Genetic Programming for Age-at-death Estimation from the Pubic Symphysis

Enrique Bermejo, Oscar Cordón
DECSAI, University of Granada
`{ebermejo,ocordon}@decsai.ugr.es`

Antonio David Villegas
Panacea Coop. Research
`antonio.villegas.panacea@gmail.com`

Javier Irurita
Physical Anthropology Lab
University of Granada
`javieri@ugr.es`

Sergio Damas
Software Engineering Dpt.
University of Granada
`sdamas@ugr.es`

Abstract

Skeleton-based age-at-death estimation is an arduous task in human identification based on characteristics such as appearance, morphology or ossification patterns. This process is performed manually, although in recent years there have been several studies that attempt to automate it. This study proposes a semi-automatic method for estimating age-at-death using nine pubic symphysis traits derived from Todd's method. By employing genetic programming and symbolic regression, simple mathematical expressions are generated to estimate age-at-death. To address the imbalance in the data, oversampling methods are implemented. The method achieves state-of-the-art performance while maintaining interpretability, allowing validation of existing knowledge and discovery of new forensics insights.

1 Introduction

Human identification in forensic anthropology (FA) is crucial in scenarios of individual and mass casualties [1]. The accurate estimation of biological profile (BP) (i.e. age, sex, ancestry, and stature) is essential in narrowing down potential matches, with DNA analysis used for final identification if possible. Among other bones, the pubic symphysis is highly reliable for age-at-death estimation [2].

Historically, methods for age estimation have evolved from phase-based to numerical approaches. Phase-based methods, like the Suchey-Brooks extension of Todd's method [3], are simple and widely used but suffer from subjectivity and reduced accuracy. Meanwhile, methods obtaining the estimation from an overall analysis of the morphological characteristics associated to the different pubic symphysis changes (usually performed with a visual inspection of the bone) and methods analyzing each pubic bone trait in isolation and then aggregating the partial observations to take the final decision (i.e., component-scoring and component-based methods). Component-scoring methods, first proposed by Gilbert and McKern [4], offer a more objective approach by evaluating each trait individually before aggregating the observations. The use of either scoring or other component-based methods had already shown a significant reduction in both intra- and inter-observer error. This is because labeling each component separately can be done more objectively than assigning a general age-at-death estimation (phase or number) to the entire pubic symphysis.

The development of automatic, precise, and robust age-at-death estimation methods is currently a significant focus in FA [5]. Modern approaches include advanced computer vision and machine learning (ML) techniques, which, although accurate, often produce complex models that are not easily interpretable. This study [6] focuses on developing transparent and accurate models.

XVI Congreso Español de Metaheurísticas, Algoritmos Evolutivos y Bioinspirados (maeb 2025).

2 Symbolic Regression with Genetic Programming

Explainable artificial intelligence (XAI) [7] is vital for human-centric decision support systems, including FA. XAI emphasizes the need for interpretable models that balance accuracy and transparency, ensuring trustworthiness in critical applications like medicine, law, and security. Symbolic regression, a form of regression analysis, seeks to represent data relationships without prior knowledge of the underlying mathematical expressions, enhancing interpretability compared to black-box models like deep learning. Evolutionary algorithms (EAs) are effective for exploring the complex space of such expressions. Genetic programming (GP) [8], a tree-based EA, represents mathematical expressions as tree structures with variables as terminal nodes and operators as inner nodes. Niching genetic algorithm-programming (GA-P) [9] hybridizes GP with genetic algorithms (GAs) for parameter optimization, enhancing expression estimation. GP and GA-P both suffer from *bloat*, an excessive growth in individual size that increases evaluation costs and overfitting risk. Various bloat control methods, such as limiting tree depth or using semantic approximation (GP-DA) [10], address this issue by optimizing expression subtrees for semantic similarity. This method optimizes the semantic vector of subtrees, enhancing diversity while maintaining performance.

3 Results and analysis

3.1 Experimental design

The data set includes 960 annotated pubic symphysis samples from 600 individuals, aged 17-82, collected by forensic experts. Nine traits, identified from Todd's method, are categorized and annotated. The imbalanced data set, with more older samples, requires oversampling for effective model training. Random replication oversampling re-balances the data set, creating 2780 training and validation samples. The final test set remains original and unseen, ensuring unbiased evaluation. The experimental setup involves 5-fold cross-validation (5-CV) to avoid overfitting and ensure robust model selection. Multiple runs with different seeds assess method robustness. Twenty-five models are learned for each method (GP, GP-DA, GA-P), with the best validation model tested on the unseen data set. Parameters for GP-DA include a maximum depth of 20 nodes. GP and GA-P configurations vary tree depths at 20, 40, and 60 nodes. Other parameters include variable generation probability (0.3), crossover (0.75) and mutation (0.05) probabilities, intra-niche crossover probability (0.03), and population size (1000). The stopping criterion is 1,000,000 evaluations.

3.2 Model validation and selection

Performance metrics include root mean squared error (RMSE) and mean absolute error (MAE). GP-DA shows slight performance advantages with smaller tree depths. Despite similar accuracies, simpler models with fewer parameters generalize better. Seven best-performing models from 5-CV are evaluated on the test set, compared with classical ML methods like linear regression (LR), support-vector machines (SVM), decision trees (DT), and random forests (RF). GP-DA and GA-P achieve the best results, with simpler models offering better generalization. The finally selected model is expressed by the following equation[1], including only 5 different pubic symphysis variables:

$$Age = 6.06 \cdot I_P + U_{SE} + 6.06 \cdot L_{SE} + 5.06 \cdot V_M + \frac{V_B}{I_P}, \tag{1}$$

where I_P represents the irregular porosity, U_{SE} the upper symphysial extremity, L_{SE} the lower symphysial extremity, V_M the ventral margin, and V_B the ventral bevel. We select this model due to its simplicity, which helps to make it more interpretable. The expression is simple as in all the cases but one the variables are simply multiplied by a factor and added together. The only compound trait is a ratio between the ventral bevel and the irregular porosity values, that is directly incorporated into the remaining expression. The most influential features are I_P and L_{SE}, according to the coefficients in the expression. Notably, the former plays a double role, as it also acts as a modifier of the ventral bevel trait in the denominator of the $\frac{V_B}{I_P}$ variable. From the extensive experimentation, we can confirm that key traits for age-at-death estimation include irregular porosity, lower symphysial extremity, and ventral margin, frequently used in GA-P models, while less relevant traits include bony nodule and dorsal plateau, aligned with previous work [11].

---

[1] The interested reader is referred to [6] for a graphical description of the variables involved in this equation.

3.3 Benchmarking and overview

As a final comparison, Table 1 summarizes the relation of the best models found in this contribution for GP, GA-P, and GP-DA with the age estimation methods reviewed earlier. Though the proposed benchmarking is a rough and generic comparison, it will allow us to draw a valuable overview on the accuracy of the methods. First of all, the comparison involves methods of different typology (phase-based and numerical, as well as global and component-scoring) which were tested by using a different validation methodology (leave-one-out cross validation (LOOCV), single 50% training-test partition, use of the samples of one pubic symphysis laterality for training and those of the other laterality for test, and 5-fold CV). Moreover, the size, age range, and distribution according sex or ethnic groups of the samples are key differences. Setting aside such differences, our proposal stands out in the comparison.

Table 1: Comparison between the best proposed methods, and the state-of-the-art results. Acronyms used: **Method type**: CS=component-scoring, PB=phase-based, N=numeric; **Experimental setup**: LOOCV: leave-one-out cross validation, 50%-50% split: single 50% training-test partition, tra: XXX-test:XXX: use of the samples of one pubic symphysis laterality for training and those of the other laterality for test; 5-CV: 5-fold cross validation; AD: Alternative Distribution.

Method	**Type**	**Exp setup**	**# Samples**	**Age range**	**RMSE**	**MAE**
Slice and Algee-Hewitt [12]	N	LOOCV	41	19-96	17.15	
Stoyanova et al. [13]	N	LOOCV	56	16-100	19	
Stoyanova et al. [14]	N	50-50% split	93	16-90	13.7-16.5	
Kotěrová et al. [15]	CS,N	5-CV (w/o test)	941	19-100	12.1	9.7
Kotěrová et al. [16] SAAS	CS,N	5-CV (w/o test)	483	18-92	14.3	11.7
Kotěrová et al. [16] AANNESS	N	5-CV (w/o test)	483	18-92	12.9	10.6
Gámez-Granados et al. [11]	PB	tra: right lat.-test: left lat.	892 (439-453)	18-60	13.19	10.38
Gámez-Granados et al. [11]	PB	tra: right lat.-test: left lat.	960 (487-473)	18-82	14.61	11.62
GP (Depth 20)	CS,N	5-CV tra-val-test	960 (614-154-192)	18-82	10.82	8.56
GA-P (Depth 20)	CS,N	5-CV tra-val-test	960 (614-154-192)	18-82	**10.81**	**8.55**
GP-DA (Depth 20)	CS,N	5-CV tra-val-test	960 (614-154-192)	18-82	10.84	8.55
GA-P (Depth 20) AD	CS,N	5-CV tra-val-test	668 (381-95-192)	18-82	**9.54**	**7.51**

3.4 Designing a new model

For a more thorough exploration of the tentative expression space, we will follow a different approach by combining age-targeted undersampling and oversampling strategies. This results in a new uniform distribution in which each age value is represented by 21 samples. Specifically, the undersampling step reduces the dataset to 476 samples in the middle-age range, while the oversampling step yields a total of 1113 samples for training purposes to avoid individuals over 64 years being considered outliers. We follow the same experimental setup as in our previous experiment (5-CV) considering the new data distribution. The behavior of the algorithms is similar and GA-P (Depth 20) also achieves the best performance among the GP methods. The last row of Table 1 (GA-P AD, standing for Alternative Distribution) summarizes the results for an overview comparison. In particular, the test results are **9.54** and **7.51** years according RMSE and MAE, respectively. Hence, the refined preprocessing allows the method to improve its performance and to achieve an even lower test error. Meanwhile, the resulting equation can be expressed as follows:

$$Age = V_M \left(I_P + 4.88 B_N + 4.88 D_M - V_M + 3.22 + \frac{I_P - V_M}{V_B} \right). \tag{2}$$

These observations further support the notion that there exist certain traits more appropriate for specific age ranges, in agreement with [11] and Castillo et al. [17].

4 Conclusions

This study [6] presents a robust, interpretable method for age-at-death estimation, combining traditional forensic knowledge with modern ML. Symbolic regression with GP and GA-P offers accurate, transparent models, supporting forensic anthropologists' work. Future work should explore hierarchical models and ensemble expert assessments to further enhance reliability and accuracy.

Acknowledgments and Disclosure of Funding

This work was supported by grant CONFIA (PID2021-122916NB-I00) funded by MCIN/AEI/10.13039/501100011033, funded by “ERDF A way of making Europe”. Additionally, E.B.’s work has been supported by the Regional Government of Andalusia as postdoctoral fellow (DOC_01130). AD.V.’s work has been supported by the Regional Government of Andalusia under the Recovery, Transformation and Resilience Plan (GR/INV/0004/2022).

References

[1] D. H. Ubelaker, *Biological Anthropology of the Human Skeleton*. New Jersey: John Wiley & Sons, Ltd, 2008, ch. Forensic Anthropology: Methodology and Diversity of Applications, pp. 41–69.

[2] B. Dudzik and N. Langley, “Estimating age from the pubic symphysis: A new component-based system,” *Forensic Science International*, vol. 257, pp. 98–105, 2015.

[3] S. Brooks and J. M. Suchey, “Skeletal age determination based on the os pubis: A comparison of the acsádi-nemeskéri and suchey-brooks methods,” *Human Evolution*, vol. 5, no. 3, pp. 227–238, 1990.

[4] B. Gilbert and T. W. McKern, “A method for aging the female os pubis,” *American Journal of Physical Anthropology*, vol. 38, no. 1, pp. 31–38, 1973.

[5] D. H. Ubelaker, Q. R. Cordero, and N. F. Linton, “Recent research in forensic anthropology,” *European Journal of Anatomy*, vol. 24, pp. 221–227, 2020.

[6] E. Bermejo, A. D. Villegas, J. Irurita, S. Damas, and O. Cordón, “Interpretable machine learning for age-at-death estimation from the pubic symphysis,” *Expert Systems*, 2025.

[7] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, “Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible ai,” *Information Fusion*, vol. 58, pp. 82–115, 2020.

[8] J. R. Koza, *Genetic programming: on the programming of computers by means of natural selection*. Cambridge: MIT Press, 1992.

[9] L. M. Howard and D. J. D’Dangelo, “The GA-P: A Genetic Algorithm and Genetic Programming Hybrid,” *IEEE Expert-Intelligent Systems and their Applications*, vol. 10, no. 3, pp. 11–15, 1995.

[10] Q. U. Nguyen and T. H. Chu, “Semantic approximation for reducing code bloat in genetic programming,” *Swarm and Evolutionary Computation*, vol. 58, p. 100729, 2020.

[11] J. C. Gámez, J. Irurita, A. González, S. Damas, I. Alemán, and O. Cordón, “Automating the decision making process of Todds age estimation method from the pubic symphysis with explainable machine learning,” *Information Sciences*, vol. 612, pp. 514–535, 2022.

[12] D. Slice and B. Algee-Hewitt, “Modeling bone surface morphology: A fully quantitative method for age-at-death estimation using the pubic symphysis,” *Journal of Forensic Sciences*, vol. 60, no. 4, pp. 835–843, 2015.

[13] D. Stoyanova, B. Algee-Hewitt, and D. Slice, “An enhanced computational method for age-at-death estimation based on the pubic symphysis using 3D laser scans and thin plate splines,” *American Journal of Physical Anthropology*, vol. 158, no. 3, pp. 431–440, 2015.

[14] D. K. Stoyanova, B. F. B. Algee-Hewitt, J. Kim, and D. E. Slice, “A computational framework for age-at-death estimation from the skeleton: Surface and outline analysis of 3D laser scans of the adult pubic symphysis,” *Journal of Forensic Sciences*, vol. 62, no. 6, pp. 1434–1444, 2017.

[15] A. Kotrová, D. Navega, M. tepanovský, Z. Buk, J. Brek, and E. Cunha, “Age estimation of adult human remains from hip bones using advanced methods,” *Forensic Science International*, vol. 287, pp. 163–175, 2018.

[16] A. Kotěrová, M. Štepanovský, Z. Buk, J. Brůžek, N. Techataweewan, and J. Velemínská, “The computational ageatdeath estimation from 3d surface models of the adult pubic symphysis using data mining methods,” *Scientific Reports*, vol. 12, no. 1, p. 10324, Jun 2022.

[17] A. Castillo, I. Galtés, S. Crespo, and X. Jordana, “Technical note: preliminary insight into a new method for age-at-death estimation from the pubic symphysis,” *International Journal of Legal Medicine*, vol. 135, no. 3, pp. 929–937, 2021.