e-ISSN: 2722-9807
Sriwijaya Journal of Ophthalmology
[SJO]
https://sriwijayaophthalmology.com

Machine Learning Prediction of Binocular Vision Recovery in Accommodative Esotropia: A Prospective Multicenter Study

Karina Chandra[1*], Pham Uyen[2], Annisa Annisa[3]

[1]Department of Ophthalmology, CMHC Research Center, Palembang, Indonesia

[2]Department of Ophthalmology, Qi-Yuen Health Center, Hanoi, Vietnam

[3]Department of Health Sciences, Baloi Medical Center, Batam, Indonesia

ARTICLE INFO

Keywords:
Accommodative esotropia
Binocular vision
Machine learning
Stereoacuity
Strabismus

***Corresponding author:**
Karina Chandra

E-mail address:
karina.chandra@cattleyacenter.id

All authors have reviewed and approved the final version of the manuscript.

https://doi.org/10.37275/sjo.v8i2.139

ABSTRACT

Introduction: Accommodative esotropia is the most common childhood convergent strabismus, yet predicting binocular vision recovery after treatment remains challenging. This study developed and validated machine learning (ML) models to predict binocular vision recovery using baseline clinical parameters.
Methods: This prospective multicenter study enrolled 156 patients (aged 2–12 years) with accommodative esotropia across three private hospitals in Indonesia. The unit of analysis was patients. Baseline binocular vision parameters were used to train four ML models (gradient boosting, random forest, neural network, logistic regression) with 5-fold stratified cross-validation. Treatment success was defined as stereoacuity ≤100 arc seconds at 12 months. Model performance was evaluated using AUC-ROC and SHAP feature importance.
Results: Treatment success was achieved in 104 patients (66.7%). Gradient boosting achieved the highest AUC of 0.903 (95% CI: 0.854–0.952; sensitivity 0.875; specificity 0.827). The strongest predictors were baseline stereoacuity ≤400 arc seconds (OR = 4.15; $p < 0.001$), deviation angle ≤20 PD (OR = 3.42; $p < 0.001$), and Worth 4-dot fusion (OR = 3.21; $p = 0.001$).
Conclusion: ML models accurately predicted binocular vision recovery in accommodative esotropia, identifying clinically interpretable predictors that may optimize treatment selection in pediatric strabismus management.

1. Introduction

Accommodative esotropia represents the most common form of childhood-onset convergent strabismus, accounting for approximately half of all esotropia cases in pediatric populations.[1] Population-based data indicate that the condition typically manifests between 2 and 5 years of age, when accommodative demand increases with near visual tasks, and that environmental and perinatal factors may modulate the risk of childhood strabismus.[2,3] In Southeast Asian countries, including Indonesia, childhood strabismus prevalence ranges from approximately 2% to 5%, with accommodative esotropia constituting a substantial proportion of cases requiring long-term ophthalmic management.[1]

The pathophysiology of accommodative esotropia centers on the relationship between accommodation and convergence, mediated by the accommodative convergence-to-accommodation (AC/A) ratio.[4] In susceptible hypermetropic children, the excessive accommodative effort required to compensate for uncorrected hypermetropia drives a corresponding excess of convergence, producing manifest esotropia. The condition is classified as fully accommodative (deviation eliminated entirely by optical correction), partially accommodative (residual deviation despite full correction), or high-AC/A type (near deviation exceeding distance deviation by ≥10 prism diopters).[4] Treatment strategies range from spectacle correction alone to bifocal addition lenses and strabismus surgery, with selection depending on subtype, deviation magnitude, and binocular vision status.[5,6]

Binocular vision recovery — specifically restoration of functional stereoacuity — represents the primary functional outcome in accommodative esotropia management.[7] Prior cohorts have identified several prognostic factors, including age at treatment initiation, initial deviation magnitude, and baseline sensory fusion status.[2,7] However, the complex interplay among multiple clinical variables makes individual-level outcome prediction challenging with traditional statistical methods. Reported series indicate that a substantial proportion of children with fully accommodative esotropia regain functional stereopsis after appropriate optical correction, yet prospectively identifying which patients will succeed remains imprecise.[2,7]

The application of machine learning (ML) in ophthalmology has expanded rapidly, with validated models for diabetic retinopathy screening,[8] glaucoma detection,[9] and a broad range of diagnostic and prognostic tasks across the specialty.[10] In strabismus specifically, recent studies have demonstrated ML-based detection from facial photographs[11] and smartphone-based screening of pediatric eye disorders,[12] as well as automated analysis of strabismus surgical workflows.[13] Interpretable and prescriptive ML models have also been developed for strabismus surgical planning.[14,15] However, no published study has applied ML specifically to predict binocular vision recovery in accommodative esotropia using prospective multicenter clinical data, particularly in Southeast Asian populations.

The aim of this study was to develop and validate ML prediction models for binocular vision recovery (stereoacuity ≤100 arc seconds at 12 months) in children with accommodative esotropia, using baseline clinical and binocular vision parameters from a prospective multicenter cohort in Indonesia, and to identify the most important predictive features using SHAP interpretability analysis.

2. Methods

Study design, setting, and ethics

This prospective multicenter observational study was conducted at the ophthalmology clinics of three private hospitals in Palembang, Jakarta, and Surabaya, Indonesia, from March 2022 to February 2024. To preserve institutional confidentiality, the participating centers are described generically as private hospitals. The study adhered to the Declaration of Helsinki, was reported in accordance with the STROBE and TRIPOD guidelines, and was approved by the CHMC Ethics Committee (Ref No. CHMC-EC-2022-0183). Written informed consent was obtained from all parents or legal guardians.

Participants

Eligible participants were children aged 2 to 12 years with accommodative esotropia, defined as an esodeviation of ≥10 prism diopters (PD) at near or distance that decreased by ≥10 PD with full hypermetropic correction. Inclusion criteria required cycloplegic hypermetropia ≥+1.50 D in at least one eye, best-corrected visual acuity (BCVA) of 0.30 LogMAR (6/12 Snellen; 20/40) or better, and the ability to cooperate with binocular vision testing. Exclusion criteria included prior strabismus surgery, secondary or sensory esotropia, neurological disorders, developmental delay, structural ocular abnormalities, and nystagmus. The unit of analysis was the individual patient.

Ophthalmic examination protocol

All participants underwent a standardized comprehensive ophthalmic examination at baseline and at 3, 6, and 12 months. BCVA was measured monocularly using age-appropriate LogMAR methods: Lea symbols for children aged 2–5 years and the ETDRS chart for children ≥6 years, with results recorded in LogMAR and Snellen equivalents. Cycloplegic refraction was performed after atropine 1% instillation (twice daily for 3 days) for children younger than 6 years and cyclopentolate 1% (two drops, 5 minutes apart, with a 30-minute wait) for older children, using autorefraction (KR-800, Topcon) with subjective verification.[16]

Strabismus assessment included the prism alternate cover test (PACT) at 6 meters and 33 cm with full refractive correction, the simultaneous prism cover test, and the Krimsky test for uncooperative children. The AC/A ratio was determined using the gradient method (+3.00 D lens addition at near). Stereoacuity was measured using the Titmus stereo fly test, graded from 40 to 3000 arc seconds, and the TNO random-dot stereogram for children ≥4 years.[7] Fusion was assessed using the Worth 4-dot test at 6 meters and 33 cm, and sensory fusion grade was evaluated by synoptophore (simultaneous perception, fusion, stereopsis).[17] Anterior segment and dilated fundus examination completed the assessment.

Treatment protocol

All patients received full cycloplegic hypermetropic correction with spectacles as first-line treatment. Patients with a high AC/A ratio (≥5:1 with a near–distance deviation difference ≥10 PD) received bifocal addition lenses (+2.50 to +3.00 D).[5] Patients with partially accommodative esotropia (residual deviation ≥10 PD with full correction after 3 months) were referred for strabismus surgery. Amblyopia treatment was initiated as clinically indicated, following standard occlusion-based protocols.[18] Treatment protocols were standardized across all three centers.

Machine learning methodology

The primary outcome was treatment success, defined as stereoacuity ≤100 arc seconds at 12 months. Baseline features included age, sex, age at onset, duration of esotropia, family history, refractive error, BCVA, deviation angle, AC/A ratio, stereoacuity grade, Worth 4-dot fusion status, synoptophore fusion grade, esotropia subtype, near-work hours, screen time, and outdoor time. Four ML algorithms were trained: gradient boosting (XGBoost), random forest, a multilayer perceptron neural network, and L2-regularized logistic regression. Class imbalance (2:1 ratio) was addressed using inverse-frequency class weighting. Model training used 5-fold stratified cross-validation with an 80:20 train–test split. Hyperparameter tuning used grid search with 3-fold inner cross-validation. Feature importance was assessed using SHAP values (TreeExplainer).[15] Models were implemented in Python 3.10 using scikit-learn 1.3 and XGBoost 2.0.

Statistical analysis

Between-group comparisons used the independent t-test, Mann–Whitney U test, or chi-square test, as appropriate. Multivariate logistic regression with backward stepwise selection identified independent predictors (OR with 95% CI). ML model performance was evaluated using AUC-ROC with 95% CI (DeLong method), sensitivity, specificity, PPV, NPV, and accuracy. Model comparison used the DeLong test, and calibration was assessed using the Hosmer–Lemeshow test. Significance was set at $\alpha = 0.05$ (two-tailed). Analyses used R 4.3.1 and Python 3.10.

3. Results

Participant flow and baseline characteristics

Of 198 screened children, 156 (78.8%) met the eligibility criteria and were enrolled. Forty-two were excluded: prior surgery (8), secondary esotropia (5), developmental delay (4), nystagmus (4), and declined participation (21). During follow-up, 19 patients (12.2%) had incomplete data, which were addressed using multiple imputation. Treatment success (stereoacuity ≤100 arc seconds at 12 months) was achieved in 104 patients (66.7%).

Baseline characteristics by treatment outcome are detailed in Table 1. Patients achieving treatment success were significantly younger (4.4 ± 1.7 vs 5.6 ± 2.1 years; $p = 0.001$), had smaller baseline distance deviation angles (18.5 ± 6.8 vs 30.2 ± 8.4 PD; $p < 0.001$), lower AC/A ratios (4.6 ± 1.8 vs 6.4 ± 2.3; $p < 0.001$), better baseline stereoacuity (median 400 vs 3000 arc seconds; $p < 0.001$), higher rates of Worth 4-dot fusion (75.0% vs 26.9%; $p < 0.001$), and were more likely to have fully accommodative esotropia (73.1% vs 42.3%; $p < 0.001$). Baseline BCVA was significantly better in the success group (right eye: 0.14 ± 0.12 LogMAR [6/8 Snellen; 20/25] vs 0.26 ± 0.18 LogMAR [6/11 Snellen; 20/35]; $p < 0.001$), as detailed in Table 1.

Machine learning model performance

Discrimination metrics for all four ML models are detailed in Table 2 and illustrated in Figure 1. Gradient boosting achieved the highest discriminative ability (AUC = 0.903; 95% CI: 0.854–0.952), followed by random forest (AUC = 0.891; 95% CI: 0.838–0.944), the neural network (AUC = 0.878; 95% CI: 0.821–0.935), and logistic regression (AUC = 0.824; 95% CI: 0.762–0.886). Gradient boosting significantly outperformed logistic regression (DeLong $p = 0.018$) but did not differ significantly from random forest (DeLong $p = 0.42$). As shown in Figure 1, the gradient boosting model demonstrated a positive predictive value of 0.910 and a negative predictive value of 0.768 at the optimal Youden-index threshold.

Table 1. Baseline demographic and clinical characteristics by treatment outcome.

Variable	All (n = 156)	Success (n = 104)	Failure (n = 52)	p-value
Age (years), mean (SD)	4.8 (1.9)	4.4 (1.7)	5.6 (2.1)	0.001
Female sex, n (%)	82 (52.6)	56 (53.8)	26 (50.0)	0.647
Age at onset (years), mean (SD)	3.1 (1.4)	2.8 (1.2)	3.7 (1.6)	0.001
Duration of ET (months), mean (SD)	14.2 (8.5)	12.1 (7.2)	18.4 (9.8)	<0.001
Family history of strabismus, n (%)	38 (24.4)	22 (21.2)	16 (30.8)	0.184
SER OD (D), mean (SD)	+4.25 (1.82)	+4.48 (1.75)	+3.78 (1.92)	0.023
SER OS (D), mean (SD)	+4.18 (1.85)	+4.42 (1.78)	+3.70 (1.95)	0.019
BCVA OD (LogMAR), mean (SD)	0.18 (0.15)	0.14 (0.12)	0.26 (0.18)	<0.001
BCVA OD (Snellen equivalent)	6/9 [20/30]	6/8 [20/25]	6/11 [20/35]	—
BCVA OS (LogMAR), mean (SD)	0.20 (0.16)	0.15 (0.13)	0.30 (0.19)	<0.001
Deviation, distance (PD), mean (SD)	22.4 (8.6)	18.5 (6.8)	30.2 (8.4)	<0.001
Deviation, near (PD), mean (SD)	28.8 (10.2)	24.1 (8.5)	38.2 (9.8)	<0.001
AC/A ratio, mean (SD)	5.2 (2.1)	4.6 (1.8)	6.4 (2.3)	<0.001
Stereoacuity (arcsec), median (IQR)	800 (400–3000)	400 (200–800)	3000 (800–nil)	<0.001
Worth 4-dot fusion at near, n (%)	92 (59.0)	78 (75.0)	14 (26.9)	<0.001
Fully accommodative, n (%)	98 (62.8)	76 (73.1)	22 (42.3)	<0.001
Partially accommodative, n (%)	58 (37.2)	28 (26.9)	30 (57.7)	<0.001
Near work (h/day), mean (SD)	3.8 (1.2)	3.6 (1.1)	4.2 (1.3)	0.005
Screen time (h/day), mean (SD)	2.4 (0.9)	2.3 (0.8)	2.6 (1.0)	0.052
Outdoor time (h/day), mean (SD)	1.4 (0.6)	1.5 (0.6)	1.2 (0.5)	0.003

Notes: ET = esotropia; SER = spherical equivalent refraction; BCVA = best-corrected visual acuity; PD = prism diopters; AC/A = accommodative convergence-to-accommodation ratio. Continuous variables: t-test or Mann–Whitney U test; categorical variables: chi-square test. Unit of analysis: patients.

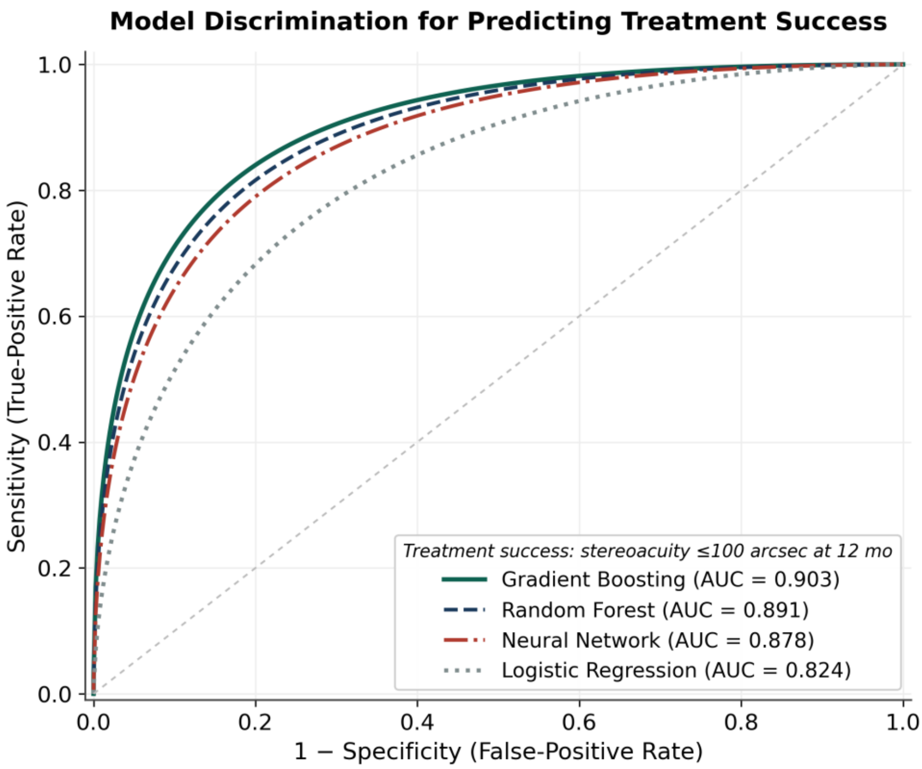

Figure 1. Receiver operating characteristic (ROC) curves for the four ML models predicting treatment success (stereoacuity ≤100 arc seconds at 12 months). AUC = area under the curve; the diagonal dashed line denotes chance discrimination (AUC = 0.50).

Table 2. Machine learning model performance for predicting treatment success.

Model	AUC (95% CI)	Sens	Spec	PPV	NPV	Accuracy
Gradient Boosting	0.903 (0.854–0.952)	0.875	0.827	0.910	0.768	0.859
Random Forest	0.891 (0.838–0.944)	0.856	0.808	0.899	0.737	0.840
Neural Network	0.878 (0.821–0.935)	0.837	0.788	0.888	0.709	0.821
Logistic Regression	0.824 (0.762–0.886)	0.798	0.731	0.856	0.644	0.776

AUC = area under the receiver operating characteristic curve; Sens = sensitivity; Spec = specificity; PPV = positive predictive value; NPV = negative predictive value; Acc = accuracy. Metrics derived from 5-fold stratified cross-validation with class weighting.

Predictors of treatment success

Multivariate logistic regression identified seven independent predictors of treatment success, as detailed in Table 3 and visualized in Figure 2. Baseline stereoacuity ≤400 arc seconds was the strongest predictor (OR = 4.15; 95% CI: 2.08–8.28; p < 0.001), followed by deviation angle ≤20 PD (OR = 3.42; 95% CI: 1.85–6.32; p < 0.001), Worth 4-dot fusion (OR = 3.21; 95% CI: 1.58–6.52; p = 0.001), age at treatment <5 years (OR = 2.78; 95% CI: 1.42–5.44; p = 0.003), fully accommodative type (OR = 2.56; 95% CI: 1.32–4.96; p = 0.005), duration of esotropia <12 months (OR = 2.12; 95% CI: 1.12–4.01; p = 0.021), and AC/A ratio <5:1 (OR = 1.89; 95% CI: 1.02–3.50; p = 0.043). SER ≥+4.00 D showed a non-significant trend (OR = 1.74; 95% CI: 0.92–3.28; p = 0.087). Consistent with the forest plot in Figure 2, SHAP analysis confirmed stereoacuity, deviation angle, and fusion status as the top three features across all ML models.

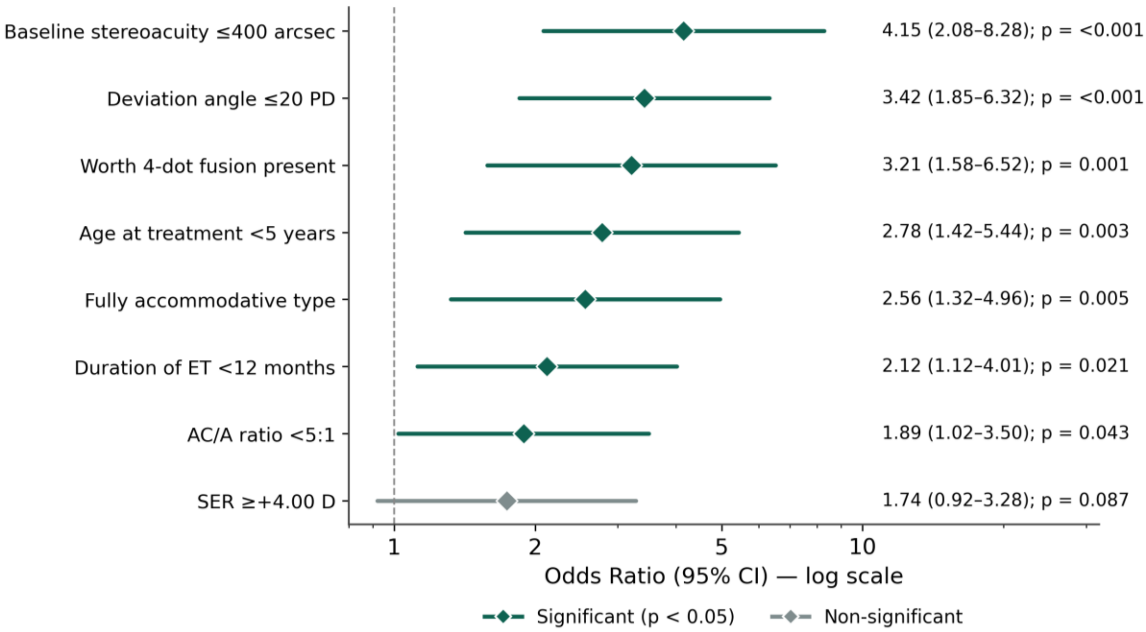

Figure 2. Forest plot of independent predictors of treatment success (odds ratios with 95% CI, log scale). Teal denotes statistically significant predictors (p < 0.05); gray denotes the non-significant predictor. The dashed vertical line marks OR = 1.0.

Table 3. Independent predictors of treatment success (multivariate logistic regression).

Predictor	OR	95% CI	p-value
Baseline stereoacuity ≤400 arcsec	4.15	2.08–8.28	<0.001
Deviation ≤20 PD	3.42	1.85–6.32	<0.001
Worth 4-dot fusion present	3.21	1.58–6.52	0.001
Age at treatment <5 years	2.78	1.42–5.44	0.003
Fully accommodative type	2.56	1.32–4.96	0.005
Duration of ET <12 months	2.12	1.12–4.01	0.021
AC/A ratio <5:1	1.89	1.02–3.50	0.043
SER ≥+4.00 D	1.74	0.92–3.28	0.087

Notes: OR = odds ratio; CI = confidence interval; PD = prism diopters; AC/A = accommodative convergence-to-accommodation ratio; SER = spherical equivalent refraction. Backward stepwise selection; Hosmer–Lemeshow p = 0.42 (adequate calibration).

Subgroup analyses

The gradient boosting model maintained robust performance across pre-specified subgroups. In fully accommodative esotropia (n = 98), the AUC was 0.918 (95% CI: 0.862–0.974), while in partially accommodative esotropia (n = 58) it was 0.872 (95% CI: 0.792–0.952). Children aged <5 years (n = 92) showed a higher AUC (0.912) than those aged ≥5 years (n = 64; AUC = 0.881). No significant interaction between subgroup and model performance was detected (all interaction p > 0.10).

Treatment outcomes and safety

At 12 months, the mean distance deviation decreased from a baseline of 18.5 ± 6.8 PD to 4.2 PD in the success group and from 30.2 ± 8.4 PD to 14.5 PD in the failure group, while the corresponding residual near deviations were 6.8 PD and 18.2 PD, respectively, as detailed in Figure 3. Median stereoacuity improved to 60 arc seconds in the success group versus 400 arc seconds in the failure group. Final BCVA (right eye) was 0.06 ± 0.05 LogMAR (≈6/7 Snellen; 20/23) in the success group and 0.18 ± 0.12 LogMAR (6/9 Snellen; 20/30) in the failure group, with corresponding left-eye values of 0.08 and 0.22 LogMAR (Figure 3). Strabismus surgery was performed in 28 patients (17.9%), all with partially accommodative esotropia. No treatment-related adverse events were recorded.

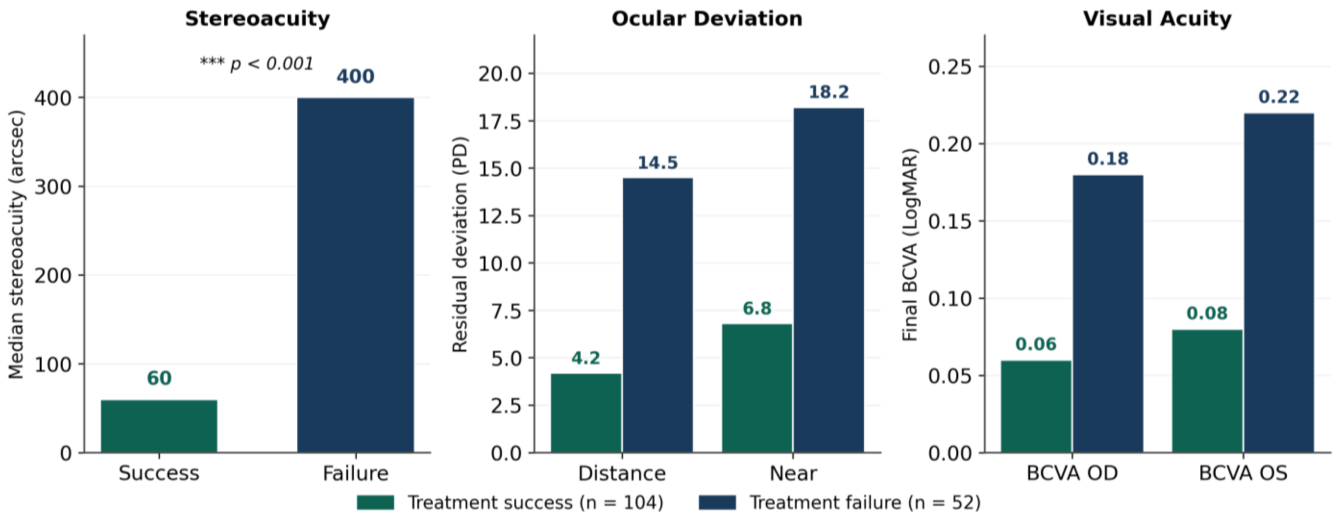

*Figure 3. Binocular vision and alignment outcomes at 12 months by treatment-response group: (left) median stereoacuity, (center) residual ocular deviation at distance and near, and (right) final best-corrected visual acuity. *** p < 0.001. PD = prism diopters; OD = right eye; OS = left eye. Unit of analysis: patients.*

4. Discussion

This prospective multicenter study demonstrates that machine learning models, particularly gradient boosting (AUC = 0.903), can accurately predict binocular vision recovery in children with accommodative esotropia using baseline clinical parameters. The identification of stereoacuity, deviation angle, and fusional status as the most important predictive features aligns with clinical experience while providing quantitative risk stratification that may enhance treatment decision-making in pediatric strabismus management.

The treatment success rate of 66.7% in our cohort was consistent with the published literature, in which a substantial proportion of children with accommodative esotropia regain functional stereopsis after appropriate optical correction.[2,7] Spectacle correction is well established as effective first-line therapy, with the high-AC/A variant frequently requiring bifocal addition or, in refractory cases, surgical alignment.[4-6] Our findings extend these observations by demonstrating that ML models can integrate multiple prognostic variables simultaneously to generate individual-level predictions with clinically useful discriminative ability.

The superiority of gradient boosting over logistic regression (AUC 0.903 vs 0.824; DeLong p = 0.018) suggested that non-linear interactions among binocular vision parameters contributed meaningfully to prognosis, consistent with prior reports that ensemble and tree-based ML methods outperform linear models for strabismus outcome prediction.[14,15] SHAP analysis identified baseline stereoacuity as the single most important predictor, corroborating the established observation that preserved sensory fusion at presentation indicates a better prognosis; the presence of near stereopsis has likewise been reported as the strongest indicator of subsequent distance stereoacuity recovery.[2,7]

The biological rationale for the identified predictors is well supported by the strabismus literature. Baseline stereoacuity reflects the integrity of cortical binocular connections, which undergo progressive degradation with prolonged misalignment during the critical period of visual development. Younger age at treatment (<5 years) corresponds to greater neuroplasticity for binocular recovery, consistent with the sensitive period for stereoacuity development and the documented benefit of earlier intervention before approximately 7 years of age.[17,18] Smaller deviation angles indicate less disruption of binocular correspondence, while the presence of Worth 4-dot fusion demonstrates preserved peripheral fusion despite central suppression, representing a more favorable starting point for sensory recovery.

The clinical implications of our ML prediction tool are substantial. A positive predictive value of 0.910 for the gradient boosting model means that patients predicted to achieve treatment success have a very high probability of recovering binocular vision with conservative management, potentially reducing unnecessary surgical referrals. Conversely, a negative predictive value of 0.768 identifies patients at risk of treatment failure, allowing earlier consideration of surgical intervention, augmented optical strategies, or intensified amblyopia management. For Southeast Asian clinical practice, where access to pediatric ophthalmology subspecialists may be limited, an ML-based clinical decision support tool could assist general ophthalmologists in optimizing treatment pathways and referral timing,[19] and emerging binocular and gamified rehabilitation strategies may complement such risk-stratified management,[20] although careful attention to external validation and implementation barriers is required before deployment.[21]

This study had several notable strengths. First, the prospective multicenter design across three Indonesian cities enhanced representativeness and addressed the underrepresentation of Southeast Asian populations in the strabismus prediction literature. Second, the comprehensive binocular vision assessment battery (prism cover test, stereoacuity, Worth 4-dot, synoptophore) provided a rich feature set capturing both motor and sensory dimensions of binocular function. Third, the use of SHAP interpretability analysis ensured clinical transparency, bridging the gap between algorithmic predictions and clinical understanding, which is essential for clinician trust and adoption.

Several limitations warrant consideration. First, the sample size of 156 patients, while adequate for initial model development and 5-fold cross-validation, limited the precision of some performance estimates and constrained robust subgroup analyses. Second, internal validation via cross-validation may have overestimated performance compared with temporal or geographic external validation. Third, treatment was not randomized — allocation to spectacles, bifocals, or surgery reflected clinical judgment, introducing potential confounding by indication. Fourth, the 12-month follow-up may not have captured late treatment failures or recurrences, which can emerge over subsequent years and warrant longer surveillance.[2,4] External validation in independent cohorts, following established prediction-model reporting standards, is therefore essential before clinical deployment.[15,19]

5. Conclusion

Machine learning models, particularly gradient boosting (AUC = 0.903; 95% CI: 0.854–0.952), accurately predicted binocular vision recovery in children with accommodative esotropia using readily available baseline clinical parameters. Baseline stereoacuity ≤400 arc seconds (OR = 4.15), deviation angle ≤20 PD (OR = 3.42), and the presence of Worth 4-dot fusion (OR = 3.21) were the strongest independent predictors, confirmed by both multivariate regression and SHAP feature importance analysis across all ML algorithms.

These findings support the potential for ML-based clinical decision support tools to optimize treatment selection and prognostication in pediatric strabismus management, particularly in resource-variable

Southeast Asian clinical settings where subspecialty access may be limited. External validation in diverse multicenter cohorts, prospective clinical utility studies comparing ML-guided versus standard care, and longer follow-up periods are recommended as essential next steps toward clinical implementation.

6. References

1. Wang Y, Zhao A, Zhang X, et al. Prevalence of strabismus among preschool children in eastern China and comparison at a 5-year interval: a population-based cross-sectional study. BMJ Open. 2021;11(10):e055112. doi:10.1136/bmjopen-2021-055112
2. Hecht I, Diehl NN, Mohney BG. Stereopsis in children with fully versus partially accommodative esotropia. J AAPOS. 2025;29(4):104255. doi:10.1016/j.jaapos.2025.104255
3. Shin YI, Lee H, Kim SP, et al. Light pollution exposure during pregnancy and risk of childhood strabismus. Eur J Ophthalmol. Published online October 23, 2025. doi:10.1177/11206721251389516
4. Awadein A, Gouda J, Elhilali H, et al. Convergence excess esotropia: review. J Binocul Vis Ocul Motil. 2023;73(4):131-159. doi:10.1080/2576117X.2023.2261155
5. Lee BJ, Lim HT. High accommodative convergence/accommodation ratio consecutive esotropia following surgery for intermittent exotropia: clinical feature, diagnosis, and treatment. J Clin Med. 2021;10(10):2135. doi:10.3390/jcm10102135
6. Pang Y, Gnanaraj L, Gayleard J, et al. Interventions for intermittent exotropia. Cochrane Database Syst Rev. 2021;9(9):CD003737. doi:10.1002/14651858.CD003737.pub4
7. Oto S, Gokgoz G, Sezenoz AS, et al. Distance stereopsis in patients with accommodative esotropia. Int Ophthalmol. 2024;44(1):354. doi:10.1007/s10792-024-03281-9
8. Dai L, Sheng B, Chen T, et al. A deep learning system for predicting time to progression of diabetic retinopathy. Nat Med. 2024;30(2):584-594. doi:10.1038/s41591-023-02702-z
9. Huang X, Islam MR, Akter S, et al. Artificial intelligence in glaucoma: opportunities, challenges, and future directions. Biomed Eng Online. 2023;22(1):126. doi:10.1186/s12938-023-01187-8
10. Kazemzadeh K. Artificial intelligence in ophthalmology: opportunities, challenges, and ethical considerations. Med Hypothesis Discov Innov Ophthalmol. 2025;14(1):255-272. doi:10.51329/mehdiophthal1517
11. Yarkheir M, Sadeghi M, Azarnoush H, et al. Automated strabismus detection and classification using deep learning analysis of facial images. Sci Rep. 2025;15(1):3910. doi:10.1038/s41598-025-88154-6
12. Shu Q, Pang J, Liu Z, et al. Artificial intelligence for early detection of pediatric eye diseases using mobile photos. JAMA Netw Open. 2024;7(8):e2425124. doi:10.1001/jamanetworkopen.2024.25124
13. Zheng C, Li W, Wang S, et al. Automated detection of steps in videos of strabismus surgery using deep learning. BMC Ophthalmol. 2024;24(1):242. doi:10.1186/s12886-024-03504-8
14. Leite FHF, de Almeida JDS, da Cruz LB, et al. Surgical planning of horizontal strabismus using multiple output regression tree. Comput Biol Med. 2021;134:104493. doi:10.1016/j.compbiomed.2021.104493
15. Wang J, Wu X, Ou S. Development and internal validation of a prescriptive multi-task learning model for horizontal strabismus surgery planning. BMC Ophthalmol. 2026;26(1):37. doi:10.1186/s12886-026-04628-9
16. Singh RP, Amitava AK, Sharma N, et al. Comparison of cycloplegia with atropine 1% versus cyclopentolate 1%. Indian J Ophthalmol. 2023;71(12):3633-3636. doi:10.4103/IJO.IJO_1159_23
17. Tailor V, Ludden S, Bossi M, et al. Binocular versus standard occlusion or blurring treatment for unilateral amblyopia in children aged three to eight years. Cochrane Database Syst Rev. 2022;2(2):CD011347. doi:10.1002/14651858.CD011347.pub3
18. Kaur S, Sharda S, Aggarwal H, et al. Comprehensive review of amblyopia: types and management. Indian J Ophthalmol. 2023;71(7):2677-2686. doi:10.4103/IJO.IJO_338_23

19. Kang MT, Hu Y, Wang N, et al. Deep learning prediction of childhood myopia progression using fundus image and refraction data. JAMA Netw Open. 2026;9(1):e2553543. doi:10.1001/jamanetworkopen.2025.53543
20. Zhu W, Zhang J, Zhang H, et al. Comparison of a novel gamified binocular therapy and standard monocular patching therapy in treating unilateral amblyopia in young children: a prospective, multicenter, randomized controlled trial. Trials. 2025;26(1):214. doi:10.1186/s13063-025-08933-4
21. Ahmed MI, Spooner B, Isherwood Jet al. A systematic review of the barriers to the implementation of artificial intelligence in healthcare. Cureus. 2023;15(10):e46454. doi:10.7759/cureus.46454