Manuscript – EICON 2026
ESOFT International Conference
ESOFT Uni, Sri Lanka

Trilingual Offline Smart AAC with On-Device Facial Expression Recognition for Autism in Sri Lanka

Fernando E. Y. P.[#], Fonseka N., Akalanka P. D. A. U.
School of Computing, ESOFT Metro Campus, Galle, Sri Lanka
[#]Corresponding author: yasaspasindufernando@gmail.com, Contact No: +94 776905654

Abstract

Introduction: Some children with autism spectrum disorder (ASD) have limited or unreliable speech and depend on augmentative and alternative communication (AAC). In Sri Lanka, AAC must support Sinhala, Tamil and English, work offline and be affordable for families who cannot afford commercial products.

Objectives: To design, implement and verify an offline-first Android AAC prototype combining trilingual AAC, caregiver-customisable vocabulary and on-device facial expression recognition (FER) framed as supportive observation, not diagnosis.

Methods: The prototype was built in Flutter for Android 8.0 and above with compiled trilingual vocabulary, local storage and device text-to-speech. On-device FER was performed through TensorFlow Lite using a fine-tuned EfficientNetB0 with MobileNetV2 fallback, trained on three public datasets with six emotion classes. Uncertain states were produced by a face-detection gate, five-frame burst averaging and confidence-margin checks. Verification was conducted through unit tests, tensor-contract checks, real-device walkthroughs, Google Play testing and a questionnaire from 11 testers. No children were recruited and no pilot was conducted.

Results: All 41 unit tests and 10 walkthroughs passed and both models satisfied the tensor contract. Validation accuracy reached 70.5% after fine-tuning, but accuracy on the held-out test set was 52.9%, with weighted F1-score 0.54: joy reached F1-score 0.84 and sadness only 0.19. Nine of 11 judged the application could help a child express basic needs.

Conclusions: The prototype is technically feasible and locally relevant, but measured recognition performance supports use only as a cautious cue under caregiver supervision. Clinical evaluation and formal usability testing remain outstanding.

Keywords: *Augmentative and Alternative Communication, Autism spectrum disorder, Facial expression recognition, Sinhala and Tamil, TensorFlow Lite.*

Acknowledgment: This research received no specific grant or external funding.

Introduction

Autism spectrum disorder (ASD) involves differences in social communication, restricted or repetitive behaviour and varied support needs. Some children have limited, delayed or unreliable speech, which makes everyday requests, choices, discomfort and social messages hard to express [1,2]. Augmentative and alternative communication (AAC) gives another route to communicate; it supports participation and does not replace speech, therapy or caregiver judgement.

The Sri Lankan setting adds barriers that general AAC products do not solve. Specialist services remain uneven and tied to larger hospitals or urban centres [3,4], so families face travel costs, few appointments and difficulty finding materials that match home routines. Mobile AAC is therefore attractive, because Android phones are far more available than dedicated speech-generating devices.

Language and cultural fit are central. Sinhala and Tamil are the official languages and English is common in education and clinical settings, so one child may need Sinhala at home, Tamil in the community and English in school. AAC research stresses participation and long-term access over symbol translation [5,6], and mobile speech-generating tools help children with ASD only when they fit the language, vocabulary and family context [7].

Table 1 compares three established products with the proposed system. Proloquo2Go is Apple-only, in four languages, at about USD 250 [8]; TD Snap installs free on iPad and Windows and covers more than 20 languages, but speech output is a paid upgrade and neither Sinhala nor Tamil is listed [9]; Avaz AAC runs on Android and iOS, with a Sri Lankan edition supporting Sinhala and Tamil for about USD 90 [10]. Offline core operation is already standard, a requirement rather than a distinguishing feature. Localised Sinhala and Tamil AAC exists commercially, but at a price many families supported through state clinics cannot meet, and none of these products offers emotion observation.

Table 1. Established AAC applications compared with the proposed system.

System	**Platform**	**Sinhala and Tamil**	**Cost to the family**	**Emotion observation**
Proloquo2Go [8]	iPad, iPhone, Mac	No	USD 250, one-time	None
TD Snap [9]	iPad, Windows	No	Free; paid speech upgrade	None
Avaz AAC, international [10]	Android, iOS	No	USD 200, one-time	None
Avaz for Sri Lanka [10]	Android, iOS	Yes	USD 90, one-time	None
Smart AAC (this work)	Android 8.0 and above	Yes, with English	No purchase cost	On-device, non-diagnostic, uncertainty-gated

Emotion-aware support could help caregivers notice when a child is uncomfortable, tired or unable to explain a feeling, but facial expression recognition (FER) demands caution. Emotion regulation matters in ASD [11], yet facial movement is not direct evidence of internal state [12], and children with ASD produce expressions more ambiguous than standard categories assume [13]. FER is therefore treated as a supportive cue only: it does not diagnose ASD, classify mental state clinically or replace human interpretation.

The contribution is an integration rather than a new algorithm. Commercial trilingual AAC now exists, so the distinguishing combination is affordability, caregiver customisation, a privacy-aware offline architecture and integrated emotion observation: Sinhala, Tamil and English AAC in one Android application at no purchase cost, vocabulary caregivers can match to a child's routines, and on-device FER that reports uncertainty rather than forcing a label. The objectives were to implement this prototype for caregiver-supervised use, integrate a mobile-safe FER model with explicit safety controls, verify the build through unit, device and release testing, and document the ethical conditions any future child-facing evaluation must satisfy.

Methods

Study design and setting

This implementation and prototype-verification study measured technical behaviour rather than clinical outcomes. No children participated, no clinical pilot ran and no sample-size calculation applied. The intended setting is caregiver-supervised use in Sri Lankan homes, schools or therapy sessions; any child-facing pilot depends on ethical clearance and health-sector approval.

System architecture

The system was built in Flutter for Android 8.0 (API level 26) and above [14], offline-first: the core workflow needs no server, user account or connectivity. Figure 1 shows the implemented structure: a presentation layer of screens; a service layer isolating local storage, connectivity, text-to-speech (TTS), sensory feedback and supportive audio; a data layer holding the trilingual vocabulary across 16 categories; and a model-asset layer bundling both TensorFlow Lite files with their labels. The solution has two applications: the main one providing structured AAC and on-device FER, and a companion for caregiver-created cards with backup, restore and sharing.

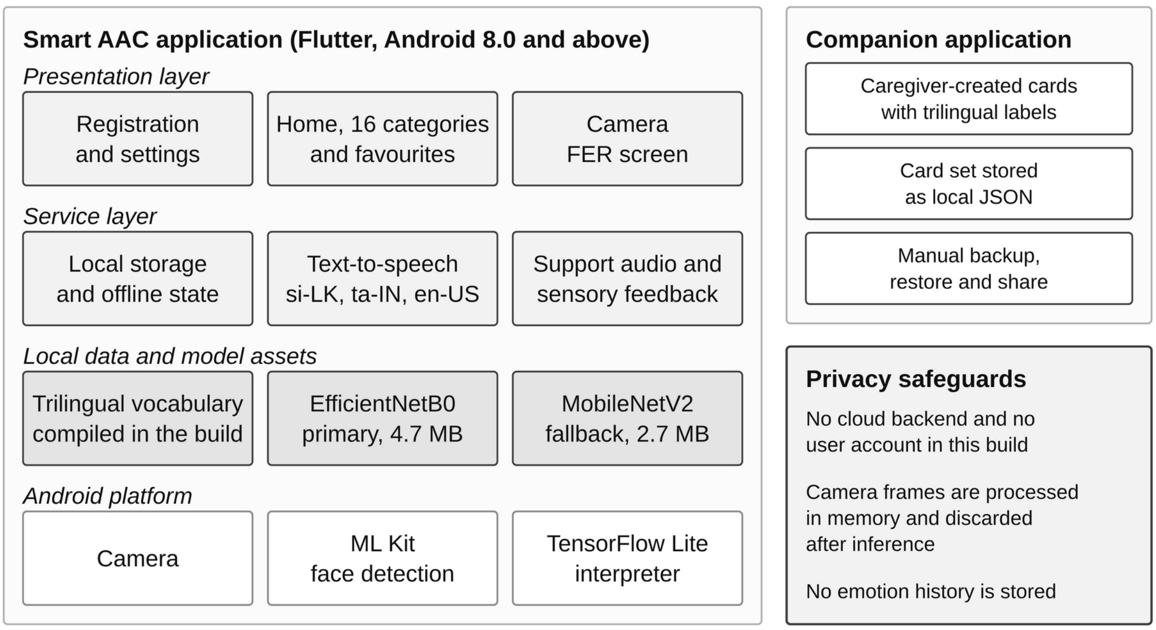

Figure 1. Implemented architecture, showing the application layers, bundled model assets and privacy safeguards.

AAC communication layer and local data management

The main application presents AAC cards in Sinhala, Tamil and English across categories such as needs, feelings, food and family, with a favourites list. TTS is requested from the installed device engine using the active language code; where no Sinhala or Tamil voice exists, the request returns silently. User state, comprising the child's name, theme, registration flag, favourites, vibration preference and model-load mode, is held in a local key-value store; the fixed vocabulary is compiled in rather than held in a database. The companion application serialises cards as local JSON and exports them in a manual backup archive. Card content and flow were refined with informal speech-therapy and stakeholder feedback.

Model strategy and training

EfficientNetB0 was chosen as the primary model for its balance of capability against mobile deployment cost [15], with MobileNetV2 as a fallback [16]. Both were trained by transfer learning on three public facial-expression datasets [17-19] with six classes: anger, fear, joy, natural, sadness and surprise. The 8,579 labelled training images were split by stratified sampling into 6,863 training and 1,716 validation images, with a separate 1,800-image test set. The distribution was unbalanced, with 3,720 joy images against 494 for fear, so class weights were applied, from 0.38 for joy to 2.90 for fear. Images were converted to RGB, resized to 224 × 224 pixels and rescaled to [-1, +1], the identical transform to the camera path. Because EfficientNetB0 with ImageNet weights expects the original pixel scale, an early run failed until a rescaling layer was added inside the model graph to convert [-1, +1] back to [0, 255], keeping one preprocessing path for both models. Its head is global average pooling, batch normalisation, dropout, a 128-unit dense layer and a six-way output, 4.2 million parameters. Fine-tuning unfroze the top 40 of the 238 base layers, giving 32 trainable layers with 49 batch-normalisation layers frozen, at a learning rate of 1e-5 with early stopping. No Sri Lankan or ASD-specific facial-expression data was collected.

TensorFlow Lite integration and model loading

Inference runs locally through TensorFlow Lite [20]. Both models export with built-in operators only and were verified before integration against the tensor contract in Table 2. At runtime the application loads the primary model and validates the interpreter tensors against that contract; if it fails it loads the fallback, and if both fail the user sees a localised error. A stored setting selects automatic, primary-only or fallback-only loading, making each path testable. The bundled models occupy about 4.7 MB and 2.7 MB, roughly 7.4 MB in total.

Safety controls for the FER feature

Figure 2 shows the decision logic of the camera feature, which runs under caregiver supervision. On-device face detection [21] in fast mode, with minimum face size 0.15, acts as a gate: the screen reports no face detected, or face not clear when the largest face spans less than 0.18 of the frame's short side. Five frames are captured at 120 ms intervals and their softmax vectors averaged, reducing the influence of one poor frame. A result is accepted only if the top-1 probability reaches a minimum confidence and the gap to the second class a minimum margin; the verified build uses 0.40 and 0.08, provisional pending real-device calibration towards 0.50–0.65. If either check fails, the screen reports an uncertain state. Raw frames are processed in memory and discarded, no emotion history is stored, and supportive audio plays only on a caregiver press. These controls address the sensitivity of FER to lighting, head angle, occlusion and expression ambiguity [22].

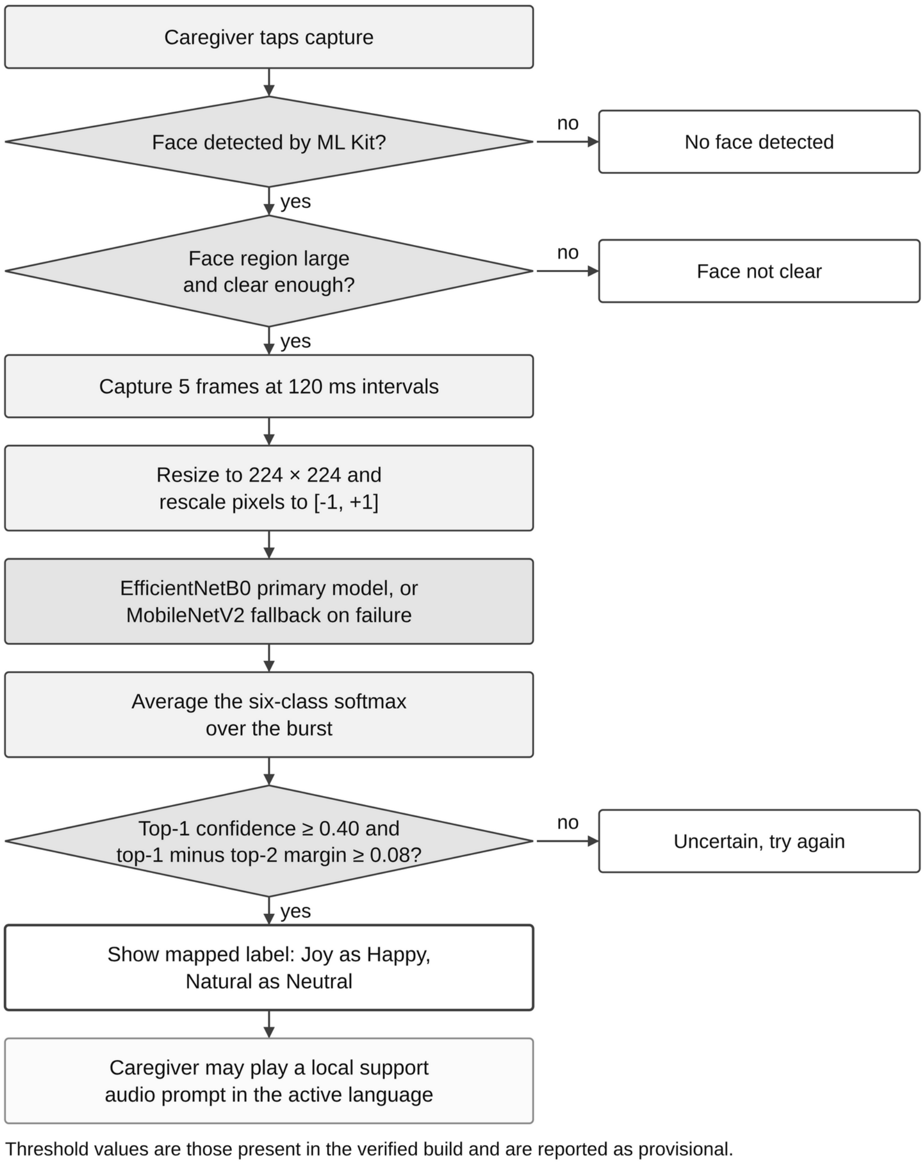

Figure 2. Decision logic of the on-device FER feature, including the gate, burst averaging, threshold checks and uncertain state.

Ethics, privacy and safeguarding

Processing stays on the device, only the minimum child information needed to operate is stored, and the build declares only camera, internet and network-state permissions with no cloud database or analytics. No identifiable child data was collected. Any future child-facing study would require written guardian consent, age-appropriate assent, and treatment of any sign of distress as withdrawal. Responsible-AI practice rests on the supportive framing, visible uncertainty states and caregiver override, so the model informs attention rather than asserting a mental state [23]. Bias remains an open risk, since the training data are public and not locally representative. Safeguarding rests on caregiver supervision, no retention of face images and anonymisation of supporting material. The prototype is not a medical device and makes no diagnostic claim; wider clinical use would require health-sector authorisation under Sri Lankan health-research guidance, the Personal Data Protection Act and international ethics principles [24-26].

Evaluation and verification

Unit tests exercised the model-load resolver, tensor-shape validator, confidence and margin gate and display mapper. Scripted walkthroughs on physical Android devices covered the full user journey, and an airplane-mode walkthrough tested offline operation. The signed release package was verified from the build output and distributed through Google Play internal testing for invited review only. Model performance was quantified on the held-out test set as accuracy with per-class precision, recall and F1-score. Early feedback came from a short anonymous questionnaire, not a standardised instrument or a clinician study.

Results

Delivered system

The prototype implements registration, trilingual card navigation across 16 categories, favourites, settings, TTS where a device voice exists, supportive audio and the camera FER screen, all offline. The companion application delivers caregiver-created cards with backup, restore and sharing.

Verification outcomes

Table 2 summarises the evidence: every unit test and scripted walkthrough passed, both models satisfied the tensor contract, and the airplane-mode walkthrough confirmed the workflow does not depend on connectivity. One quantity could not be established: inference latency.

Table 2. Implementation and verification evidence.

Verification activity	Scope	Outcome
Automated unit tests	Model resolution, tensor validation, confidence and margin gate, label mapping, storage and audio fallback	41 of 41 passed
Scripted UI walkthroughs	Registration, languages, categories, favourites, TTS, camera	10 of 10 completed
Offline operation	Airplane-mode walkthrough of the AAC workflow	Passed
Model verification	Tensor contract, operator support, dummy inference, both models	13 checks passed
TFLite export	Primary 4.73 MB, fallback 2.68 MB; input [1, 224, 224, 3] float32, output [1, 6] float32	Verified; no select operators required
Release build	Signed release package from the verified build	Verified, version 1.1.0
Distribution	Google Play internal testing, invited testers only	Active; production access under review
Platform coverage	Android 8.0 and above	Delivered; no iOS release
Real-device testing	At least two Android devices of different form factors	Completed; models and Android versions not recorded
Inference latency	Not measured in this study	Reported as a limitation

Facial expression recognition performance

Figure 3 shows the two-stage training history, Figure 4 the confusion matrix on the held-out test set of 1,800 images and Table 3 the per-class metrics. Validation accuracy reached 67.2% at the best frozen-base checkpoint and 70.5% after fine-tuning, but test accuracy was 52.9% (952 of 1,800), with weighted F1-score 0.54 and macro 0.39. Performance was strongly uneven: joy, the largest class, reached F1-score 0.84 whereas sadness reached 0.19, with 124 of its 414 test faces read as anger and 85 as natural. The fallback model reached 47.9% on the same test set. Macro one-vs-rest ROC AUC from the same evaluation probabilities was 0.82, ranging from 0.94 for joy to 0.60 for sadness. The gap between validation and test accuracy reflects the difficulty of generalising to unseen faces from limited, imbalanced public data [22]; the test figures are the performance of record.

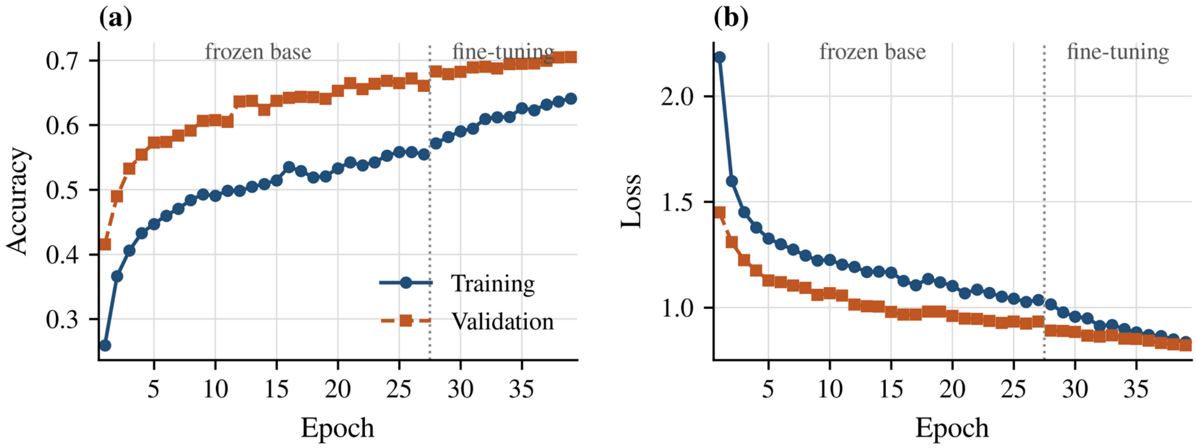

Figure 3. Training history of the primary model from the saved logs: (a) accuracy and (b) loss over 27 frozen-base and 12 fine-tuning epochs.

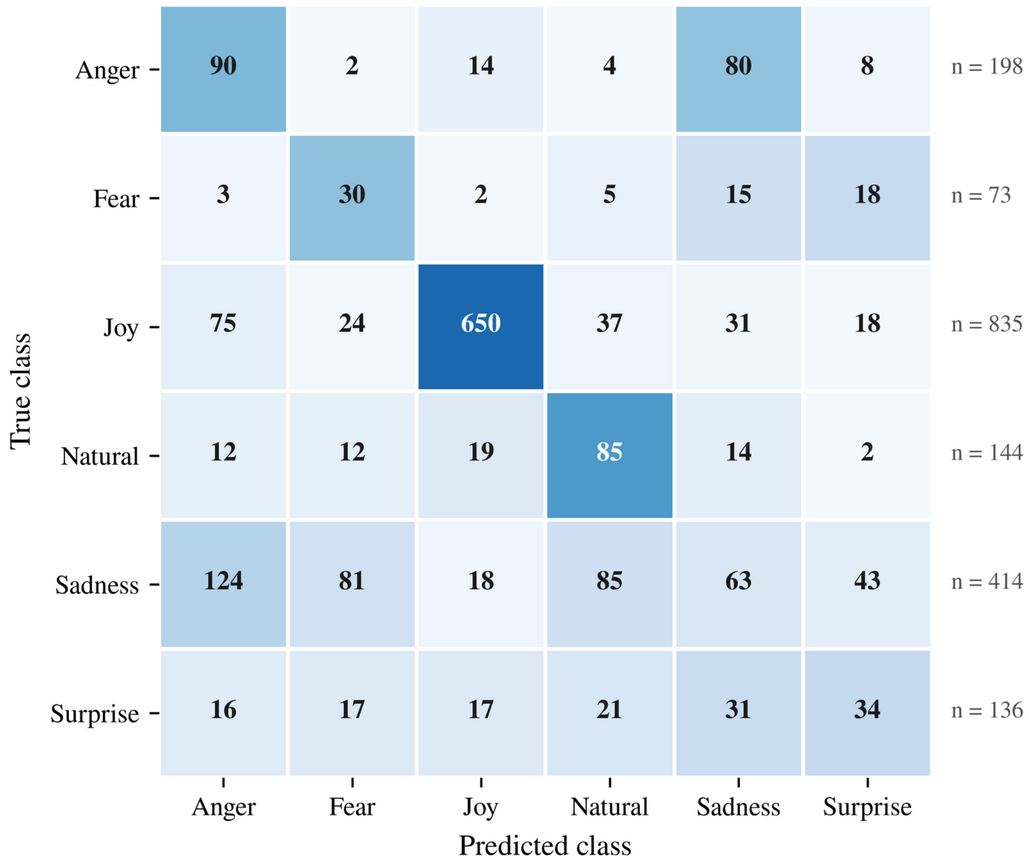

Figure 4. Confusion matrix of the primary model on the held-out test set; shading is row-normalised and n is test images per class.

Table 3. Per-class test-set performance of the primary model.

Class	Precision	Recall	F1-score	Support
Anger	0.28	0.45	0.35	198
Fear	0.18	0.41	0.25	73
Joy	0.90	0.78	0.84	835
Natural	0.36	0.59	0.45	144
Sadness	0.27	0.15	0.19	414
Surprise	0.28	0.25	0.26	136
Macro average	0.38	0.44	0.39	1800
Weighted average	0.57	0.53	0.54	1800

Early tester feedback

An anonymous questionnaire was completed by 11 invited testers; no children took part and it was not a clinician study. Nine (81.8%) judged that the application could help a child express basic needs or feelings, with one "maybe" and one "not sure". The features most often selected as useful were the communication cards (8 of 11), English support and TTS (7 each), audio prompts (6), and the camera feature and offline use (5 each). Sinhala was selected by 4 respondents and Tamil by 2, reflecting this sample's language profile rather than the value of either language. Six respondents answered a free-text question on what to improve first; those responses were not preserved and are not reported. Pragathi Centre and National Hospital Galle have provided a formal letter of collaboration making any child-facing pilot conditional on ethical clearance, which is not health-sector approval [27].

Discussion

Interpretation of model performance

Test accuracy of 52.9% is far below the 80% originally targeted and caps what the feature may claim. Joy dominated the training data and is recognised well, while sadness, the second largest test class, is recognised in only 15% of cases and read most often as anger or natural; anger, fear and surprise remain weak. This is consistent with residual class imbalance and the difficulty of separating negative-valence expressions in still images. The AUC values show the scores still rank faces usefully, which is what the confidence and margin gates rely on; they will often return an uncertain state instead of a label. That is the intended failure mode: a visible refusal is safer than a confident error, though caregivers should expect frequent refusals. The FER layer supports cautious attention-prompting only; no clinical or diagnostic interpretation is warranted.

Contribution in context

As Table 1 shows, localised Sinhala and Tamil AAC is already sold commercially, so what distinguishes this work is the combination of no purchase cost, caregiver-editable vocabulary and privacy-preserving emotion observation, not localisation alone.

Generalisability of public-dataset training

The model was trained on public datasets that are not Sri Lankan, are weighted towards adult faces and were not collected from children with ASD, whose expressions are more ambiguous than standard categories assume [13]. The per-class figures therefore describe this model on this public test set, not expected performance with Sri Lankan children. Whether it performs equitably across local skin tones, ages and presentations cannot be answered from present evidence, so revalidation on consented local data is a precondition for child-facing use.

Implementation challenges

The preprocessing mismatch described above was the largest single source of error, resolved inside the model graph. A convolutional block attention module [28] performed better in testing but could not be deployed, because its export depended on select TensorFlow operators the mobile interpreter does not load reliably; a loadable model was worth more than a higher score that would not run. On a real device, poor lighting produced false positive happy detections, motivating the burst averaging and margin check. Supportive audio was incomplete for some Sinhala and Tamil entries, and TTS availability varies by device. The release package is large at about 126 MB, distribution remains limited to the internal testing track, and long-term maintenance is unresolved: release ownership, incident response and content updates would need a named custodian.

Limitations

No children participated, no clinical pilot ran and no standardised usability instrument was used, so no claim of clinical or educational effectiveness is possible. The questionnaire drew 11 self-selected testers, adequate for early signal only. Inference latency was not measured, and although testing used at least two Android devices of different form factors, their models and versions were not recorded, so responsiveness on low-end hardware is unknown. The thresholds remain provisional, the build is Android-only, and because no emotion history is stored the system cannot support longitudinal observation.

Future validation and development

The immediate next step is technical: benchmark latency on documented low-end and mid-range devices and calibrate the thresholds against that evidence. Next, with institutional ethics approval and health-sector authorisation, consented data collection with children with ASD under supervision would allow the FER layer to be revalidated locally and audited for bias. A supervised pilot with speech therapists should then evaluate usability with standardised instruments and caregiver-reported outcomes, with the sample size planned in advance. Further development should extend vocabulary in co-design with therapists, strengthen Tamil content and audio coverage, and add iOS support. A supervised workflow in which a therapist marks a prediction reasonable, uncertain or incorrect could later support model improvement, and the descoped caregiver dashboard could be revisited, but only under ethics approval, guardian consent and health-sector authorisation. Any move towards diagnostic claims would place the system under medical-device regulation, which this project does not pursue.

Conclusions

This work produced an offline-first trilingual AAC prototype for Sri Lankan use, with caregiver-editable vocabulary and on-device emotion observation that keeps camera data on the phone and reports its own uncertainty. Technical verification was consistent: every unit test and walkthrough passed, both models met the deployment contract, and the signed release build reached invited testers. Measured recognition performance was modest and highly uneven, at 52.9% test accuracy with dependable behaviour only for joy, so the feature is offered as a cautious supportive cue under caregiver supervision and not as a diagnostic aid. Early feedback from 11 testers was encouraging but is no substitute for clinical evaluation. The prototype is best understood as a verified engineering foundation, affordable and locally relevant, that requires ethical clearance, local validation data and a supervised pilot before child-facing deployment.

Acknowledgements

The authors thank Ms. Niruni Fonseka and Mr. P. D. A. U. Akalanka for supervision, and Pragathi Centre and National Hospital Galle for their institutional interest.

References

1. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th ed. Washington (DC): American Psychiatric Publishing; 2013.

2. Beukelman DR, Light JC. Augmentative and alternative communication: supporting children and adults with complex communication needs. 5th ed. Baltimore (MD): Paul H. Brookes; 2020.

3. Perera H, Wijewardena K, Aluthwelage R, Seneviratne S, Buddhika K. Prevalence of autism spectrum disorder in Sri Lanka: a population-based study. Sri Lanka J Child Health. 2019;48(2):133-138.

4. Samad A, Razick S, De Silva MVC, Hettiarachchi S. Challenges and opportunities for autism services in Sri Lanka. J Autism Dev Disord. 2020;50(8):3023-3029.

5. Alant E, Bornman J. Augmentative and alternative communication: engagement and participation. San Diego (CA): Plural Publishing; 2021.

6. Light JC, McNaughton D. The changing face of augmentative and alternative communication: past, present, and future challenges. Augment Altern Commun. 2012;28(4):197-204.

7. Lorah ER, Parnell A, Whitby PS, Hantula D. A systematic review of tablet computers and portable media players as speech generating devices for individuals with autism spectrum disorder. J Autism Dev Disord. 2015;45(12):3792-3804.

8. AssistiveWare. Proloquo2Go: AAC app with symbols [Internet]. Amsterdam: AssistiveWare; 2026. Available from: https://www.assistiveware.com/products/proloquo2go [cited 2026 Jul 27].

9. Tobii Dynavox. TD Snap user’s manual [Internet]. Stockholm: Tobii Dynavox; 2026. Available from: https://download.mytobiidynavox.com/Snap/documents/TD_Snap_UsersManual_en-US.pdf [cited 2026 Jul 27].

10. Avaz Inc. Download Avaz apps [Internet]. Chennai: Avaz Inc; 2026. Available from: https://www.avazapp.com/download/ [cited 2026 Jul 27].

11. Mazefsky CA, Herrington J, Siegel M, Scarpa A, Maddox BB, Scahill L, et al. The role of emotion regulation in autism spectrum disorder. J Am Acad Child Adolesc Psychiatry. 2013;52(7):679-688.

12. Barrett LF, Adolphs R, Marsella S, Martinez AM, Pollak SD. Emotional expressions reconsidered: challenges to inferring emotion from human facial movements. Psychol Sci Public Interest. 2019;20(1):1-68.

13. Grossard C, Dapogny A, Cohen D, Bernheim S, Martinerie J, Janvier M, et al. Children with autism spectrum disorder produce more ambiguous and less socially meaningful facial expressions. Mol Autism. 2020;11(1):5.

14. Flutter. Flutter documentation [Internet]. Available from: https://docs.flutter.dev [cited 2025 Jan 15].

15. Tan M, Le Q. EfficientNet: rethinking model scaling for convolutional neural networks. Proc Mach Learn Res. 2019;97:6105-6114.

16. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. MobileNetV2: inverted residuals and linear bottlenecks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway (NJ): IEEE; 2018. p. 4510-4520.

17. Rajasreechaiti. FERAC dataset [Internet]. Kaggle. Available from: https://www.kaggle.com/datasets/rajasreechaiti/ferac-dataset [cited 2026 May 18].

18. Hasibur013. Autism facial emotion recognition dataset [Internet]. Kaggle. Available from: https://www.kaggle.com/datasets/hasibur013/autism-facial-emotion-recognition [cited 2026 May 18].

19. Talaat FM. Autistic children emotions dataset [Internet]. Kaggle. Available from: https://www.kaggle.com/datasets/fatmamtalaat/autistic-children-emotions-dr-fatma-m-talaat [cited 2026 May 18].

20. TensorFlow. TensorFlow Lite documentation [Internet]. Available from: https://www.tensorflow.org/lite [cited 2025 Jan 15].

21. Google. ML Kit face detection [Internet]. Mountain View: Google; 2026. Available from: https://developers.google.com/ml-kit/vision/face-detection [cited 2026 Jul 27].

22. Li S, Deng W. Deep facial expression recognition: a survey. IEEE Trans Affect Comput. 2022;13(3):1195-1215.

23. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. Nat Mach Intell. 2019;1(9):389-399.

24. Ministry of Health, Sri Lanka. National guideline for ethics review of health research in Sri Lanka. Colombo: Ministry of Health; 2020.

25. Parliament of the Democratic Socialist Republic of Sri Lanka. Personal Data Protection Act, No. 9 of 2022. Colombo: Department of Government Printing; 2022.

26. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA. 2013;310(20):2191-2194.

27. Pragathi Centre, National Hospital Galle. Letter of institutional interest in the Pragathi AAC application. Galle: National Hospital Galle; 2026 Apr 7. Unpublished document.

28. Woo S, Park J, Lee JY, Kweon IS. CBAM: convolutional block attention module. In: Computer Vision - ECCV 2018. Lecture Notes in Computer Science, vol. 11211. Cham: Springer; 2018. p. 3-19.