between 65.46 and 80.85% accurate, depending on the cultivar. Due to the low accuracy (<70%), the random forest feature importance analysis was repeated but with the "to" grades removed to increase the distinction between the grades. This resulted in an increase to the Out-Of-Bag prediction accuracy to between 80.17 and 97.31%, depending on the cultivar. The new models produced a similar conclusion by estimating that the impurity feature is at least estimated to be a minimum of between 3.03 and 14.57 times more important than the colour features for classifying cotton lint grade.

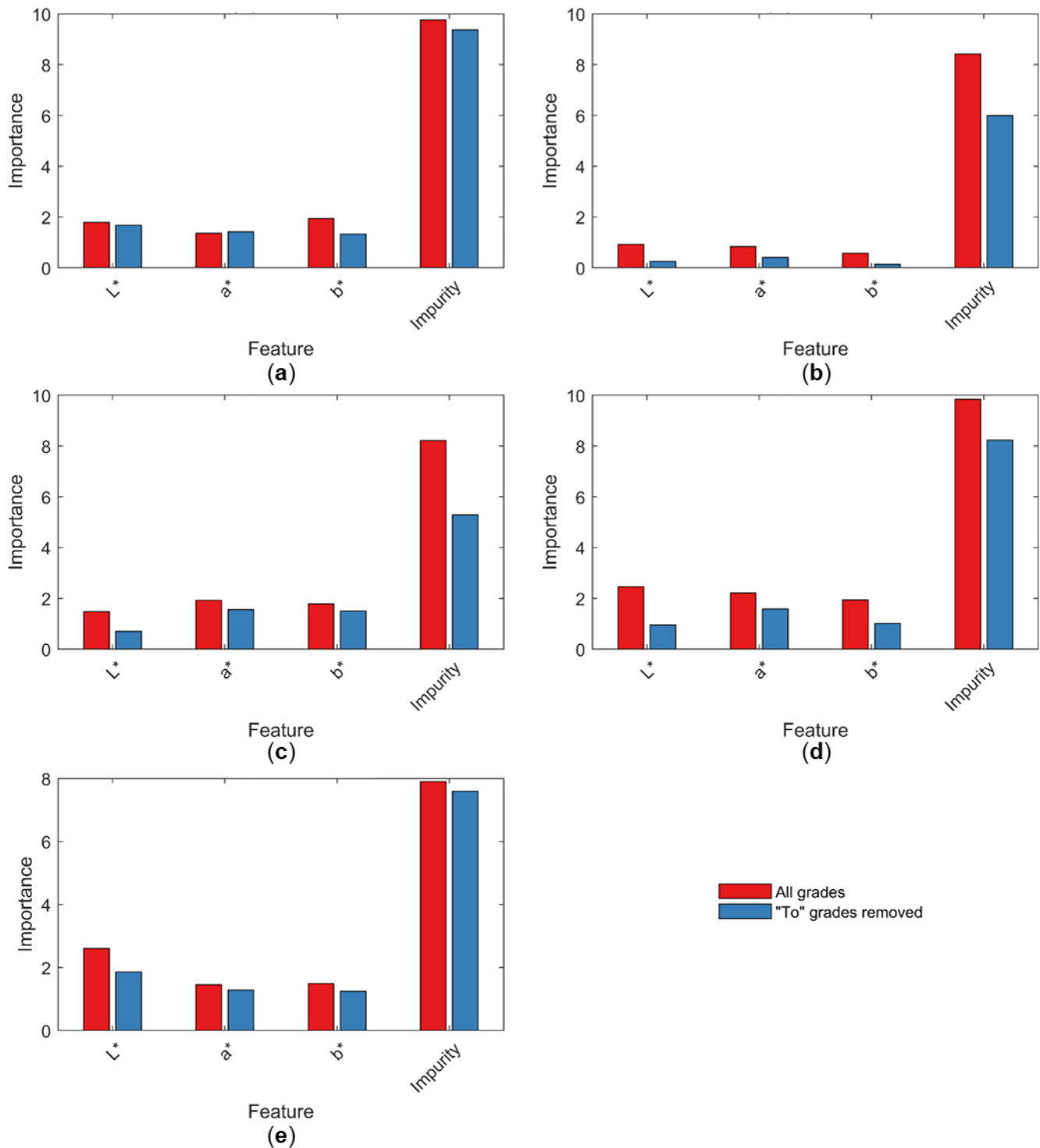


Figure 5. The relative importance of the extracted image features on the cotton lint grades for the cultivars (a) Giza 86, (b) Giza 87, (c) Giza 90, (d) Giza 94 and (e) Giza 96 determined using a random forest model developed for each cultivar.

In addition to quantifying the importance of the image features on cotton lint grading, the analysis can be used to identify which features significantly contribute to improving the accuracy of the classification models. This can reduce the computational time involved both in the training of the models and using the models for future predictions [27]. The random forest analysis in Figure 5 predicts that a model developed using solely the percentage of impurity feature will achieve similar classification accuracy as one developed using all the extracted features from the cotton lint image. However, when comparing the prediction accuracy of the models built with solely the impurity feature models to

those built using all features and the, only the models built to classify the Giza 87 grade had similar accuracies, as reported in Table 2. The remaining classification models developed using only the impurity feature reported a decrease in accuracy of 16.43-24.55 when trained on all the grades and 4.82-20.83 when trained on the data with “to” grades removed compared to the models developed using all the extracted features. As previously discussed, these reduced accuracies may be explained by the overlap in the distributions of the impurity feature between the grades highlighted in Figure 4(D) and Figures A1-4(D). In addition, the computational time was only reduced by a maximum of 0.45 seconds for the training time and 0.31 seconds for the prediction time for the models using on impurity as a feature. Therefore, due to the negligible increase in computational time and decrease in the accuracy, the decision was taken to develop the classification models using all the extracted features.

Table 2 Comparison of classification performance for models built using all the features extracted from the cotton lint images against those built using only the extracted percentage of impurity feature

Cultivar	Accuracy (%)		Training time (s)		Prediction time (s)	
	All features	Only impurity	All features	Only impurity	All features	Only impurity
<i>All grades</i>						
Giza 86	71.86	47.31	5.16	5.39	1.59	1.90
Giza 87	77.78	74.44	4.67	4.90	1.27	1.35
Giza 90	64.29	47.86	5.27	5.21	1.65	1.66
Giza 94	73.65	53.38	5.14	4.98	1.40	1.71
Giza 96	62.59	45.32	5.18	5.19	1.54	1.56
<i>“to” grades removed</i>						
Giza 86	95.88	83.51	4.32	4.30	1.90	1.44
Giza 87	95.56	95.56	4.03	3.94	1.35	1.12
Giza 90	85.92	73.24	4.30	4.30	1.66	1.40
Giza 94	95.18	90.36	4.26	4.71	1.71	1.27
Giza 96	76.04	55.21	4.55	4.65	1.56	1.50

3.2 Predicting cotton lint grade

3.2.1 Supervised machine learning modelling approach

The supervised machine learning models were first developed to classify the cotton lint samples into one of nine grades. However, when evaluated on the test data the highest accuracy achieved was only 83.04% by the LDA model developed to classify the Giza 87 samples. For the cultivar Giza 96, the highest classification accuracy achieved was 65.32% by the ensemble model, meaning a sensor developed using this model would misclassify the Giza 96 cotton lint samples 35 times out of 100. Classification models built to grade non-Egyptian cotton have achieved accuracies up to 98.9% [11]. However, these models were built to classify cotton lint into only seven grades, compared to the nine used to grade Egyptian cotton lint. It was decided to evaluate the impact of reducing the number of grades has on the prediction accuracy of models. Therefore, the transition grades (i.e., those containing the word “to” in Table 1, Section 2.1) were removed from the data. Although the granularity of the sensor would be reduced, it would still provide benefit by (1) improving the inspection efficiency by reducing the possible grades the cotton lint belongs to and (2) providing Egyptian cotton growers valuable information about the quality of their samples.

By reducing the number of grades, the classification models’ accuracies across all cultivars improved, as shown in Figure 6. Accuracies of greater than 95% were achieved by

cultivars Giza 86, 87 and 94, which is on par with classification accuracies reported by other image models built to grade cotton lint [11]. The accuracy of the models built to grade Giza 90 and 96 were 89.89% and 77.31% respectively. The ensemble models proved best at predicting cotton lint grade reporting the highest accuracy across all cultivars (Figure 6). This is in agreement with other recent work that has found ensemble learning to improve the accuracy to model cotton related applications [33]. Ensemble models generally have improved prediction accuracy when compared to single models because they reduce the model variance, which is the amount a model would change if developed using different training data [25].

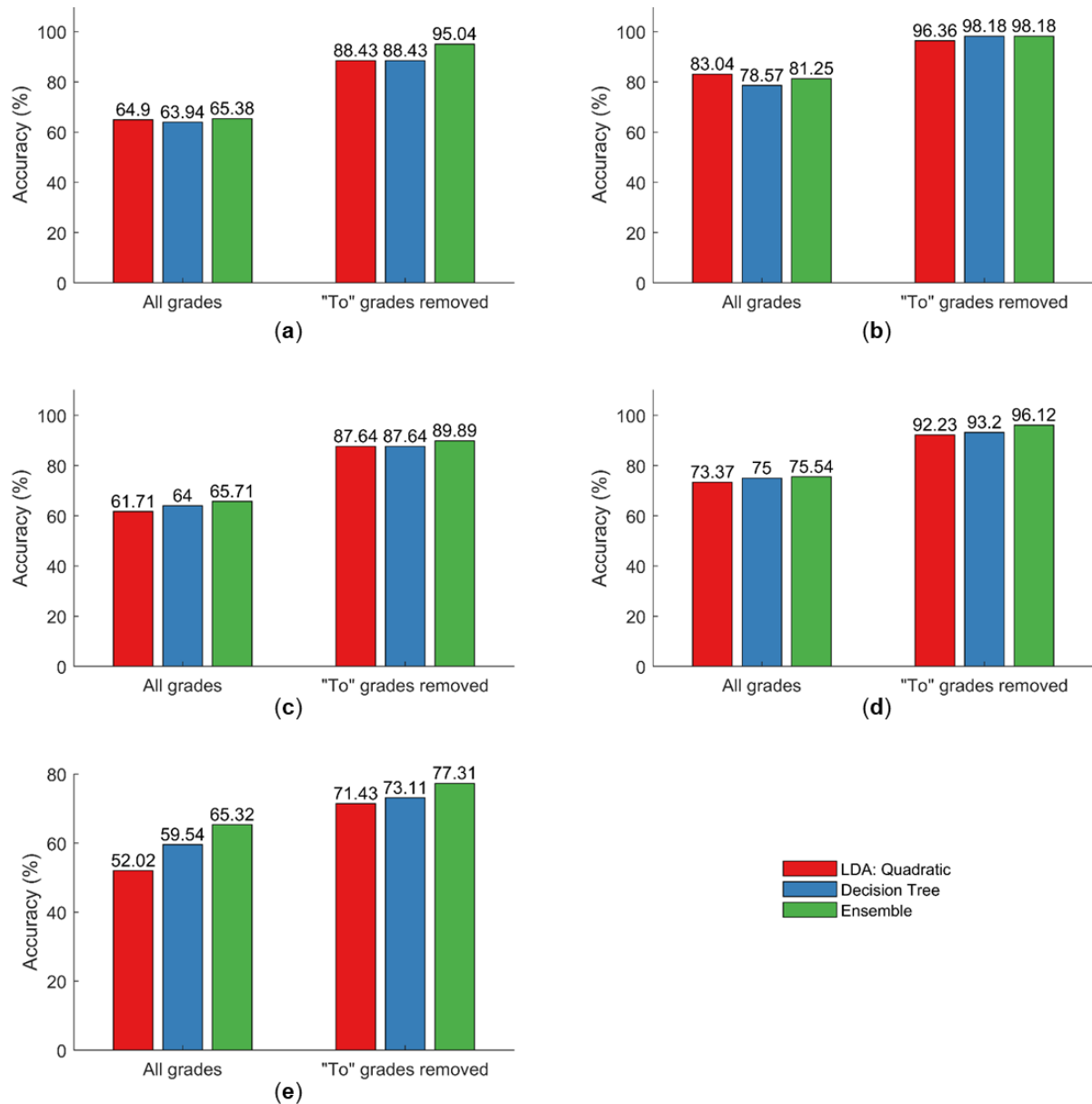


Figure 6 Accuracy of classification models of cotton lint developed for the cultivars(a) Giza 86, (b) Giza 87, (c) Giza 90, (d) Giza 94 and (e) Giza 96 evaluated using the test data

To understand the strengths and weakness of the ensemble models, the confusion matrixes were plotted for each culture in Figure 7. The column on the far right of the plot shows the percentages of all the examples predicted to belong to each class that are correctly classified. This metric is often referred to as the precision (or positive predictive value) and the false discovery rate referred to those that were incorrectly classified [34]. For example, Figure 7(A) shows that for cultivar Giza 86 the model was 90% precise at

predicting the images belonging to grade 7, but falsely labelled 3 images as grade 7 representing a 10% false discovery rate. The row at the bottom of the plot shows the percentages of all the examples belonging to each class that are correctly classified, known as the recall (or true positive rate) [34]. For example, the Giza 86 model correctly classified 89.2% of the image samples belonging to grade 5, and incorrectly labelled three as grade 7 and one as grade 3. This information can be used to identify which grades the models struggles to classify and may require additional data to help define. The bottom right percentage is the overall accuracy of the models.

Of all the cultivars, Giza 90 and 96 had the lowest classification accuracies (89.89% and 77.31% respectively). Therefore, the focus of the confusion plot analyses is on these cultivars to understand why and where they have underperformed, when compared to the models developed for other cultivars. Figure 7(C) shows that all the grades have a similar recall rate (88.0-90.9%), but a range of precisions (82.4-96.8%). The model has a low false discovery rate for grades 3 and 7 (3.2% and 8.3% respectively) meaning that when the model labels an image as grade 3 or 5 it is rarely false. Instead, the error occurs from incorrectly labelling images that belong to grades 3 and 7, as grade 5 (17.6% false discovery rate). The Giza 96 model was able to label 84.0% and 96.2% of the images belonging to grades 1 and 3 respectively correctly, higher than the 77%.3 overall classification accuracy. The error for this model occurred mainly when classifying images belonging to grades 5 and 9, where the model tended to incorrectly label the images belonging to grade 7. This meant the model was only 64.7% precise when classing images as grade 7. This analysis illustrates where the features extracted from the cotton lint images have been less effective at providing information to define the grades. Options to improve the accuracy of these models include either collecting and labelling additional samples of the grades with higher false-negative rates or exploring additional feature extraction methods that may help to define these grades.

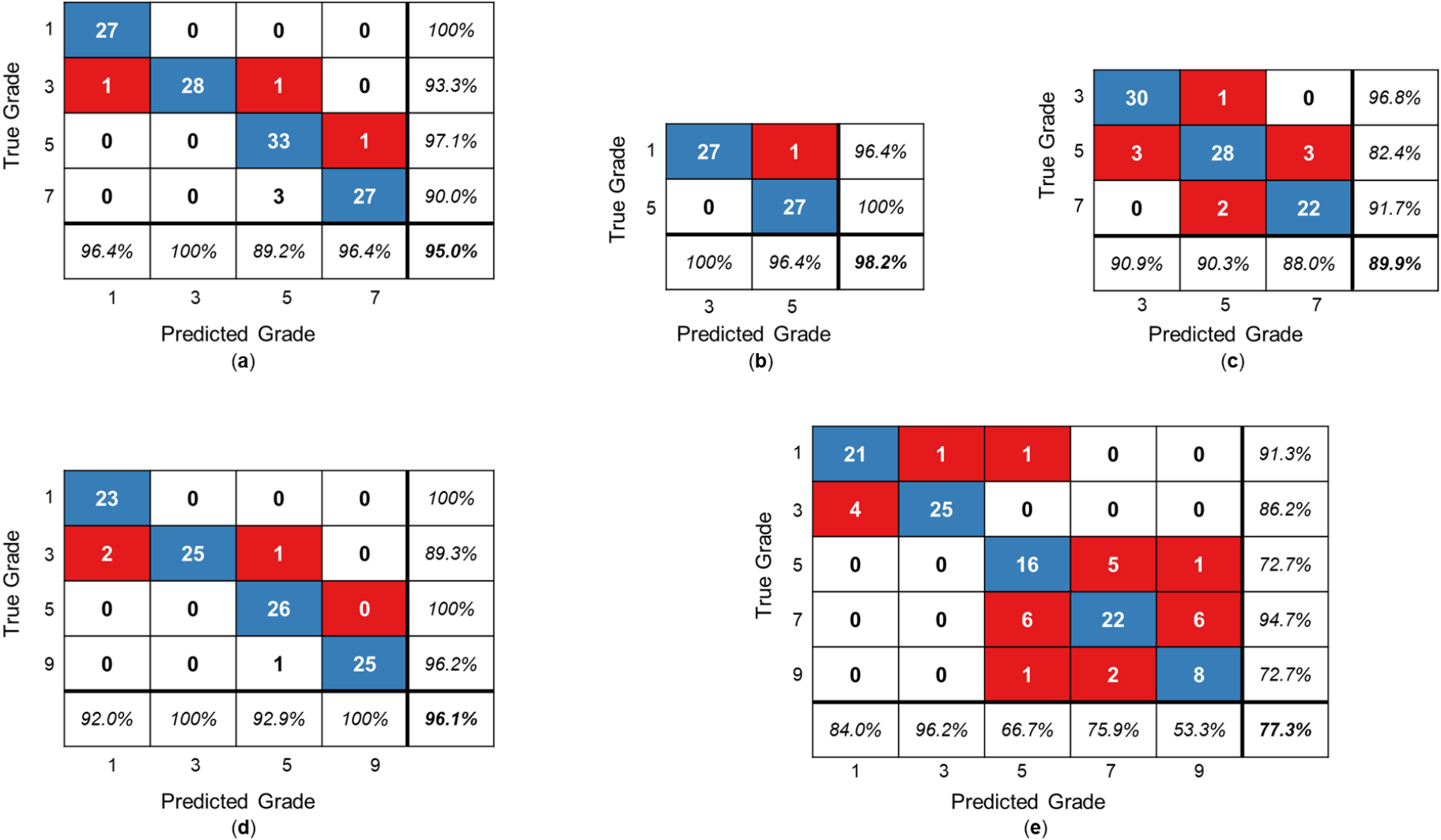


Figure 7 Confusion matrices of classifying cotton lint using ensemble models for the cultivars (a) Giza 86, (b) Giza 87, (c) Giza 90, (d) Giza 94 and (e) Giza 96.

3.2.2 Unsupervised machine learning modelling approach

Whilst the final ensemble models were able to achieve high accuracies (>95%) for cultivars Giza 86, 87 and 94, it was reliant on the collection of a labelled dataset. The cotton lint quality varies each harvest season due to seasonal fluctuations in environments, soil quality, biotic and abiotic stresses [22]. Therefore, models would require retraining each season on new labelled data in order to account for this variability. The labelling of data is typically an expensive and time-consuming task [35]. This is especially true for this instance, as grading of Egyptian cotton is a very intricate and complex subject, as it depends upon human perceptions of sight and touch and requires a high degree of critical judgement on the part of the officials responsible [9]. The complexity of this task leads to inevitable human error in the labelling of cotton lint grades. Unsupervised learning can be used to understand where human error is likely to occur by grouping together similar samples into clusters. Clusters that contain samples from multiple grades, indicates these grades show similar characteristics, in terms of percentage of impurity and colour, and are likely to be susceptible to human error. If the unsupervised clusters show a good correlation with the cultivar grades, it may also then be possible to use an unsupervised machine learning modelling approach to label future data, significantly reducing the models' development time and cost.

The K-means unsupervised algorithm was used to identify clusters in the feature data extracted from the cotton lint images. Similar to the supervised learning algorithm, the K-means data was developed on the training data and then applied to the testing data for evaluation. Plots of the labelled data were then compared to the K-means cluster plots to investigate if the clusters identified by the K-means models corresponded to the cotton lint grades, which are shown in Figure 8 and Figures S.5-8. Figure 8 shows that for Giza 86 the K-means algorithm grouped the data belonging to grades 1 and 3 into one cluster. Grouping these grades into one cluster resulted in the algorithm incorrectly splitting grade 7 into two clusters. This left cluster 4, which did show reasonable overlap with the data belonging to grade 5. The unsupervised analysis shows a high level of similarity between grades 1 and 3, when characterised by the percentage of impurity and colour features, which could cause a higher likelihood of human error when misclassifying cotton lint samples belonging to these grades. Figures S.7-8 show that the cultivars Giza 90, 94 and 96 also experienced the same problem of the K-means algorithm grouping the best grades together in one cluster, which then caused subsequent incorrectly identified clusters. The chance of human error appears relatively lower in the Giza 87 cultivar, as Figure S.5 indicates a good overlap between the K-means clusters and cotton grades for the Giza 87 cultivar. This likely explains the high accuracy of the supervised classification model predicting Giza 87 grades (98.2%). The success of the K-means labelling is likely attributed to the impurity feature showing a clear distinction between the grades, as previously reported in Section 3.1. This success comes with the caveat that only two grades were defined within the Giza 87 data.

Overall, the clusters defined by the unsupervised learning technique show that the manual inspection of Egyptian cotton is likely to contain human mislabelling, particularly in the higher quality grades. This highlights the need to develop an intelligent sensor, trained via objective labelling, to replace manual inspection. Additionally, as the clusters did not correspond to the labels within the cotton lint image data, relying on exclusively unsupervised learning is not a feasible option to label future data. Instead, it can be used alongside human labelling to identify potential errors and indicate new grade boundaries. Future work should explore the feasibility of other techniques, such as semi-supervised or active learning, to reduce the volume of training data required.

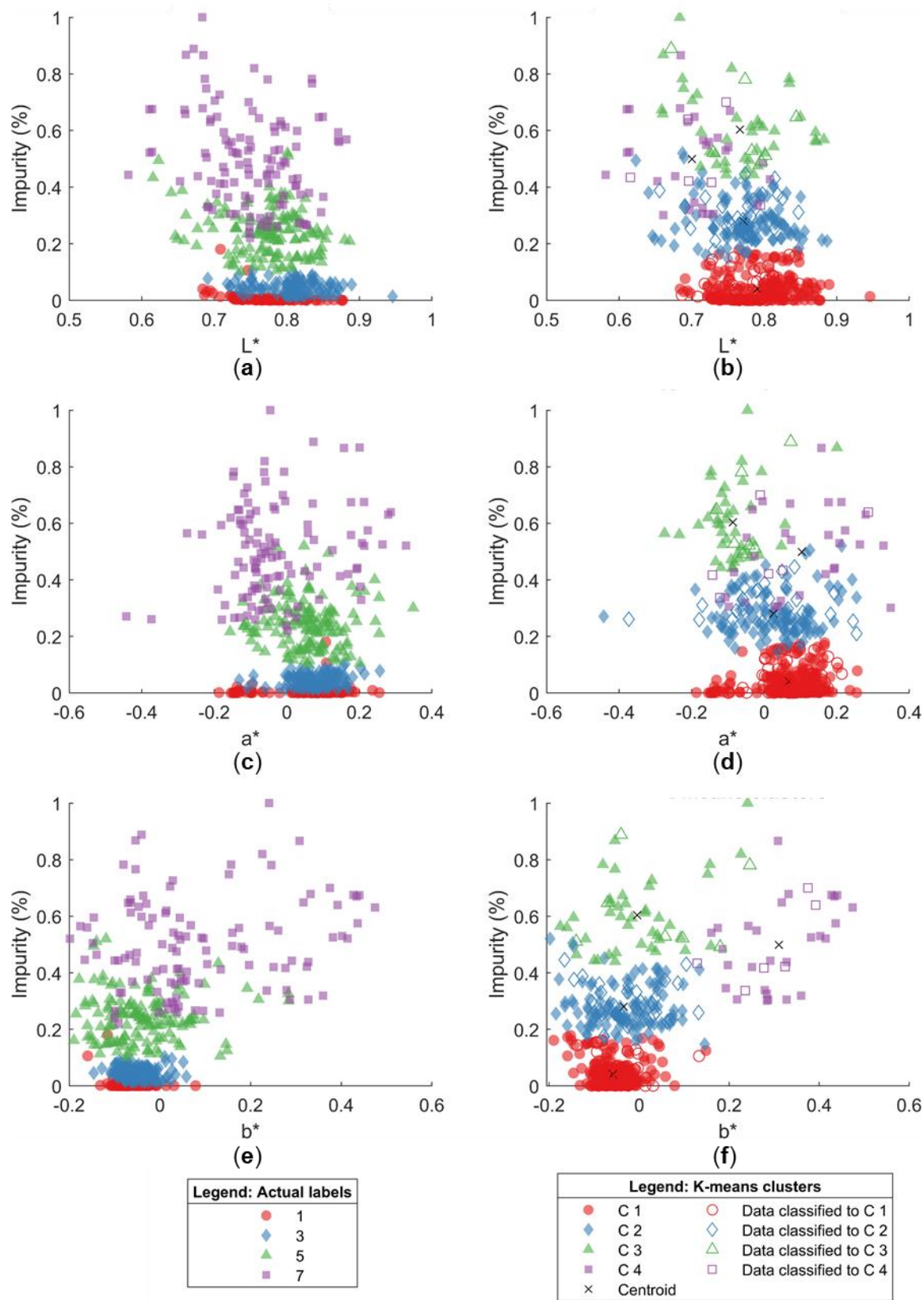


Figure 7 The feature impurity plotted against L^* (a,b), a^* (c,d) and b^* (e,f) features and labelled using the actual cotton lint labels (a, c and e) and the K-means clusters (b, d and f) for cultivar Giza 86.

This work has demonstrated that colour vision systems combined with machine learning models (intelligent sensors) have the potential to improve the productivity of the

Egyptian cotton processing industry. These could have a particular impact on SMEs cotton processors that cannot afford HVI instrumentation to grade their cotton lint. However, for the system to be commercially viable future research is required in the two areas.

- **Improve the classification accuracy of transitional grades.** As previously mentioned, the classification of the models decreased when the transitional grades were included within the training data. Future research should look into (a) either improving the feature extraction methods to increase the distinction between grades within the feature data or automating the feature extraction method via techniques like deep learning and (b) additional unsupervised learning to investigate and identify human labelling errors contained within the data.
- **Reduce the volume of labelled data:** The largest barrier to deploying the solution presented in this work, is the requirement to obtain and label cotton fibre samples to be used for training the classification models embedded in the colour vision system. This is because the labelling must be performed by Egyptian officials from the Cotton Arbitration & Testing General Organization (CATGO) in order for the training data, and therefore models, to be trusted by the Egyptian cotton industry. Future research should explore the techniques, such as semi-supervised learning and active learning, which are able to reduce the volume of labelled data required. Semi-supervised does this by combining a small amount of labelled data with learning from unlabelled data to train a model [36]. Multiple semi-supervised frameworks may apply to this work but are currently untested, including combining unsupervised with supervised learning [37], self-training [38], and co-training [38]. Alternatively active learning methods, which requests the user to label a data point if the model's confidence in its prediction is below a specified confidence score [38]. Thus, the overall volume of data requiring labelling is reduced, as only the data that will be most useful to the model is labelled.
- **Transfer learning between cultivars:** Another option that may lead to both an increase in classification accuracy and a reduction in the overall volume of data required is to incorporate knowledge from each of the cultivars into one model, via transfer learning. Transfer learning, which is when the learning gained from one task is applied to a different but related problem [39].
- **Framework to share data between cotton lint processors:** Generally, the accuracy of a classification model increases as more data is made available to learn from. Consequently, if data from multiple Egyptian cotton lint processors were to be shared between one another, the overall accuracy of the system could be improved to the benefit of all users. However, cotton processors are unlikely to be willing to share information about their products or processes with their competitors. Future research should explore the possibility of using frameworks, such as federated learning, that are capable of sharing learning from multiple users without ever exposing the raw data belonging to each user with other users [40].

5 Conclusion

The growing, harvesting and processing of Egyptian cotton is an industry that still uses traditional manual processes and intelligent sensors have the potential to provide increased efficiency, sustainability, and productivity. Currently, cotton lint samples are graded by manual inspection, which has several drawbacks including significant labour requirement, low inspection efficiency, and influence from inspection conditions such as light. This work has presented an intelligent sensor, which would remove the need for manual classification of cotton lint samples. This work showed that an intelligent sensor that uses features extracted from a colour vision system, combined with machine learning models, is able to accurately grade the cotton lint samples. The highest accuracy models (77.3-98.2%) used an ensemble modelling technique to classify the cultivars Giza 86, 87, 90, 94 and 96 samples within the Egyptian cotton grades: Fully Good, Good, Fully Good Fair, Good Fair and Fully Fair. However, when the transitional grades were included the models' accuracy was reduced (65.3-81.3%). Unsupervised machine learning analysis

suggested where human error could have occurred during the supervised classification of cotton lint and highlighted the need for an intelligent automated system to replace manual inspection. Finally, four areas of future research are identified to progress the development of the system so that it is fit for commercial use:

- Improve the classification accuracy of transitional grades;
- Reduce the volume of labelled data;
- Transfer learning between cultivars;
- Framework to share data between cotton processors.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Figure S1: Distribution of the cultivar Giza 87 extracted features (A) L*, (B) a*, (C) b* and (D) impurity data across the Egyptian cotton lint grades, Figure S2: Distribution of the cultivar Giza 90 extracted features (A) L*, (B) a*, (C) b* and (D) impurity data across the Egyptian cotton lint grades, Figure S3: Distribution of the cultivar Giza 94 extracted features (A) L*, (B) a*, (C) b* and (D) impurity data across the Egyptian cotton lint grades, Figure S4: Distribution of the cultivar Giza 96 extracted features (A) L*, (B) a*, (C) b* and (D) impurity data across the Egyptian cotton lint grades, Figure S5: The feature impurity plotted against L* (A,B), a* (C,D) and b* (E,F) features and labelled using the actual cotton lint labels (A, C and E) and the K-means clusters (B, D and F) for cultivar Giza 87, Figure S5: The feature impurity plotted against L* (A,B), a* (C,D) and b* (E,F) features and labelled using the actual cotton lint labels (A, C and E) and the K-means clusters (B, D and F) for cultivar Giza 87, Figure S6: The feature impurity plotted against L* (A,B), a* (C,D) and b* (E,F) features and labelled using the actual cotton lint labels (A, C and E) and the K-means clusters (B, D and F) for cultivar Giza 90, Figure S7: The feature impurity plotted against L* (A,B), a* (C,D) and b* (E,F) features and labelled using the actual cotton lint labels (A, C and E) and the K-means clusters (B, D and F) for cultivar Giza 94, Figure S8: The feature impurity plotted against L* (A,B), a* (C,D) and b* (E,F) features and labelled using the actual cotton lint labels (A, C and E) and the K-means clusters (B, D and F) for cultivar Giza 96.

Author Contributions: Conceptualization, Ahmed Rady, Aly A. A. El-Banna and Haitham H. Emaish; methodology, Oliver J. Fisher, Ahmed Rady, Aly A. A. El-Banna, Nicholas J. Watson and Haitham H. Emaish; software, Oliver J. Fisher and Ahmed Rady; validation, Oliver J. Fisher, Ahmed Rady, Aly A. A. El-Banna, Nicholas J. Watson and Haitham H. Emaish; formal analysis, Oliver J. Fisher, Ahmed Rady and Nicholas J. Watson; investigation, Oliver J. Fisher, Ahmed Rady, Aly A. A. El-Banna, Nicholas J. Watson and Haitham H. Emaish; resources, Oliver J. Fisher, Ahmed Rady, Aly A. A. El-Banna, Nicholas J. Watson and Haitham H. Emaish; data curation, Ahmed Rady, Aly A. A. El-Banna and Haitham H. Emaish; writing—original draft preparation, Oliver J. Fisher; writing—review and editing, Oliver J. Fisher, Ahmed Rady and Nicholas J. Watson; visualization, Oliver J. Fisher; supervision, Nicholas J. Watson; project administration, Ahmed Rady and Nicholas J. Watson; funding acquisition, Nicholas J. Watson. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Engineering and Physical Sciences Research Council (EPSRC) [EP/S036113/1], Connected Everything II: Accelerating Digital Manufacturing Research Collaboration and Innovation.

Conflicts of Interest:

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Yu, C. Chapter 2 - Natural Textile Fibres: Vegetable Fibres. In *Woodhead Publishing Series in Textiles*; Sinclair, R.B.T.-T. and F., Ed.; Woodhead Publishing, 2015; pp. 29–56 ISBN 978-1-84569-931-4.
2. Dhehibi, B.; El-Shahat, A.A.I.A.; Frija, A.; Hassan, A.-A. Growth in total factor productivity in the Egyptian agriculture sector: Growth accounting and econometric assessments of sources of growth. *Sustain. Agric. Res.* **2016**, *5*, p38, doi:10.5539/SAR.V5N1P38.
3. Ahmed, Y.N.; Delin, H. Current situation of Egyptian cotton: Econometrics study using ARDL model. *J. Agric. Sci.* **2019**, *11*, 88–97, doi:10.5539/jas.v11n10p88.

4. Abdel-Salam, M.E.; Negm, M.; Ardabb, C.S. The Egyptian Cotton; current constraints and future opportunities. *Text. Ind. Hold. Co., Mod. Press.* **2009**.
5. Gallico, D. The Cotton Value Chain: Improving Its Sustainability, Inclusiveness and Value Adding Capabilities: Italy-Egypt Collaborative Sustainability-Based Innovation. In *Sustainability Awareness and Green Information Technologies*; Issa, T., Issa, T., Issa, T.B., Isaias, P., Eds.; Springer Nature, 2020; pp. 199–214.
6. Food and Agriculture Organization of the United Nations Crops and livestock products Available online: <http://www.fao.org/faostat/en/#data/TP> (accessed on Jun 23, 2021).
7. Wally, A.; Tate, B. *2018 Cotton and products annual*; 2018;
8. The United Nations Industrial Development Organisation From cotton seeds to clothing: Enhancing the sustainability, inclusiveness and value addition of the cotton value chain in Egypt Available online: https://open.unido.org/projects/EG/projects/160068?_ga=2.69868693.447703373.1612258725-931943205.1612258725 (accessed on Feb 2, 2021).
9. Hussein, M.G.; El-Marakby, A.M.; Tolb, A.M.; Amal, Z.A.M.; Ebaido, I. Relationship between fiber cotton grade and some related characteristics of long and extra-long staple Egyptian cotton varieties (*Gossypium barbadense* L). *Arab Univ. J. Agric. Sci.* **2020**, *28*, 191–205, doi:10.21608/ajs.2020.28192.1197.
10. Wang, X.; Yang, W.; Li, Z. A fast image segmentation algorithm for detection of pseudo-foreign fibers in lint cotton. *Comput. Electr. Eng.* **2015**, *46*, 500–510, doi:https://doi.org/10.1016/j.compeleceng.2015.06.022.
11. Lv, Y.; Gao, Y.; Rigall, E.; Qi, L.; Gao, F.; Dong, J. Cotton appearance grade classification based on machine learning. *Procedia Comput. Sci.* **2020**, *174*, 729–734, doi:10.1016/j.procs.2020.06.149.
12. Majumdar, A.; Mangla, R.; Gupta, A. Developing a decision support system software for cotton fibre grading and selection. *Indian J. Fibre Text. Res.* **2010**, *35*, 195–200.
13. Shankar, C.; Siddhartha, B. Cotton fibre selection and grading – a PROMETHEE-GAIA-based approach. *Int. J. Cloth. Sci. Technol.* **2017**, *29*, 646–660, doi:10.1108/IJCST-12-2016-0137.
14. Ballard, Z.; Brown, C.; Madni, A.M.; Ozcan, A. Machine learning and computation-enabled intelligent sensor design. *Nat. Mach. Intell.* **2021**, *3*, 556–565, doi:10.1038/S42256-021-00360-9.
15. Pathan, M. V.; Ponnusami, S.A.; Pathan, J.; Pitongsawat, R.; Erice, B.; Petrinic, N.; Tagarielli, V.L. Predictions of the mechanical properties of unidirectional fibre composites by supervised machine learning. *Sci. Rep.* **2019**, *9*, 13964, doi:10.1038/s41598-019-50144-w.
16. Krishnakumar, P.; Rameshkumar, K.; Ramachandran, K.I. Acoustic emission-based tool condition classification in a precision high-speed machining of titanium alloy: a machine learning approach. *Int. J. Comput. Intell. Appl.* **2018**, *17*, 1850017.
17. Amiri, N.; Farrahi, G.H.; Kashyzadeh, K.R.; Chizari, M. Applications of ultrasonic testing and machine learning methods to predict the static & fatigue behavior of spot-welded joints. *J. Manuf. Process.* **2020**, *52*, 26–34, doi:https://doi.org/10.1016/j.jmapro.2020.01.047.
18. Mulrennan, K.; Donovan, J.; Creedon, L.; Rogers, I.; Lyons, J.G.; McAfee, M. A soft sensor for prediction of mechanical properties of extruded PLA sheet using an instrumented slit die and machine learning algorithms. *Polym. Test.* **2018**, *69*, 462–469, doi:10.1016/J.POLYMERTESTING.2018.06.002.
19. Huang, M.-L.; Fu, C.-C. Applying image processing to the textile grading of fleece based on pilling assessment. *Fibers* **2018**, *6*, 73.
20. Zhang, X.; Xin, B.; Zheng, Y.; Shi, M.; Lin, L.; Gao, C.; Yi, Y.; Yang, Z.; Li, H. Fiber recognition with machine learning methods by fiber tensile fracture via acoustic emission method. *Text. Res. J.* **2020**, *90*, 2552–2563, doi:10.1177/0040517520924130.
21. Bishop, C.M. *Pattern recognition and machine learning (information science and statistics)*; Springer-Verlag: Berlin, Heidelberg, 2006; ISBN 0387310738.
22. Kothari, N.; Dever, J.; Hague, S.; Hequet, E. Evaluating intraplant cotton fiber variability. *Crop Sci.* **2015**, *55*, 564–570, doi:10.2135/CROPSCI2014.01.0077.

23. Rafael C. Gonzalez; Richard E. Woods *Digital image processing: Pearson international edition*; 3rd ed.; Pearson, 2008;
24. Vergés Llahí, J. Color Constancy and Image Segmentation Techniques for Applications to Mobile Robotics, Universitat Politècnica de Catalunya, 2005.
25. Rokach, L. *Ensemble learning: Pattern classification using ensemble methods*; 2nd ed.; World Scientific: Singapore, 2019;
26. Probst, P.; Boulesteix, A.-L. To tune or not to tune the number of trees in random forest? *J. Mach. Learn. Res.* **2017**, *18*, 1–8.
27. Louppe, G.; Louis Wehenkel; Antonio Suter; Pierre Geurts Understanding variable importances in forests of randomized trees. In Proceedings of the Proceedings of the 26th International Conference on Neural Information Processing Systems; Burges, C.J.C., Bottou, L., Welling, M., Ghahramani, Z., Weinberger, K.Q., Eds.; Curran Associates Inc.: Red Hook, US, 2013; pp. 431–439.
28. Al-Fattah, S.M.; Al-Naim, H.; Al-Naim, H.A. Artificial-intelligence technology predicts relative permeability of giant carbonate reservoirs. *SPE Reserv. Eval. Eng.* **2009**, *12*, 96–108, doi:10.2118/109018-PA.
29. Thakur, G.S.; Daigle, B.J.; Qian, M.; Dean, K.R.; Zhang, Y.; Yang, R.; Kim, T.-K.; Wu, X.; Li, M.; Lee, I.; et al. A Multimetric Evaluation of Stratified Random Sampling for Classification: A Case Study. *IEEE Life Sci. Lett.* **2016**, *2*, 43–46, doi:10.1109/lsls.2016.2615086.
30. Kim, S. *MATLAB Deep learning with machine learning, neural networks and artificial intelligence*; 1st ed.; Apress, 2017; ISBN 978-1-4842-2844-9.
31. Heng, C.; Chen, L.; Shen, H.; Wang, F. Study on the measurement and evaluation of cotton color using image analysis. *Mater. Res. Express* **2020**, *7*, 75101, doi:10.1088/2053-1591/aba3e4.
32. Matusiak, M.; Walawska, A. Important aspects of cotton colour measurement. *FIBRES Text. East. Eur.* **2010**, *18*, 17–23.
33. Fue, K.G.; Porter, W.M.; Barnes, E.M.; Rains, G.C. Ensemble Method of Deep Learning, Color Segmentation, and Image Transformation to Track, Localize, and Count Cotton Bolls Using a Moving Camera in Real-Time. *Trans. ASABE* **2021**, *64*, 341–352, doi:10.13031/trans.13112.
34. Alvarez, S.A. An exact analytical relation among recall, precision, and classification accuracy in information retrieval. *Bost. Coll. Boston, Tech. Rep. BCCS-02-01* **2002**, 1–22.
35. Sager, C.; Janiesch, C.; Zschech, P. A survey of image labelling for computer vision applications. *J. Bus. Anal.* **2021**, doi:10.1080/2573234X.2021.1908861.
36. Levatić, J.; Ceci, M.; Koccev, D.; Džeroski, S. Self-training for multi-target regression with tree ensembles. *Knowledge-Based Syst.* **2017**, *123*, 41–60, doi:10.1016/j.knosys.2017.02.014.
37. Forestier, G.; Wemmert, C. Semi-supervised learning using multiple clusterings with limited labeled data. *Inf. Sci. (Ny)*. **2016**, *361–362*, 48–65, doi:10.1016/j.ins.2016.04.040.
38. Sheikh Hassani, M.; Green, J.R. Multi-view Co-training for microRNA Prediction. *Sci. Rep.* **2019**, *9*, 1–10, doi:10.1038/s41598-019-47399-8.
39. Bowler, A.L.; Watson, N.J. Transfer learning for process monitoring using reflection-mode ultrasonic sensing. *Ultrasonics* **2021**, *115*, 106468, doi:10.1016/j.ultras.2021.106468.
40. Khan, L.U.; Saad, W.; Han, Z.; Hossain, E.; Hong, C.S. Federated learning for Internet of Things: Recent advances, taxonomy, and open challenges. **2020**, doi:10.1109/comst.2021.3090430.