Journal of Research High School

Vol. 2026 (11) 863 – 869
ISSN 2688-3651 [online]

Correcting Ultrasonic Sensed Distance Using Supervised Machine Learning

Joshua Cho[1] *, Myoungwoo Kim[2]

[1]North Penn High School, North Wales, PA, USA
[2]Blue Bell Envirobotics Lab, Ambler, PA, USA
*Corresponding Author: yeasungc1104@gmail.com

Advisor: Myoungwoo Kim, information@bluebellenvirobotics.com

Received November 20, 2025; Revised August 4, 2026; Accepted August 31, 2026

Abstract

Ultrasonic sensors such as the HC-SR04 are widely used for low-cost distance measurement, but their readings become inaccurate when the detected object lies away from the perpendicular (90°) direction of the sensor. This study quantified how angle and distance affected the measurement bias of a common ultrasonic sensor. An HC-SR04 sensor was fixed at the origin of a printed directional grid and measured a target placed at every 2° from 60° to 120° and at nine prescribed distances from 5.0 cm to 25.0 cm, and the angular bias of each reading was taken as its difference from the reading obtained at 90° for the same distance. A supervised Classification and Regression Tree (SCART) model was then applied, in which the measurements were classified first by distance and then by angle, a quadratic regression of bias against angle was fitted within each node, and the fitted bias was subtracted from the raw reading. Under leave-one-out cross-validation the mean absolute error decreased from 0.84 cm to 0.16 cm and the maximum error from 3.20 cm to 0.88 cm, a 73% reduction, while an otherwise identical correction that omitted the classification by distance reached only 1.44 cm. A simple machine learning model can therefore significantly improve the reliability of inexpensive ultrasonic sensing systems, and the classification step is what makes the improvement possible. This correction method can be applied to various applications including robotics and driver-assistance systems.

Keywords: IoT, Arduino, Ultrasonic sensor, Distance error, Machine learning, CART model, Radar

1. Introduction

An ultrasonic sensor measures distance by emitting a short burst of high-frequency sound, waiting for that burst to reflect from a nearby object, and converting the round-trip travel time into a distance (Aliew, 2022). Because the method requires no contact with the target and can be built from inexpensive components, ultrasonic sensors including the HC-SR04 are used in a wide range of Internet of Things (IoT) applications, such as object detection, automation, vehicle rear distance measurement, smart canes for people who are blind, AI-driven robotics, drone height control above crops, and defense-related systems.

However, it has been reported that distance error occurs due to the angular loss of reflection (Aliew, 2022; Carotenuto et al., 2024; Qiu & Lu, 2022; Smaili et al., 2021; Van Scoy et al., 2023). Depending on the surface and the orientation of an object, part of the transmitted wave is reflected away from the receiver, so the echo that returns is weaker and arrives later, and the sensor reports a distance that differs from the true distance. When the line from the sensor to an object is not perpendicular to the sensor face, the reported distance therefore deviates systematically from the true distance. Although the deviation is often only one to three centimeters, it is large enough to matter for applications such as obstacle avoidance in small robots.

Prior work has established that these angular errors exist and can be partly compensated, but two questions remain unresolved. First, most characterizations have reported error along a single axis or at a few angles, so the combined influence of angle and distance across a two-dimensional measurement field is not well described. Second, published

Vol. 2026 (11) 863 – 869
ISSN 2688-3651 [online]

corrections have generally relied on fixed physical or empirical formulas rather than on a correction surface learned from measurements taken with the same low-cost hardware that is deployed in the field. The present study addressed both gaps by generating a dense angle-distance error map, by applying a supervised CART (SCART) model whose classification structure was imposed on physical grounds rather than selected automatically, and by comparing that model directly with a correction that uses angle alone, so that the contribution of the classification step could be measured.

A test setup of a fixed ultrasonic sensor was assembled and placed onto a paper model that had a degree range from 0 to 180 and a distance range from 0 cm to 30 cm, and measurements of a target placed at prescribed angles and distances were recorded. The measurement error before and after correction is presented and evaluated in the sections that follow.

2. Background

Research on low-cost ultrasonic rangefinders has followed three broad directions. The first has characterized the error of the sensor itself. Smaili et al. (2021) examined errors in the distance and angle measurements of the HC-SR04 and reported that accuracy degrades as a target moves away from the sensor axis, and Van Scoy et al. (2023) compared low-cost ultrasonic and infrared sensors and showed that the "reality gap" between nominal and actual sensor behavior is large enough to affect robot simulation.

The second direction has sought improved measurement procedures. Aliew (2022) proposed an approach for more precise distance measurement with ultrasonic sensors based on the trigger and pulse-echo timing method. Qiu and Lu (2022) reviewed the available ultrasonic ranging methods and summarized the challenges that remain, including temperature dependence, beam spreading, and the geometry of reflection.

The third direction has applied corrections to the readings after they are acquired. Carotenuto et al. (2024) developed a method for correcting signal aberrations in ultrasonic indoor positioning, demonstrating that systematic distortions can be compensated in post-processing rather than eliminated in hardware.

Taken together, these studies establish that reflection geometry is a dominant source of error and that post-processing correction is feasible. What they do not provide is a correction that varies simultaneously with angle and with distance for a single low-cost sensor and that is derived from measurements made with that same sensor.

Machine learning offers a way to construct such a correction directly from measured data. Classification and Regression Tree (CART) analysis partitions a dataset into subgroups whose members behave similarly and fits a simple model within each subgroup. A fully automated tree, however, selects its own splitting variables and thresholds, which can produce partitions that are statistically convenient but physically difficult to interpret, and which tend to overfit when the training set is small, as it is in a single-sensor benchtop experiment. In the supervised CART (SCART) model used here, the splitting variables and their order were specified in advance on physical grounds, which keeps the tree shallow and every node interpretable. SCART was therefore preferred to ensemble methods such as Random Forest or gradient boosting, which require larger training sets before they can be trained and validated reliably.

3. Materials and Methods

The study was carried out in four stages that together formed a single experimental design. A measurement platform was first assembled from an Arduino Mega 2560 board and an HC-SR04 sensor. A directional grid was then printed and used as the reference coordinate system, so that every reading could be paired with a known prescribed distance and angle, the difference between the two defining the bias for that grid cell. The resulting bias field served as the training data for the SCART model, and the modeled bias was finally subtracted from the raw readings, whose remaining error was then evaluated by cross-validation.

3.1 Data Generation

Arduino Setup

To collect the ultrasonic-sensed distance data with the HC-SR04, an Arduino Mega 2560 board was connected to

Vol. 2026 (11) 863 – 869
ISSN 2688-3651 [online]

a laptop running the Windows operating system, together with an ultrasonic sensor that provides trigger and echo channels as well as 5 V and ground wires (Elegoo Inc., 2017). Figure 1 displays the electromechanical circuit design for the Arduino board and the ultrasonic sensor. The trigger channel was set to pin 12 and the echo channel was wired to pin 11 on the Arduino Mega.

Ultrasonic Sensor

The ultrasonic sensor is an inexpensive and easy-to-use device that provides non-contact measurement over a range of 2 cm to 400 cm. When an input/output trigger signal is held high for at least 10 µs, the module transmits eight 40 kHz pulses and then detects whether a pulse returns; the duration of the high output on the echo channel corresponds to the travel time, from which distance is calculated. The trigger (left) and echo (right) channels are separated by a gap of 0.1 inch (2.54 mm) (Elegoo Inc., 2017). It has been reported that measurement error can develop from this gap, because the path from the trigger to an object is not identical to the path from that object back to the echo receiver.

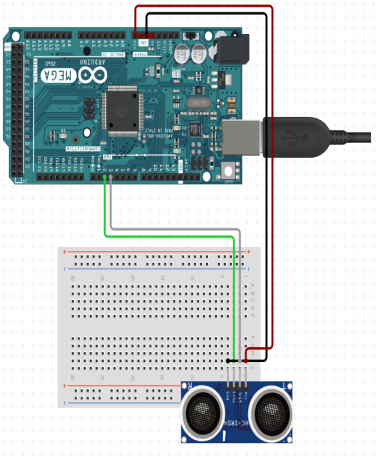


Figure 1. Electromechanical design of Arduino Mega and an ultrasonic sensor (HC-SR04) used in this study. The diagram was created by the authors using Fritzing.

Directional Grid Coordinates

As shown in Figure 2, the ultrasonic sensor was fixed at the origin of the printed grid and measured an object placed at prescribed positions between 0° and 180°. Readings outside the range of 60° to 120° were returned as out-of-range values, because the reflected pulse no longer reached the receiver, so the analysis was restricted to that angular window. Measurements were recorded at every 2° from 60° to 120° and at nine prescribed radial distances from 5.0 cm to 25.0 cm in 2.5 cm steps, producing the 31 by 9 grid summarized in Table 1. For each grid cell, the bias was defined as the difference between the distance sensed at that angle and the distance sensed at 90° for the same prescribed distance, so that the deviation caused by the angle was isolated from the fixed offset of the sensor itself.

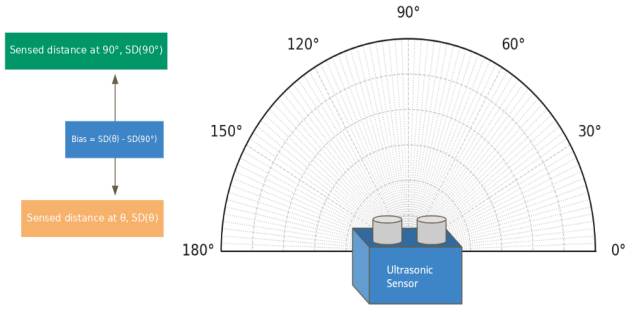


Figure 2. Directional grid coordinates and bias calculation protocol.

Machine Learning: Supervised Classification and Regression Tree (SCART)

The supervised CART model was adopted to correct the biased values. CART analysis divides a dataset into subgroups whose members behave similarly and then fits a simple regression model within each subgroup. In the supervised form used here, the classification variables and the order in which they were applied were specified in advance rather than selected automatically. The first split was made on the prescribed distance, because the area covered by the sensor beam differs at each radius, and the second split was made on the angle relative to the sensor axis ($60° \le \theta < 90°$ and $90° \le \theta < 120°$), because the proportion of the emitted wave returned to the receiver depends on the incidence geometry. A quadratic regression of bias against angle was fitted within each node. As the decision tree went down, the correlation coefficient increased while the sample size in each node decreased, so the tree was limited to two levels, leaving 15 or 16 measurements in each terminal node.

Model Training and Evaluation

The 279 paired readings of Table 1 formed the dataset used to build the model. Because each terminal node contained only 15 or 16 measurements, the model was validated by resampling rather than by setting aside a single test set. Two schemes were used. In the first, the 31 angular positions of each node were split into two interleaved

Vol. 2026 (11) 863 – 869
ISSN 2688-3651 [online]

halves, the regression was fitted on one half and evaluated on the other, and the procedure was repeated with the halves exchanged. In the second, leave-one-out cross-validation was applied, in which each of the 279 cells was removed in turn, the node equation was refitted on the remaining cells, and the bias of the removed cell was predicted. Every reported correction is therefore obtained from a model that has not seen the cell it corrects.

The correction itself was applied by subtracting the bias predicted for a grid cell from the raw reading in that cell, so that the error remaining after correction is the difference between the measured and the predicted bias. Two correction schemes were compared in order to isolate the contribution of the classification step. In the SCART scheme, the regression equation of the node to which a cell belongs, defined by its distance and then by its angle, was used. In the angle-only scheme, the measurements of all nine distances were pooled and a single pair of equations, one on each side of 90°, was applied to every cell.

Performance was assessed with the coefficient of determination (R^2) of the regression within each node, the mean absolute error across all grid cells, the maximum absolute error, and the proportion of cells with an error below 0.8 cm. The reduction in the maximum error relative to the uncorrected readings was reported as the overall metric, and cross-validated values were reported alongside fitted values so that the predictive accuracy of the model could be separated from its fit to the training data.

4. Results

4.1 Results from Data Analysis

Table 1 displays the ultrasonic distance (cm) measurements obtained on the directional grid. At a prescribed distance of 5 cm, for example, the sensor reported 5.1 cm at 90° and 6.3 cm at 60°, so the bias at 60° was 1.2 cm. The readings taken at 90° differed from the prescribed distances by between −0.2 cm and 0.3 cm, and referencing every reading to the 90° value removed this fixed offset from the analysis.

Table 1. Distance measurement of ultrasonic sensor in the direction of the coordinate system

Distance	**Degree**															
	60	**62**	**64**	**66**	**68**	**70**	**72**	**74**	**76**	**78**	**80**	**82**	**84**	**86**	**88**	**90**
5cm	6.3	6.4	6.3	6.1	6.1	5.8	5.8	5.6	5.7	5.7	5.3	5.3	5.4	5.4	5.3	5.1
7.5cm	9.6	9.4	9.2	8.8	8.9	8.9	8.9	8.5	8.4	8.3	7.8	7.9	7.7	7.6	7.4	7.3
10cm	11.4	10.9	11	10.9	10.8	10.6	10.7	10.8	10.6	10.7	10.5	10.2	10.3	10.1	10	10
12.5cm	13.8	13.5	13.4	13.2	13.2	13.1	13.2	13.1	13.0	13.0	12.6	13.0	12.8	12.9	12.6	12.5
15cm	16.9	16.5	16.4	16.3	16.2	16.2	15.7	15.8	15.6	15.3	15.3	15.6	15.3	15.7	15	15.2
17.5cm	19.4	18.8	18.8	18.4	18.2	17.9	17.8	17	17.5	16.9	17.7	17.7	17.7	17.4	17.3	17.4
20cm	22.6	22.3	21.9	22	22	21.9	21.3	21.2	20.8	20.7	20.1	20.6	19.9	20.1	20.1	20.3
22.5cm	25.1	24.6	24.7	24	23.9	23.5	23.1	22.7	22.5	22.6	22.3	22.1	22.4	22.5	22.4	22.6
25cm	26.9	27.0	26.8	26.4	26.0	25.5	25.4	25.6	25.5	25.5	24.8	25.1	25.2	25.1	25.1	25.0
Distance	**Degree**															
	92	**94**	**96**	**98**	**100**	**102**	**104**	**106**	**108**	**110**	**112**	**114**	**116**	**118**	**120**	
5cm	4.9	5.2	5.4	5.4	5.4	5.3	5.4	5.7	5.8	5.8	6	6.1	6.2	6.2	6.3	
7.5cm	7.4	7.4	7.8	7.9	7.9	8	8.2	8.2	8.5	8.7	9	9.5	9.6	9.6	9.7	
10cm	9.9	10.1	10.2	10.4	10.2	10.2	10.3	10.5	10.7	10.9	11.3	11.4	11.5	11.9	11.9	
12.5cm	12.6	12.6	12.7	12.6	12.7	13.2	13.3	13.4	13.3	13.5	13.7	13.8	14.1	14.2	14.4	
15cm	15.3	15.2	15.6	15.6	15.3	15.4	15.5	15.4	15.6	15.7	15.7	16.2	16.1	16.4	16.7	
17.5cm	17.4	17.4	17.6	17.8	17.9	17.5	17.8	17.9	18.2	18.7	18.6	18.7	19.2	19.3	19.5	
20cm	20	20.1	21	21	21.8	22	22.2	22.6	22.7	22.8	23	23.1	23.3	23.5	23.5	
22.5cm	22	22.2	23.1	23.6	23.5	23.6	23.7	23.7	23.6	24.1	24.2	24.5	24.3	24.8	25.0	
25cm	25	25.1	25.1	25.2	25.1	25.3	25.7	25.6	25.7	25.9	26.1	26.2	26.5	26.5	26.6	

Across the 279 grid cells, the readings deviated from the on-axis value in almost every cell, and the mean absolute bias was 0.84 cm with a maximum of 3.20 cm (Figure 4A). The bias varied primarily with angle: it averaged 0.13 cm within the 86° to 94° band, rising to 1.34 cm between 60° and 70° and to 1.62 cm between 110° and 120°. Values recorded at the same angle were similar across distances, whereas values recorded at the same distance changed substantially with angle, and the largest bias, 3.20 cm, occurred at 118° and 120° at a prescribed distance of 20 cm.

Vol. 2026 (11) 863 – 869
ISSN 2688-3651 [online]

The bias field was not symmetric about 90°, averaging 0.77 cm on the 60° to 88° side and 0.96 cm on the 92° to 120° side.

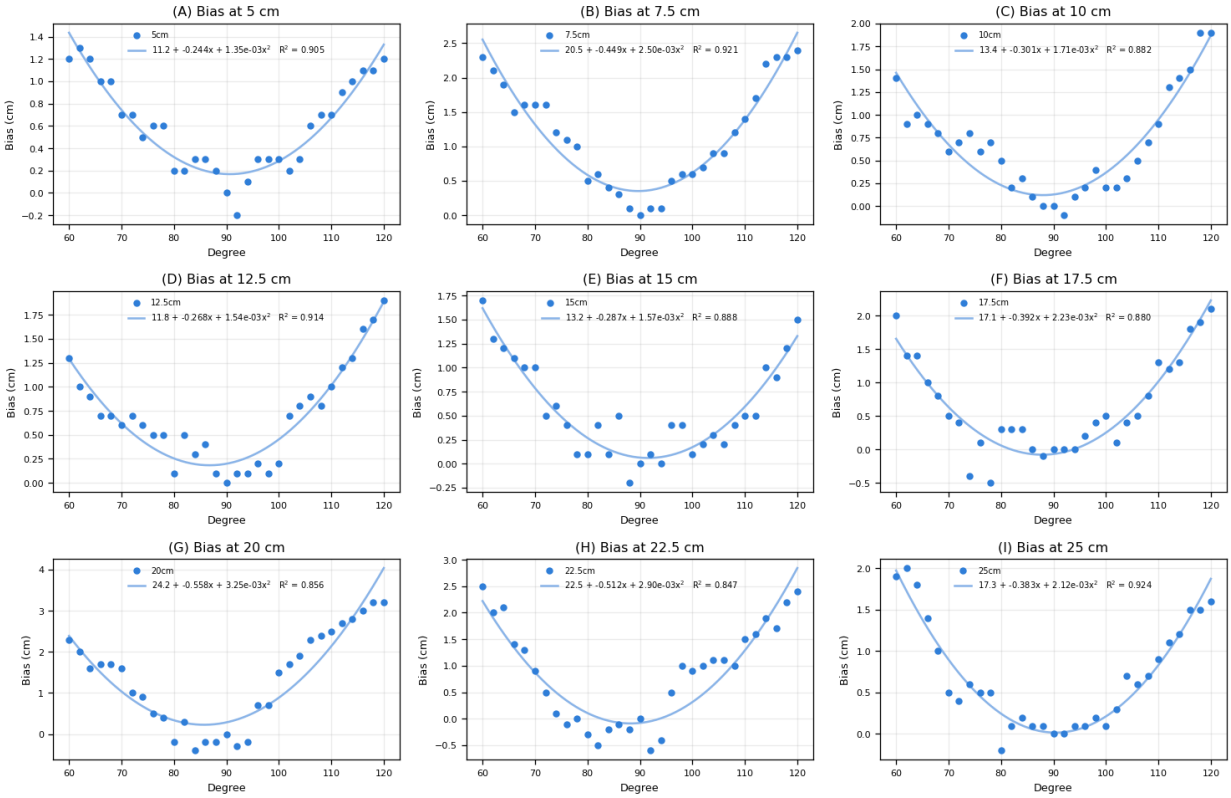


Figure 3. Bias versus angle at each prescribed distance, with the fitted quadratic regression equation and R^2 value for each node of the first classification level.

Figure 3 presents the bias plotted against angle for each of the nine prescribed distances, together with the quadratic regression fitted at the first classification level. In every panel the bias followed a U-shaped curve with its minimum near 90°, and the coefficient of determination ranged from 0.847 at 22.5 cm to 0.924 at 25 cm.

Splitting each distance node further by angle raised the coefficient of determination to between 0.85 and 0.97, with 15 measurements in the $60° \leq \theta < 90°$ node and 16 in the $90° \leq \theta < 120°$ node. These terminal equations were the ones used for correction. In cross-validation they predicted the bias of a withheld cell with a mean absolute error of 0.16 cm under leave-one-out cross-validation and 0.16 cm under the two-fold interleaved split, compared with 0.84 cm for the uncorrected readings.

Applying the SCART correction reduced the mean absolute error from 0.84 cm to 0.16 cm and the maximum error from 3.20 cm to 0.88 cm under leave-one-out cross-validation, a 73% reduction in the maximum error, and 99.6% of the grid cells were left with an error below 0.8 cm, compared with 55.9% before correction (Figure 4C). When the equations were fitted to the complete dataset the corresponding values were 0.13 cm and 0.55 cm. The angle-only correction, which pooled all nine distances, reduced the mean absolute error to 0.32 cm and the maximum error to 1.44 cm (Figure 4B), leaving its largest residual errors at 20 cm, the distance at which the bias departed most from the pooled average.

5. Discussion

The results indicate that the dominant source of error was the geometry of the reflection rather than the distance to the object. Because the sensor transmits from one channel and receives on another, separated by 2.54 mm, the outgoing and returning paths are not identical, and the difference between them grows as the object moves away from

Vol. 2026 (11) 863 – 869
ISSN 2688-3651 [online]

the sensor axis. At angles far from 90°, a smaller fraction of the emitted wave is reflected back toward the receiver and the detected pulse corresponds to a longer travel path, which the sensor converts into an overestimated distance. This mechanism is consistent with the observed pattern, in which the error was near zero at 90° and increased as the angle departed from the sensor axis while remaining almost unchanged when only the distance was varied.

The bias field was slightly larger on the 92° to 120° side than on the 60° to 88° side, 0.96 cm compared with 0.77 cm. A likely explanation is that the trigger and echo channels are not arranged symmetrically about the center of the module, so reflection from one side of the thin cylindrical target returned less energy than reflection from the other. Testing this interpretation would require targets of different shapes and orientations.

The comparison between the two correction schemes shows where the improvement comes from. Correcting with angle alone removed a little more than half of the maximum error, because a single pair of curves cannot follow nine bias profiles that differ in magnitude; its residuals were largest at 20 cm, where the bias was roughly twice the average of the other distances. Classifying by distance before fitting the angular regression reduced the maximum error from 1.44 cm to 0.88 cm and the mean from 0.32 cm to 0.16 cm. The classification step of the SCART model, rather than the regression itself, is therefore what accounts for most of the accuracy gained, which supports the choice of a supervised tree over a single global fit.

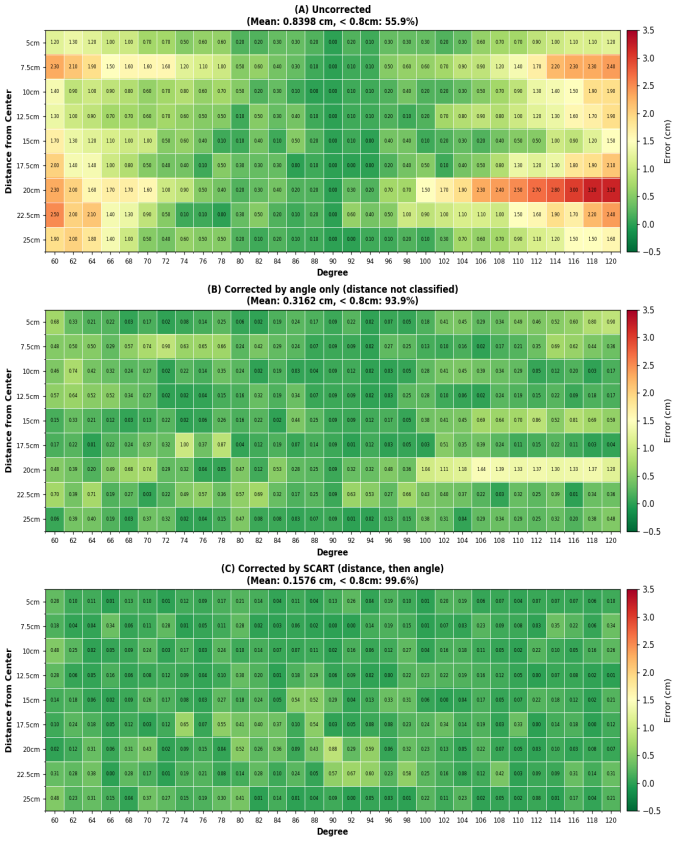


Figure 4. Absolute error distribution over the directional grid: (A) uncorrected readings, (B) after correction by angle alone, and (C) after correction by the SCART model, which classifies by distance before angle. Panels B and C show cross-validated errors.

These findings can be compared quantitatively with earlier reports. Smaili et al. (2021) measured the HC-SR04 against a reference rule and found a standard deviation of about 1.6% between the sensor readings and the true distances, with no deviation in the angle measurements themselves. The maximum error observed in the present study, 3.2 cm at a prescribed distance of 20 cm, corresponds to about 16% of the true distance, an order of magnitude larger, which indicates that characterizations performed near the sensor axis substantially understate the error encountered off axis. Carotenuto et al. (2024) noted that time-of-arrival estimates based on the envelope peak remain subject to errors on the scale of the acoustic wavelength even at high signal-to-noise ratio; the maximum error measured here is roughly four times the 0.86 cm wavelength of a 40 kHz pulse in air, consistent with that scale. Aliew (2022) reached an accuracy of about 0.4 mm, but only by operating on the raw echo waveform with a smart filter signal detection algorithm and by compensating for air temperature, which requires access to the signal that a packaged HC-SR04 module does not expose. The correction developed here works on the distance value that the module already reports and therefore trades absolute accuracy for the ability to be applied to hardware that is already deployed. Qiu and Lu (2022) identified reflection geometry as an outstanding challenge in ultrasonic ranging, and the present results show that a substantial part of it can be addressed in software, which also offers one way to narrow the simulation gap reported by Van Scoy et al. (2023).

Several limitations should be noted. The residual error after correction, 0.16 cm on average, is close to the 0.1 cm resolution at which the sensor reports distance, so the accuracy achieved should be read as approaching the reporting

Vol. 2026 (11) 863 – 869
ISSN 2688-3651 [online]

limit of the device rather than as an unlimited improvement. The correction removes the deviation caused by angle but not the fixed offset of the sensor at 90°, which ranged from −0.2 cm to 0.3 cm and would require a separate calibration against a reference length. The grid resolution of 2° and 2.5 cm limits how finely the correction can vary, and all measurements used a single sensor unit, a single target, and one set of ambient conditions, so unit-to-unit variation, target material and shape, and the temperature dependence of the speed of sound were not examined.

The practical implication is that a low-cost ultrasonic sensor can be made several times more accurate without any additional hardware, using a calibration procedure that a student can reproduce in a few hours and that requires only the sensor itself and a printed grid. Future work should extend the grid to a second sensor unit and to targets of different shapes, refine the angular resolution beyond 110°, and compare the SCART model with ensemble methods such as Random Forest or XGBoost.

6. Conclusion

This study measured the directional error of a low-cost HC-SR04 ultrasonic sensor over a 31 by 9 grid of angles and distances and corrected it using a supervised Classification and Regression Tree model. The uncorrected readings deviated from the on-axis value by 0.84 cm on average and by as much as 3.20 cm, and the deviation depended strongly on angle and only weakly on distance. Classifying the measurements by distance and then by angle and fitting quadratic regressions within each node produced R^2 values between 0.85 and 0.97 and, under leave-one-out cross-validation, reduced the mean absolute error to 0.16 cm and the maximum error to 0.88 cm, a 73% reduction. An otherwise identical correction that omitted the classification by distance reached only 1.44 cm, which identifies the classification step as the source of most of the improvement. The main contribution of this work is the demonstration that a small, interpretable machine learning model trained on data the sensor itself produces can recover most of the accuracy lost to reflection geometry, using no additional hardware.

References

Aliew, F. (2022). An approach for precise distance measuring using ultrasonic sensors. *Engineering Proceedings*, 24(8). https://doi.org/10.3390/IECMA2022-12901

Carotenuto, R., Iero, D., & Merenda, M. (2024). A method for correcting signal aberrations in ultrasonic indoor positioning. *Sensors*, 24(6), 2017. https://doi.org/10.3390/s24062017

Elegoo Inc. (2017). *The most complete starter kit tutorial for MEGA2560* (Version 1.0.17.7.9). Elegoo Inc.

Qiu, Z., & Lu, Y. (2022). Review of ultrasonic ranging methods and their current challenges. *Micromachines*, 13(4), 520. https://doi.org/10.3390/mi13040520

Smaili, F., Beqiri, P., & Lubovci, R. (2021). Errors in distance and angle measurements of ultrasonic sensor HC-SR04. *UBT International Conference*, 298.

Van Scoy, B., et al. (2023). Low-cost distance sensor characterization of ultrasonic and infrared to evaluate the "reality gap" in robot system simulation. In *Proceedings of the 10th International Conference of Control Systems, and Robotics (CDSR'23)*, Ottawa, Canada, Paper No. 202. https://doi.org/10.11159/cdsr23.202