Application of Machine Learning for Product Stability Characterization and Control

May 12, 2026

Clark Merchant, PhD

Lumetics

Session Description and Objectives

- Particle classification algorithm options for isolating and reporting changes to stability-indicating particle populations
- Pros and cons of machine learning algorithms for routine particle classification
- Example use cases

Biography and Contact Information

Clark Merchant, PhD
CTO, Lumetics Inc.

Working in the field of imaging-based particle analysis, particle classification algorithm development, and deployment of production-scale data management and analysis solutions in the pharmaceutical industry for nearly two decades.

Email: clark.merchant@lumetics.com
www.lumetics.com

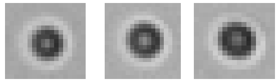
Problem statement

- Complex particle systems make isolating and reporting statistically significant changes to stability indicating particle populations (e.g. protein aggregates) very difficult.

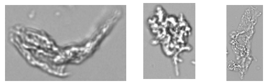
- Advanced classification algorithms are often required but can present deployment challenges.

Wide range of particle types (many not indicative of degradation)

Inherent

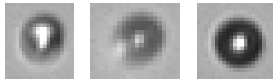


Fatty Acids

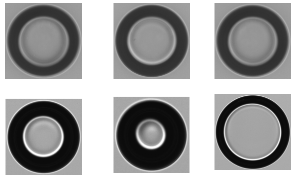


Protein Aggregates

Intrinsic

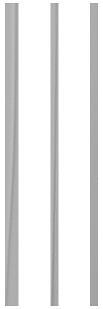


Silicone Oil

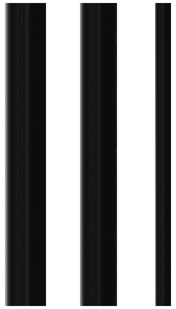


Air Bubbles

Artifacts



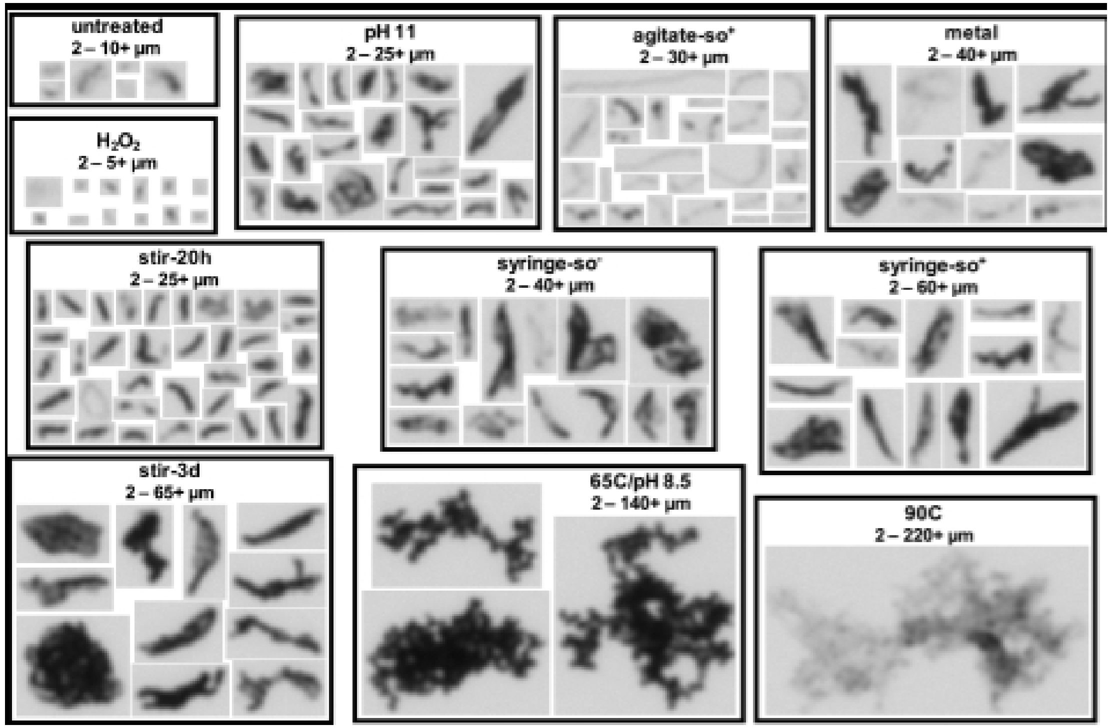
Schlieren Lines



‘Edges’

‘Transparent’

Wide range of subspecies (e.g. aggregates due to stress mechanism)



Classification and characterization of therapeutic antibody aggregates

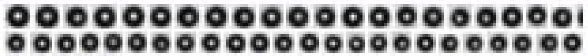
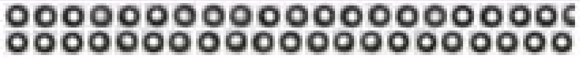
Marisa K Joubert [1], Quanzhou Luo, Yasser Nashed-Samuel, Jette Wypych, Linda O Narhi

Affiliations + expand

PMID: 21454532 PMCID: PMC3137085 DOI: 10.1074/jbc.M110.160457

Wide range of image variations (e.g. due to fluid/particle delta RI)

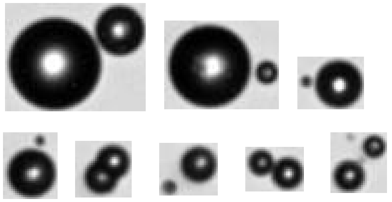
Table 1 **Delta RI in solutions spiked with silicone (100% expected)**

Sample Type	RI at 589 nm	Delta RI from Silicone (1.41)	MFI Counts/mL >= 5 µm	% Separated	Silicone images
Water	1.333	0.078	15381	99%	
0.2% w/v PS80	1.333	0.077	16485	99%	
200 mM Histidine	1.340	0.070	22378	97%	
Placebo	1.345	0.065	16833	98%	
High-concentration non-aggregated protein	1.379	0.032	10728	61%	
41% w/v Trehalose	1.383	0.028	156	48%	
55% w/v Glycerol	1.398	0.012	1243	63%	
65% w/v Gliceról	1.411	0.000	531	63%	Not detected

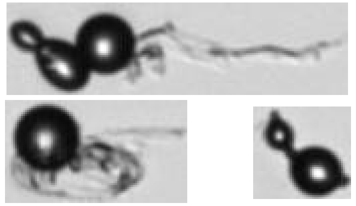
Aspects of silicone separation in protein-silicone mixtures using S-factor-style electronic filter

E. Torres [1], A. Mehta [1] M. Lisowski [1], L. Rosner [1], P. Kolhe [1], C. Merchant [2], D. Thomas [2]
[1] Pfizer Pharmaceutical R & D; [2] Lumetics Inc.

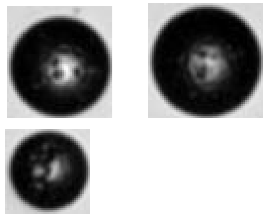
And agglomerates…!



Silicone Oil
Doublets



Silicone Oil/Protein
Agglomerates



Silicone Oil/Precipitate
Agglomerates

Particle classification using morphological parameter filters

Sub Populations

1 **Protein (AR)** : AspectRatio <=0.9

2 **Silicone Oil (AR)** : AspectRatio >0.9

3 **Silicone Oil (SF, MFI 5200)** : ((ECD >= 1 AND ECD < 2.75 AND SFMinusCutoff4 >= 0) OR (ECD >= 2.75 AND ECD < 6 AND SFMinusCutoff5 >= 0) OR (ECD >= 6 AND SFMinusCutoff6 >= 0))

4 **Protein (SF, MFI 5200)** : ((ECD >= 1 AND ECD < 2.75 AND SFMinusCutoff4 < 0) OR (ECD >= 2.75 AND ECD < 6 AND SFMinusCutoff5 < 0) OR (ECD >= 6 AND SFMinusCutoff6 < 0))

5 **Protein (Geodesic AR, Circ)** : GeodesicAspectRatio < 0.8 AND Circularity < 0.9

6 **Silicone Oil (Geodesic AR, Circ)** : GeodesicAspectRatio >= 0.8 AND Circularity >= 0.9

Add Sub-Population

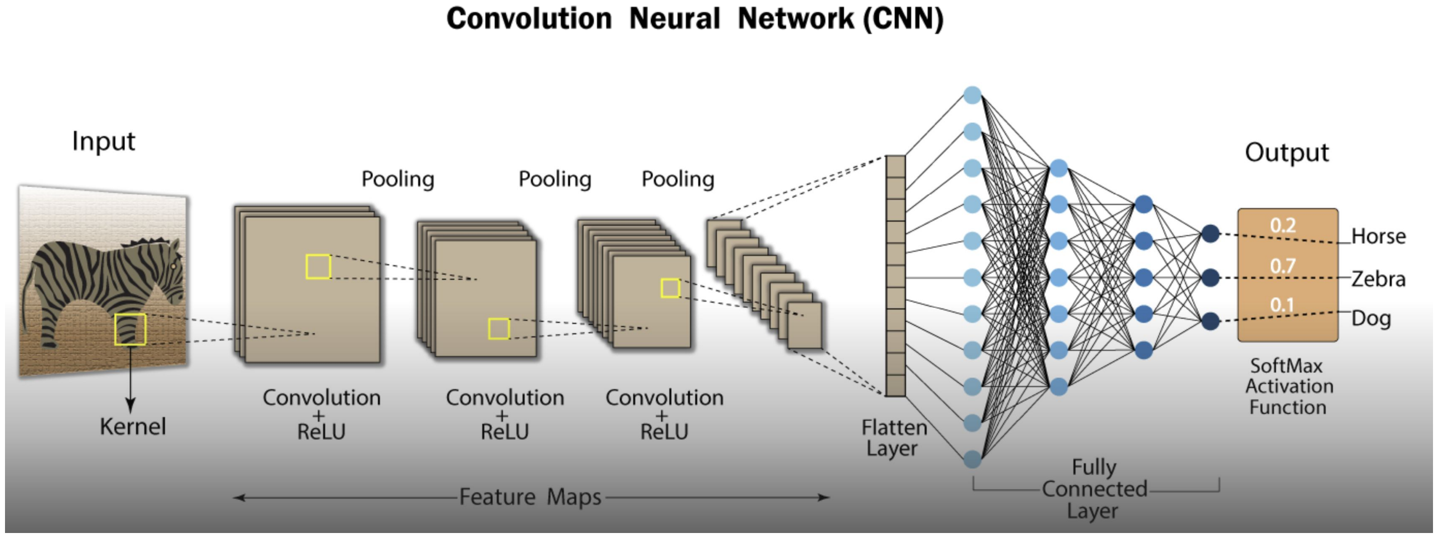
* Lumetics LINK import method option

Easily programmed into automated data management systems, and very fast data processing speeds! Often highly effective for ‘simple’ particle systems such as irregular shaped particles vs. circular/round However, struggle with particles with similar morphology and smaller particles reaching instrument detection limits, or complex particle systems with many particle types

Particle classification using machine learning

CNN Advantages (Convolution Neural Network)

- Utilize entire particle image vs. extracted morphological parameters
- Ideal for multiple particle populations
- Ideal for particles that appear by the human eye to be similar (e.g. ‘small’ particles)
- Proven to deliver high classification accuracies
- Potential to obtain classification accuracies closer to the detection limit of imaging system



Accuracy of machine learning

- Differentiation of challenging particle types for smaller size ranges or similar morphologies can yield surprisingly good results!
  - Protein Aggregates, Silicone Oil, Free-Fatty Acids down to 5 µm [1]
  - Protein Aggregates generated from different stress conditions / containers [2,3]

- Models can detect differences that may be undetectable by humans, especially for small particle sizes

- Small particle sizes represent the vast majority of particles in a sample, so challenges are in creating fast, accurate models that can process thousands or millions of particles with very low image quality/resolution

Examples of 2-5 um particles with different stress conditions

1. Image Classification of Degraded Polysorbate, Protein and Silicone Oil Sub-Visible Particles Detected by Flow-Imaging Microscopy in Biopharmaceuticals Using a Convolutional Neural Network Model, Fedorowicz et al. (2023)
2. Machine Learning & Statistical Analyses for Extracting and Characterizing "Fingerprints" of Antibody Aggregation at Container Interfaces from Flow Microscopy Images , Daniels et al. (2020)
3. Particle formation in response to different protein formulations and containers: Insights from machine learning analysis of particle images, Milef et al. (2024)

Machine learning deployment considerations

Implementation Challenges (for rapid accurate routine application)

1. Utilize imaging system output (image frame + data file)
2. Process many 10's of thousands of particles quickly
3. Support multiple models (imaging system, sub-species, size range, formulation)
4. Permit comparing to conventional particle classification techniques
5. Detect and correct model inaccuracies
6. Aggregate for comprehensive analysis/reporting (incl. metadata)
7. Overlay and compare to complimentary techniques to aid in RCA
8. Detect when model updates are required
9. Facilitate thumbnail image selection and downloads to support model updates
10. 21 CFR Part 11 compliance

Scalable production machine learning particle analysis deployment systems exist and should be considered…!

Example processing speed comparison…

For illustration only using typical hardware, processing times depend on server hardware, network latencies, etc.

Algorithm	Particle Count	Processing time/particle	Total Processing Time
Default Import (No Populations)	100k	--	--
Existing filter using morphological param filter – 3 Subpops	100k	--	+2.5mins
ML optimized for deployment (3 subpops)	100k	6ms	+10 mins
ML optimized for accuracy only (3 subpops)	100k	41ms	+65 mins

Optimizing models for speed is important for model deployment during routine production testing of many samples

Do this by reducing model complexity (lower number of parameters, reduced input image size), while maintaining high accuracy

If requirements can be met without deploying ML models, overall processing/analysis time is quicker!

Take Aways…

- Utilize ‘simple’ morphological parameter filters when classification accuracy permits this (fast and easy).

- Consider ‘complex’ morph param filters such as S-factor or other techniques[1]; they are also fast and easy, and data management systems exist to aide in easily developing and applying cutoff functions!

- If your particle system/research objectives require machine learning level classification accuracies, ensure that you optimize the model for speed as well as accuracy, and deploy in a ‘production’ data management and analysis environment optimized for ML application.

1. Assessment of subvisible particles in biopharmaceuticals with image feature extraction and machine learning, Maharjan et al. (2024)

Questions

Please contact:

Clark Merchant, clark.merchant@lumetics.com
David Thomas, dave.thomas@lumetics.com

Thank You!

Lumetics