Using artificial intelligence and imaging flow cytometry to create a yeast identification model

Vincent Genty[1], Alessandro Serra[2], Livine Duban[2]
[1]Amarok Biotechnologies, Saint Malo, France.
[2] Luminex, a DiaSorin company, 's-Hertogenbosh, the Netherlands.

Introduction

The objective of this preliminary work was to create a statistically significant classification model of mixed yeast populations, using Imaging flow cytometry with Amnis® ImageStream®X MKII coupled with Amnis® AI and IDEAS® ML artificial intelligence software.

Five yeast strains were analyzed in order to generate an identification model determined by Artificial Intelligence (AI) using convolutional neural networks or linear discriminant analysis (LDA).The models are built by feeding the software with the morphometric characteristics of each channel for each acquired cell.
We combined two models: one based on brightfield features validated by statistical analysis of the identity of each strain predicted by the model and its actual class; the second using an LDA algorithm adding the use of autofluorescence measurements.
The calculated 'super-parameters' allow for maximum separation between the different strains when analyzing mixed populations.

Materials and Methods

Yeast Strains

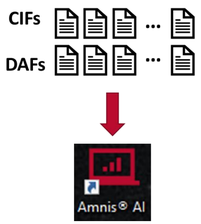
Five strains were used to create the analysis model: German Ale, B. Brut, Pichia, LG Monaco, 1H.

Image Acquisition

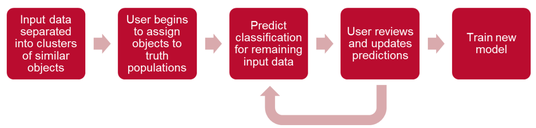
Files were collected using the Amnis® brand ImageStream®X MkII imaging flow cytometer equipped with the 405 nm, 488 nm and 642 nm lasers with 20X, 40X and 60x magnification.
Each strain was ran individually to collect sample morphometric features which have been used in analysis to create the analysis model.

Image Analysis

The collected files were analyzed with the Amnis®AI and Amnis IDEAS® software.

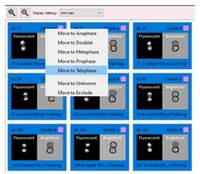


RIF/CIF files are processed in IDEAS® to remove debris and optionally identify truth populations. Those files are then opened in Amnis® AI and truth populations are assigned, step 1 to 4. User reviews and updates predictions as much as needed until training a model. Then, user will validate the model by reviewing the statistics.

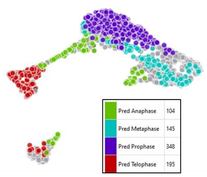


Step 1: Cluster
Unsupervised clustering of similar images

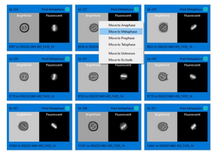
Step 2: Assign Truth Images
Using clustered data



Step 3: Predict
Supervised prediction of truth label



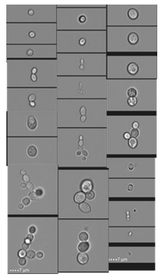
Step 4: Assign Truth Images
Using predicted data



Results

Amnis® AI workflow

	Researcher	Amnis® AI
Define Experiment	X	
Label Truth Data	X	X
Create Classification Model		X
Apply Model		X
Report Accuracy		X
Validate Results	X	



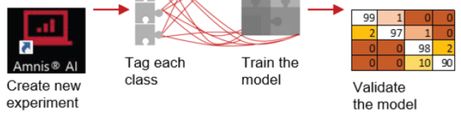


Figure 1. **Amnis® AI analysis workflow**. User will define the experiment, acquire the data, load them into Amnis® AI, train and validate the model

Yeast strains Morphology

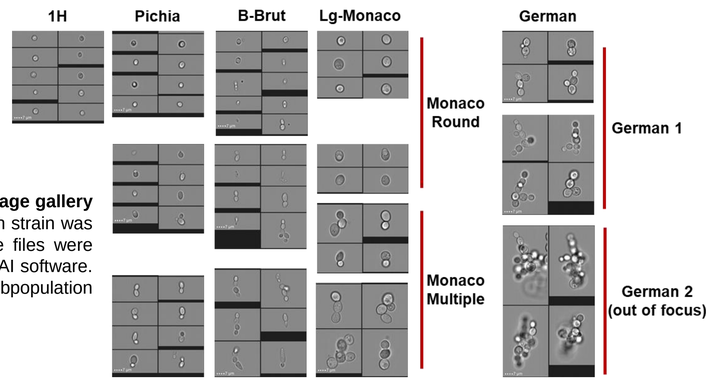
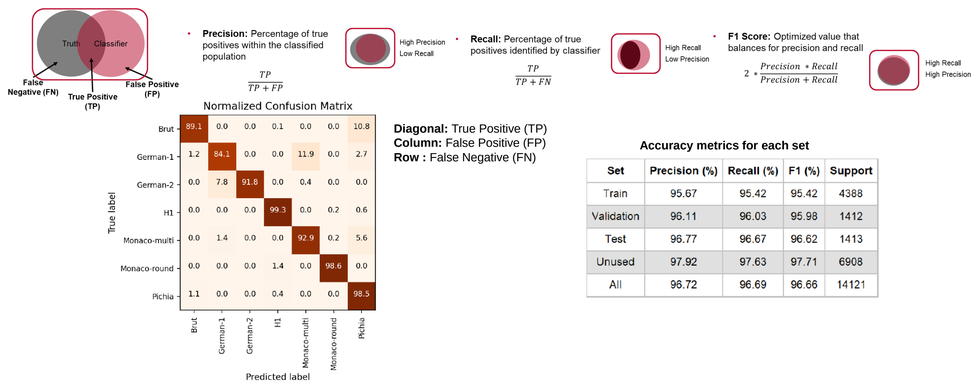


Figure 2. **Sample brightfield image gallery of the five different strains**. Each strain was ran individually and Amnis Image files were collected and analyzed in Amnis® AI software. Each five population and/or subpopulation was used to define a class.

Model Validation: Interpreting the statistics



Diagonal: True Positive (TP)
Column: False Positive (FP)
Row : False Negative (FN)

Accuracy metrics for each set

Set	Precision (%)	Recall (%)	F1 (%)	Support
Train	95.67	95.42	95.42	4388
Validation	96.11	96.03	95.98	1412
Test	96.77	96.67	96.62	1413
Unused	97.92	97.63	97.71	6908
All	96.72	96.69	96.66	14121

Figure 3. **Amnis® AI software results' interface**. The software provides statistical tools used to validate the model:
1- the classification results table is a confusion matrix that displays the distribution of the classified images with respect to the truth populations
and 2- the accuracy table which provides information on the classification efficiency of the model.
A model is generally considered accurate when all predicted probability in the Confusion Matrix and F1 Accuracy values are high for each set or class.

Amnis® AI software Identification Model

Class	Precision (%)	Recall (%)	F1 (%)	Support
Brut	96.27	89.10	92.54	1825
German-1	90.49	84.15	87.20	328
German-2	100.00	91.82	95.74	269
H1	93.37	99.25	96.22	1206
Monaco-multi	93.19	92.87	93.03	589
Monaco-round	99.93	98.60	99.26	4508
Pichia	95.54	98.50	97.00	5396
weighted avg	96.72	96.69	96.66	14121

Table 2: Accuracy metrics for each class

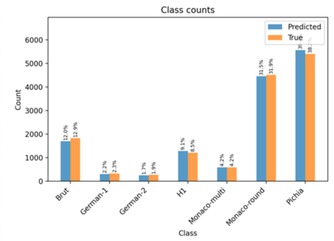


Figure 4. Accuracy results for each class. For each class, F1 values are all above 85%. The precision of the model is 97% (weighted avg) suggesting that it is accurate. The graph shows a comparison between the results for prediction and truth.

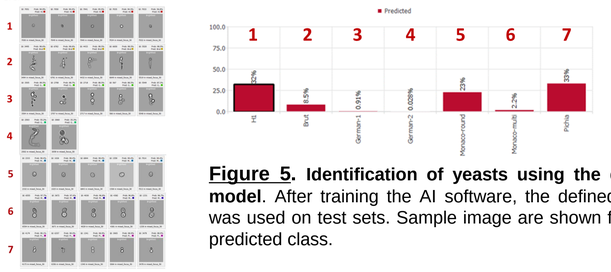
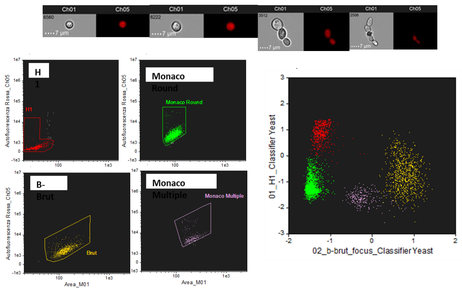


Figure 5. Identification of yeasts using the defined model. After training the AI software, the defined model was used on test sets. Sample image are shown for each predicted class.

Figure 6. Impoving the model by using autofluorescence. As the Amnis® ImageStream®X is recording fluorescence in all available channels, as a preliminary work we have improved the Amnis® AI model using the yeast strain specific autofluorecence and area in µm. This was done using our Amnis® IDEAS® ML software. Specific classifyer (multiparametric features) was computed to separate each of the 5 strains tested.



Conclusions

The statistical evaluation of the accuracy of the first model is 97% (weighted average), i.e. true predictions of class membership. This model is therefore able to identify strains of different sizes and aggregation states. No obvious errors were found on the part of the AI, and the statistical performance was rated at 95%.
As some yeast strains have similar morphometric characteristics, we have removed the ambiguity by adding individual characteristics of autofluorescence and size (area in µm²).

Each strain is traced back to its own 'multiparametric digital signature' based on the characteristics identified as most significant. The AI model can be re-trained by adding additional strains and additional parameters such as specific antibody-type fluorescent markers (Bretta Test), viability staining.

Vincent Genty v.genty@amarokbiotech.com