**Deep learning-based non-intrusive detection of instabilities in formulated liquids**

Maurizio De Micco1, Diego Gragnaniello2, Fabio Zonfrilli3, Massimiliano M. Villone1\*, Giovanni Poggi4, Luisa Verdoliva5, Vincenzo Guida3

*1 Dip. di Ingegneria Chimica, dei Materiali e della Produzione Industriale, University of Naples Federico II, 80125 Naples, Italy; 2 Dipartimento di Ingegneria dell'Informazione ed Elettrica e Matematica applicata, University of Salerno, 84084 Fisciano (SA), Italy; 3 The Procter and Gamble Company, Brussels Innovation Center, 1853 Grimbergen, Belgium; 4 Dip. di Ingegneria Elettrica e delle Tecnologie dell’Informazione, University of Naples Federico II, 80125 Naples, Italy; 5 Dip. di Ingegneria Industriale, University of Naples Federico II, 80125 Naples, Italy*

*\*Corresponding author E-Mail: massimiliano.villone@unina.it*

**1.Introduction**

Emulsions are metastable systems composed by at least two immiscible liquids and they have many industrial applications. Because of their thermodynamic instability, they have always represented a challenge for formulators and manufacturers. Emulsion stability is influenced by both chemical and physical factors and can be promoted by using emulsifying agents such as surfactants. This approach leads to the generation of formulated liquids, the prediction of whose long-term stability can be quite tricky due to both their complex chemical composition and the uncertain conditions to which they may be exposed. Instabilities lead to changes in the appearance, consistency, and performance of the products, thus an important aspect in their design is ensuring physical stability over shelf life. In the industrial field, the assessment of the physical stability of formulated liquids is typically done by placing samples of products in transparent jars and observing them at regular intervals of time under known operating conditions, e.g., temperature and pressure [1]. Observations are typically done by expert researchers and can be hardly repeated. This method allows to monitor potential long-term problems that might arise at given operating conditions. Visual observations can be eventually coupled with rheological measurements. This is a time-consuming task, which is ubiquitous in all the industrial sectors involving formulated liquids.

In order to automatize the process of stability assessment, time-lapse videos of the samples can be acquired and analyzed at a later time, which guarantees the repeatability of observations and allows studying also the temporal evolution of phenomena. In this work we leverage a large dataset of frames extracted by such time-lapse videos and use it to train several deep CNNs to perform frame-level instability detection. The dataset is strongly unbalanced, since it includes a very large number of stable samples with respect to unstable ones. Hence, we investigate different possible approaches to address this issue in order to improve the performance of the model. We also add results on interpretability to better understand which parts of the image impacted most on the final decision.

In recent years, image analysis methods have proven effective for addressing a number of chemical engineering-related problems, such as the prediction of physical phenomena and properties. This trend has further accelerated with the advent of deep learning, with its unprecedented performance. For example, Viitanen et al. [2] trained a CNN to predict the occurrence of specific events in dry foams by using just a single input frame. Likewise, a CNN combined with an auto- encoder has been employed by Khor et al. [3] to characterize microfluidic systems. The CNN has been trained to extract shape descriptors of emulsion droplets, whereas the auto- encoder has been trained to predict their break-up.

Besides good performance, deep learning-based image analysis is highly appreciated in several industrial applications because of its non-invasiveness. As an example, Hadikhani et al. [4] proposed a non-intrusive method to measure with high accuracy the flow rate of a water/oil emulsion and the concentration of each component of the system through a deep CNN. A hybrid deep CNN - Long Short Term Memory (LSTM) network has been proposed by Lin et al. in [5] with the aim of predicting the membrane elasticity and viscosity of microcapsules from their dynamic deformation when flowing in a branched microchannel.

The high throughput of CNNs at inference time make them especially suitable for real-time tasks. Chu et al. [6] employed a CNN to classify newly-formed droplets in a microencapsulation system. In their experimental setting, a class weighting scheme is adopted in order to penalize misclassifications of under-represented classes in the loss function. A similar problem has been addressed by Anagnostidis et al. [7], who trained a deep CNN to count the number of micro-objects contained in droplets, cells or other systems at the time of image acquisition.

Instabilities of liquid formulations can be interpreted as product defects. Defect detection is a growing topic in the field of industrial applications and there are great efforts to automatize this process. CNNs have been investigated for this task by Wang et al. [8], who proposed a CNN composed of two major parts, one for classification and one for defect detection, while Zhou et al. [10] implemented a Region-based CNN to improve accuracy and speed of fabric defect detection.

Several image analysis techniques have been explored for the assessment of emulsion instability. Hosseini et al. [11] have used image processing and visual observations to maximize the stability of Arabic gum emulsions by calculating parameters such as creaming index, color, and average dimension of the droplets. Unnikrishnan et al. [12] adopted different machine learning methods, such as Random Forests, Multinomial Logistic Regression, and Neural Networks to perform classification into 4 categories on a set of emulsion micrographs previously labelled by experts. In this work, we study the effectiveness of CNN-based classifiers to detect instabilities in liquid formulations through the analysis of macroscopic images. To the best of our knowledge, this is the first time low-resolution images are automatically analyzed to detect liquid formulation instabilities.

**2. Methods**

We collected a large dataset of images extracted from time- lapse videos and labeled them thanks to expert researchers. In particular, the evolution of formulated liquids with various compositions when exposed to different operative conditions was observed. These experiments are of great importance to assess the so-called ‘shelf life’ of a product before its marketing. To this aim, images of the vials were periodically acquired to store snapshots of the content status. Images were cropped and stored by a fully automatic monitoring system, thus obtaining a time series of images for each experiment, with spatial resolution and temporal span changing from a series to another. For each series, a subset of 10 images equally spaced in time was selected and these were manually labeled as stable or unstable by a team of experts. Note that, during this process, the experts took advantage of the full series to reliably tell if and when an instability appears. Among the unstable samples, cases of phase separation, creaming, flocculation, cracking, and crystallization are present. Some examples of early, mid, and late snapshots of different failures are reported in Fig. 1. After discarding a small subset of ambiguous images for which the experts did not reach an agreement, as well as those erroneously cropped, we obtained a dataset comprising 32,078 stable and 4,857 unstable images, extracted from 6,341 time series. The dataset classes are heavily unbalanced, with 85% stable and 15% unstable samples, most of which show phase separation. This is a common problem in industrial applications since non-defective products outnumber defective ones [13].



**Figure 1.** Early (top), mid (center), and late (bottom) stage of different experiments. From left to right: a stable sample and samples showing cracking (Cra), creaming (Cre), flocculation (F), and phase separation (PS).

The models we used rely on EfficientNet [14], a class of recently proposed CNNs that have been successfully employed in various computer vision applications. In order to train and validate the networks, a pre-processing pipeline was designed. Firstly, cropped images of different dimensions were resized to the same resolution, namely, 224 × 224 pixels. Since the average crop resolution is 150 × 240 pixels, this resizing operation did not produce significant information loss for the vast majority of the samples. After that, the whole dataset was split into training, validation, and test subsets by adopting a 70%-15%-15% proportion. The partitioning was carried out on whole time series rather than on individual images, so as to ensure that images from the same sequence, likely similar to one another, belong to the same subset. This avoids biases due to the presence of similar images in both training and validation/test sets.

Our dataset is heavily unbalanced, which may prevent the trained models from correctly generalizing to test data. To overcome this problem, we considered several approaches proposed in the literature. On one hand, we tested different loss functions beyond the classical cross-entropy loss. More specifically, we weighted the cross-entropy loss by using weights proportional to the inverse of the number of samples for each class, so as to discourage errors on the minority class. Alternatively, we adopted the focal loss [15] to down-weigh examples whose labels are easily guessed by the network, likely those of the majority class. On the other hand, we synthetically increased the number of samples of the minority class. At each iteration, the mini-batch is composed by sampling images of the training set so as to balance the classes.

To further reduce the risk of poor generalization due to the unbalanced dataset, we combined the above mentioned strategies with augmentation techniques. We tested both geo- metrical transformations, like horizontal flipping and rotation, and color/illumination variation. In particular, we introduced random variations in hue, saturation, contrast, and brightness.

**3. Results and discussion**

For a preliminary set of experiments, we selected EfficientNet-B0, which ensures a performance almost as good as the deeper architectures of the family, but has a smaller number of parameters, and therefore is trained much faster. We trained it by combining different strategies in terms of loss functions, dataset balancing, and augmentation profiles. In this regard, we defined a *light* profile, involving only geometrical transformations, and a *strong* profile, which combines geo- metrical and color transformations. For comparison, we also trained the network without any form of augmentation. Results are reported in Tab. I. To assess the various solutions, we measure performance in terms of per-class weighted accuracy, which is not affected by the class unbalance, F-measure, and Pd@Q%. The latter is the probability of correctly detecting the minority class of unstable sample (true positive rate) when the probability of false alarm (false positive rate) is fixed at Q%, where Q here is 1 or 5. When considering the network trained using the classical cross-entropy loss, the light augmentation uniformly improves performance with respect to the case without any augmentation. In particular, this configuration achieves the highest F-measure, Pd@1%, and Pd@5% (boldface entries) among all the preliminary experiments. On the contrary, the training does not benefit from the strong augmentation. Probably, random color/illumination variations distort important features for this specific task. As for the mini- batch balancing, it appears to improve the performance in the absence of augmentation, providing more controversial results in the other cases, with improvements in terms of weighted accuracy achieved in particular with strong augmentation. Weighted cross-entropy loss and focal loss do not improve performance.

We then proceeded to explore alternative architectures. We selected two different training configurations, light augmentation with cross-entropy loss and strong augmentation with balanced batches. The former provides best FM, Pd@1% and Pd@5%, while the latter achieves comparable results with respect to other configurations, nevertheless it is interesting to understand to which extent including balanced batches can effectively represent an improvement.



**Table I.** Preliminary experiments using EfficientNet-B0. For each augmentation setting, the classical cross-entropy (XE) loss, its weighted version (W. XE), the focal loss, and the balanced batch approach (BB XE loss) are tested. Results with † are statistically similar to the top score (bold).

We trained deeper and wider EfficientNet models, as well as the popular ResNet50, and the architecture proposed in [17]. We also added two baseline approaches based on the Random Forest classifier trained on handcrafted features, namely the Local Binary Patterns (LBP) [18] and the Scale Invariant Descriptor (SID) [19] encoded using the Bag of Words (BoW) paradigm [20]. Results are reported in Tab. II. First of all, they confirm our initial preference for the EfficientNet architectures which dominate ResNet50 under all performance metrics and training settings. In any case, all CNN-based methods are much superior to handcrafted based approaches. A second notable result is that EfficientNet-B0 outperforms almost uniformly the deeper versions B1-B3, with both light and strong augmentation. This was somewhat unexpected, considering results reported in the literature for similar tasks, but can be explained with the relatively small size of the available training set, especially for the unstable class, probably insufficient to optimize a complex network with a large number of parameters. the comparison between training configurations confirms light augmentation to be preferable in all cases. This result suggests to pay special attention on testing proper forms of augmentation for this problem.



**Table II.** Comparisons of architectures under two selected augmentation / loss settings: light augmentation with cross- entropy loss (top), strong augmentation with balanced mini- batch cross-entropy loss (bottom). Results with † are statistically similar to the top score (bold).

Besides the quantitative performance metrics, we also pro- vide a visual interpretation of the result provided by the network. By relying on the Grad-CAM [16] method, we obtain a heat-map spotting the image regions that mostly contributed to the network decision. Example results from the test set are shown in Fig. 2 where images characterized by different types of instability are displayed together with the corresponding heatmaps. It appears that the network, although trained in a weakly supervised manner, with no information on the location of the instability, correctly focuses on the small regions of the input image where instability artifacts arise.



**Figure 2.** Sample images from the test set together with their GradCAM heatmaps. From left to right: stable (S), creaming (Cre), flocculation (F), and phase separation (PS). In all unstable cases artifacts are correctly spotted in the heatmap, while for the stable case there is no specific region that is highlighted.

**4. Conclusions**

We investigated CNN-based image analysis methods for non-intrusive detection of instabilities in formulated liquids. To the best of our knowledge, this is the first time deep learning is used to address this problem. Results are already very encouraging, confirming the potential of this approach. Nonetheless, there is much room for improvements. First of all, a larger and more refined dataset is necessary to train modern CNN architectures, and more specific losses and forms of augmentation need to be tested. After that, our focus will shift on more challenging problems, such as fine-grained instability classification, and early prediction of instability onset over time series.

**References**

1. G. Zografi, J. Soc. Cosmet. Chem. 33 (1982)
2. L. Viitanen, J.R. Mac Intyre, J. Koivisto, A. Puisto, and M. Alava, Phys. Rev. Research 2 (2020)
3. J.W. Khor, N. Jean, E.S. Luxemberg, S. Ermon, and S. K. Y. Tang, Soft Matter 15 (2019) 1361-1372
4. P. Hadikhani, N. Borhani, S.M.H. Hashemi, and D. Psaltis, Sci. Rep. 9 (2019)
5. T. Lin, Z. Whang, W. Wang, and Y. Sui, Soft Matter 17 (2021) 4027-4039
6. A. B. Chu, D. Nguyen, A. D. Kaplan, and B. Giera, Applications of Machine Learning (2019).
7. V. Anagnostidis, B. Sherlock, J. Metz, P. Mair, F. Hollfelder, and F. Gielen, Lab Chip 20 (2020) 889-900
8. T. Wang, Y. Chen, M. Qiao, and H. Snoussi, Int. J. Adv. Manuf. Technol. 94 (2018) 3465–3471
9. M. Wieler, and T. Hahn, DAGM (2007)
10. H. Zhou, B. Jang, Y. Chen, and D. Troendle, AI4I (2020) 52-55
11. A. Hosseini, S.M. Jafari, H. Mirzaei, A. Asghari, and S. Akhavan, Carbohydr. Polym. 168 (2015) 1-8
12. S. Unnikrishnan, J. Donovan, R. Macpherson, and D. Tormey, Chem. Eng. Res. Des. 166 (2021) 281-294
13. R. Shimizu, K. Asako, H. Ojima, S. Morinaga, M. Hamada, and T. Kuroda, AI4I (2018) 27-30
14. M. Tan, and Q. V. Le, ICML (2019) 6105-6114
15. T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, ICCV (2017) 2999-3007
16. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, ICCV (2017) 618-626
17. T. Wang, Y. Chen, M. Qiao, and H. Snoussi, Int. J. Adv. Manuf. Technol. 94 (2018) 3465–3471
18. T. Ojala, M. Pietikainen, and T. Maenpaa, IEEE PAMI 24 (2002) 971–987
19. I. Kokkinos, M. Bronstein, and A. Yuille, INRIA RR-7914 (2012)
20. D. Gragnaniello, G. Poggi, C. Sansone, and L. Verdoliva, Pattern Recognit. Lett. 82 (2016) 251-257