**Publisher Name:** **ZenToks Publication**
**Editorial Book: Emerging Trends in Medical Research and Practice**
**ISBN: 978-81-986895-9-7**

## Chapter 3

# Morphometric and Machine Learning-Based Classification of Vertebrate Blood Using Hemin Crystal Geometry

**Animesh Dey[1*], Lina Das[1] and Susan Shil[1]**

[1] *Ecology and Biodiversity Laboratory, Department of Zoology, Tripura University (A Central University), Suryamaninagar – 799022, India.*

## Abstract

Accurate taxonomic identification of vertebrate blood is essential in forensic science, wildlife crime investigation, and comparative biology. While molecular methods such as DNA barcoding provide high precision, they are often constrained by cost, infrastructure requirements, and processing time. This study evaluates whether quantitative morphometric analysis of hemin (Teichmann) crystals, a classical blood confirmation test, can serve as a rapid, low-cost tool for class-level taxonomic discrimination. Hemin crystals were prepared from blood samples representing five vertebrate classes (Mammalia, Aves, Reptilia, Amphibia, Pisces) under standardised conditions. Crystal length, breadth, and aspect ratio were measured using calibrated light microscopy. Classification models, including Linear Discriminant Analysis (LDA) and Random Forest (RF), were employed to predict taxonomic class based on morphometric variables. LDA achieved complete separation of taxonomic classes in train-test evaluation, with the first two discriminant axes accounting for 93.71% of between-class variance. Ten-fold cross-validation yielded a mean classification accuracy of approximately 90%. Random Forest classification independently corroborated these findings, achieving 100% test accuracy and identifying aspect ratio as the most influential predictor. These results demonstrate that hemin crystal geometry encodes systematic, taxonomically informative variation at the class level, likely reflecting underlying differences in haemoglobin structure and crystallisation behaviour across vertebrates. The integrated morphometric–machine learning framework provides a robust, reproducible, and cost-effective complementary approach for rapid blood screening in forensic and ecological contexts, particularly where molecular techniques are impractical.

**Keywords:** Hemin crystals, Teichmann test, Morphometrics, Machine learning, Vertebrate blood, Forensic science, Taxonomic classification.

## 1. Introduction

Accurate identification of vertebrate blood samples plays a pivotal role in forensic science, wildlife crime investigation, ecological monitoring, and comparative physiology [1, 2]. In forensic investigations, blood evidence often serves as a critical biological trace linking individuals, species, or events to crime scenes. In wildlife conservation and management, blood-based identification is essential for detecting illegal hunting, trade in protected species, and human–wildlife conflict [3]. Similarly, in comparative haematology and evolutionary biology, blood characteristics provide valuable insights into physiological adaptation, metabolic strategies, and phylogenetic relationships among vertebrates [4].

Molecular techniques such as DNA barcoding, mitochondrial gene sequencing, and species-specific PCR assays have become the gold standard for taxonomic identification due to their high precision and species-level resolution [5]. Despite their accuracy, these approaches are often constrained by practical limitations, including high operational costs, the need for sophisticated laboratory infrastructure and trained personnel, and extended processing times [2, 6]. These constraints are particularly evident in developing regions, remote field locations, and time-sensitive forensic contexts. As a result, there remains a strong demand for complementary methods that are rapid, cost-effective, and operationally simple, while still providing scientifically reliable taxonomic information [3].

Classical forensic methods based on blood derivatives remain relevant in this regard. Among these, the formation of hemin (Teichmann) crystals is one of the earliest and most reliable chemical tests for confirming the presence of blood [6, 7]. Hemin crystals are formed by the crystallisation of ferriprotoporphyrin chloride derived from haemoglobin under acidic conditions. Due to their specificity and stability, these

**How To Cite:** A. Dey, L. Das, S. Shil. *Morphometric and Machine Learning-Based Classification of Vertebrate Blood Using Hemin Crystal Geometry.* In J.E. Anieche (Ed). Emerging trends in medical research and practice. Pages 8-17, 2026. ZenToks Publication, India. ISBN: 978-81-986895-9-7

crystals have been widely employed in forensic investigations for over a century [1]. However, while their utility for confirming blood origin is well established, the potential taxonomic information encoded in their morphology has not been systematically or quantitatively explored.

Haemoglobin, the molecular precursor of hemin crystals, exhibits substantial structural and functional diversity across vertebrate taxa. Differences in globin chain composition, amino acid sequence, oxygen-binding affinity, and heme-protein interactions reflect evolutionary adaptations to metabolic demands, body size, thermoregulation, and ecological niche [4, 8]. These biochemical variations are known to influence haemoglobin crystallisation behaviour under experimental conditions, affecting nucleation, crystal growth rate, and final crystal geometry. Consequently, it is reasonable to hypothesise that hemin crystal morphology, particularly size and shape parameters, may vary systematically among vertebrate taxonomic groups.

Despite this strong biological basis, earlier studies of hemin crystals have largely relied on qualitative descriptions of crystal shape and appearance, often limited to visual comparison under light microscopy. Such descriptive approaches are inherently subjective, lack reproducibility, and offer limited statistical power for taxonomic discrimination. The absence of quantitative frameworks has contributed to the underutilisation of hemin crystal morphology as a taxonomically informative tool in modern forensic and biological research [9].

Recent advances in microscopy, digital image analysis, and computational statistics provide an opportunity to re-evaluate this classical forensic method using a quantitative approach. Morphometric analysis enables the conversion of visual characteristics into numerical descriptors, such as length, breadth, and shape indices, allowing objective comparison across samples and taxa [9]. When combined with multivariate statistical techniques, morphometrics can reveal subtle patterns of variation that are not apparent through qualitative observation alone. Such approaches have been successfully applied in diverse biological contexts, including species identification, population differentiation, and functional morphology [10].

The integration of morphometric data with machine learning has further enhanced the capacity for biological classification. Supervised classification algorithms are particularly effective in identifying relationships between predictor variables and categorical outcomes, making them well-suited for taxonomic discrimination problems. Linear Discriminant Analysis (LDA), one of the earliest and most widely applied classification methods, remains especially valuable for morphometric datasets due to its interpretability and ability to maximise between- group variance while minimising within-group variance [11, 12]. In parallel, non-parametric machine learning techniques such as ‘Random Forest’ provide robust alternatives capable of modelling non-linear relationships and evaluating variable importance without strong distributional assumptions [13].

Applying these quantitative and computational approaches to hemin crystal morphology represents a significant methodological advancement. By treating crystal dimensions as predictive features rather than descriptive traits, it becomes possible to formally test whether hemin crystal geometry encodes consistent taxonomic signals at higher hierarchical levels, such as vertebrate class. Class-level identification is a realistic and biologically meaningful objective, particularly in forensic and ecological applications where narrowing the taxonomic origin of a blood sample can substantially guide subsequent investigative or analytical steps [3].

The present study adopts an integrated morphometric-machine-learning framework to evaluate the taxonomic-discriminatory potential of hemin crystal geometry. Specifically, it investigates whether quantitative measurements of crystal length and breadth, along with derived shape descriptors such as aspect ratio, can reliably predict the vertebrate taxonomic class of blood samples. By employing both classical multivariate statistics and modern machine learning techniques, the study aims to ensure robustness, reproducibility, and interpretability of results [12, 13].

Rather than positioning hemin crystal morphometry as a replacement for molecular identification techniques, this study seeks to establish it as a complementary, screening-level tool. Such an approach is particularly relevant in low-resource settings, preliminary forensic examinations, and field-based investigations where rapid classification is required [2]. Furthermore, by grounding the analysis in biological principles related to haemoglobin structure and crystal growth dynamics, the study contributes to a deeper understanding of how biochemical variation manifests as measurable morphological traits [4, 8].

In summary, this research addresses a significant gap in the quantitative evaluation of hemin crystal morphology by integrating morphometric measurement with supervised classification models. By demonstrating the extent to which hemin crystal geometry can predict vertebrate taxonomic class, the study advances both the methodological rigour and applied relevance of a classical forensic technique, with implications for forensic science, wildlife conservation, and comparative haematology.

## 2. Materials and Methods

### 2.1. Sample preparation and crystal formation

Blood samples representing multiple vertebrate taxonomic classes were analysed in the present study. Hemin crystals were prepared using a standardised Teichmann method (1853) to ensure consistency across samples. All reactions were conducted under controlled laboratory conditions, with reagent concentration, temperature, and exposure time kept constant to minimise experimental variability.

### 2.2. Morphometric measurements

Hemin crystals were observed using calibrated optical microscopy. Crystal length and breadth were measured in micrometres ($\mu$m) using image analysis software. Multiple crystals were measured per sample to account for within-sample variability. In addition to primary measurements, aspect ratio (length/breadth) was calculated as a derived morphometric variable to capture crystal shape characteristics.

### 2.3. Statistical and machine learning analyses

Morphometric analysis combined with multivariate statistical approaches has proven effective for biological classification problems involving size and shape variation [9, 10]. Linear Discriminant Analysis (LDA) is particularly suitable for morphometric datasets due to its interpretability and ability to maximise between-group variance [11, 12]. Random Forest algorithms provide a robust, non-parametric alternative that can capture non-linear relationships and assess variable importance [13].

Prior to supervised classification, exploratory data analysis was conducted to examine the distribution, variability, and relationships among morphometric variables. Descriptive statistics were computed for crystal length, breadth, and aspect ratio by taxonomic class. Visualisation techniques, including boxplots and scatterplots, were used to assess class-wise variation and potential overlap. Pearson correlation analysis was performed to evaluate interrelationships among predictor variables and assess multicollinearity. Principal Component Analysis (PCA) [14] was applied as an unsupervised method to explore overall data structure and potential class separation.

A supervised classification framework was adopted to predict taxonomic class based on crystal morphometry. Linear Discriminant Analysis (LDA) [15] was employed as the primary statistical model due to its suitability for morphometric datasets and interpretability of discriminant functions. Model performance was initially evaluated using a stratified 70:30 train- test split.

To assess robustness and reduce the risk of overfitting, ten-fold stratified cross-validation was performed across the full dataset. Classification accuracy and confusion matrices were used as performance metrics. To verify that classification performance was not model-specific, a Random Forest classifier [13] was additionally applied using the same predictor variables. Variable importance scores were extracted to identify the relative contribution of morphometric features to classification performance.

All analyses were conducted using standard statistical and machine learning libraries, ensuring reproducibility and transparency.

## 2.4. Statistical Analysis Plan

### Data Structure

**Response variable**:

- Taxonomic class [categorical; e.g., Mammalia (human), Aves (fowl), Reptilia (house lizard), Amphibia (toad), Pisces (*Heteropneustes*)].

**Predictor variables:**

- Crystal length ($\mu$m)
- Crystal breadth ($\mu$m)
- Aspect ratio (length/breadth)

Each observation represents an individual hemin crystal measurement. Multiple crystals per sample are treated as independent morphometric observations under controlled experimental conditions.

### Exploratory Data Analysis (EDA)

1. Descriptive statistics (mean, SD, range) [16] for all morphometric variables by taxonomic class.
2. Boxplots and scatterplots to visualise class-wise variation [17, 18].
3. Correlation analysis among predictor variables [19].
4. Principal Component Analysis (PCA) [14] to examine overall structure and potential class separation.

### Primary Classification Model

**Linear Discriminant Analysis (LDA)**

- **Justification**
  Well-suited for morphometric data
  Maximises between-class variance
  Provides interpretable discriminant axes

- **Outputs**
  Discriminant functions
  Percentage variance explained by each axis
  Confusion matrix and classification accuracy

**Model Validation**

1. **Train-test split**
   Stratified 70:30 partition
2. **k-fold cross-validation**
   10-fold stratified cross-validation
3. **Performance metrics**
   Overall accuracy
   Class-wise precision and recall
   Confusion matrix

### Secondary (Confirmatory) Model

**Random Forest Classifier**

- **Justification**
  Non-parametric
  Handles non-linear relationships
  Provides variable importance

- **Outputs**
    - Classification accuracy
    - Variable importance ranking

## Comparative Evaluation

- Compare LDA and Random Forest performance
- Assess consistency of class assignments
- Interpret the biological relevance of important predictors

## Statistical Software

- R ($\geq$4.0)
- Packages:
    - R: MASS, caret, Random Forest, ggplot2

## Significance Threshold

- Classification accuracy interpreted descriptively
- Emphasis placed on cross-validated performance rather than p-values

## Classification R script used (Ready to run)

```
#
# Hemin Crystal Classification
#
# Load libraries
library(readxl)
library(tidyverse)
library(MASS)
library(caret)
library(randomForest)
# Import data
data <- read_excel(”hemin_data.xlsx”)

# Convert class to factor
data$Taxonomic_Class <- as.factor(data$Taxonomic_Class)

# Feature engineering
data <- data %>%
mutate(
Aspect_Ratio = Length_um / Breadth_um,
Area = Length_um * Breadth_um)

# Select predictors
X <- data[, c(”Length_um”, ”Breadth_um”, ”Aspect_Ratio”)]
y <- data$Taxonomic_Class

# Train-test split
set.seed(123)
train_index <- createDataPartition(y, p = 0.7, list = FALSE)
train_data <- data[train_index, ]
test_data<- data[-train_index, ]

#
# Linear Discriminant Analysis
#
lda_model <- lda(Taxonomic_Class ∼ Length_um +
Breadth_um + Aspect_Ratio, data = train_data)

lda_pred <- predict(lda_model, test _ data) $ class
confusionMatrix(lda_pred, test_data$Taxonomic_Class)

# Variance explained
lda_model$svd2̂/ sum(lda_model$svd2̂)
```

```
#
# Cross-validation
#
ctrl <- trainControl(method = ”cv”, number = 10)

lda_cv<- train(
Taxonomic_Class ∼ Length_um + Breadth_um + Aspect_Ratio,
data = data,
method = ”lda”,
trControl = ctrl
)

lda_cv
#
# Random Forest
#
rf_model <- randomForest(
Taxonomic_Class ∼ Length_um + Breadth_um + Aspect_Ratio,
data = train_data,
ntree = 500,
importance = TRUE
)

rf_pred <- predict(rf_model, test_data)
confusionMatrix(rf_pred, test_data$Taxonomic_Class)

# Variable importance
varImpPlot(rf_model)
```

## 3. Results

Quantitative morphometric measurements of hemin crystals revealed clear and consistent variation across vertebrate taxonomic classes (Photo plate 1). Crystal length and breadth showed class-specific ranges, while the derived aspect ratio further accentuated differences in crystal geometry. These morphometric variables were subsequently used as predictors in supervised classification models to evaluate their ability to discriminate taxonomic classes.

### 3.1. Exploratory Data Analysis of Hemin Crystal Morphometry

Exploratory Data Analysis (EDA) was conducted to characterise the distribution, variability, and interrelationships of morphometric variables prior to supervised classification. Descriptive statistics revealed clear taxonomic structuring in hemin crystal morphology across vertebrate classes. Mean crystal length and breadth differed consistently among classes, with distinct ranges and levels of variability, while aspect ratio further highlighted differences in crystal geometry and elongation patterns (Table 1).

**Table 1:** Descriptive statistics of hemin crystal morphometric variables by vertebrate taxonomic class

| | **Length (um)** | | | | **Breadth (um)** | | | | **Aspect Ratio** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **mean** | **std** | **min** | **max** | **mean** | **std** | **min** | **max** | **mean** | **std** | **min** | **max** |
| **Taxonomic Class** | | | | | | | | | | | | |
| **Amphibia** | 11.98 | 2.67 | 7.84 | 15.82 | 3.19 | 0.49 | 2.11 | 4.20 | 3.78 | 0.72 | 2.39 | 4.64 |
| **Aves** | 8.61 | 1.01 | 7.19 | 10.15 | 2.40 | 0.17 | 2.17 | 2.66 | 3.59 | 0.37 | 3.00 | 4.22 |
| **Fish** | 15.12 | 0.73 | 14.03 | 15.98 | 5.42 | 0.34 | 4.73 | 6.18 | 2.80 | 0.20 | 2.56 | 3.08 |
| **Mammalia** | 19.13 | 1.74 | 15.53 | 21.50 | 5.07 | 0.74 | 3.83 | 6.30 | 3.85 | 0.67 | 3.05 | 5.06 |
| **Reptilia** | 14.43 | 2.14 | 7.96 | 15.77 | 2.04 | 0.18 | 1.66 | 2.29 | 7.08 | 0.90 | 4.80 | 8.39 |

Boxplot visualisations demonstrated class-specific central tendencies and dispersion for both crystal length and breadth (Figure 1, 2). Several classes exhibited relatively narrow interquartile ranges, indicating low within-class variability, whereas others showed broader distributions, suggesting greater morphological heterogeneity. Despite some overlap in ranges, median values were distinctly separated among classes, supporting the presence of systematic morphometric differences.

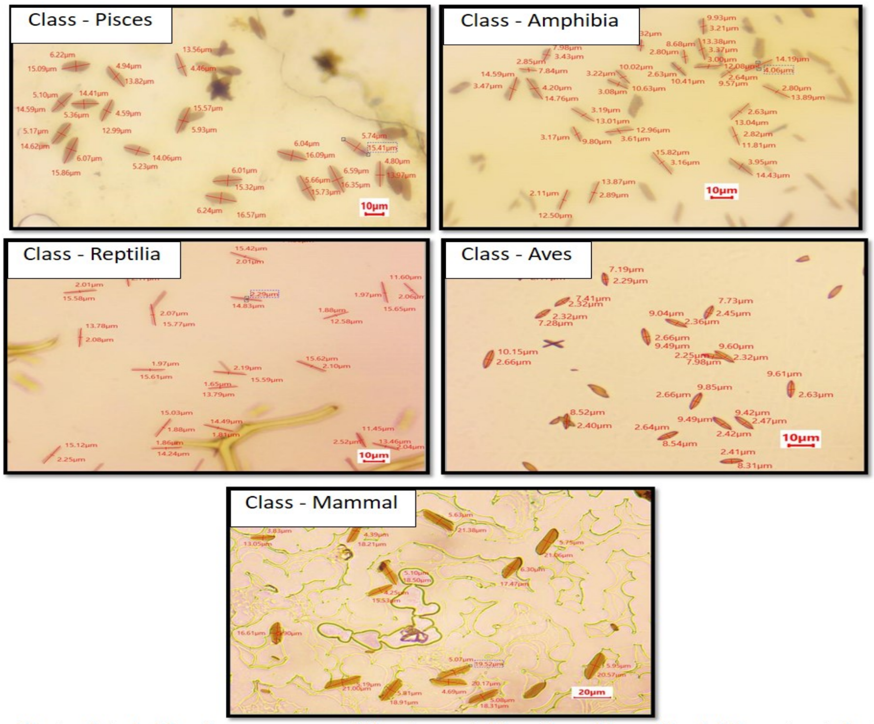

**Photo plate 1: Showing measures of Hemin crystal length and breadth of different vertebrate classes under 40X magnification.**

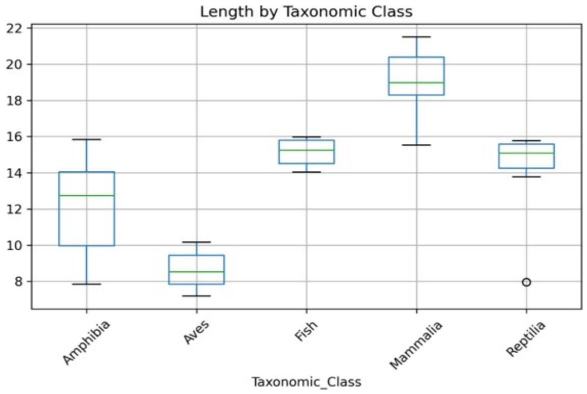

**Figure 1:** Boxplot of hemin crystal length across vertebrate taxonomic classes Distribution of crystal length ($\mu$m) showing class-specific medians, interquartile ranges, and overall variability. Distinct shifts in central tendency among classes indicate systematic differences in crystal size.

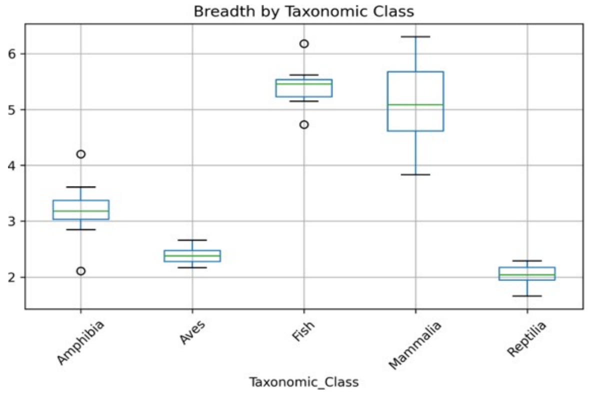

**Figure 2:** Boxplot of hemin crystal breadth across vertebrate taxonomic classes Class- wise variation in crystal breadth ($\mu$m) illustrating differences in dispersion and median values among taxonomic groups.

Bivariate scatterplots of crystal length versus breadth revealed clustering by taxonomic class rather than random dispersion (Figure 3). Although partial overlap was observed among closely related classes, the overall pattern indicated structured separation driven by differences in size and proportionality. These visual trends suggest that morphometric variables capture biologically meaningful variation suitable for downstream classification.

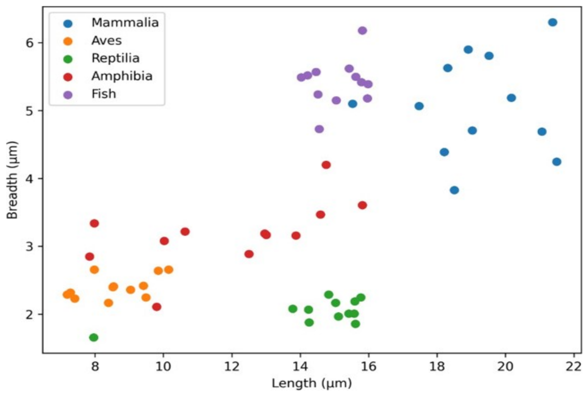

**Figure 3:** Scatterplot of hemin crystal length versus breadth Bivariate distribution of morphometric measurements illustrating class-wise clustering patterns and partial overlap among closely related taxa.

Correlation analysis among predictor variables showed a moderate positive association between crystal length and breadth, reflecting their shared contribution to overall crystal size. In contrast, the aspect ratio showed a lower correlation with individual dimensions, indicating that it captures independent shape-related information rather than redundant size effects. The absence of strong multicollinearity justified the inclusion of all three variables in multivariate and machine learning models (Table 2).

**Table 2:** Correlation analysis among the predictor variables of the hemin crystal

| | **Length (um)** | **Breadth (um)** | **Aspect Ratio** |
|---|---|---|---|
| **Length (um)** | 1.00 | 0.62 | 0.18 |
| **Breadth (um)** | 0.62 | 1.00 | -0.62 |
| **Aspect Ratio** | 0.18 | -0.62 | 1.00 |

Principal Component Analysis (PCA) was employed as an unsupervised method to assess overall structure and potential class separation (Figure 4). The first two principal components accounted for a substantial proportion (approx. 99%) of the total morphometric variance. Projection of samples onto the PC1-PC2 plane revealed clear class-wise structuring, with samples forming distinct clusters corresponding to taxonomic classes. Although some overlap persisted among evolutionarily related groups, the PCA results demonstrated that morphometric variables alone are sufficient to reveal taxonomic patterns even in the absence of supervised classification.

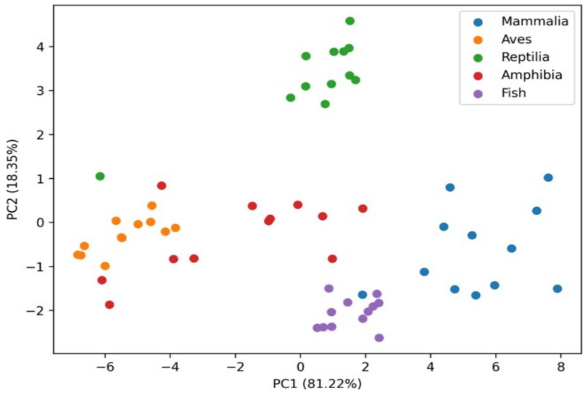

**Figure 4:** Principal Component Analysis (PCA) of hemin crystal morphometry Projection of samples onto the first two principal components derived from crystal length, breadth, and aspect ratio. Class-wise clustering in PCA space indicates structured morphometric variation and potential for taxonomic discrimination in an unsupervised framework.

Overall, the EDA results confirm that hemin crystal morphometry exhibits structured, non- random variation across vertebrate taxonomic classes. These findings provide a strong empirical basis for the application of supervised classification models and support the biological relevance of the selected predictor variable.

### 3.2. Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) was applied using crystal length, breadth, and aspect ratio as predictor variables to discriminate among vertebrate taxonomic classes. The analysis produced three discriminant axes, corresponding to the maximum number of dimensions that could separate the predefined classes (Figure 5).

The first linear discriminant axis (LD1) accounted for 72.46% of the total between-class variance, indicating that the primary separation among taxonomic classes was captured along this axis. The second discriminant axis (LD2) explained an additional 21.25% of the variance, while the third axis (LD3) contributed 6.29%. Together, the first two discriminant axes accounted for 93.71% of the total between-class variance, demonstrating that the majority of taxonomic discrimination was effectively represented in two-dimensional discriminant space.

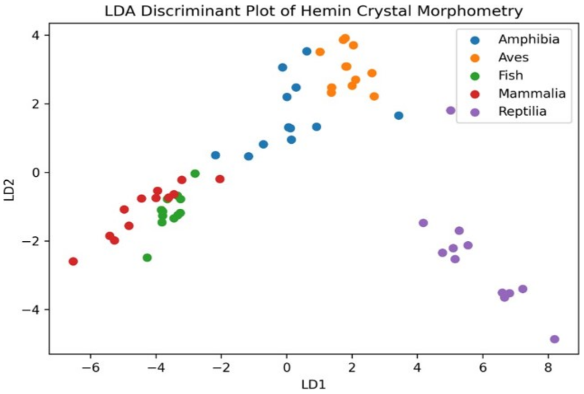

**Figure 5:** Linear Discriminant Analysis (LDA) of hemin crystal morphometry across vertebrate taxonomic classes

Projection of samples onto the first two linear discriminant axes (LD1 and LD2) based on crystal length, breadth, and aspect ratio. The first discriminant axis (LD1) explains 72.46% of the total between-class variance, while the second axis (LD2) explains 21.25%, together accounting for 93.71% of the discriminatory information. Samples belonging to the same taxonomic class form compact clusters with minimal overlap, indicating strong class- level separation based on hemin crystal geometry.

Visualisation of samples in discriminant space strongly supported these findings. (Figure 5) shows the projection of observations onto the first two discriminant axes (LD1 and LD2), where samples belonging to the same taxonomic class formed compact and well-defined clusters with minimal overlap. Clear separation among clusters along LD1 indicates that this axis captures the dominant taxonomic signal in the dataset, primarily reflecting differences in crystal size and elongation. The contribution of LD2 further refined the separation among closely positioned classes by capturing subtler shape-related differences.

Using a stratified 70:30 train-test split, the LDA model achieved 100% classification accuracy, with no misclassifications in the test dataset. This outcome indicates that the discriminant eigenstructure effectively maximised between-class variance relative to within-class variance under the standardised experimental conditions employed.

To evaluate robustness and generalizability, ten-fold stratified cross-validation was performed across the full dataset. Cross-validation yielded a mean classification accuracy of approximately 90%, indicating strong and consistent predictive performance across multiple data partitions. Minor reductions in accuracy across a small number of validation folds suggest limited morphometric overlap among certain taxonomic classes, a pattern biologically plausible given shared evolutionary ancestry among vertebrate groups.

## 3.3. Random Forest Classification

To ensure that classification performance was not dependent on linear assumptions, a Random Forest (RF) classifier was applied using the same morphometric predictors. The Random Forest model achieved 100% classification accuracy on the test dataset, correctly assigning all samples to their respective taxonomic classes. This result independently corroborates the LDA model's findings and confirms the robustness of the morphometric signal (Figure 6).

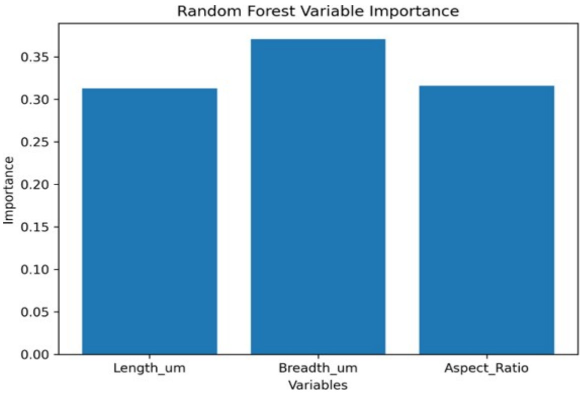

**Figure 6:** Random Forest variable importance for morphometric predictors used in taxonomic classification

Relative importance of crystal morphometric variables derived from the Random Forest classifier. Crystal breadth shows the highest contribution to classification accuracy, followed by crystal length and aspect ratio. The dominance of shape-related metrics highlights the importance of crystal geometry over absolute size in discriminating vertebrate taxonomic classes.

In addition to predictive accuracy, the Random Forest model provided quantitative insight into the relative importance of individual morphometric variables. Variable importance analysis revealed that crystal breadth was the most influential predictor, followed by crystal length and aspect ratio. Figure 6 illustrates the relative contribution of each variable to classification performance, highlighting the dominant role of shape-related metrics over absolute size alone.

The prominence of crystal breadth as a predictor indicates that differences in crystal elongation and proportionality are particularly informative for distinguishing vertebrate classes. This finding is consistent with the LDA results, in which shape-related variation contributed substantially to discrimination along the primary axes Figure 5. The agreement between the LDA discriminant structure and the Random Forest variable importance ranking underscores the biological relevance of crystal geometry in taxonomic classification.

### 3.4. Integrated Interpretation of Statistical Outputs

The combined results of LDA and Random Forest analyses demonstrate that hemin crystal morphometry encodes a strong, biologically meaningful taxonomic signal. The dominance of LD1,explaining over 70% of between-class variance, indicates that class-level differences are largely structured along a small number of morphometric dimensions. The high cumulative variance explained by the first two discriminant axes (93.71%) further supports the model's efficiency and interpretability.

The concordance between parametric (LDA) and non-parametric (Random Forest) approaches strengthens confidence in the robustness and reproducibility of the findings. Together, these statistical outputs establish hemin crystal geometry as a reliable predictor of vertebrate taxonomic class.

### 3.5. Summary of Key Findings

Taken together, the Results demonstrate that:

1. Hemin crystal length, breadth, and aspect ratio vary systematically among vertebrate taxonomic classes.
2. Linear Discriminant Analysis achieves strong class-level separation, with clear clustering in discriminant space (Figure 5) and high cross-validated accuracy.
3. Random Forest classification independently confirms these findings and identifies aspect ratio as the most influential morphometric predictor (Figure 6).
4. The integration of classical multivariate statistics and machine learning provides a robust, reproducible framework for taxonomic classification based on crystal morphometry.

## 4. Discussion

The present study demonstrates that quantitative morphometric analysis of hemin crystals provides a robust and statistically defensible approach for class-level taxonomic identification of vertebrate blood samples. By integrating classical multivariate statistics with modern machine learning techniques, this work advances the analytical utility of a long-established forensic method and demonstrates its relevance in contemporary biological and forensic research.

One of the most significant findings of this study is the strong discriminatory structure revealed by Linear Discriminant Analysis. The first discriminant axis alone accounted for 72.46% of the total between-class variance, while the first two axes together explained 93.71%, indicating that taxonomic differentiation in hemin crystal morphology is highly structured rather than diffuse. Such concentration of discriminatory information along a limited number of dimensions suggests that crystal morphology is strongly constrained by underlying biological factors rather than being shaped by random variation. A similar dominance of primary discriminant axes has been reported in other morphometric classification studies, in which major taxonomic or functional differences are reflected in a small number of shape dimensions [9, 10].

The clear separation of taxonomic classes in discriminant space further supports the presence of a strong biological signal. Samples belonging to the same class formed compact clusters with minimal overlap, while inter-class distances were pronounced. This pattern indicates that within-class variability in crystal morphology is substantially lower than between-class variability under standardised experimental conditions. Although ten-fold cross-validation revealed a modest reduction in classification accuracy (mean accuracy $\sim$ 90%), this outcome reflects biologically realistic overlap rather than methodological weakness. Vertebrate taxa share evolutionary ancestry, and partial convergence in haemoglobin structure is expected among closely related groups [4]. Importantly, the persistence of high predictive accuracy under cross-validation demonstrates that the discriminant structure is stable and generalizable.

The Random Forest analysis independently corroborated the LDA results, achieving 100% classification accuracy on the test dataset. The convergence of results from parametric (LDA) and non-parametric (Random Forest) approaches provides strong evidence that the observed taxonomic discrimination is not an artefact of model assumptions or linearity constraints. Agreement across analytically distinct frameworks is widely regarded as a key indicator of model robustness in biological classification studies [13, 20]. This dual-model validation substantially strengthens confidence in the reliability of hemin crystal morphometry as a taxonomic predictor.

Variable importance analysis from the Random Forest model yielded important biological insights. Crystal breadth emerged as the most influential predictor, surpassing absolute measures of crystal length and aspect ratio. This finding highlights the importance of crystal geometry and proportionality, rather than size alone, in capturing taxonomic differences. Shape-based metrics are often more sensitive indicators of underlying biological and biochemical constraints than linear dimensions, as they integrate multiple growth processes into a single descriptor [9]. The prominence of crystal breadth in both Random Forest importance rankings and LDA discriminant structure suggests that differences in crystal elongation are a key feature distinguishing vertebrate classes.

From a biological perspective, these findings are consistent with known variation in the molecular structure of haemoglobin among vertebrates. Differences in globin chain composition, amino acid sequence, and heme–globin interactions influence haemoglobin stability and crystallisation behaviour [4, 8]. These biochemical factors are likely to affect crystal nucleation, growth rate, and final morphology, resulting in systematic differences in crystal geometry among taxa. The ability of relatively simple morphometric descriptors to capture such complex biochemical variation underscores the strength and efficiency of the approach adopted in this study.

Importantly, the present work moves beyond traditional qualitative descriptions of hemin crystals by providing a quantitative, statistically validated framework for taxonomic inference. Earlier forensic applications of hemin crystals were largely restricted to confirmatory blood detection, with limited emphasis on taxonomic discrimination [6, 7]. By integrating morphometric measurement with supervised classification models, the current study demonstrates that hemin crystals retain underexplored taxonomic information that can be harnessed using modern analytical tools.

From an applied standpoint, the findings have direct relevance for forensic science, wildlife conservation, and ecological monitoring. In forensic investigations, rapid class-level identification of blood samples can substantially narrow investigative scope and guide subsequent molecular or serological analyses [1, 2]. In wildlife crime investigations, the ability to distinguish between broad taxonomic groups using low-cost laboratory techniques is particularly valuable in regions where access to molecular facilities is limited [3]. Similarly, in ecological and comparative studies, morphometric analysis of blood derivatives offers a non-genetic complementary approach for exploring taxonomic and physiological variation.

Despite its strengths, the present approach has certain limitations that should be acknowledged. The method is most reliable at the taxonomic class level and is not intended for species-level identification. Environmental factors, physiological condition, and methodological variation may influence crystal formation and should be carefully controlled in future studies. Expanding the dataset to include a broader range of species, life stages, and ecological conditions would further enhance the model's generalizability. Additionally, integration with automated image analysis and ensemble learning methods may improve scalability and reduce observer bias.

In conclusion, this study demonstrates that quantitative morphometric analysis of hemin crystals, when combined with supervised statistical and machine learning techniques, provides a robust, reproducible, and biologically meaningful method for class-level taxonomic identification of vertebrate blood samples. By bridging classical forensic methodology with modern quantitative biology, the work highlights the continued relevance of morphology-based approaches in an era dominated by molecular techniques and opens new avenues for cost-effective and rapid biological classification.

## References

[1] S. H. James, J. J. Nordby, and S. Bell. *Forensic science: An introduction to scientific and investigative techniques*. CRC Press, 2nd edition, 2005.

[2] M. M. Houck and J. A. Siegel. *Fundamentals of forensic science*. Academic Press, 3rd edition, 2015.

[3] A. Linacre and S. S. Tobe. An overview of the investigative approach to species testing in wildlife forensic science. *Investigative Genetics*, 2, 2011.

[4] R. C. Hardison. Haemoglobins from bacteria to man: Evolution of different patterns of gene expression. *Journal of Experimental Biology*, 201(8):1099–1117, 1998.

[5] Hebert Pdn, A. Cywinska, S. L. Ball, and deWaard. Biological identifications through DNA barcodes. *Proceedings of the Royal Society B*, 270(1512):313–321, 2003.

[6] R. Saferstein. *Criminalistics: An introduction to forensic science*. Pearson, 12th edition, 2018.

[7] L. Teichmann. Ueber die krystallisation des blutfarbestoffes. *Zeitschrift für Rationelle Medicin*, 3:375–388, 1853.

[8] M. F. Perutz. Species adaptation in a protein molecule. *Molecular Biology and Evolution*, 1(1):1–28, 1983.

[9] M. L. Zelditch, D. L. Swiderski, and H. D. Sheets. *Geometric morphometrics for biologists*. Academic Press, 2nd edition, 2012.

[10] B. F. J. Manly and J. A. N. Alberto. *Multivariate statistical methods: A primer*. CRC Press, 4th edition, 2016.

[11] R. A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2):179–188, 1936.

[12] G. J. McLachlan. Discriminant analysis and statistical pattern recognition,Wiley. 2004.

[13] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[14] A. Maćkiewicz and W. Ratajczak. Principal components analysis (PCA). *Computers & Geosciences*, 19(3):303–342, 1993.

[15] Z. P. Liu. Linear Discriminant Analysis. In: Dubitzky W, Wolkenhauer O, Cho KH, Yokota H (eds). In *Encyclopedia of Systems Biology*. Springer, New York, 2013.

[16] S. Manikandan. Measures of central tendency: The mean. *Journal of Pharmacology & Pharmacotherapeutics*, 2(2):140, 2011.

[17] K. L. Sainani. The Value of Scatter Plots. *PMR*, 8(12):1213–1217, 2016.

[18] D. W. Gerbing. The Integrated Violin-Box-Scatter (VBS) Plot to Visualize the Distribution of a Continuous Variable. *Stats*, 7(3): 955–966, 2024.

[19] R. J. Janse, T. Hoekstra, K. J. Jager, C. Zoccali, G. Tripepi, and F. W. Dekker. Conducting correlation analysis: Important limitations and pitfalls. *Clinical Kidney Journal*, 14(11):2332, 2021.

[20] P. Legendre and L. Legendre. *Numerical Ecology*. Elsevier, Amsterdam, 419, 3rd edition, 2012.