

# Determining viability of image processing models for forensic analysis of hair for related individuals

Yuqing Olivia Wang<sup>1</sup>, David S.H Funes<sup>2</sup>, Candice Bridge<sup>2,3</sup>

<sup>1</sup> Tesla STEM High School, Union Hill-Novelly Hill, Washington

<sup>2</sup> University of Central Florida Department of Chemistry, Orlando, Florida

<sup>3</sup> National Center for Forensic Science, University of Central Florida, Orlando, Florida

## SUMMARY

Errors in testimonies of microscopical hair examination have led to dozens of false convictions, since hair strands from different people can look similar. Mitochondrial DNA (mtDNA) can be obtained from hairs; however, because mtDNA is identical along the maternal line, it is ineffective in distinguishing between family members. While nuclear DNA (nDNA) can be extracted from the hair's root, it is not always present, and nDNA in the hair shaft is limited and often degraded. This limitation has led to extensive research on objective methods of differentiating individuals using microscopical imaging and statistical analysis. We aimed to determine if there was a machine learning algorithm with the highest accuracy in individual classification using imaging of microscopical analysis of hair. Images of 644 hair sections were analyzed by objectively quantifying the hair's medulla, width, color, texture, and pigment distribution. To determine the highest accuracy, we used four supervised models: discriminant analysis (DA), multinomial logistic regression, K-nearest neighbor, and support vector machine; and two unsupervised models: principal component analysis and hierarchical cluster analysis. The accuracy was tested on isolated data from nine families to test cases where mtDNA would fall short in differentiating related individuals. Unlike the unsupervised models, the supervised models provided comprehensive results of accuracy determination. The DA model yielded the highest accuracy of 88.89%, with 3/5 families with twins meeting or exceeding this accuracy. Our results demonstrated DA as the strongest statistical model for future research in differentiating between related individuals, which may increase the probative value of the evidence.

## INTRODUCTION

In 2015, a report from the Federal Bureau of Investigation found that in 35 cases where testimony on microscopic hair analysis was used to sentence a suspect to the death penalty, 33 (94%) contained erroneous statements, leading to the execution of nine individuals and five dying on death row (1). Other national organizations, such as the National Academy of Sciences, whose mission is to provide objective advice on science and technology matters, and the President's Council of Advisors on Science and Technology, which is tasked

with developing evidence-based recommendations to the President of the United States, have also argued that the subjectivity of the hair analysis makes it unreliable (2-3). These criticisms have fueled skepticism toward microscopically comparing hairs in forensic fields. Forensic hair examiners currently perform examinations by microscopically analyzing around 25 hairs from a suspect to compare them to ones found at the crime scene (4). They compare categorical features such as hair color, medulla type (the dark band of pigment that often runs through the middle of the hair), and pigment density (how concentrated the pigmentation is in the hair strand) (5). Even with extensive training, this work can be prone to bias and inconsistencies across examiners (2). For these reasons, the skepticism towards this science primarily comes from the subjective nature of qualitative statements and the lack of statistical support (6).

Due to the lack of quantitative methods used for hair analysis, research into improving methods is important and necessary, as it may generate investigative leads in forensic cases (7). Hair is often the most abundant evidence found at crime scenes, as people naturally shed over 100 hairs daily (1). Hair strands can be found if they are ripped out of the assailant's head by the victim or if left behind passively, such as on an abandoned piece of clothing. Improper testimony of hair examinations leads to the danger of erroneous verdicts, whether the suspect is convicted or acquitted. Therefore, it is crucial to develop quantitative methods to accompany testimony for clarity and reliability of the method.

Currently, the most widely used method for hair analysis is DNA analysis. Nuclear DNA (nDNA) can distinguish between all individuals but, until recently, was thought to only be in the skin tag located at the hair's root. A skin tag is only present when the hair is forcefully removed, which may not apply to most hair samples found in crime scenes. Recent research has shown the strong potential of extracting nDNA from the hair shaft, but low yields and varied results complicate this finding and require further study (6). Alternatively, mitochondrial DNA (mtDNA) is found all throughout the hair strand and can be extracted, although doing so destroys the hair (6). Unlike nDNA, completely unrelated individuals can share mtDNA haplogroups (6). Most importantly, mtDNA is identically passed down the maternal lineage, making it nearly impossible to distinguish between related individuals using mtDNA alone. For these reasons, mtDNA is a complementary method to microscopical analysis.

Previous studies obtained microscopical images of hairs and distinguished between individuals through machine learning, neural networks, and statistical models (6, 8, 9). These methodologies resolved the issues of subjectivity

through quantitative and statistical analysis; however, none have sought to compare the accuracy of computational models as their primary purpose. The statistical hair comparisons discussed in this paper have been derived from the successes of previous computational models, including the addition of a novel pigment distribution feature (10). Previous studies have shown color measurements as a potential for discrimination using red, green, and blue (RGB) measurements, obtaining F-ratios (a statistical measure of sample and group variance) up to 69.24 (9-13). The F-ratio is the ratio of the inter-sample variability to the intra-sample variability, with higher values indicating a significant difference between the groups. Additionally, cortical texture has yielded statistically significant results through homomorphic filtering and gray-level co-occurrence matrix calculations (GLCM) (11,14). GLCM shows the spatial relationship of image pixels and their intensities, which can describe the texture of the hair cortex, the middle layer of the hair (11, 14). Another feature that has been measured is pigment distribution along the hair strand, which can be distributed evenly, unevenly, or clustered toward the medulla, a pigmented band sometimes found at the center of the hair strand. This feature was investigated in a survey of forensic hair examiners, and pigment distribution was found to be an important feature used by 96% of trained hair examiners (5,15,16). A previous study also used pigment distribution to differentiate individuals, but its methodology relied more on matching pigment than quantification (17). Analysis of variance (ANOVA) tests were performed on the data to calculate the statistical significance of variables.

Statistical models are used to aid in interpreting large datasets and understanding how datasets are related or differentiated. These models can be either unsupervised or supervised. Unsupervised models are typically used for visualizing patterns and similarities within the dataset in order to classify samples into groups. These models use the most discriminatory variables to group similar samples and separate different samples. Thus, these models cannot predict which individual a hair strand came from. Still, they can provide visual and mathematical aids to determine how similar or different individuals' data points are. However, unsupervised models rely on the analyst to accurately interpret the results to know why samples grouped together. Therefore, unsupervised models are not used to make predictions on how new samples in a dataset would be classified. In this study, hierarchical cluster analysis (HCA), principal component analysis (PCA) and k-means clustering were used to evaluate whether the hair measurements collected from each participant can be used to associate or differentiate the samples with a family group. HCA is a technique that clusters data points together based on similarity distances, such as Euclidean distance (18). When the distance is small between two data points, they are clustered together, and when the distance is further away those two data points are placed into different groups. PCA operates in a similar way by grouping similar items together into one cluster (19). Notably, PCA minimizes the impact of similarities and maximizes that of differences, as the differences are what places the data points into groups. K-means clustering operates similarly by grouping data points into a pre-determined number of clusters based on the smallest distance between the data point and the center of the group (20).

Conversely, supervised models can yield concrete

predictions and perform accuracy testing. This is because supervised models require that the classification of each data point be included in the training dataset so that the algorithm can determine which features should be used to best group samples into pre-determined classifications. The resulting classifications can then be compared to the pre-determined classification to calculate accuracy. The supervised models tested were discriminant analysis (DA), which is similar to PCA but attempts to separate classes of data (all data from the same individual) as much as possible (21); multivariate logistic regression (MLR), which calculates the probability of classes through the linearization of variables (22); support vector machine (SVM), which performs pairwise comparisons between data, attempting to maximize the distance between the data points (23), and K-nearest neighbor (KNN), which tests to see which class the nearest values are from (24). It should be noted that because SVM can properly compare two classes, it was used to compare the datasets from the pairs of twins. Although these models all operate in a similar fashion, there are a few notable differences. DA uses an equation to calculate the class to which data point should be assigned (21). MLR requires a probability threshold to be inputted into the model to determine which data points should go into which class (25). SVM attempts to find the best line that separates data into different classes. KNN uses the closest points to the center of the group to aid in classifying data points, similar to K-means clustering.

The data measured from the hairs by the digital imaging program were separated by family and then input into the various statistical analysis algorithms. This project sought the best methodology for microscopical hair analysis by determining which statistical model yielded the highest accuracy in an individual identification scenario focusing on related family members. Of the models, we hypothesized the supervised models would be the most useful as they could give actual identity predictions.

This project's aims were to improve hair examination methods by comparing the accuracies of statistical models, using digital imaging measurements of features, to differentiate family members. Our study focused on related family members to simulate a scenario where mtDNA evidence may fall short because of the previously mentioned limitations. It is worth noting that a scenario where microscopical analysis is the only possible source of evidence to convict an individual is very unlikely. Since hair is only one type of evidence collected from a crime scene, a comprehensive analysis of the scene would involve recovering many types of evidence including trace evidence, biological evidence, and other types of physical evidence. Determining the best statistical model may help to create a more objective methodology for hair analysis that can be applied to other scenarios. We have found that statistical models can aid in comparisons and discrimination between hair samples and that the DA model provided the strongest accuracies of association. Through further innovations, hair evidence technology may not only help with assault, burglary, or sexual assault cases but also with cases such as identifying missing persons and locating human trafficking victims. However, it is to be noted that no statistical analysis for hair comparisons has been validated yet and nor is one currently in use for casework. This study further aids in developing microscopical examination of hair into a viable tool for a variety of cases where individual

identification could be useful.

Our null hypothesis was that there would be no statistical classification algorithm able to distinguish between hairs analyzed using the developed novel microscopical hair analysis methodology. After analyzing hairs from 23 participants and testing the data on numerous supervised and unsupervised models, the Discriminant Analysis supervised provided the highest accuracy in distinguishing between individuals. This procedure and data collected from this project can then be incorporated into future forensic technologies to help with an array of cases such as assault, burglary, sexual assault, and missing persons.

## RESULTS

The goal of our study was to compare different statistical models assessing the ability to use hair features to discriminate between donors and related individuals. This was done using both unsupervised and supervised methods that can aid in the microscopical examination of hair features. The samples for these models consisted of measurements of color, entropy, correlation, homogeneity, energy, contrast, and pigment distributions. These measurements were collected at the midshaft of the hair. Three hairs per donor were used. Each measurement was compiled into a single data matrix, each as an individual variable. 16 families responded to the initial form, and after equally distributing the population, nine families and 23 participants were selected. Families with twins were kept to include individuals with maximum similarity. Data was not collected on whether the twins were identical or fraternal to limit identifying information, and because we were more focused on studying familial relationships (**Table 1**).

The hair strands submitted with the surveys were evaluated through a series of quantitative measurements, including contrast, correlation, homogeneity, energy, entropy, hair width and the presence of the medulla. The distribution at 75x74 as well as the distribution 12x8 were also measured. Additionally, the red, blue, and green values were measured for each hair strand. The mean value for the red color, blue color and green colors, as well as the mode, were calculated across the twenty-five hair strands for each individual. The compiled data for each individual as well as each family unit was subsequently inputted into several unsupervised and supervised models to determine which model yielded

the most accurate association/differentiation of known hair strands.

## Unsupervised Models

We evaluated the following unsupervised models in this study: HCA, PCA and K-Means clustering. These models are given the measured variables but not given which individual the data point comes from. Unsupervised models can be used as a mathematical aid or data visualization technique and the techniques used in this study were chosen to evaluate their use in forensic analysis of hair characteristics (31). All of the measurements taken were used to observe the clustering and similarities of each of the samples (32).

We used PCA as a data dimensionality deduction technique and for data visualization. Score plots were used by converting the measured variables into fewer variables, called principal components, designed to reduce dimensionality while observing the relationship between original variables (**Figure 1A**). The first two principal components were calculated to explain the most cumulative variance of the variables. For example, F1 was primarily composed of and most affected by the color values and the distribution variables (contained 54.22% of the variance), while F2 was primarily composed of the Gray-Level Co-Occurrence Matrix (GLCM) variables and contained 25.12% variance (**Figure 1A**). The plotted points show how similar or different hairs are, just as a questioned hair can be visually compared to a known suspect's hair. The PCA data also provide further insights into the variables' viability and variability. This model showed how closely each of the samples and donors were related.

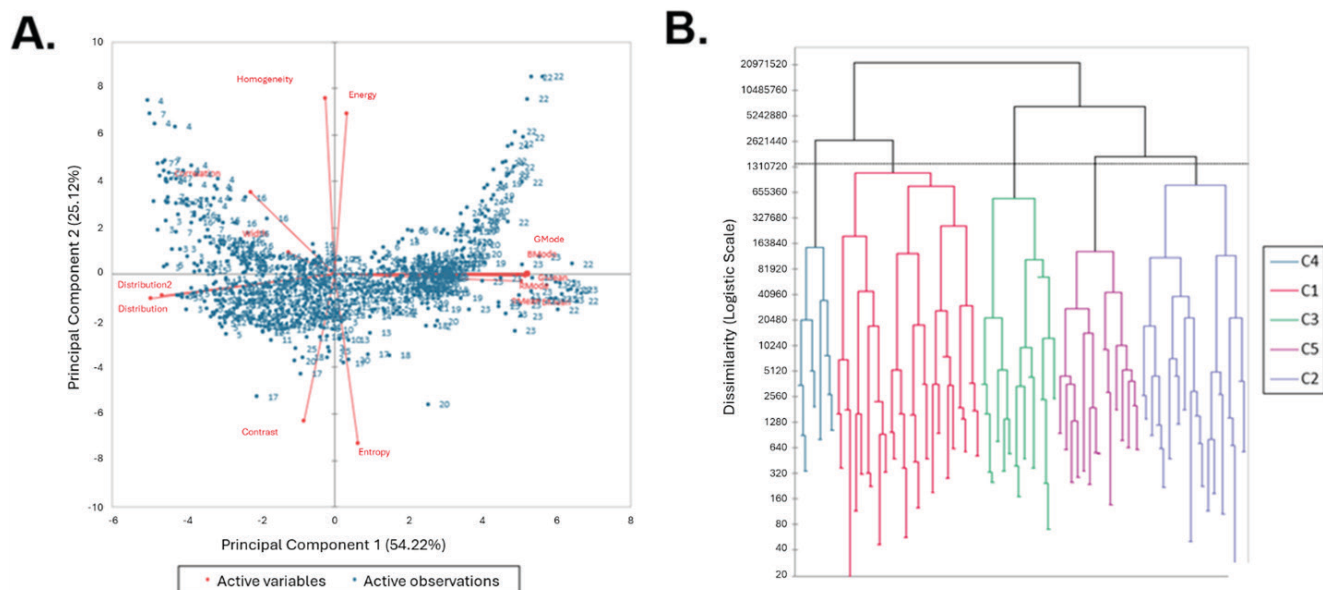
We developed a dendrogram using the HCA algorithm, which was formed by continuously linking the closest two data points until all the data was linked together. Then, the clusters can be formed by splitting the dendrogram into branches at a specific level (e.g., into five individuals), indicating all the data points under one branch belong in one cluster (as seen under the dotted line in **Figure 1B**). The dendrograms can visualize the relative similarity between two hair samples.

The tests we performed using k-means clustering yielded a large overlap of grouping results and significant misclassifications. When data from all the individuals were tested, hairs from the same individual were classified into many different clusters. For example, cluster 1 contained

Family #	Letter	Self-Identified Race	# of members	Twins
1	A	Caucasian	3	Y
2	B	East Asian	2	Y
3	C	Caucasian	4	Y
4	E	Caucasian & Hispanic	3	N
5	F	South Asian	4	Y
6	G	East Asian	3	Y
7	H	Caucasian	2	N
8	J	Hispanic	2	N
9	M	South Asian	2	N

**Table 1: Demographic of Participating Families.** Information about the nine families who participated in this study is presented, including: the family identification (i.e., Letter), the families' self-identified race, how many members of the family participated and which families had biological twin participants.





**Figure 1: Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA).** A) Principal Component Analysis graph of the data points from every individual,  $n=644$ . Each point indicates one hair from an individual, with the individual that hair is from as the numbered label. F1 and F2 are values calculated by the model to be based on which variables, when multiplied by a constant, help to distinguish the most between data points. Each active variable shows how each variable (e.g., medulla, width) influences each respective data point, which are the active observations (e.g., one hair from individual A1). B) HCA dendrogram of individuals 15, 25, 5, 11, and 19. These were randomly chosen individuals from Random Group 4 when testing using the model on unrelated individuals. The brackets extend until they reach one of those individuals' data points. The dotted line indicates the point where the data splits into five clusters for five individuals. C1-C5 indicate the different clusters the data points classified into.

hairs from 11 different individuals, and cluster 2 contained hairs from 6 different individuals. This became less of a problem when the model was used on only 5 individuals, but each cluster would still contain hairs from 2-3 different individuals. Therefore, we did not use this statistical model in subsequent analysis.

### Supervised Models

We also investigated several supervised statistical models, specifically DA, MLR, KNN and SVM. We used these statistical models to classify the individual from whom each data point came. We tried to find the supervised model that yielded the highest accuracy when testing data from related family members.

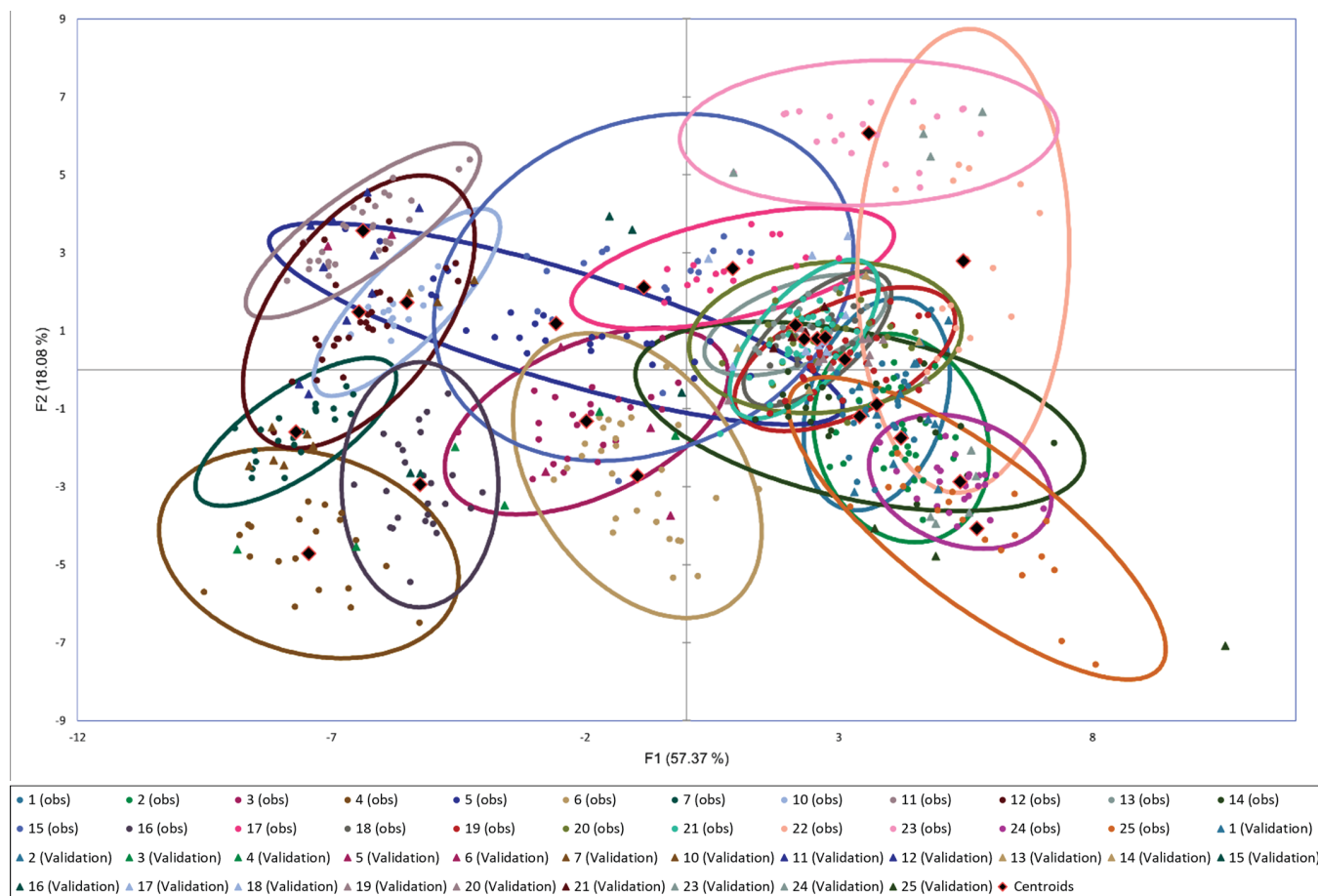
We investigated the accuracy potential of differentiating members of the same family using these four models. Each model was tested on the data from one family, and the accuracy was tested using three-six data points per individual based on how many family members there were and how many hairs were analyzed per individual. While taking images of the hairs, some individuals had fewer data points as some hairs were not suitable for analysis. This was only the case for a few individuals, and on average, most of the study participants yielded 36 measurements per individual. For example, the accuracy of Family A was tested with nine randomly chosen points because there are three individuals in the family, and three points were taken per individual. If possible, the accuracy of the data from all 23 individuals was taken. SVM was only tested on twins because this model is designed to compare two datasets.

The DA model yielded the highest accuracy in distinguishing between related individuals, with a family average of 88.89% (Figure 2). MLR had a family average

of 80.6%, KNN of 84.3%, and SVM of 84.0%. Additionally, DA aids data visualization, making it easier to interpret and understand (Table 2). Although all models yielded results greater than 80%, the DA model yielded the highest accuracy, so we proceeded with it.

DA models reduce the dimension of the data into just two variables, F1 and F2, which are based on the variables most helpful in distinguishing between individuals and are plotted on a graph. This graph can then be used to visually present how sections of the entire data set group together based on the input variables. DA was first performed on the data from all 23 individuals and validated using 92 random hair sections (four (4) strands each for 23 individuals) to test the classification accuracy, using the selected hairs as a random test set (Figure 2A). The model had an average accuracy of 72.83%. Of course, a case with 23 related suspects would be unlikely in the real world, and we present this result to illustrate the model's performance in an extreme case. While testing the DA model on families, the accuracy ranged from 75% to 100%, with an average of 88.9% accuracy. When testing the model on the five families with twins, 3/5 pairs yielded an 88.9% or higher accuracy (Figure 2).

To better understand how testing the model on related individuals impacted its performance, the DA model was also tested on unrelated individuals. Four or five individuals from the dataset were randomly selected using a number generator, and the DA model was run on the data. Groups of four or five were chosen as was considered the size of a reasonable suspect pool a theoretical model would have to differentiate between. The random individuals' average accuracy was higher than the family's. The graphs generated showed a clear distinction between individuals' data points, more so than the figures for related individuals (Figure 4A).



**Figure 2: Results of Discriminant Analysis (DA) model analysis on data from all individuals, n = 644.** Each circular plot point represents one hair from one individual that has been analyzed and graphed by the DA, and the data points from the same color indicate hairs from the same individual. Each colored oval represents the grouped data of one individual. The triangle plotted points indicate data points that were randomly chosen to test the validity of the model (4 hairs per individual).

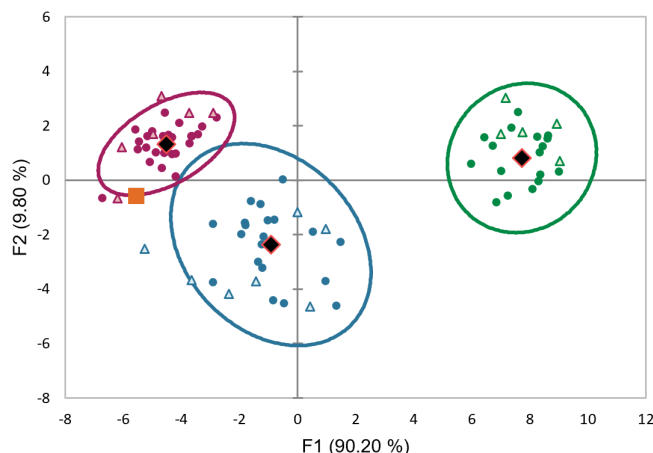
Statistical Method	A	B	C	E	F	G	H	J	M	Average	All Individuals
Discriminant Analysis	88.9%	75.0%	75.0%	88.9%	100%	88.9%	91.7%	100%	91.7%	88.9%	72.8%
Multinomial Logistic	66.7%	66.7%	68.3%	83.3%	100%	100%	100%	50.0%	100%	80.6%	N/A
K-Nearest Neighbor	66.7%	100%	58.3%	83.3%	100%	100%	50.0%	100%	100%	84.3%	39.4%
Support Vector Machine	90.0%	70.0%	60.0%		100%	100%				84.00%	N/A

**Table 2: Accuracy of Supervised Models on family data.** Accuracies of each of the models (rows) for each of the families (A-M) the model was tested on. The accuracies for all 23 individuals combined are also shown, with the exception of the Multinomial Logistic Model and Support Vector Machine (SVM), as the models were unable to support the large data size. Support vector machine (SVM) was only used for pairs of twins, italicized, because it performs comparisons between two classes.

If the questioned hair was not from any suspected individuals, it is essential to confirm that the model will not incorrectly assign the hair to one of the suspects. We evaluated this by using the model to predict an additional six hairs from an individual not used to train the model. The tests showed that while the model still provided a prediction, the F1 and F2 values calculated correctly indicated that the hair was from a different individual. The graph plotted the tested hairs, labeled "O", separately from the other individuals (**Figure 4B**).

## DISCUSSION

In this study, we assessed a qualitative approach to hair comparisons using distinct statistical models. We observed that while overlap in the measurement of different features occurred, the methodology could still aid in distinguishing both related and unrelated individuals. Our study showed the potential of using hair evidence to link to possible donors; however, it is to be noted that complete individualization using hair analysis was not possible. Our method showed how hair analysis is scientifically valid and can be made more objective by measuring features.



**Figure 3: Plotted successful individual prediction using Discriminant Analysis.** The model, trained with the three individuals from family G (in which 4 and 5 are twins), was asked to predict data from one hair from individual 5 (called the "Predicted Individual"). The model calculated the values of -5.534 for F1 and -0.589 for F2 for the unknown individual and plotted the value, as seen by the orange square. As the point fell into Individual 5's group, the model correctly predicted the hair to be from Individual 5.

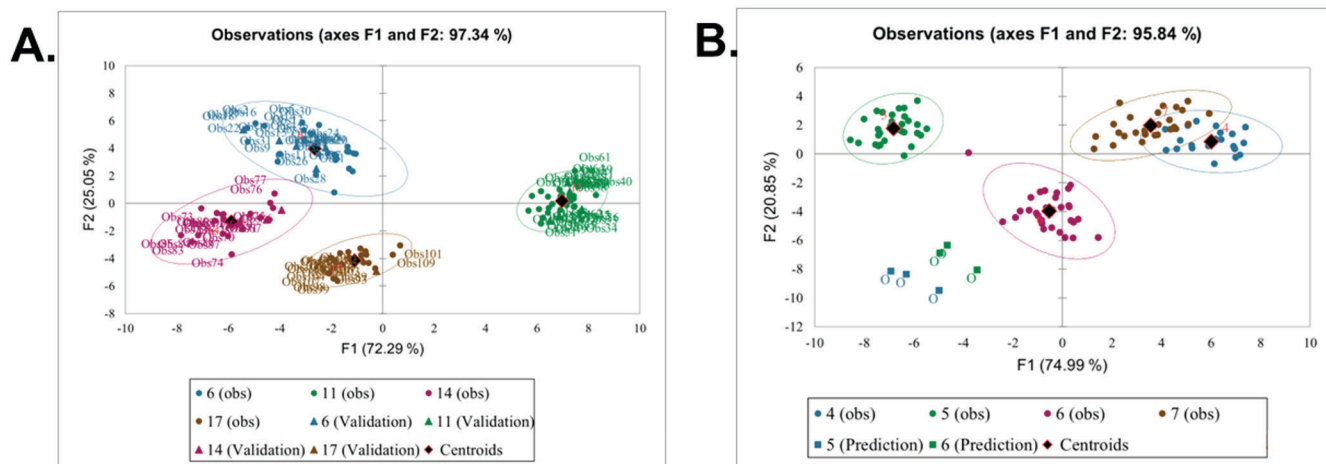
Discriminant analysis was the most accurate model when distinguishing between related individuals, with an accuracy of 88.89% for combined families. While the accuracy for all 23 individuals was relatively lower at 72.83%, it was expected that the discrimination power of each feature would be reduced due to increased variance overlap that occurred because of the similarities observed between individuals from the same family. In comparison, MLR had an average of 80.56%, KNN of 84.26%, and SVM of 84.00%. Therefore, as the technology for digitalizing microscopic hair analysis evolves, our results suggest that DA would be the most promising choice for the primary form of statistical analysis. We also found that DA remained reliable even when comparing hair strands from related individuals. Hence, this method can still be used

in cases where mtDNA does not suffice because hairs still visually vary amongst related individuals, making this methodology viable.

The unsupervised models, HCA and PCA, demonstrated showed potential by illustrating how similar or different hairs are using relationships between the input variables. These models group hair together based on the data, regardless of which individual the hair strand came from. This conclusion showed hair color and texture were viable hair features that unsupervised models could use in distinguishing hairs. Hair features such as pigment size, cross-section ratio, and width variation were also found to be additional variables measured for the comparisons. The results from the k-means clustering yielded poor results with significant incorrect groupings of sample. This likely occurred because the distance between different donors was small, resulting in incorrect groups of individuals. This method would lead to a large number of misclassifications, false positives, and false negatives. This method would lead to many misclassifications, false positives, and false negatives.

As a result of the success of discriminant analysis, further tests on the same sample population and associated demographic information were conducted. We found that the DA model yielded a high accuracy in distinguishing between unrelated individuals. Furthermore, hair from individuals not used to train the model were plotted as clear outliers. This aspect would be important in supporting real-world forensic examinations. If a DA model were created, for a specific forensic investigation, using data collected from the suspect's known hair, that model could then predict if the unknown strand collected from the crime scene can be included with the suspect's hair or if it is an outlier.

The misclassifications seemed centralized in a few families (Figure 2). While there may be individuals in which there will always be limitations in distinguishing between them, i.e., identical twins, complete exclusions or associations are not possible due to large similarities in observable and measured features. Thus, these misclassifications showed the values of qualitative and quantitative features or methods that could



**Figure 4: Observations for the unrelated and randomly chosen individuals from Group 3.** A – Graph of DA model trained with randomly selected unrelated individuals: 6 (South Asian Female), 11 (Hispanic Male), 14 (Hispanic Female), and 17 (Caucasian Male). B - Graph of DA model trained with related individuals 4-7 (individuals from families G and F). The model was asked to predict hairs from individual "O" (an individual not used to train the model), and the points at the bottom left show how the model can properly identify a hair as not belonging to a training set.



be considered in future research to improve our model's performance. The variability in accuracy between families also indicated the importance of testing the model's accuracy before its application. Once the process has been refined, it could be implemented into real forensic practices to support a hair examiner's testimony, especially when mtDNA does not suffice. The methodology could also be used in other areas involving individual identification, such as missing persons or human trafficking. The DA method could be refined by increasing the model's precision by sampling more families. Additional models and methods could be developed and used to provide a comprehensive evaluation of familial hair analysis. Note: some of the families consistently had much higher accuracies across all models (e.g., Families F, G, and M), whereas some families had much lower accuracies across all models (e.g., Families A, B, C).

The visual representation from both unsupervised models and the supervised DA model, coupled with the numerical data points generated from the image processing program, could assist a hair examiner in determining whether a hair strand may belong to a suspect. We note that no statistical methods for hair analysis have been yet validated (32). Altogether, the results obtained in this study and possible future developments may improve the forensic field of hair microscopy by making it more objective. With the continual improvement of methodology and technology, objective microscopical imaging of hair may improve the role of hair evidence in crime scenes, missing persons, and human trafficking cases.

## MATERIALS AND METHODS

An initial form was posted online through Facebook and shared by word of mouth starting from my high school, looking for people with brown/black hair only. This was to test a more realistic scenario where microscopic analysis would be needed because differentiating between black and blonde hair can be done with the naked eye. Those interested filled out a form where they provided their email, self-identified race, number of family members willing to participate, and whether there was a pair of participating twins. Families with four or more participating members, twins, or mixed race were qualified to be part of the study. In total, 23 participants from 9 families were selected from an original 16 participants (Table 1).

An envelope containing a questionnaire, instructions, and the informed consent form was given to the qualified family participants. Participants were asked to section their hair into four parts to obtain three hairs from each section. They were told hairs could be removed without force, i.e., by combing the section to retrieve hair strands. Once the hair samples were collected by the families, sealed in an envelope and they were returned to me by hand.

### Sample Collection

The middle of each collected hair strand was determined, and 1.5 cm was cut from each side to obtain a 3 cm hair length. Hairs were then mounted onto a slide using Kleermount Xylene mounting solution (refractive index of ~ 1.5) and covered with a glass cover slip.

A Canon T7i camera was mounted to an AmScope microscope-to-camera adapter (15x) and placed on an AmScope microscope. The camera was set to ISO 800,

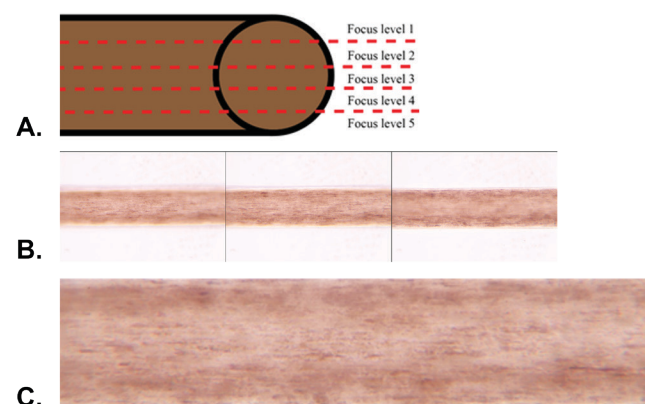
shutter speed 1/10, white balance to Tungsten, and picture style to Fine Detail. Kohler Illumination was performed before each session to ensure consistency in lighting and contrast (14). This consisted of closing the aperture to the smallest part and adjusting its height until it is clearest. The halogen light on the microscope was set to the highest setting. The hair was located by moving the stage and then adjusting until the hair was seen under the 40x lens.

Photos of each hair strand were taken at each end of the strand and in the middle. At 40x magnification, the hair was adjusted by turning the knob to be in focus at the highest part. The knob was then gradually turned so that the focus moved down the hair. Photos were taken at each of these focal points on the T7i.

### Hair Image Preparation

All images along the focus of one section of hair were then depth composited using PICOLAY (26), which combines the parts of each picture that were in focus (Figure 5A). This was done because hair is a 3D object, meaning as the focus changes, different hair parts would come into focus (Figure 5B). In PICOLAY, noise suppression was set to 1 so that the program only took parts of the image with detail; the filter was set to "smart" to keep both the finer and the bigger structures in the hair in focus; and the speed was set to "Slow, fine" to yield the clearest image. These settings were determined to yield the optimum clear hair image. The image was then white balanced through PICOLAY to ensure even color distribution across samples.

The image was then transferred to the ImageJ program, where it was cropped to only contain the hair's cortex (27, 28).



**Figure 5: Diagram and images of hair focus levels.** A) At the beginning, the hair was adjusted to focus level one so that only the top of the hair was in focus. Then the magnification was adjusted to see each of the hair sections in focus (focus levels 2 through 5). At each of the focus levels, a photo is taken. Note focus levels are not concrete; some hairs may require more or less focus levels depending on the shape of the hair. B) Raw images of hair at different focus levels from participant C-18. An image is taken where the strand is most in focus in the middle of the hair (left). This matches focus level 1 from Panel A. Images are then continuously taken along the focus (middle), around focus level 2, until the edges are in focus (right), around focus level 3. Note these are only 3 photos of the 7 taken for the hair section. C) Cropped image using ImageJ of individual C-18's hair. The 7 total images from the hair were depth composited together using PICOLAY so that each hair could have one finalized photo that showed every part of the hair clearly.

An example of the image of a hair strand without the medulla which was inputted into the program is presented (**Figure 5C**). If the hair had a medulla, a darker pigmented section that runs down the center of the hair, the hair was cropped into two images: one above and one below the medulla. The data calculated from both sides of the medulla was then averaged when input into the data table.

### Feature Quantification

The feature quantification program was written using Python and compiled in a Jupyter Notebook (29, 30). Before any calculations were taken, hair width and whether a medulla was present was recorded (0 = no medulla, 1 = medulla). The width was the number of pixels on the width of the cropped hair cortex. In total, 644 hair sections were analyzed.

### Color Analysis

A histogram of the color distribution of the cortex image was generated. Each color distribution histogram was developed by graphing every pixel's red, blue, and green values from 0 to 255. The mean and mode intensity values were measured and recorded for each RGB value. While the full graph of each RGB value was not used in this analysis to discriminate hair samples, it has potential in future research and serves as a helpful visual aid.

### Texture Calculations – Entropy & Grey-Level Co-Occurrence Matrix (GLCM)

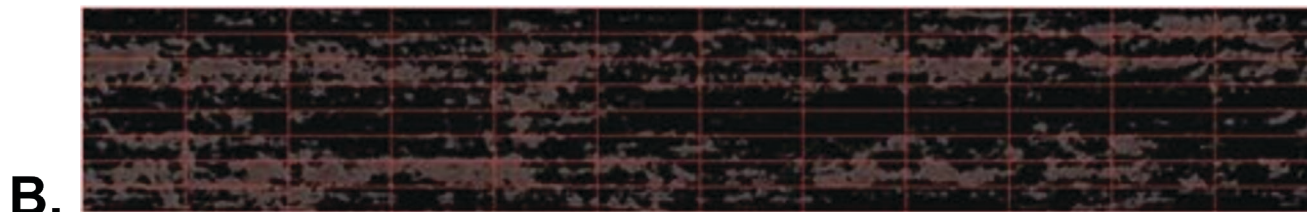
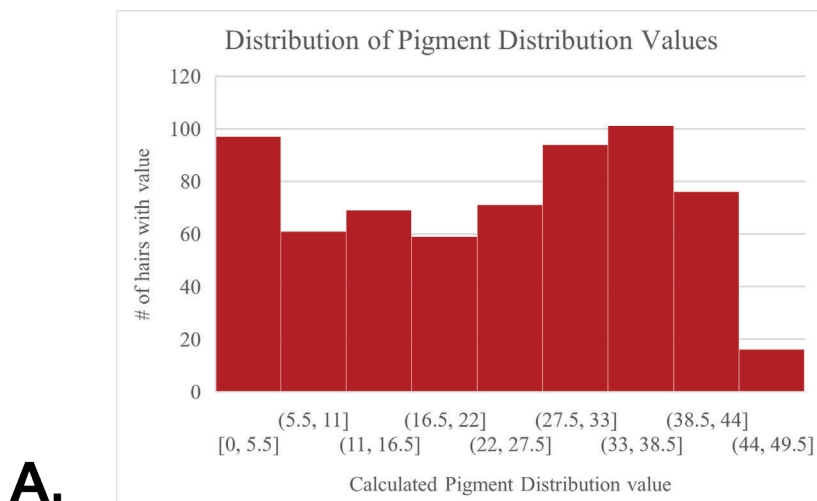
For the texture of the hair cortex to be quantifiably

measured, the image of the hair was first converted using homomorphic filtering. This is a method of image enhancement that increases contrast and was shown in a prior study to yield improved results when paired with image texture calculations (14). After the image was filtered, five calculations, as presented by Funes (10) and McWhorter (14), were performed on these images. The five calculations used were:

- Entropy: measures the amount of structure and chaos in an image (14)
- Correlation: the probability of the occurrence of specified pixel pairs (14)
- Homogeneity: the closeness of the distribution of elements along the image's diagonal (14)
- Energy: measures the uniformity of elements through the sum of the squared elements (14)
- Contrast: compares the intensity of each pixel to its neighboring pixel (14)

### Pigment Distribution

Pigment distribution along the cortex was quantified by first isolating the pigmented pixels. Pigmentation was analyzed by converting the image into hue, saturation, and value (as opposed to RGB) because previous research has shown the viability of using the lightness value (V) to isolate the darker and, therefore, pigmented pixels (11). The average lightness value across each pixel was measured, and the top 90% of the darkest pixels were isolated as the pigmented pixels. The value of the top 90% of pixels was chosen through multiple



**Figure 6: Distribution and visualization of pigment distribution variable.** A) Distribution of the calculated pigment distribution values for the 644 hairs from 23 individuals inputted into the model. B) Example of the isolated pigmented pixels and how they would be split. The percentage of pigmented pixels in each grid would be calculated. For example, the top left grid has a 3% pigmentation ratio, and the bottom left has 36%. Then the standard deviation of the values would be taken as the Pigment Distribution value of the hair.



tests of adjusting HSV values to isolate the pigmented pixels.

The cortex was then split up into 75x74 pixels ("Distribution") and 12x8 pixels ("Distribution2") (Figure 6B). These values were chosen because tests using these splits on a smaller scale yielded the highest F-values, the ratio of between sample variance to the within sample variance. The ratio of pigmented pixels to total pixels for each section was calculated. The standard deviation of all the values was then calculated to determine the evenness of the pigments in the cortex.

An image with a very uneven pigment distribution would yield a higher distribution value due to having a larger standard deviation (SD), and a hair with an even pigment distribution would have a low SD (Figure 6A).

### Statistical Models

After the images were run through the Python program, the data was recorded into an Excel spreadsheet. One-way ANOVA was calculated to determine the most statistically significant variables. All statistical calculations were done through a trial version of XLStat (33). Several variables had high F-values, indicating that there was more similarity between the same group than there were differences between different groups for a given variable. The variables with F-values greater than 100 are listed with their individual F-values presented in the parenthesis: red mean (191.398), blue mean (160.626), green mean (151.988), blue mode (130.107), green mode (122.698), and red mode (102.972). The variables with F-values between 10 and 100 were: homogeneity (54.700), distribution - 75x74 (55.128), energy (50.834), entropy (37.776), distribution - 12x8 (36.925), correlation (25.965), hair width (24.151), and contrast (20.644). The last variable evaluated, medulla, had the lowest F-value of 4.959 which indicated that the presence or absence of medulla did not help associate or differentiate hair samples of related family members.

The unsupervised models used in this study were HCA, PCA, and K-means clustering. In unsupervised models, the known group of an individual data point is not provided, and the model attempts to group data points based on similarity. Supervised models require known group information for each data point, and the model uses that information when grouping similar data points.

The supervised models that evaluated the data sets were DA, MLR, and KNN. The supervised models tested were DA, which is similar to PCA but attempts to separate classes of data (all data from the same individual) as much as possible (21); K-nearest neighbor, which tests to see which class the nearest values are from (24); multivariate logistic regression, which calculates the probability of classes through the linearization of variables (22); and SVM, which performs pairwise comparisons between data, attempting to maximize the distance between the data points (23). It should be noted that because SVM can properly compare two classes, it was used to compare the datasets from the pairs of twins. While these models provide discrete class assignments, it is harder to understand the processing done by the model, thus making the data more difficult to explain in court and should be used in conjunction with an unsupervised model. We could correctly discriminate between both related and unrelated donors using this model.

For each model, the algorithm was tested on data isolated from individuals of one family. This reduced the number of

classes the model must distinguish between and was more realistic to an actual forensic scenario. Comparison of accuracy across the different algorithms determined which model yielded the highest accuracy and the lowest possibility of false positives.

The model with the highest accuracy was further experimented on through tests such as having the model distinguish between unrelated individuals. Additionally, we tested a scenario where the model analyzed a hair strand from an individual that was not used to train the model. This simulated a situation where the questioned hair comes from a suspect not included in the questioned pool.

### ACKNOWLEDGMENTS

We want to thank Dr. Sandra Koch, Dr. Allyce McWhorter, and Ms. Heather Deel for their continued support and encouragement throughout the project's development. We would also like to thank Ms. Allender for her extensive help and guidance throughout the project and for inspiring this project in the first place. Finally, we would also like to thank Yuqing Olivia Wang's family and friends who supported her throughout this endeavor. Tesla STEM High School provided funding for this project.

**Received:** August 3, 2024

**Accepted:** October 23, 2024

**Published:** February 4, 2025

### REFERENCES

1. Houck, M. and B. Budowle. "Correlation of Microscopic and Mitochondrial DNA Hair Comparisons." *Journal of Forensic Sciences*, vol. 47, no. 05, Sep. 2002, pp. JFS15515J. <https://doi.org/10.1520/jfs15515j>.
2. National Research Council. *Strengthening Forensic Science in the United States*. The National Academies Press, 2009, <https://doi.org/10.17226/12589>.
3. Executive Office of The President's Council of Advisors on Science and Technology. "Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods Executive Office of the President President's Council of Advisors on Science and Technology." President's Council of Advisors on Science and Technology, 09 2016, [obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast\\_forensic\\_science\\_report\\_final.pdf](http://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf). Accessed 02 Feb. 2023.
4. Oien, C. "Forensic Hair Comparison: Background Information for Interpretation" *Forensic Science Communications*, vol. 11, no. 2, Apr. 2009. [archives.fbi.gov/archives/about-us/lab/forensic-science-communications/fsc/april2009/review/2009\\_04\\_review02.htm](http://archives.fbi.gov/archives/about-us/lab/forensic-science-communications/fsc/april2009/review/2009_04_review02.htm)
5. Wilkinson, L., and C. Gwinnett. "An International Survey into the Analysis and Interpretation of Microscopic Hair Evidence by Forensic Hair Examiners." *Forensic Science International*, vol. 308, Mar. 2020, pp. 110158. <https://doi.org/10.1016/j.forsciint.2020.110158>.
6. Grisedale, K.S., et al. "Successful Nuclear DNA Profiling of Rootless Hair Shafts: A Novel Approach." *International Journal of Legal Medicine*, vol. 132, Oct. 2017, pp. 107-115. <https://doi.org/10.1007/s00414-017-1698-z>.
7. Minnesota Department of Public Safety. "Hair." Bureau of Criminal Apprehension, 2020, [dps.mn.gov/divisions/bca/](http://dps.mn.gov/divisions/bca/)

- bca-divisions/forensic-science/Pages/trace-hair.aspx. Accessed 02 Feb. 2023.
8. Verma, M.S., et al. "Hair-MAP: A Prototype Automated System for Forensic Hair Comparison and Analysis". *Forensic Science International*, vol. 129, no. 3, Oct. 2002, pp. 168-186. [https://doi.org/10.1016/s0379-0738\(02\)00277-3](https://doi.org/10.1016/s0379-0738(02)00277-3).
9. Mills, M., et al. "Differentiation of Human Hair by Colour and Diameter Using Light Microscopy, Digital Imaging and Statistical Analysis". *Journal of Microscopy*, vol. 270, no. 01, Sep. 2017, pp. 27-40. <https://doi.org/10.1111/jmi.12646>.
10. Funes, D.S. H., and C. Bridge. "Analysis of Hair Color and Texture for Forensic Examinations." *Journal of Forensic Sciences*, vol. 66, no. 02, Mar. 2020, pp. 520-533. <https://doi.org/10.1111/1556-4029.14640>.
11. Koch, S.L., et al. "Microscopical Discrimination of Human Head Hairs Sharing a Mitochondrial Haplogroup". *Journal of Forensic Sciences*, vol. 66, no. 01, Sep. 2020, pp. 56-71. <https://doi.org/10.1111/1556-4029.14560>.
12. de la Mettrie, R., et al. "Shape Variability and Classification of Human Hair: A Worldwide Approach." *Human Biology*, vol. 79, no. 3, Jun. 2007, pp. 265-281. <https://doi.org/10.1353/hub.2007.0045>.
13. Birngruber, C., et al. "The Color(s) of Human hair Forensic Hair Analysis with SpectraCube®." *Forensic Science International*, vol. 185, no. 1-3, 24 Mar. 2009, pp. e19-23. <https://doi.org/10.1016/j.forsciint.2008.12.018>.
14. McWhorter, A.S. "Development and Evaluation of an Objective Method for Forensic Examination of Human Head Hairs Using Texture-Based Image Analysis". *Graduate Theses, Dissertations, and Problem Reports*, vol. 6207, 2015. <https://doi.org/10.33915/etd.6207>.
15. Hicks, J.W. *Microscopy of Hairs: A Practical Guide and Manual*, vol. 2, U.S. Department of Justice, 1977, pp. 1-57.
16. Airlie, M., et al. "Forensic Hair Analysis – Worldwide Survey Results." *Forensic Science International*, vol. 327, Oct. 2021, pp. 110966. <https://doi.org/10.1016/j.forsciint.2021.110966>.
17. Brooks, E., et al. "Digital Imaging and Image Analysis Applied to Numerical Applications in Forensic Hair Examination" *Science & Justice*, vol. 51, no. 01, 31 Jul. 2010, pp. 28-37. <https://doi.org/10.1016/j.scijus.2010.06.008>.
18. Bridges Jr., C. C., et al. "Hierarchical Cluster Analysis." *Psychological Reports*, vol. 18, no. 03, Jun. 1966, pp. 851-854. <https://doi.org/10.2455/pr0.1966.18.3.851>.
19. Wold, S., et al. "Principal Component Analysis." *Chemometrics and Intelligent Laboratory Systems*, vol. 2, no. 1-3, Aug. 1987, pp. 37-52. [https://doi.org/10.1016/0169-7439\(87\)80084-9](https://doi.org/10.1016/0169-7439(87)80084-9).
20. Kanungo, T., et al. "An efficient k-means clustering algorithm: analysis and implementation." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 07, July 2002, pp. 881-892. <https://doi.org/10.1109/TPAMI.2002.1017616>.
21. Ramayah, T., et al. "Discriminant Analysis: An Illustrated Example." *African Journal of Business Management*, vol. 4, no. 09, 4 Aug 2010, pp. 1654-1667.
22. Kwak, C., et al. "Multinomial Logistic Regression." *Nursing Research*, vol. 51, no. 06, Nov. 2002, pp. 404-410.
23. Noble, W. S. "What is a support vector machine?" *Nature Biotechnology*, vol. 24, 01 Dec. 2006, pp. 1565-1567. <https://doi.org/10.1038/nbt1206-1565>.
24. Cunningham, P., et al. "k-Nearest Neighbour Classifiers – A Tutorial." *ACM Computing Surveys*, vol. 54, no. 06, 13 Jul. 2021, pp. 1-25. <https://doi.org/10.1145/3459665>.
25. Gurden, S. P., et al. "Quantitative Analysis and Classification of AFM Images of Human Hair." *Journal of Microscopy*, vol. 215, no. 1, Jul. 2004, pp. 13-23. <https://doi.org/10.1111/j.0022-2720.2004.01350.x>.
26. Cypionka, H. *PICOLAY*. [www.picolay.de](http://www.picolay.de). Accessed 02 Feb. 2023.
27. Schindelin, J., et al. "Fiji: An Open-Source Platform for Biological-Image Analysis." *Nature Methods*, vol. 9, 28 Jun. 2012, pp. 676-682. <https://doi.org/10.1038/nmeth.2019>.
28. Strock, C.F. "Batch Cropping Images in ImageJ/Fiji." *Zenodo*, vol. 10, 9 Oct 2021, pp. 1. <https://doi.org/10.5281/zenodo.5559119>.
29. Virtanen, P., et al. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python." *Nature Methods*, vol. 17, 03 Feb. 2020, pp. 261-272. <https://doi.org/10.1038/s41592-019-0686-2>.
30. Kluyver, T., et al. "Jupyter Notebooks-a Publishing Format for Reproducible Computational Workflows." *International Conference on Electronic Publishing*, Jun. 2016. <https://doi.org/10.3233/978-1-61499-649-1-87>.
31. Misra, A., et al. "A comparison of supervised and unsupervised pre-training of end-to-end models." *Interspeech 2021*, 30 Aug. 2021, pp. 731-735. <https://dx.doi.org/10.21437/Interspeech.2021-654>.
32. Funes, David S. H., et al. "Morphological and chemical profiling for forensic hair examination: A review of quantitative methods." *Forensic Science International*, vol. 346, May 2023, pp. 111622. <https://doi.org/10.1016/j.forsciint.2023.111622>.
33. Addinsoft. *XLSTAT Statistical and Data Analysis Solution*. 2022, [www.xlstat.com](http://www.xlstat.com). Accessed 02 Feb. 2023.

**Copyright:** © 2025 Wang, Funes, and Bridge. All JEI articles are distributed under the attribution non-commercial, no derivative license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>). This means that anyone is free to share, copy and distribute an unaltered article for non-commercial purposes provided the original author and source is credited.