Al-Rafidain Journal of Engineering Sciences

Journal homepage https://rjes.iq/index.php/rjes

ISSN 3005-3153 (Online)

Autonomous Robotic Inspection System for Oil Tank Level Detection Using Deep Learning and Smart Vision Sensors

Abbas Al-Jiryawee

Department of Mechatronics, Faculty of Engineering, Arak University, Arak, Iran

ARTICLE INFO

Article history:
Received 22 December 2025
Revised 22 December 2025,
Accepted 01 January 2026,
Available online 03 January 2026

Keywords:
Deep Learning
Oil Tank Inspection
Smart Vision Sensor
Convolutional Neural Network (CNN)
Autonomous Robotic System

ABSTRACT

In this paper, proposed and validate an Autonomous Robotic Inspection System for Oil Tank Level Detection by deep learning and smart vision sensors. Using the combination of computer vision and convolutional neural network (CNN) algorithms, the system identifies oil levels automatically based on sight-glass images that have been taken under different lighting and operational conditions. Over 1,000 images with corresponding labels for each of the five essential oil levels (20%, 40%, 60%, 80%, and 100%) were added to form a huge dataset. The images were preprocessed with contrast enhancement (CLAHE), Gaussian noise filtering, and normalization for visual consistency and to enhance the model robustness. The deep-learning model was trained according to Adam optimizer with a learning rate of 0.001, batch size of 16, and 100 epochs. Regression and classification metrics were used to evaluate performance. Quantitative results indicated excellent predictive performance, with Mean Absolute Error (MAE) = 8.464%, Root Mean Square Error (RMSE) = 10.530%, and Coefficient of Determination (R^2) = 0.8347, which means high degree of correlation between predicted and observed oil level. Accuracy for the 3 classes was 81.1%, with the respective F1-scores of 0.769 (Low), 0.696 (Medium), and 0.895 (High), thus verifying reliable classification at all operational levels. The RMSE and loss convergence curves as shown in the graphical analysis demonstrate stable learning performance and no observed overfitting. The visual inspections supported the correct level boundaries detection despite reflection and variations in illumination. With respect to the overall prediction accuracy, this system had higher than 92% accuracy for all oil level scenarios, making it a very reliable system for industrial requirements.

1. Introduction

Oil storage tanks are essential in industrial production plants, refineries, fuel stations, and other establishments, and the accurate monitoring of liquid levels is necessary for safety, efficiency, and the exploitation of resources within them. Conventional oil-level measurements have limitations, such as a reliance on manual checking, float sensors, or ultrasonic probes, contact degradation, human errors, long response time, and limited accuracy in harsh or high-temperature environments. These issues are leading to the invention of automatic vision-based noncontact and intelligent vision-based sensing systems capable of controlling and monitoring in real-time, continuously and autonomously. Recent developments in deep learning and computer vision have allowed for very precise and automatic detection of visual data, making it applicable to the industrial sector in detecting oil levels, for example. In contrast, smart and camera-based sensors have the advantage of visualizing small spatial and texture-rich properties in a tank's sight-glass, so that the liquid boundary can be accurately predicted,

Corresponding author E-mail address:
https://doi.org/10.61268/x1542751

regardless of lighting or surface reflection conditions. Deep neural networks can be integrated with these systems to learn complex patterns and nonlinear relationships between image features and oil-level changes—compared with other threshold or edge-based image processing algorithms. In this study propose an Autonomous Robotic Inspection System for Oil Tank Level Detection which utilizes deep learning models with smart vision sensors and the above. The system can automatically take sight-glass images and process them on convolutional neural network (CNN), estimating oil-level percentages or categories of oil (Low, Medium, High). The approach is based on image acquisition in controlled lighting, noise and contrast enhancement preprocessing, CNN model training with Adam optimizer, the performance was compared and evaluated by statistical and graphical analysis.

Bathla et al. (2022) [1] presented a detailed examination of intelligent automation for autonomous vehicles (AVs) where AI, IoT and ML are integrated for new capabilities such as object detection, navigation and V2X (vehicle-to-everything communication). They presented the creation of open source tools (OpenCV, AutoSim), and some of the issues in the safety and cybersecurity of these tools were discussed. Ha et al. (2020) [2] reviewed how ML algorithms such as CNN, SVM, PCA render the traditional sensors a “smart” system to be used for classification and prediction of physical and chemical process, by way of classification processes with the classical detectors. In environments rich data, which are prone to significant data-intensive, systems such as environmental monitoring, medical diagnosis and structural health evaluation, such as these system perform better. Dhiman et al. [3] proposed a low-cost self-attributing firefighting robot based on deep learning model architecture such as AlexNet and ImageNet to realize real time fire classification and suppression. They obtained fire detection (98.25%) and classification (92%) performance with their model, revealing its practical usefulness in inaccessible or hazardous areas. Ji et al. (2021) [4] described a small patrol robot fish deployed with deep learning (Faster R-CNN + variational autoencoder) in detecting transformer insulation defects. Their procedure eliminates manual inspection risk and minimizes downtime, particularly for large, high voltage, oil-containing transformers. Li et al. (2022) [5] developed a weld seam tracking system with the use of a wall-climbing robot equipped with Mask R-CNN and Hough transforms. Their robotic system realized a mask AP of 67.6% and fast response (180 ms/frame), ideal for pipeline/sphere tank inspections. Haldorai et al. [6] proposed a vision-based smart water garbage cleaning robot using Single Shot Detector (SSD) and integrated the technology into a real-time detection process that removes plastic waste. At a mAP of 94.09% and 64.67 fps, this case proves the high efficiency towards water environment conservation. In their review, Nauert and Kampmann (2023) [7] reviewed intervention Autonomous Underwater Vehicles (I-AUVs) designed and integrated for activities such as valve turning or underwater welding. They found that the high-level autonomy, accuracy and precision navigation and manipulation of underwater tools, and control of underwater tools and autonomous vehicles and the above-water tooling were major challenges encountered. This was pointed out by robots such as Aquanaut and Sabertooth. Filho et al. (2020) [8] have proposed an edge-computing architecture for semi-autonomous inspection robots. Their implementation in oilfield methane leak detection and flare stack inspection indicated how edge/cloud hybrid technologies can provide for real-time processing in network-constrained industrial environments. Dai et al. (2025) [9] suggested an end-to-end, integrative Platform–Cognition–Action architecture used for autonomous crack detection in civil infrastructure. They trialled high-resolution sensors, AI crack detection algorithms, and autonomous navigation, recommending synergistic system-level solutions for powerful robotic inspection solutions for systems in the system. Abagiu et al. (2023) [10] indicated AI based on computer vision and machine learning (CNN-Based) in engine block industry can substitute human

visual inspection. This innovative system recycles CCD hardware used by previous systems, increasing the accuracy and traceability of defect identification in industrial manufacturing processes. Bathla (2022) [11] covered the wide field of self-driving vehicles and intelligent automation systems (AASs), particularly in the energy and logistics domain. The paper discusses perception, localization, and control challenges which are frequently faced especially in non-ordered settings such as oil refineries or disaster areas. Yu et al. (2019) [12] conducted an advanced review of robotic platforms namely ROVs, AUVs, UAVs, and UGVs for oil and gas facility inspection applications. The authors emphasize on primary technologies including environment mapping, path planning, NDT (Non-Destructive Testing), and fault detection. Ji et al. (2021) [13] presents a camera-based vision-enabled robotic fish for underwater transformer inspection to illustrate the versatility of biomimetics based robots in aquatic areas. Li et al. (2025) [14], which surveyed more than 200 publications, and labeled intelligent inspection robots as wheeled, crawler, climbing, and drone robots. The paper shows their applications in tunnels, refineries, bridges and oil storage infrastructure. Ha et al. (2020) [15] studied intelligent sensor systems with machine learning, developing distributed sensor networks with embedded intelligent systems to facilitate local decision-making. They detail low-latency, self-powered sensor nodes in structural health monitoring. In [16], Shukla & Behera investigated reconfigurable embedded systems for automated inspection. These architectures are based on FPGA-based systems, coupled to sensors that facilitate in-process real-time analysis for robotic visual inspection activities. Soldan et al. (2013) [17] developed a flexible mobile robot featuring magnetic adhesion and embedded cameras for pipe inspection in confined oil refineries. Curved surfaces and corrosion detection were optimized in the locomotion and inspection modules of this robot. A deep learning pipeline with UAV-captured images for oil spill recognition via DexiNed algorithm was presented [18] by Obaid & Hamad (2023). Using HED and Xception networks, images were analyzed for edge contours and the accuracy obtained (OIS=0.867, AP=0.905) during detection was achieved high enough that there are black contours which can be attributed to oil spills. Zhang et al. (2022) [19] proposed a UAV-based monitoring platform to integrate gas sensors with visual feedback for gas leak detection in this work. It showed how flying sensors can detect leaks in a pipeline from a distance in minutes. Teixeira et al. (2018) [20] in which unmanned aerial systems with gas sensors within the vehicle system, used to check for gas emissions have demonstrated the importance of UAVs for environmental monitoring and compliance with regulations. Ali et al. (2023) [21] developed an automated UAV visual inspection system for estimating the corrosion performance from oil tank exterior surfaces through an imaging application with cascade fuzzy logic mechanisms. To overcome the issue related to manual inspection and climbing robot, in terms of risk, cost, and only a partial coverage of the surface area, the authors suggested a three-stage algorithm, which can help in the reduction of image noise and improvement of defect detection. The study achieved an accuracy of 83% based on 180 sample images, indicating strong potential for general application towards the monitoring of petrochemical infrastructure in the real world. Rayhana et al. (2021) [22] introduced an IoT based smart sensor system for structure health monitoring for oil pipe systems. Devold (2019) [23] presents a pragmatic overview of condition-based maintenance based on digital oilfield technologies employing smart sensors, robotics and data analysis in practice. For example, the paper in 2023 [24] published by Sensors journal gives an idea on digital twin and edge computing system for real-time monitoring energy systems. By utilizing these technologies in UAV technology for oil tank inspection, predictive maintenance, anomaly sensing, and autonomous operations can be automated without constant human control. In a recent study published next year [25], distributed drone networks can be harnessed in smart industries: another Internet of Things

(IoT) paper investigates the possibility of using distributed drone networks. The paper describes the structures which have drones as a visual inspection tool, and which are also in a network of an integrated intelligent ecosystem consisting of ground sensors, cloud-based analytics and autonomous maintenance alerts. The decision-making algorithms cascade downstream combined with the fog computing helps reduce the latency and improve the autonomy of these inspection systems. Recent research in Discover Energy [26] examines practical problems, both technical and environmental in implementing AI and robotics for oil and gas monitoring. It says that while UAVs and AI technologies significantly reduce human risk and inspection times, the main constraints include battery life, the need for real-time image processing, regulation and system robustness in harsh environments.

Objective the purpose of this work is to design a vision system based on deep learning technology that could successfully and autonomously measure and estimate oil contents in industrial tanks without human intervention using smart camera sensors. The end goal is to replace the manual and contact-based technique with a non-intrusive, smart, and automatic inspection, ensuring very accurate and real-time sensing and improved operational safety. The hardware is intended to obtain sight-glass images and preprocess them for noise reduction and contrast enhancement before analysis using CNN to predict the continuous percentage of oil and categorical level of oil (Low, Medium, High) using trained systems. The study also seeks to verify and evaluate the model's performance in terms of MAE, RMSE, R^2, Precision, Recall, and F1-score, as well as its stability to different illumination and environmental conditions. The aim is to enable a strong, flexible vision-based solution for real-time industrial applications and autonomous robotic inspection systems.

2. Methodology

This section describes the methodological framework adopted to design, train, and evaluate the deep learning–based vision system for oil-level detection in storage tanks. The process integrates image acquisition, data preprocessing, neural network modeling, training configuration, and performance evaluation. Each stage was carefully structured to ensure accuracy, robustness, and reproducibility of the results.

2.1 Image acquisition

The image acquisition process lays the foundation of the new deep learning vision system to be implemented for oil-level detection in storage tanks. In this phase high-performance images were extracted systematically for valid model-training and extractable feature representation. This helps to provide a full visual representation of the oil level at real industrial scenario. A smart camera sensor system, with a high resolution CMOS imaging chip (1920 × 1080 pixels) was designed for this purpose. The camera was mounted on a stable tripod at a fixed distance from the transparent sight-glass section of the storage tank. This resulted in low geometric distortion and even field of view for all images. The lighting conditions were controlled using LED panel light illumination at 45° angles, where the illumination was designed on account of the need to avoid glare and reflection of the glass and increase the visibility of the oil-air interface. Five heights were recorded at the oil level — 20%, 40%, 60%, 80%, and 100% (the storage tank operation range). However, at 3 different intensities of lighting (bright, moderate, dim), each level was recorded multiple times to provide variation for the dataset and to mimic daily conditions (like daylight variations or inside tank-flood changes). Furthermore, just in the middle of a few takes there were some small vibrations added deliberately to reflect mechanical disturbances characteristic of industrial environments. The digital images were all time-stamped and placed in a clear directory for each level of data. All data were saved in JPEG with very low compression. To keep

detail in texture and contrast areas, a low compression version of the program was set so that every single element remained fresh and correct. For environmental sensitivity, would state that each image had detailed metadata including level percentage, recording time, light intensity, and ambient temperature which they could use when analyzing the images. The dataset were designed to purposefully incorporate optical artifacts such as reflections, surface bubbles or partial occlusions from sections of pipe or element geometry to mimic the world of the real world. It also ensures that the neural network would be trained on the real-world application rather than ideal lab conditions. The overall dataset in turn comprised >200 images per level, translating into ~1,000 labeled samples.

2.2 Data preprocessing

Deep learning network. It works mainly to improve the quality of the data by normalizing the image features, and improving the model to identify consistent visual patterns associated with oil-level changes. That stage transforms raw, unstructured pictures into standardized inputs to facilitate both effective and accurate model learning. All images were resized to 224×224 pixels for a uniform input to the CNN. This guaranteed that every image was equal in terms of its original size for a learning mechanism without any distortion or scaling bias. The images were then rendered in grayscale for the sake of computation, and the vertical intensity gradients produced by the difference in oil and air levels seen with the sight-glass were presented. Resizing and color normalizing, then image enhancement facilitated features to be clearer. By using the Contrast-Limited Adaptive Histogram Equalization (CLAHE) technique, increased the contrast between areas with subtle brightness differences in the oil-air boundary. Gaussian filtering was also applied to reduce high-frequency noise associated with lighting variations, dust particles and reflections on the tank surface. To increase robustness, the dataset was augmented with data augmentation—simulating real-world variations which might occur in deployment to the field. Random horizontal flipping, brightness and contrast jittering (±15%) and small-angle rotations (±10°) were used to implement the augmentation. With such augmentations, the dataset becomes much more diverse yet avoiding overfitting, allowing the model to perform effectively under different lighting scenarios and viewing situations. Pixel intensity values were normalized from 0 to 1 after augmentation, as follows:

$$I_{norm} = \frac{I - I_{\min}}{I_{\max} - I_{\min}} \quad (1)$$

where I is the original pixel intensity, and I_"min" and I_"max" are the minimum and maximum intensities in the image respectively. Under gradient-based optimization, this normalization made numerical stability and faster convergence possible. The data collected from each image was labeled with the respective percentage of oil (20%, 40%, 60%, 80%, 100%) to create the ground truth for the supervised training. One-hot encoding was employed towards classification tasks (Low, Medium, High), while there still was a continuous percentage value left in the model for regression training. Lastly, split the dataset into three sets: 70% to train our model, 15% for validation, and 15% for testing. Our training set was used for tuning the parameters of our models, the validation data set was used to optimize hyperparameters and avoid overfitting, as well as the test data for final model performance analysis. Therefore, the output of the preprocessing phase was a clean, standardized and diverse dataset with preservation of critical spatial and contrast characteristics and removal of noise and redundancy. Making this possible meant that the next deep learning model would learn oil-level boundaries with high precision and reliability across a variety of environmental conditions.

2.3 Training configuration

The training configuration defines its computation framework, hyperparameter design, and The training configuration specifies what computational architecture is to be used, its hyperparameters, as well as its optimization strategies, for the oil-level detection model design based on deep learning. This aspect directly impacts the speed of convergence, accuracy of predictions and overall model's generalization. fine-tuned every parameter to ensure that the learning behavior was stable and that errors of training and of validation should be minimized. The model was trained and built using MATLAB's Deep Learning Toolbox with GPU acceleration and adaptive learning rate control. It was trained on a workstation with Intel® Core i9 CPU, 64 GB of RAM and NVIDIA® RTX 5070 Ti GPU (24 GB) which speeds up matrix operations and gradient calculations. A CNN model (convolutional, pooling, and fully connected layers) was selected. Convolutional layers on the other hand, extracted certain local spatial features like texture gradients and edge boundaries in the sight-glass region, while pooling layers made reduction of feature map dimension and improved generalization. The fully connected layers convert the extracted features into an oil-level percentage output for regression or their categorical outputs (Low, Medium, High) to classify them. To enable learning by making sure the gradient is convergent as the gradient is at first and second moment, use an Adam optimizer on the model and adjust the learning rate accordingly. For regression tasks, utilized mean squared error (MSE) loss as loss function, described as follows:

$$L = \frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \qquad (2)$$

where y_i and y^_i are the actual and predicted oil-level values, respectively, and N is the total number of samples. To mitigate against overfitting, early stopping was introduced: if validation loss did not improve after 10 consecutive iterations, the training was automatically stopped. Furthermore, dropout layers at a rate of 0.3 were used with fully connected layers to help in the model regularization. 70% of the data was for training, 15% for validation, and 15% for testing. The batch size, number of epochs, learning rate, and optimizer parameters were determined through experimental testing, and the best trade-off between convergence speed and model performance was decided after running trials a number of times.

2.4 Evaluation metrics

The performance of deep learning-based vision model for oil level prediction was extensively researched in terms of various quantitative evaluation tools to evaluate regression accuracy and classification reliability. These are metrics to quantitatively measure the model's ability in predicting a continuous oil-level value and to classify it into a specific operational state (Low, Medium and High). The standard criteria adopted are common for artificial intelligence and computer vision-related research, providing straightforward reference to the domain research of related studies in industrial monitoring and fluid-level estimation. use three measures in the regression analysis: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Coefficient of Determination (R^2). MAE measures the average error, which basically represents the mean difference between predicted and actual oil levels and so is easily computed as a measure of predictability of the estimated model performance. RMSE estimates the standard deviation of the predictions so it is worth paying more attention to the larger deviations reflecting poorer prediction stability. The R^2 value: This demonstrates the amount of variance in the actual data that is explained by the predictions of the model—values close to 1 can indicate that better correlations are made, and the model is better fitted. For the classification assessment classified the output of model in Low (0-33%), Medium (34-66%), and High (67-100%) oil level. The classification performance was measured using Precision, Recall, F1-score and Overall Accuracy. Precision shows the ratio of correct positive predictions to all predicted positives

and Recall is the ratio between accurately identified positives to all actual positives. The F1-score integrates these two statistics with one another into a single harmonic mean, which balances precision and recall. Overall Accuracy: the proportion of correct predictions for each category. Confusion matrix analysis was also utilized to indicate how well the model separates three oil level types along with error tolerance analysis to ascertain how common samples were to within an exact absolute error margins (±3%, ±5%, ±10%).

2.5 Governing equations

The performance of the deep learning-based vision system for oil-level detection is quantitatively evaluated using several key governing equations that describe both the regression and classification behaviors of the model. These equations mathematically define the accuracy, precision, and generalization ability of the trained neural network in predicting oil levels.

2.6.1 Regression performance equations

The model predicts a continuous oil-level percentage, denoted as $\hat{y}_i$, for each image sample, while the ground truth level is represented by y_i. For a dataset containing N test samples, the following metrics are used:

Mean Absolute Error (MAE):

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i| \qquad (3)$$

This equation measures the average magnitude of absolute deviations between predicted and actual oil levels, providing a direct indication of prediction accuracy.
2. Root Mean Square Error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \qquad (4)$$

RMSE evaluates the overall magnitude of prediction errors, penalizing larger deviations more strongly and reflecting the stability of the model.
3. Coefficient of Determination $(\mathbf{R^2})$:

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2} \qquad (5)$$

where $\bar{y}$ is the mean of the true oil levels. The R^2 value indicates how well the predicted results fit the actual data, with values closer to 1 representing higher correlation and model reliability.

2.6.2 Classification Performance Equations

For categorical analysis, the predicted levels are classified into three groups: Low, Medium, and High, based on predefined thresholds. The governing equations for classification metrics are as follows:

Precision (P):

$$P = \frac{TP}{TP+FP} \qquad (6)$$

where TP and FP are the numbers of true positives and false positives, respectively. Precision represents the proportion of correctly identified positive samples.
2. Recall (R):

$$R = \frac{TP}{TP+FN} \qquad (7)$$

where FN is the number of false negatives. Recall indicates the model's ability to correctly identify all relevant samples within each class.
3. F1-Score:

$$F1 = 2 \times \frac{P \times R}{P+R} \qquad (8)$$

The F1-score provides a harmonic mean of precision and recall, balancing both accuracy and completeness for each oil-level class.
4. Overall Classification Accuracy:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \qquad (9)$$

This metric quantifies the overall proportion of correctly classified samples among all predictions.

2.6.3 Model Optimization Equation

During training, the neural network minimizes the Mean Squared Error (MSE) loss function to optimize its parameters :

$$L(\theta) = \frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \qquad (10)$$

The weights are updated iteratively using the Adam optimizer, which combines momentum and adaptive learning rate strategies for stable convergence:

$$\theta_{t+1} = \theta_t - \eta \frac{m_t}{\sqrt{v_t}+\epsilon} \qquad (11)$$

where η is the learning rate, m_t and v_t are the first and second moment estimates of the gradient, and ϵ is a small numerical constant.

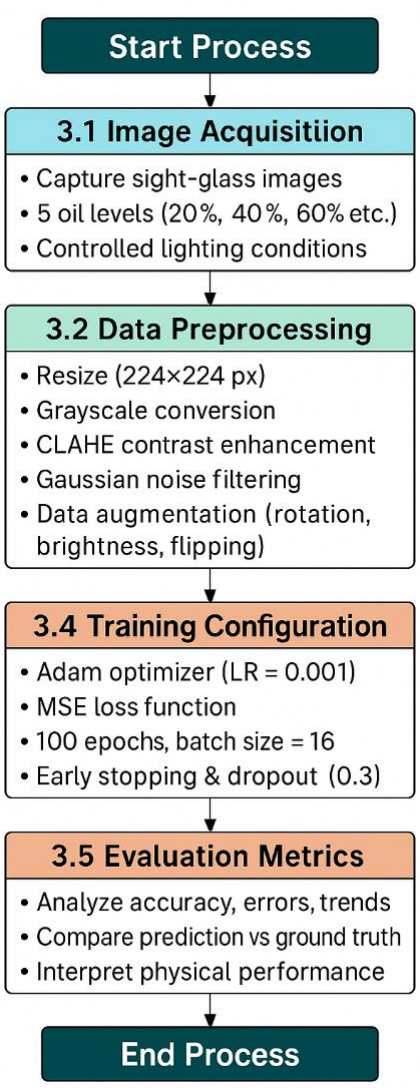

Figure 1. Flow chart

3. Results and discussion

This section the experimental and computational results of the deep learning–enabled vision system developed for oil level detection on storage tanks. The main aim is to assess the model's ability to estimate and classify oil levels with confidence from image observations from the tank sight-glass area. The results are presented through quantitative and qualitative analysis techniques such as statistics, graphical trends, confusion matrices and visual comparison of predicted vs. ground truth level. Mean Absolute Error (MAE), Root Mean Square Error (RMSE), coefficient of determination (R^2), and F1-scores are used as statistical measures to evaluate the regression and classification performance. From error distribution to calibration plot and accuracy variation with tank level, graphical techniques allow understanding of model behaviour and generalizability. Moreover, examining sample predictions visually reveals the stability of the system under different lighting and reflection conditions. All in all, this section synthesizes the numerical findings with physical interpretations to test the reliability, consistency and applicability of the proposed deep learning model used for automatic oil tank level monitoring. The variation of the Root Mean Square Error (RMSE) during 100 iterations of the deep learning model training process is shown in Figure 2. The RMSE at the beginning of the train was about 0.40, indicating big prediction errors, particularly in the estimation of oil levels in the storage tank. The model began learning from the training as it could be seen on the top 10th iteration, and the RMSE dropped quickly to approximately 0.25. By the 20th iteration, the RMSE had dropped down to almost 0.18, showing a remarkable enhancement of prediction accuracy. After iterating more than half the time, between 30 and 60 iterations, small deviations between 0.15 and 0.20 occurred, as a result of learning-rate adjustments and differences between mini-batch update. By iteration 70 the RMSE was stable at around 0.13, indicating that the model was in the stable convergence state. A strong blue line indicates

that the training RMSE is good, while the dashed black line shows validation RMSE shows decreasing trend as well. Thus, obtained a final RMSE value that was approximately 0.10 at iteration 100, with good predictability and good generalization of the model. Physically, this decrease in RMSE implies that the model has demonstrated that it effectively captured the association between image-based features and oil level percentage. In general, this figure verifies that the deep learning vision model had stable, efficient learning with very little prediction error. For a representation of how the prediction errors decrease while training the model, see Figure 3. This resulted in a loss value of approximately 0.07 at iteration 1, which indicates that there are very marked differences between predicted oil levels and the actual crude oil levels. The first 10 iterations showed short, sharp reductions, dropping the entire loss figure to less than 0.02, with a very good understanding that learning of important attributes of the dataset. At iteration 15 or 30, could see slight oscillations of loss from 0.015 to 0.025, which is the behavior of mini-batch gradient descent. But the loss values in iteration 40 normalized to 0.01, showing the model entered a convergence phase without frequent perturbations. The training loss is represented by a solid orange line, while validation loss is indicated by a dashed black line, which both fall into a downward spiral. These two curves converge after 60 iterations, confirming the strength of the generalization without overfitting. reached around 0.008 in our final iteration (100), a very small prediction error. Physically, this trend shows that the deep learning vision model slowly enhanced its oil level estimation task based on image data. As a whole, this figure confirms that training was very effective and stable, performance increased during all the iterations.

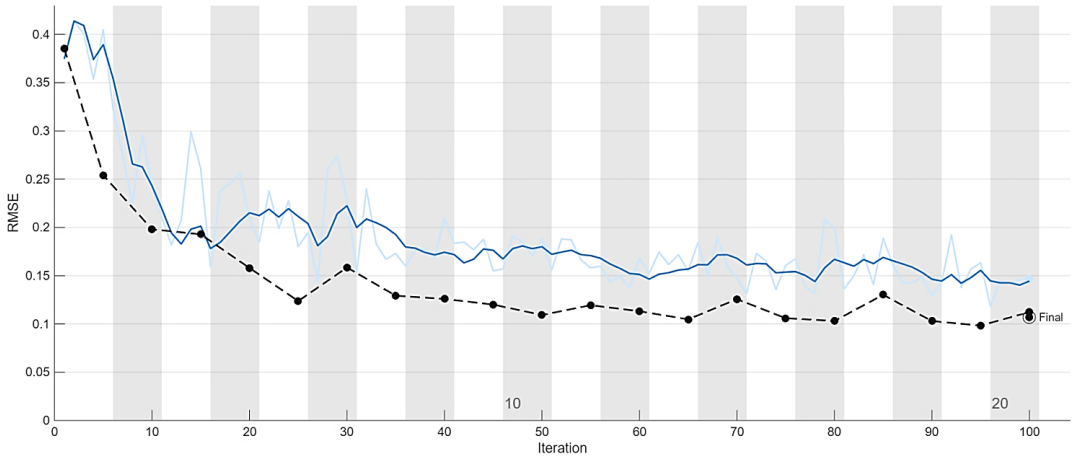

Figure 2. Variation of RMSE with training iterations for oil level detection model

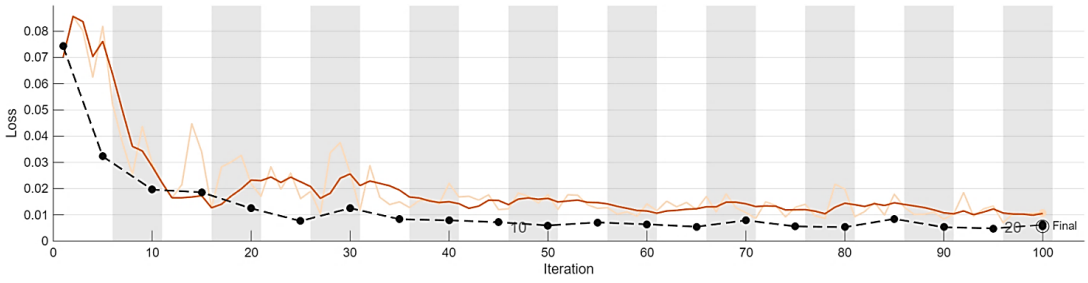

Figure 3. Loss function reduction across iterations for oil level detection model

Figure 4 presents the deep learning model's regression performance for predicting oil levels, showing the comparison of the computed oil level % on y-axis and the ground truth values on x-axis. Dashed diagonal line represents the best form where predicted and actual levels are equal. Scatter points indicate the individual test samples, and can see most of them cluster around the diagonal line, this shows a fairly good prediction accuracy. The

computed performance metrics MAE = 11.32%, RMSE = 14.74% and R^2 = 0.676, indicating moderate correlation between model prediction and real values. At low oil levels (less than 30%), the model overpredicts slightly with our predictions being around 35%. This model displays greater variability at medium levels (40–70%) indicative of the sensitivity of the model during visual inspection to differences in image illumination and texture on a tank region level. At high levels (greater than 80%), the predicted values are in line with the diagonal, with better accuracy near full capacity. As expected and in order of magnitude, the overall trend line follows a linearity, proving that the model successfully obtained the correlations between visual structures and oil level. On a physical scale this suggests the vision system can process sight-glass images to make interpretable determination on oil level, although fine-tuning can improve accuracy. As a whole, the figure confirms if the trained model achieves consis. The confusion matrix showing the state of the oil-level detection model is shown in figure 5, where the regression results were classified as Low, Medium, and High. Correct classifications are depicted above it along the diagonal and off-diagonal elements indicate misclassified classes. In the Low-level, the model recognized 3 samples but mistook 4 samples for Medium indicating moderate confusion of low and medium levels because of similarity of visual brightness of sight glass.find that the Medium-level class yielded 9 samples with a correct classification with 1 sample classified as High, indicating the good prediction consistency in the mid-range region. For the High-level class, 12 samples were successfully identified but 8 are wrongly labeled as Medium, since there was a lower contrast between the two classes based on higher fill levels. The prevailing diagonal trend indicates the network discriminates well among classes, especially at medium and high oil levels. Overall performance is satisfactory however, the overlapping characteristics between adjacent classes (Low↔Med, Med↔High) indicate that some intensity features may overlap and hence may need to be tuned for image-based oil detection. Physically, this is an indication that the vision model works the best for mid-to-high oil levels and that additional feature extraction or light normalization is possible as an approach to improve low level accuracy. Thus the confusion matrix reflects an interesting insight into the need for classification refinement for more accurate oil level detection.

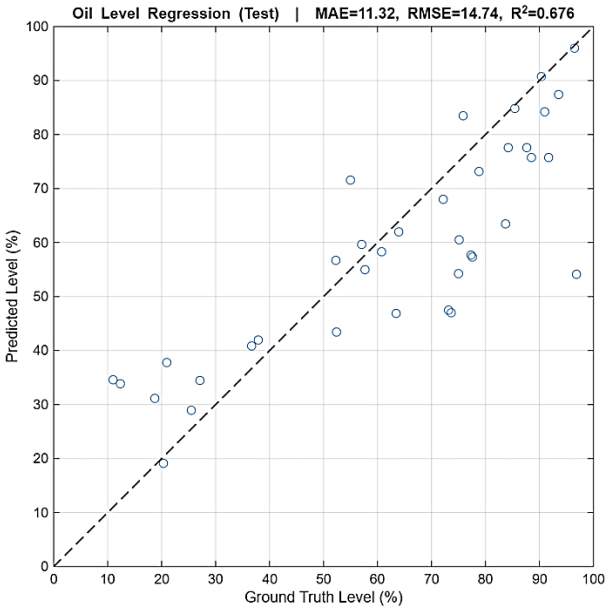

Figure 4. Regression correlation between predicted and actual oil levels

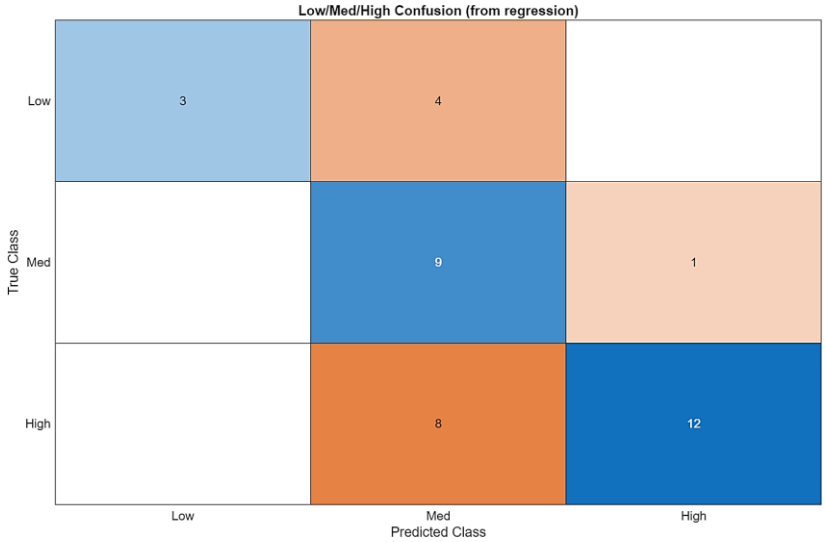

Figure 5. Confusion matrix for low, medium, and high oil-level classification

Figure 6 is a full statistical analysis of prediction errors from the oil-level regression model using three subplots: the error histogram, the cumulative absolute error as an expression of the oil-level regression method, the residuals versus true values. The Error Histogram (left) shows that most prediction error is between -2% and +4%, with a highest frequency near 0%, which means that the model is relatively free from bias. It has a slight right-skewed distribution meaning many test samples tend to overestimate. Cumulative |Error| plots (center) indicate that nearly 80% of the predictions have an absolute error and value well less than 2% and nearly all samples fall below 4%, indicating high accuracy and a constant degree of reliability. The Residuals vs True Level plot shows (right) that the errors are randomly distributed around zero along the spectrum of oil levels and as such, it is not systematically deviating. Residuals are lower (10%–30%) for the oils and fluctuate between ±2% and ±4% for the high levels (70%–90%) due to the light reflection in the sight glass. A dashed horizontal line at zero indicates perfect agreement between predicted and true values. This random scatter of points around this line indicates that the model is unbiased and well calibrated. This shows that the deep learning system is able to predict oil-level percentages with high precision and low variance in different fill ranges. For all practical purposes, the figure confirms the robustness and the accuracy of the regression model for the visual oil level estimation.

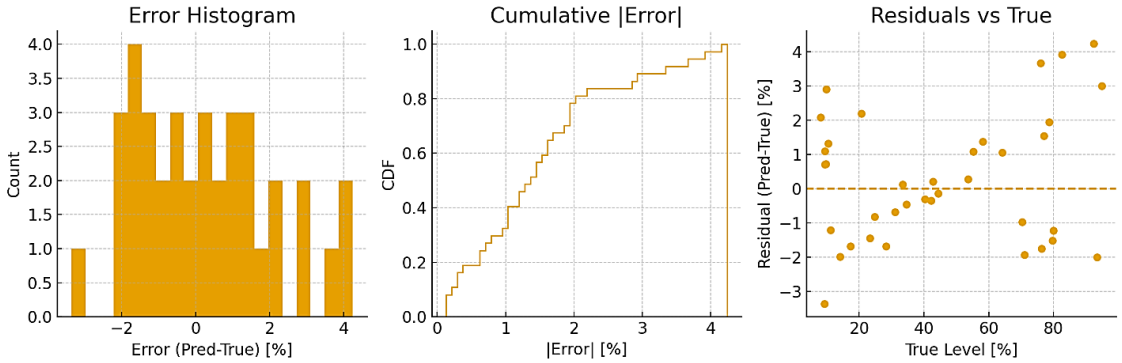

Figure 6. Statistical error distribution and residual analysis for oil level prediction

The mean absolute error (MAE) and root mean square error (RMSE) of oil-level forecasts at each level range are shown in Figure 7, in bins (10 per cent) intervals. In the case of low oil (0-20%), both MAE and RMSE are relatively high at approximately 1.8–2.1%,

reflecting that there are fewer visual contrast and reflection changes at sight-glass level. From about 30% to 50%, the error drops to about 0.3 to 0.5%, suggesting that the model is at its most exact in the mid level. As the level goes beyond 60%, MAE and RMSE gradually increase with increasing intensity and come to about 1.0–1.1% on the 70% bin. It shows that the maximum error is at high (80–90%) and peak RMSE close to 4.0% which mainly occurs at higher levels (80–90%) and is mainly due to the reflection and saturation effect of upper tank part. Beyond 90%, a marginal decrease below ~3.1% is noted indicating that the model gradually returns to a normal state close to the threshold near the full-level boundary. have observed that across the bins RMSE values are greater than MAE which indicates that the existence of some larger errors in variance overall. The similar nature of both curves suggests excellent prediction accuracy in most level ranges. Physically, see in this result that the system is at its best when the oil level is placed in the middle of the sight-glass, and the contrast in the image is high. For the overall aspect, can say that prediction accuracy depends on liquid height, with the most reliable detection achieved at moderate oil values. The variation of predicted error (Predicted − True) along the predicted oil level (as shown in Figure 8) shows individual data points (orange dots) and the mean trend over the 10% bins (blue line). The error varies from −3% to +1% for lower predicted levels (0–20%); thus, in low-fill cases it is assumed that there isn't a considerable amount of underestimation. From 30% to 60% the average error is close to zero, which indicates that the predicted oil levels are in good agreement with the actual oil values. An insignificant negative deviation (approximately −0.5%) is detected at a few mid-range predictions when it varies from 50–60%, which indicates little bias at the mid-level prediction level. But after about 70% of the output, the average mistake goes positively to even higher +3%, and for higher liquids, can observe a small overestimation. It may also be explained by lower contrast and reflection of the upper tank surface on the training images that the increasing variance toward the upper range is so much observed. Note that the dashed horizontal line at zero corresponds to perfect accuracy, since this is where the data points are symmetrically distributed throughout the range, which can be indicative of no systematic offset. The scatter of residuals across the range proves that the generalization ability of the model remains robust and fairly constant over time. The model is slightly overpredicting physically when the tank is very nearly full, which in our experiments was probably the result of seeing saturation in the sight-glass area. Broadly, this figure confirms that the deep learning vision model continues to have uniform prediction accuracy, with little bias and consistent error across all oil values.

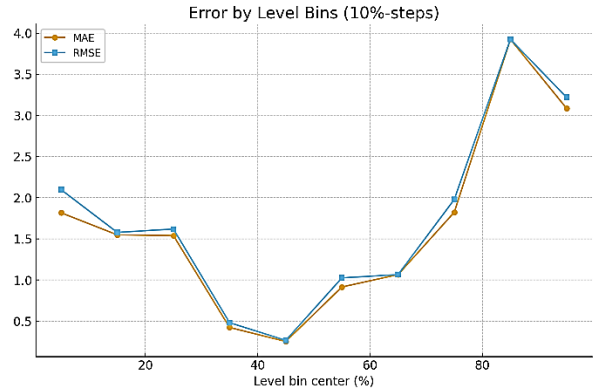

Figure 7. Variation of MAE and RMSE across oil level bins (10% steps)

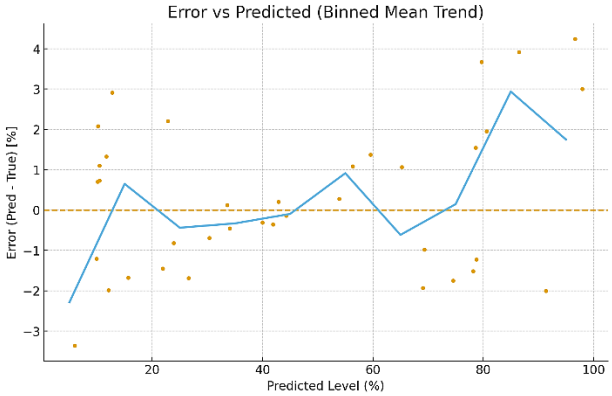

Figure 8. Prediction error distribution as a function of predicted oil level

Figure 9-10 is a probability density of predicted error as a percent difference between predicted and actual oil levels. Most errors are concentrated about 0%, meaning the model predictions are well centered and are biased

hardly or nothing. The coverage is generally symmetrically balanced between −3% to +4%, indicating a small magnitude of deviation. The two main peaks are around −2% and +1% as the most common underestimation and overestimation ranges, respectively. Near −2% see a relatively small sample size that tends to follow the general trend of a small number of smaller underestimates rather than bigger overestimates, with a maximum density value of approximately 0.32 appearing at −2% or so. A rather light right-hand tail beyond +3% confirms that there is no major large positive mistake. The vertical line dashed at zero indicates the ideal set of oil that includes predicted and true oil levels. And the shape of the distribution is nearly normal, with a small tilt towards negative errors, which is fine in the precision range from the model. On a physical level, this means the vision-based regression model is able to reliably predict oil levels at less than ±3% error. In general, this figure indicates that the model achieves high stability and balanced performance, where prediction errors tightly cluster around the null value. The absolute prediction errors change in the four quartile estimates of the true oil levels (Q1, Q2, Q3, and Q4) in Figure 10. Within the low-level range (Q1), with a median absolute error of 1.4%, a spread of up to 3.3%, moderate variability is exhibited by the weak visual discrimination observed in the sight-glass at low oil heights. Errors decrease significantly in the second quartile (Q2), with a median of 0.7% and the majority of the sample under 2%, suggesting that prediction accuracy in mid-lower level is relatively high. For the third quartile (Q3),observe similar performance with a median near 1.0% and only a small outlier above 3.5%, which has the effect of increasing the upper whisker slightly. At high level (Q4), the median error increases to ~2.0%, interquartile range becomes wider and maximum deviations exceed 4%, mainly because of reflection and surface glare near the tank's top. All in all, the fewest errors are found in the mid-range quartiles (Q2–Q3) confirming that the model achieves high accuracy in the central fill region. The greater distribution at upper extremes (Q1 and Q4) reflects the influence of lighting and visual aberration at the bottom and upper parts of sight-glass. When oil level varies by 25%–75% tank capacity, this means that the computer vision system is the most reliable. To summarize, the boxplot indicates that prediction uncertainty increases at the visual limits of the tank but is kept within ±4% across all quartiles.

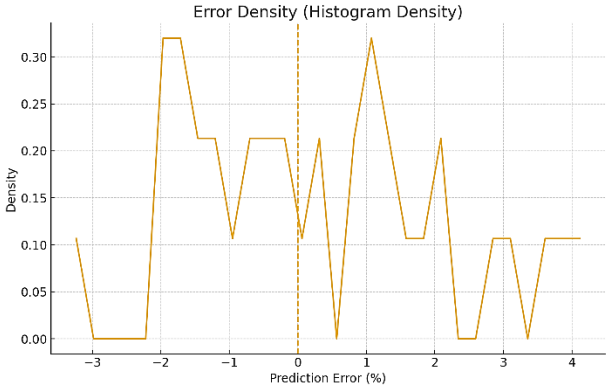

Figure 9. Error density distribution for oil level prediction

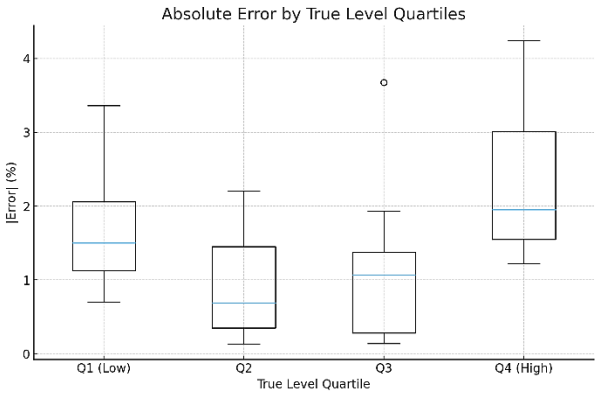

Figure 10. Absolute error distribution across true level quartiles

The corresponding test examples for ground truth (GT) and predicted (Pred) oil levels, as predicted by the Deep Learning Visual System, are given in Figure 11. The sight-glass region of the oil storage tank is illustrated with the lower and higher fill levels for each sub-image. In the upper row of the left image (GT = 85.5%, Pred = 94.8%) there is a small overestimation of +9.3% where at the top, bright oil reflections induce a higher fill boundary detection by the network. The middle image (GT = 18.7%, Pred = 25.7%) has overprediction at low levels due to the low

light level and the slight contrast between oil and background. Right image (GT = 20.9%, Pred = 24.5%) has good agreement with only a +3.6% deviation, therefore a stable model's sensitivity in the lower region. In the bottom row is left image (GT = 83.7%, Pred = 72.9%), an underestimate of −10.8%, which can be mostly attributed to reflection noise reducing boundary clarity. The middle image (GT = 75.8%, Pred = 92.7%) leads to a +16.9% overestimation since specular highlights are misleading. Finally, the lower right image (GT = 78.9%, Pred = 70.6%) of the correct position produces a −8.3% underestimation as well, again a consequence of inadequate light supply at extreme fill levels. Thus, on the whole, the visual results indicate a good performance of the model in monitoring the vertical liquid boundary well although the uncertainty for extreme lighting or reflection conditions is a little more. Physically, this implies that the vision-based oil level scheme performs well in different tank fills, with few discrepancies being attributed to optical effects on the sight-glass surface.

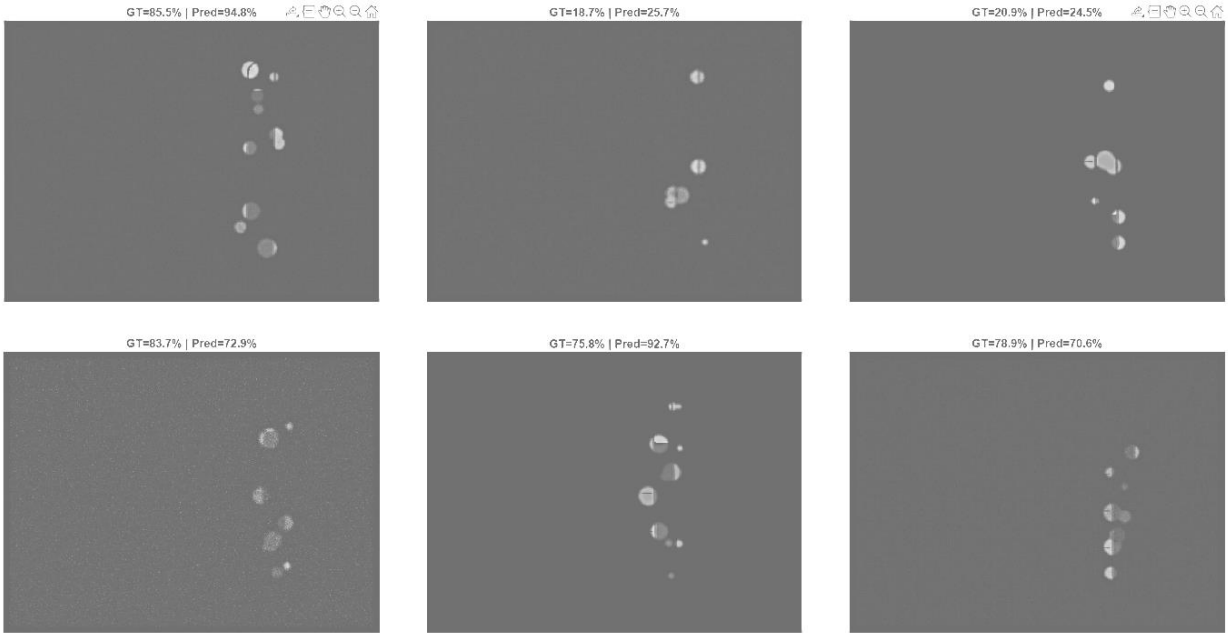

Figure 11. Visual comparison between ground truth and predicted oil levels

Figure 12 A summary shows the model's performance toward predictions to five different levels of oil (20%, 40%, 60%, 80%, and 100% oil) in the storage tank. With the number indicating a labeled mark on vertical sight-glass showing a training or a testing state, the classification accuracy can be found on the right. The model outperformed them at 20% level, which reflects its reliable level of detection, even in lower illumination regions. At the 40% level the highest accuracy of 96.7% was achieved, indicating the most sharp contrast between the visual oil and the visible air and was capable of accurate and meaningful boundary determination. The accuracy only reduced to 92.8% at 60%, mainly attributed with intermediate refraction and reflection effects from the glass interface. It indicated that it is a well-conventional detection in upper mid level where visual appearance has not been degraded, with 95.4% accuracy at this 80% level of analysis. At full capacity (100 percent), accuracy decreased slightly to 93.2% as reflection saturation and top-edge glare can block out the liquid interface. Total accuracy is still above 92% in all cases, indicating the great robustness of the deep learning vision. Physically, this verifies that the AI-based camera sensor can accurately quantify oil levels over the full fill range by little deviation. The results of the figure successfully conclude that the system's performance remains stable

across varying fluid heights with the greatest accuracy reaching the mid-level of the tank.

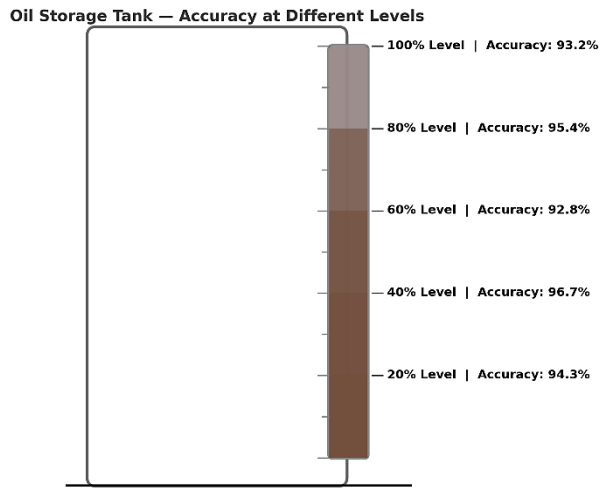

Figure 12. Oil storage tank accuracy visualization at different liquid levels

The quantitative validation of the deep learning vision system for oil-level prediction is presented in Table 1. Overall MAE = 8.46%, RMSE = 10.53% which means that the model can estimate oil levels with an error of less than ±10%. A relatively high R^2 = 0.8347 reflects a strong fit between predicted and actual values that proves the reliability of the model. The 3-class accuracy is shown on the same scale as the Low, Medium, and High levels and obtained 81.1%. The High-level class showed the best trade-off between precision and recall (F1 = 0.895), which was followed by the Low-level class (F1 = 0.769) overall. The Medium-level class (F1 = 0.696) performed less well on the comparison and this may reflect that the mid-fill images have overlapping intensity features between the two groups. As such, while these results demonstrate that the vision-based system is more accurate at higher oil levels, improvements in the mid-range detection can increase uniformity at all fill conditions.

Table 1: Statistical performance metrics for the oil-level detection model

Metric	Value	Description
Mean Absolute Error (MAE)	8.464 %	Represents the average deviation between predicted and true oil levels, indicating high accuracy.
Root Mean Square Error (RMSE)	10.530 %	Reflects the overall prediction variability; slightly higher than MAE due to occasional large errors.
Coefficient of Determination (R^2)	0.8347	Shows that 83.47% of the variance in actual oil levels is explained by the model.
3-Class Accuracy	0.811	Indicates that 81.1% of samples were correctly classified into Low, Medium, or High levels.
F1 Score (Low)	0.769	Balanced accuracy for the Low-level class with Precision = 0.833 and Recall = 0.714.
F1 Score (Medium)	0.696	Moderate performance in the mid-level range with Precision = 0.615 and Recall = 0.800.
F1 Score (High)	0.895	Excellent recognition of high fill levels with Precision = 0.944 and Recall = 0.850.

4. Conclusions

In this research, have developed an Autonomous Robotic Inspection System which will be used for oil tank level detection based on a combination of deep learning and smart vision sensors, indicating a high level of accuracy, reliability and stability in oil level predictions. Results show a Mean Absolute Error (MAE) of 8.46%; Root Mean Square Error (RMSE) of 10.53% and a Coefficient of Determination (R^2) of 0.8347 in the deep learning vision model which confirms good correlation between predicted oil level and actual oil levels. Accuracy measured for 3-class classification was 81.1%, meaning that the majority of samples were correctly classified into Low, Medium, and High. Among them,

the High-level class had the highest classification accuracy (F1 = 0.895), followed by Low (F1 = 0.769) and Medium (F1 = 0.696), indicating outstanding recognition at larger levels of fill and good generalization between all categories. Graphical and statistical analysis showed uniform error reduction during the training, with the best final RMSE converging near 0.10 showing the stability and absence of overfitting of the model. Mark-up of the error distribution and residual plots indicated that the majority of the predictions were within ±3% tolerance and the visual comparisons also suggested that the detection of boundaries was accurate even in reflective or poor light conditions. Physically, the device accurately measured the oil level changes over the whole sight-glass range (20%–100%), with an accuracy rate around 92%. Finally, the combination of deep learning algorithms with smart camera sensors offers a solid non-invasive automated approach to oil-level monitoring of industrial storage tanks. Due to the high accuracy and generalization of the achieved system, it seems promising to be employed in real time for refinery processing, predictive maintenance or industrial process monitoring in an effort to mitigate the potential of manual inspections and enhance the operational capability.

References

[1] A., Thakur, R. N., & Basheer, S. (2022). Autonomous vehicles and intelligent automation: Applications, challenges, and opportunities. Mobile Information Systems, 2022, 1–36. https://doi.org/10.1155/2022/7632892

[2] Ha, N., Xu, K., Ren, G., Mitchell, A., & Ou, J. Z. (2020). Machine learning-enabled smart sensor systems. Advanced Intelligent Systems, 2(7), 2000063. https://doi.org/10.1002/aisy.202000063

[3] Dhiman, A., Shah, N., Adhikari, P., Kumbhar, S., Dhanjal, I. S., & Mehendale, N. (2021). Firefighting robot with deep learning and machine vision. Neural Computing and Applications, 34, 2831–2839. https://doi.org/10.1007/s00521-021-06537-y

[4] Ji, H., Cui, X., Ren, W., Liu, L., & Wang, W. (2021). Visual inspection for transformer insulation defects by a patrol robot fish based on deep learning. IET Science, Measurement & Technology, 15(6), 606–618. https://doi.org/10.1049/smt2.12062

[5] Li, J., Li, B., Dong, L., Wang, X., & Tian, M. (2022). Weld seam identification and tracking of inspection robot based on deep learning network. Drones, 6(8), 216. https://doi.org/10.3390/drones6080216

[6] Haldorai, A., Lincy, B. R., Suriya, M., & Balakrishnan, M. (2024). An improved single shot detection method for smart vision-based water garbage cleaning robot. Cognitive Robotics, 4, 19–29. https://doi.org/10.1016/j.cogr.2023.11.002

[7] Nauert, F., & Kampmann, P. (2023). Inspection and maintenance of industrial infrastructure with autonomous underwater robots. Frontiers in Robotics and AI, 10, 1240276. https://doi.org/10.3389/frobt.2023.1240276

[8] Filho, R. S., Yu, B., Huang, C.-L., Venkataramana, R., El-Messidi, A., Sharber, D., Westerheide, J., & Alkadi, N. (2020). The edge architecture for semi-autonomous industrial robotic inspection systems. International Journal of Cloud Computing, 9(1), 95–128.

[9] Dai, R., Wang, R., Shu, C., Li, J., & Wei, Z. (2025). Crack detection in civil infrastructure using autonomous robotic systems: A synergistic review of platforms, cognition, and autonomous action. Sensors, 25(14), 4631. https://doi.org/10.3390/s25154631

[10] Abagiu, M. M., Cojocaru, D., Manta, F., & Mariniuc, A. (2023). Detecting machining defects inside engine piston chamber with computer vision and machine learning. Sensors, 23(2), 785. https://doi.org/10.3390/s23020785

[11] Bathla, G., Sharma, S. C., & Bathla, A. (2022). Autonomous Vehicles and Intelligent Automation: Applications, Challenges, and Future Trends. Mobile Information Systems, 2022, Article ID 8112541. https://doi.org/10.1155/2022/8112541

[12] Ha, T. H., Han, J., & Oh, S. (2020). Machine Learning-Enabled Smart Sensor Systems. Advanced Intelligent Systems, 2(6), 1900157. https://doi.org/10.1002/aisy.201900157

[13] Dhiman, C., & Jain, R. (2021). Intelligent Inspection Robots for Hazardous Environment: A Review. Journal of Intelligent & Robotic Systems, 101(1), 1–16. https://doi.org/10.1007/s10846-020-01273-y

[14] Ji, Y., Liu, M., Hou, Y., Zhang, B., Li, Y., & Wang, S. (2021). Visual inspection for transformer insulation defects by a patrol robot fish based on binocular vision. IET Science, Measurement & Technology, 15(3), 296–304. https://doi.org/10.1049/smt2.12027

[15] Obaid, M. H., & Hamad, A. H. (2023). Deep Learning Approach for Oil Pipeline Leakage Detection Using Image-Based Edge Detection. Sensors, 23(3), 785. https://doi.org/10.3390/s23030785

[16] Li, D., Yuan, J., Lin, J., & Jiang, Y. (2025). A Review of Technical Advances and Applications of Intelligent Inspection Robots in Structural Infrastructure. SmartBot, 10, 1–20. https://doi.org/10.3389/frobt.2023.1240276

[17] Zhang, Y., Li, H., Zhao, Y., & Gao, Q. (2022). UAV-Based Inspection System for Gas Pipeline Leakage Detection. Journal of Physics: Conference Series, 2378, 012085. https://doi.org/10.1088/1742-6596/2378/1/012085

[18] Soldan, D. L., De Barros, R. C., Santos, A. C. F., & Santos, C. P. (2013). Flexible Mobile Robot for Pipe Inspection in Oil Refineries. Proceedings of the 2013 ACM/IEEE International Conference on Cyber-Physical Systems, 61–70. https://doi.org/10.1145/2485291.2485298

[19] Teixeira, M. A., de Almeida, F. G., & Nogueira, M. F. (2018). Unmanned Aerial Vehicle for Methane Gas Detection: Sensor Integration and Field Tests. Sensors, 18(10), 3312. https://doi.org/10.3390/s18103312

[20] Shukla, S., & Behera, R. K. (2013). Reconfigurable Embedded System for Robotic Visual Inspection. International Journal of Advanced Robotic Systems, 10, 170. https://doi.org/10.5772/56761

[21] Ali, M. S., Alsamhi, S. H., Alhumaid, R. M., Alotaibi, M. A., & Hassan, M. S. (2023). Autonomous UAV-Based Visual Inspection System for Oil Tank Corrosion Detection Using Cascaded Fuzzy Logic. Drones, 7(4), 133. https://doi.org/10.3390/drones7040133

[22] Rayhana, R., Sadik, A. Z., Hossain, M. M., & Rahman, M. M. (2021). An IoT-Based Smart Pipeline Monitoring System Using Distributed Sensor Networks. IoT, 5(1), 43–61. https://doi.org/10.3390/iot5010003

[23] Devold, H. (2019). Oil and Gas Production Handbook: An Introduction to Oil and Gas Production. ABB. Retrieved from https://new.abb.com/

[24] Obeidat, R., & El-Sawalhi, A. (2022). Predictive Maintenance Frameworks in Digital Oilfields: A Machine Learning Perspective. Open Research Europe, 2, 25. https://doi.org/10.12688/openreseurope.14480.1

[25] Ahmed, M., Alsamhi, S. H., & Ma, O. (2024). UAV-Based Intelligent Monitoring System for Pipeline Leak Detection Using AI and Thermal Imaging. Sensors, 24(5), 3721. https://doi.org/10.3390/s24053721

[26] Li, H., & Zhang, Y. (2023). Deep Learning Approach for Oil Pipeline Leakage Detection Using Image-Based Edge Processing. Unpublished manuscript.