SMU Data Science Review

Volume 7 | Number 3 Article 5

2023

Differentiation of Human, Dog, and Cat hair Fibers using DART TOFMS and Machine Learning

Laura Ahumada
Southern Methodist University, lauraah10@gmail.com

Erin R. McClure-Price
Southern Methodist University, erinmcclureprice@gmail.com

Chad Kwong
Southern Methodist University, kwongchad7@gmail.com

Edgard O. Espinoza
US National Fish and Wildlife Forensic Laboratory, ed_espinoza@fws.gov

John Santerre
Southern Methodist University, john.santerre@gmail.com

Recommended Citation

Ahumada, Laura; McClure-Price, Erin R.; Kwong, Chad; Espinoza, Edgard O.; and Santerre, John (2023) "Differentiation of Human, Dog, and Cat hair Fibers using DART TOFMS and Machine Learning," *SMU Data Science Review*: Vol. 7: No. 3, Article 5.
Available at: https://scholar.smu.edu/datasciencereview/vol7/iss3/5

Differentiation of Human, Dog, and Cat Hair Fibers using DART TOFMS and Machine Learning

Laura Ahumada[1], Erin McClure-Price[1], Chad Kwong[1], Ed Espinoza[2], John Santerre[1]
[1] Master of Science in Data Science, Southern Methodist University,
Dallas, TX 75275 USA
[2] US Fish and Wildlife National Forensics Laboratory

Abstract. Hair is found in over 90% of crime scenes and has long been analyzed as trace evidence. However, recent reviews of traditional hair fiber analysis techniques, primarily morphological examination, have cast doubt on its reliability. To address these concerns, this study employed machine learning algorithms, specifically Linear Discriminant Analysis (LDA) and Random Forest, on Direct Analysis in Real Time time-of-flight mass spectra collected from human, cat, and dog hair samples. The objective was to develop a chemistry- and statistics-based classification method for unbiased taxonomic identification of hair. The results of the study showed that LDA and Random Forest were highly effective in separating mass spectra collected from hair samples with accuracies ranging from 94-98%. This approach holds significant promise for forensic investigations, archaeology, and artifact analysis.

1 Introduction

According to Locard's Exchange Principle, "every contact leaves a trace", meaning that whenever two things touch, evidence is left behind (Petherick et al., 2010). This is especially true in the case of hair fibers, which are estimated to be found in over 90% of crime scenes (Liu et al., 2023). This theory has held an important place in the ongoing field of forensic investigations, and FBI Unit Chief Oien noted that "[h]air evidence is one of the most common types of evidence encountered in criminal investigations" (Oien, 2009, p. 1).

Hair fibers are shed regularly by humans and household pets, which makes the presence of both often analogous (Gwinnet, 2016). Hair fiber analysis has been a traditional forensic tool since its first publication in 1857 (Cole, 2017). Since then, its usage has expanded globally, and it became an accepted practice worldwide. During forensic investigations, morphological features are generally the first tools that are utilized to identify hair fibers found at crime scenes (Zafarina & Panneerchelvam, 2009). Unfortunately, in a report by Garret (2016), it was found that Federal Bureau of Investigation (FBI) examiners were not able to differentiate between human and dog hair fibers which, in addition to other faulty hair fiber analyses, resulted in the unlawful incarceration of an innocent man.

Numerous problems have been found to have occurred through morphological hair fiber analysis. In 2015, Cates et al. reported that the FBI found that erroneous testimony had been made by analysts in over 90% of court room trials in a review of 3,000 cases conducted prior to 2000 although we note that many of these errors involved individualization of hairs rather than class character analysis. Due to similar reviews, microscopic hair analysis has been largely disesteemed in favor of genetic analysis (Norton et al., 2016). However, in an international survey by Wilkinson & Gwinnet (2020) that included data from nine countries, the authors reported that human hair identification was the most requested analytical service, and the top three tools for answering this question were stereomicroscopy, transmitted light microscopy, and comparison microscopy. Clearly, new methods are needed to identify hair fibers collected from crime scenes.

The purpose of this study was to apply machine learning algorithms to Direct Analysis in Real Time time-of-flight mass spectra collected from human, cat, and dog hair samples with the purpose of developing a chemistry- and statistics-based classification method that could be used to make unbiased and objective species identifications. This work is important for forensics and will have utility in archeology and for the analysis of museum artifacts.

2 Literature Review

2.1 Keratin

The solid matrix of hair fibers is predominantly made up of a fibrous protein called keratin. The keratin found in mammalian hair, horns, hooves, and skin is classified as α-keratin, due to the α-helical conformation the secondary structures emulate (Kreplak el al., 2002), while birds and reptiles enjoy the strength benefits from pleated sheets of β-keratin in their scales, beaks, feathers, and talons (Greenwold et al., 2014). Both α- and β-keratins have been important materials used throughout human history; horns were commonly used as cups, sea turtle shells were made into combs, feathers were used as quills, and we enjoy cloth made from the keratinous wool of sheep and other mammals (O'Connor et al., 2015).

2.2 DART TOFMS

Direct Analysis in Real Time time-of-flight mass spectrometry (DART TOFMS) was developed in 2005 with the goal of using it to analyze a myriad of substances (Cody et al., 2005), since then, it has been successfully applied in a variety of forensic applications. From drug and explosives analysis to timber species identification (Sisco & Forbes, 2021), DART TOFMS is rapidly gaining popularity in the forensics community. In addition, it has also been successfully used to differentiate between rhinoceros and bovine keratin (Price et al., 2018), and for species identification of Camelid hair fibers (Price et al., 2020; Jahanbanifard et al., 2023).

DART TOFMS, as described by Cody et al. (2005), is a technique employed to collect mass spectra through several concurrent processes that occur resulting in the generation of ions, which are electrically charged molecules. Within the DART system, a carrier gas—typically helium or nitrogen—is subjected to an electric potential to create plasma, producing high-energy atoms. When a sample is exposed to this high-energy plasma, several processes occur concurrently. If the gas is heated, the sample burns as atoms from the gas react with atmospheric molecules to produce charged ions. These charged ions are subsequently measured using time-of-flight mass spectrometry, providing valuable information about the composition of the sample.

2.3 LDA

Linear discriminant analysis (LDA) is a model first offered by Ronald Fischer in 1936 as a technique to discriminate between taxonomic populations using linear functions. Nowadays, this algorithm is considered one of the most popular methods used for modeling classification of continuous data suffering from high dimensionality.

The model works by first ensuring a maximum separation between samples through maximizing the ratio of the “between class” variance to the “within class” variance (Balakrishnama & Ganapathiraju, 1995), where the “within class” variance is the variance between individual vectors within a class and the between class variance is the variance between features of a population. Once this maximum separation is ensured, the observations are divided for classification by a function fit using metrics such as Euclidean distance or RME. In summary, LDA focuses on predicting categorical classification using continuous data by maximizing the distance between observations and then configuring a linear function that best divides the data.

As for its use in practice, Jahanbanifard et al. (2022) found that LDA outperformed other algorithms, such as convolutional neural networks, kernel discriminant analysis, and quantum-enhanced support vector Machine, in the classification of camelid hair samples. Furthermore, previous research completed by Vaclavik et al. (2011) showed that LDA applied to DART TOFMS data could be used to identify the validity of animal fats such as lard and tallow.

2.4 Random Forest Models

Random forest is an algorithm that was created as a parallelizing counterpart to the Adaboost model in January 2001 by Leon Breiman. The algorithm is an accumulation of decisions tree predictors made of independent random samples and random set of variables that go through a voting process when making predictions (Breiman, 2001). While decision trees alone have their advantages, it is known that they tend to overfit and are sensitive to the data itself, which is where the Random Forest algorithm comes into play.

Random Forest models are a huge improvement over individual decision trees in several ways. They use bagging and random feature selection to enhance the ability of decision trees by reducing correlation, preventing overfitting, and making it robust to outlier and noise. Randomization within the algorithm also largely helps it obtain better accuracy and converge via the Law of Large Numbers (Dinov et al., 2009). Features of each individual tree within the algorithm help to obtain information about correlation and variables importance, and additionally, the sheer structure of the algorithm makes it easy to parallelize (Breiman, 2001). With all these optimizations and advantages in mind, the Random Forest model has certainly earned its accolades in practice.

Currently, Random Forest is one of the most common algorithms used in machine learning for both classification and regression tasks. As for its use with mass spectrometry, Random Forest was applied to DART high resolution mass spectra collected from psychoactive plants with the goal of developing a method for identifying illicit materials by sampling headspace (Appley et al., 2019). In addition, Lennert & Bridge (2018) used Random Forest to successfully classify smokeless powders and found that the data collected by both DART TOFMS and gas chromatography mass spectrometry performed similarly.

3 Methods

3.1 Sample collection

Individual human hair samples were collected via donation from individuals and hair salons. No identifying information was taken or retained for these samples. Dog and cat hair samples were taken from the authors' own pets, their friends' pets, dog grooming salons, and animal shelters with no inclination towards color, breed, size, or length of hair fiber. The samples themselves consisted of approximately 40-90 hair fibers without roots. All human (n = 113), cat (n = 83), and dog (n = 87) samples were obtained from three primary locations in the United States: Ashland, OR, Austin, TX, Des Moines, IA, and from Bogota and Cundinamarca, Colombia. Human hair samples were collected from each region as follows: Ashland, OR (n = 38), Austin, TX (n = 23), Des Moines, IA (n = 24), and Bogota and Cundinamarca, Colombia (n = 28).

3.2 Sample preparation

All hair samples underwent the following process. Approximately 20 hair fibers were placed in borosilicate glass culture tubes and immersed in deionized water before undergoing sonication for 15 minutes. All samples were rinsed and allowed to dry overnight prior to analysis.

3.3 Data collection

Mass spectra of all hair samples were collected using a DART-SVP ion source (IonSense, Saugus, MA, USA) paired with a JEOL AccuTOF 4G LC Plus time-of-flight mass spectrometer (JEOL USA, Peabody, MA, USA). The DART TOFMS system was operated at 250°C, and a reference standard (poly(ethylene glycol) 600 (Ultra Scientific, Kingstown, RI, USA)) was collected at the beginning, middle, and end of each sample set of 10. Detailed system settings are described in Price et al. (2018).

Sample mass spectra were collected by twisting several hair fibers together and holding them in the sampling gap using forceps, resulting in total ion current chromatograms. Using msAxel (version 1.0.5.0, JEOL Ltd.), each mass spectrum was collected by averaging a chromatogram and subtracting the background before exporting the data as centroided text files. Spectra were collected over a range of 90 – 1000 *m/z*.

3.4 Data structure & preparation

The data used for analysis and interpretation consisted of mass spectra samples, wherein each spectrum text file contained a column for the mass of each ion, and a column for its intensity, measured by the number of times that a particular mass hit the detector within the mass spectrometer.

Each mass spectrum was combined into a single dataset consisting of 1500 features, or ions, and 274 rows representing each sample, making the overall structure of the data "short and fat". The data consisted of mass spectra collected from 78 dogs, 83 cats, and 113 humans. After removing null values, the total number of remaining features, or ions, was 950.

For the exploratory data analysis (EDA) and machine learning modeling, the spectra were rounded to their nominal values and split into 80% training and 20% testing groups to ensure that the classes within both test and training classes maintained the same distribution as the data itself.

Machine learning models were evaluated by accuracy, F1 (the harmonic mean of precision and recall), recall, and precision. The metric that carried the most importance for this research was the F1 score, due to its reflecting both precision and recall which refines the focus to true positives while reducing the number of false positives and false negatives.

3.5 EDA

We began our exploratory data analysis by looking at the distributions of the features/ions. We initially found evidence to suggest multimodal distributions of ion densities after we individually analyzed the distribution of the mean intensity per ion for each class. The multimodal distribution inferred potential feature importance between the ions while the difference in the distributions themselves indicated evidence to suggest distinction between classes. The results from our density analysis are represented below in Figure 1.

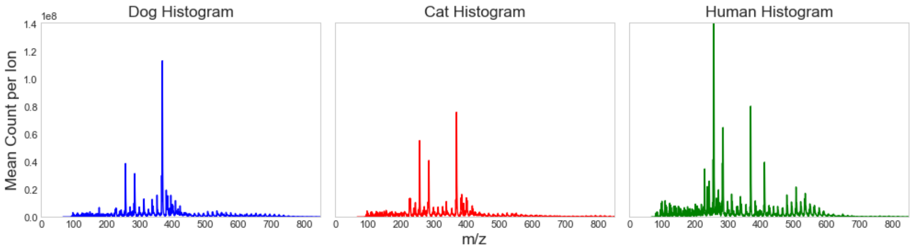


Figure 1. The mean distribution of ions from dog, cat, and human classes.

The density plots (Figure 1) indicated a potential pattern that could be leveraged to classify different groups within the datasets. To further visualize the data without focusing on the mean densities, a t-distributed stochastic neighbor embedding (TSNE) analysis was performed. The data was reduced to 2 components and the results can be found in Figure 2. While there is notable intermixing between the classes, overall, there was evidence to suggest that there were distinct differences between each of the three classes.

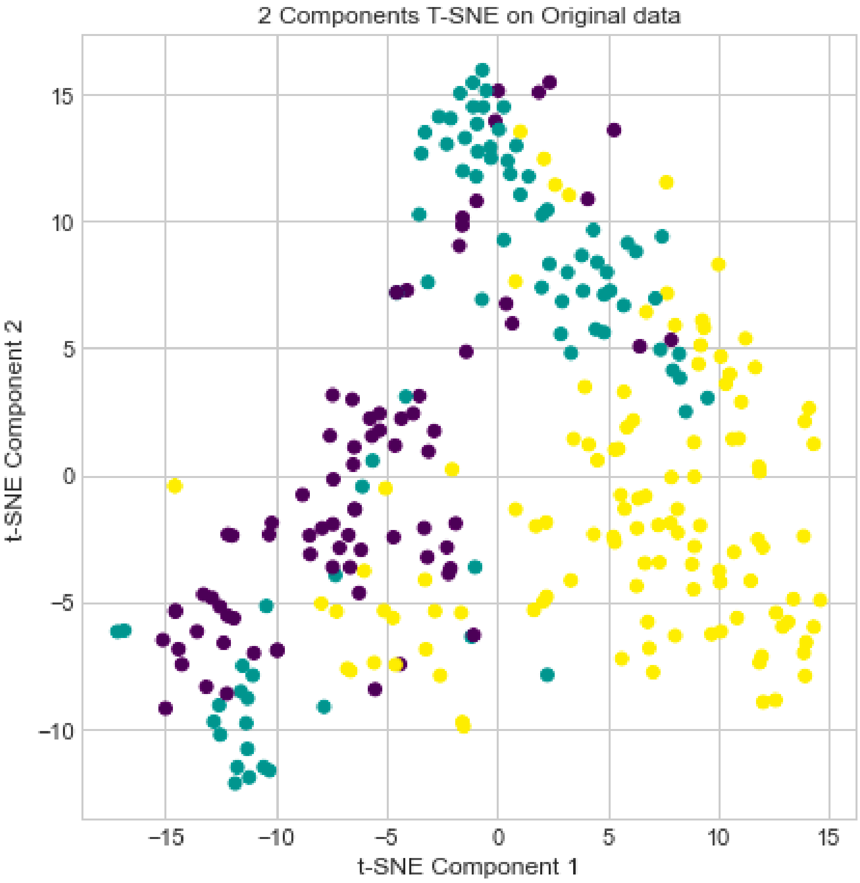
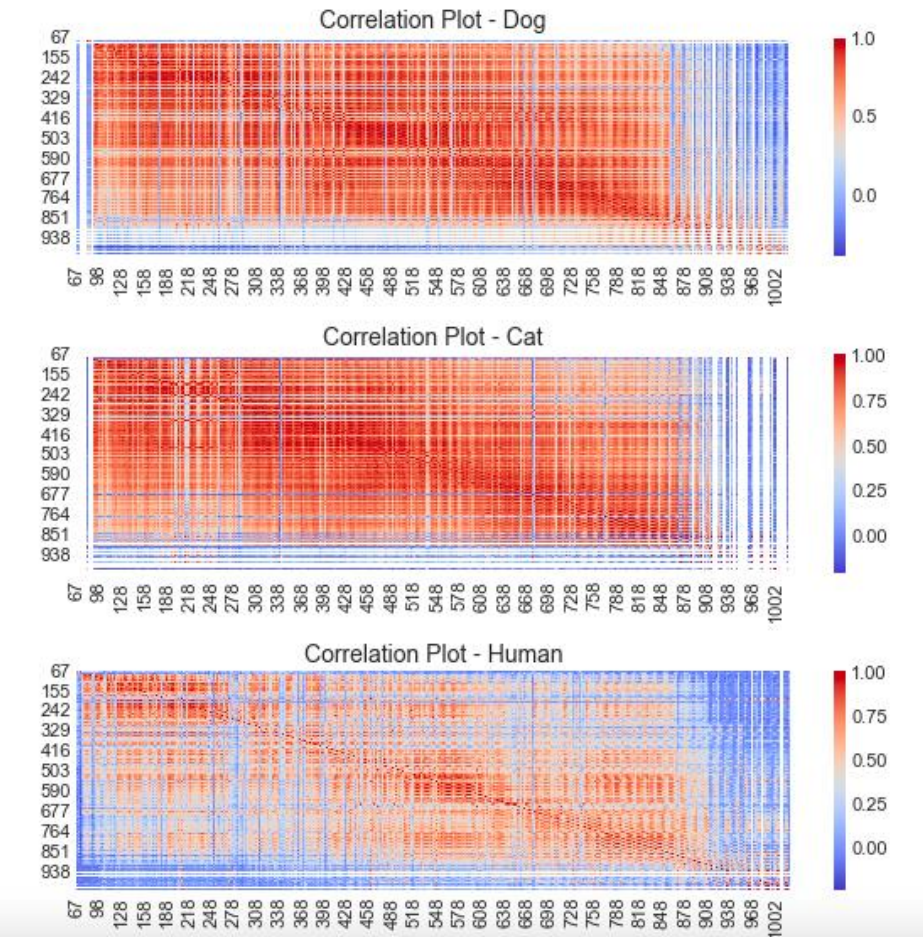


Figure 2. Visualization of the 2 component TSNE plot.

In an investigation into the relationships between the features and classes, we constructed a correlation plot (Figure 3) which indicated correlations existed between several ions. While there was evidence to suggest similar hotspots that overlapped between classes, we also found evidence to suggest differences between classes due to the different distributions of hotspots. In addition, the cat and dog classes showed this interlap more so than the human class. With reference to Figure 3, cats and dogs displayed higher correlations between ions below 600 *m/z*. More specifically we found that when the cat and dog classes were grouped together, ions in the range of 500-530 *m/z* showed stronger correlation to ions in the range of 430-450 *m/z*.

Figure 3. Correlation plots per species.



The investigation into ion correlation led to a further search of the unique ions found per class and between classes. We binned the ions by their occurrence between classes and these results can be found below in Figure 4. In this investigation, we discovered that there were 28 ions unique to humans and 41 ions unique to dogs and cats when grouped into a single class. The evidence of similarity between classes presented by

Figure 4 helped to explain the intermixing observed in the TSNE plot (Figure 2). However, the ion frequency could potentially have more of an impact, so we then investigated which ions were unique to each class by focusing on the rate of the occurrence of ions vs analyzing the specific densities of the ions found.

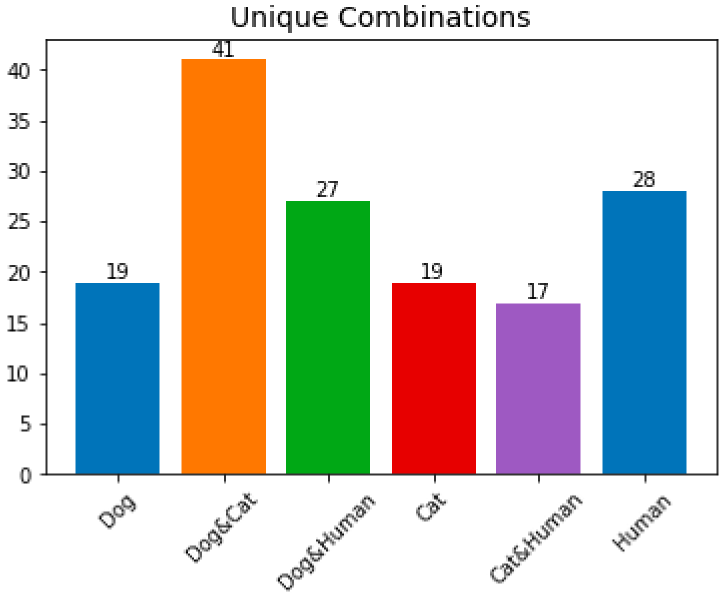


Figure 4. Unique ion combinations per individual class and grouped classes.

In this investigation into unique ions, we began by analyzing the overall distribution of the ions detected within the samples through a simple histogram of the data separated by class. From the histogram presented in Figure 5, one interesting observation was that the most frequently occurring ions were found between 90 and 600 *m/z*.

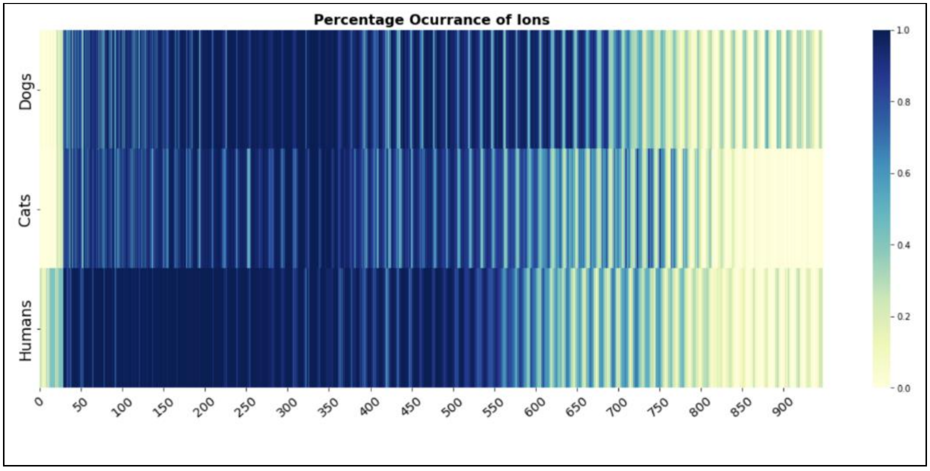


Figure 5. Percentage occurrence of Ions

Continuing our investigation into specific ions, we searched for individual ions unique to each class. We determined which ions were present in 100% of the samples for each class and displayed this information in Figure 6. Using this information, we could hypothesize on which ions are truly distinct between humans, cats, and dogs.

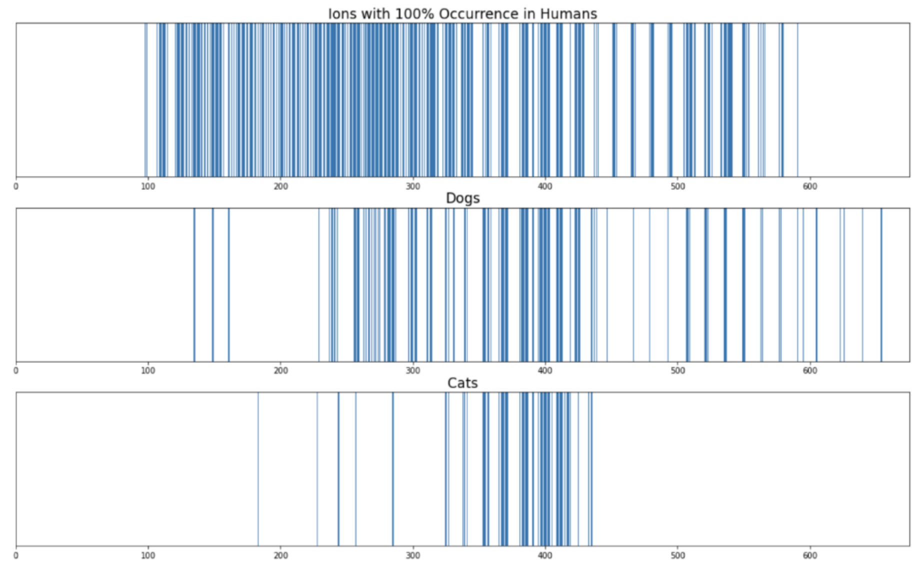


Figure 6. Ions that were present in all individual samples per class.

With evidence and insight from the density and correlations investigations, we then decided to visualize the data once again through a dimensionality reduction algorithm.

Given that the class data followed a multivariate Gaussian distribution, we chose to use Linear Discriminant Analysis (LDA) as a tool to analyze the variance between the classes. As shown in Figure 7, the LDA model produced a clear linear combination of features that indicated a clear separation between all three classes. These results showed that the data was well-structured and did not contain substantial noise, which increased the confidence for the classification model's performance.

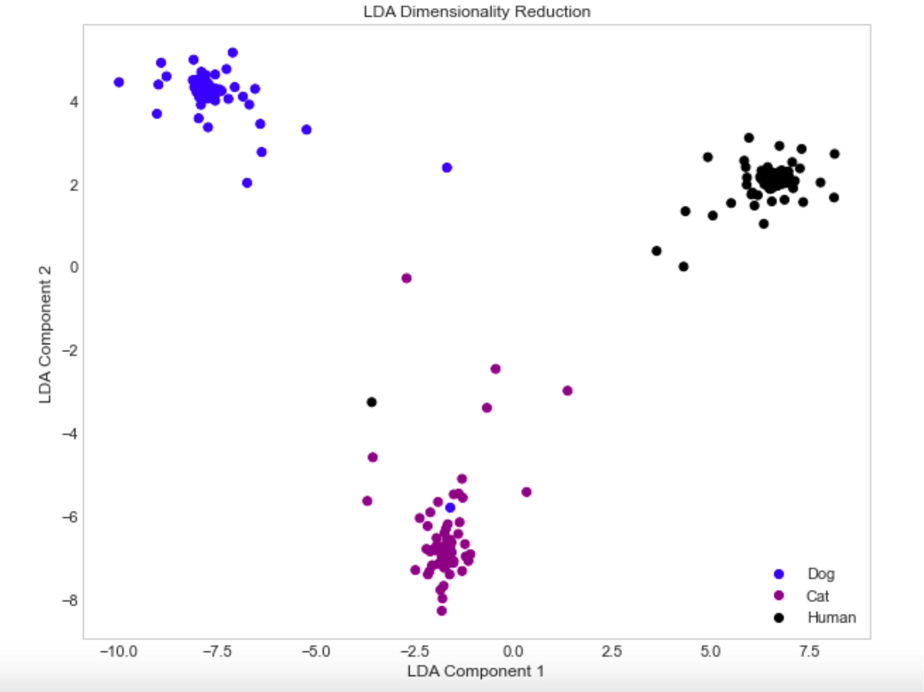


Figure 7. LDA plot of the dog, human and cat spectra.

Since we had broad regional information available for the human class, further analysis of the human distribution per region was performed. The correlation plots per location showed different patterns while still maintaining specific hotspots among them (Figure 8).

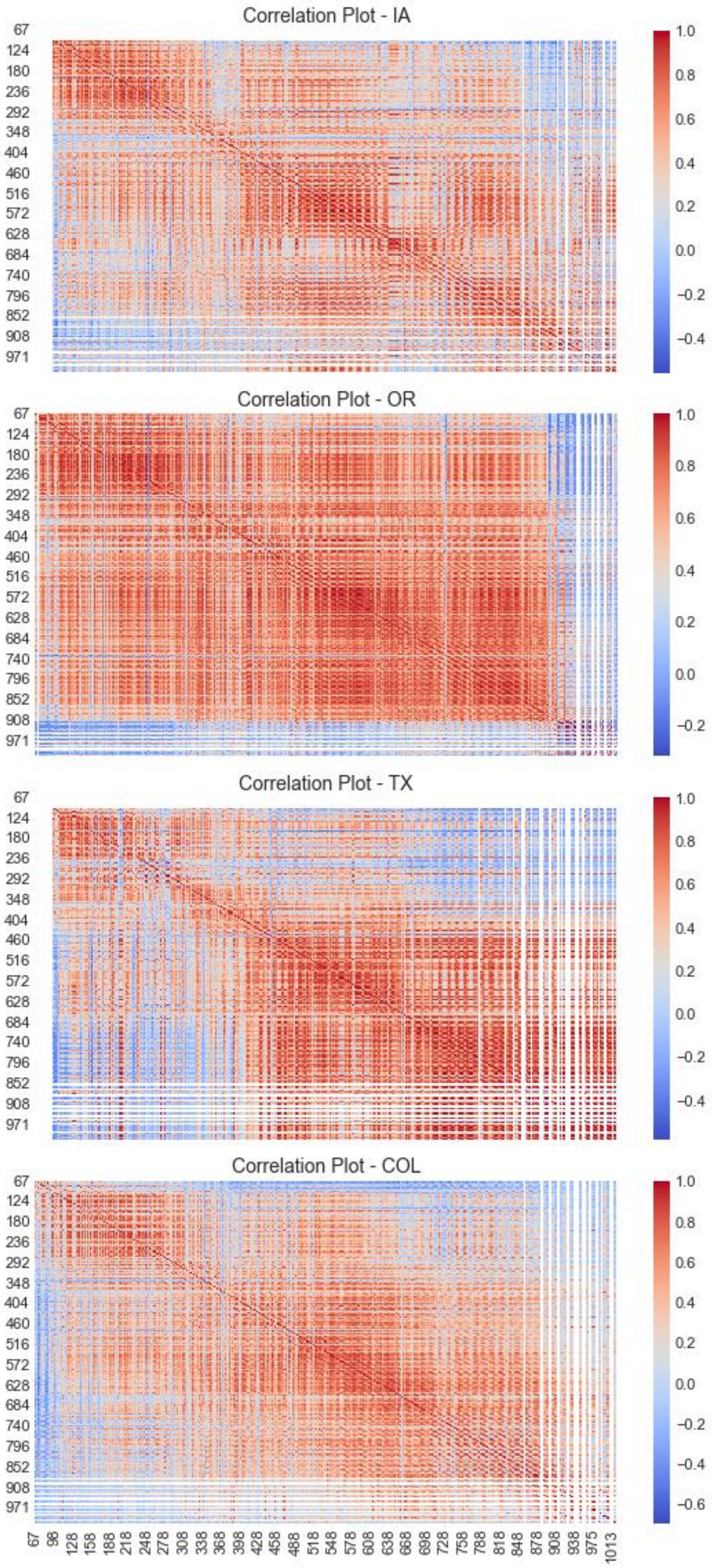

Correlation Plot - IA
Correlation Plot - OR
Correlation Plot - TX
Correlation Plot - COL
67
124
180
236
292
348
404
460
516
572
628
684
740
796
852
908
971
67
98
128
158
188
218
248
278
308
338
368
398
428
458
488
518
548
578
608
638
668
698
728
758
788
818
848
878
908
938
975
1013
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6

Figure 8. Correlation plots per regional human group: Iowa (IA), Oregon (OR), Texas (TX) & Colombia (COL).

Thus, an LDA was performed to find if there was a pattern that could be detected based on differences between the regions. The results (Figure 9) showed that there was indeed a clustering trend with the samples gathered from Colombia forming the most distinct group while still distinguishing those from Texas, Iowa and Oregon, indicating that there is indeed significant variation across the human class samples.

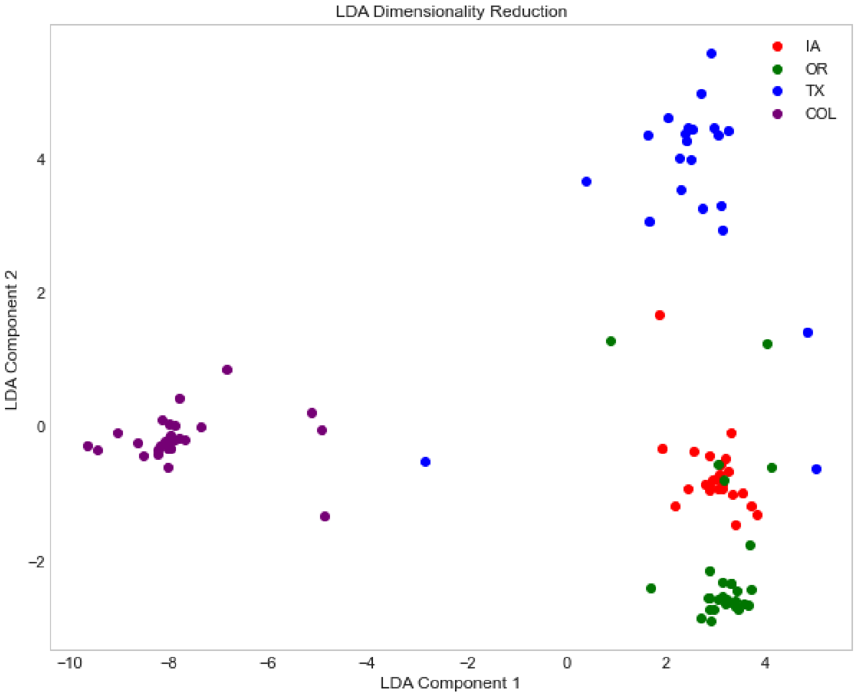


Figure 9. LDA plot of the dog, human and cat spectra.

4 Results

4.1 Random Forest

Random Forest was selected due to the sparsity of the data as well as its complex relationships. The first Random Forest model was created with default parameters of 100 decision trees, Gini as the criterion, and a minimum sample split of 2. Given that it led to 100% scores for all scores (accuracy, precision, F1 and recall), overfitting immediately became a concern. For that reason, a 10-fold cross validation was added to the model to verify its performance while using the same parameters. The resulting mean accuracy, precision, F1, and recall scores of 97% (Table 1) supported the theory that this model was robust to overfitting. To further verify the performance of the model and check the bias-variance trade off, a Leave-One-Out-Cross-Validation (LOOCV) was performed. The LOOCV achieved the same high scores of 97- 98% (Table 1). These results showed that the model performed well and was capable of generalizing without overfitting to the training set. Additionally, there was evidence to support the claim that the model had low bias, was robust and had a relatively low standard deviation.

4.2 Logistic Regression

Given that the random forest algorithm performed well, a multiclass Logistic Regression model was created to better interpret the random forest model and gain statistical inferences regarding its performance. Based on the results of the EDA, the Logistic Regression model was not expected to perform as well as the random forest models due to the high scale of variability within the classes. Unlike random forest, Logistic Regression is sensitive to high variance, making it necessary to scale the data. Our logistic regression model was created using the default L2 penalty with an alpha/lambda value of 1. These parameters and corresponding values were selected due to the high dimensional nature of the data and to help control the complexity of the model while still reducing the impact of features that are statistically insignificant. This model ultimately led to lower results when compared to Random Forest, but still performed adequately with accuracy, recall and precision scores of 94% (Table 1). To assess the Logistic Regression performance, Cross Validation, using the same parameters, was applied. This led to the same mean scores of 94% , but a different standard deviation of +/- 10, showing that scores could be as low as 84% and as high as 100%. This makes intuitive sense as we previously noted that there are some complexities in the data that the Logistic Regression model would not be able to define. While the Logistic Regression model was not as robust as the various, previous Random Forest models applied, it showed that there are distinct patterns between the classes that even Logistic Regression was able to retrieve. More importantly, applying Logistic Regression to the data showed that ions in the range of 653 and 708 *m/z* were the most important features for differentiating between each class.

Table 1. Test results from different models

Method	**Accuracy**	**Precision**	**F1 Score**	**Recall**
Random Forest	97%	98%	98%	98%
Cross Validation Random Forest	97%	97%	97%	97%

LOO Cross Validation Random Forest	98%	99%	98%	100%
Logistic Regression	73%	94%	94%	94%
Cross Validation Logistic Regression	94%	94%	93%	94%

5 Discussion

Utilizing machine learning approaches to distinguish between DART TOF mass spectra collected from human, cat, and dog hair fibers was a successful approach. Random Forest modeling was able to classify the taxonomic origin of hair fibers with scores as high as 97%, indicating that this approach is reliable and a good alternative to, or at the very least an additional verification to, microscopic analysis. In addition, utilizing these tools can provide unbiased and accurate information in a short amount of time.

From a taxonomic perspective, it is logical that the classification of these species was correct given that humans, cats, and dogs exist in different taxonomic orders. More surprising was that we observed distinct trends within the humans based on the location where the samples were obtained, however we acknowledge that there are no statistical grounds to indicate or assume causality. However, this did suggest that the random forest model was robust to the differences found in the human hair groups collected from various regions. Simply put, this would imply that the model would remain effective if applied to hair fibers collected from different regions and warrants further studies focused on the geological differences between a larger population of samples. However, while the distinction between human groups across broad geographic ranges is important to the global forensics' community, the goal of this research was to create an accurate method that could be used to differentiate cat, dog and human hair fibers. Therefore, we could not further pursue the meaning of this interesting trend observed within this study.

Given the complexities in each of the class datasets, it was interesting to find that Logistic Regression was able to predict with an F1 score of 94%, which presents better performance in the true positives. The Logistic Regression results provide a unique opportunity for further research to determine the identity of those unique ions that were the most important for distinguishing the classes. In this research, the ions were binned to their nominal masses, and we did not attempt to identify any of the ions found within the datasets. However, from our observations we could see that the ions at 679, 654, 680, 693, 707, 395, 647, 682 and 381 *m/z* were the most statistically significant for distinguishing between the human, cat, and dog classes.

While our analysis proved successful, there were limitations that need to be considered for larger scale implementation. The dataset was limited by size, while we had sufficient samples to create reliable models, we acknowledge that we would see improvement in the standard deviation scores by increasing sample sizes.

From an ethical perspective, all samples were collected with anonymity towards each donor (i.e., no personal identifying information such as names, race, age, or sex, was collected) and the samples were destroyed during the mass spectra collection process. Therefore, no other instrumental analyses can be accomplished with the samples themselves, and the samples used cannot be connected back to their original donors. In addition, samples were collected using scissors which did not harm any dog, cat, or human within this study. The authors feel that conducting future replicate studies with such datasets offer no potential for harm.

6 Conclusion

The findings of this research successfully demonstrated that DART TOF mass spectra in conjunction with machine learning algorithms can be used for distinguishing the species of hair samples taken from humans, cats, and dogs. This study further contributes to the field of forensics but also offers a method that could be used to gain insightful information from museums or archaeological items.

To further enhance the robustness of these findings, future research should include increasing the sample size for each group and by including samples from a broader range of geographic locations. This approach would help capture more potential variation for each of the groups.

Acknowledgments

The authors would like to thank the Friends of the Animal Shelter, Talent, OR., the Animal Shelter, Georgetown, TX., as well as all people who donated their own and their pets' hair to this study. In addition, we would like to thank Jacquelyn Cheun, PhD, SMU, for her thoughtful and thorough advice for this capstone. The authors declare that they have no competing interests.

References

1. Appley, M.G., Beyramysoltan, S., & Musah, R.A. (2019). "Random Forest Processing of Direct Analysis in Real-Time Mass Spectrometric Data Enables Species Identification of Psychoactive Plants from Their Headspace Chemical Signatures". ACS Omega 2019 4 (13), 15636-15644 DOI: 10.1021/acsomega.9b02145
2. Balakrishnama, S., & Ganapathiraju, A. (1995, August 3). LINEAR DISCRIMINANT ANALYSIS - A BRIEF TUTORIAL. http://www.music.mcgill.ca/~ich/classes/mumt611_07/classifiers/lda_theory.pdf
3. Garrett, B.L. Bad Hair: The Legal Response to Mass Forensic Errors, 42 Litigation 32-36 (2016)
4. Breiman, L. (2001). 1 random forests - University of California, Berkeley. https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf
5. Cates, P., Dominguez, I., Pierce, E., & Kortan, M. P. (2015, April 20). FBI Testimony on Microscopic Hair Analysis Contained Errors in at Least 90 Percent of Cases in Ongoing Review. FBI. https://www.fbi.gov/news/press-releases/fbi-testimony-on-microscopic-hair-analysis-contained-errors-in-at-least-90-percent-of-cases-in-ongoing-review
6. Cody, R.B., Laramée, J.A., & Durst, H.D. (2005). Versatile New Ion Source for the Analysis of Materials in Open Air under Ambient Conditions. Anal. Chem. 77(8):2297-2302. 10.1021/ac050162j
7. Cole, K. (2017). Intrinsic Factors Affecting Decomposition Changes in Archaeological Head Hair from Kellis 2 Cemetery, Egypt. University of Central Florida STARS. https://stars.library.ucf.edu/etd/5351/
8. Dinov, I.D., Christou, N., Gould, R. (2009). Law of Large Numbers: the Theory, Applications and Technology-based Education. J Stat Educ. 2009 Mar;17(1):1-19. doi: 10.1080/10691898. 11889499.
9. Greenwold, M. J., Bao, W., Jarvis, E. D., Hu, H., Li, C., Gilbert, M. T. P., … Sawyer, R. H. (2014). Dynamic evolution of the alpha (α) and beta (β) keratins has accompanied integument diversification and the adaptation of birds into novel lifestyles. BMC Evolutionary Biology, 14(1). doi:10.1186/s12862-014-0249-1

10. Gwinnett, C. (2016). “Chapter 6: Forensic Examination of Animal Hair”. Practical Veterinary Forensics (ed. D. Bailey). CAB International.
11. Jahanbanifard, M., Price, E., González, B.A., Raggi, L.A., Javanmardi, S., Lens, F., Gravendeel, B., Espinoza, E., Verbeek, F.J. (2023). A novel method to analyse DART TOFMS spectra based on Convolutional Neural Networks: A case study on methanol extracts of wool fibres from endangered camelids. 489. 1-8. https://doi.org/10.1016/j.ijms.2023.117050
12. Kreplak, L., Franbourg, A., Briki, F., Leroy, F., Dallé, D., & Doucet, J. (2002). A New Deformation Model of Hard α-Keratin Fibers at the Nanometer Scale: Implications for Hard α-Keratin Intermediate Filament Mechanical Properties. Biophysical Journal, 82(4), 2265–2274. doi:10.1016/s0006-3495(02)75572-0
13. Lennert, E., & Bridge, C. (2018). Analysis and classification of smokeless powders by GC-MS and DART-TOFMS. Forensic Sci Int. 2018 Nov; 292:11-22. doi: 10.1016/j.forsciint.2018.09.003.
14. Liu, Z., Simayijiang, H., Wang, Q., Yang, J., Sun, H., Wu, R., & Yan, J. (2023, February 3). DNA and protein analyses of hair in forensic genetics - International Journal of Legal Medicine. SpringerLink. https://link.springer.com/article/10.1007/s00414-023-02955-w
15. Norton, H.J., Anderson W.E. & Divine, G. (2016). Flawed forensics: Statistical failings of microscopic hair analysis. In Detail https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1740-9713.2016.00897.x
16. Oien, C. (2009). Forensic Hair Comparison: Background Information for Interpretation. Federal Bureau of Investigation: Forensic Science Communications. https://archives.fbi.gov/archives/about-us/lab/forensic-science-communications/fsc/april2009/review/2009_04_review02.htm
17. O'Connor, S., Solazzo, C., & Collins, M. (2015) Advances in identifying archaeological traces of horn and other keratinous hard tissues, Studies in Conservation, 60:6, 393-417, DOI: 10.1179/2047058414Y.0000000134
18. Petherick, W. A., Turvey, B. E., & Ferguson, C. E. (2010). Edmond Locard. The Forensics Library. http://aboutforensics.co.uk/edmond-locard/
19. Price, E. R., Larrabure, D. M., Gonzales, B., McClure, P. J., & Espinoza, E. O. (2020). Forensic Identification of the Keratin Fibers of South American Camelids by Ambient Ionization Mass Spectrometry: Vicuña, Alpaca and Guanaco. Rapid communications in mass spectrometry: RCM, e8916. https://pubmed.ncbi.nlm.nih.gov/32770752/
20. Price, E. R., McClure, P. J., Jacobs, R. L., & Espinoza, E. O. (2018). Identification of rhinoceros keratin using direct analysis in real time time-of-flight mass spectrometry and multivariate statistical analysis. Rapid communications in mass spectrometry: RCM, 32(24), 2106-2112. https://pubmed.ncbi.nlm.nih.gov/30230063/

21. Sisco, E., & Forbes, T. (2021, November 19). Forensic applications of DART-MS: A Hair histology as a tool for forensic identification of some domestic animal species review of recent literature. Forensic Chemistry, 22. https://www.sciencedirect.com/science/article/pii/S2468170920300825
22. Vaclavik, L., Hrbek, V., Cajka, T., Rohlik, B.-A., Pipek, P., & Hajslova, J. (2011). Authentication of Animal Fats Using Direct Analysis in Real Time (DART) Ionization−Mass Spectrometry and Chemometric Tools. Journal of Agricultural and Food Chemistry, 59(11), 5919–5926. doi:10.1021/jf200734x
23. Wilkinson, L. & Gwinnett, C. (2020). An international survey into the analysis and interpretation of microscopic hair evidence by forensic hair examiners. Forensic Science International. 308:1-15. https://doi.org/10.1016/j.forsciint.2020.110158.
24. Zafarina Z, Panneerchelvam S. Analysis of hair samples using microscopical and molecular techniques to ascertain claims of rare animal species. Malays J Med Sci. 2009 Jul;16(3):35-40. PMID: 22589663; PMCID: PMC3329137.

If you made it this far, you’re amazing.