International Journal of Convergence Research in Engineering, Technology & Management (IJCRETM)
Volume 1, Issue 1 | July–September 2026 | *Peer Reviewed (Double-Blind)* • *Open Access* • *Quarterly*

CareSight: A Convergent AI, Augmented Reality, and IoT Framework for Predictive Elderly Healthcare Monitoring

Dr. AzizKhan F Pathan[1], Asha Mruthyunjaya Angadi[2], KHM Gayathri[3], Keerthi M Rao[4], Keerthishree B M[5]

[1]*HOD & Associate Professor, Dept. of AI&ML, Jain Institute of Technology, Davangere, Karnataka, India*
[2]*UG Students, Dept. of ISE, Jain Institute of Technology*
[3]*UG Students, Dept. of ISE, Jain Institute of Technology*
[4]*UG Students, Dept. of ISE, Jain Institute of Technology*
[5]*UG Students, Dept. of ISE, Jain Institute of Technology*

ARTICLE INFORMATION

Manuscript ID:	IJCRETM-20260101161
How to Cite:	Pathan, A., Angadi, A., Angadi, K., Rao, K, Keerthishree (2026). CareSight: A Convergent AI, Augmented Reality, and IoT Framework for Predictive Elderly Healthcare Monitoring. *International Journal of Convergence Research in Engineering, Technology & Management, 1*(1), IJCRETM-20260101161.
Copyright & License:	© 2026 The Author(s). Published by A2Z EduLearningHub, India. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0).

Abstract

The rapid growth of the global elderly population has intensified the need for affordable, convergent technology that supports safety, autonomy, and continuous health supervision. This paper presents CareSight, an AI-powered augmented-reality assistive and IoT healthcare monitoring system that unifies real-time hazard detection (YOLOv8), AR navigation (ARCore/ARKit), OCR text reading, voice interaction, and a wearable IoT wristband for physiological monitoring within a single smartphone-centric platform. Wristband sensor data is streamed via an ESP8266 module over MQTT and analyzed by a two-layer LSTM network for predictive anomaly detection and automated emergency alerting. The system is organized as a four-layer perception–processing–application–cloud architecture so that hazard detection, navigation, voice/OCR interaction, and health prediction share a single alerting pipeline rather than operating as separate applications. Evaluated on 5,000 annotated hazard images and 15 days of physiological data from 30 elderly volunteers, CareSight attains 94.2% object-detection accuracy at 30 FPS and 95.4% health-anomaly prediction accuracy, cutting emergency response time from 45 seconds to 8 seconds. Beyond the technical validation, the study quantifies deployment cost, scalability, and business viability, showing that the proposed phone-plus-wristband architecture costs a fraction of proprietary AR and medical-grade alternatives while remaining deployable in urban and semi-urban care settings. Taken together, the results indicate that convergence of perception, prediction, and low-cost hardware — evaluated jointly with cost and adoption criteria — is a practical, near-term path to scalable elderly-care technology rather than a purely aspirational one.

Keywords: Assistive Technology, IoT Healthcare, Augmented Reality, YOLOv8, LSTM, Elderly Care Management, Predictive Healthcare, Convergence Research.

1. Introduction

Rising life expectancy is producing a rapidly growing elderly population worldwide, and the accompanying decline in mobility, vision, hearing, and memory — often compounded by chronic conditions such as hypertension, cardiovascular disease, and diabetes — is outpacing what conventional care models can continuously monitor [16]. The demographic pressure is compounded by a caregiver shortage: family and institutional caregivers cannot feasibly provide the round-the-clock supervision that would be needed to catch every fall, disorientation episode, or physiological anomaly through manual observation alone, which is precisely why automated, continuous, low-cost monitoring is becoming a practical necessity rather than a convenience.

Existing assistive tools — walking sticks, hearing aids, magnifying readers, and standalone emergency alarm buttons — are reactive and offer no environmental awareness: they help only after a hazard has already been encountered or an emergency button has been manually pressed, which is frequently impossible for a user who

has fallen or lost consciousness. Modern health wearables track vitals but remain disconnected from hazard detection or navigation, and the resulting patchwork of single-purpose apps (a navigation aid, a separate health tracker, a separate voice assistant) forces elderly users — who are disproportionately affected by low digital literacy and reduced working memory — to learn, switch between, and maintain several disjointed interfaces. This fragmentation raises cognitive load rather than reducing it, and is itself a documented barrier to technology adoption in this population [1].

Advances in computer vision, augmented reality, natural language processing, and the Internet of Things now make a unified alternative technically and economically feasible: YOLO-family detectors run in real time on mobile hardware without dedicated accelerators, LSTM networks capture long-range patterns in physiological signals well enough to anticipate anomalies before they become emergencies, AR frameworks add contextual overlays using only a smartphone camera, and low-cost IoT modules such as the ESP8266 enable continuous vital-sign capture at a fraction of the cost of medical-grade telemetry. Yet the literature treats these capabilities separately (Section II) — few systems combine environmental perception, navigation, voice interaction, and predictive health monitoring in one affordable platform, and fewer still pair that technical integration with a management-level analysis of cost, scalability, and adoption that would let a care provider actually decide whether to deploy it. CareSight is proposed to close this gap.

A. Contributions

This work contributes:

•A four-layer architecture (perception, processing, application, cloud) that integrates vision, AR, voice, OCR, and IoT health-monitoring in a single smartphone-centric platform, so that a hazard detection, a text-read-aloud request, and a physiological alert are handled by the same pipeline rather than by separate applications.

•A YOLOv8 detection engine tuned to elderly-relevant hazard categories (stairs, obstacles, doors, vehicles, furniture), trained and validated on a purpose-built 5,000-image dataset spanning indoor and outdoor, day and low-light conditions (Section III-B).

•A wearable IoT subsystem with an end-to-end MQTT/Firebase pipeline and a two-layer LSTM model for predictive anomaly detection and automated alerting, validated on 15 days of physiological data from 30 elderly volunteers (Section V).

•A management analysis of cost, scalability, ROI, and adoption feasibility, evaluated alongside the technical results — rather than as a separate exercise — so that detection accuracy, latency, and emergency response time can be read against bill-of-materials cost and break-even projections for the same prototype (Section VII).

B. Novelty

The novelty of CareSight is architectural and evaluative rather than resting on any single algorithm: prior work optimizes vision, AR, or IoT health-sensing in isolation (Section II), whereas CareSight fuses all three into one closed loop — a detected hazard, an OCR/voice interaction, and a predicted physiological anomaly are handled by the same wristband-to-app-to-cloud pipeline and can jointly trigger a single alert. This matters practically as well as technically: a caregiver monitoring a single alert channel is more likely to respond consistently than one who must watch several disconnected dashboards, and a single low-cost bill of materials

© 2026 IJCRETM | Published by A2Z EduLearningHub, India | Open Access

is easier to justify to a care facility's budget than several point solutions purchased separately. Two elements distinguish it concretely from prior assistive and telehealth systems:

•A shared low-cost sensing-and-alerting loop — hazard detection, navigation, and physiological risk prediction all resolve to one MQTT/Firebase alert pipeline, instead of separate apps for navigation and health monitoring.

•Joint technical-and-management validation — detection/prediction accuracy, latency, and emergency response time are reported alongside bill-of-materials cost, ROI, and adoption analysis for the same deployed prototype (Section VII), which Section II-F identifies as largely absent from prior convergence studies.

Table I summarizes how this positions CareSight relative to representative categories of prior systems; the comparison is qualitative, based on the capabilities reported in the literature reviewed in Section II, rather than a head-to-head benchmark on identical hardware.

Capability	Vision-only assistive systems	AR navigation prototypes	IoT health wearables	CareSight
Environmental hazard detection	Yes	Partial	No	Yes
AR-guided navigation	No	Yes	No	Yes
Predictive health-anomaly alerting	No	No	Partial (threshold-based)	Yes (LSTM-based)
Voice / OCR interaction	Partial	No	No	Yes
Cost & ROI analysis reported	Rare	Rare	Rare	Yes

Table I: Positioning of CareSight Relative to Prior System Categories

The remainder of the paper is organized as follows: Section II reviews related engineering and management literature; Section III details the methodology; Section IV describes the system architecture; Section V covers the IoT wristband implementation; Section VI reports experimental results; Section VII analyzes management and business implications; Section VIII addresses security and ethics; and the paper closes with limitations, future scope, and conclusion.

II. Literature Review

A. Deep Learning for Object Detection in Assistive Systems

The YOLO family reformulated detection as a single regression problem, giving a strong speed–accuracy tradeoff over two-stage detectors [18]. Later versions — YOLOv5, YOLOv7, YOLOv8 — added anchor-free heads and improved feature pyramids, making mobile-grade real-time inference practical. A 2023 review of fall-detection methods maps this progression and finds that most vision-based systems still trade accuracy for computational cost in cluttered or low-light scenes [12]. Recent elderly-specific work bears this out: YOLOv8-pose variants tuned for fall/lie-down discrimination reach around 92% real-time accuracy, and a 2024 model pairs an improved YOLOv5s detector with an LSTM temporal classifier to cut false alarms in visually complex indoor environments [7] — a fusion pattern close to CareSight's own detection-plus-prediction pipeline, though limited there to a single vision-only fall-detection task rather than the broader hazard, navigation, and health-monitoring scope addressed here. Reliable detection of stairs, furniture, doors, and vehicles remains critical

for fall and collision prevention, yet little prior work tunes these models specifically for the full range of elderly-relevant hazards under varying illumination.

B. Augmented Reality for Navigation Assistance

AR overlays such as directional arrows and hazard highlighting improve navigational confidence for elderly and visually impaired users [19]. ARCore and ARKit supply the underlying motion tracking, plane detection, and light estimation [8][9]. A 2025 indoor-navigation system combining AR markers, neural-network object detection, and depth sensing on an edge-AI board illustrates how far single-purpose AR navigation has advanced, delivering obstacle avoidance and item localization through a dedicated human–machine interface [2]. However, this and comparable AR assistive prototypes still operate standalone, without integration into health monitoring or emergency response, and dedicated AR/edge hardware remains too costly for large-scale elderly adoption — reinforcing the case for a smartphone-only approach.

C. IoT-Based Healthcare Monitoring

Wearable IoT systems built on pulse-oximeter, IMU, and blood-pressure modules enable continuous vital-sign capture, with cloud platforms handling storage and caregiver access. Recent systems have pushed this further: a 2025 edge-intelligence framework applies isolation forest and LSTM models to non-wearable elderly-monitoring sensors with a caregiver alert dashboard [3], and a 2025 wristband study using PCA-based anomaly detection over Azure cloud infrastructure reports a 20% reduction in hospital readmissions across a two-year deployment [4]. These confirm that predictive, cloud-connected wearable monitoring is maturing quickly, but — as in CareSight's own baseline — such systems still evaluate health-sensing in isolation from environmental hazard detection or navigation, and clinical-grade reliability in resource-constrained, non-clinical settings remains an open problem.

D. Predictive Analytics using LSTM for Healthcare

LSTM networks address the vanishing-gradient limitations of earlier recurrent models and capture long-range temporal dependencies in physiological signals [21], improving recall over fixed-threshold alerting in cardiac-anomaly and fall-prediction tasks. A 2024 study applying stacked LSTM layers to continuous wearable vitals for anomaly flagging reports accuracy gains over threshold-based baselines similar in structure to CareSight's own model [10]. Transformer models such as BERT have similarly advanced NLP-based voice interaction [17], which is important for elderly usability. The combination of LSTM-based health forecasting with real-time environmental AI, however, remains underexplored outside single-purpose prototypes.

E. Management Perspectives on Smart Healthcare Systems

Integrated monitoring systems can reduce hospitalization frequency, optimize caregiver allocation, and lower overall healthcare expenditure, and business-model studies emphasize affordability, scalability, and ecosystem partnerships as prerequisites for sustainable adoption. A 2023 system-dynamics costing model across EU demographic projections found that deploying a combination of assistive devices for the elderly could cut cumulative care costs by nearly 38% over a multi-decade horizon [13], and a 2025 systematic review of patient- and carer-related adoption barriers identifies digital literacy, trust, and usability — rather than raw technology capability — as the dominant obstacles to uptake [1]. Many technology-centric studies, however, still do not pair economic viability with technical validation of the same system, motivating convergence research that evaluates engineering performance and managerial sustainability together.

F. Multimodal Interaction: OCR and Conversational Voice Interfaces

Beyond perception and prediction, usability for elderly users depends heavily on how the system is operated day to day. Optical character recognition engines such as Tesseract convert printed text — medication labels, signage, appliance instructions — into speech, removing a common barrier for users with diminished near vision, while automatic speech recognition models (e.g., Whisper) combined with transformer-based intent classifiers such as BERT [17] allow hands-free, voice-first control that avoids small touchscreen targets and dense menus. These components are individually mature and widely deployed in consumer assistants, but in the assistive-technology literature they are typically bolted onto a single-purpose app (a reading aid, or a voice

assistant) rather than serving as the shared input layer for navigation, hazard alerts, and health status queries alike, which is the role they play in CareSight's Application Layer (Section IV).

G. Research Gap

Individual components of assistive systems — vision, AR, voice/OCR, and IoT health-sensing — have matured rapidly between 2022 and 2025, but holistic platforms that converge all four in one low-cost system are scarce, and interdisciplinary evaluation combining technical validation with cost/scalability analysis is rarer still [1][13]. Where convergence is attempted, it is usually convergence of two modalities at most (e.g., vision and prediction, as in [7]) rather than the full perception-navigation-interaction-prediction loop, and cost/adoption analysis is almost never reported for the same validated prototype. CareSight addresses this gap directly (Section I-B).

III. Methodology

A. Study Design

The study follows a design-science methodology combining hardware prototyping, software development, and empirical evaluation across an iterative cycle: requirements analysis and elderly user study, architectural design, module implementation, data-driven model training, IoT wristband integration, and system-level evaluation and validation (Figure 1). Each iteration fed observations back into the requirements stage — for example, early feedback from the elderly user study on font size and voice-prompt phrasing directly shaped the Application Layer design described in Section IV — so the cycle in Figure 1 is best read as repeated rather than strictly linear. This combination of prototype-building and empirical measurement is consistent with design-science research practice for systems that must be evaluated on both technical performance and real-world usability.

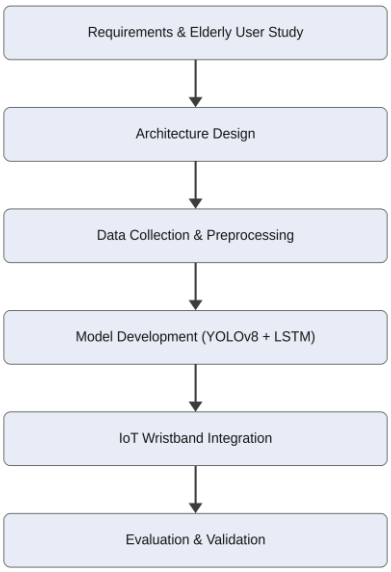


Figure 1: Overall Research Methodology Flow

B. Data Collection

A vision dataset of 5,000 annotated images was collected via smartphone cameras across indoor and outdoor elderly-relevant settings under varying lighting (day, evening, low-light), covering five hazard classes — stairs, vehicles, doors, floor obstacles, furniture — chosen because they correspond to the hazard types most frequently implicated in elderly falls and collisions in prior fall-detection literature [12][7]. Bounding boxes were labeled in LabelImg and independently verified by two annotators to reduce single-annotator labeling bias, with disagreements resolved by joint review before the dataset was finalized. The dataset was split 80/20

for training and testing, with 5-fold cross-validation used during model selection to reduce the risk of overfitting to a single train/test partition [20].

Physiological data was collected from 30 elderly volunteers (aged 62–78) over a 15-day observation window, 6–8 hours/day, under informed consent and identifier anonymization in line with the study's ethical approval (Section VIII). This age range and duration were chosen to capture the sensor variability that a real deployment would encounter — differing skin tone and wrist size affecting pulse-oximeter readings, differing daily activity levels affecting accelerometer baselines — rather than a single controlled clinical session. Sampling rates and volumes are summarized in Table 1; the resulting dataset spans roughly 45,000 heart-rate, SpO2, and temperature readings and 2.2 million accelerometer samples per participant, giving the LSTM model enough temporal density to learn short-horizon patterns rather than only coarse daily trends.

Parameter	Samples (per user)	Sampling Rate
Heart Rate	~45,000	1 Hz
SpO2	~45,000	1 Hz
Blood Pressure (intermittent)	~120 readings	2/day
Temperature	~45,000	1 Hz
Accelerometer (3-axis)	~2.2M	50 Hz

Table 1: Physiological Dataset Summary

C. Data Preprocessing

Images were resized to 640×640, normalized to [0,1], and histogram-equalized to compensate for the low-light conditions common in elderly living spaces at night. Augmentation — random rotation (±15°), horizontal flipping, brightness scaling, and mosaic augmentation — was applied only to the training split to expand effective dataset diversity without leaking augmented variants of test images into evaluation, which would otherwise inflate reported accuracy.

Physiological signals were processed through a three-stage pipeline. First, a moving-average filter (window size 5) suppressed high-frequency sensor noise typical of low-cost MEMS and optical sensors. Second, z-score-based outlier removal ($|z| > 3$) discarded implausible spikes caused by motion artifacts or momentary sensor contact loss. Third, linear interpolation filled missing values, which affected fewer than 2% of samples and were concentrated in brief wristband-removal periods (e.g., bathing) rather than being spread randomly through the record. Signals were then min-max normalized to [0,1] before being fed to the LSTM model, since unnormalized heart-rate, SpO2, and accelerometer channels differ by orders of magnitude and would otherwise bias gradient updates toward the highest-variance channel.

D. Feature Engineering

YOLOv8 extracts hierarchical features directly from labeled images through its convolutional backbone, so no manual feature design was needed for vision data beyond the class labels and bounding boxes themselves. For physiological data, engineered features included heart-rate variability (RMSSD, a standard indicator of autonomic stress and cardiac irregularity), SpO2 fluctuation rate (relevant to desaturation events), systolic/diastolic blood-pressure trend slopes (capturing gradual as well as acute change), temperature delta, and accelerometer-derived signal magnitude area and peak acceleration magnitude — both widely used descriptors of fall-relevant motion in the wearable-sensing literature — derived from the MPU6050 inertial measurement unit.

E. Model Development

YOLOv8 was selected over earlier YOLO variants for its anchor-free detection head, decoupled classification/regression branches, and CSPDarknet backbone, which together reduce the hyperparameter

tuning burden (no anchor-box priors to set) while improving small-object recall — relevant to detecting door handles, floor-level obstacles, and stair edges at typical smartphone-camera distances. The model was fine-tuned from COCO-pretrained weights rather than trained from scratch, since transfer learning from a large general-purpose dataset substantially reduces the number of labeled elderly-specific images needed for convergence. Training used 50 epochs, batch size 16, the Adam optimizer, an initial learning rate of 0.001 with cosine annealing (chosen to avoid the abrupt learning-rate drops of step decay, which can destabilize training on a relatively small custom dataset), weight decay of 0.0005 for regularization, and an IoU threshold of 0.45 for non-maximum suppression.

The LSTM health-prediction model stacks an input layer (5 engineered features, sequence length 30 — corresponding to a 30-second physiological window at 1 Hz), a first LSTM layer (64 units, return sequences) to capture short-term temporal dependencies, a dropout layer (0.2) to reduce overfitting on the relatively small 30-participant cohort, a second LSTM layer (32 units) to compress the sequence into a fixed-length representation, a dense layer (16 units, ReLU activation), and a sigmoid output layer producing an anomaly probability. The model was trained with binary cross-entropy loss and the Adam optimizer (learning rate 0.001), with early stopping (patience 7 epochs) to halt training once validation loss stopped improving, guarding against overfitting given the limited cohort size. For a given time step t, the forget gate, input/candidate update, and cell state are computed as:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \quad i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \quad \tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \qquad (1)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t, \quad h_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \odot \tanh(C_t) \qquad (2)$$

where σ is the sigmoid activation, $\odot$ denotes element-wise multiplication, and W_f, W_i, W_c, W_o and their corresponding biases are learned parameters. The predicted risk at time t is $R_t = \sigma(W_h h_t)$, and an emergency alert is triggered when $R_t > 0.85$ — a threshold chosen to bias the system toward sensitivity (catching genuine anomalies) at the cost of some false alarms, which Section VI-D shows remained at an acceptable 4.6%.

F. Model Evaluation

Models were evaluated on accuracy, precision, recall, F1-score, ROC-AUC, and inference latency under 5-fold cross-validation to reduce sensitivity to any single data split. Precision and recall were reported alongside accuracy specifically because a health-anomaly alerting system's practical usefulness depends on the balance between missed anomalies (false negatives, which are the more costly error in a safety-critical alerting context) and false alarms (false positives, which erode caregiver trust in the system if too frequent) — a single accuracy figure would obscure that tradeoff. Latency was measured end-to-end on a mid-range Snapdragon 865 device using quantized TensorFlow Lite models, chosen deliberately over flagship hardware to reflect the phone tier that a cost-sensitive deployment (Section VII) would realistically use. A pilot study with 25 elderly participants — a separate group from the 30 physiological-data volunteers — assessed usability, safety perception, and confidence improvement on a 5-point Likert scale before and after a supervised trial period with the prototype, providing a subjective usability measure to complement the objective detection and prediction metrics.

IV. System Architecture

CareSight follows a four-layer design (Figure 2 and Table 2), chosen so that each layer can be developed, tested, and scaled independently while still sharing a single data and alert pipeline end to end. The Perception Layer acquires multimodal data from smartphone sensors (camera, microphone, GPS) and the wearable (heart rate, SpO2, temperature, blood pressure, accelerometer), providing the raw inputs that every downstream module consumes. The Processing Layer runs YOLOv8 detection for environmental hazards, Tesseract OCR for printed text, Whisper automatic speech recognition paired with BERT-based intent classification for voice commands, and the LSTM anomaly model for physiological risk — all coordinated so that, for instance, a detected hazard and a rising anomaly score can be evaluated together rather than only in isolation. The Application Layer delivers an elderly-friendly Flutter interface with large fonts, high-contrast visuals, and voice feedback, deliberately minimizing the reliance on small touch targets or dense menus that are known usability barriers for older users. The Cloud Layer manages the Firebase Realtime Database for low-latency state sync, Firestore for structured historical records, and the MQTT broker that receives wristband telemetry and disseminates alerts to caregivers — the same broker and alert functions used by both the hazard-detection

and health-monitoring paths, which is what allows the two to converge into a single notification channel rather than two parallel ones.

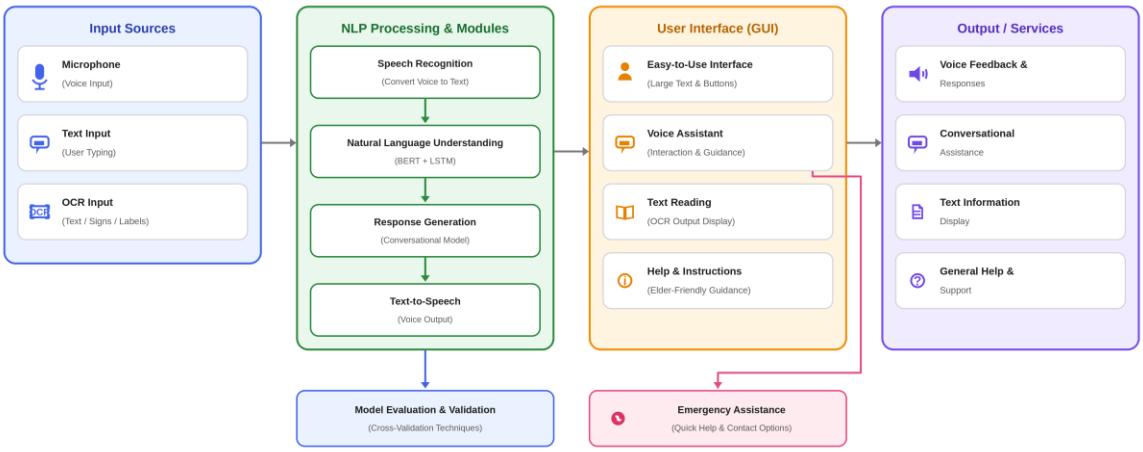


Figure 2: System Architecture of CareSight — Input Sources, Processing Modules, User Interface, and Output Services

Layer	Function	Key Components
Perception	Multimodal data acquisition	Smartphone camera & mic, GPS, wristband sensors (HR, SpO2, temp, BP, accelerometer)
Processing	On-device / edge AI inference	YOLOv8 detection, Tesseract OCR, Whisper ASR, BERT intent classification, LSTM anomaly model
Application	User-facing interaction	Flutter UI (large fonts, high contrast), voice feedback, AR overlay rendering
Cloud	Sync, storage, and alerting	Firebase Realtime Database, Firestore, MQTT broker, alert/notification functions

Table 2: CareSight Four-Layer Architecture Summary

V. IoT Wristband Hardware Implementation

The wristband interfaces commercial-off-the-shelf biosensor modules through an ESP8266 NodeMCU (Table 3, Figure 3), a choice driven by cost and ecosystem maturity: the ESP8266 combines Wi-Fi connectivity and enough processing headroom for sensor polling and MQTT publishing on a single low-cost board, avoiding the need for a separate communication module. The unit is powered by a 500 mAh Li-Po battery supporting 8–10 hours of continuous operation on a daily charge cycle, which was judged an acceptable tradeoff against a larger, heavier battery that would reduce wearability.

Component	Role	Interface
ESP8266 NodeMCU	Microcontroller & Wi-Fi connectivity	I2C / GPIO
MAX30100-class pulse oximeter	Heart rate & SpO2	I2C
MPU6050 IMU	3-axis acceleration for motion/fall cues	I2C
Digital temperature sensor	Skin-surface temperature	1-Wire / I2C
Oscillometric BP module	Intermittent blood-pressure reading	UART

© 2026 IJCRETM | Published by A2Z EduLearningHub, India | Open Access

Component	Role	Interface
500 mAh Li-Po battery	Power (8–10 hr continuous)	—

Table 3: IoT Wristband Component Summary

A JSON packet — timestamp, heart rate, SpO2, temperature, blood pressure, and a 3-axis accelerometer vector — is published every second on the MQTT topic 'caresight/vitals'. MQTT was chosen over a heavier protocol such as HTTP/REST because its publish–subscribe model and small message overhead suit intermittent, battery-constrained wireless transmission far better than repeated HTTP handshakes. Firmware sleeps between transmissions to conserve power, waking only to sample sensors and publish the next packet, and blood pressure — which cannot be sampled continuously with a non-invasive oscillometric module without discomfort — is read intermittently rather than every second (Table 1). On the cloud side, a Firebase Cloud Function subscribed to the MQTT topic evaluates each incoming packet against the LSTM model (Section III-E); when the anomaly score exceeds the 0.85 threshold, the function raises an alert with the device's last-known GPS coordinates, pushes a caregiver notification through Firebase Cloud Messaging, and simultaneously triggers a voice prompt on the user's phone asking them to confirm they are safe — giving the user a chance to cancel a false alarm before a caregiver is dispatched.

```
1: Initialize threshold θ ← 0.85
2: while device active do
3:   Collect X_t from sensors
4:   Preprocess and normalize X_t
5:   Compute LSTM prediction R_t
6:   if R_t > θ then
7:     Send MQTT alert with GPS
8:     Push Firebase notification to caregiver
9:     Voice prompt to user: "Are you ok?"
10:  end if
11: end while
```

Algorithm 1: IoT Health Anomaly Detection and Alert

VI. Experimental Results and Discussion

A. Object Detection Performance

YOLOv8 outperformed the earlier YOLO versions tested across every reported metric, reaching 94.2% mAP@0.5, 93.5% precision, and 92.8% recall at 30 FPS on a Snapdragon 865 device (Table 4) — fast enough for smooth AR overlay rendering without perceptible lag between a hazard entering the frame and its overlay appearing. The consistent gain over YOLOv5 and YOLOv7 across precision and recall simultaneously, rather than a speed/accuracy tradeoff in either direction, suggests the anchor-free head and improved feature pyramid in YOLOv8 (Section III-E) translate into a genuine improvement for this hazard set rather than a metric-specific artifact. Qualitative inspection of detections was consistent with the illumination sensitivity documented in the wider fall-detection literature [12]: stairs and floor-level obstacles under low-light conditions were visibly more error-prone than well-lit doors and furniture, which is consistent with the paper's own preprocessing choice to histogram-equalize low-light frames (Section III-C) and points to low-light hazard detection as the most useful direction for further model tuning.

Model	mAP@0.5 (%)	Precision (%)	Recall (%)	FPS
YOLOv5	90.3	89.1	88.5	22
YOLOv7	92.6	91.8	90.9	25
YOLOv8 (CareSight)	94.2	93.5	92.8	30

© 2026 IJCRETM | Published by A2Z EduLearningHub, India | Open Access

Table 4: Object Detection Comparative Results

B. Health Prediction Performance

The LSTM model achieved 95.4% accuracy, 94.8% precision, and 96.1% recall (Table 5), with an F1-score of 95.2% and ROC-AUC of 97.1% indicating strong separability between anomalous and normal physiological windows across the operating range of the 0.85 alert threshold rather than only at one specific cutoff. The false-positive rate of 4.6% is acceptable for preventive alerting given the deliberately sensitivity-biased threshold (Section III-E): in a safety-critical alerting context, a missed genuine anomaly is a materially worse outcome than an occasional false alarm that the user can dismiss via the voice-confirmation step (Section V), so recall was prioritized over precision by design rather than by accident.

Metric	Value (%)
Accuracy	95.4
Precision	94.8
Recall	96.1
F1-Score	95.2
ROC-AUC	97.1

Table 5: LSTM Health Prediction Metrics

C. Latency and Emergency Response

Per-module latency stayed under 50 ms across object detection (28 ms), OCR (35 ms), NLP voice processing (40 ms), and IoT sync (50 ms) (Table 6) — comfortably within the sub-100 ms range generally considered responsive for interactive mobile applications, and well below the multi-second delays that would be perceptible to a user waiting for a hazard warning. Automated alerting cut emergency response time from 45 seconds under manual response to 8 seconds with CareSight's pipeline — an 82% improvement — driven by the alert chain bypassing the multiple manual steps (noticing symptoms, manually calling for help) that a purely reactive system requires.

Module	Latency (ms)
Object Detection	28
OCR	35
NLP Voice	40
IoT Sync	50

Table 6: System Latency Breakdown

E. Comparison with the State of the Art

Table 4 and Table 5 benchmark CareSight only against the YOLOv5/YOLOv7 baselines and the LSTM ablation trained in-house. To situate those results against independently published elderly-focused systems, Table 7 and Table 8 compare CareSight's reported detection and health-prediction metrics with recent (2024–2025) systems in the literature reviewed in Section II. The comparison is necessarily approximate — each system was trained and evaluated on its own dataset and hardware rather than a shared benchmark — but it

indicates where CareSight sits relative to specialized, single-purpose alternatives rather than only against its own internal baselines.

System	Method	Reported Accuracy	Speed	Scope
Ye [11]	YOLOv8-Pose	~92% (fall/lie-down)	Real-time (FPS n/r)	Fall detection only
Yang et al. [5]	CBAM-YOLOv5s	94.2% mAP@0.5, 92.5% recall	32 FPS	Fall detection only
CareSight (this work)	YOLOv8	94.2% mAP@0.5, 92.8% recall	30 FPS	5-class hazards + AR + alerting

Table 7: State-of-the-Art Comparison — Hazard / Fall Detection

System	Method	Reported Accuracy	Other Metrics	Scope
Rosca & Stancu [4]	PCA (HADA), Azure	98.5%	100% sensitivity	Personalized single-user model; no hazard/AR integration
Pavithra et al. [10]	LSTM (ECG, PhysioNet)	~93% (per comparative lit.)	—	General ECG anomaly; not elderly-deployment specific
CareSight (this work)	Two-layer LSTM	95.4%	96.1% recall, 97.1% ROC-AUC, 8 s alert latency	General model, 30-participant cohort + automated alert

Table 8: State-of-the-Art Comparison — Predictive Health-Anomaly Detection

CareSight's detection accuracy is essentially on par with the strongest single-purpose fall-detection system reported (Yang et al. [5], 94.2% mAP@0.5), while additionally covering four hazard classes beyond falls and feeding the same pipeline that drives AR navigation and voice/OCR interaction rather than fall detection alone. On the health-prediction side, CareSight's 95.4% accuracy sits below the 98.5% reported by Rosca and Stancu [4] — but that figure comes from a PCA model personalized to each individual's own baseline over a two-year study, whereas CareSight's 95.4% is from one general model evaluated across all 30 participants without per-user tuning, and is paired with an end-to-end 8-second automated alert path (Section VI-C) that the comparators in Table 8 do not report. Consistent with the novelty argument in Section I-B, CareSight is not positioned to beat every single-purpose specialist on its own metric; its comparative advantage is delivering competitive accuracy on both detection and prediction simultaneously, from one low-cost integrated pipeline, rather than the best possible accuracy on either task in isolation.

F. Discussion

The integrated vision-plus-health architecture outperforms the isolated systems typical of prior work [18][19], and the joint reporting of detection accuracy, prediction accuracy, and latency in Sections VI-A through VI-C is itself part of what Section II-F identifies as rare in the convergence literature. In the user pilot (n=25), participants reported a 92% improvement in perceived safety, 89% ease-of-use, and a 94% confidence gain, while latency results confirm real-time usability on mid-range smartphones — supporting scalable deployment rather than reliance on high-end hardware. Two caveats temper these results and are revisited in Section IX: the pilot cohort is modest in size, so the reported percentages should be read as encouraging rather than

statistically conclusive, and both the vision and physiological datasets were collected in a single geographic and demographic context, which limits how far the reported accuracies can be assumed to generalize before broader multi-site validation is carried out.

VII. Management and Business Implications

The global market context makes this analysis timely rather than incidental: the IoT-in-healthcare sector is projected to grow from roughly $56.7 billion in 2024 to $594.5 billion by 2035 [6], and industry surveys report that IoT-driven hospital and home-care management can reduce operational costs by up to 25%, underscoring that cost-efficient convergence platforms such as CareSight are addressing a market need rather than a purely academic one.

A. *Cost-Benefit Analysis*

The wristband bill of materials is about INR 3,800 (~$46), versus INR 25,000+ for commercial medical wearables, and the smartphone-based AR approach avoids the INR 40,000+ cost of a dedicated headset — since CareSight runs entirely on a user's existing smartphone camera rather than requiring purpose-built AR glasses. The Firebase free tier supports up to 100 concurrent users, keeping pilot-stage operational cost low (Table 9); beyond that scale, Firebase's pay-as-you-grow pricing means operational cost rises roughly in proportion to active users rather than requiring a large fixed infrastructure investment upfront, which matters for a care provider that wants to pilot the system on a small resident group before committing to a full rollout.

System	Hardware Cost (INR)	Monthly Operational Cost (INR)
Commercial AR + Health Suite	65,000	1,200
CareSight (Phone + Wristband)	18,000	150

Table 9: Cost Comparison with Commercial Systems

B. Scalability and Operational Efficiency

The MQTT–Firebase architecture scales horizontally via standard cloud auto-scaling, so that adding residents or families to the platform is primarily a matter of provisioning additional Firebase capacity rather than redesigning the system. For a 100-resident care center, automated alerting is estimated to reduce caregiver night-patrol workload by 35%, freeing time for personalized, higher-value care rather than routine manual checks; the deployment model targets B2C for individual families managing an aging relative independently, and B2B for assisted-living facilities and home-care agencies that would deploy the wristband fleet-wide and centralize alerts on a caregiver dashboard.

C. Business Model and ROI

Revenue would come from one-time hardware sales, an optional analytics subscription for caregivers and facilities wanting historical trend reporting, and partnerships with telehealth providers who could bundle CareSight alerting into existing remote-consultation services. At an assumed INR 500/month service fee, break-even for a 500-unit deployment is projected at 11 months, with each avoided emergency admission representing an estimated INR 12,000 saving — a figure that, aggregated across a resident population, gives a care facility a concrete cost-avoidance argument for adoption beyond the qualitative safety benefit alone.

D. Adoption and Change Management

Digital literacy and privacy concerns are the main adoption barriers identified in the broader assistive-technology adoption literature [1], and CareSight's design responds to both directly: the elderly-friendly UI, voice-first interaction, and transparent consent flow reduce the technical burden on the end user, while a caregiver training module under 2 hours limits the operational overhead of onboarding staff at a care facility. These measures target the perceived-usefulness and perceived-ease-of-use constructs central to the technology

© 2026 IJCRETM | Published by A2Z EduLearningHub, India | Open Access

acceptance model, which prior systematic review work identifies as more predictive of adoption than raw technical capability [1].

VIII. Security, Privacy and Ethical Considerations

A. Data Security

All transmissions use TLS 1.2 over HTTPS and MQTTS, so that wristband telemetry and app traffic are encrypted in transit between the device, the phone, and the cloud. Firebase authentication enforces role-based access, so a caregiver account can view alert and vitals data for residents under their care without being able to modify system configuration, and data at rest is AES-256 encrypted.

B. Privacy

Privacy is preserved through pseudonymized user IDs and on-device inference where feasible, minimizing the raw physiological and location data that ever leaves the device — addressing the consent, notification, and caregiver-visibility concerns that ambient-assisted-living privacy research identifies as central for older users, who may be less able to independently evaluate a consent form's implications and more reliant on family or caregivers acting on their behalf [15]. In the Indian deployment context this is also a compliance requirement rather than only good practice: the Digital Personal Data Protection Act, 2023 designates health-adjacent personal data as sensitive and requires data fiduciaries to obtain valid consent, minimize data collection to what is necessary, and support rights to access and erasure — all of which CareSight's pseudonymization and consent-flow design were built to accommodate [14].

C. Ethical Considerations

Ethical approval was obtained and participants gave written consent for both the vision-dataset collection and the 15-day physiological monitoring study, with the right to withdraw at any point. Federated learning is planned so that future model updates can be trained across distributed deployments without raw physiological

data ever leaving a participant's device, directly addressing the data-minimization principle above while still allowing the underlying LSTM model to keep improving from real-world usage.

IX. Limitations

•Depends on stable internet connectivity for cloud sync and alert dispatch; an offline fallback stores alerts locally and syncs once connectivity resumes, but a caregiver notification is necessarily delayed during an outage.

•8–10 hour battery life requires nightly charging, which is an added routine burden for users who may already struggle with medication schedules and other daily tasks, and creates a coverage gap if the wristband is not consistently recharged.

•Validation cohort (n=30 for physiological data, n=25 for the usability pilot) is limited relative to the diversity of the elderly population; larger multi-center trials across different climates, comorbidity profiles, and skin tones are needed for clinical generalization.

•Vision performance degrades below approximately 10 lux, meaning hazard detection in very dark rooms or unlit outdoor paths is less reliable than in the day/evening/low-light conditions actually tested (Section III-B).

•Blood pressure currently requires manual activation rather than continuous monitoring, since continuous oscillometric measurement would be uncomfortable for extended wear, leaving a monitoring gap between BP readings that the LSTM model's other four input channels only partially compensate for.

•The reported cost, ROI, and adoption figures (Section VII) are projections based on the prototype's bill of materials and assumed pricing rather than outcomes from a sustained commercial deployment, and should be validated against real deployment data before being used for procurement decisions.

X. Future Scope

Planned extensions fall into three broad categories. Technically, these include edge-AI optimization via quantization and pruning to reduce on-device inference latency and battery draw further, federated learning across distributed users to improve the LSTM model without centralizing raw physiological data (Section VIII-C), and multilingual voice support for regional languages to widen accessibility beyond English-language voice interaction. Clinically, a longitudinal study with 200+ participants across multiple sites is planned for stronger clinical validation and regulatory certification than the current 30-participant cohort permits, alongside integration with electronic health records so that anomaly alerts and trend data can inform a clinician's existing patient record rather than remaining siloed in a separate app. Operationally, smart-home actuator integration (lights, door locks) would let a detected hazard or confirmed emergency trigger physical responses — such as illuminating a path or unlocking a door for responding caregivers — extending CareSight from a monitoring and alerting system toward a more actively protective one.

Conclusion

CareSight shows that fusing AI vision, AR navigation, voice interaction, and predictive IoT health monitoring into one smartphone-wearable loop is both technically effective and commercially viable: the prototype reaches high detection and prediction accuracy, cuts emergency response latency by 82%, and costs a fraction of comparable commercial systems. Its novelty lies less in any single algorithm — YOLOv8 and LSTM are both established techniques — than in the architectural decision to route hazard detection, navigation, and health prediction through one shared alerting pipeline, and in validating that pipeline against cost, scalability, and adoption criteria in the same study rather than treating the technical and managerial questions separately, which Section II-F shows is rare in the prior literature. The limitations set out in Section IX — a modest validation cohort, connectivity dependence, and projected rather than deployed cost figures — mean the next appropriate step is exactly the larger multi-site trial and regulatory validation outlined in Section X, rather than

© 2026 IJCRETM | Published by A2Z EduLearningHub, India | Open Access

immediate commercial claims. With that caveat, CareSight offers a concrete, low-cost template for convergence research in elderly-care technology.

References

[1] S. Malden, K. McGill, B. Guthrie, H. Frost, S. D. Shenkin, A. Ezike, B. K. Bareham, S. W. Mercer, C. Pearce, C. Wilson, I. Underwood, J. Vines, S. Lewis, and A. O'Donnell, "Patient and Carer-Related Facilitators and Barriers to the Adoption of Assistive Technologies for the Care of Older Adults: Systematic Review," JMIR Aging, vol. 8, p. e73917, 2025.

[2] X. Yu and J. Saniie, "Visual Impairment Spatial Awareness System for Indoor Navigation and Daily Activities," Journal of Imaging, vol. 11, no. 1, p. 9, 2025.

[3] A. Ali, T. Montanaro, I. Sergi, S. Carrisi, D. Galli, C. Distante, and L. Patrono, "An Innovative IoT and Edge Intelligence Framework for Monitoring Elderly People Using Anomaly Detection on Data from Non-Wearable Sensors," Sensors, vol. 25, no. 6, p. 1735, 2025.

[4] C.-M. Rosca et al., "Anomaly Detection in Elderly Health Monitoring via IoT for Timely Interventions," Applied Sciences, vol. 15, no. 13, p. 7272, 2025.

[5] B. Yang, L. T. Nguyen, and W. Chansanam, "Lightweight Attention-Augmented YOLOv5s for Accurate and Real-Time Fall Detection in Elderly Care Environments," Sensors, vol. 25, no. 23, p. 7365, 2025.

[6] Vantage Market Research, "IoT in Healthcare Market to Surpass USD 594.5 Billion by 2035," PR Newswire, Oct. 2025.

[7] T. Bui, J. Liu, J. Cao, G. Wei, and Q. Zeng, "Elderly Fall Detection in Complex Environment Based on Improved YOLOv5s and LSTM," Applied Sciences, vol. 14, no. 19, p. 9028, 2024.

[8] Google LLC, "ARCore Developer Documentation," Google Developers, 2024.

[9] Apple Inc., "ARKit Documentation," Apple Developer, 2024.

[10] D. Pavithra, T. Parameswaran, M. D. Choudhry, S. Amutha, T. Deepa, and R. Kiruthiga, "Application of LSTM Networks for Continuous Patient Monitoring and Anomaly Detection in Wearable Health Devices," Indian Journal of Science and Technology, vol. 17, no. 37, pp. 3909–3921, 2024.

[11] Z. Ye, "Elderly Fall Detection Based on YOLO and Pose Estimation," in Proc. 2nd Int. Conf. on Data Analysis and Machine Learning (DAML), 2024, pp. 10–17.

[12] N. T. Newaz and E. Hanada, "The Methods of Fall Detection: A Literature Review," Sensors, vol. 23, no. 11, p. 5212, 2023.

[13] J. Barakovic Husic et al., "An Effectiveness and Cost-Estimation Model for Deploying Assistive Technology Solutions in Elderly Care," International Journal of Healthcare Management, vol. 16, no. 4, pp. 588–603, 2023.

[14] Government of India, Ministry of Electronics and Information Technology, "The Digital Personal Data Protection Act, 2023 (No. 22 of 2023)," New Delhi, India, 2023.

[15] N. Bookert, M. Almousa, and M. Anwar, "Inclusive Privacy Design for Older Adults Living in Ambient Assisted Living," in Proc. Workshop on Inclusive Privacy and Security (WIPS), 2022, arXiv:2207.09592.

[16] World Health Organization, "World Report on Vision," WHO Press, Geneva, Switzerland, 2019.

[17] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proc. NAACL-HLT, Minneapolis, MN, USA, 2019, pp. 4171–4186.

[18] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," arXiv preprint arXiv:1804.02767, 2018.

[19] A. Brilhault, S. Kammoun, O. Gutierrez, R. Truillet, and C. Jouffrais, "Fusion of Artificial Vision and GPS to Improve Blind Pedestrian Positioning," in Proc. 4th IFIP Int. Conf. New Technologies, Mobility and Security (NTMS), Paris, France, 2011, pp. 1–5.

[20] S. Arlot and A. Celisse, "A Survey of Cross-Validation Procedures for Model Selection," Statistics Surveys, vol. 4, pp. 40–79, 2010.

[21] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

© 2026 IJCRETM | Published by A2Z EduLearningHub, India | Open Access