

Artificial Intelligence in hair research: a proof-of-concept study on evaluating hair assembly features

Gabriela Daniels	Cosmetic Science Research Group, London College of Fashion, London, UK	g.n.daniels@fashion.arts.ac.uk
Slobodanka Tamburic	Cosmetic Science Research Group, London College of Fashion, London, UK	d.tamburic@fashion.arts.ac.uk
Sergio Benini	Department of Information Engineering, University of Brescia, Brescia, Italy	sergio.benini@unibs.it
Jane Randall	Sensory Dimensions Ltd, UK	jane@sensorydimensions.com
Tracey Sanderson	Sensory Dimensions Ltd, UK	tracey.sanderson@sensorydimensions.com
Mattia Savardi	Department of Information Engineering, University of Brescia, Brescia, Italy	m.savardi001@unibs.it

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/ICS.12706](#)

This article is protected by copyright. All rights reserved

MRS. GABRIELA DANIELS (Orcid ID : 0000-0002-1754-5184)

Article type : Original Article

Artificial Intelligence in Hair Research: a Proof-of-Concept Study on Evaluating Hair Assembly Features

ABSTRACT

OBJECTIVE: The first objective of this study was to apply computer vision and machine learning techniques to quantify the effects of haircare treatments on hair assembly and to identify correctly whether unknown tresses were treated or not. The second objective was to explore and compare the performance of human assessment with that obtained from artificial intelligence (AI) algorithms.

METHODS: Machine learning was applied to a dataset of hair tress images (virgin and bleached), both untreated and treated with a shampoo and conditioner set, aimed at increasing hair volume whilst improving alignment and reducing the flyway of the hair. The automatic quantification of the following hair image features was conducted: local and global hair volumes and hair alignment. These features were assessed at three time points: t_0 (no treatment), t_1 (two treatments), t_2 (three treatments). Classifier tests were applied to test the accuracy of the machine learning. A sensory test (paired comparison of t_0 vs t_2) and an online front-image based survey (paired comparison of t_0 vs t_1 , t_1 vs t_2 , t_0 vs t_2) were conducted to compare human assessment with that of the algorithms.

RESULTS: The automatic image analysis identified changes to hair volume and alignment which enabled the successful application of the classification tests, especially when the hair images were grouped into untreated and treated groups. The human assessment of hair presented in pairs confirmed the automatic image analysis. The image assessment for both virgin hair and bleached only partially agreed with the analysis of the subset of images used in the online survey. One hypothesis is that treatments changed somewhat the shape of the hair tress, with the effect being more pronounced in bleached hair. This made human assessment of flat images more challenging than when viewed directly in 3D. Overall, the bleached hair exhibited effects of higher magnitude than the virgin hair.

CONCLUSIONS: This study illustrated the capacity of artificial intelligence for hair image detection and classification, and for image analysis of hair assembly features following treatments. The human assessment partially confirmed the image analysis, and highlighted the challenges imposed by the presentation mode.

Keywords: artificial intelligence, bleached hair, virgin hair, sensory assessment, machine learning, hair detection, hair segmentation

Introduction

The shape, dimensions and mechanical properties of scalp hair fibres have an important impact on the overall individual's appearance. To modify this appearance according to changing personal preferences, consumers can select from a large array of products. Assessing the suitability of the products for consumers with different types of hair is an essential part of the process of developing cosmetics. Such assessments are commonly conducted on hair swatches, also referred to as tresses, prepared from human hair that closely represent the target consumer.

A typical hair fibre of individuals from European descent (Caucasian hair) is often elliptical, with a mean diameter of $70\mu\text{m}$ (1), corresponding to a perception of fine hair as defined by a study dedicated to fine hair evaluation (2). Furthermore, approximately 50% of Caucasian hair can be classified as straight and wavy (3), which, in combination with fibre size, may generate a visual impression of a flat, limp hair. Therefore, consumers with such type of hair would seek products which cause changes to individual fibres that generate an overall appearance of more hair, whilst hair remains in the desired shape and appearing well-conditioned.

Hair assembly volume, alignment and flyaway

Hair volume, sometimes referred to as "body", is of particular relevance to Caucasians with straight fine hair. Due to the lack of curvature, the main inherent contributors to the volume effect of the "head of hair" are the density of fibres on the scalp, the shape of fibres and the average fibre diameter. These factors are, in turn, dependent on the size, shape and function of the hair follicle and are not typical targets of cosmetic treatments. On the other hand, the physical and optical properties of the individual fibres could affect the appearance of the hair assembly in terms of its volume; such interventions are cosmetic by definition.

The individual fibre properties related to the hair volume are bending and torsional stiffness, as well as surface roughness, which promotes inter-fibre friction. Low bending elasticity (high bending stiffness) and high torsional stiffness of the fibre are expected to contribute positively to the assembly volume (4).

Bending elasticity is dependent on the fibre cross-sectional area and shape (degree of ellipticity). Bending stiffness is low for elliptical fibres and increases substantially with the increased size of the fibre minor diameter, hence more round fibres are stiffer. In addition, the cuticle has been found to contribute up to 50.9% of the bending stiffness of hair (Caucasian and Asian), with high stiffness creating hair thickening effect (5)(6). Another study, comparing the bending Young's moduli of Japanese and Caucasian hair, found no difference despite the different fibre shapes, but found a decreased stiffness in fibres of individuals from the older groups who had perceptions of low volume hair appearance (7). Such inconsistency in results between studies could be attributed to the different methods used for the

bending stiffness measurements and calculations. Fibre stiffness has also been reported as an end point for measuring the effect of styling polymer treatments as well as for treatments with actives anticipated to have penetrated the fibre structures following a prolonged contact time. The instruments and methods used vary, from measuring bending force (8) through loop deformation force (9) to measuring bending angle due to gravitational force (10). Atomic Force Microscopy nano-indentation and nano-scratch methods have been used to examine cuticle hardness and elasticity *in situ*, concluding that chemical and conditioning treatments cause cuticle softening in Caucasian hair (11). In summary, the hair fibre cuticle hardness/softness appears to be of a direct relevance to volume formation.

The torsional properties of the hair fibre were firstly studied in the context of hair setting and perming, as high rigidity could better support the acquired style. The cross-sectional area of the fibre has a notable effect, with smaller fibres having lower rigidity, translating into lower capacity to maintain style (12). Chemical treatments such as bleaching were found to increase torsional modulus of dry hair. The cuticle and cortex contributions to this effect were estimated theoretically, based on an assumption of the cuticle thickness (appropriate to the hair type and area tested), and the cuticle torsional module was found to be 6 times larger than that of the cortex, thus causing the increased torsional stiffness of the fibres after bleaching (13).

Therefore, increasing the assembly volume of fine hair requires treatments which increase cuticle and cortex resistance to bending and torsional deformation, as well as the inter-fibre friction. At the same time, in line with prevalent fashion styles, there exist an expectation for the hair assembly to also look smooth and lustrous, which requires high fibre alignment.

Fibre alignment, resulting from fibres falling in a parallel orientation, is expected to be inversely related to the volume effect in fine hair. This is because more spaces and cross-junctions between fibres, supported by higher inter-fibre friction, will generate higher assembly volume. Inter-fibre friction cannot be measured directly; however, the condition of the cuticle is a main contributor to surface friction, and the more uneven the cuticle surface is, the more friction will be generated at the points where fibres make contact. Cosmetic products can increase or decrease the fibre surface friction, depending on the desirable effect, and by doing that, will increase or decrease the inter-fibre friction, too.

In addition, the generation of electrostatic charges on dry hair surface, referred to as “flyaway”, has an unappealing effect on straight fine hair. Commonly, flyaway occurs after contact between dry hair surface and a comb or fabric, while the coating the hair fibres with hydrophobic materials during shampooing and post-wash conditioning elevates this condition.

In summary, the volume and fibre alignment effects are expected to be inversely related, with the cuticle condition and its interaction with cosmetic treatments making a substantial contribution to the overall effect. The early volume/body assessment methods involved measuring the work involved while

vertically pulling hair through templates with decreasing size of their openings, thus assessing the resistance of the hair bulk to compression (14). Other methods use laser stereometry, by scanning hair tresses laid flat and generating 3d surface plots (15). Currently, the high quality image analysis methods are prevalent, with direct or back illumination being used to generate images for volume and flyaway assessment, whilst polarised light illumination has been used for alignment (16). Finally, sensory assessments of hair are conducted in order to ensure that such instrumentally measured effects are perceivable by trained and, ultimately, naïve panels who represent the average consumer.

Hair assembly properties and Artificial Intelligence

Recently, different applications of Artificial Intelligence (AI), such as computer vision and machine learning algorithms, have been utilised in the cosmetics industry for the automated detection and analysis of image-based data, mostly as a face skin diagnostic tool (17, 18, 19). In those assessments, the AI was validated by parallel assessments conducted by expert dermatologists, who normally use validated scales for skin ageing assessment.

So far, the published research on hair detection and classification has focused on the automatic segmentation and classification of hair into hairstyles e.g. long, short, straight/curly, dreadlocks (20), as well as the construction of the 3D structure of a head of hair based on a single image (21). Hence, there is a lack of studies on AI methods designed for the assessment of small assembly alterations, which are more appropriate reflections of the consumer experience following the routine usage of hair shampoos and conditioners. At the same time, all research on machine learning in the cosmetics industry has been carried out via proprietary datasets (not publicly available).

This study had two objectives. *The first objective was to apply computer vision and machine learning techniques to quantify the effects of haircare treatments on hair assembly and to identify correctly whether unknown tresses were treated or not.* The outcomes of these analyses were compared with the **expected theoretical behaviour (ETB)** of the treatment and were envisaged to offer more accurate methods for the assessment of hair assembly. As the experiment was based on fine straight hair, the **ETB** of the treatments was as follows:

- the treatments would cause a hair volume increase, a decrease in fibre alignment and a reduction of flyaway;
- the virgin and bleached hair would display the same behaviour, but the magnitude of the effects in bleached hair would be higher.

The rationale for these assumptions were as follows:

- A range of ingredients would deposit on the hair during the treatment. Some would cause hair stiffening to increase volume; some would cause surface smoothing to manage alignment and prevent flyway of dry hair. The relative magnitude of these opposing effects would be balanced in favour of volume; but good alignment and low flyway were expected to be maintained in order to produce an orderly hair assembly;

- Bleached hair's surface proteins were expected to display higher electric charge density, resulting in the increased electrostatic interactions between hair and charged ingredients in the cosmetic products and, therefore, higher treatment substantivity and effects were expected.

The second objective was to explore and compare the performance of human assessment with that obtained by AI algorithms. In order to do that, after performing automatic image analysis, we carried out a sensory test with naive assessors and an online survey based on tress images from the employed dataset. A general alignment of these two types of data would provide the proof-of-concept for this area of the artificial intelligence application. The proposed approach exploited a public repository¹ of 1080 hair images which we had published already (22). Furthermore, to allow replicability and encourage further research, we have released the code and other relevant materials.

Materials and Methods

Hair tresses and treatment

Two types of hair tresses were tested, virgin ($n_v = 60$) and lightly bleached ($n_b = 60$) Caucasian hair (Banbury Postiche, UK). Each tress was 10cm long and weighed 3g. Bleached hair was obtained by subjecting virgin hair to light oxidative treatment (Wella Professionals, UK). The products used for the hair treatment were a shampoo and a conditioner available on the high street in the UK and marketed as a system. The treatment products were chosen based on claims to give volume without producing tangles and weighing the hair down. In accordance to the European Union Cosmetic Regulation (23), such claims would be based on testing and comparing the combined effect of the shampoo and conditioner on fine hair, using the comparison of before and after treatment values. Therefore, the product choice was intended to reflect the ETB appropriate for the type of hair used in the experiment.

Each type of tress was subjected to a reproducible treatment process structured around three time points: t_0 =no treatment; t_1 =two consecutive treatments; t_2 =three consecutive treatments. Each treatment comprised the following steps: the application of 2g of a commercial shampoo to wet hair; working the shampoo into the hair for 20 seconds; rinsing under running water for 20 seconds; applying 2g of a conditioner; working the conditioner into the hair for 20 seconds; resting the hair for 2 minutes; rinsing the tresses for 20 seconds under running water; naturally drying for approximately 12 hours at 35° C. The treatment process aimed to replicate exposure times for the products typical for the consumer use and to eliminate the influence of blow drying on the measured effects.

Image data set

Three images of each hair tress, from three different angles (-45° , 0° , $+45^\circ$) were taken at each treatment point t_0 , t_1 , and t_2 (Figure 1), resulting in 9 images per tress (image size: 3264pixels × 2448 pixels). Each tress was combed twice just before the photograph was taken. A total of 1080 images were then used for analysis (540 for each hair type).

¹ <https://data.mendeley.com/datasets/37dkygd6sm/2>

Automatic hair segmentation

Automatic hair analysis, consisting of hair detection and segmentation, was performed by the method published in (20). This framework was able to detect the presence of human hair in images taken from unconstrained view by relying only on image textures, without a priori information on head shape and location or using body-part classifiers. Implementing a coarse-to-fine method, this approach first derived a hair probability map by classifying image elements into hair vs non-hair, the overlapping image patches described by features extracted from a Convolutional Neural Network (CNN). The network was trained using data in Figaro1K (20), a dataset containing more than 1000 annotated hair images. The result of such binary patch classification, which employs Random Forest classifier (24), is shown in Figure 1.

The accurate model for the specific subject's image, hair vs non-hair, was created from high vs low probability regions. As shown in Figure 1, finer segmentation happens only in uncertain areas, and it was performed at pixel level by using Linear Ternary Pattern (LTP) features (25) and Support Vector Machine (SVM) classifier (26). The segmentation accuracy achieved by this method on the Figaro1K dataset (around 90%) was superior to the known state-of-the-art. Examples of hair segmentation masks obtained on images with virgin (first pair) and bleached (second pair) tresses are shown in Figure 2 (original on the left, segmentation mask on the right).

Automatic quantification of hair assembly features

Based on the above, the **hair volume** (or “body”) and **fibre alignment** (or “straightness”) of the hair assembly were identified as the two key endpoints of the treatments. The following extracted features were therefore automatically analysed and computed as described below, starting from the segmented hair mask:

Hair volume

- **Global Hair Volume (GHV):** mean volume (in pixels) across the three views of each tress: (-45° , 0° , $+45^\circ$);
- **Local Hair Volume (LHV):** evaluation (in pixels) of the hair volume in thirds of the vertical length of the segmented image (upper 1/3, middle 1/3, and lower 1/3 of the tress, as shown in Figure 3) and averaged between the three views.

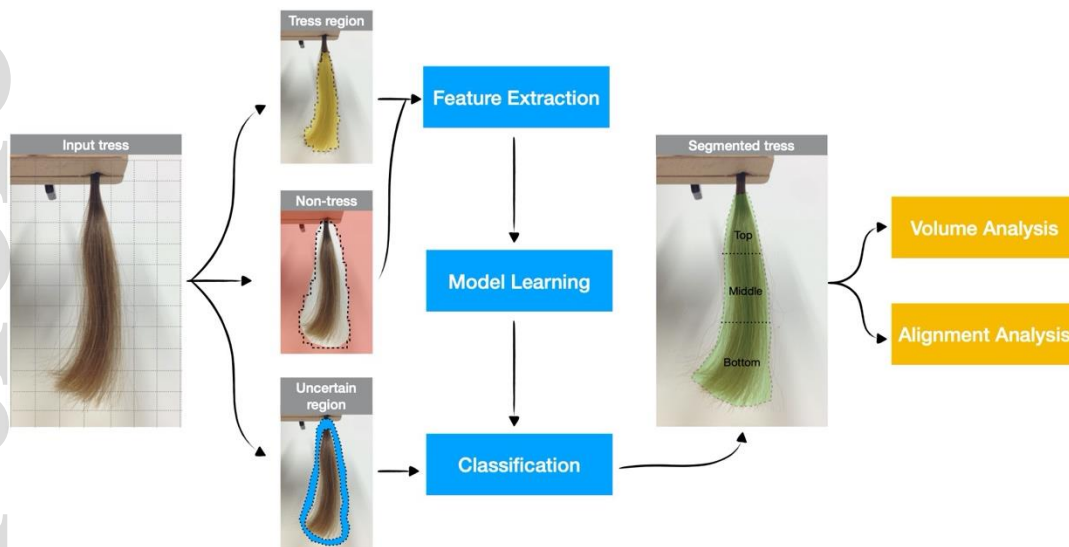


Figure 1: Segmentation workflow. Coarse segmentation: (left) a hair probability map was created on image patches for identifying high probability regions with hair vs. non-hair; as a result, a high-probability hair region was returned (top, in yellow), together with a high-probability non-hair region (middle, red). Finer segmentation was achieved by developing a more specific subject's hair model, by classifying image pixels in the uncertain area only (in blue, bottom) thus refining the final hair region (right, in green).

Fibre alignment

- **Histogram of Oriented Gradients (HOG):** originally developed for pedestrian detection (27), HOG technique counts occurrences of gradient orientation in image local regions. As shown in Figure 4, it was extracted from hair segmentation mask on a squared sliding window, using the following parameters: 8 orientation bins, 64×64 pixels per cell, and 1×1 cells per block. The final straightness index was calculated as the standard deviation of the histogram of oriented gradients (HOG) of the tress.

Treatment order test on hair image pairs with AI

The computed features (global and local hair volumes, HOG) for numerically quantifying the effects of haircare treatments on hair assembly, could also be used to perform **the treatment order test**. This is the classification experiment on images of the same tress taken at two different timepoints, each from 3 different angles (-45° , 0° , $+45^\circ$) for a total of 6 views. This test tried to answer the following question: *could images of the same tress from two unknown treatment points be correctly ordered by an AI algorithm?*

A Support Vector Machine (SVM) classifier was adopted for such test after a proper model selection procedure. It uses a feature vector composed by the concatenation of the three features $f = [\text{GVH}, \text{LHV}, \text{HOG}]$ from the three views of the two images, which was first normalised and reduced in dimensions by a PCA (Principal Component Analysis) algorithm (28). The dataset was composed of tress image pairs (A,B), with 3 views each, and with an associated binary label (0,1) that specified whether the tresses appeared in the correct treatment order. All SVM parameters are tuned by means of 5-fold cross-validation. Since tresses were 60 for each type (virgin and bleached), the possible pairs (t_0 vs t_1 , t_0 vs t_2 , and t_1 vs t_2) were

$60 \times 3 = 180$. The training process was carried out on 70% of the dataset (126 pairs for each type), while the results were evaluated on the remaining, never seen 30% (54 pairs for each type).



Figure 2: Examples of hair segmentation masks obtained from images with virgin (the two on the left) and bleached tress (the two on the right).

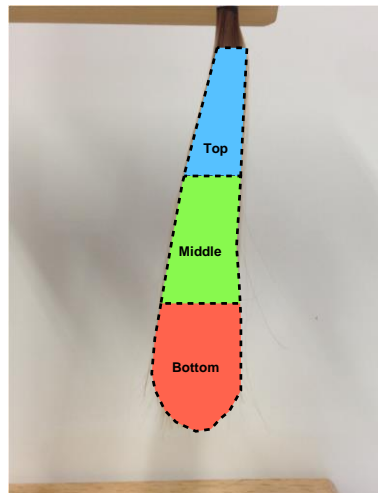


Figure 3: Tress subdivision in thirds for the computation of Local Hair Volume.

Timepoint recognition on single hair images with AI

By using the same feature vector f , we also performed the following classification experiment on single hair images (from the three views): **the timepoint recognition test**. This test tries to answer the following question: *could a tress image from unknown treatment point be assigned to the correct treatment point?*

Specifically, considering $A=3$ views of a tress taken at an unknown time, the task required to estimate whether A was photographed at timepoint t_0 , t_1 or t_2 . In this case the dataset was composed by a single feature vector (from the three views) with the label expressing the correct number of treatment cycles.

Through a procedure of model selection, we adopted Naïve Bayes (29) as the best classifier. All the parameters for the model selection were tuned by means of 5-fold cross-validation. As in the previous scenario, the training process was carried out on 70% of the dataset (42 images for each type), while the results were evaluated on the remaining, never seen 30% (18 images for each type).

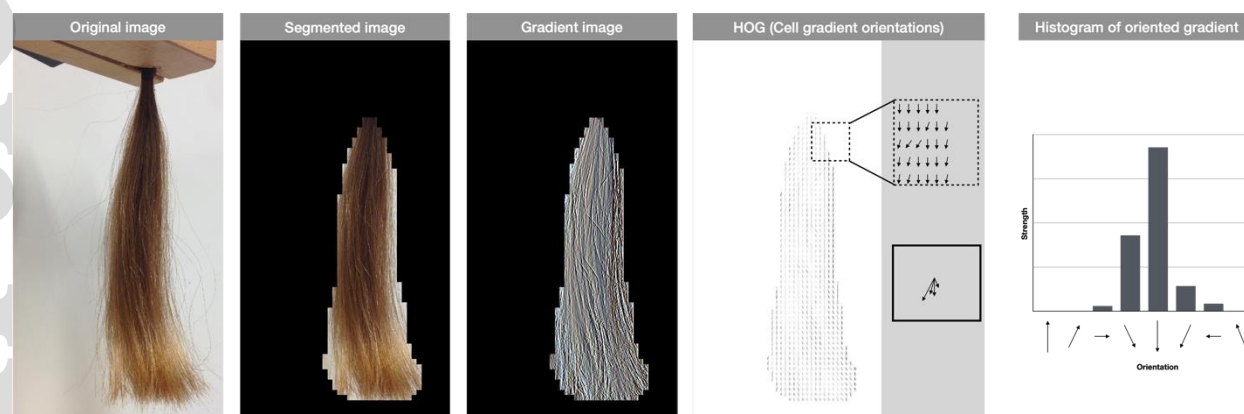


Figure 4: Histogram of oriented gradients (HOG) descriptor, from left to right: the original image, the segmented mask, the related HOG feature, a detail on gradient orientation and the final HOG feature.

Online paired image-comparison test with naïve assessors (n=100)

An online survey was conducted using Qualtrics XM (SAP, USA). A subset made of the frontal (0°) images of 10 virgin and 10 bleached tresses, representing the three time points (total of 60 images), was randomly selected for this survey. Virgin and bleached hair were compared in pairs, representing the combinations of all time points. The images were trimmed using Adobe Photoshop CC2018 crop tool and further resized in a consistent manner using the Qualtrics XM image manipulation tool.

Each assessor was asked to view 2 bleached and 2 virgin tresses, which were presented (see Fig. 5) in pairs of: t_0 vs t_1 , t_0 vs t_2 , t_1 vs t_2 for each tress respectively, and to answer the following questions: 1) *Which tress has more volume?* 2) *Which tress was straighter?* 3) *Which tress has more flyway?* The positions of the tresses were randomised, so for each assessor half of the viewed pairs had the earlier time point tress appearing on the left and half on the right.

Paired difference test with naïve assessors (n=50)

A pair of each hair type representing t_0 and t_2 was suspended on a horizontal bar positioned within a colour assessment cabinet, under illumination D65 (VeriVide, UK) (Figure 6). Each assessor was asked to view one pair of virgin and one pair of bleached hair tresses, while the presentation of t_0 and t_2 tresses (left and right) was randomised between assessors. A total of 120 tresses were prepared for this experiment: 30 in each category (virgin t_0 and t_2 , bleached t_0 and t_2); some of them were used more than once in a randomised manner.

Each tress was combed through twice just before the assessment was made. The assessors were instructed to freely view, but not to touch the hair and to answer the following three questions: 1) *Which tress has more volume?* 2) *Which tress was straighter?* 3) *Which tress has more flyaway?*

Statistical analysis

The hair volume data was analysed by comparing each portion's volumes and the GHVs at the three time points respectively by ANOVA and Tukey HSD tests, using StatsModels 0.12.2 (Python Software Foundation, USA), with $p < 0.05$ threshold being considered significant.

The sensory and the online survey data were analysed by a binomial distribution using Microsoft Excel, (Microsoft, USA). The probability of $p < 0.05$ was taken as a threshold for statistical significance.

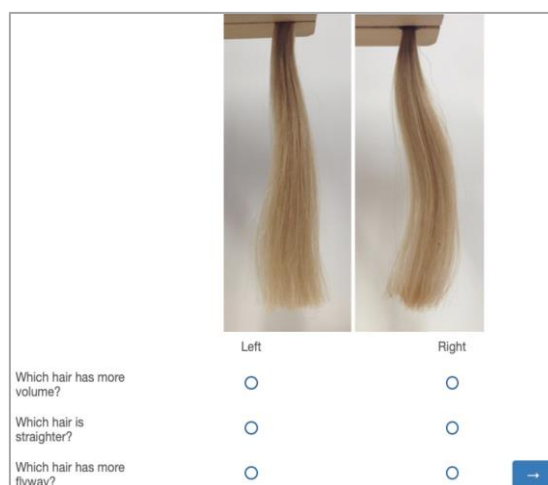


Figure 5: The online survey screen appearance.



Figure 6: The set up for the visual paired difference test

Results

Hair volume analysis

The treatment effect on hair volume, measured by AI, has been presented numerically in Table I. The volume type is marked as follows: GHV (global hair volume); UHV (upper portion); MHV (middle portion) and LHV (lower portion).

In virgin hair tresses, the treatment caused statistically significant reduction (vs t_0) in GHV ($t_1 = -5.2\%$ and $t_2 = -3.6\%$). The LHV reduction was not statistically significant ($t_1 = 0.8\%$; $t_2 = -1.4\%$). However, the UHV showed significant increases ($t_1 = +1.4\%$; $t_2 = +1.5\%$).

In bleached hair, a reduction in GHV after treatments was also noted ($t_1 = -4.1\%$ and $t_2 = -2.2\%$) but only t_0 vs t_1 was statistically significant. This trend was mirrored by the LHV for the lower portion ($t_1 = -2.8\%$ and $t_2 = -0.3\%$), whilst the UHV increased with treatments ($t_1 = +3.2\%$ and $t_2 = +1.1\%$). Notably, t_1 vs t_0 and t_2 vs t_1 were statistically significant but not t_2 vs t_0 .

In summary, the GHV hair volume changes for the two types of hair followed the same trend: a notable decrease ($t_1 - t_0$), followed by a relative increase ($t_2 - t_1$), but still resulting in an overall relative decrease ($t_2 - t_0$). The lower hair volumes appeared to follow these trends, whilst the upper volume did not. The bleached hair returned higher magnitude of volume changes than the virgin hair.

Fibre alignment analysis

The standard deviation (SD) computed on the histogram of gradients (HOG) of each hair tress was adopted as an index for hair straightness. In particular, the HOG features were extracted on the segmented region only, thus discarding the background. The lower the SD value, the straighter the tress, since in this case the hair fibre was mainly aligned along one dominant orientation.

Table II shows the numerical values of the fibre alignment indices. For both types of hair, the treatments resulted in improved straightness, as the SD values were reduced between t_0 and t_1 and t_2 , respectively. However, alignment reduction between t_1 and t_2 was noted, which was higher for bleached hair (3%) than that of virgin hair (1%). This outcome corroborated the volume results for virgin and bleached hair.

Machine learning: treatment order test

Based on the above analysis, in the **treatment order test**, correct results were returned for 84% of virgin pairs and 92% of pairs of bleached hair tresses.

Machine learning: timepoint recognition

On images from virgin hair tresses, the best classifier (Naïve Bayes) reached a 71% accuracy, with the confusion matrix shown in the upper part of Table III. Grouping t_1 and t_2 together ($t_1 + t_2$), we could classify t_0 (before using the product) vs ($t_1 + t_2$) (after using the product 2 or 3 times) and reach an accuracy score of about 94%, as shown in the lower part of Table III

On images from bleached hair tresses, the best classifier (Naïve Bayes) reached a 74% accuracy, with the confusion matrix shown in Table III (upper part). Again, after grouping t_1 and t_2 together we could classify t_0 (before using the product) vs $(t_1 + t_2)$ (after using the product 2 or 3 times) reaching an accuracy score of about 93%, as shown in the lower part of Table III (lower part).

It was notable that for both hair types the predictive power was fairly similar, hence the different magnitude of volume and alignment effects for virgin and bleached hair did not cause confusion.

Table I: Global and local hair volumes for the three time points of the training dataset. UHV=volume of the upper hair portion; MHV=volume of the middle hair portion; LHV=volume of the lower hair portion; GHV= total (global) volume of the whole hair tress.

time	region	Virgin hair volume (%)			Bleached hair volume (%)		
		volume	std	change	volume	std	change
0	UHV	20.4	2.3	0.0	21.7	2.9	0.0
	MHV	35.0	3.1	0.0	34.3	2.0	0.0
	LHV	44.0	3.4	0.0	44.0	3.3	0.0
1	UHV	21.8	3.2	1.4	25.0	3.5	3.2
	MHV	34.4	4.0	-0.7	33.8	3.5	-0.5
	LHV	43.2	4.1	-0.8	41.2	5.1	-2.8
2	UHV	21.9	2.8	1.5	22.8	3.2	1.1
	MHV	35.4	2.5	0.4	33.5	3.4	-0.8
	LHV	42.7	3.3	-1.4	43.6	4.9	-0.3
0	GHV	36.9	6.6	0.0	21.5	7.9	0.0
1	GHV	31.7	8.4	-5.2	17.4	10.8	-4.1
2	GHV	33.3	8.9	-3.6	19.3	6.0	-2.2

Brace brackets denote pairs with statistically significant difference.

Table II: Fibre alignment indices for three time points of the training dataset.

time	Virgin hair		Bleached	
	SD	% change vs t_0	SD	% change vs t_0
t_0	0.1482		0.1217	
t_1	0.1407	-5%	0.1131	-7%
t_2	0.1428	-4%	0.1171	-4%

		Predicted					
		Virgin hair			Bleached hair		
		t_0	t_1	t_2	t_0	t_1	t_2
Ground truth	t_0	17	1	1	15	0	2
	t_1	0	10	7	1	8	2
	t_2	2	3	13	0	9	11
		t_0	$(t_1 + t_2)$		t_0	$(t_1 + t_2)$	
		t_0	17	2	15	2	
		$(t_1 + t_2)$	2	33	1	30	

Table III: Confusion matrices. Correct results in bold.

Online paired-image comparison test: image analysis

The selected subsets of images used for the online paired image-comparison test were separately analysed using the AI methods described already (based on three images per tress, per time point). Global and local hair volumes for the three time points for the online survey dataset are shown in Table IV. This small data set did not show the same trends as the large data set, but such variations in the scale of the local and global volume effects are not atypical for hair tress testing, hence the larger number of images used for the training set. The alignment analysis for the subset of images displayed the same trend for both types of hair (Table VI) - a reduction of alignment from t_0 to t_1 , and a relative increase for bleached hair from t_1 to t_2 .

time	region	Virgin hair volume (%)			Bleached hair volume (%)		
		volume	std	change	volume	std	change
0	UHV	20.9	3.1	0.0	22.3	2.2	0.0
	MHV	32.4	4.0	0.0	34.3	1.6	0.0
	LHV	43.3	5.1	0.0	43.4	3.2	0.0
1	UHV	21.8	2.8	0.8	23.9	3.1	1.6
	MHV	35.1	1.8	2.7	34.5	3.2	0.2
	LHV	43.2	2.1	-0.2	41.6	5.8	-1.8
2	UHV	20.6	3.4	-0.3	22.3	2.6	0.0
	MHV	35.8	2.2	3.4	33.2	3.7	-1.2
	LHV	43.6	3.5	0.3	44.6	4.9	1.2
0	GHV	33.8	6.9	0.0	24.6	10.5	0.0
1	GHV	31.4	8.7	-2.4	18.9	13.8	-5.6
2	GHV	38.0	12.4	4.2	20.3	7.8	-4.3

Table IV: Results of the online paired-image comparison test.

Table V: Fibre alignment indices for the selected image subset used in the online survey, based on three images for each tress.

time	Virgin hair		Bleached	
	SD	% change vs t_0	SD	% change vs t_0
t_0	0.1457		0.1280	
t_1	0.1395	-4.5% (the straightest)	0.1161	-10% (the straightest)
t_2	0.1487	+1.9%	0.1194	-7.3%

Online paired image-comparison test: human assessment

Results of the online paired-image comparison test are shown in Table VI. The volume of virgin hair at t_2 was seen as lower than t_0 but higher than t_1 . Bleached hair after treatments was assessed as having higher volume than at t_0 .

The two types of hair were assessed as showing opposing alignment trends. Virgin hair's straightness increased after treatments in comparison with t_0 , whereas for bleached hair it decreased.

The flyway reduction for virgin at all three time points was as predicted. For bleached hair, the flyway reduction prediction was only confirmed in t_0 vs t_1 and t_1 vs t_2 .

In summary, according to the human eye assessment of 2D images, the virgin hair lost volume and flyway and gained alignment after the treatments, whilst the bleached hair was more volumised and lost alignment.

The human assessment was then compared to the volume and straightness trends from the image analysis of the subset used for the on-line survey (Table VII). For bleached hair. humans saw volume increase and straightness reduction, whilst the measured overall effect of the treatments was volume decrease and straightness increase. The human assessment of the virgin hair's volume was for a decrease

following all treatments and an increase for t_1 vs t_2 , but only the latter was corroborated by the AI analysis. The straightness assessment elicited results inconsistent with the measured overall effect.

Table VI: Results of the online paired-image comparison test.

Paired comparisons		Number of responses selecting the earlier time point	Number of responses selecting the later time point	Total number of valid responses	p - value
Virgin hair tresses					
Which tress is more voluminous?	t_0 vs t_1	110	91	201	0.102
	t_0 vs t_2	128	73	201	0.000
	t_1 vs t_2	68	131	199	0.000
Which tress is straighter?	t_0 vs t_1	85	116	201	0.017
	t_0 vs t_2	82	119	201	0.005
	t_1 vs t_2	96	103	199	0.335
Which tress has more fly away?	t_0 vs t_1	140	61	201	0.000
	t_0 vs t_2	113	88	201	0.045
	t_1 vs t_2	126	73	199	0.000
Bleached hair tresses					
Which tress is more voluminous?	t_0 vs t_1	86	116	202	0.021
	t_0 vs t_2	85	117	202	0.014
	t_1 vs t_2	88	114	202	0.039
Which tress is straighter?	t_0 vs t_1	171	31	202	0.000
	t_0 vs t_2	149	53	202	0.000
	t_1 vs t_2	109	93	202	0.146
Which tress has more fly away?	t_0 vs t_1	128	74	202	0.000
	t_0 vs t_2	96	106	202	0.263
	t_1 vs t_2	127	75	202	0.000

Table VII: Volume and hair straightness comparisons between the different time points reported in the survey (Table VI) and

their agreement with AI-generated GHV. ✓ denotes agreement between human comparisons reported in the AI analysis.

Treatment	Virgin		Bleached	
	Volume	Straightness	Volume	Straightness
t_0 vs t_1	No difference	Increase ✓	Increase	Decrease
t_1 vs t_2	Increase ✓	No difference	Increase ✓	No difference
t_0 vs t_2	Reduction	Increase	Increase	Decrease

Visual paired difference test

The t_2 treated hair (both virgin and bleached) was identified as having less volume, less flyaway and being straighter than t_0 , a result which agrees with the AI (Table VIII). This result corroborates the GHV statistical analysis of virgin hair, and the GHV trend for bleached hair, albeit not statistically significant. Hence, the human visual assessment of hair, suspended in a way similar to the way straight hair hanged down, appeared sensitive enough to identify changes in volume and straightness of both hair types tested. The flyway showed a trend of reduction, but was only significant for bleached hair.

Table VIII: Results of the visual paired difference test (n=50 responses)

Question	Virgin hair			Bleached hair		
	t_0	t_2	p-value	t_0	t_2	p-value
Which tress is more voluminous?	33	17	0.016	32	18	0.032
Which tress is straighter?	11	39	0.000	18	32	0.032
Which tress has more flyaway?	22	28	0.240	31	19	0.059

Discussion

Image data analysis and machine learning

The results highlighted that the lower and the upper portions of the tresses showed opposite directions of volume change following the treatments: whilst the lower portion of the hair appeared to lose volume, the upper part was gaining. However, it was noted that the treatments caused an overall reduction in GHV for both types of hair suggesting that the LHV reduction compensated for the gains in the UHV (Table 1). For virgin hair the UHV increase and GHV decrease between t_0 and post treatments are statistically significant, hence the above explanation was partially confirmed. For bleached hair, the statistical analysis confirmed the UHV and LHV trends between t_0 and t_1 and t_1 and t_2 inferring that the hair firstly lost volume and became more aligned, but gained back volume after the third treatment. The GHV showing the statistical difference for t_0 vs t_1 partially corroborates this explanation.

The hair types were not directly statistically compared statistically, but the comparison of trends suggests that bleached hair leans more towards volumising than the virgin hair. In summary: a) the virgin hair firstly (t_0 vs t_1) showed a volume loss of 5.2%* (alignment), followed (t_1 vs t_2) by a gain of 1.6% (less alignment and some volumising) but remained with 3.6%* less volume at (t_0 vs t_2) (overall alignment); b) the bleached hair firstly (t_0 vs t_1) showed a loss of 4.1%* (alignment), followed by an increase of 1.9% (the balance moved towards volumising) and a total volume loss of 2.2% (* denotes that the two time compared time points showed statistically significant differences).

The changes in alignment measurements, both overall and between time points, mirrored the volume trends. The overall trend was an increased alignment between t_0 and each post treatment timepoint by 4-7% (Table 2), indicating that the fibre surfaces exhibited reduced inter-fibre friction. However, the increased alignment between t_0 and t_1 was partially lost between t_1 and t_2 , by 1% for virgin and 3% for bleached hair. This data corroborates the volume data and the overall trend of the alignment effect at t_1 being offset by volumising at t_2 .

These outcomes infer that the treatment effect only partially followed the ETB, which predicted that all treatments would cause a volume increase and an alignment reduction. Between t_0 and t_1 , the effect was opposite to the ETB, but between t_1 and t_2 , it confirmed the ETB. One possible explanation for this result is that altering the assembly shape might have been the product efficacy objective, rather than a GHV increase. Another explanation is a possible cumulative effect of the treatments which could not be predicted precisely via the ETB. If such effect occurred despite the robust washing procedure, it could be impacting on the volumising and alignment with somewhat different magnitudes between timepoints.

The ETB that volume and alignment are inversely related was confirmed. This suggests that a balance between the two is needed to achieve a desirable appearance. The ETB for higher magnitude of treatment effect on bleached hair was also confirmed, which supports the assumption that the fibre surface interacts with the treatments via different mechanisms than those for virgin hair.

Classifiers tests

Both hair types exhibited good accuracy in the treatment order test, inferring that the chosen treatments were sufficient to elicit an adequate training dataset and that the chosen features reflected the distinctive differences between time points. The treatment order test supported the ETB prediction that bleached hair would interact more effectively with the treatments and display more notable effects, as the accuracy of ordering tresses for bleached hair was higher than that for virgin hair (Table III). The time order test performed well when images were tested as treated vs non-treated. The correct allocation of a tress between t_1 and t_2 was more challenging, which could be due to the smaller magnitude of changes. It could also be attributed to the challenges the image-based data presented due to the context within which the images were acquired. Although the conditions like the lighting, distances and handling (i.e. combing the hair before photographs are taken) were kept consistent, they reflected ambient context similar to the way people take photographs of themselves. One avenue for improvement would be to better control the image acquisition conditions. However, whilst this step would reflect controlled laboratory conditions,

the application of computer vision in cosmetics is shifting towards using selfies for automated image analysis.

The online survey and AI

The online survey structure aimed to assess if humans were able to see differences in the images of the same hair tress at different time points. The survey was not intended to provide a verification of the AI, as it only included the frontal image of each tress and used terminology which was not a direct equivalent to the measured features. It was envisaged that it could offer insights into how close humans could come to observing changes in images of hair tresses which were quantified using the machine learning approach for hair analysis.

The set of images were selected randomly, in order to prevent any bias if the small set was adapted to match exactly the large one. As a result, its volume and alignment characteristics were not the same as those of the larger dataset, however the objective of the test was to assess whether humans can detect any changes.

The volume and straightness assessment of bleached hair by humans was not consistent with the image analysis but was consistent with the ETB. We suggest that this could be due to the human's judgement being affected by the tress shape changes, i.e. more even distribution of volume between UHV and LHV created a perception of volume increase. It is feasible that despite having to give a separate judgement for flyaway, humans found it difficult to distinguish volume and flyway on a hair image. This explains the somewhat inconsistent flyaway results for bleached hair and is underlined by the ETB that the two types of hair interact differently with the treatments.

For the virgin hair, the human perceptions of volume and straightness changes showed less conclusive assessment and partial alignment with the image analysis. As indicated above, one reason for it could be the different interactions with the actives, resulting in virgin hair undergoing less notable shape alterations i.e. the virgin tresses retained a more triangular shape than the bleached tresses. This more consistent shape could be easier to compare for humans when viewing two images. The flyway assessment supports this explanation, as its assessment is in line with expectations. In addition, the optical properties of the hair such as colour might have influenced the shape/volume perceptions, too. In this case, virgin hair was notably darker than that of the bleached.

In summary, accurate visual assessment of hair from a single view point was very challenging for humans, yet it is a common everyday occurrence for people to judge hair appearance in such way: in the mirror or on a selfie. As the shape and colour of the hair are likely to impact on the human perceptions more than discreet parameters such as volume or fibre alignment, human assessment remains an important element of hair tress assessment methodology.

Visual paired difference test

This test, including t_0 and t_2 tresses only, aimed to compare before and post treatment effects (reflecting total cumulative changes), when viewing hair tresses directly. The result agreed with the AI prediction for volume reduction and straightness increases for both types of hair. This inferred that the

human judgements of hair volume and straightness, when hair was presented vertically and viewed directly, may not be the same as when viewing hair photographs. In this experiment, the human perception of flyaway was also included and the result broadly agreed with the flyway ETB and with the on-line test results. One explanation for the fit of these data with the AI analysis is that humans could observe and compare the 3D shape of the hair tresses more accurately than the 2D image. In another test, not reported in the paper, the tresses from all three time points were ranked by trained assessors, but the results were less conclusive than those by naïve assessors when judging the suspended tresses. This supports the conclusion that hair presentation has a significant impact on the results of sensory assessments of hair assembly for volume and alignment.

Conclusion

We conclude that the training dataset created for this experiment and the related machine learning process were able to detect and successfully quantify important hair assembly features and to place unknown hair images in the correct before or after-treatment category. The AI also offered valuable insights into the build of hair volume by the tested products, whilst maintaining hair alignment. It was based on the simultaneous analysis of features such as upper, mid and lower tress volume and overall fibre alignment. This analysis was achieved using images generated in a consistent, but not strictly controlled manner, more akin to photographs taken “in the wild” than to instrumental image collection and analysis.

The project also introduced the concept of ETB, based on the published technical knowledge of fibres and assembly treatments aimed at increasing hair volume. The ETB was used to triangulate the AI analysis, not as its validation. The AI analysis appeared to be more sensitive and discriminatory tool than a theoretical prediction based on the assembly behaviour, as the treatments themselves were complex mixtures which appeared to interact differently with the different hair types.

Finally, the human perceptions of the above hair assembly features were another equally important triangulating factor. The different tests with human assessors highlighted the challenges associated with the hair presentation and perception. The range of tests did not confirm or reject the AI analysis completely, but highlighted the role of hair presentation i.e. live tress assessment in vertical position delivered the results most consistent with the AI for both types of tested hair.

In summary, this proof-of-concept study combined machine learning and human testing in order to explore treatment effects on hair assembly features and to compare the capacity of AI and humans to detect and classify them in contexts which are more akin to daily hair viewings than instrumental tress assessment. The AI analysis was proven to be informative, but further tests are needed to clarify how the texture and possibly optical features are interpreted by humans, so that AI can one day be considered as a substitute for human assessment.

Acknowledgements

References

1. Loussouarn G, Lozano I, Panhard S, Collaudin C, El Rawadi C, Genain G. Diversity in human hair growth, diameter, colour and shape. An in vivo study on young adults from 24 different ethnic groups observed in the five continents. *Eur J Dermatology*. 2016;26(2):144–54.
2. Bouabbache S, Galliano A, Littaye P, Leportier M, Pouradier F, Gillot E, et al. What is a Caucasian ‘fine’ hair? Comparing instrumental measurements, self-perceptions and assessments from hair experts. *Int J Cosmet Sci*. 2016;38(6):581–8.
3. Loussouarn G, Garcel AL, Lozano I, Collaudin C, Porter C, Panhard S, et al. Worldwide diversity of hair curliness: A new method of assessment. *Int J Dermatol*. 2007;46(SUPPL. 1):2–6.
4. Hough PS, Sc B, Hue JE. *Hair Body*. 1976;578(November):571–8.
5. SWIFT JA. Some simple theoretical considerations on the bending stiffness of human hair. *Int J Cosmet Sci*. 1995;17(6):245–53.
6. Ezure M, Tanji N, Nishita Y, Mizooku T, Nagase S, Osumi T. The Secrets of Beautiful Hair: Why is it Flexible and Elastic? *Cosmetics*. 2019;6(3):40.
7. Mamada A, Nakamura K. A study of the volume and bounce decrease in hair with aging using bending elasticity measurements. *J Cosmet Sci*. 2007;58(5):485–94.
8. Mamada A, Ishihama M, Fukuda R, Inoue S. Changes in hair properties by Eucalyptus extract. *J Cosmet Sci*. 2008;59(6):481–96.
9. Jachowicz J, McMullen R. Mechanical analysis of elasticity and flexibility of virgin and polymer-treated hair fiber assemblies. *J Cosmet Sci*. 2002;53(6):345–61.
10. Jung IK, Park SC, Lee YR, Bin SA, Hong YD, Eun D, et al. Development of a stiffness-angle law for simplifying the measurement of human hair stiffness. *Int J Cosmet Sci*. 2018;40(2):157–64.
11. Wei G, Bhushan B, Torgerson PM. Nanomechanical characterization of human hair using nanoindentation and SEM. In: *Ultramicroscopy*. 2005.
12. Wolfram LJ, Albrecht L. (II) (III). 1985;99(February):87–99.
13. Wortmann FJ, Wortmann G, Haake HM, Eisfeld W. Analysis of the torsional storage modulus of human hair and its relation to hair morphology and cosmetic processing. *J Cosmet Sci*. 2014;65(2):59–68.
14. Robbins CR, Crawford RJ. A method to evaluate hair body. *J Soc Cosmet Chem*. 1984;35(November):369–77.
15. McMullen R, Zisa F, Jachowicz J. Measurements of hair volume by laser stereometry. *J Cosmet Sci*. 2009;60(2):171–85.
16. Evans TA, Ph D. *Body Talk : Defining Volume , A Technical Dissertation on Hair Body and Volume Scientifically Speaking*. 2020;1–8.
17. Jiang R, Kezele I, Levinshtein A, Flament F, Zhang J, Elmoznino E, et al. A new procedure, free from human assessment that automatically grades some facial skin structural signs. Comparison with

- assessments by experts, using referential atlases of skin ageing. *Int J Cosmet Sci.* 2019;41(1):67–78.
18. Messaraa C, Richard TJC, Walsh M, Doyle L, Connor CO, Mansfield A, et al. Perceived age and perceived health among a Chinese cohort: does it mean the same thing?
19. Polanía LF, Bala R, Purwar A, Matts P, Maltz M. Skin Chromophore Estimation from Mobile Selfie Images using Constrained Independent Component Analysis. *Electron Imaging.* 2020;2020(14):357-1-357-6.
20. Muhammad UR, Svanera M, Leonardi R, Benini S. Hair detection, segmentation, and hairstyle classification in the wild. *Image Vis Comput.* 2018;
21. Zhang M, Zheng Y. Hair-GAN: Recovering 3D hair structure from a single image using generative adversarial networks. *Vis Informatics [Internet].* 2019;3(2):102–12. Available from: <https://doi.org/10.1016/j.visinf.2019.06.001>
22. Savardi M, Daniels G, Tamburic S, Muhammad UR, Benini S. Figaro-tresses: A dataset for evaluating hair assembly features before and after cosmetic treatment. *Data in Brief* 31 (2020) 105964.
23. S.-W. G. on Claims, Technical document on cosmetic claims, Tech. rep., Tech Doc Cosmet claims (July 2017).
24. Ho TK. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 832-844, Aug. 1998, doi: 10.1109/34.709601.
25. Tan X, Triggs B. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Transactions on Image Processing* 19 (6) (2010) 1635–1650. doi:10.1109/TIP.2010.2042645.
26. Cortes C, Vapnik V. Support-vector networks. *Mach. Learn.* 20 (3) (1995) 273-297. <https://doi.org/10.1023/A:1022627411411>
27. Dalal N, Triggs B. Histograms of oriented gradients for human detection. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), Vol. 1, 2005, pp. 886–893 vol. 1. doi:10.1109/CVPR.2005.177.
28. Jackson J. A User's Guide to Principal Components. Wiley Series in Probability and Statistics, Wiley, 2005. URL <https://books.google.it/books?id=f9s6g6cmUTUC>
29. Lewis DD. Naive (bayes) at forty: The independence assumption in information retrieval. Nédellec C, Rouveirol C (Eds.), *Machine Learning: ECML-98*, Springer Berlin Heidelberg, Berlin, Heidelberg, 1998, pp. 4–15.