



# SPEEDUP OF DETECTION AND ESTIMATION OF E. COLI CONTAMINATION USING AI-BASED OBJECT DETECTION

Niven Ahmad Alibraheem<sup>1</sup>, Mohamad Shady Alrahal<sup>2</sup>

<sup>1</sup>Faculty of Technical Engineering, Department of Food Engineering Technologies, Aleppo University, Syrian Arab Republic

<sup>2</sup>AIMD Lab, College of Computing, Department of Computer Science, Fahad Bin Sultan University, Saudi Arabia

Corresponding Author: Niven Ahmad Alebrahim

Article DOI: <https://doi.org/10.36713/epra24415>

DOI No: 10.36713/epra24415

## ABSTRACT

**Problem:** Detection of E.coli contamination and speedup the detection process, while providing quantifiable measure to estimate the growth level of contamination.

**Used Techniques:** Convolutional Neural Network (CNN) supported by object detection modeling based on directions of the shape of the E.coli bacteria.

**Used Data:** E.coli contamination images taken from Universe online repository and the milk is used as a food to be infected.

**Results:** E. coli contamination poses a significant health risk, making its detection crucial. Leveraging advanced technologies like artificial intelligence (AI) and deep learning (DL), this study introduces an intelligent system designed to identify E. coli contamination in milk. The system employs object detection modeling to achieve higher accuracy in recognizing contaminants, while k-medoids clustering groups the detected objects. The cluster size serves as an indicator to estimate the contamination's growth level. Compared to existing methods, the proposed system demonstrates superior performance, achieving 96% detection accuracy and a 7-second reduction in detection time.

**KEYWORDS:** E. coli, Object Detection, Clustering, Accuracy, Artificial Intelligence

## 1. INTRODUCTION

There is a wide spectrum of bacteria that leads to contamination of food. E.coli is considered one of the most important reasons of food contamination. Escherichia coli (E. coli) is a type of bacteria commonly found in the intestines of humans and animals. While most strains are harmless, some can cause severe foodborne illnesses [1]. There are several harmful strains of E. coli as summarized in the following points:

- Shiga toxin-producing E. coli (STEC): The most dangerous type, including E. coli O157:H7, which can cause severe illness.
- Enterotoxigenic E. coli (ETEC): Causes traveler's diarrhea.
- Enteropathogenic E. coli (EPEC): Affects infants and children.

The traditional ways to prevent E. coli contamination includes cook meat thoroughly (ground beef to 160°F/71°C), avoid raw milk/unpasteurized juices, wash hands, utensils, and surfaces after handling raw meat, and store food properly (keep raw meat separate from ready-to-eat foods) [2].

### 1.1. Statement of Problem

Recently, Artificial Intelligence (AI) techniques has been employed to detect E. coli contamination, such as image processing based on color detection or based on the signature of gases issued by the E.coli bacteria, where in both solutions Deep Learning (DL) is used. However, depending on such solutions suffer from some issues, such as:

- 1- They are computationally cost, as the process requires applying complex filter to extract colors as well as applying color indicators to detect the contamination.
- 2- They are expensive, as the process requires conducting intensive experiments to gather gases and monitoring them before applying signature based methods.
- 3- They are lack of a mechanism of estimating the growth of the E.coli contamination.

The issues mentioned above leads to low accuracy as well as wasting time, where speedup the detection process maybe critical to avoid spread of contamination. The best way to respond to the after mentioned issues is to use object detection to quickly identify and detect E.coli contamination [3].

### 1.2. Research Questions

The research questions that are linked to the problem stated above are as follows:

- 1- How to speedup the detection of E.coli contamination through object detection?
- 2- How to estimate the level of contamination accurately?

## 2. LITERATURE REVIEW

This section provides a review of related works that aim at detection of E. coli contamination using artificial intelligence techniques, where one of them will be used to compare the results of this work with it.

**First Study:** Authors of work [4] proposed a system to detect E. coli contamination using color image processing with K-means AI-based method. To extract good features, they used augmentation phase to show the images from different views. The achieved accuracy is 93%. It is worth mentioning that this study is selected to compare its results with the proposed system in this work.

**Second Study:** Authors of work [5] showed that the real-time object detection and classification algorithm You Only Look Once version 4 (YOLOv4) can accurately identify the presence of E. coli contamination. They used object detection based on classification and localization. The accuracy they achieved is 94%.

**Third Study:** Authors of work [6] proposed an environmentally friendly method for detecting Escherichia coli (E. coli) in fresh pork using hyperspectral imaging (HSI) and deep learning. Convolutional Neural Network (CNN) with Support Vector Machine (SVM) classification method is used, where the accuracy that was achieved is 87.5%.

**Forth Study:** Authors of work [7] stated that In order to detect E. coli, the most used and the cheapest technique takes approximately  $22 \pm 2$  hours and the results are obtained with naked eye. However, this is too long and during this test, pathogenic bacterial contaminated water can be consumed. In this study, analysis tools, which can automatically perform the bacteria detection process and can give an immediate alarm when the detection provided were proposed. They used R-CNN framework to build the system, where the accuracy they achieved is 95%.

## 3. METHODOLOGY

The methodology used in this work is illustrated in Figure 1, where each stage has its own goal and the output of each stage is considered an input to the next one.

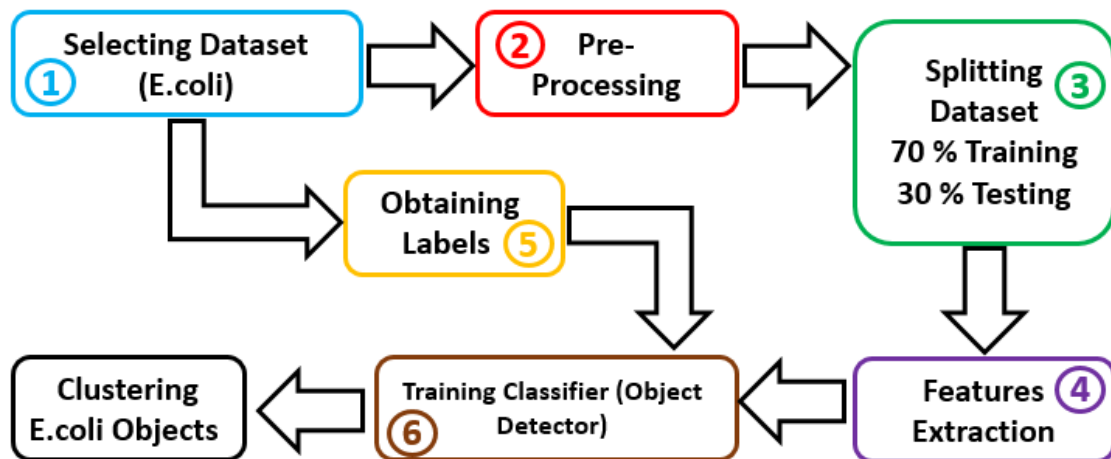


Figure 1. Flowchart of used methodology.

### 3.1. Selection of Dataset

The dataset is taken from the universe website, which is a common online repository that gather many databases from different domains to be used by researchers. The dataset is called EcoliDetection [8]. The size of the dataset is 49 images of E.coli contamination in milk. The images are jpg type with different sizes. Figure 2 provides a sample of E.coli contamination in milk dishes.

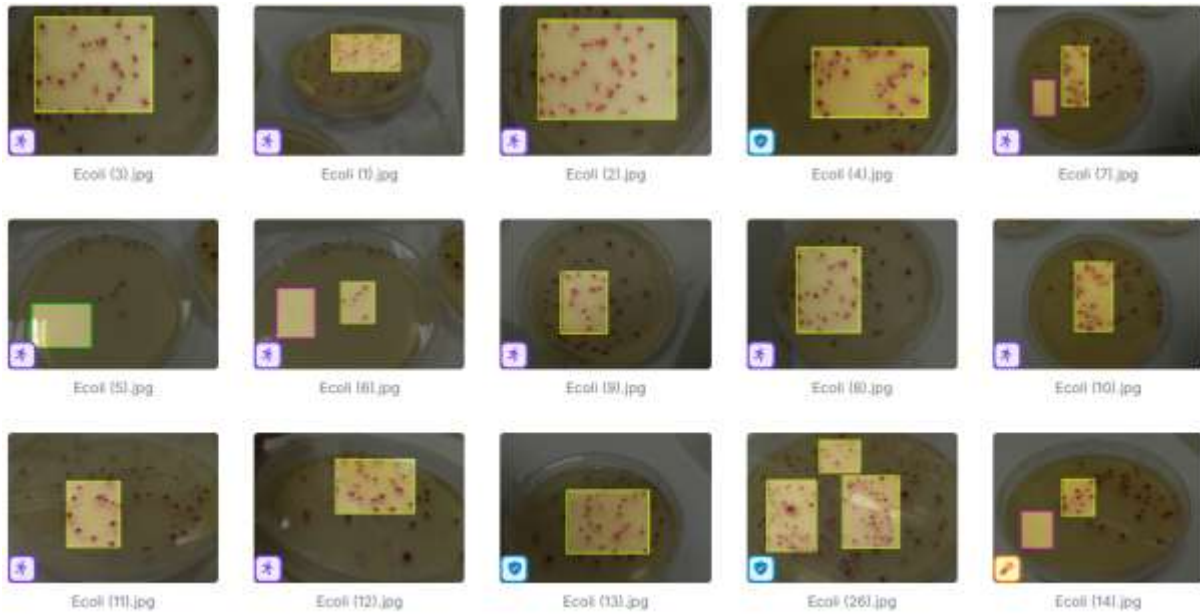


Figure 2. Samples of E.coli contamination taken form the EcoliDetection dataset.

### 3.2. Pre-Processing

The input of this stage is the images contained in the selected dataset. The goal of this stage is to prepare the images to facilitate further processing that leads to achieve higher accuracy detection. In this stage, the input images are rescaled to uniform the size (the new size is 416x416 resolution). In addition, the region of interest is identified by surrounding the infected area by a colored rectangle, as shown in Figure 3.

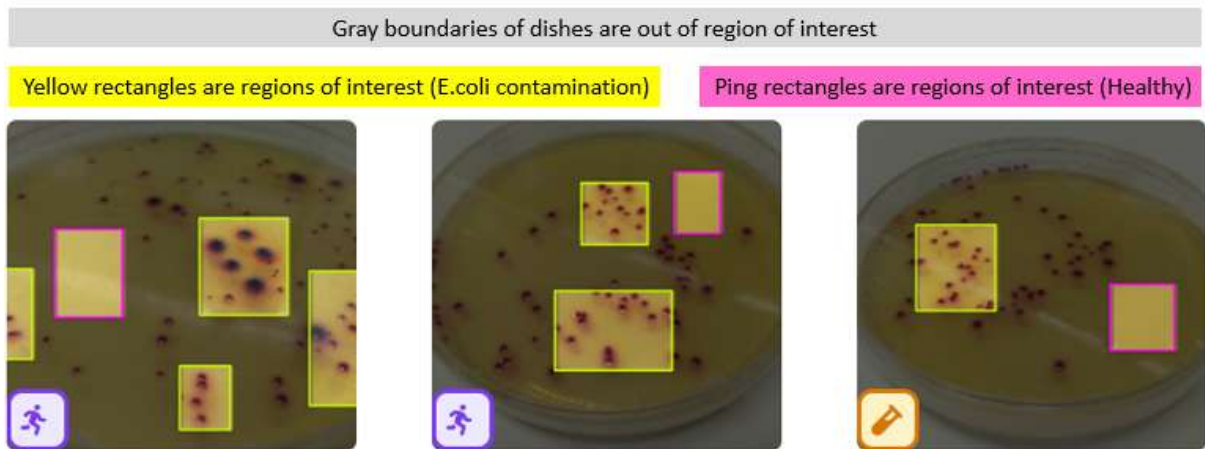


Figure 3. Regions of interest.

### 3.3. Dataset Division

The pre-processed images are the input of this stage. The goal of this stage is to obtain training part and testing part to enable training and testing process. In this context, the dataset is divided into training part (70% of original dataset, which means about 34 images) and testing part (30% of original dataset, which means 15 images).

### 3.4. Features Extraction

The input of this stage is the regions of interest obtained from the previous stage. The goal of this stage is to extract features used to train the intelligent model. To extract good features, the following steps are performed, as illustrated in Figure 4.

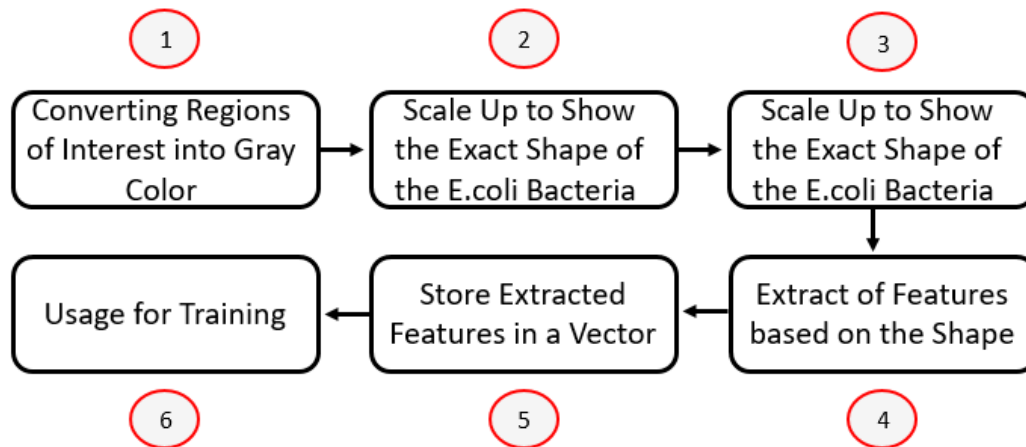


Figure 4. Process of extraction of features.

As shown in Figure 4, the first step is to convert the regions into gray color for the purpose of simplified process. The next step is to scale up the gray images to enable wide space of work. The third step is to shape the exact shape of the E. coli to be ready for extracting of features step. Figure 5 combines the results of the first three steps.

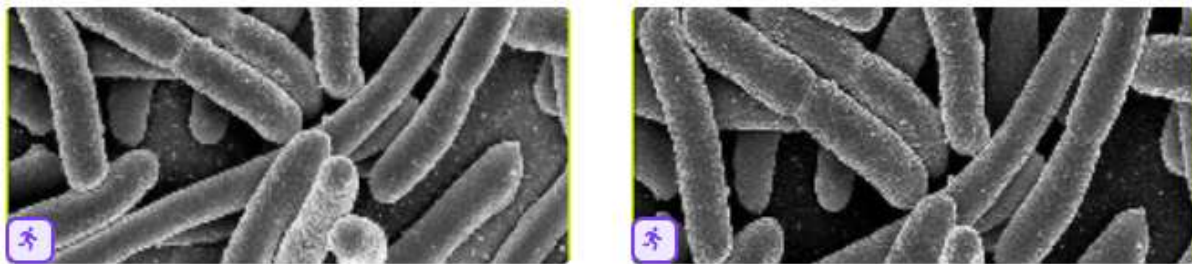


Figure 5. Resultant output of the first three steps.

To enable good features extractions (step 4), modeling is required. In this context, Histogram of gradient orientations is used [9], where the directions of the edges in shapes are modeled based on the direction cycle. Figure 6 illustrates the process.

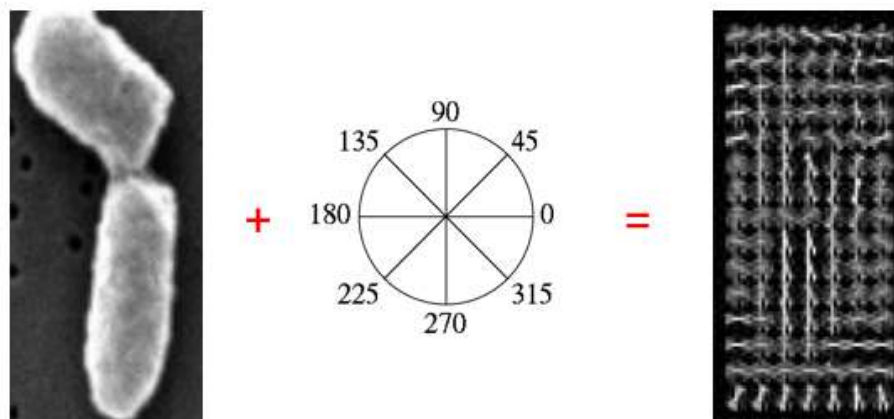


Figure 6. Modeling of features.

Step 5 includes arranging the extracted features in a vector to form the repository of training. However, the extracted features are not labeled. The labels are processed in the next stage.

### 3.5. Obtaining Labels

The input of this stage is the modeled features that are stored in the vector. The labels are obtained from the original dataset as it is used for classification (i.e., supervised learning). Therefore, the output of this stage is a pairs of features and class (E.coli contamination region or healthy region). At this point, the training stage can

start to train the detector as described in the next stage.

### 3.6. Training of E.coli Contamination Detector

The input of this stage is the labeled features. The goal of this stage is to train the detector based on the extracted features. In this context, Convolutional Neural Network (CNN) is used. Figure 7 illustrates the architecture of the CNN in terms of convolutional layers, pooling layers, fully connected layer, and used activation function (Softmax).

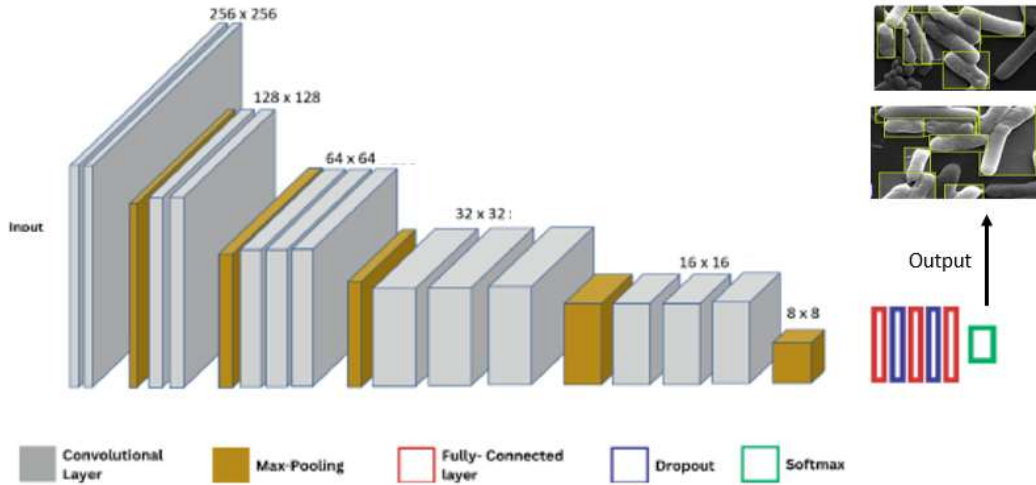


Figure 7. Architecture of used CNN.

### 3.7. Clustering E.coli Objects

The output of the CNN (detected objects) will be the input of the clustering process. The goal of this stage is to cluster similar objects in one group and then calculate the size of the cluster to be used as indicator to the growth of the E.coli contamination. In this context, the K-Medoids algorithm is used [10], which is illustrated in Figure 8.

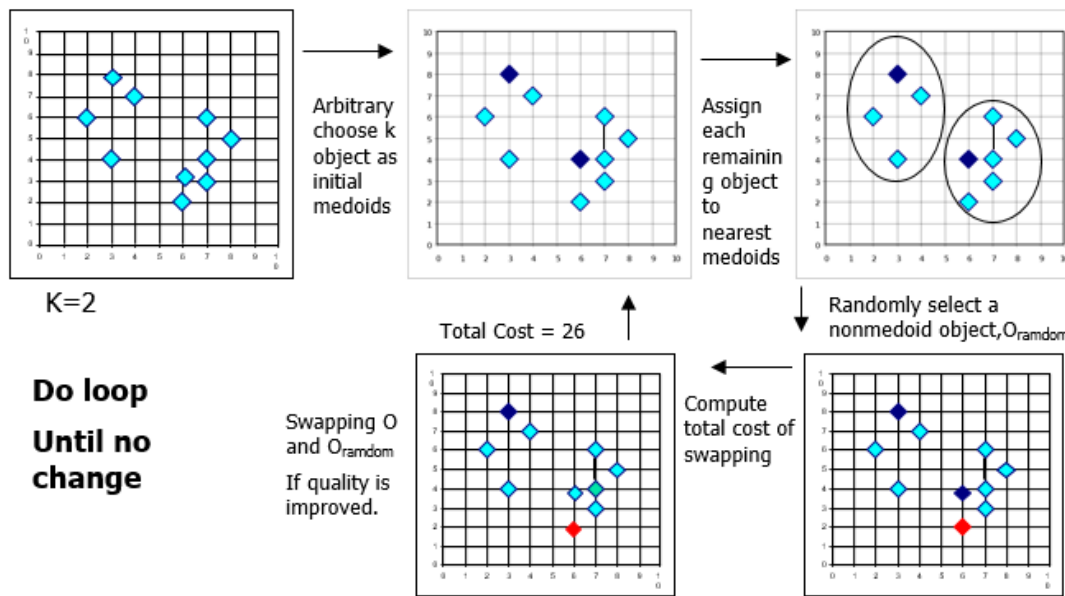


Figure 8. K-Medoids clustering method.

## 4. RESULTS AND DISCUSSION

The implementation of the intelligent system is developed using Matlab programming language. AIMD lab [11] within the college of computing at Fahad Bin Sultan University (FBSU) is used to conduct experiments. Accuracy of object detection as well as accuracy of clustering are used to evaluate the performance of the intelligent system. They are inspired from the confusion matrix [12], which is a common benchmark used in artificial intelligence domain. Table 1 shows the confusion matrix used to infer accuracy measure.



**Table 2: Confusion matrix.**

Predicted class \ Actual class	Confusion matrix		
	<i>E.coli Object</i>	$\neg$ <i>E.coli Object</i>	Total
<i>E.coli Object</i>	True Positives (TP)	False Negatives (FN)	TP + FN = P
$\neg$ <i>E.coli Object</i>	False Positives (FP)	True Negatives (TN)	FP + TN = N

Accuracy =  $\frac{(TP+TN)}{\text{number of all images in } E.coli \text{ dataset}}$

Table 2 summarizes the value of accuracy over time, which refers to the days of keeping the milk in an open space.

**Table 2: Results of Accuracy.**

Day Number	Accuracy of Object Detection	Accuracy of Clustering
1	93%	97%
2	94%	98%
3	95%	98%
4	95.5%	98%
5	96%	98%

**Discussion of Results.** The accuracy of *E.coli* detection in the first day reported the lowest value as it is in the emerging stage, where the intelligent system suffers from recognizing them from the healthy patterns. Compared to the other days, the accuracy increases to reach 96%. This can be justified by the growing of the *E.coli* bacteria that cause contamination. In regards to the accuracy of clustering, it reports steady value as the system is configured to generate tow clusters. This indicates to a good quality of clustering process. For the first day, the accuracy is 97% as the system is in the initial phase.

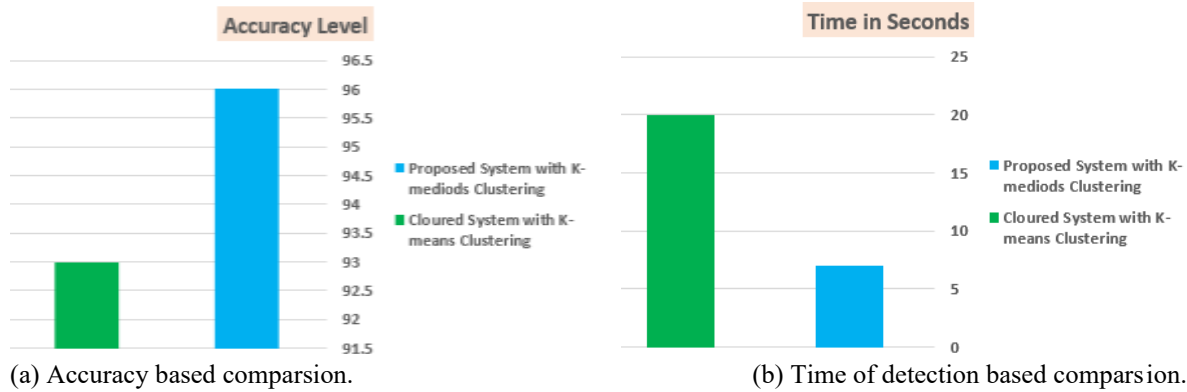
To measure the growth level of *E.coli* contamination, the size of clusters are considered. The size of a given cluster can be calculated based on the number of objects that are correctly detected and grouped in one cluster. Table 2 summarizes the results.

**Table 2: E.coli growth level based on size of clusters.**

Day Number	Configuration Number of Clusters	Number of Healthy Clusters	Size of Clusters	Number of Contamination Clusters	Size of Clusters
1	2	1	60	1	17
2	4	3	53	1	30
3	6	2	41	4	43
4	8	1	20	7	70
5	10	1	10	9	78

**Discussion of Results.** Under different configuration of the k-medoids algorithm (number of clusters), the size of healthy clusters (number of healthy detected objects) reported the highest value during the first day. That is because the *E.coli* is still in emerging phase. Moving to the further days, the size is decreased as the contamination dominated the situation. On opposite side, the size of the clusters that refers to the growth of the *E.coli* contamination reported the lowest value, and the value increased over time. This refers to the indicator that is used to measure the level of *E.coli* level of contamination.

**Comparison with similar work.** The proposed system is compared to the work [1] presented in the literature review section in terms of accuracy and time of detection. Figure 9 illustrates the results.



**Figure 9. Comparison to similar work.**

**Discussion of Results.** As shown in Figure 9 (a), the proposed system outperforms the other work by 3% accuracy. The reason behind this is related to using the histograms technique supported with a good modeling of the objects, which in turn leads to detect more objects with a higher level of accuracy. Figure 9 (b) showed that the time of detection of the proposed system is shorter than the other work by 13 seconds. That is because the used CNN depends mainly on the detected objects for clustering, while the other work mainly depends on color manipulation which requires more time processing.

## 5. CONCLUSION

E.coli contamination is a major cause of illness. In the light of existing advanced technologies, such as AI and DL, this work proposed an intelligent system to detect E.coli contamination in milk. The detection of objects used adopted modeling that leads to identify objects more accurately, while clustering based on k-medoids is employed to group the detected objects. The size of the clusters is used as a measure to estimate the growth level of the E.coli contamination. When compared to similar works, the proposed system showed better accuracy of detection with 96% and a speed up of detection by 7 seconds. The limitation of this work is related to the small size of training dataset. In future work, we intend to expand the experiments on larger datasets.

## REFERENCES

1. Smith, J. L., and P. M. Fratamico. "Escherichia coli as a Pathogen." *Foodborne diseases*. Academic Press, 2017. 189-208.
2. Allocati, Nerino, et al. "Escherichia coli in Europe: an overview." *International journal of environmental research and public health* 10.12 (2013): 6235-6254.
3. Deisingh, A. K., and M. Thompson. "Strategies for the detection of Escherichia coli O157: H7 in foods." *Journal of applied microbiology* 96.3 (2004): 419-429.
4. Naz, Bushra, et al. "A Machine Learning Framework for E. coli Bacteria Detection and Classification." *UMT Artificial Intelligence Review* 3.1 (2023): 17-30.
5. Ma, Luyao, et al. "Accelerating the detection of bacteria in food using artificial intelligence and optical imaging." *Applied and Environmental Microbiology* 89.1 (2023): e01828-22.
6. Liu, Ying, et al. "Detection and visualization of Escherichia coli in pork using hyperspectral imaging-based deep learning." *Journal of Agriculture and Food Research* (2025): 102107.
7. Yanik, Hüseyin, A. Hilmi Kaloğlu, and Evren Değirmenci. "Detection of escherichia coli bacteria in water using deep learning: A faster r-cnn approach." *Tehnički glasnik* 14.3 (2020): 273-280.
8. Universe Website. Online [Available]: <https://universe.roboflow.com/sabarishvt/ecolidetection/browse?queryText=&pageSize=50&startIndex=0&browseQuery=true> (Accessed 25-9-2025)
9. Tomasi, Carlo. "Histograms of oriented gradients." *Computer Vision Sampler* 1 (2012): 1-6.
10. Park, Hae-Sang, and Chi-Hyuck Jun. "A simple and fast algorithm for K-medoids clustering." *Expert systems with applications* 36.2 (2009): 3336-3341.
11. FBSU Website. Online [Available]: <https://fbsu.edu.sa/All/sections/Artificial-Intelligence,-Machine-Learning-and-Data-Security-Lab> (Accessed 27-9-2025)
12. Mona Alfi, Mohamad Shady Alrahal, Samir Bataineh and Mohammad Mezher. "Enhanced Artificial Intelligence System for Diagnosing and Predicting Breast Cancer using Deep Learning". *International Journal of Advanced Computer Science and Applications (IJACSA)* 11.7 (2020). <http://dx.doi.org/10.14569/IJACSA.2020.0110763>