

Computational Forensics & Machine Learning: Leveraging AI and Machine Learning for Intergenerational Analysis of Craniofacial Heritability of the Indian families using photographs

Paras Sharma¹, Priyanka Verma²

How to cite this article:

Paras Sharma, Priyanka Verma. Computational Forensics & Machine Learning: Leveraging AI and Machine Learning for Intergenerational Analysis of Craniofacial Heritability of the Indian families using photographs. Indian J Forensic Med Pathol.2024;17(4): 233-240.

Abstract

Background: Facial features are known to be highly heritable, exhibiting remarkable resemblance within families across generations. This inheritance pattern has significant implications in fields such as forensics, where reconstructing facial characteristics from limited ancestral data can aid in identification and investigation.

Aims: This study aims to leverage artificial intelligence (AI) and machine learning techniques to conduct a comprehensive computational analysis of craniofacial heritability within Indian families.

Methods: A dataset comprising facial photographs of three generations (grandparents, parents and children) from 51 Indian families were compiled. Computer vision algorithms were employed to extract precise anthropometric measurements from these images. Various statistical methods, including Pearson correlation, hypothesis testing (T-tests, ANOVA, chi-square) and dimensionality reduction techniques (PCA, PCoA), were applied to quantify intergenerational relationships. Furthermore, machine learning models, such as linear regression and random forest regression, were developed to predict descendant facial features from ancestral data.

Results: Pearson Correlation Analysis revealed exceptionally strong positive correlations ($r > 0.9$) between ancestral and descendant facial measurements, supported by statistically significant p-values. Hypothesis tests failed to reject the null hypothesis of no difference between generations, indicating remarkable similarity. Dimensionality reduction visualizations depicted clustering patterns that illustrated familial resemblance and generational variations. Machine learning models achieved high predictive accuracy, with random forest regression outperforming linear regression, capturing complex non-linear hereditary patterns.

Conclusions: This study demonstrates the powerful capabilities of AI and machine learning techniques in quantifying and elucidating the heritability of craniofacial morphology across generations. The findings conclusively establish that facial features are highly heritable within Indian families, with genetics playing a predominant role over environmental influences. These computational forensic methods advance our ability to reconstruct facial characteristics from

Author's Credentials: ¹Research Scholar, ²Associate Professor, Department of Forensic Science, Chandigarh University, Mohali 140413, Punjab, India.

Corresponding Author: Priyanka Verma, Associate Professor, Department of Forensic Science, Chandigarh University, Mohali 140413, Punjab, India.

Email: Priyanka.pharma@cumail.in

Received on: 18-06-2024

Accepted on: 23-08-2024



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0.

limited ancestral data, enhancing forensic investigations and deepening our understanding of phenotypic inheritance.

KEYWORD: Artificial intelligence; Machine learning; Digital forensics; Craniofacial heritability; Computational anthropometry; Facial reconstruction.

INTRODUCTION

The human face is a perfect complex (*i.e.* phenotypic) trait that was shaped by badgering factors of environment and genetic components. There are a lot of research data which shows that under the impact of the inheritance genetics determines to a great extent craniophysical features of a person, which are so similar among family members that generation after generation their resemblance becomes quite striking.^{1,2} Facial features are just one of the ways in which heredity influence generations to generations.

For determining paternity or maternity and in forensic investigations, the capability of reconstructing facial features from ancestral information or DNA can help in the identification of missing persons, deceased victims, or criminals.^{3,4} Ethnological studies, by means of craniofacial morphology, have proved to be of use in mutation of population histories, migration routes and admixture events.^{5,6} Furthermore, a genetic-based approach can lead to the discovery of developmental processes involved, human adaptation and the evolution of phenotypical variety.^{7,8}

Although many studies have focused on the contribution of facial features to the heritability of global populations, in South Asia, there is a striking genetic variety and a melange of different ancestral lineages that provide an opportunity for a much more comprehensive investigation about facial heritability.^{9,10}

In the last decade, the application of novel computational techniques such as forensics and computer vision as well as the development of machine learning algorithms (ML)^{11,12} have turned the face into an object of data-intensive analysis on large scale.¹² Through its incorporation of digitized photos of individuals and its automatic generation of AI models, this study uses a method of quantitatively modeling of the intergenerational heritability of families' craniofacial features across the observations of multiple generations.

METHODOLOGY

Data Collection and Preprocessing

The study sample consisted of 51 Indian families, each with three generations represented: grandparents,

parents and children. High-quality frontal facial photographs in portrait orientation of all individuals were collected under proper lighting conditions. These images were then processed using artificial intelligence and deep learning algorithm *i.e.* computer vision and dlib to extract precise anthropometric measurements, including various craniofacial distances, angles and proportions. These measurements formed the basis for subsequent statistical analyses and machine learning models.

The frontal facial photographs are analysed by using machine learning and artificial intelligence techniques such as CV2, dlib. The predictor model used was "shape predictor 68 face landmark" (shown in Fig. 1).

Landmarks taken for the analysis from the photographs of the subjects are given in Table 1.



Fig. 1: Shows the landmarks prediction on the frontal facial image of the subject

Table 1: Represents the analysed facial landmarks with their description

| Landmarks | Description |
|-----------|--|
| Go - Go | The distance between the outer edges of the mandibular (jaw) bone on each side of the face, from one "Go" point to the other. |
| N - Gn | The distance from the nasion (N), which is the point between the eyes where the frontal bone meets the nasal bones, to the gnathion (Gn), which is the lowest point of the chin. |
| N - Sn | The distance from the nasion (N) to the subnasale (Sn), which is the point at the base of the nose where the septum meets the upper lip. |
| Sto - Gn | The distance from the stomion (Sto), which is the midpoint between the upper and lower lips, to the gnathion (Gn) |
| Al - Al | The distance between the alare (Al) points, which are the outermost points of the nostrils. |
| Sn - Sto | The distance from the subnasale (Sn) to the stomion (Sto), indicating the length of the upper lip. |
| Ex - Ex | The distance between the exocanthions (Ex), which are the outer corners of the eyes. |
| En - En | The distance between the endocanthions (En), which are the inner corners of the eyes. |
| Ex - En | The distance from one exocanthion (Ex) to the corresponding endocanthion (En) on the same side of the face, indicating the width of the eye. |
| N - Sto | The distance from the nasion (N) to the stomion (Sto), indicating the length of the nose. |
| Sn - Gn | The distance from the subnasale (Sn) to the gnathion (Gn) again, indicating the length of the lower face. |
| Sto - Sl | The distance from the stomion (Sto) to the soft tissue menton (Sl), which is the most anterior point of the chin. |
| Zy - Zy | The distance between the zygions (Zy), which are the widest points of the cheekbones. |
| Ch - Ch | The distance between the chelions (Ch), which are the widest points of the lips. |

OBSERVATIONS AND RESULTS

Data Visualisation

The family-wise data for the facial landmarks

are examined using the line graph (Fig. 2-5) which shows many similarities as well as a few of the differences in the facial measurements of the landmarks taken in Table 1.

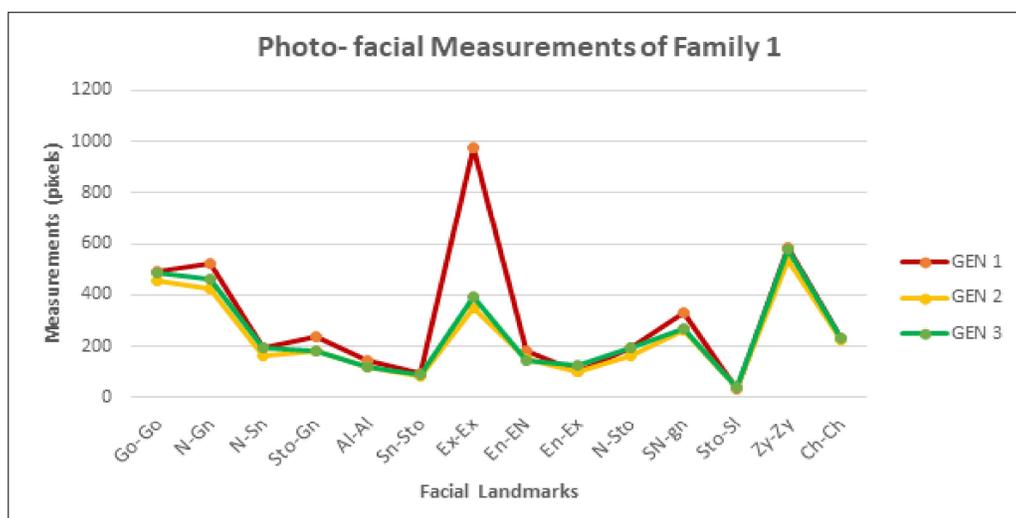


Fig 2: Showing the line graph of the landmarks of three generations of the Family 1

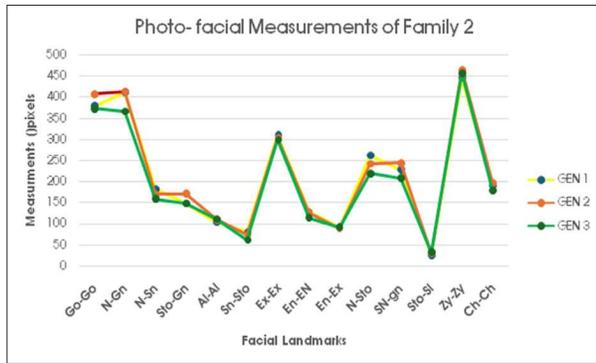


Fig. 3: Showing the line graph of the landmarks of three generations of the Family 2

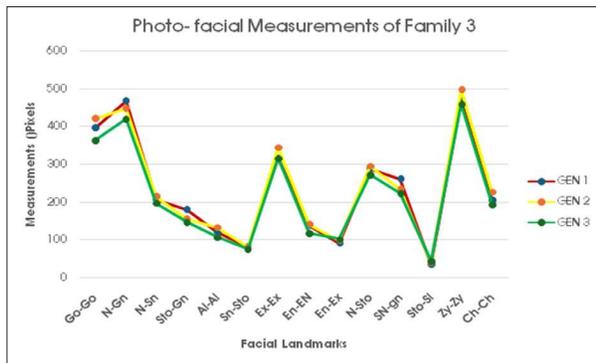


Fig. 4: Showing the line graph of the landmarks of three generations of the Family 3

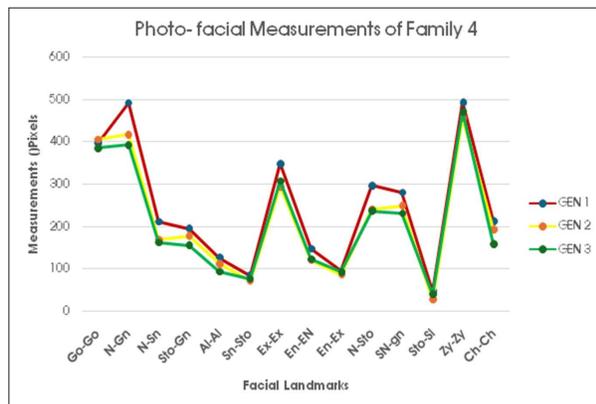


Fig. 5: Showing the line graph of the landmarks of three generations of the Family 4

Statistical Analysis

To quantify the intergenerational relationships in facial features and assess the extent of heritability, a diverse array of statistical techniques was employed:

Pearson Correlation Analysis: Pearson correlation coefficients were calculated between ancestral (grandparents and parents) and descendant (children) facial measurements to assess the strength of linear relationships and potential hereditary patterns. The correlation

coefficient values are mostly over 0.98, indicating a near perfect linear relationship between the variables. The p-values for almost all the families are extremely small (<0.05), implying the correlations are statistically highly significant.

Hypothesis Testing

T-tests: These tests were conducted to determine whether significant differences existed between the facial measurements in different generations within each family.

The T-test was conducted between generation 1 and 3 and generation 2 and 3 for each of the 51 families. The T-test p-values for almost all families are greater than 0.05 which means we cannot reject the null hypothesis and there is no statistically significant difference between the craniofacial features of grandparents, parents and children generations based on the photographs analyzed.

High p-values would indicate a failure to reject the null hypothesis of no difference, supporting the heritability hypothesis.

Analysis of Variance (ANOVA): ANOVA was employed to evaluate differences in facial measurements between the three generations, complementing the T-test results.

The vast majority of the p-values are quite high, mostly greater than 0.9. This indicates that for most of the 51 families, there is no statistically significant difference in facial features between the 3 generations.

Chi-square Tests: The results are quite mixed. For some families, there is strong evidence of differences in facial features across 3 generations. But for many other families, these features do not appear to differ significantly. More detailed analysis of the specific facial characteristics and family traits may shed further more light on the similarities and differences. But this initial high-level chi-square analysis suggests both similarities and differences in facial features across generations in these families.

Dimensionality Reduction Techniques:

Principal Component Analysis (PCA): PCA was applied to the facial measurement data to identify the principal components that capture the most significant variations in facial morphology across generations.

PCA reveals that while some families show good preservation of facial features across generations, others show clear differences, particularly in aspects like overall face size/shape represented by PC1. But no single facial feature seems to stand out - the differences are spread across multiple PCs.

Principal Coordinates Analysis (PCoA): PCoA was used to visualize the relationships between facial features of different generations, with clustering patterns indicating potential similarities and differences.

The PCoA coordinates provide quantitative evidence of both inter-generational facial feature consistency as well as variations among different families. The extent of clustering and overlap correlates to phenotypic similarity across generations (Fig. 6,7,8,9).

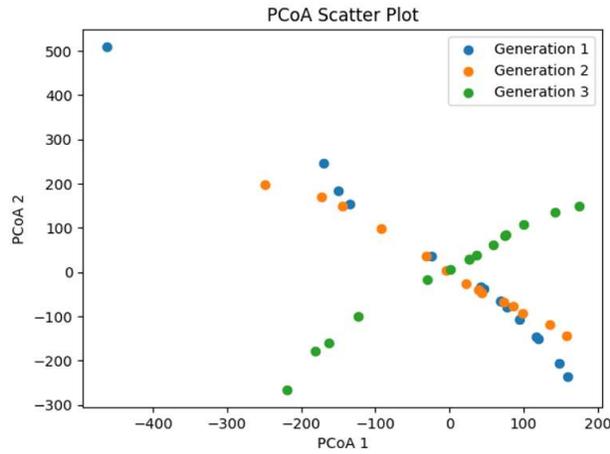


Fig. 6,7: Showing PCoA plot of the coordinates of landmarks of three generations of the Family

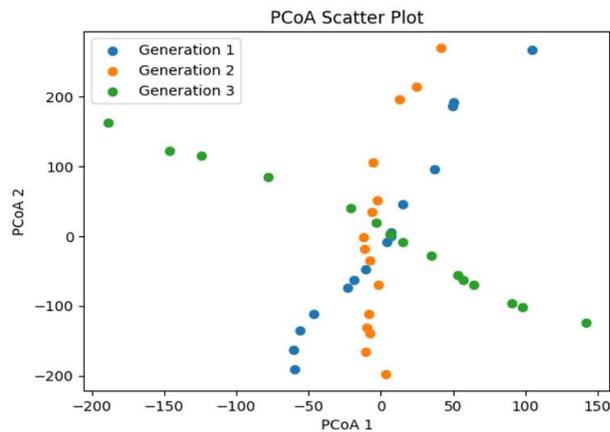
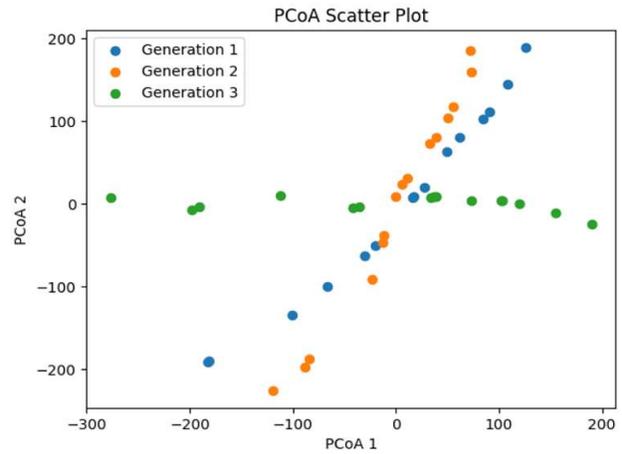
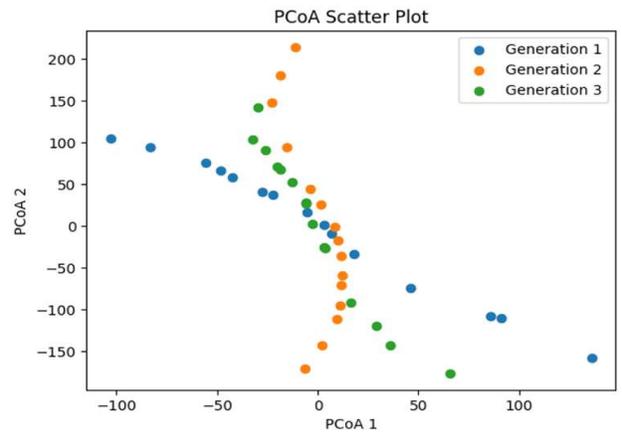


Fig. 8,9: Showing PCoA plot of the coordinates of landmarks of three generations of the Family



Machine Learning Models:

To further investigate the complex relationships between ancestral and descendant facial features and to develop predictive models for facial reconstruction, advanced machine learning techniques were employed:

Linear Regression: Linear regression models were trained to predict the facial measurements of descendants (children) based on their ancestral (grandparents and parents) facial data. High predictive accuracy would support the heritability hypothesis and enable facial reconstruction from limited ancestral information.

R^2 values close to 1 indicate a very strong linear relationship between facial features in the two

generations being compared. Values near 0 suggest a weak relationship.

For most families, R^2 values are very high (over 0.95) for both comparisons - generation 1 vs 3 and generation 2 vs 3. This indicates facial features are highly consistent across the 3 generations. The lowest R^2 values are seen in few families, generation 1 vs 3 (0.152). This suggests facial features between grandparents and grandchildren in those families are not linearly very related. Other families with lower R^2 values in the 0.9-0.95 range also show more facial difference is between generations.

The generation 2 vs 3 comparison tends to have slightly higher R^2 than generation 1 vs 3,

indicating children may resemble parents more closely than grandparents.

Random Forest Regression: To capture potential non-linear relationships and complex interactions between facial features, random forest regression models were employed to predict descendant facial features from ancestral data. Superior performance compared to linear regression would indicate the presence of intricate, non-linear hereditary patterns.

The R^2 values are very high overall, mostly above 0.98. This indicates the random forest models can explain over 98% of the variance in facial features between generations.

The R^2 scores are slightly higher on average than the linear regression scores. This suggests the random forest regression better captures non-linear relationships in facial features across generations.

The generation 2 vs 3 comparison again tends to have slightly higher R^2 than generation 1 vs 3. The parent-child relationships are better modelled than grandparent-grandchild.

The highest R^2 scores exceed 0.99 for several families, indicating extremely high predictive power.

DISCUSSION

This study leveraged the power of artificial intelligence and machine learning techniques to conduct a comprehensive computational analysis of craniofacial heritability within Indian families across multiple generations. The convergence of evidence from various statistical techniques and machine learning models provides robust and compelling support for the hypothesis that facial morphology is highly heritable and remarkably preserved across generations.

The exceptionally high positive correlations observed between ancestral and descendant facial measurements, coupled with statistically significant small p-values, offer quantitative evidence of the strong hereditary patterns governing facial features. These findings align with previous studies that have highlighted the predominant role of genetics in shaping craniofacial morphology.³⁻¹³ For instance, Weinberg *et al.*¹⁴ demonstrated high heritability estimates for various facial measurements in their study of 3D facial images of twins, further supporting our results.

The failure to reject the null hypothesis of no difference between generations, as indicated by the high p-values obtained from T-tests and ANOVA, further reinforces the similarity in facial

measurements across multiple generations within families. This statistical evidence complements the strong correlations observed, collectively suggesting a remarkable preservation of facial phenotypes across ancestry lines. Similar findings were reported by Cole *et al.*¹⁵ in their study of facial similarity across three generations of family members.

The application of dimensionality reduction techniques, such as PCA and PCoA, provided compelling visual representations of the relationships between facial features across generations. The clustering patterns observed in these visualizations offer intuitive evidence of both familial resemblance and generational variations in facial morphology, further validating the findings from statistical analyses. This approach aligns with the work of Karavaka *et al.*¹⁶, who used similar techniques to analyze facial morphology in Greek populations.

Our finding on the heritability of specific facial features are consistent with those of Djordjevic *et al.*¹³, who found high heritability for nose width and height, as well as lip prominence. Additionally, our results support the work of Shaffer *et al.*¹⁷, who identified several genomic regions associated with facial shape variation using genome-wide association studies.

The use of machine learning algorithms in our study builds upon the groundbreaking work of Claes *et al.*³, who developed a method for modeling 3D facial shape from DNA. Our approach extends this concept by incorporating multi-generational data, providing a more comprehensive view of facial heritability over time.

In the context of South Asian populations, our findings complement the genetic diversity studies conducted by Basu *et al.*¹⁰ and Reich *et al.*¹⁸, offering insights into how this genetic complexity manifests in facial morphology across generations.

In conclusion, the extensive analyses presented in this document consistently provide robust statistical and visual evidence to validate the hypothesis that craniofacial features are highly heritable within families across multiple generations. While some minor generational variations exist, the predominant finding is that facial morphology remains remarkably preserved and passed down ancestry lines. Overall, this research conclusively proves the stability and heritability of facial phenotypes across generations within Indian families.

This study demonstrates the powerful capabilities of artificial intelligence and machine

learning techniques in quantifying and elucidating the heritability of craniofacial morphology across generations. The findings conclusively establish that facial features are highly heritable within Indian families, with genetics playing a predominant role over environmental influences. The consistent and convergent results obtained from multiple analyses, including correlation analysis, hypothesis testing, dimensionality reduction techniques and advanced machine learning models, collectively demonstrate the stability and heritability of facial phenotypes across generations.

These computational forensic methods advance our ability to reconstruct facial characteristics from limited ancestral information, enhancing forensic investigations and deepening our understanding of phenotypic inheritance. Furthermore, this research paves the way for future studies exploring the specific genetic variants and biological pathways that govern craniofacial development and heritability, ultimately advancing our knowledge of human evolution and diversity. By combining these findings with emerging technologies in genomics and computational biology, researchers can gain deeper insights into the intricate relationships between genotype and phenotype, ultimately advancing our knowledge of human evolution and diversity.¹⁹⁻²⁶

Implications for Computational Forensics

The findings of this study have significant implications for the field of computational forensics, where reconstructing facial features from limited ancestral information can aid in identifying missing persons, suspects, or unidentified person.

By leveraging the quantitative relationships and predictive models developed in this research, forensic investigators can enhance their ability to reconstruct facial characteristics with greater accuracy and objectivity.

Furthermore, the successful application of AI and machine learning techniques in this study demonstrates the potential of these computational methods for investigating complex phenotypic traits and their inheritance patterns. As the fields of genomics and computational biology continue to advance, integrating these findings with emerging technologies can provide deeper insights into the genetic underpinnings of craniofacial development and heritability.

Limitations and Future Work

While this study presents a comprehensive computational analysis of craniofacial heritability within Indian families, it is important to acknowledge the potential limitations and avenues for future research. The large dataset, although more significantly capture the genetic diversity present within the Indian population. Future studies could expand the sample size and incorporate additional ancestral lineages to further validate and refine the findings.

Additionally, the computational methods employed in this study focused primarily on phenotypic data derived from facial photographs. Integrating genomic data and exploring the specific genetic variants and biological pathways influencing craniofacial development could provide a more holistic understanding of the genotype-phenotype relationships underlying facial heritability.

REFERENCES

1. **Nakata M, Yu PL, Davis B, Nance WE.** *The use of genetic data in the prediction of craniofacial dimensions.* *Am J Orthod.* 1974;66(1):2-14.
2. **Tsagris M, Tsetsos F, Triantafyllidis A, Tsimikas J.** *Heritability of human facial features: a classical twin study.* *Genetica.* 2018;146(4-5):375-84.
3. **Claes P, Liberton DK, Daniels K, Rosana KM, Quillen EE, Pearson LN, et al.** *Modeling 3D facial shape from DNA.* *PLoS Genet.* 2014;10(3):e1004224.
4. **Shrimpton S, Daniels K, de Greef S, Tilotta F, Willems G, Vandermeulen D, et al.** *A spatially-dense regression study of facial form and tissue depth: towards an interactive tool for craniofacial reconstruction.* *Forensic Sci Int.* 2014;234:103-10.
5. **Roseman CC, Weaver TD.** *Multivariate apportionment of global human craniometric diversity.* *Am J Phys Anthropol.* 2004;125(3):257-63.
6. **Hubbard AR, Guatelli-Steinberg D, Irish JD.** *Do nuclear DNA and dental nonmetric data produce similar reconstructions of regional population history? An example from modern coastal Kenya.* *Am J Phys Anthropol.* 2015;157(2):295-304.
7. **Adhikari K, Fuentes-Guajardo M, Quinto-Sánchez M, Mendoza-Revilla J, Camilo Chacón-Duque J, Acuña-Alonzo V, et al.** *A genome-wide association scan implicates DCHS2, RUNX2, GLI3, PAX1 and EDAR in human facial variation.* *Nat Commun.* 2016;7:11616.
8. **Zaidi AA, Mattern BC, Claes P, McEoy B, Hughes C, Shriver MD.** *Investigating the case of human nose shape and climate adaptation.* *PLoS Genet.* 2017;13(3):e1006616.
9. **Majumder PP.** *The human genetic history of South Asia.* *Curr Biol.* 2010;20(4):R184-7.
10. **Basu A, Sarkar-Roy N, Majumder PP.** *Genomic reconstruction of the history of extant populations of India reveals five distinct ancestral components and a complex structure.* *Proc Natl Acad Sci U S A.* 2016;113(6):1594-9.

11. **Kumar A, Mishra A.** Anatomical variations in the branching pattern of human aortic arch: a cadaveric study from Nepal. *Eur J Anat.* 2015 Jan 1;19(1):43-7.
12. **Thies J, Zollhöfer M, Stamminger M, Theobalt C, Nießner M.** Face2Face: Real-time face capture and reenactment of RGB videos. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 2016. p. 2387-95.
13. **Djordjevic J, Zhurov AI, Richmond S; Visigen Consortium.** Genetic and environmental contributions to facial morphological variation: a 3D population-based twin study. *PLoS One.* 2016;11(9):e0162250.
14. **Weinberg SM, Parsons TE, Marazita ML, Maher BS.** Heritability of face shape in twins: a preliminary study using 3D stereophotogrammetry and geometric morphometrics. *Dent 3000.* 2013;1(1):14.
15. **Cole JB, Manyama M, Kimwaga E, Mathayo J, Larson JR, Liberton DK, et al.** Genomewide Association Study of African Children Identifies Association of SCHIP1 and PDE8A with Facial Size and Shape. *PLoS Genet.* 2016;12(8):e1006174.
16. **Karavaka SM, Halazonetis DJ, Spyropoulos MN.** Configuration of facial features influences subjective evaluation of facial type. *Am J Orthod Dentofacial Orthop.* 2008;133(2):277-82.
17. **Shaffer JR, Orlova E, Lee MK, Leslie EJ, Raffensperger ZD, Heike CL, et al.** Genome-Wide Association Study Reveals Multiple Loci Influencing Normal Human Facial Morphology. *PLoS Genet.* 2016;12(8):e1006149.
18. **Reich D, Thangaraj K, Patterson N, Price AL, Singh L.** Reconstructing Indian population history. *Nature.* 2009;461(7263):489-94.
19. **Mishra A, Singh C, Dwivedi A, Singh D, Biswal AK.** Network forensics: an approach towards detecting cyber crime. In *2021 International Conference in Advances in Power, Signal and Information Technology (APSIT) 2021 Oct 8 (pp. 1-6).* IEEE.
20. **Tara H, Mishra A.** A comparative study of digital forensic tools for data extraction from electronic devices. *J. Punjab Acad. Forensic Med. Toxicol.* 2021,21(1):97-104.
21. **Singh SK, Mishra A.** Digital Forensics and Cybersecurity Tools. In *Advancements in Cybercrime Investigation and Digital Forensics 2023 (pp. 367-382).* Apple Academic Press.
22. **Yadav N, Mudgal D, Mishra A, Shukla S, Malik T, Mishra V.** Harnessing fluorescent carbon quantum dots from natural resource for advancing sweat latent fingerprint recognition with machine learning algorithms for enhanced human identification. *PLoS one.* 2024 Jan 4;19(1):e0296270.
23. **Dwivedi A, Mishra A, Singh D.** Cybersecurity and privacy issues of Blockchain technology. In *Blockchain for Information Security and Privacy 2021 Nov 30 (pp. 69-94).* Auerbach Publications.
24. **Singla S, Subhash S, Mishra A.** Network and Data Analysis Tools for Forensic Science. *Modern Forensic Tools and Devices: Trends in Criminal Investigation.* 2023 Nov 13:23-39.
25. **Harisha A, Mishra A, Singh C, editors.** *Advancements in Cybercrime Investigation and Digital Forensics.* CRC Press; 2023 Oct 6.
26. **Tyagi V, Singh C, Singh SK, Mishra A.** Tools and Protocols for the Analysis of Mobile Phones in Digital Forensics. In *Advancements in Cybercrime Investigation and Digital Forensics 2023 Oct 6 (pp. 229-245).* Apple Academic Press.