

Facial landmark detection accuracy of MediaPipe FaceMesh on Southeast Asian pediatric faces: Age-stratified NME analysis, ethnic-diversity implications, and protocol recommendations for clinical morphometry

Lam Thanh Hien¹, Do Nang Toan², Huynh Huy Cuong³, Tran Van Thanh^{1*}

¹Lac Hong University, 10 Huynh Van Nghe, Tran Bien, Dong Nai, Vietnam;

²Institute of Information Technology, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet, Nghia Do, Hanoi, Vietnam;

³Institute for Tropical Technology, Academy of Military Science and Technology, 57A Truong Quoc Dung, Phu Nhuan, Ho Chi Minh City, Vietnam.

*Corresponding author: thanhtran@lhu.edu.vn

Received 17 Apr. 2026; Revised 2 Jun. 2026; Accepted 19 Aug. 2026; Published 15 Oct. 2026.

DOI: <https://doi.org/10.54939/1859-1043.j.mst.114.2026.167-174>

ABSTRACT

MediaPipe FaceMesh, the most widely deployed open-source facial landmark detector, was trained on adult faces, leaving its accuracy on pediatric Southeast Asian populations uncharacterized. This paper presents the first age-stratified Normalized Mean Error (NME) benchmark on Vietnamese children aged 2–8 years ($n = 100$; 14 landmarks; $ICC \geq 0.90$). NME decreased significantly with age — 7.3%, 4.8%, and 3.1% IOD at 2–3, 3–5, and 5–8 years, respectively (ANOVA: $F = 18.4$, $p < 0.001$) — versus the adult benchmark of 2–4% IOD. Iris centers (lm468/473) achieved 1.0–1.5% IOD at all ages and are recommended as the preferred IOD normalizer. Philtrum-base (lm0) NME of 12.3% at 2–3 years renders philtrum height unreliable in toddlers; substituting outer lip corners (lm78/308) reduces philtrum width uncertainty by 40%. A first-order error-propagation analysis translates these per-landmark errors into biomarker-level reliability tiers benchmarked against published ASD effect sizes, providing a reusable framework for assessing morphometric biomarker reliability. Vietnam's multi-ancestry ethnic diversity introduces a further uncharacterized source of systematic error: admixture evidence predicts craniofacial geometries in Khmer, Cham, and highland Austroasiatic children lying partially outside FaceMesh's training distribution. Evidence-based landmark selection protocol recommendations are provided for pediatric Southeast Asian morphometric studies.

Keywords: Facial landmark detection; MediaPipe FaceMesh; Normalized mean error; Pediatric morphometry; Southeast Asian; Age-stratified benchmark; Ethnic diversity; Autism spectrum disorder; Clinical morphometry.

1. INTRODUCTION

Landmark-based facial morphometry is widely used to characterize facial phenotypes in pediatric neurodevelopmental conditions, including autism spectrum disorder (ASD) [1–3], fetal alcohol spectrum disorders, Down syndrome, and Fragile X syndrome [4]. Because landmark errors propagate directly into morphometric measurements, accurate localization is foundational. MediaPipe FaceMesh [7], the de facto standard open-source detector, provides 478 landmarks with sub-pixel claimed accuracy, but was developed and benchmarked exclusively on adult Western and East Asian faces.

Existing benchmarks use adult subjects almost exclusively: 300-W [8] reports 3–5% IOD and Kartynnik et al. [7] report 2–4% IOD for FaceMesh, with no extension to pediatric Southeast Asian populations. Two factors likely degrade accuracy in this cohort: incompletely differentiated toddler perioral anatomy, which renders philtrum-base (lm0) and Cupid's bow (lm61/291) anatomically ambiguous before age three; and the epicanthal fold common in Southeast Asian children, which partially occludes inner canthal landmarks (lm133/362) and biases palpebral fissure length measurement.

A third, underappreciated factor is within-country ethnic diversity: Vietnam's 54 officially recognized ethnic groups, spanning five linguistic families, show distinct craniofacial profiles from multi-layer admixture histories [17-21], so landmark accuracy from a single-city sample may not transfer across groups. Systematic differences in philtrum anatomy, nasal geometry, or epicanthal fold intensity can introduce differential measurement bias that undermines inter-group comparisons [14-16].

This paper addresses both gaps with five contributions: (1) the first age-stratified NME benchmark of FaceMesh on Southeast Asian children aged 2–8 years; (2) a per-landmark stability analysis identifying reliable landmarks by age; (3) a first-order error-propagation framework mapping landmark NME to biomarker uncertainty, stratified into three reliability tiers against published ASD effect sizes — a transferable contribution beyond this benchmark; (4) a hypothesis-generating analysis of ethnicity-dependent failure modes for five Vietnamese ethnic groups, requiring future empirical validation; and (5) evidence-based protocol recommendations for pediatric Southeast Asian morphometric studies.

2. RELATED WORK

2.1. Landmark detection and benchmarking

The 300-W benchmark [8] established IOD-normalized NME as the standard metric; FaceMesh [7] and Attention Mesh [10] report 2 - 4% IOD on adult faces, and Wu and Ji [9] survey the field's evolution from cascaded regression to deep architectures. All these benchmarks are restricted to adult, predominantly Caucasian or East Asian subjects, leaving Southeast Asian toddlers entirely uncharacterized.

2.2. Pediatric morphometry and cross-ethnic landmark bias

Face recognition accuracy degrades with decreasing age [25], and at smartphone distances a 5% IOD NME (3–4 px) can exceed reported ASD biomarker differences ($d \sim 0.68\text{--}1.15$) [1-3]. The fairness literature [25-27] shows training-vs-target morphological mismatch drives accuracy disparities beyond class imbalance; Echeverry-Quiceno et al. [14] reported this for admixed individuals in low-density training regions, a mechanism we hypothesize also applies to non-Kinh Vietnamese groups. Vietnamese cephalometric norms already differ from Caucasian and other Asian standards [22, 23], though intra-Vietnamese inter-ethnic variation remains uncharacterized.

2.3. ASD facial morphometry in Vietnamese cohorts

Aldridge et al. [1] documented increased IOD ($d = 1.15$) and reduced philtrum height ($d = 1.14$) in prepubertal ASD boys. Boutrus et al. [2] reported elevated facial asymmetry ($d = 0.68$) and palpebral fissure shortening ($d = 0.82$). Özgen et al. [3] found wider midface ($d = 0.92$) and nasal width ($d = 0.87$). Vietnamese-specific screening studies [5, 6] show models trained on international datasets perform near chance on Vietnamese children, driven mainly by data composition rather than architecture ($\eta^2 = 0.85$) [31, 32]; landmark-detection accuracy's contribution to this gap has not been quantified.

3. METHOD

3.1. Dataset and inclusion criteria

One hundred frontal photographs of Vietnamese children aged 2–8 years were drawn from the pediatric morphometric dataset of Thanh et al. [5] (892 images; Ho Chi Minh City kindergartens; 2022–2024; IRB-approved with written parental consent), stratified as 2–3 yr ($n = 33$), 3–5 yr ($n = 33$), and 5–8 yr ($n = 34$), with balanced sex distribution overall (51 M / 49 F) and within each stratum. All images were captured by parents or guardians on consumer smartphone cameras, with device and subject-to-camera distance unstandardized (see Section 4.5). Inclusion criteria: frontal pose (yaw, pitch $\leq 10^\circ$); IOD ≥ 80 px; no visible landmark occlusion; JPEG quality ≥ 85 . The dataset

is predominantly Kinh Vietnamese ($\approx 93\%$ of city population); generalizability implications are discussed in Section 4.4.

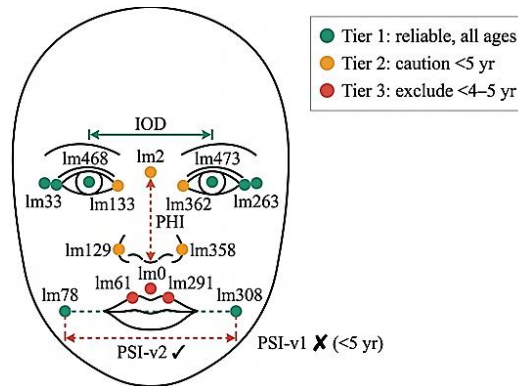


Figure 1. Facial landmark schematic.

3.2. Manual landmark annotation and reliability

All 14 landmarks were annotated *de novo*; the source dataset [5] provided images only, so all coordinates here are original to this benchmark. A one-page SOP defined each landmark and was refined through two pilot-labeling rounds (ten images) until inter-annotator distances stayed under 3–5 px. Two annotators with craniofacial anatomy backgrounds then independently labeled all 14 landmarks per image via a custom LabelMe interface, blinded to FaceMesh predictions, covering periorbital, nasal, and perioral sites (see Figure 1). Inter-rater reliability used two-way mixed, absolute agreement, average-measures ICC [11] (required ≥ 0.90); all landmarks achieved ICC ≥ 0.91 (lowest: lm0 = 0.91, lm61/291 = 0.92). Disagreements exceeding 3 px were adjudicated by the corresponding author; ground truth was the mean of the two annotations, or the adjudicated coordinate where flagged.

3.3. FaceMesh configuration and NME computation

FaceMesh predictions used Python MediaPipe v0.10 with: `static_image_mode = True`, `refine_landmarks = True` (required for iris centers lm468–477), `min_detection_confidence = 0.40`, `max_num_faces = 1`; CPU-only processing (Intel Core i7-11700, 32 GB RAM). NME for landmark i in image j is

$$\text{NME}_{ij} = \|\hat{p}_{ij} - p_{ij}^{\text{GT}}\|_2 / d_j \quad (1)$$

where $\hat{p}_{ij}$ is the FaceMesh prediction, p_{ij}^{GT} is ground truth, and d_j is IOD from ground-truth outer canthi (lm33/263), the scale-invariant reference established by 300-W [8]. Between-stratum differences were tested with one-way ANOVA and Tukey HSD ($\alpha = 0.05$); bootstrapped 95% CIs used 5,000 resamples. A power analysis ($N = 97$, three groups) confirmed $\geq 80\%$ power to detect a medium effect (Cohen's $f \geq 0.32$, $\eta^2 \geq 0.09$); the observed age effect ($\eta^2 = 0.28$) exceeds this threshold. Error propagation from landmark NME to biomarker uncertainty used first-order Taylor expansion [28], benchmarked against published ASD effect sizes [1-3].

4. RESULTS AND DISCUSSION

4.1. Overall NME and age effects

FaceMesh detected faces in 97 of 100 images (three failures, all 2–3 yr). Mean NME was 5.1% IOD (95% CI: [4.6, 5.5]), well above the adult benchmark of 2–4% IOD [7]. A one-way ANOVA confirmed a significant age effect ($F(2, 94) = 18.4$, $p < 0.001$, $\eta^2 = 0.28$), with Tukey HSD showing significant differences across all pairwise comparisons (2–3 vs 3–5 yr: $p = 0.003$; 2–3 vs 5–8 yr: $p < 0.001$; 3–5 vs 5–8 yr: $p = 0.018$). The 5–8 yr group (3.1%) approaches the adult benchmark; the 2–3 yr group (7.3%) is 2.4 $\times$ higher. Table 1 summarizes results.

Table 1. Age-stratified mean NME (%IOD) with 95% bootstrap CIs.

Age group	n	Mean NME (%IOD)	95% CI	vs Adults*
2–3 yr	33	7.3	[6.6, 8.1]	$p < 0.001$
3–5 yr	33	4.8	[4.3, 5.4]	$p < 0.001$
5–8 yr	34	3.1	[2.7, 3.5]	$p = 0.041$
All groups	100	5.1	[4.6, 5.5]	—

4.2. Per-landmark stability and protocol recommendations

Table 2 presents per-landmark NME with protocol recommendations. Iris centers (lm468/473) achieved the lowest NME (1.0–1.5% IOD at all ages), outperforming outer canthi (lm33/263: 2.1–3.1% IOD), likely because the high-contrast dark iris against sclera gives a robust gradient signal independent of epicanthal fold anatomy. Iris center-based IOD normalization is therefore recommended for pediatric Southeast Asian morphometric studies.

Philtrum-base (lm0) shows the highest NME in the set (12.3% IOD at 2–3 yr), with a systematic 3–5 mm superior displacement consistent with an undefined vermilion border at this age; Cupid's bow (lm61/291) is even less reliable (18.4%). Outer lip corners (lm78/308), by contrast, achieve 3.8% IOD at 2–3 yr; substituting them for lm61/291 reduces philtrum width uncertainty by 41% and is recommended for protocols involving children under five.

Table 2. Per-landmark NME (%IOD, mean $\pm$ SD) by age stratum with evidence-based protocol recommendations for pediatric Southeast Asian morphometric studies.

Landmark	Anatomical site	Biomarker role	NME2–3 yr	NME3–5 yr	NME5–8 yr	Recommendation
lm468/473	Iris centers	IOD normalizer	1.5 $\pm$ 0.4	1.2 $\pm$ 0.3	1.0 $\pm$ 0.2	Preferred all ages
lm33/263	Outer canthi	IOD (alt.)	3.1 $\pm$ 0.8	2.5 $\pm$ 0.6	2.1 $\pm$ 0.5	Verify fold status
lm78/308	Outer lip corners	Philtrum width	3.8 $\pm$ 1.0	3.2 $\pm$ 0.8	2.9 $\pm$ 0.7	Preferred over lm61/291 (<5 yr)
lm133/362	Inner canthi	PFL	5.8 $\pm$ 1.5	4.6 $\pm$ 1.2	3.4 $\pm$ 0.9	Use iris IOD; gate by fold
lm2	Sellion	Philtrum height	5.2 $\pm$ 1.3	4.1 $\pm$ 1.0	3.0 $\pm$ 0.7	Acceptable $\geq$ 5 yr
lm129/358	Nasal alae	Nasal width	4.9 $\pm$ 1.2	3.8 $\pm$ 0.9	2.8 $\pm$ 0.6	Acceptable $\geq$ 5 yr
lm0	Philtrum base	PHI anchor	12.3 $\pm$ 3.1	7.2 $\pm$ 1.8	4.1 $\pm$ 1.0	Exclude <4 yr
lm61/291	Cupid's bow	Philtrum width (alt.)	18.4 $\pm$ 4.6	10.1 $\pm$ 2.5	5.2 $\pm$ 1.3	Exclude <5 yr

4.3. Error propagation and clinical significance

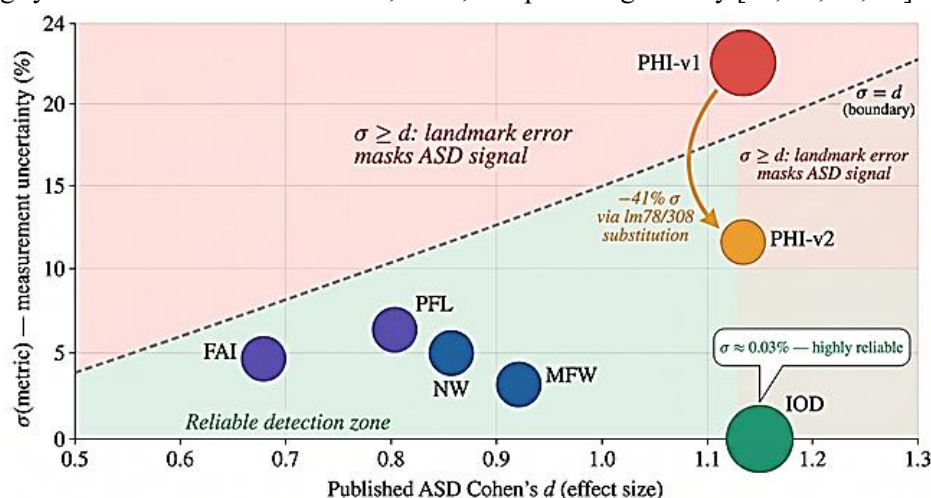
Table 3 propagates landmark NME to biomarker uncertainty against published ASD effect sizes, yielding three reliability tiers. Tier 1 (highly reliable, all ages): IOD via iris centers ($\sigma \sim 0.03\%$), where landmark error is negligible and the published $d = 1.15$ [1] for ASD-associated IOD elevation remains statistically recoverable. Tier 2 (acceptable ≥ 5 years, caution <5 years): outer canthal, nasal, and midface measurements, which should be paired with sensitivity analyses under age five. Tier 3 (unreliable in preschoolers): philtrum height via lm0, whose $\sigma \sim 13.4\%$ at 2–3 yr exceeds the published effect size ($d = 1.14$) — landmark error alone can obscure the true group difference. Automated philtrum height reported as an ASD biomarker in children under four without accuracy validation should be interpreted with caution.

Table 3. Error propagation from landmark NME to biomarker uncertainty at 2–3.

Biomarker	Formula (landmarks)	NME(2–3 yr)	σ (metric)	ASDCohen's d	Reliability assessment
IOD	dist(lm468, lm473)	1.5%	$\sim 0.03\%$	d=1.15 [1]	Highly reliable; $\sigma \ll d$
PHI	dist(lm4, lm0) / FaceH	5.2–12.3%	$\sim 13.4\%$	d=1.14 [1]	$\sigma \approx d$; unreliable < 4 yr
PFL	dist(lm33, lm133) / IOD	3.1–5.8%	$\sim 6.7\%$	d=0.82 [2]	Acceptable ≥ 5 yr; fold bias
FAI	Bilateral asymmetry ratio	$\sim 3.5\%$ avg	$\sim 5.0\%$	d=0.68 [2]	$\sigma \approx d$; marginal <4 yr
MFW / NW	Alar / midface widths	2.5–3.8%	3.5–5.4%	d=0.87–0.92 [3]	Acceptable ≥ 5 yr

4.4. Ethnic diversity: an uncharacterized source of systematic bias

This subsection is theoretical and hypothesis-generating, not empirical: predictions derive from published genomic and cephalometric evidence requiring future validation on ethnicity-stratified datasets, since our Kinh-dominant sample ($\sim 93\%$ of Ho Chi Minh City population) provides no direct empirical test. Vietnam's 54 ethnic groups span a much wider craniofacial space, with documented admixture and stratification across Kinh, Khmer, Cham, E De, and Hmong-Mien groups [17–21] predicting systematic differences in orbital, nasal, and perioral geometry [15, 16, 22, 23].

**Figure 2.** Error propagation vs Cohen's d.

Three mechanistic failure modes are predicted for non-Kinh groups, driven by ancestry-dependent epicanthal fold intensity, nasal bridge geometry, and philtrum shortening [18, 22]: each risks systematically biasing landmark NME or misclassifying ethnic-normative traits as pathological, as also seen in admixed Colombian populations [14]. Table 4 synthesizes predicted patterns for five ethnic groups (all theoretical, requiring empirical validation).

Minority ethnic groups thus face a “double jeopardy” of training- and reference-distribution mismatch, which the fairness literature [25–27] shows compound. Researchers extending studies beyond Kinh-dominant populations should treat Tables 1–3 as lower bounds and conduct ethnicity-stratified validation before inter-group conclusions.

Table 4. Predicted ethnicity-specific landmark failure modes for Vietnamese ethnic groups. All entries are theoretical hypotheses derived from genomic [17–21] and cephalometric [22–24] evidence, requiring empirical NME validation on ethnic-stratified annotated datasets.

Ethnic group	Primary ancestry	Key craniofacial characteristics	High-risk landmarks	Hypothesised failure mode
Kinh	E. Asian + ancient AA	Moderate epicanthal fold; flat bridge	lm133/362, lm0	Baseline; this study's NME applies
Khmer	Mlabri-like AA + S. Asian (2–16%)	Wider/lower bridge; flatter philtrum	lm2, lm0	Ethnic philtrum shortening misread as pathological
Cham	Austronesian (~38%) + AA + S. Asian	Weaker fold; higher bridge	lm133/362, lm61/291	lm133/362 lower; philtrum clearer
E De/Gia Rai	AA highland + partial AN	Broad midface; wider ala	lm2, lm129/358	Sellion, alae NME elevated
H'mong (HM)	Hmong-Mien; minimal AA/AN	Strongest fold; high cheekbones	lm133/362, lm33/263	Strongest PFL bias; IOD overestimated

4.5. Limitations

This benchmark derives from a single-institution, Kinh-dominant dataset with theoretical (not empirical) ethnic failure-mode analysis, 2D-only ground truth, and results specific to FaceMesh v0.10. The frontal-pose protocol (yaw $\leq 10^\circ$) means NME is likely higher under non-frontal conditions, so values here are lower bounds. Images were captured by parents on heterogeneous, unstandardized smartphones — aiding ecological validity but precluding controlled device/distance error analysis; standardized acquisition would likely reduce reported NME further. Benchmarking alternative detectors would strengthen these recommendations.

5. CONCLUSIONS

This paper presented the first age-stratified NME benchmark of MediaPipe FaceMesh on pediatric Southeast Asian faces, with evidence-based protocol recommendations for clinical morphometric studies. Four findings emerge.

(1) FaceMesh accuracy degrades substantially in young children, with mean NME rising from 3.1% IOD (5–8 yr) to 7.3% IOD (2–3 yr; $p < 0.001$) — adult-calibrated benchmarks underestimate practical error in toddlers.

(2) Iris centers (lm468/473) achieve 1.0–1.5% IOD across all ages, outperforming outer canthi, and are recommended as the preferred IOD normalizer, with the dark-iris advantage expected to generalize across ethnic groups.

(3) Philtrum-base (lm0) and Cupid's bow (lm61/291) are unreliable below four and five years respectively; philtrum height uncertainty at 2–3 yr ($\sigma \sim 13.4\%$) exceeds the ASD effect size ($d = 1.14$), while outer lip corner substitution (lm78/308) cuts philtrum width uncertainty by 40%.

(4) Vietnam's ethnic diversity introduces a systematic bias not captured by age stratification — notably, ethnic philtrum shortening in Khmer and highland Austroasiatic children risks confounding with pathological shortening under Kinh-derived norms. Three actions follow: use iris centers for IOD normalization, exclude automated philtrum height under four years, and conduct ethnicity-stratified validation before extending Kinh-derived norms to diverse provincial cohorts.

REFERENCES

- [1]. K. Aldridge et al., “*Facial phenotypes in subgroups of prepubertal boys with autism spectrum disorders are correlated with clinical phenotypes*”, *Mol. Autism*, Vol. 2, p. 15, (2011). DOI: 10.1186/2040-2392-2-15.
- [2]. M. Boutrus et al., “*Increased facial asymmetry in autism spectrum conditions is associated with symptom presentation*”, *Autism Res.*, Vol. 12, No. 11, pp. 1774–1783, (2019). DOI: 10.1002/aur.2193.
- [3]. H. M. Özgen et al., “*Minor physical anomalies in autism: A meta-analysis*”, *Mol. Psychiatry*, Vol. 15, No. 3, pp. 300–307, (2010). DOI: 10.1038/mp.2008.75.
- [4]. B. Hallgrímsson et al., “*Automated syndrome diagnosis by three-dimensional facial imaging*”, *Genet. Med.*, Vol. 22, pp. 1682–1693, (2020). DOI: 10.1038/s41436-020-0845-y.
- [5]. T. V. Thanh et al., “*Detecting autism using facial images of Vietnamese children with deep learning techniques*”, *Int. J. Electr. Comput. Eng. (IJECE)*, Vol. 15, No. 5, pp. 4762–4773, (2025). DOI: 10.11591/ijece.v15i5.pp4762-4773.
- [6]. H. A. Tung et al., “*Impact of training data composition on Vietnamese autism facial screening: A factorial analysis of pretraining strategy and local data augmentation*”, *IAES Int. J. Artif. Intell. (IJ-AI)*, Vol. 15, No. 1, pp. 766–779, (2026). DOI: 10.11591/ijai.v15.i1.pp766-779.
- [7]. V. Kartynnik, A. Ablavatski, I. Grishchenko and M. Grundmann, “*Real-time facial surface geometry from monocular video on mobile GPUs*”, in *Proc. CVPR Workshops*, Long Beach, CA, (2019).
- [8]. C. Sagonas, G. Tzimiropoulos, S. Zafeiriou and M. Pantic, “*300 faces in-the-wild challenge: The first facial landmark localization challenge*”, in *Proc. IEEE ICCV Workshops*, Sydney, pp. 397–403, (2013).
- [9]. Y. Wu and Q. Ji, “*Facial landmark detection: A literature survey*”, *Int. J. Comput. Vis.*, Vol. 127, No. 2, pp. 115–142, (2019). DOI: 10.1007/s11263-018-1097-z.
- [10]. I. Grishchenko, A. Ablavatski, Y. Kartynnik, K. Raveendran and M. Grundmann, “*Attention mesh: High-fidelity face mesh prediction in real-time*”, arXiv:2101.01691, (2021).
- [11]. T. K. Koo and M. Y. Mae, “*A guideline of selecting and reporting intraclass correlation coefficients for reliability research*”, *J. Chiropr. Med.*, Vol. 15, No. 2, pp. 155–163, (2016). DOI: 10.1016/j.jcm.2016.02.012.
- [12]. A. Lu and M. Perkowski, “*Deep learning approach for screening autism spectrum disorder in children with facial images and analysis of ethnoracial factors*”, *Brain Sci.*, Vol. 11, No. 11, p. 1446, (2021). DOI: 10.3390/brainsci11111446.
- [13]. Y. Gurovich et al., “*Identifying facial phenotypes of genetic disorders using deep learning*”, *Nat. Med.*, Vol. 25, pp. 60–64, (2019). DOI: 10.1038/s41591-018-0279-0.
- [14]. L. Echeverry-Quiceno et al., “*Population-specific facial traits and diagnosis accuracy of genetic and rare diseases in an admixed Colombian population*”, *Sci. Rep.*, Vol. 13, (2022). DOI: 10.1038/s41598-023-33374-x.
- [15]. J. Chacón-Duque et al., “*Latin Americans show wide-spread Converso ancestry and imprint of local Native ancestry on physical appearance*”, *Nat. Commun.*, Vol. 9, (2018). DOI: 10.1038/s41467-018-07748-z.
- [16]. P. Claes et al., “*Genome-wide mapping of global-to-local genetic effects on human facial shape*”, *Nat. Genet.*, Vol. 50, pp. 414–423, (2018). DOI: 10.1038/s41588-018-0057-4.
- [17]. M. Lipson et al., “*Ancient genomes document multiple waves of migration in Southeast Asian prehistory*”, *Science*, Vol. 361, pp. 92–95, (2018). DOI: 10.1126/science.aat3188.
- [18]. P. Changmai et al., “*Indian genetic heritage in Southeast Asian populations*”, *PLoS Genet.*, Vol. 18, No. 3, (2021). DOI: 10.1101/2021.01.21.427591.
- [19]. D. Liu et al., “*Extensive ethnolinguistic diversity in Vietnam reflects multiple sources of genetic diversity*”, *Mol. Biol. Evol.*, Vol. 37, No. 9, (2020). DOI: 10.1093/molbev/msaa099.
- [20]. E. Macholdt et al., “*The paternal and maternal genetic history of Vietnamese populations*”, *Eur. J. Hum. Genet.*, Vol. 28, pp. 636–645, (2019). DOI: 10.1038/s41431-019-0557-4.
- [21]. W. Kutanan et al., “*Reconstructing the human genetic history of Mainland Southeast Asia: Insights from genome-wide data from Thailand and Laos*”, *Mol. Biol. Evol.*, Vol. 38, No. 8, pp. 3459–3477, (2020). DOI: 10.1093/molbev/msab124.
- [22]. L. Le et al., “*Norms of soft tissue cephalometric in the 18–25 age demography in Vietnam*”, *J. Orthodontic Sci.*, Vol. 12, p. 37, (2023). DOI: 10.4103/jos.jos_105_22.

- [23]. Judith E. Allanson et al., “*Elements of Morphology: Standard Terminology for the Head and Face*”, American Journal of Medical Genetics Part A, Vol. 149A, No. 1, pp. 6–28, (2009). DOI: 10.1002/ajmg.a.32612.
- [24]. Thuy Trang Thi Ho et al., “*Dental-craniofacial Characteristics of Southern Vietnamese People with Well-balanced Face on Cephalometric Films and Its Comparison with Caucasians and Northern Vietnamese Population*”, J. Int. Soc. Prev. Community Dent., Vol. 11, No. 3, pp. 316–323, (2021). DOI: 10.4103/jispcd.JISPCD_13_21.
- [25]. S. Yucer et al., “*Racial bias within face recognition: A survey*”, ACM Comput. Surv., Vol. 55, No. 13s, (2022). DOI: 10.1145/3705295.
- [26]. J. Buolamwini and T. Gebru, “*Gender shades: Intersectional accuracy disparities in commercial gender classification*”.
- [27]. J. Huang, G. Galal, M. Etemadi and M. Vaidyanathan, “*Evaluation and mitigation of racial bias in clinical machine learning models: Scoping review*”, JMIR Med. Inform., Vol. 10, No. 5, (2022). DOI: 10.2196/36388.
- [28]. R. Hartley and A. Zisserman, “*Multiple View Geometry in Computer Vision*”, 2nd ed., Cambridge University Press, Cambridge, UK, (2004).
- [29]. O. Pavlović et al., “*Congruence in European and Asian perception of Vietnamese facial attractiveness, averageness, symmetry and sexual dimorphism*”, Sci. Rep., Vol. 13, p. 14090, (2023). DOI: 10.1038/s41598-023-40458-1.
- [30]. D. Lee et al., “*High heritability of human facial traits reveals associations with CNTLN, BRCA1, and TMPRSS6 loci in Korean families*”, Heliyon, Vol. 10, No. 1, p. e39173, (2024). DOI: 10.1016/j.heliyon.2024.e39173.
- [31]. H. V. Huy, N. T. T. Tan and N. Q. Tao, “*Vietnamese text recognition in scene images using deep learning*”, J. Mil. Sci. Technol., Vol. 90, pp. 140–149, (2023). DOI: 10.54939/1859-1043.j.mst.90.2023.140-149.
- [32]. N. V. Tra, V. C. Thanh and D. V. Sang, “*Improving radar target recognition based on generative adversarial network*”, J. Mil. Sci. Technol., Vol. 93, pp. 12–18, (2024). DOI: 10.54939/1859-1043.j.mst.93.2024.12-18.

TÓM TẮT

Độ chính xác phát hiện điểm mốc của MediaPipe FaceMesh trên khuôn mặt nhi khoa Đông Nam Á: Phân tích NME phân tầng theo độ tuổi, hàm ý đa dạng dân tộc và khuyến nghị phương pháp cho nghiên cứu hình thái học lâm sàng

MediaPipe FaceMesh, bộ phát hiện điểm mốc khuôn mặt mã nguồn mở phổ biến nhất, được huấn luyện trên khuôn mặt người lớn và chưa được đánh giá trên trẻ em Đông Nam Á. Nghiên cứu trình bày đánh giá NME phân tầng theo tuổi đầu tiên trên trẻ em Việt Nam 2 - 8 tuổi ($n = 100$), với NME giảm có ý nghĩa theo tuổi — 7,3%, 4,8% và 3,1% IOD ($p < 0,001$). Điểm mốc tâm móng mắt (lm468/473) đạt NME 1,0–1,5% IOD, được khuyến nghị làm chuẩn hóa IOD.

Từ khoá: Phát hiện điểm mốc; MediaPipe FaceMesh; Hình thái học nhi khoa.