RESEARCH ARTICLE

OPEN ACCESS

An Edge AI-Vision Model for Elderly Social Assistance Robots Using Edge Impulse

Luke Liu
The Governor's School for Science and Technology, Hampton, Virginia, United States
Email: luke.liu1245@gmail.com

Moses Garuba, Ph.D., LL.M
Department of Electrical and Computer Engineering, Hampton University, Hampton, Virginia, United States
Email: moses.garuba@hamptonu.edu

----------**********----------

Abstract: **Social robots are increasingly deployed in elderly care to provide companionship, emotional support, and assistive monitoring. AI-based vision models have enabled these robots to detect and interpret facial features in support of personalized engagement, but many current systems rely on cloud-based architectures that raise privacy and ethical concerns around the transmission of sensitive biometric data. Beyond interception risk, algorithmic bias stemming from insufficient dataset diversity may produce inferences that perform unequally across demographic groups. This study proposes Edge AI vision as a privacy-preserving framework for elderly care robotics. An Edge AI vision model was developed using Edge Impulse and trained to identify the eyes and mouth of male and female elderly subjects aged 55–94. The model achieved classification accuracy of 98.04% on male subjects and 99.51% on female subjects, with no statistically significant difference between groups ($z = 1.43$, $p = 0.15$), suggesting that edge-based inference can deliver equitable and privacy-preserving vision capabilities for elderly care robotics.**

Keywords — edge artificial intelligence, social robotics, elderly care, AI vision, Edge Impulse, FOMO, MobileNetV2, facial feature detection, algorithmic fairness, privacy.

----------**********----------

I. BACKGROUND

Care for the elderly has always relied on some form of support, from family assistance in the earliest human societies to mechanical aids and, later, electronic systems designed to maintain independence and safety. Walking canes have been used since approximately 4000 BCE [1], and as technology evolved, devices for promoting elderly independence and empowerment evolved alongside them. Following the technological revolution, assistive technology shifted toward comfort and emotional grounding through wheelchairs, music boxes, and wind-up toys [1] that appealed to both children and elderly users. By the late 20th century, electric mobility systems enabled communication between subjects and caregivers [1]. It is worth noting that in the historical context of the 19th and 20th centuries, elderly caretakers were both more likely to be immediate family of the elderly subject and more likely to be able to dedicate substantial time to the elderly subject — a configuration that has largely dissolved in the present day.

The demographic landscape that frames this evolution has changed dramatically. The proportion of adults aged 65 and older in the United States rose from roughly 8% in 1950 to approximately 19% in 2026, and is projected to continue climbing toward 22–23% by 2050 [19]. Combined with declining birth rates and increasingly dispersed families, this has altered who provides elderly care: fewer family members are available as informal caretakers, while the formal care workforce faces persistent shortages [20], [21]. International estimates suggest a shortfall of millions of healthcare workers globally by 2030, with elderly care among the hardest-hit categories. These pressures have driven both industry and academia to look toward intelligent assistive technologies as part of a broader response to the long-term care gap.

Technological advances in elderly care have therefore shifted toward companionship. Early companionship robots were developed specifically for dementia patients — PARO (the personal seal robot), for example, improves sociability in some elderly dementia subjects and increases positive factors such as medicine retention [2]. In the past decade, social robots have become widely used to promote the mental and physical well-being of elderly residents in care homes. All such robots share some element of personalization in how they interact with subjects; they differ in the form that interaction takes [3]. Modern social assistive robots can be loosely grouped into three overlapping categories based on the form of interaction: zoomorphic companion robots that mimic the appearance and behavior of an animal; humanoid or semi-humanoid robots designed for conversational engagement; and task-oriented assistive robots that monitor activities of daily living, deliver medication reminders, or alert caregivers in the event of a fall. Most deployed systems span more than one category, and the boundary between them is blurred by the integration of artificial intelligence into the perception, dialogue, and decision-making layers.

AI vision has also become increasingly utilized in elderly care robotics. Transfer learning and lightweight neural network architectures have enabled vision models to run efficiently on resource-constrained edge devices [4]. These models can detect and interpret facial features in real time, allowing robots to

elderly subjects. The integration of AI vision into social robots therefore represents a significant step toward more responsive, personalized, and privacy-conscious elderly care systems. However, integrating vision capabilities into devices that operate in private living spaces — bedrooms, bathrooms, and other sensitive environments — surfaces a distinct set of ethical and architectural questions that this study engages with directly.

II. INTRODUCTION

A. Interaction Types

Each social interaction robot has some specific element regarding how it interacts with the elderly patient. PARO has been shown to promote a small but substantial positive reaction from patients, including a standardized mean difference (SMD) of −0.63 for medication adherence [2]; models like it use uniform outputs across subjects. Other models such as Andromeda Robotics' Abi (Figure 1) use AI-driven interaction to encourage more personalized engagement [5].

Fig. 1 Abi assistive robot for elderly care developed by Andromeda Robotics.

The Abi model, integrated with OpenAI's ChatGPT-4, has been recorded to recognize residents, answer questions, and conduct conversations based on previous interactions [5]. Residents above the age of 55 expect initially to have positive benefits from companion robots [6], showing that AI used in conjunction with caregivers has the potential to aid in general caretaking without aggregating subjects. Elderly individuals are most accepting of AI-based robots that function as reminder tools, with improved adherence to daily activities (taking medication, maintaining personal hygiene, contacting family) when prompted by an AI-based robot [3]. Non-invasive social robots therefore present a promising application of artificial intelligence in elder care by supporting independence while minimizing the ethical and psychological concerns associated with more intrusive technologies.

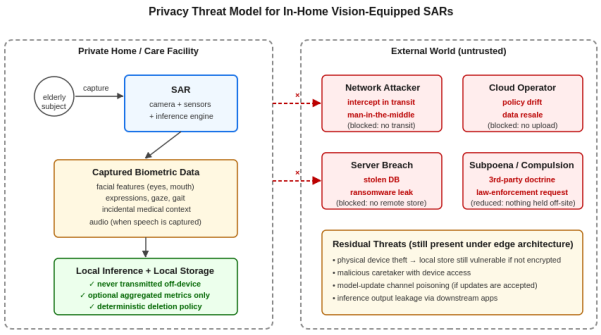
The personalization dimension of these systems is particularly important. Standardized, non-adaptive interactions can produce a meaningful but limited benefit. Adaptive systems that adjust their responses based on the user's identity, history, or current emotional state have the potential to deliver substantially deeper engagement, but they do so by collecting and processing data that would otherwise never have been broader data collection is the central design tension in contemporary social robotics, and it is the tension that motivates the architectural choices examined in this paper.

Results are context-specific and apply only to the spheres in which they were conducted, such as elderly care homes or distinct cultural environments [3], [6]–[8]. This limitation is especially relevant for studies in China, Taiwan, or Japan, where cultural attitudes toward aging and technology may differ significantly from other regions. Comparative work has shown, for example, that acceptance of humanoid robots is notably higher in several East Asian populations than in North American or European samples, in part because of differing cultural framings of automation. Health conditions, comorbidities, and individual factors can also drastically influence results — particularly for dementia and similar neurological diseases, where the approach must account for cognitive decline, memory impairment, and varying levels of autonomy. Most robots used to monitor dementia subjects therefore lack true anthropomorphic qualities and collect little to no health data, since these robots function alongside caretakers [7], [9].

B. Privacy and Ethical Concerns

There are considerable privacy risks intrinsically tied to robot Internet of Things (IoT) data storage. Health data collection and transfer to various databases raise numerous concerns [10], and the extent to which some risks can be conceded for the sake of security gain is unsettled [10]. Social robots that appear more zoomorphic and use standard interactions devoid of AI decision-making have been reported to reduce stress and ease communication for some subjects [2], [9], [11], [12].

The privacy threat model for a vision-equipped social assistive robot differs from that of a typical smart-home device. A camera-equipped robot observes the user across most waking hours, in spaces where the expectation of privacy is highest. The data it captures — facial expressions, gaze, gait, and incidental views of medical interventions or personal hygiene — is among the most sensitive any sensor in the home can collect. When transmitted to a cloud service, it traverses network infrastructure and rests on third-party servers, each an additional surface for interception, accidental disclosure, or subpoena. Figure 2 summarizes this threat model and indicates which classes of risk are eliminated, reduced, or left untouched by moving to an edge-based architecture.



robots. An edge-based architecture eliminates several major risk surfaces while leaving a smaller residual set that must be addressed at the device level.

Training AI vision models for elderly care introduces additional concerns beyond privacy. Insufficient dataset diversity, particularly the under-representation of elderly subjects across demographics, can produce biased or unreliable models. Class imbalance and mislabeling during annotation may further degrade performance, leading to inconsistent facial feature detection — with meaningful consequences in healthcare contexts.

Algorithmic bias has been documented in facial recognition systems broadly. The influential Gender Shades audit of commercial gender-classification systems found error rates as much as 34 percentage points higher for darker-skinned women than for lighter-skinned men [22]. Subsequent work using the Berkeley DeepDrive (BDD) dataset, an annotated image dataset developed for object detection, found that person-detection models performed significantly worse for individuals with darker skin tones [13]. Both findings point to the same root cause: training corpora that fail to represent the populations the deployed model will eventually be applied to. This poses a particular concern in elderly care settings where consistent and equitable performance is essential to user safety and dignity. Older adults are themselves underrepresented in many widely used computer-vision benchmarks, and the visual characteristics of aging — wrinkles, changes in skin pigmentation, prosthetics, eyewear, and altered facial musculature — interact with model training in ways that are still imperfectly understood. A vision system that performs well on a benchmark of working-age adults cannot be assumed to perform comparably on a population aged 70 and above without explicit evaluation.

Regarding AI-integrated elderly social assistive robots (SARs), vision-based tracking and monitoring has been reported to be unpopular for certain subjects who reside in elderly homes [3]. Sensitive data collected by vision-based SARs is also deemed confidential under the Health Insurance Portability and Accountability Act (HIPAA). HIPAA's privacy and security rules establish minimum standards for the handling of protected health information by covered entities and their business associates, and biometric data captured by an in-home robot will frequently fall within the scope of these protections when the device is deployed in clinical or post-acute care environments. Parallel legal regimes in other jurisdictions — notably the EU GDPR and several state-level biometric privacy laws within the United States such as Illinois' BIPA — impose comparable or in some respects stricter requirements, including explicit consent for the processing of biometric identifiers and obligations to minimize the data retained [14].

Beyond the legal floor, ethical considerations extend to the question of meaningful consent. Many elderly users have not grown up with always-on networked sensors and may have only a partial understanding of how cloud-based vision systems handle their data. Cognitively impaired users — a population that overlaps heavily with the population most likely to benefit comprehending and authorizing data collection. These considerations suggest that, wherever feasible, the privacy properties of an assistive robot should be a function of its architecture rather than of disclosures users must read and assent to.

C. Edge AI-Vision

Two methods have been identified for storing the data collected by AI-based SARs: cloud-based and edge-based storage [15]. Edge AI uses AI-enabled IoT devices deployed on-site to process and store data locally, thereby earning the designation "edge AI". Edge AI has been proven to provide robust protection against data breaches due to its inherent architectural nature [16], and the utilization of edge-based IoT storage devices can help prevent potential data leakage [17]. There is, however, a significant cost differentiation between edge AI and cloud AI. Edge AI IoT devices can be customized for specific tasks on-site, often resulting in mixed upfront hardware costs but reduced long-term expenses related to data transmission, cloud storage, and latency [15], [17]. This level of localization allows the caretaker to easily create an intuitive edge-based SAR using on-device processing and locally stored data, enabling responsive interactions while reducing reliance on external cloud infrastructure. Figure 3 compares the two pipelines.

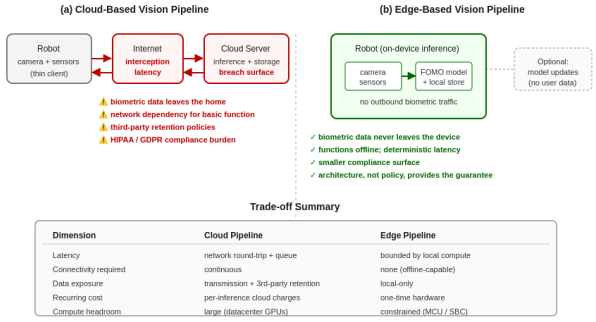


Fig. 3 Cloud-based vs. edge-based vision pipelines. The edge pipeline keeps biometric data on-device and functions offline, at the cost of operating within tighter compute and memory budgets.

The architectural distinction has practical consequences that extend well beyond data security. Cloud pipelines introduce network latency that scales with connection quality and server load; for an assistive robot whose value depends on fluid, real-time interaction, even a few hundred milliseconds of round-trip delay can degrade the perceived naturalness of the system. Cloud inference also requires continuous network connectivity, meaning that any interruption in service — temporary outages, network maintenance, or a household losing internet access — degrades or disables the robot's functionality. Edge inference avoids both of these failure modes: the model executes locally with deterministic latency bounded by the device's compute capacity, and the system can continue to operate in fully disconnected environments. The emergence of standardized benchmarks for ultra-low-power on-device inference, such as MLPerf Tiny [23], has made it possible to characterize the performance of microcontroller-class platforms in a way that

risk associated with committing to an edge-first design.

Edge AI vision's effectiveness is not solely determined by its architectural advantages in privacy and latency, but is fundamentally contingent on the fairness and representativeness of the model underlying it. A model deployed on an edge device that has been trained on a non-representative dataset will reproduce and perpetuate the same demographic biases regardless of where inference takes place [13], [22]. Moving inference from the cloud to the edge does nothing, by itself, to address bias baked into the model weights. This is of particular concern in elderly care settings, where vision models must perform equitably across diverse subjects varying in age, gender, and skin tone in order to deliver reliable and dignified assistive interactions. The privacy guarantee provided by edge architecture and the fairness guarantee provided by a representative training process are therefore complementary requirements: each is necessary, neither is sufficient on its own.

D. Research Goals and Objectives

This study addresses two primary objectives. First, to develop and evaluate an Edge AI vision model capable of reliably detecting facial features — specifically the eyes and mouth — in elderly subjects aged 55–94 using the Edge Impulse platform. Second, to assess whether the model performs equitably across male and female subjects by determining whether any statistically significant disparity exists in classification accuracy.

The eyes and mouth were chosen because they are the dominant carriers of facial expression in nearly all standard taxonomies of emotion recognition, including those derived from the Facial Action Coding System (FACS). A model that can robustly localize these features across the visual variability characteristic of older adult faces is positioned to serve as a foundation for emotion-recognition, attention-tracking, and engagement-monitoring modules in a fielded SAR.

III. MATERIALS AND METHODS

A. Edge Impulse

Edge Impulse — an industry-standard platform for developing, managing, and deploying machine learning models on edge devices — was used to develop the AI vision model. The model was trained on the FOMO (Faster Objects, More Objects) MobileNetV2 0.35 algorithm with a training pool of 315 male and 315 female subjects. Edge Impulse allowed efficient labeling, training, testing, and deployment within a single integrated platform. The platform streamlined the overall development process by providing tools for dataset management, automated preprocessing, training, and performance evaluation. Through Edge Impulse, images could be uploaded directly into the training environment, where annotations and labels were applied to each image sample. The platform also enabled rapid iteration during model development by allowing training parameters, learning rates, and model architectures to be modified and tested efficiently.

visualizing precision, recall, and F1 score across training iterations, allowing performance comparisons between training runs across both groups. The platform supports lightweight edge-optimized deployment, making the trained model computationally efficient while maintaining accurate feature detection. Edge Impulse was selected over alternative frameworks for three reasons. First, its native support for FOMO and constrained-device architectures meant the trained model could be deployed to a microcontroller-class device without further re-engineering of the inference pipeline. Second, its integrated labeling, training, and evaluation tooling reduces inconsistencies between training and inference preprocessing. Third, its automatic export of training metrics simplified the longitudinal monitoring of model convergence reported in the Results section.

B. The FOMO MobileNetV2 0.35 Architecture

FOMO is a deep-learning architecture developed by Edge Impulse for object detection on resource-constrained devices. It combines an image feature extractor — by default, a truncated MobileNetV2 [24] — with a fully convolutional classifier that reframes object detection as classification over a spatial grid. Rather than emitting bounding boxes, FOMO predicts a centroid heatmap in which each grid cell is classified as either background or as containing the center of an object. Figure 4 summarizes the architecture.

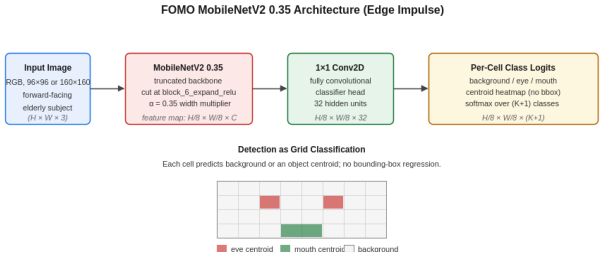


Fig. 4 FOMO MobileNetV2 0.35 architecture. A truncated MobileNetV2 backbone (α = 0.35) feeds a 1×1 convolutional classifier head that emits per-cell class logits over an 8×-reduced spatial grid.

FOMO uses MobileNetV2 as its trunk and by default performs a 1/8 spatial reduction from input to output, achieved by cutting MobileNetV2 at an intermediate expansion layer (block_6_expand_relu in the standard Edge Impulse configuration). The 0.35 alpha scales the number of channels by 0.35 relative to the full MobileNetV2, dramatically reducing parameter count and memory footprint at a moderate cost in expressive capacity. MobileNetV2 itself uses inverted-residual blocks with linear bottlenecks [24] to achieve a favorable accuracy-to-parameter ratio; combined with the lightweight FOMO classifier head, this produces a model that has been reported to run at approximately 30 frames per second on a Cortex-M7 microcontroller with on the order of 240 KB of RAM available — the regime that MLPerf Tiny was designed to characterize [23].

Because FOMO classifies grid cells rather than regressing bounding boxes, it does not produce object size estimates. For the present study this is an acceptable trade-off: the goal is to localize the eyes and mouth of an elderly subject in front of the

sufficient to anchor subsequent processing stages such as facial expression recognition or gaze estimation. The architectural compromise that FOMO accepts — losing per-object size in exchange for radically lower compute and memory requirements — is precisely the compromise that makes on-device inference feasible at the kind of price point and power budget that a fielded elderly care robot would have to operate within.

C. AGFW Dataset

The dataset was taken from the Aging Faces in the Wild (AGFW) database, a public database containing still, forward-facing images of men and women aged 10–94 [18]. A total of 1,040 images were taken (519 men and 521 women), split into a roughly 60/40 train/test split, with a training pool of 315 for each group and a testing pool of 206 women and 204 men. The AGFW database was selected because it contains a diverse collection of facial images with variations in age, facial structure, lighting conditions, image quality, and minor pose differences. Only elderly subjects aged 55–94 were included to ensure consistency with the research objective of analyzing feature detection accuracy in older adults. The 55–94 age range is broad enough to capture the heterogeneity of late-life facial morphology — including the changes in skin texture, eyelid drooping, and bone-structure prominence that often differentiate older from younger adult faces — while remaining narrow enough that the resulting model can be characterized as an elderly-faces model rather than a general adult-faces model.

Prior to training, all images were reviewed to confirm that the subjects' eyes and mouth were visible and that the image quality was sufficient for annotation and model analysis. Images with severe obstructions, excessive blur, or incomplete facial visibility were excluded to reduce potential inaccuracies during training and testing. The balanced distribution between male and female subjects was maintained to minimize gender bias within the dataset and to allow the model to evaluate facial features consistently across both groups. Maintaining numerical parity between the male and female subsets was a deliberate methodological choice, because the central equitability analysis described later in this paper requires that any disparity in measured accuracy not be attributable to a corresponding imbalance in the training distribution.

The AGFW database has limitations that bear acknowledgment. Like most publicly available face databases, it is not perfectly balanced across racial and ethnic groups, and the demographic distribution of its elderly subset reflects whatever distribution was sampled at the time of compilation. The present study examines equitability along the male/female axis specifically because this axis is well-supported by the dataset; equitability across other demographic axes — including skin tone, racial background, and the presence or absence of facial accessories such as eyeglasses — remains an open question that the Limitations section addresses in greater detail.

D. Labeling and Training Procedure

mouth (Figures 5 and 6). Labeling used Edge Impulse's AI Labeling feature, which leverages a zero-shot object detector based on OWL-ViT to automatically draw bounding boxes; every annotation was manually checked and corrected to address mislabeled boxes or missing labels. The hybrid pipeline balanced throughput against label quality: pure manual annotation of more than a thousand face images, each requiring three bounding boxes (two eyes and a mouth) plus a class label, would have been prohibitively time-consuming, while a fully automated pipeline without human review would have propagated systematic errors from the zero-shot detector into the training set.

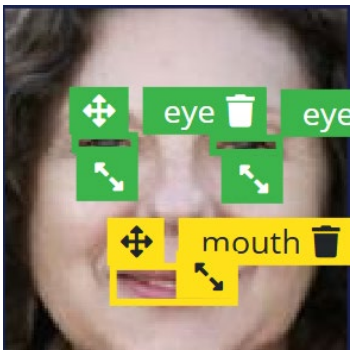


Fig. 5 Female sample from the AGFW database with eye and mouth annotations.

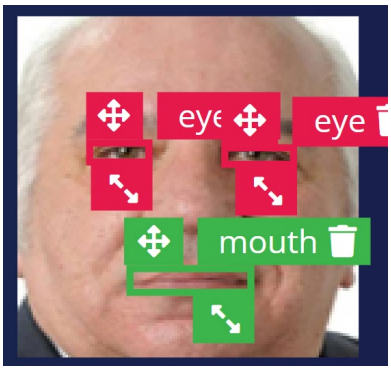


Fig. 6 Male sample from the AGFW database with eye and mouth annotations.

FOMO was selected because it is specifically designed for lightweight object detection tasks on edge-based systems while maintaining computational efficiency and fast inference speeds. The selection criteria reflected both the constraints of the target deployment environment — a microcontroller-class or small single-board computer attached to a social assistive robot — and the nature of the detection task itself, which does not require precise bounding-box dimensions and is well-served by the centroid-classification output that FOMO produces.

Training was conducted in the Edge Impulse cloud environment with cross-entropy loss and the Adam optimizer. Epochs were selected by monitoring loss curves for both subsets until they had visibly converged and ceased to show further meaningful improvement. Hyperparameters were held constant between the male and female subsets so that any

attributed to dataset properties rather than to differential model configurations. Figure 7 summarizes the complete pipeline, from dataset selection through final equitability testing.

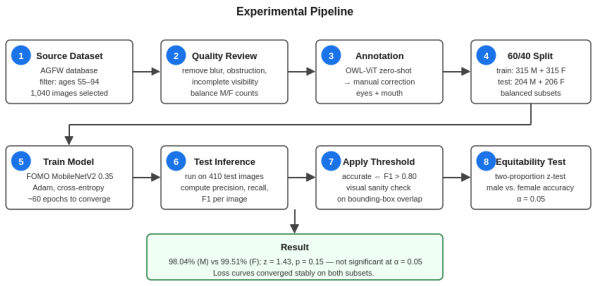


Fig. 7 End-to-end experimental pipeline. Steps 1–4 prepare the data, steps 5–7 train and score the model, and step 8 performs the equitability test.

E. Analysis Procedure

After training, the model was validated on its ability to identify labeled eyes and mouths in the testing dataset of 206 female and 204 male subjects, and a per-image accuracy percentage was calculated. An "accurate" identification was defined as any image in which the facial features were classified with an F1 score — the harmonic mean of precision and recall — greater than 80%. Predictions were compared against the manually verified ground-truth annotations created during the labeling phase.

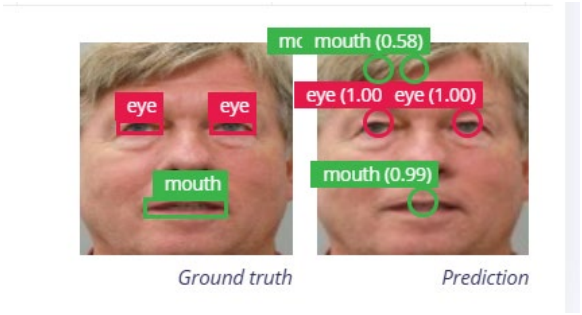


Fig. 8 Actual inference output (right) vs. ground-truth annotations (left). The model outputs per-feature confidence scores; here both eye centroids are detected at confidence 1.00 and the mouth at 0.99.

During the evaluation process, the trained model processed each testing image individually and generated predicted bounding boxes for the labeled facial features. Precision values were used to determine the proportion of correctly identified facial features relative to all predicted features, while recall values measured the proportion of correctly detected features relative to the total number of actual labeled features present in the image. The F1 score combined both precision and recall into a single metric, allowing overall model performance to be evaluated more effectively.

The 80% F1 threshold was chosen to provide a clear, interpretable accept/reject criterion for whether a given inference would be acceptable in a downstream consumer-facing system. The harmonic-mean structure of F1 means the threshold cannot be cleared by a model with high precision through under-prediction or high recall through indiscriminate prediction; both must be reasonably high simultaneously. An evidence that the model had localized the relevant facial features with enough fidelity to be useful for downstream tasks. In addition to numerical evaluation, prediction outputs were visually inspected to confirm that bounding boxes aligned with the eyes and mouth of each subject, catching failure modes that pure metrics can mask.

To formally evaluate the equitability question, a two-proportion z-test was used to compare the rate of accurate (F1 > 0.80) inferences across the male and female testing subsets. The test was selected because each subset's outcome can be naturally expressed as a proportion (accurate vs. inaccurate inferences), and the null hypothesis — that the true accuracy rate is equal across groups — is exactly the hypothesis the two-proportion z-test is designed to evaluate. A significance threshold of $\alpha = 0.05$ was used, consistent with conventional practice in applied machine-learning fairness evaluation.

IV. RESULTS

A. Equitability Results

A two-proportion z-test was conducted to compare classification accuracy between male and female subjects. Of the 204 male inferences, 200 facial feature detections achieved F1 > 0.80, while 205 out of 206 female inferences did. Figure 9 summarizes the per-group accuracy rates with 95% Wilson confidence intervals.

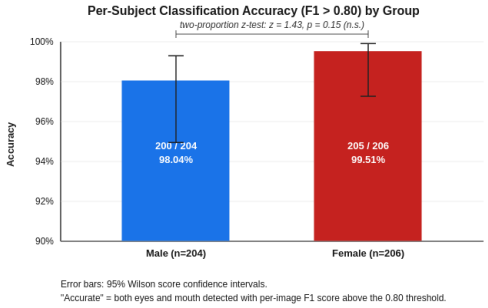


Fig. 9 Per-subject classification accuracy by demographic group. Error bars are 95% Wilson score CIs. The two-proportion z-test yields $z = 1.43$, $p = 0.15$, which does not reach significance at $\alpha = 0.05$.

Classification accuracy for female subjects (99.51%) was slightly higher than for male subjects (98.04%). However, the difference was not statistically significant ($z = 1.43$, $p = 0.15$) at $\alpha = 0.05$. These results suggest that the AI vision model performed consistently across both demographic groups and did not exhibit significant gender-based disparities. The model exhibited high classification performance for both groups.

Several aspects of this result merit comment. First, the absolute accuracy levels — over 98% in each case — are comparable to or better than reported figures for general-purpose face detectors evaluated on adult populations, despite the model being constrained to run on resource-limited devices and trained on an underrepresented population (55–94 years old). The lightweight FOMO MobileNetV2 0.35 configuration is not, on its own, a limiting factor for usable performance on elderly faces.

higher accuracy on female subjects — is the opposite of what is most commonly reported in the broader face-recognition literature, which has historically documented worse performance on women [22]. The fact that the gap is small, non-significant, and reversed provides modest additional evidence that the male/female axis is not a major source of performance disparity, though it says nothing about axes that were not measured.

Third, the failure cases (four male, one female) were examined qualitatively. Figure 10 reports the breakdown and dominant failure modes. The most common male failure mode involved subjects whose eyes were partially occluded by heavy eyeglasses frames or hair, leading the model to predict only one eye or a single region spanning both. The single female failure case involved a partially obscured mouth. These failure modes are consistent with FOMO's coarse spatial reduction, which can struggle when two instances of the same class fall close together.

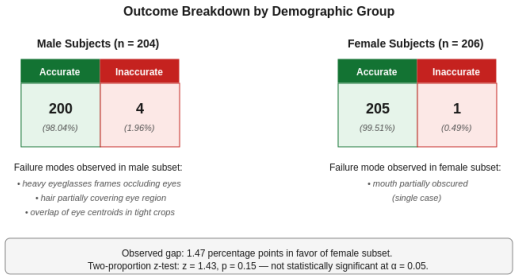


Fig. 10 Outcome breakdown by group, with qualitative summary of dominant failure modes in each subset.

B. Accuracy Results

Epoch loss was analyzed to evaluate model convergence. Loss was measured using Cross-Entropy Loss, which quantifies the difference between predicted outputs and ground-truth labels; lower values indicate better convergence.

The epoch loss curves, visualized in TensorBoard for both training instances (Figure 11), demonstrated a consistent downward trend, indicating stable learning. Both curves exhibited the characteristic shape of a healthy training run: a steep initial drop as the model learns the broad structure of the task, followed by a gradual asymptotic decline. Neither curve exhibited divergent or oscillating behavior.

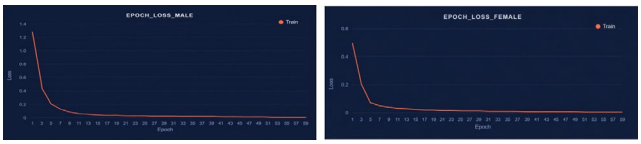


Fig. 11 Epoch loss curves for male (left) and female (right) training instances. Both curves reach near-zero loss within roughly twenty epochs. The female curve begins at a lower initial loss; both converge to comparable terminal values.

The male curve began at approximately 1.4 and decayed to near zero within twenty epochs, with the remaining forty epochs contributing only marginal improvement. The female curve began at a substantially lower initial loss — approximately 0.6 in the first epoch — and converged in a subset reflects a happenstance of the random initialization and early-epoch dynamics rather than any structural difference between the two subsets; both curves reach comparable terminal loss values, which is what matters for the downstream accuracy comparison. The fact that both curves converge to similar terminal values is itself a check on the equitability claim: if the model were systematically struggling to fit one of the two subsets, that struggle would typically register as a higher terminal loss on that subset, which it did not.

Taken together, the convergence behavior of the loss curves and the high terminal accuracy on the held-out test set support the conclusion that the elderly eye and mouth object detection models developed in Edge Impulse are accurate and can be used in future applications in elderly care. The trained model meets the technical objective of reliably localizing the relevant facial features across the elderly age range examined, and it meets the equitability objective of doing so without statistically significant disparity between male and female subjects.

V. DISCUSSION

A. Technical Feasibility of Edge-Based Elderly Vision

The headline finding is that a lightweight, microcontroller-deployable object detection architecture — FOMO with a MobileNetV2 0.35 backbone — can achieve facial feature detection accuracy in excess of 98% on elderly faces aged 55–94 when trained on a curated subset of AGFW. This has practical implications for SARs that must operate within the power, thermal, and cost envelopes of consumer hardware. A model of this size can plausibly run on a single-board computer or high-end microcontroller with no GPU acceleration, eliminating the recurring cost of cloud inference and the engineering complexity of maintaining a stable network link between a robot and a remote inference server.

The result also informs the broader question of whether the apparent trade-off between model size and accuracy is as sharp as is sometimes suggested. For the narrowly scoped task of localizing the eyes and mouth of a clearly visible elderly face, the trade-off appears to be quite forgiving: a model approximately one or two orders of magnitude smaller than the kinds of detection networks typically used in cloud-based pipelines reaches accuracy comparable to those larger systems on the population of interest. Broader detection tasks — multi-face localization in crowded scenes, fine-grained expression classification, or detection under extreme occlusion — would likely require a larger model or additional architectural support, and the favorable performance reported here should not be over-generalized to those settings.

The applicability of this finding to elderly care robotics specifically is reinforced by the fact that the most common deployment scenario — a robot stationed in front of, or moving slowly around, a seated or standing user — naturally produces the kind of frontal, well-lit imagery that the model was trained on. The mismatch between training distribution and deployment distribution that often degrades model performance in real-world use is, in this application,

characterized empirically before any system based on this work is fielded.

B. Implications for Fairness and Equitability

The equitability analysis returned a non-significant difference between male and female accuracy ($z = 1.43$, $p = 0.15$). This is the desired outcome from a fairness perspective, but the claim it supports is narrow: there is no detectable accuracy disparity along the male/female axis given the present sample sizes and the present threshold of statistical significance. It does not establish that the model is unbiased in any stronger sense, and it does not extend to demographic axes that were not measured.

Equitability across skin tone, ethnicity, the presence of medical accessories (eyeglasses, oxygen tubing, hearing aids), and the presence of facial hair are all relevant axes along which the model's behavior should be characterized before deployment in a real care setting. The Gender Shades audit demonstrated that an apparent overall accuracy figure can conceal severe intersectional disparities once results are disaggregated by skin tone and gender simultaneously [22]; analogous intersectional audits along age × skin tone or age × ethnicity may surface gaps that the gender-only analysis here would not reveal. AGFW does not provide the demographic metadata required for such an analysis, and dedicated dataset construction is likely necessary for a full fairness audit. The present study should be read as evidence that the proposed approach is not obviously unfair along one important axis, while remaining agnostic about others.

It is also worth noting that algorithmic fairness in a healthcare-adjacent context has dimensions beyond the kinds of per-image accuracy comparisons performed here. Equity of access — whether the device is affordable, available, and culturally acceptable to all the populations it is designed to serve — is at least as important as equity of inference. Edge-deployed inference contributes to access equity in that it removes the recurring cost of cloud inference and the requirement of broadband connectivity, both of which can be barriers in rural, low-income, or institutionally underserved settings.

C. Privacy Architecture as a First-Class Design Constraint

The motivation for on-device inference was framed in this paper as a privacy property. The experimental work demonstrated that this property does not come at the cost of accuracy on the target population. That conjunction — privacy-preserving by architecture and accurate in practice — is what makes edge AI vision a credible architectural choice for elderly SARs.

Privacy properties that follow from architecture are systematically more reliable than those that follow from policy. A robot that physically does not transmit image data outside the home cannot leak that data through a server-side breach, regardless of the operator's practices, regulatory changes, or whether users read the privacy disclosures. This is the same principle that motivates client-side encryption and on-device biometric matching in mobile phones; applying it to in-home class of devices that today still largely depends on cloud-side guarantees. The residual-threat panel in Figure 2 enumerates risks that remain even under an edge-only architecture — physical device theft, malicious caretakers, model-update channel compromise, downstream output leakage — each of which requires its own mitigation but is independent of where inference runs.

VI. LIMITATIONS AND FUTURE STEPS

This study proposed utilizing Edge AI vision to collect and process data on elderly subjects in a safe and efficient manner. Storage and efficiency metrics regarding inference time, peak RAM usage, and flash usage also need to be characterized in dedicated follow-up work. The present study established that the trained model achieves the accuracy and equitability targets that motivated its development; characterizing its runtime behavior on the specific target hardware — measured in milliseconds per inference, in peak RAM consumed during inference, and in flash consumed by the deployed model artifact — is the natural next step toward a fielded system. Edge Impulse exposes these measurements directly as part of its deployment tooling, and a downstream paper should report them across the candidate target devices (Cortex-M7 microcontrollers, Cortex-A class single-board computers, and small embedded GPUs) that would plausibly be used to host the vision subsystem of an elderly care robot. Standardized benchmarks such as MLPerf Tiny [23] provide an established methodology for reporting these figures in a way that can be compared across hardware platforms and across competing model architectures.

Future research should also expand beyond the demographic limitations of the current dataset. Although this study focused on elderly subjects aged 55–94, broader datasets containing increased racial, ethnic, geographic, and age diversity would improve the generalizability of the AI vision model and reduce potential demographic bias. The equitability analysis reported here is specific to the male/female axis, and analogous analyses along skin-tone, age-bracket, and accessory-presence axes are needed before the model can be characterized as broadly equitable. Such analyses depend on datasets that include the corresponding demographic metadata, and constructing or curating such datasets is itself a substantial research effort. An intersectional audit modeled on the Gender Shades methodology [22], conducted on an elderly-faces dataset stratified by skin tone, would be a particularly informative next step.

Additional studies should evaluate model performance under varying environmental conditions, including different lighting environments, camera resolutions, facial orientations, partial occlusions, and real-time video inputs. The AGFW database consists primarily of well-lit, forward-facing still images, which is a tractable starting point but is not representative of the imagery a robot operating in a real home or care facility would encounter. Variability in ambient lighting (warm interior bulbs, mixed natural and artificial light, low-light evening conditions), camera positioning (a robot looking

from above), and motion (a subject turning their head, walking past the camera, or being partially occluded by furniture) will each affect detection performance, and dedicated evaluation against each of these variability sources is needed.

A further limitation concerns the threshold-based definition of accuracy used in the present study. An 80% F1 threshold provides an interpretable, conservative cutoff, but the binary accept/reject framing it produces discards information about how close to that threshold each inference fell. Future analyses should also report distributions of F1 scores, average precision and recall in their continuous form, and per-feature accuracy disaggregations (separating eye detection from mouth detection rather than collapsing both into a single per-image score). The two-proportion z-test, while appropriate for the binary-outcome formulation used here, inherits the limitations of any frequentist hypothesis test conducted on a single dataset; a larger testing pool would either provide stronger evidence of equitability or surface a difference that the present sample size cannot detect.

Looking further forward, this work would allow for the next stage of development, which would involve implementing real-time emotion prediction and analysis on elderly subjects using Edge AI vision systems. By accurately identifying facial features such as the eyes and mouth, the trained model could serve as a foundational component for future emotion-recognition algorithms capable of detecting emotional expressions in real time. Such systems could potentially be applied in healthcare, assisted living environments, and elderly monitoring technologies to help assess emotional well-being, detect signs of distress, and improve patient care through continuous noninvasive observation. The privacy-preserving architectural choice that motivates the present work would be especially valuable in this downstream context, since emotion data is among the most personal categories of information a sensor can collect, and processing it locally would substantially reduce the surface area for misuse.

Finally, the integration of the trained vision model into a complete robotic platform is itself an engineering project that this paper does not attempt. Embedding the model in a robot also requires designing the camera subsystem, the inference scheduler, the policy that converts inference outputs into robot behaviors, the user-facing controls that govern when the vision subsystem is active, and the auditing and logging infrastructure that supports oversight by caregivers and family members. Each of these layers raises further design questions that future work will need to address.

VII. CONCLUSIONS

This study proposed and evaluated an Edge AI vision model for elderly social assistance robots, developed using Edge Impulse and trained on a curated subset of AGFW (ages 55–94). The model, built on the FOMO MobileNetV2 0.35 architecture, achieved 98.04% accuracy on male subjects and 99.51% on female subjects, with the difference failing to reach statistical significance ($z = 1.43$, $p = 0.15$) at $\alpha = 0.05$.

edge-deployable vision for elderly care robotics, demonstrated at the scale of feature localization on a constrained model, supports a broader architectural shift away from cloud-based vision pipelines and toward on-device inference. Second, the absence of a detectable male/female accuracy disparity is encouraging evidence that lightweight, edge-deployed vision systems can be both privacy-preserving and demographically equitable, though equitability along other axes remains to be characterized. Continued work along the directions outlined in the previous section will be necessary to translate this feasibility result into a deployed system that can responsibly support elderly users in their daily lives.

ACKNOWLEDGMENT

The author thanks the Governor's School for Science and Technology and Hampton University for their support, Dr. Moses Garuba for his continuous guidance throughout the research process, and Dr. Nina Semenova and Laura Vobrak for their feedback and assistance during the completion of this study.

REFERENCES

[1] M. Nielsen, "The evolution of assistive technology for seniors," DEV Community, Jan. 2025. [Online]. Available: https://dev.to/vela-chairs/the-evolution-of-assistive-technology-for-seniors-58eb

[2] N. L. A. Rashid, Y. Leow, P. Klainin-Yobas, S. Itoh, and V. X. Wu, "The effectiveness of a therapeutic robot, 'PARO', on behavioural and psychological symptoms, medication use, total sleep time and sociability in older adults with dementia: A systematic review and meta-analysis," Int. J. Nurs. Stud., vol. 145, p. 104530, Sept. 2023.

[3] T. Huang and C. Huang, "Attitudes of the elderly living independently towards the use of robots to assist with activities of daily living," Work, vol. 69, pp. 55–65, May 2021.

[4] O. O. Aramide, "Edge AI and its impact on resilient AI fabric design: Distributed intelligence and data locality," SAMRIDDHI J. Phys. Sci., Eng. Technol., vol. 17, pp. 12–24, 2025.

[5] Andromeda Robotics, "Andromeda | Personalised robot companions for aged care home residents in Australia," 2025. [Online]. Available: https://andromedarobotics.ai/

[6] J. Liu, X. Wang, and J. Zhang, "Investigating elderly individuals' acceptance of artificial intelligence (AI)-powered companion robots: The influence of individual characteristics," Behav. Sci., vol. 15, p. 697, May 2025.

[7] B. Deusdad, "Ethical implications in using robots among older adults living with dementia," Front. Psychiatry, vol. 15, Sept. 2024.

[8] L. Battistuzzi, A. Sgorbissa, C. Papadopoulos, I. Papadopoulos, and C. Koulouglioti, "Embedding ethics in the design of culturally competent socially assistive robots," IEEE, Feb. 2020.

[9] V. Karami, M. J. Yaffe, G. Gore, A. Moon, and S. Abbasgholizadeh Rahimi, "Socially assistive robots for individuals with Alzheimer's disease: A scoping review," Arch. Gerontol. Geriatr., vol. 123, pp. 105409, Mar. 2024.

[10] M. Leineweber, C. V. Keusgen, M. Bubeck, J. Haltaufderheide, R. Ranisch, and C. Klingler, "Ethical aspects of the use of social robots in elderly care – a systematic qualitative review," arXiv:2505.09224, 2025.

[11] C. Moro, S. Lin, G. Nejat, and A. Mihailidis, "Social robots and seniors: A comparative study on the influence of dynamic social features on human–robot interaction," Int. J. Soc. Robot., vol. 11, pp. 5–24, Aug. 2018.

[12] J. A. Dosso, J. N. Kailley, G. K. Guerra, and J. M. Robillard, "Older adult perspectives on emotion and stigma in social robots," Front. Psychiatry, vol. 13, Jan. 2023.

[13] K. Alikhademi, E. Drobina, and J. D. Louis, "Person detection through the lens of algorithmic bias," OpenReview, 2025. [Online]. Available: https://openreview.net/forum?id=tC1b9DBWww

[14] C. Wang, Z. Yuan, P. Zhou, Z. Xu, R. Li, and D. O. Wu, "The security and privacy of mobile-edge computing: An artificial intelligence perspective," IEEE Internet Things J., vol. 10, pp. 22008–22032, Dec. 2023.

directions," arXiv:2407.04053, 2024.

[16] M. Mukherjee, R. Matam, C. X. Mavromoustakis, H. Jiang, G. Mastorakis, and M. Guo, "Intelligent edge computing: Security and privacy challenges," pp. 26–31, Sept. 2020.

[17] A. Yao, G. Li, X. Li, F. Jiang, J. Xu, and X. Liu, "Differential privacy in edge computing-based smart city applications: Security issues, solutions and future directions," Array, vol. 19, pp. 100293, Sept. 2023.

[18] "Face aging | The AGFW database," Github.io, 2026. [Online]. Available: https://dcnhan.github.io/projects/agingproject/the-agfw-database.html

[19] Statista Research Department, "Percentage of the U.S. population aged 65 years and over from 1950 to 2050," Statista, 2026. [Online]. Available: https://www.statista.com/statistics/457822/

[20] Y. Yu, J. Zhang, et al., "Response of global health towards the challenges presented by population aging," China CDC Wkly., 2023.

[21] B. Bardach et al., "Healthcare on the brink: Navigating the challenges of an aging society in the United States," npj Aging, vol. 10, 2024.

[22] J. Buolamwini and T. Gebru, "Gender Shades: Intersectional accuracy disparities in commercial gender classification," Proc. Mach. Learn. Res., vol. 81, pp. 77–91, 2018.

[23] C. Banbury et al., "MLPerf Tiny benchmark," Proc. NeurIPS Datasets and Benchmarks Track, 2021. [Online]. Available: https://arxiv.org/abs/2106.07597

[24] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 4510–4520.