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], and Costantino Grana[1]

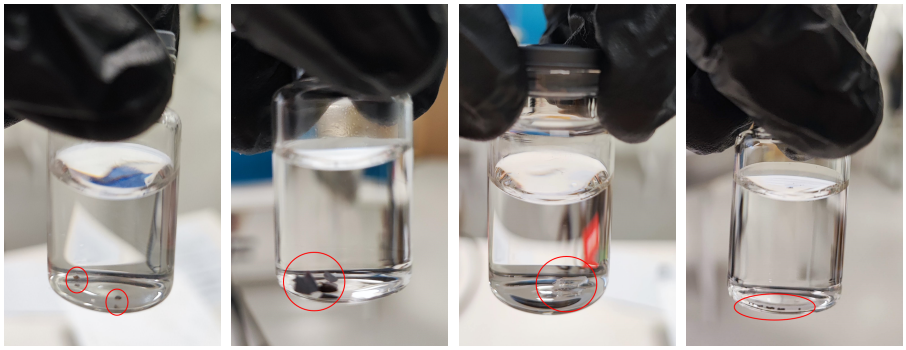
[1] Università degli Studi di Modena e Reggio Emilia, Modena, Italy
{name.surname}@unimore.it
[2] PBL S.r.l., Parma, Italy
{name.surname}@pblsrl.it

Abstract. The presence of visible particles in pharmaceutical products is a critical quality issue that demands strict monitoring. Recently, Convolutional Neural Networks (CNNs) have been widely used in industrial settings to detect defects, but there remains a gap in the literature concerning the detection of particles floating in liquid substances, mainly due to the lack of publicly available datasets. In this study, we focus on the detection of foreign particles in pharmaceutical liquid vials, leveraging two state-of-the-art deep-learning approaches adapted to our specific multiclass problem. The first methodology employs a standard ResNet-18 architecture, while the second exploits a Multi-Instance Learning (MIL) technique to efficiently deal with multiple images (sequences) of the same sample. To address the issue of no data availability, we devised and partially released an annotated dataset consisting of sequences containing 19 images for each sample, captured from rotating vials, both with and without impurities. The dataset comprises 2,426 sequences for a total of 46,094 images labeled at the sequence level and including five distinct classes. The proposed methodologies, trained on this new extensive dataset, represent advancements in the field, offering promising strategies to improve the safety and quality control of pharmaceutical products and setting a benchmark for future comparisons.

Keywords: Vial Liquid inspection · Multi-Instance Learning · Convolutional Neural Network · Classification · Prediction

1 Introduction

Control over visible particles represents an important aspect in various fields, such as pharmaceuticals, food and beverages, and manufacturing, because they have a significant effect on the quality of the products. Impurities found in food can have different forms: physical, chemical, and biological contaminants, like small metal fragments or pesticides [30]. These impurities pose significant health risks to humans, potentially leading to severe illnesses and affecting the



(a) Plastic (b) Rubber (c) Glass (d) Sand

Fig. 1. An example of some impurities (circled in red) that can occur in a liquid vial. (a) shows brown plastic particles, in (b) black rubber particles are present, (c) illustrates the presence of a piece of glass, and (d) shows residual sand at the bottom of the vial.

quality and taste of food. In manufacturing processes, impurities can alter desired product characteristics and performance. For example, metallic contaminants in semiconductor manufacturing can alter product mechanical properties, leading to the development of weak points, thus impairing the functionality of final products [9]. Our work is focused on foreign particles in pharmaceutical products. These impurities can lead to various consequences, including reduced effectiveness due to interference with active ingredients, safety risks from toxic substances or allergens, and regulatory issues resulting in product recalls or legal penalties [3]. These particles can arise from injection of the bottles, packaging, collisions, or filtration, and they can pose serious health risks when injected into the bloodstream, potentially resulting in thrombosis, phlebitis, tumors, and anaphylactic reactions [15]. The detection of these particles is particularly challenging because they can occur in various forms, such as dust, plastic, rubber and silicone particles, glass fragments, and sand residues as illustrated in Fig. 1.

Traditionally, identifying particles and impurities relied on manual inspection, which has been proven to be inefficient due to its time-consuming nature, subjectivity, low repeatability, and susceptibility to errors. Several factors influence the likelihood of visually detecting particles, such as the particle's size, composition, and shape, as well as the product formulation, the vials, the filled volume, and inspection conditions [29]. In manual detection, typically, inspectors position the injection bottle under a high brightness and planar light source, then rotate and tilt the container manually (or with the assistance of machinery) to observe any visible foreign substances inside. Based on their inspection experience, they decide whether these substances are acceptable or not. Such an approach often exhibits poor efficiency since it strongly depends on light conditions and other external factors, and it is not exactly repeatable [10].

Advancements in imaging technology and computer vision have led to the establishment of automated particle detection systems, which are increasingly reliable, removing human error. These systems typically leverage image processing [21], machine learning techniques [42], and well-known vision algorithms [18]. The main challenge to overcome is to find a method capable of efficiently extracting fine-grained features from images that may be captured under suboptimal lighting conditions and contain various sources of interference or noise. In summary, the issues to be tackled when designing particle detection methods in liquid vials are as follows:

1. The **appearance of particles** can be influenced by different lighting conditions, leading to variations in color, especially if some particles are transparent, as is often the case with glass fragments. Images of pharmaceutical containers are taken using a camera positioned beneath a mobile tracking device, operating synchronously, and changes in illumination within the image may occur due to ambient light conditions and vibrations from the machine;
2. Particles come in a **variety of forms**, ranging from small spots to bigger shapes. They can exhibit different textures and surface characteristics, from smooth to rough or irregular. The diversity in particle properties poses a significant challenge for detection and classification systems, requiring robust algorithms capable of effectively distinguishing between different particle types under various circumstances;
3. The presence of **noisy elements** on the bottle wall and bubbles [40] within the liquid can pose challenges in classifying the foreign particles, as they share similar visual characteristics.

In recent years, Convolutional Neural Networks (CNNs) have been extensively used for multiple applications [4, 7, 33, 34, 38, 39], including industrial defect detection [8, 11, 20, 41]. They have shown promising results in overcoming the aforementioned issues. CNN models can perform various tasks thanks to their strong capability to represent robust features. While recent literature focused on developing tools for detecting particles in liquids using deep learning methods, there is an absence of publicly available datasets for this task, mainly due to the preservation of industrial secrets. For this reason, previous research in this area relied only on private datasets, making the comparison with existing approaches impractical.

Paper Contributions. To partially cope with this literature gap, this paper releases a small set of images that can be employed for future comparison.[3] Unfortunately, for the same aforementioned reasons, the entire training set cannot be released.

More specifically, this paper tackles the problem of identifying different kinds of impurities in pharmaceutical vial liquid by smartly leveraging two existing state-of-the-art deep learning approaches, namely ResNet-18 [16] and DSMIL [22]. To cope with the previously identified issues 1 and 2, instead of dealing with a single image per sample, we opted for acquiring sequences of 20 images for each

[3] Test data are available at https://ditto.ing.unimore.it/residual.

vial, suitably subjected to machinery-supported rotation. Such an approach, which is also feasible in modern inline injection machines, allows for mitigating particle appearance issues and the presence of noisy elements. However, it introduces additional challenges in the automatic detection algorithms. In order to achieve satisfactory performance without sacrificing computation time, our approach advocates for ResNet-18 by directly feeding it with multiple channels, each corresponding to a sequence frame.

Additionally, to achieve similar results, although tackling it in a different way, a Multi-Instance Learning Approach (MIL) is employed by treating each sequence as a *bag* composed of multiple images *instances*. This way, the model can deal with moving objects in the sequence without requiring expensive tracking strategies as previously proposed in literature [48].

In both cases, our proposed pipeline achieves outstanding results without requiring pixel-level annotations.

2 Related works

Product quality is crucial for pharmaceutical products, given their impact on people's health. To ensure this quality, various works have been made on vial inspection, with the goal of detecting defects such as tilting and sinking of the cap or cracks in the glass, which may negatively affect the product quality [44]. Although this is a slightly different task with respect to the detection of liquid defects, our approach follows similar steps and employs comparable techniques to those used in these studies.

The first works in this field employed traditional computer vision techniques. Liu *et al.* [25] have proposed an inspection method that used the watershed transform to find defective areas and a fuzzy SVM ensemble combined with an ensemble of genetic algorithms to classify the type of imperfection. Also, Liu *et al.* [27] used the SVM classifier to inspect vials for flaws, fed with local binary pattern (LBP) features extracted from the region of interest of the image, grouped using k-means clustering to have a compact representation of them. Several other studies have utilized SVM for classifying defects on the surface of the rolled steel [19], in the industrial pavements [28], and in textile materials [1]. The key difference in existing approaches lies in the method used for feature extraction. More recently, Zhou *et al.* [49] proposed two different techniques to find defects in glass bottles using traditional vision algorithms: a template-matching-based method with multiscale filtering, and a region-growing Euclidean saliency method, with the integration of superpixel segmentation and geodesic saliency detection algorithms.

Regarding the analysis of liquid solutions, Wang *et al.* [45] developed a method to find unwanted glass fragments in the liquid by shaking the container, exploiting the fact that the glass pieces are heavier, so they cannot move smoothly with liquid and other particles. Thus, they took several images in sequence and used the optical flow algorithm to perform the detection. In the same year, Ge *et al.* [12] presented an automated system for checking ampoule

injections for tiny foreign particles. They developed a custom hardware platform for transportation and agitation, capturing images for analysis. The computation of trajectories of moving objects within liquid allowed them to differentiate foreign particles in the images; then impurity types were classified through multiple features, including particle area, mean gray value, and geometric invariant moments.

The advent of deep learning has been a breakthrough in visual detection tasks, including defect detection [41]. Its ability to autonomously learn complex features from datasets enabled algorithms to accurately identify patterns and objects with more precision. One of the first approaches regarding foreign particle inspection is another work of Ge *et al.* [11]. They successfully explored the usage of a modified version of Pulse-Coupled Neural Networks (PCNN) [20] to identify undesired particles in glucose or sodium chloride injection liquids. PCNNs are non-trained neural networks where each neuron receives as input the corresponding pixel intensity and other inputs from its neighboring neurons. These stimuli are added together, accumulating them until they surpass a dynamic threshold, triggering a pulse output. This process, iteratively performed, generates a series of binary images as outputs. Neighboring neurons' connections lead to pixels of the image with similar intensity values pulsing together. Thus, it is possible to obtain image segmentation by identifying pixels corresponding to synchronously pulsing neurons. The main drawback of this technique lies in its dependence on the choice of thresholds. The author of the paper suggested an adaptive approach to find the best hyperparameters.

Since the middle of the 2010s, many neural network architectures have been developed for detection tasks, such as R-CNN [14], Faster R-CNN [36], YOLO [35], SSD [26], and ResNet [16] and have become widely popular. Examples of application of these networks can be found in defect detection addressing various domains, such as the inspection of flat surfaces [46] using a combination of Fast and Faster R-CNN, the particle detection in complex biomedical images [13] through a ResNet-based architecture and the detection of cracks in aircraft structures through the usage of YOLOv3-Lite [23]. In the work of Ding *et al.* [8], a defect-detecting Single-Shot Detector (SSD) is devised for wood inspection, using DenseNet [17] as the backbone to improve the extraction of deep features and mitigate gradient vanishing issues of the original SSD backbone. Furthermore, the integration of a feature fusion function to combine multi-layer feature maps from the backbone enhances the classification of wood defects. Ritter *et al.* [37] presented a new method to identify and track fluorescent particles in microscopy images. Their approach leveraged the Deconvolution Network [31], a CNN similar to an encoder-decoder architecture for particle detection, along with a bidirectional long short-term memory for tracking, which also aided in particle classification. A less conventional deep learning approach was used by Zhang *et al.* [48], who developed a particle inspection system for liquid vials. They captured eight sequential images and used fuzzy cellular neural networks for precise position and segmentation, introducing an adaptive tracking system based on a sparse model for determining the presence of foreign particles. In

one of the most recent works, Yi *et al.* [47] explored the usage of the attention mechanism on pharmaceutical foreign particle detection. They developed an end-to-end deep architecture with adaptive convolution and multiscale attention to identify and classify foreign particles.

Based on the results reported in the aforementioned papers, we can state that deep learning detection methods outperform traditional approaches in particle detection liquids. For these reasons, in this work, we choose to employ two state-of-the-art deep learning architectures: ResNet [16], which we employed in a new fashion to handle sequences rather than individual images, and DSMIL [22], a method not previously investigated for multiclass particle detection. The specific details of our architectures and the results on our dataset are outlined in the following sections.

3 Methods

As said, to face the task of recognizing defective vials, we decided to explore two different paths; the former is based on the use of ResNet [16] in a slightly different way than the standard one, in order to deal with the entire sequence of images, the latter is a Multi-Instance Learning (MIL) [5] based technique.

3.1 ResNet-18

Residual Neural Network (also known as ResNet [16]) is a family of deep learning models in which the weights layers learn residual functions based on the layer inputs. This is possible through the residual connections that execute identity mappings and are added to layer outputs. In our study we employed ResNet-18.

As ResNet operates on individual images, we had to adapt its architecture to our problem, where we deal with a sequence of images for each rotating vial. Our goal was to capture the collective information across the sequence of frames acquired during the vial's rotation. To perform classification at the sequence level, we explored two different aggregating methods. In the first approach, we learned to predict a class for each frame within the sequence and subsequently determined the class for the entire sequence through a majority voting approach. Secondly, we investigated an alternative approach wherein we independently extracted features from each frame using ResNet convolutional layers. Then, these features were concatenated along a new dimension before being fed to the fully connected layers, resulting in a single prediction for the entire sequence. As a loss function, we used cross-entropy.

3.2 MIL-based Approach

Multiple instance learning is an extensively used weakly supervised learning algorithm [2,24,32] where a subset of examples from the training set is arranged as

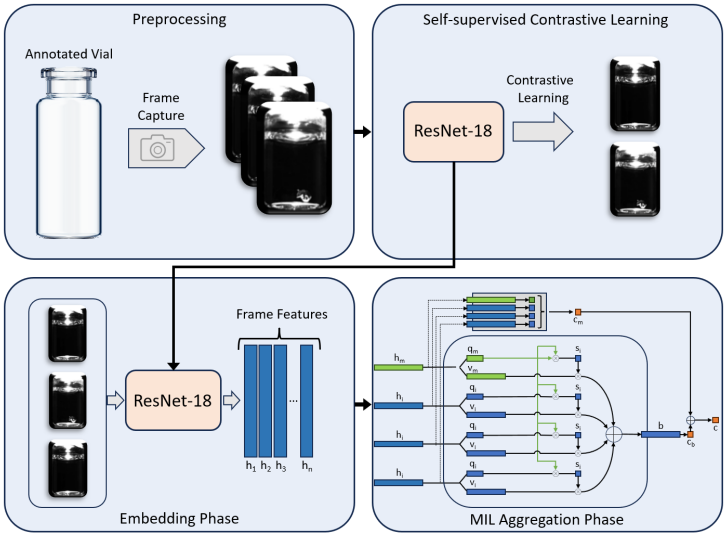


Fig. 2. Representation of the proposed MIL-based pipeline, divided into four main steps: preprocessing, self-supervised training of ResNet-18, Embedding phase, and the MIL phase.

a set (bag) composed of multiple instances. If we deepen the case of binary classification, let $B = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ be a bag where $x_i \in X$ are instances with labels $y_i \in \{0, 1\}$, the label of B is given by:

$$c(B) = \begin{cases} 1, & \text{if } \exists\ y_i \in B : y_i = 1 \\ 0, & \text{otherwise} \end{cases}$$

The challenge of detecting defects in the liquid inside vials can be seen as a multiple-instance learning problem if the detection method involves capturing a series of images of the rotating vials and the labeling is done based on the entire sequence. The sequence-level labeling method is usually the standard because while foreign particles may not always be visible in every frame if they appear at least in one frame, the vial should be classified as defective.

The problem of multi-instance learning for a bag-level classification can be approached by training a model that assigns a probability $c(X)$ of the bag being labeled as positive ($Y = 1$). The function $c(X)$, can be formulated as follows:

$$c(X) = g(\sigma(f(x_1), \ldots, f(x_n)))$$

where the function f is a feature extractor transforming single instances into a lower-dimensional embedding; σ is a permutation-invariant aggregation function (often referred to as MIL pooling), which derives the bag representation; and g apply a final transformation to obtain the bag probability. Both functions f and

g can be parameterized by neural networks, which can be trained end-to-end through backpropagation. The only other requirement is that the MIL pooling operation σ must be differentiable.

In our case, each image sequence is considered a bag, while each single frame composing the sequence is treated as an instance. We used the MIL architecture developed by Li *et al.* [22] called Dual-Stream Multiple Instance Learning (DSMIL). This network, depicted in Fig. 2, learns from both instances and bag embeddings at the same time. The first stream works at instance-level. It extracts an embedding from each instance and classifies each embedding, giving a single score in case of a binary classification problem. Then, the classification step is followed by a max-pooling operation to identify the instance with the highest score, referred to as the *critical instance*.

In a more exhaustive way, let $X = x_1, ..., x_n$ denote a sequence (bag) of frames of a rotating vial. Given f as feature extractor, each frame x_i can be projected into an embedding $h_i = f(x_i) \in \mathbb{R}^{L\times 1}$. The first stream uses a frame classifier on each frame embedding, followed by max-pooling on the scores:

$$c_m(X) = g_m(f(x_1), \ldots, f(x_n)) = \max\{W_0 h_1, \ldots, W_0 h_n\}$$

where W_0 is a weight vector. The max-pooling stream provides the frame with the highest score (the *critical instance*).

The second stream aggregates the above frame embeddings into a single sequence embedding, which is further scored by a bag classifier. It transforms each instance embedding h_i, obtained in the first stream (including the critical instance embedding h_m) into two vectors, query $q_i \in \mathbb{R}^{L\times 1}$ and information $v_i \in \mathbb{R}^{L\times 1}$, which are given respectively by:

$$q_i = W_q h_i, \quad v_i = W_v h_i, \quad i = 0, \ldots, N-1$$

where W_q and W_v are learnable weight matrices. Then, a distance measurement U, which has a similar structure and meaning of the attention operation used in Transformers architecture [43], is defined as follows:

$$U(h_i, h_m) = \frac{\exp(\langle q_i, q_m \rangle)}{\sum_{k=0}^{N-1} \exp(\langle q_k, q_m \rangle)}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product of two vectors. As we can see from the formulation, the distance is computed only between the critical instance and all the instances in the bag. This ensures a linear complexity of O(n) rather than quadratic like the attention mechanism.

Overall bag representation b is computed by combining the information vectors v_i of all instances using a weighted sum, where the weights are determined by the distances to the critical instance:

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

Fig. 3. Sample images from a sequence of a vial containing no impurities.

The bag score of the second stream is obtained through a final linear layer. This score, averaged with the one of the first stream $c_m(B)$, produced the final score.

Since our research tackles a multiclass problem, DSMIL has been adapted accordingly. We use max-pooling to determine the critical instance of each class, and then we compute attention weights for each class individually with respect to the corresponding critical instance. As a result, the bag embedding b becomes a matrix with dimensions $L \times C$, where C represents the number of classes. In this matrix, each entry is a weighted sum of the instance information vectors v_i. The final fully connected layer for the classification has C output channels.

DSMIL exploited SimCLR [6], which stands for Simple Contrastive Learning Representation, to produce a robust feature extractor in an unsupervised learning setting. In our case, SimCLR trains a ResNet-18 to drastically reduce the input size of each frame by embedding it into a vector. It randomly selects pairs of images from the sequences, applies random augmentations to improve the robustness, and trains the network to maximize similarity between images belonging to the same sequence while minimizing similarity between images from different sequences. After training, ResNet-18 is used to generate the embeddings for single frames within the first stream of DSMIL.

4 Experiments and Results

Dataset. The samples under examination are glass vials with silicone caps filled with distilled water. Image acquisition was performed on a rotating test bench using a Matrix Vision camera with a bottom white illuminator and a single LED. The vials were rotated at a speed of 200rpm with an acceleration and

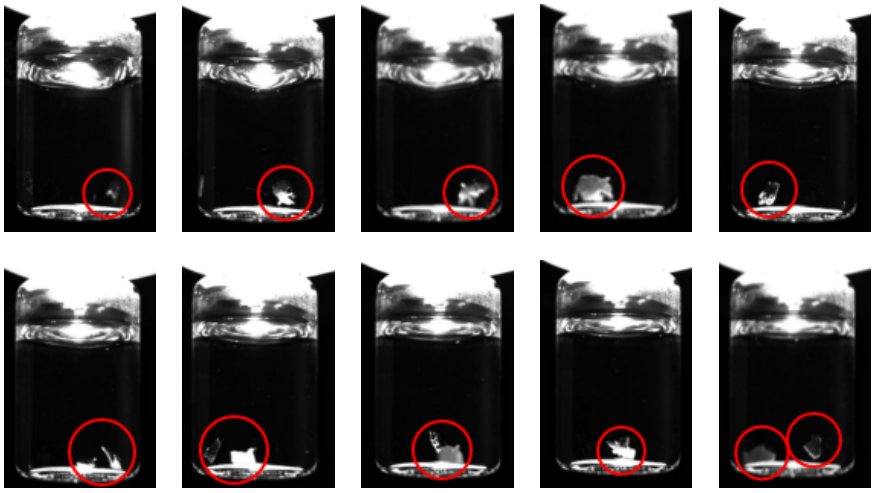


Fig. 4. Sample images from a sequence of a vial containing glass impurities.

deceleration time of 400ms. Image acquisitions of each sample occurred after a rotation of the vial, with a delay of approximately 100ms.

Before the acquisition procedure, each vial was cleaned on the outside with alcohol to remove marks and residual particles from the glass. The dataset used to train and evaluate our models is composed of 2,426 vial sequences, where each sequence consists of 19 frames, for a total of 46,094 images.

The dataset contains annotations for five different classes. One class represents *good* vials, indicating the absence of impurities. The other four classes refer to different types of foreign particles: *brown* impurities, corresponding to burnt plastic particles, *black* defects, corresponding to rubber or silicone particulates, a class is for *glass* pieces of various sizes, and the last class for *sand* residues. These are essentially the defects shown in Fig. 1. Samples from a clean sequence are reported in Fig. 3, while images extracted from sequences containing glass and sand impurities are depicted in Fig. 4 and Fig. 5.

Pre-processing. Each frame in the dataset encompassed a pre-processing phase consisting of a center crop to a fixed dimension of 325×268 pixels to isolate the vial, followed by a rotation to ensure a consistent vial alignment.

Implementation Details. The experiments were conducted for both the presented methods by dividing the dataset into 4 separate and non-overlapping sequence splits. For each split, each training set consists of $2,000$ sequences, while each test set consists of 426 sequences.

For what concerns ResNet-18 with voting and concat, we used SGD with momentum as optimizer, a learning rate of 0.01, ReduceLROnPlateau as scheduler, and a batch size of 4 sequences. In this case, convergence is achieved after a total of 30 epochs.

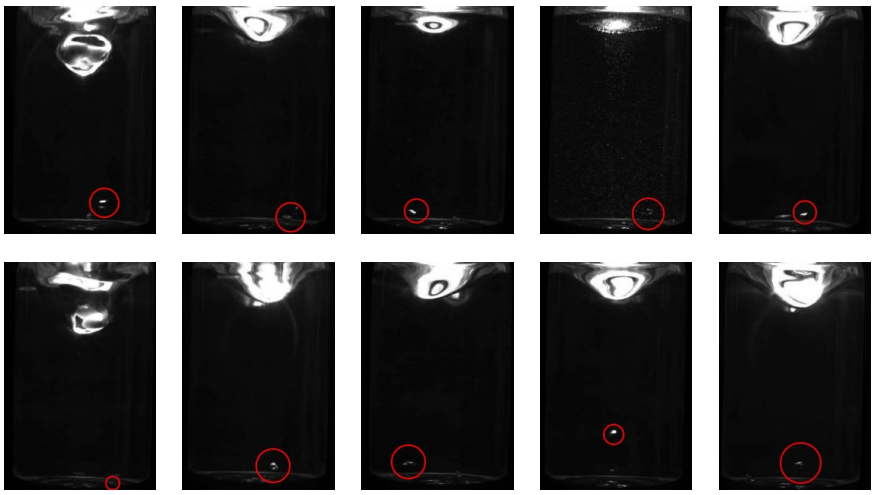


Fig. 5. Sample images from a sequence of a vial containing sand impurities.

For DSMIL, instead, we used Adam as optimizer, keeping a fixed learning rate of 0.0002 during the training and a batch size of 1. The feature extractor (ResNet-18) is trained using the SimCLR framework on each frame of all the sequences. To achieve convergence, DSMIL is trained for a total of 120 epochs. Both DSMIL and ResNet are trained using NVIDIA Tesla K80 as GPU.

Results. The classification results are summarized in Tab. 1. For each method, we reported accuracy, precision, recall, and F1-score computed on the test set, averaged across the five classes. Additionally, we computed the average inference time on a single sequence. We used 4-fold cross-validation to evaluate the model's performance more robustly and mitigate the risk of overfitting to a specific subset of the data. Thus, the reported results consist of the average metrics computed across all the folds. The results suggest that all models reach good performance on this classification task; in particular, the best-performing method is DSMIL, which reaches an accuracy of 99.53%. DSMIL misclassifies only a few sequences confusing *brown* particles sample as vials' without impurities. This occurred because these types of impurities consist of very tiny burnt plastic pieces. Comparing the two aggregation methods used for ResNet experiments, we observed that concatenation is slightly more effective than majority voting.

Table 1. Comparison of different methods on our dataset.

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

We noticed that instances where majority voting failed were due to a misclassification of vials with an impurity as pure vials. This happens because very small impurities (Fig. 5) are only visible in specific frames of the sequence, leading to most of them being assigned the "no impurities" class. Thus, we can conclude that, particularly for challenging-to-detect defects, using concatenation before the ResNet-18 fully connected layers is preferable.

5 Conclusion

In conclusion, this study addresses the critical issue of detecting visible particles in pharmaceutical liquid vials using advanced deep-learning techniques. Over the years, some traditional algorithms, such as SVM and k-means clustering have been explored. More recently, deep learning techniques have outperformed the latter, improving the safety of the final products. In this work, we introduce two methodologies, leveraging ResNet-18 and DSMIL, to classify four types of impurities. To gap the absence of publicly available dataset we also create a new dataset (which is partially released) comprising sequences of images captured from rotating vials, enhances research in this area by providing valuable data for future comparisons. Our methodologies, trained on this dataset, reaches impressive results, with a maximum accuracy of 99.53%, and 99.52% of F1-score.

Future Work. The proposed methodologies exhibited exceptional performance in the designated task, achieving near-optimal scores in multi-class classification. Future research will pivot towards the localization and detection of impurities rather than solely focusing on classification, thereby augmenting the pipeline with explanatory capabilities. Moreover, this allows to classify each detection with its own class, and identify different kind of impurities within the same sample. Another direction of research could focus on improving the inference time in order to obtain real-time performance in a production environment.

Acknowledgements. This work was supported by the University of Modena and Reggio Emilia and Fondazione di Modena, through the FAR 2023 and FARD-2023 funds (Fondo di Ateneo per la Ricerca).

References

1. Abdellah, H., Ahmed, R., Slimane, O.: Defect Detection and Identification in Textile Fabric by SVM Method. IOSR Journal of Engineering **4**(12), 69–77 (2014)
2. Bontempo, G., Porrello, A., Bolelli, F., Calderara, S., Ficarra, E.: DAS-MIL: Distilling Across Scales for MIL Classification of Histological WSIs. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. pp. 248–258. Springer (Oct 2023)
3. Bukofzer, S., Ayres, J., Chavez, A., Devera, M., Miller, J., Ross, D., Shabushnig, J., Vargo, S., Watson, H., Watson, R.: Industry Perspective on the Medical Risk of Visible Particles in Injectable Drug Products. PDA Journal of Pharmaceutical Science and Technology **69**(1), 123–139 (2015)

4. Calvo, C., Micarelli, A., Sangineto, E.: Automatic Annotation of Tennis Video Sequences. In: Joint Pattern Recognition Symposium. pp. 540–547. Springer (2002)
5. Carbonneau, M.A., Cheplygina, V., Granger, E., Gagnon, G.: Multiple instance learning: A survey of problem characteristics and applications. Pattern Recognition **77**, 329–353 (2018)
6. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A Simple Framework for Contrastive Learning of Visual Representations. In: International Conference on Machine Learning. pp. 1597–1607. PMLR (2020)
7. Cornia, M., Baraldi, L., Serra, G., Cucchiara, R.: SAM: Pushing the Limits of Saliency Prediction Models. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 1890–1892 (2018)
8. Ding, F., Zhuang, Z., Liu, Y., Jiang, D., Yan, X., Wang, Z.: Detecting Defects on Solid Wood Panels Based on an Improved SSD Algorithm. Sensors **20**(18), 5315 (2020)
9. Dobashi, K., Saito, M., Hayashi, T.: Advanced quality control of quartz parts for semiconductor equipment based on the food industry's well-established QC methodology (HACCP). In: 2008 International Symposium on Semiconductor Manufacturing (ISSM). pp. 29–32. IEEE (2008)
10. Fang, J., Wang, Y., Wu, C.: Binocular automatic particle inspection machine for bottled medical liquid examination. In: 2013 Chinese Automation Congress. pp. 397–402. IEEE (2013)
11. Ge, J., Wang, Y., Zhou, B., Zhang, H.: Intelligent Foreign Particle Inspection Machine for Injection Liquid Examination Based on Modified Pulse-Coupled Neural Networks. Sensors **9**(05), 3386–3404 (2009)
12. Ge, J., Xie, S., Wang, Y., Liu, J., Zhang, H., Zhou, B., Weng, F., Ru, C., Zhou, C., Tan, M., et al.: A System for Automated Detection of Ampoule Injection Impurities. IEEE Transactions on Automation Science and Engineering **14**(2), 1119–1128 (2015)
13. Ge, Y., Liu, Y., Xu, C.: Particle Detection of Complex Images Based on Convolutional Neural Network. In: 2022 41st Chinese Control Conference (CCC). pp. 7228–7233. IEEE (2022)
14. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 580–587 (2014)
15. Gross, M.A.: The Danger of Particulate Matter: In Solutions for Intravenous Use. Drug Intelligence **1**(1), 12–14 (1967)
16. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. In: 2016 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778 (2016)
17. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely Connected Convolutional Networks. In: 2017 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4700–4708 (2017)
18. Islam, M.J., Basalamah, S.M., Ahmadi, M., Sid-Ahmed, M.A.: Computer Vision-Based Quality Inspection System of Transparent Gelatin Capsules in Pharmaceutical Application. Am. J. Intell. Syst **2**(1), 14–22 (2012)
19. Jia, H., Murphey, Y.L., Shi, J., Chang, T.S.: An Intelligent Real-time Vision System for Surface Defect Detection. In: Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004. vol. 3, pp. 239–242. IEEE (2004)
20. Johnson, J.L., Padgett, M.L.: PCNN Models and Applications. IEEE Transactions on Neural Networks **10**(3), 480–498 (1999)

21. Kekre, H., Mishra, D., Desai, V.: Detection of defective pharmaceutical capsules and its types of defect using image processing techniques. In: 2014 International Conference on Circuits, Power and Computing Technologies [ICCPCT]. pp. 1190–1195. IEEE (2014)
22. Li, B., Li, Y., Eliceiri, K.W.: Dual-Stream Multiple Instance Learning Network for Whole Slide Image Classification With Self-Supervised Contrastive Learning. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 14318–14328 (2021)
23. Li, Y., Han, Z., Xu, H., Liu, L., Li, X., Zhang, K.: YOLOv3-Lite: A Lightweight Crack Detection Network for Aircraft Structure Based on Depthwise Separable Convolutions. Applied Sciences **9**(18), 3781 (2019)
24. Lin, T., Yu, Z., Hu, H., Xu, Y., Chen, C.W.: Interventional Bag Multi-Instance Learning On Whole-Slide Pathological Images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 19830–19839 (2023)
25. Liu, H., Wang, Y., Duan, F.: Glass Bottle Inspector Based on Machine Vision. International Journal of Computer and Information Engineering **2**(8), 2682–2687 (2008)
26. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C.: SSD: Single Shot Multibox Detector. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. pp. 21–37. Springer (2016)
27. Liu, Y., Chen, S., Tang, T., Zhao, M.: Defect Inspection of Medicine Vials Using LBP Features and SVM Classifier. In: 2017 2nd International Conference on Image, Vision and Computing (ICIVC). pp. 41–45. IEEE (2017)
28. Mathavan, S., Kumar, A., Kamal, K., Nieminen, M., Shah, H., Rahman, M.: Fast segmentation of industrial quality pavement images using laws texture energy measures and k-means clustering. Journal of Electronic Imaging **25**(5), 053010–053010 (2016)
29. Mazaheri, M., Saggu, M., Wuchner, K., Koulov, A.V., Nikels, F., Chalus, P., Das, T.K., Cash, P.W., Finkler, C., Levitskaya-Seaman, S.V., et al.: Monitoring of Visible Particles in Parenteral Products by Manual Visual Inspection—Reassessing Size Threshold and Other Particle Characteristics that Define Particle Visibility. Journal of Pharmaceutical Sciences **113**(3), 616–624 (2024)
30. Meenu, M., Kurade, C., Neelapu, B.C., Kalra, S., Ramaswamy, H.S., Yu, Y.: A concise review on food quality assessment using digital image processing. Trends in Food Science & Technology **118**, 106–124 (2021)
31. Noh, H., Hong, S., Han, B.: Learning Deconvolution Network for Semantic Segmentation. In: 2015 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1520–1528 (2015)
32. Panariello, A., Porrello, A., Calderara, S., Cucchiara, R.: Consistency-Based Self-supervised Learning for Temporal Anomaly Localization. In: European Conference on Computer Vision. pp. 338–349. Springer (2022)
33. Pollastri, F., Maroñas, J., Bolelli, F., Ligabue, G., Paredes, R., Magistroni, R., Grana, C.: Confidence Calibration for Deep Renal Biopsy Immunofluorescence Image Classification. In: 2020 25th International Conference on Pattern Recognition (ICPR). pp. 1298–1305. IEEE (Jan 2021)
34. Pollastri, F., Parreño, M., Maroñas, J., Bolelli, F., Paredes, R., Ramos, D., Grana, C.: A Deep Analysis on High Resolution Dermoscopic Image Classification. IET Computer Vision **15**(7), 514–526 (2021)

35. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You Only Look Once: Unified, Real-Time Object Detection. In: 2016 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 779–788 (2016)
36. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence **39**(6), 1137–1149 (2016)
37. Ritter, C., Spilger, R., Lee, J.Y., Bartenschlager, R., Rohr, K.: Deep Learning For Particle Detection And Tracking In Fluorescence Microscopy Images. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). pp. 873–876. IEEE (2021)
38. Roberti, I., Lovino, M., Di Cataldo, S., Ficarra, E., Urgese, G.: Exploiting Gene Expression Profiles for the Automated Prediction of Connectivity between Brain Regions. International Journal of Molecular Sciences **20**(8), 2035 (2019)
39. Roy, S., Sangineto, E., Demir, B., Sebe, N.: Deep Metric and Hash-Code Learning for Content-Based Retrieval of Remote Sensing Images. In: IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium. pp. 4539–4542. IEEE (2018)
40. Szatmári, I., Schultz, A., Rekeczky, C., Kozek, T., Roska, T., Chua, L.O.: Morphology and autowave metric on CNN applied to bubble-debris classification. IEEE Transactions on Neural Networks **11**(6), 1385–1393 (2000)
41. Tulbure, A.A., Tulbure, A.A., Dulf, E.H.: A review on modern defect detection models using DCNNs–Deep convolutional neural networks. Journal of Advanced Research **35**, 33–48 (2022)
42. Unnikrishnan, S., Donovan, J., Macpherson, R., Tormey, D.: Machine Learning for Automated Quality Evaluation in Pharmaceutical Manufacturing of Emulsions. Journal of Pharmaceutical Innovation **15**, 392–403 (2020)
43. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is All you Need. Advances in Neural Information Processing Systems **30** (2017)
44. Vishwanatha, C., Asha, V., More, S., Divya, C., Keerthi, K., Rohaan, S.: A Survey on Defect Detection of Vials. In: Proceedings of International Conference on Data Science and Applications: ICDSA 2022, Volume 1. pp. 171–186. Springer (2023)
45. Wang, S., Zhuo, Q., et al.: Detection of Glass Chips in Liquid Injection Based on Computer Vision. In: 2015 International Conference on Computational Intelligence and Communication Networks (CICN). pp. 329–331. IEEE (2015)
46. Wang, Y., Liu, M., Zheng, P., Yang, H., Zou, J.: A smart surface inspection system using faster R-CNN in cloud-edge computing environment. Advanced Engineering Informatics **43**, 101037 (2020)
47. Yi, J., Zhang, H., Mao, J., Chen, Y., Zhong, H., Wang, Y.: Pharmaceutical Foreign Particle Detection: An Efficient Method Based on Adaptive Convolution and Multiscale Attention. IEEE Transactions on Emerging Topics in Computational Intelligence **6**(6), 1302–1313 (2022)
48. Zhang, H., Li, X., Zhong, H., Yang, Y., Wu, Q.J., Ge, J., Wang, Y.: Automated Machine Vision System for Liquid Particle Inspection of Pharmaceutical Injection. IEEE Transactions on Instrumentation and Measurement **67**(6), 1278–1297 (2018)
49. Zhou, X., Wang, Y., Xiao, C., Zhu, Q., Lu, X., Zhang, H., Ge, J., Zhao, H.: Automated Visual Inspection of Glass Bottle Bottom With Saliency Detection and Template Matching. IEEE Transactions on Instrumentation and Measurement **68**(11), 4253–4267 (2019)