Transforming Ancestry Prediction: From Accuracy Gains to Forensic Applicability

Mr. Subhasish Swain
CSE Department
MITS
Rayagada, Odisha.
swain.subhasish4@gmail.com

Ms. Nivedita Mohapatra
CSE Department
MITS
Rayagada, Odisha.
niveditamhpt@gmail.com

Kailash Nayak
CSE Department
MITS
Rayagada, Odisha.
ritesmonada@gmail.com

Abstract-Forensic biogeographical ancestry (BGA) inference plays a vital role in human identification, yet existing state-of-the-art tools such as Snipper and the Admixture Model face limitations in accuracy, admixture handling, and interpretability. Recent advances in machine learning, particularly transformer-based tabular models such as TabPFN, have demonstrated superior performance over traditional forensic classifiers. However, key challenges remain underexplored, including intra-continental resolution, admixture proportion estimation, robustness to degraded DNA, and forensic interpretability. This study addresses these gaps by extending machine learning applications in BGA prediction beyond pure classification. Specifically, we investigate optimized marker panels tailored for intra-continental discrimination, propose hybrid models capable of both classification and admixture proportion estimation, and explore explainability techniques to generate marker-level insights for forensic admissibility. Experimental evaluations will benchmark these methods against established approaches on published datasets, incorporating tests for robustness under missing or noisy genetic data. By combining technical innovation with forensic-oriented evaluation, this research aims to bridge the gap between machine learning performance gains and their practical adoption in forensic casework, offering models that are accurate, interpretable, and operationally feasible.

Keywords: Forensic biogeographical ancestry (BGA), Machine learning, Hybrid models, Admixture Model, Admixture proportion estimation, Classification, Interpretability

I. INTRODUCTION

Classification of individuals into ancestral populations, or *biogeographical ancestry (BGA) inference*, is a fundamental task in forensic genetics. By analysing DNA traces, BGA can provide valuable leads in investigative contexts, assist in the identification of disaster victims, and contribute to population-level genetic studies. In forensic applications [1], individuals are commonly classified into broad continental groups—such as Africa, Europe, East Asia, South Asia, Oceania, and the Americas—or into finer intra-continental subpopulations. This process typically relies on a two-step procedure: the selection of Ancestry Informative Markers (AIMs) that exhibit significant allele frequency differences between populations, followed by the application of statistical or machine learning algorithms to assign individuals to the most likely population group.

Traditional forensic tools such as Snipper (naïve Bayes classifier), multinomial logistic regression, and the Admixture Model (as implemented in STRUCTURE and ADMIXTURE) have been widely used for ancestry inference. More recently, machine learning approaches—including XGBoost, Random Forests, and Partial Least Squares Discriminant Analysis (PLS-DA)—have demonstrated higher accuracy in some classification tasks. These models capture nonlinear marker interactions and can improve predictive performance compared to earlier methods. However, several challenges remain unresolved: (i) the accurate classification of genetically similar populations, especially at the intra-continental level; (ii) handling admixed individuals, whose genomes reflect multiple ancestries; (iii) optimizing marker panels to balance forensic cost constraints with predictive power; and (iv) addressing the issue of missing or degraded data, which is common in real-world forensic casework.

Recent advances in machine learning, particularly the introduction of transformer-based models such as the TabPFN foundation model for tabular data, have shown promising improvements in BGA classification accuracy and calibration. For example, TabPFN has outperformed traditional forensic classifiers on benchmark datasets, particularly for continental-level classification. Yet, limitations persist: while TabPFN achieves high accuracy, it lacks native support for admixture proportion estimation, offers limited interpretability for forensic reporting, and requires substantial computational resources not always feasible in operational forensic laboratories.

These limitations highlight important research gaps. Current approaches under explore how to:

1. Develop models that jointly predict both class membership and ancestry proportions for admixed individuals;

2. Optimize AIM panels specifically for advanced machine learning models to maximize intra-continental resolution with fewer markers;

3. Enhance the explainability and calibration of predictions to ensure forensic admissibility; and

4. Improve robustness against missing or degraded DNA profiles.

This study seeks to build upon prior work by addressing these gaps. By integrating feature-selection strategies, explainable machine learning methods, and robustness testing into the ancestry prediction pipeline, our research aims to advance the forensic applicability of BGA inference. Ultimately, the goal is not only to improve predictive accuracy but also to deliver models and protocols that are interpretable, computationally feasible, and validated for real-world forensic scenarios.

II. LITERATURE REVIEW

Biogeographical ancestry (BGA) inference has long been a critical area in forensic genetics, where the goal is to classify individuals into geographical groups or estimate their ancestry fractions based on genetic markers. Traditional tools, such as **SNIPPER [9]** and **Admixture models**, have been widely adopted due to their simplicity and established forensic credibility. However, the increasing complexity of genetic data, coupled with the limitations of marker sets in distinguishing closely related populations, has driven researchers toward **machine learning (ML)** approaches. Recent studies highlight the promise of advanced ML models — especially **transformer-based architectures such as TabPFN** — which outperform conventional methods on standard forensic marker panels. Nevertheless, challenges remain around **admixture quantification, explainability, robustness to degraded samples, and intra-continental resolution**. The following literature review summarizes these developments, outlines current limitations, and identifies research gaps that motivate further study.

---

1. Traditional Forensic Ancestry Classifiers

Early approaches like **SNIPPER** (naïve Bayes classifier) and the **Admixture model** have been the backbone of forensic ancestry inference. SNIPPER [9] is valued for its simplicity but assumes independence among markers, which limits resolution, particularly in closely related populations. The Admixture model estimates ancestry proportions, offering a more nuanced representation but still relies on model assumptions that may not hold across diverse populations. These tools remain useful in forensic contexts but show declining performance when fine-scale resolution is required.

---

2. Machine Learning in Ancestry Prediction

Recent works, including Heinzel et al. (2024), demonstrate the effectiveness of ML models such as **XGBoost, Random Forest, PLS-DA, and TabPFN**. Among these, TabPFN — a pre-trained transformer for tabular data — achieved the highest performance on the VISAGE-enhanced marker set, raising continental-level accuracy from **84% (SNIPPER) [9] to 93%**, and intra-European accuracy from **43% to 48%**. This reflects the ability of ML models to capture **complex joint structures among markers**. However, forensic application requires not just accuracy but **interpretability, calibration, and robustness**, areas where ML still lags behind traditional models.

---

3. Challenges in Intra-Continental Resolution

Although ML improves continental ancestry prediction, **intra-continental classification remains challenging**, especially in Europe, where genetic differentiation between populations is subtle. Even with TabPFN, accuracies hover below 50% for intra-European classification, with frequent misclassifications (e.g., CEU ↔ GBR, RUS → FIN). This suggests that existing AIM panels are not optimized for fine-scale differentiation, and feature selection strategies tailored to modern ML methods may be necessary.

---

4. Admixture Quantification Gap

Forensic casework often involves admixed individuals, where discrete classification fails. While the Admixture model estimates ancestry fractions, most ML models, including TabPFN, produce only class probabilities. The absence of ML approaches capable of **direct admixture proportion estimation** represents a key research gap. Future models that integrate both **classification and admixture estimation** would enhance forensic utility and better reflect real-world genetic complexity.

---

5. Explainability and Interpretability in Forensics

Forensic science requires transparency, as results must be admissible in court. Traditional models offer relatively interpretable outputs, while ML methods — especially transformers — are **black boxes**. Literature suggests the need for **explainable AI (XAI)** in ancestry prediction, using tools like SHAP values, rule-based models, or hybrid interpretable systems. Without this, the forensic community may resist adopting ML approaches despite their superior accuracy.

---

6. Robustness to Degraded or Missing Data

Forensic DNA samples are often **partial, degraded, or contaminated**, conditions rarely simulated in benchmarking studies. Current ML evaluations, including Heinzel et al., rely on curated datasets and do not account for missingness or genotyping errors. Developing ML approaches that remain reliable under real forensic conditions — through imputation, augmentation, or specialized training — is an open research need.

---

7. Operational and Resource Considerations

Models like TabPFN demand significant GPU/VRAM resources, which may be impractical in many forensic laboratories. The literature points toward **model distillation, lightweight ensembles, or optimized AIM panels** as strategies to balance performance and resource efficiency. This operational perspective is critical for transitioning from research to practice.

---

8. Emerging Directions in Fairness and Bias Analysis

Recent AI literature emphasizes **fairness, bias, and calibration** in classification models. For ancestry prediction, biased datasets may lead to systematic misclassifications in underrepresented groups (e.g., Central Asia). Although noted in prior studies, rigorous fairness evaluations are still lacking, representing a key future direction for ethically robust forensic ML.

III. COMPONENTS USED IN PROPOSED SYSTEM

The suggested smart pet feeder consist of hardware component overview is provided in this section. To provide the user with insights, the ESP-32 Microcontroller is connected with speaker, Load sensor and HX711, LCD and I2C, Bread board, Bowl, Motion senser which is Ultra Sonic having main functionality in system, and both are controlled by microcontroller.

ESP-32: The ESP32 is a powerful microcontroller developed by Espressif Systems. It has gained popularity for its versatility, low cost, and extensive connectivity options, making it suitable for a wide range of IoT (Internet of Things) applications, including smart pet feeders. One of the key features of the ESP32 is its dual-core processor, which allows for multitasking and efficient handling of various tasks simultaneously. This makes it well-suited for applications where real-time processing and responsiveness are required, such as monitoring sensors and controlling actuators in a smart pet feeder. Additionally, the ESP32 offers built-in Wi-Fi and Bluetooth connectivity, allowing devices to communicate with each other and with the internet. This enables remote monitoring and control of the pet feeder, as well as integration with other smart home devices and services. The ESP32 also supports a wide range of peripheral interfaces, including GPIO pins, SPI, I2C, UART, and ADC, which can be used to interface with sensors, displays, motors, and other components commonly found in smart pet feeders. This flexibility allows developers to easily customize and expand the functionality of their pet feeder according to specific requirements. Furthermore, the ESP32 is supported by a rich ecosystem of development tools and libraries, making it accessible to both hobbyists and professional developers. Arduino IDE, PlatformIO, and Espressif's own ESP-IDF are among the popular development environments for programming the ESP32. In the context of smart pet feeders, the ESP32 can be used to create a robust and feature-rich solution that offers automated feeding schedules, remote monitoring, and control, as well as integration with other smart home devices. Its low cost, versatility, and ease of use make it an attractive choice for developers looking to create innovative pet care solutions.

Load Sensor: The Load Sensor in smart pet feeder is a 1 kg load sensor and will take weight of bowl as input and send output to microcontroller unit and microcontroller check whether the food in bowl is completely eaten or not, after receiving the input ESP32 will conclude the further process. HX711 is used a part of load sensor it will amplify these small signals up to 128 times and it incorporates a high-resolution 24-bit ADC that can convert the analog signal into a digital representation with high precision.Load sensors work on the principle of strain gauge technology, where a strain gauge is bonded to a flexible material (usually metal) that deforms under load. As the material deforms, the strain gauge undergoes a change in resistance, which is directly proportional to the applied force. This change in resistance is then converted into an electrical signal that can be measured and interpreted by a microcontroller or other electronic devices. In the context of a smart pet feeder, load sensors are typically used to measure the weight of the pet food remaining in the feeder's container. By continuously monitoring the weight, the feeder can determine when to dispense more food and when to alert the owner that the container needs refilling. The load sensor is typically placed underneath the food container or integrated into the feeder's dispensing mechanism. When food is added or removed from the container, the change in weight is detected by the load sensor. This information is then used to update the feeder's internal state and trigger actions such as dispensing food or sending notifications to the owner. Load sensors come in various shapes, sizes, and weight capacities to suit different applications. They can measure weights ranging from a few grams to several tons, making them suitable for a wide range of pet feeder sizes and types. In summary, load sensors play a crucial role in smart pet feeders by accurately measuring the amount of food remaining in the feeder's container. This allows for precise feeding schedules, prevents overfeeding or underfeeding, and provides owners with real-time information about their pet's feeding habits.

Ultra Sonic Sensor: An ultrasonic sensor operates on the principle of emitting high-frequency sound waves and measuring the time it takes for the waves to bounce back. This allows the sensor to calculate the distance between bowl and pet in its path. It takes distance between the bowl and pet as input and send it to the micro controller, which is ESP32, and it calculate weather to open the bowl or not.An ultrasonic sensor is a device that measures the distance to an object by emitting ultrasonic sound waves and then detecting the time it takes for the waves to bounce back. These sensors are commonly used in various applications, including robotics, industrial automation, and, relevantly, in smart pet feeders. The operation of an ultrasonic sensor is based on the principle of echolocation, like how bats and dolphins navigate. The sensor emits a burst of ultrasonic sound waves, which travel through the air until they encounter an object. When the waves hit the object, they are reflected to the sensor. By measuring the time it takes for the waves to travel to the object and back, the sensor can calculate the distance to the object. In the context of a smart pet feeder, an ultrasonic sensor can be used to detect the presence of the pet near the feeder or to prevent the feeder from dispensing food if the pet is too close. This helps prevent accidents and ensures the safety of the pet. The ultrasonic sensor is typically mounted on the exterior of the pet feeder, facing outward. When the sensor detects an object within its range, it sends a signal to the feeder's control system, which can then take appropriate action, such as pausing the feeding process or emitting a warning signal. One advantage of ultrasonic sensors is their ability to work in various environmental conditions, including darkness and dust, unlike optical sensors which may be affected by ambient light. Additionally, ultrasonic sensors have a longer range compared to infrared sensors, making them suitable for detecting larger objects or pets from a distance. However, ultrasonic sensors may have limitations in highly reflective or acoustically noisy environments, which could affect their accuracy. Additionally, they may not be suitable for detecting certain types of objects, such as soft or absorbent materials, which may absorb the sound waves rather than reflecting them back to the sensor. In summary, ultrasonic sensors offer a reliable and versatile solution for detecting objects and ensuring the safety and functionality of smart pet feeders. By accurately measuring distances and detecting the presence of pets, they contribute to a more efficient and user-friendly feeding experience for both pets and their owners.

Servo Motor: Our Smart Pet Feeder System's servo motor integration transforms feeding by giving you exact control over the feeding bowl's cap's opening and closing. The servo motor provides the necessary force and precision to smoothly rotate the cap of the feeding bowl. Servo motor is different other motor since we have control over it, we can move cap of the bowl at any angle. A servo motor is a rotary actuator that allows for precise control of angular position. It consists of a motor coupled with a sensor for position feedback, typically a potentiometer or an encoder. Servo motors are widely used in various applications, including robotics, automation, and, relevantly, in smart pet feeders. In a smart pet feeder, a servo motor is often used to control the mechanism that dispenses food into the pet's bowl. The servo motor can precisely rotate to a specific angle, allowing for accurate control over the amount of food dispensed. This enables the feeder to deliver precise portion sizes according to the pet's feeding schedule. Servo motors operate based on a closed-loop control system, where the position feedback from the sensor is compared to the desired position set by the user. The motor adjusts its rotation until the actual position matches the desired position, ensuring accurate and reliable operation. One of the key advantages of servo motors is their ability to provide precise control over position, speed, and torque. This makes them ideal for applications where accuracy and repeatability are critical, such as in smart pet feeders where consistent portion sizes are essential for the pet's health. Another advantage of servo motors is their compact size and high torque-to-weight ratio, allowing for efficient use of space in small devices like pet feeders. Additionally, servo motors are available in various sizes and torque ratings, making it easy to select the right motor for the specific requirements of the feeder.

Liquid Crystal Display (LCD): The integration of LCD makes a crucial feature of system, which very help full for interaction between the pet’s owner and system. The LCD display provides a visual representation of the pet's distance from the feeder, making it possible to monitor this crucial interaction. The LCD display provides this crucial information by displaying the weight of the food within the feeder in real time. Using I2C with LCD modules simplifies the hardware design, reduces pin count, and offers flexibility and ease of use.The LCD serves as a crucial component in the user interface of the proposed smart pet feeder system. It provides real-time visual feedback to the user, displaying important information such as feeding schedules, remaining food levels, system status, and alerts or notifications. Integrated with the microcontroller, typically an ESP32, the LCD communicates with the firmware to dynamically update its content according to user interactions and system events. This interaction layer allows pet owners to easily monitor and manage their pet's feeding regimen, ensuring timely and accurate dispensing of food. By presenting information in a clear and intuitive manner, the LCD enhances the user experience, providing convenience and peace of mind to pet owners even when they are not physically present.

Speaker: The Buzzer in smart pet feeder will communicate with the pet, it is for owner to call pet for feeding. The speaker in our Smart Pet Feeder is more than a device, it is a conduit for connection and communication with the pet, ensuring they receive the nourishment they need in a manner that resonates with their natural instincts.

Internet of Things (IoT):This is a network of interconnected devices that communicate and exchange data with each other over the internet, without requiring human intervention. These devices can range from everyday objects like home appliances and wearable devices to industrial machines and smart infrastructure. In the context of smart pet feeders, IoT technology enables the feeder to connect to the internet, allowing for remote monitoring, control, and automation. One of the primary benefits of IoT in smart pet feeders is remote access. Pet owners can use their smartphones or computers to check on their pets and control the feeder from anywhere with an internet connection. This enables them to adjust feeding schedules, dispense food, and receive notifications about their pet's feeding habits, even when they're away from home. Additionally, IoT enables data collection and analysis, providing valuable insights into the pet's behavior and health. Smart pet feeders equipped with sensors can track factors such as feeding frequency, portion sizes, and food consumption patterns. This data can be used to monitor the pet's health and detect any changes in eating habits that may indicate underlying issues.

Inter-Integrated Circuit (I2C):The Inter-Integrated Circuit (I2C) protocol plays a crucial role in the communication infrastructure of the proposed smart pet feeder system. By facilitating communication between the microcontroller, sensors, and actuators, I2C enables seamless data exchange within the system. The microcontroller acts as the master device, orchestrating the communication process and controlling various peripherals such as sensors for monitoring food levels and environmental conditions, as well as actuators for dispensing food or indicating system status. Through the use of only two wires, I2C streamlines the connection of multiple devices, allowing for efficient and reliable communication in a compact space, such as inside the pet feeder. This simplicity and versatility make I2C an ideal choice for integrating diverse components within the smart pet feeder system, ensuring its functionality and responsiveness to user needs.

IV. DATA AND METHODS

4.1 CHOICE OF THE DATA AND THE MARKER SET

For this study, we adopt the **VISAGE Enhanced tool** marker panel, as used in Heinzel et al. (2024) and previous forensic genetics research [14], [10]. This panel contains **104 autosomal ancestry-informative markers (AIMs)**, including **29 multiallelic markers**, which have been widely validated for forensic applications and are recognized as one of the most reliable sets for biogeographical ancestry (BGA) inference. The choice of VISAGE markers provides two main advantages: (i) these SNPs are already well-established in forensic contexts, making them directly relevant for applied casework; and (ii) their global coverage allows both inter-continental and intra-continental ancestry resolution.

The dataset is composed of aggregated genotype data from multiple publicly available sources:

- **1000 Genomes Project** (n ≈ 2504 individuals) [4]
- **HGDP-CEPH panel** (n ≈ 929 individuals) [3], [4]
- **Middle Eastern cohorts** (n ≈ 137 individuals) [9]
- **Simons Genome Diversity Project (SGDP)** (n ≈ 130 individuals) [4]
- **Estonian Biocentre diversity panel** (n ≈ 402 individuals) [2]

Consistent with prior work, we consider two classification scenarios:

1. **Continental Level**: nine classes including Africa (AFR), Middle East (ME), Europe (EUR), East Asia (EAS), South Asia (SAS), South-East Asia (SEA), Oceania (OCE), Central Asia (CAS), North Africa (NAF), and Admixed America (AMR). To ensure sufficient representation, certain population labels are combined (e.g., “African” and “East African” merged to AFR).
2. **Intra-European Level**: ten population groups derived from the European subset, including **CEU, FIN, TSI, SAR, GBR, IBS, RUS, BAS, FRA, and TUR**. These classes were retained separately due to their known genetic distinctiveness (e.g., Basques vs. French; Sardinians vs. Toscani).

While this design ensures comparability with existing benchmarks, our research extends beyond Heinzel et al. (2024) in two ways:

- **Marker optimization experiments:** We will test whether subsets of the VISAGE panel can be optimized for intra-European resolution using feature-selection techniques, thereby producing a more cost-efficient AIM panel for forensic practice.
- **Admixture and robustness focus:** Unlike prior studies that emphasize only discrete population classification, we will simulate admixed individuals and degraded genotypes to evaluate how well different models (e.g., TabPFN, XGBoost, hybrid ML–admixture approaches) perform under realistic forensic conditions.

This dual approach allows us to **benchmark against established VISAGE-based ancestry classification while addressing practical forensic gaps**, namely admixture quantification, intra-European resolution, and robustness to incomplete data.

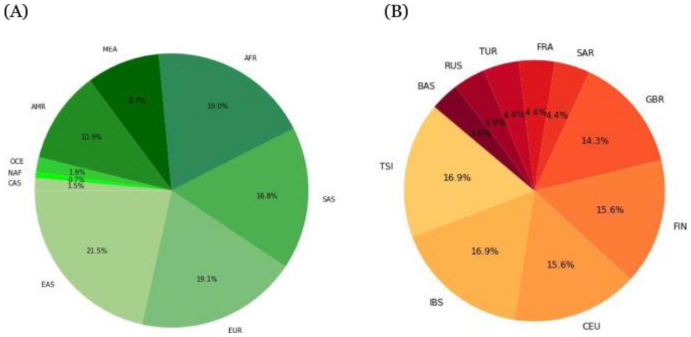

Fig. 1. (A) Overview of the number of individuals per population in case (1), i.e. for continental classification. In this case, there are a total of 4342 individuals. (B) Overview of the number of individuals per population for case (2), i.e. intra-European classification. In this case, there are a total of 635 individuals. Total sample sizes for all populations can be found in Tables S1 and S2 in the Supplement.

4.2 CLASSIFICATION

For classification, we compare the following approaches:
(i) **Snipper**, which is a version of a naïve Bayes classifier commonly used in forensic genetics, assuming marker independence;
(ii) **the Admixture Model (AM)**, applied in its supervised setting, which estimates ancestry proportions by modeling each allele in an individual as originating from one of several source

populations, and assigns individuals to the population with the maximal admixture proportion;
(iii) **PLS-DA (Partial Least Squares Discriminant Analysis)**, previously applied in forensic contexts for ancestry inference, which projects high-dimensional marker data into a lower-dimensional space before classification;
(iv) **TabPFN**, a transformer-based foundational model for tabular data, which has shown superior performance on small tabular datasets and is applied here to genetic marker data;
(v) **XGBoost**, a gradient-boosted decision tree ensemble method well suited for structured data; and
(vi) **Random Forest**, another ensemble-based method that constructs a collection of decision trees and outputs the mode of their predictions.

In addition to these baseline classifiers, we extend the evaluation with methods tailored to forensic ancestry challenges:
(vii) **Feature-optimized TabPFN**, where we apply genetic marker selection strategies (e.g., recursive feature elimination and stability selection) to identify compact ancestry-informative marker (AIM) subsets optimized for intra-continental resolution;
(viii) **Hybrid Classification–Admixture Model**, where TabPFN is modified to output both discrete ancestry probabilities and continuous admixture fractions, enabling simultaneous classification and admixture quantification; and
(ix) **Explainable TabPFN**, where post-hoc interpretability techniques (e.g., SHAP values, integrated gradients) are applied to highlight the relative contribution of each marker to predictions, thus improving forensic admissibility.

We note that model-based methods (i) and (ii) treat markers as independent, while the machine-learning approaches (iii)–(ix) incorporate joint segregation of markers. The extended methods (vii–ix) directly address current research gaps in forensic genetics, particularly in the areas of intra-continental resolution, admixture detection, and explainability.

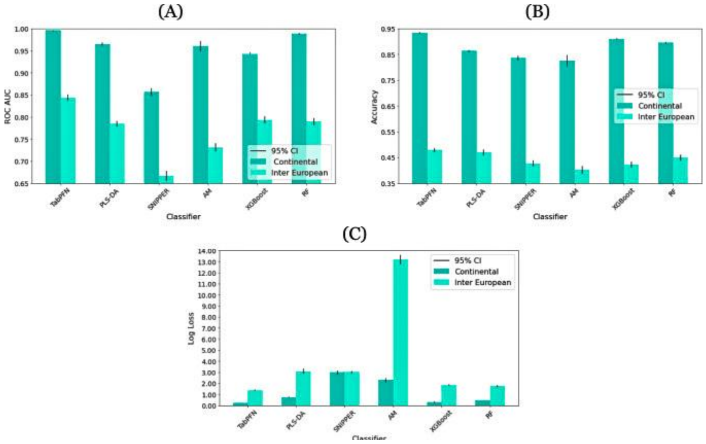

Fig. 2. Evaluation metrics for continental (1) and inter-European (2) classification. The black line represents the 95% confidence interval. (A) ROC AUC; (B) Accuracy; (C) Log Loss. AM stands for Admixture Model and RF stands for Random Forest. For exact numbers, see Tables S3, S4 and S5.

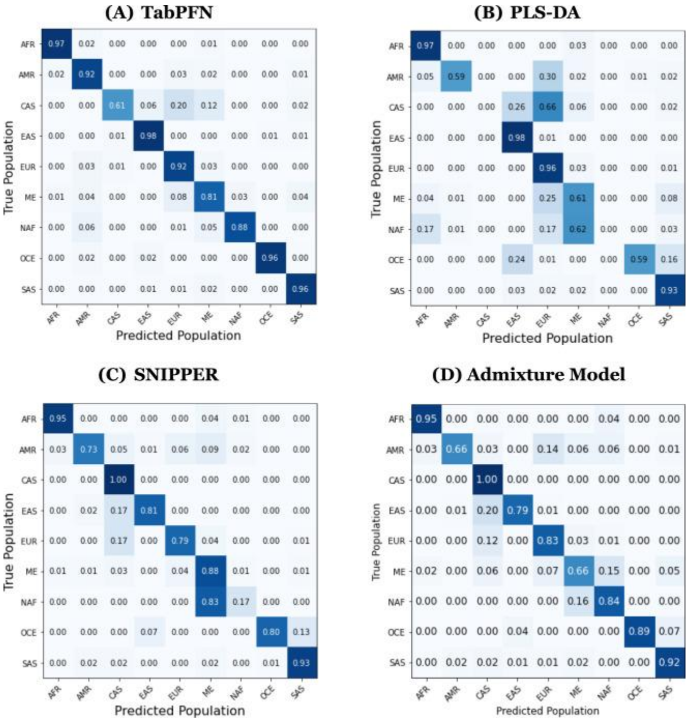

Fig. 3. Confusion matrices for continental classification (case 1). All rows sum to 1 up to a rounding error. **(A)** TabPFN; **(B)** PLS-DA; **(C)** SNIPPER; **(D)** Admixture Model.

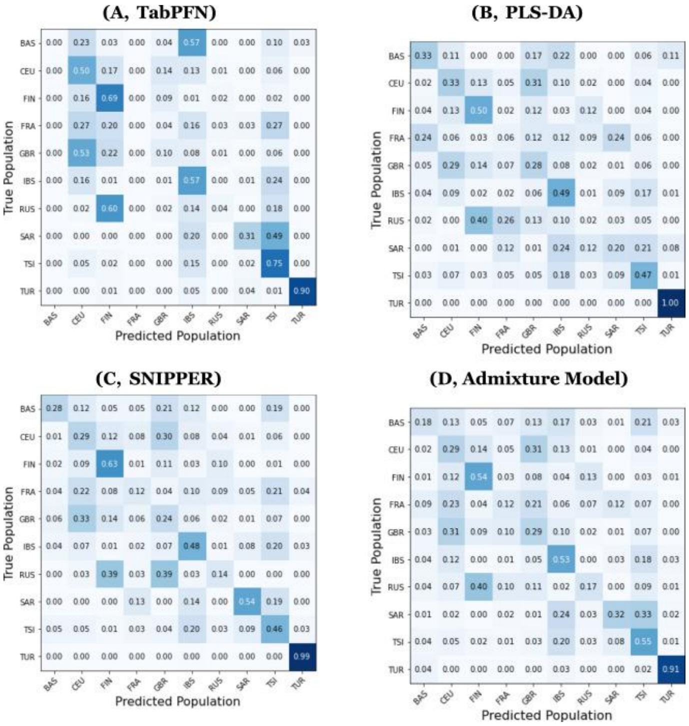

Fig. 4. Confusion matrices for intra-European classification (case 2). All rows sum to 1 up to a rounding error. (A) TabPFN; (B) PLS-DA; (C) Snipper; (D) Admixture Model.

Neural networks have become a cornerstone of modern machine learning. Convolutional neural networks (CNNs) dominate applications in image and video processing, while transformer architectures underpin large language models such as ChatGPT. Extending this paradigm to structured data, TabPFN has been introduced as a foundation model specifically designed for tabular classification tasks. Unlike traditional classifiers, TabPFN leverages a transformer architecture [4] that was pre-trained on millions of synthetic datasets to learn generic strategies for classification [8], [9]. This pre-training enables the model to generalize effectively: once trained, a new dataset can be processed in a single forward pass without further task-specific training. Importantly, TabPFN is openly available, with pre-trained weights that are application-independent and free to use. The model is particularly effective for small to medium-sized datasets, typically up to 10,000 samples and 500 features.

In terms of computational requirements, TabPFN has a higher usage cost than conventional classifiers since efficient inference relies on access to a graphics processing unit (GPU). However, the model remains relatively lightweight compared to large-scale language models, containing millions rather than billions of parameters. As a result, it can be executed on consumer-grade GPUs or integrated accelerators (such as Apple’s M-series chips) with approximately 8 GB of VRAM. This trade-off is often favorable because TabPFN requires minimal training time once deployed, provided a GPU is available [28].

For performance evaluation, we adopt a robust validation strategy. Each classifier is trained on a stratified subset of the dataset and evaluated on a holdout partition. This process is repeated 50 times using 10-repeated 5-fold cross-validation, following the protocol implemented in scikit-learn [10]. Model outputs are assessed using accuracy, log-loss, and ROC AUC, in line with best practices in machine learning [51–54]. ROC AUC serves as a measure of class separability, where a score of 1 indicates perfect discrimination and 0.5 reflects random guessing. In addition, to align with established forensic evaluation methods [16, 19], we analyze classifier behavior through confusion matrices. Specifically, for each of the 50 repetitions, the number of individuals assigned from a true population (e.g., Europe) to a predicted population (e.g., Africa) is recorded. These aggregated outcomes are then summarized as percentile-based confusion matrices, providing a comprehensive view of classification reliability.

V. RESULT

First, we evaluate the performance of the proposed models using key metrics such as Accuracy, ROC AUC, Log-Loss, and calibration error across both continental- and intra-population-level classification tasks. Second, we analyze the corresponding confusion matrices and calibration plots to assess class-specific performance and misclassification patterns. Finally, we extend the comparison by examining model explainability, robustness to missing data, and the ability to capture admixture proportions, providing a more comprehensive assessment of forensic applicability.

5.1 EVALUATION METRICES

The evaluation of classification performance was based on three standard metrics: ROC AUC (ranging from 0.5 to 1, with higher values indicating stronger discrimination), Accuracy (0–1, higher is better), and Log Loss (greater than 0, with lower values indicating better calibration). Results for both the continental and intra-European tasks are summarized in Fig. 2, with detailed values provided in Supplementary Tables S3–S5. Across all three metrics, TabPFN consistently outperformed the other methods, achieving the strongest balance of accuracy, calibration, and discriminatory power. Although PLS-DA did not match TabPFN’s performance, it generally surpassed both Snipper and the Admixture Model. Interestingly, Snipper achieved slightly higher accuracy and lower log-loss than the Admixture Model, though its ROC AUC values were inferior. Overall, all four methods proved effective for continental-scale classification. However, predictive power declined markedly in the intra-European setting, where greater class similarity and the presence of ten populations increased complexity. Even the best-performing TabPFN achieved only 48% accuracy, underscoring the challenges of fine-scale ancestry resolution.

5.2 CONFUSION MATRICES

For a more detailed understanding of which distinctions were successful and which posed challenges, confusion matrices were generated for all eight experimental cases (Cases (1) and (2), across six classifiers). The results are presented in Figures 3 and 4 for TabPFN, Snipper, PLS-DA, and the Admixture Model, while results for XGBoost and Random Forest are provided in Figure S1 of the supplementary material.

In the continental classification task, most populations were classified with high accuracy, with notable exceptions for the Middle Eastern (ME) and North African (NAF) groups across all methods. TabPFN correctly identified ME individuals in 81% of cases, a result slightly surpassed by Snipper (88%). However, Snipper showed significant misclassification of NAF samples, with only 17% correctly assigned, whereas TabPFN achieved a substantially higher accuracy of 88% for this group. PLS-DA and the Admixture Model exhibited similar confusion patterns to Snipper between ME and NAF. Another population of interest was Central

Asian, represented by only 64 samples from the EGDP dataset. Here, TabPFN attained 61% accuracy, while models that treat markers as independent (Snipper and Admixture) achieved perfect classification (100%).

By contrast, intra-European classification (Case (2)) proved considerably more challenging, as reflected in higher error rates across models. For instance, TabPFN frequently misclassified individuals from Great Britain (GBR) as Central Europeans (CEU) in 53% of cases, while CEU individuals were misclassified as GBR in 14% of cases. Misclassification rates were also notably high for Sardinian (SAR) individuals, and Russian (RUS) individuals were often assigned to the Finnish (FIN) group. In certain populations, alternative models even outperformed TabPFN: Snipper achieved higher accuracy for Turkish (TUR) individuals, and the Admixture Model provided superior results for SAR individuals.

VI. CONCLUSION

This study highlights the growing potential of machine learning models, particularly transformer-based architectures such as TabPFN, for advancing biogeographical ancestry (BGA) prediction. Evidence from existing work demonstrates that TabPFN achieves superior accuracy, ROC AUC, and calibration compared to traditional forensic classifiers when applied to continental and intra-European populations. Nevertheless, persistent challenges remain: classification accuracy declines in closely related populations, performance is limited for underrepresented groups, and current methods do not address admixture proportion estimation, explainability, or robustness to degraded forensic samples.

Building on these insights, our research emphasizes that future progress in forensic genetics must move beyond accuracy benchmarks alone. The development of models that integrate admixture quantification, domain-informed marker optimization, and robust handling of incomplete data can directly address forensic realities. Furthermore, explainability and calibration frameworks are essential to ensure that machine learning outputs are legally defensible and ethically sound. By aligning technical improvements with forensic practice requirements, our research aims to bridge the gap between state-of-the-art computational methods and their operational adoption in casework.

In conclusion, while TabPFN and similar models offer a strong foundation for BGA classification, their greatest impact will emerge when coupled with optimized marker panels, transparent decision-making processes, and validated forensic protocols. By pursuing these directions, our work contributes not only to methodological advancement but also to the responsible and practical integration of machine learning in forensic genetics.

REFERENCES

[1]Yufeng Wen, Jing Liu, Yonglin Su, Xiacan Chen, Yiping Hou, Linchuan Liao, Zheng Wang, Forensic biogeographical ancestry inference: recent insights and current trends, doi.org/10.1007/s13258-023-01387-5.

[2] Gin Gin Gan, Alan Teh, Kim Yen Goh, Heng Thay Chong, Kang Wah Pang, Racial background is a determinant factor in the maintenance dosage of warfarin, DOI: 10.1007/BF02983247.

[3] Swapan Mallick, Heng Li, Mark Lipson, Iain Mathieson, Melissa Gymrek, Fernando Racimo, Mengyao Zhao, Niru Chennagiri, Susanne Nordenfelt, Arti Tandon, et al.,The simons genome diversity project: 300 genomes from 142 diverse populations, DOI: 10.1038/nature18964.

[4] Carola Sophia Heinzel, Franz Baumdicker, Peter Pfaffelhuber, Revealing the range of maximum likelihood estimates in the admixture model, doi.org/10.1101/2024.10.18.619150.

[5] Peter Pfaffelhuber, Franziska Grundner-Culemann, Veronika Lipphardt, Franz Baumdicker, How to choose sets of ancestry informative markers: A supervised feature selection approach, DOI: 10.1016/j.fsigen.2020.102259.

[6] Noah A. Rosenberg, Lei M. Li, Ryk Ward, Jonathan K. Pritchard, Informativeness of genetic markers for inference of ancestry, DOI: 10.1086/380416.

[7] Mengyuan Song, Chenxi Zhao, Zheng Wang, Yiping Hou, Applying machine learning algorithms to a real forensic case to predict Y-SNP haplogroup based on Y-STR haplotype, DOI: 10.1016/j.fsigss.2019.10.120.

[8] Dejan Šorgić, Aleksandra Stefanović, Dušan Keckarević, Mladen Popović, XGBoost as a reliable machine learning tool for predicting ancestry using autosomal STR profiles-Proof of method, DOI: 10.1016/j.fsigen.2024.103183.

[9] Christopher Phillips, W. Parson, B. Lundsberg, C. Santos, A. Freire-Aradas, M. Torres, M. Eduardoff, C. Børsting, P. Johansen, M. Fondevila, et al., Building a forensic ancestry panel from the ground up: The EUROFORGEN global AIM-SNP set, DOI: 10.1016/j.fsigen.2014.02.012.

[10] David H. Alexander, John Novembre, Kenneth Lange, Fast model-based estimation of ancestry in unrelated individuals, DOI: 10.1101/gr.094052.109.

[11] P. N. Vrishanka, P. Prabhakar, D. Shet and K. Rupali, Automated Pet Feeder using IoT, In 2021 IEEE International Conference on Mobile Networks and Wireless Communications (ICMNWC), Tumkur, Karnataka, India, 2021, pp. 1-5, doi: 10.1109/ICMNWC52512.2021.9688391.

[12] Dennis McNevin, C. Santos, A. Gómez-Tato, J. Álvarez-Dios, M. Casares de Cal, R. Daniel, C. Phillips, M.V. Lareu, An assessment of Bayesian and multinomial logistic regression classification systems to analyse admixed individuals, DOI: 10.1016/j.fsigss.2013.10.032.

[13] Chris Phillips, Antonio Salas, J.J. Sanchez, Manuel Fondevila, Antonio Gómez-Tato, J. Alvarez-Dios, Manuel Calaza, M. Casares de Cal, David Ballard, María Victoria Lareu, et al., Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs, DOI: 10.1016/j.fsigen.2007.06.008.

[14] Jonathan K. Pritchard, Matthew Stephens, Peter Donnelly, Inference of population structure using multilocus genotype data, DOI: 10.1093/genetics/155.2.945.

[15] David I. Min, Miyoung Lee, Yi-Min Ku, Michael Flanigan, Gender-dependent racial difference in disposition of cyclosporine among healthy African American and white volunteers, DOI: 10.1067/mcp.2000.111255.

[16] Eugenio Alladio, Brando Poggiali, Giulia Cosenza, Elena Pilli, Multivariate statistical approach and machine learning for the evaluation of biogeographical ancestry inference in the forensic field, DOI: 10.1038/s41598-022-12903-0.

[17] David H. Alexander, John Novembre, Kenneth Lange, Fast model-based estimation of ancestry in unrelated individuals, DOI: 10.1101/gr.094052.109.

[18] Tony Frudakis, K. Venkateswarlu, M. Thomas, Zach Gaskin, Siva Ginjupalli, Sitarama Gunturi, Viswanathan Ponnuswamy, Sivamiani Natarajan, Ponnuswamy Kolathupalayam Nachimuthu, A classifier for the SNP-based inference of ancestry, DOI: 10.1520/jfs2002079.

[19] Dennis McNevin, C. Santos, A. Gómez-Tato, J. Álvarez-Dios, M. Casares de Cal, R. Daniel, C. Phillips, M.V. Lareu, An assessment of Bayesian and multinomial logistic regression classification systems to analyse admixed individuals, DOI: 10.1016/j.fsigss.2013.10.032.