Identifying Impurities in Liquids of Pharmaceutical Vials

Gabriele Rosati[1], Kevin Marchesini[1], Luca Lumetti[1], Federica Sartori[2], Beatrice Balboni[2], Filippo Begarani[2], Luca Vescovi[2], Federico Bolelli[1], Costantino Grana[1]

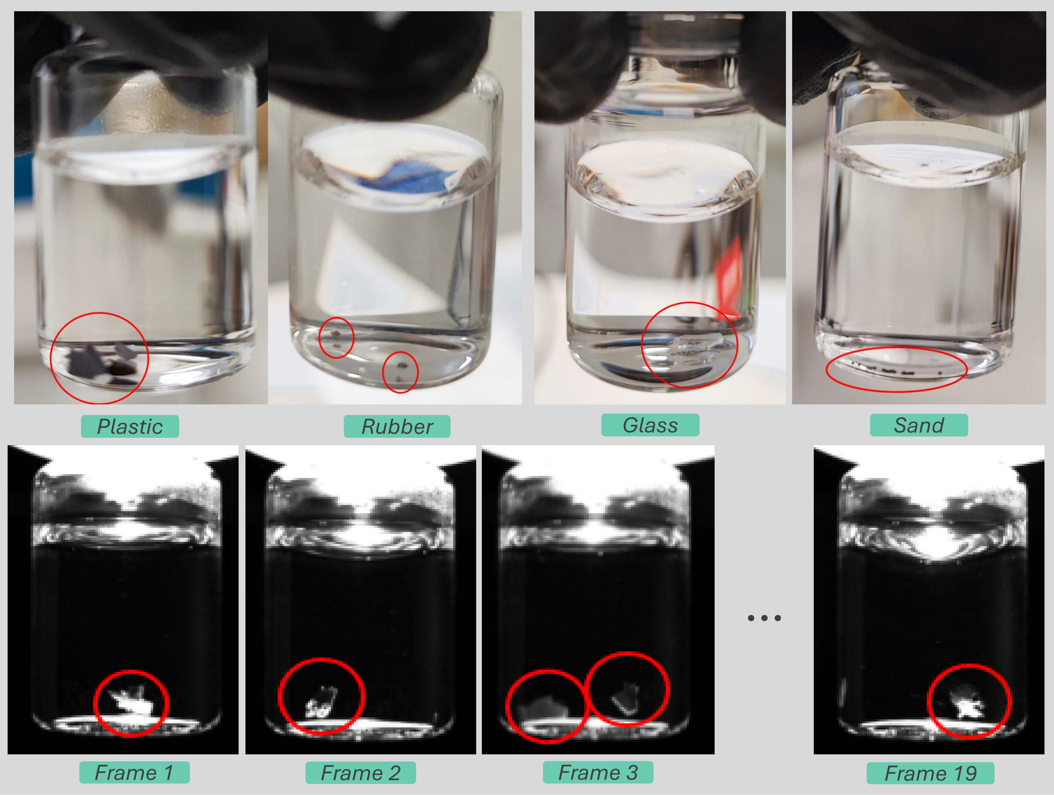
[1]University of Modena and Reggio Emilia, Italy
{name.surname}@unimore.it}

[2]Performing Beyond Limits S.r.l., Italy
{name.surname}@pblsrl.it}

Introduction

The **identification of visible particles in vials** is critical for pharmaceutical firms. Impurities such as plastic or glass can pose serious health risks and lead to regulatory non-compliance, and manual inspections are inefficient and error-prone. Modern systems, based on Deep Convolutional Neural Networks, provide a more reliable detection.

Dataset



The dataset was acquired internally by *Performing Beyond Limits (PBL)* and consists of 2,426 vial sequences filled with water, with 19 frames each.
It includes **five different classes**: one represents the absence of impurities, while the others correspond to four foreign particles: burnt plastic, rubber particulates, glass fragments, and sand debris.

Images were captured after rotating the vials at 200 rpm using a Matrix Vision camera to reveal impurities that would otherwise be invisible in static vials. Each frame was center-cropped to 325×268 pixels.

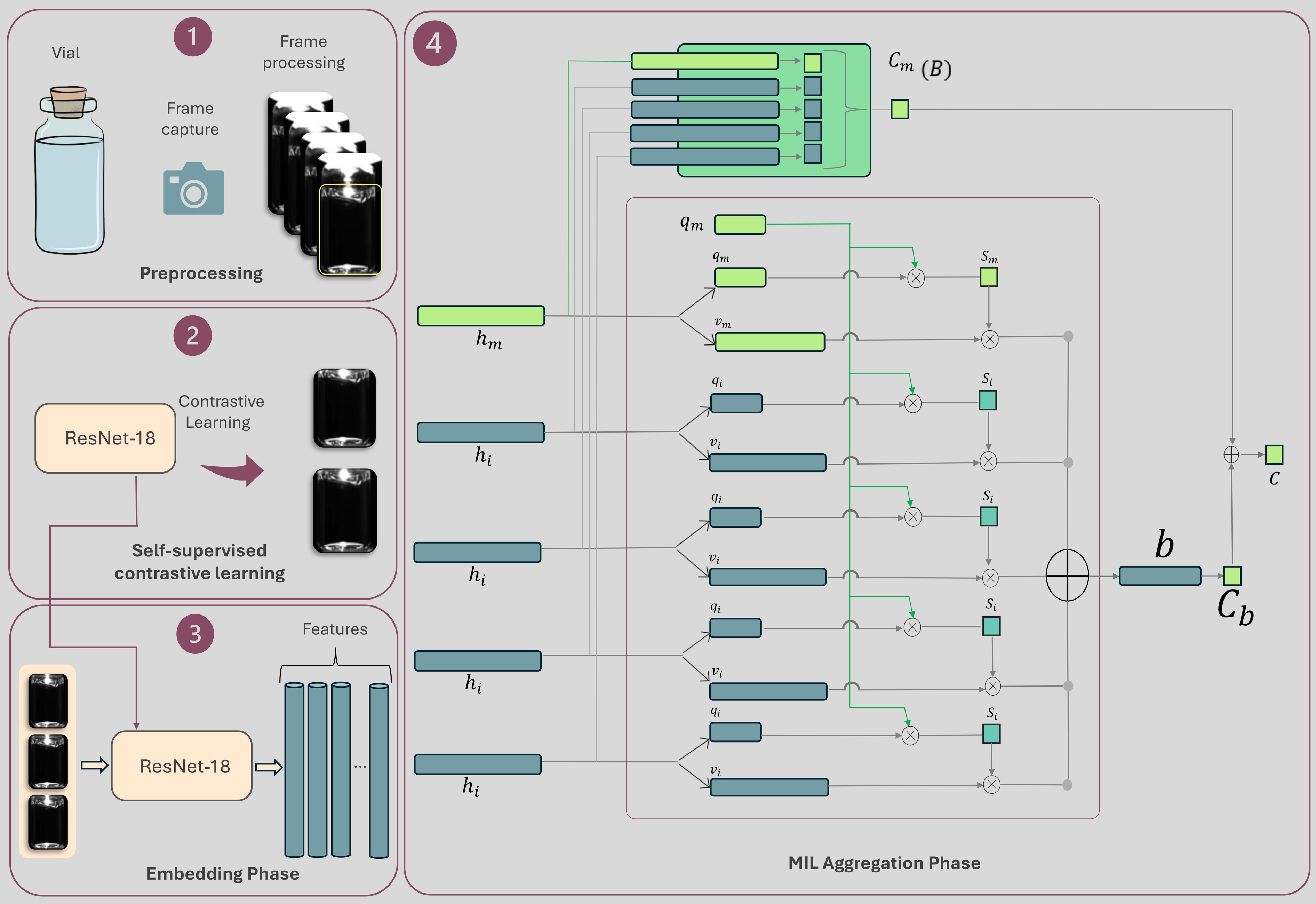
Method

Method 1. The multiclass detection challenge is tackled using a Multiple Instance Learning (MIL) architecture called **Dual-Stream Multiple Instance Learning** (DSMIL) [1,2], as MIL addresses the ambiguity of frame-by-frame labeling by enabling impurity detection at the sequence level.

First stream: extracts an embedding h_i from instances of bag B and classifies them. A max-pooling operation selects the instance with the highest score, named as the critical instance $c_m(B)$.
Second stream: transforms each instance embedding h_i into query and information vectors, which are given by: $q_i = W_q h_i,\ v_i = W_v h_i$, where W_q and W_v are learnable weight matrices.
A distance measure U is then computed between the critical instance and each of the other instances. The bag is represented as follows:

$$b = \sum_{i=1}^{n} U(h_i, h_m)\, v_i.$$

It is obtained by merging the vectors v_i of all instances using a weighted sum, where the weights are obtained by the distances to the critical instance embedding (h_m). Finally, a linear layer generates the bag score $C_b(B)$, which is averaged with $C_m(B)$, to produce the final class score.



Method 2. We used **ResNet-18** in two approaches:

a) Majority Voting: assigns a class to each frame and predicts the sequence's class based on majority voting, with the class receiving the highest votes assigned to the entire sequence.

b) Features Aggregation: extracts features from each frame using ResNet's layers, concatenates these features on a new dimension, and through fully connected layers predicts the overall sequence score.

Results

We evaluated performance using **4-fold cross-validation**, reporting the average metrics across all folds. The results suggest that the best-performing method is **DSMIL**, which reaches an accuracy of 99.53%. DSMIL misclassifies only a few sequences confusing brown particle samples as vials without impurities. Comparing the two aggregation methods used for ResNet experiments, we observed that concatenation is more effective than majority voting.

Model	Accuracy	Precision	Recall	F1-Score	Time [ms]
ResNet (voting)	0.9835 ±0.0071	0.9829 ±0.0062	0.9851 ±0.0069	0.9840 ±0.0064	**1257**
ResNet (concat)	0.9903 ±0.0046	0.9899 ±0.0042	0.9918 ±0.0048	0.9908 ±0.0046	1328
DSMIL	**0.9953 ±0.0023**	**0.9948 ±0.0020**	**0.9957 ±0.0024**	**0.9952 ±0.0022**	1639

References

[1] Li, B., Li, Y., Eliceiri, K.W.: Dual-Stream Multiple Instance Learning Network for Whole Slide Image Classification With Self-Supervised Contrastive Learning, CVPR 2021
[2] Bontempo, G., Porrello, A., Bolelli, F., Calderara, S., Ficarra, E.: DAS-MIL: Dis-tilling Across Scales for MIL Classification of Histological WSIs, MICCAI 2023.