Development of a Flexible Ensemble Classification System for Microscopy

Tomas J. McIntee[1], Mathieu Therezien [1] and Zachary E. Russell[1*]

[1.] Ion Innovations, Boone, NC, United States.
* Corresponding author: Zach.Russell@ion-innovations.com

In the present work, a highly-flexible ensemble classification system is proposed for broad use in automatic identification of objects in microscopy applications. While a number of plug-and-play solutions for machine learning and object classification exist, this platform is set apart due to being specifically tuned for microscopy applications. This ensemble classifier utilizes a modular collection of machine learning algorithms in conjunction with manually tagged data to identify objects with high accuracy and a meaningful confidence level. The use of a meaningful confidence level in the output identification allows for objects with low confidence to be flagged automatically for human review as part of a larger classification software package (illustrated in Figure 2). This can then be used reflexively to improve the quality of the training data used in the component classifiers [1].

At present, the system supports the use of different combinations of k-nearest neighbor (KNN), weighted KNN, specialized weighted binary KNN, support vector machine (SVM), Gaussian naïve Bayesian (GNB), GNB packs, decision tree, and random forest (RF) classifiers. Note that GNB packs and random forests are themselves ensemble classifiers, respectively ensembles of Bayesian and decision tree classifiers. Each of these classification systems is individually more effective at some identification tasks than others [2, 3, 4].

The classifier system relies on the use of a meaningful space of extracted object features, such as size, brightness, moment, et cetera, which can be corrected or adjusted on the basis of imaging conditions (e.g., magnification, exposure). This extraction process is carried out for both training data and for objects under classification. The modified feature vectors associated with known identified objects are used to train component classifiers. The classifiers may be trained on a subset of data or a subset of features [5].

The ensemble classifier makes use of a high-level decision tree in order to choose a classifier to apply to the object. The results of each classification (e.g., distinguishing optical artifacts from genuine objects) then inform what type of classification should be applied next. In the simplest case, illustrated in Figure 2, all classifiers are applied regardless of intermediate classifications, and decision architecture is implemented in the aggregation process.

Each component classifier has been modified in order to produce a meaningful vector of confidences associated with each possible identification instead of simply producing a categorical calculation. At the end of the process, each object under classification has been associated with a sequence of confidence vectors, which can then be aggregated in several different ways to produce a final identification with an associated confidence. The simplest aggregation process supported is taking an arithmetic or geometric mean of the classification vectors. Various ordinal voting methods and ordinal statistical analyses can also be used to produce a meaningful aggregate confidence value in the final identification by highlighting when an object is similar to two different categories [6, 7].

Because different types of classifiers fail under different circumstances due to violations of different underlying assumptions (e.g., object features being normally distributed), the use of a diverse group of classifiers can both produce higher accuracy classifications and more accurate confidence values associated with those classifications [5, 8, 9]. Since this classification system utilizes a set of object features that can be corrected for different imaging conditions, such as magnifications, it is broadly suitable for microscopy, which tends to involve known magnification levels and frequently identifiable substrates or backgrounds.
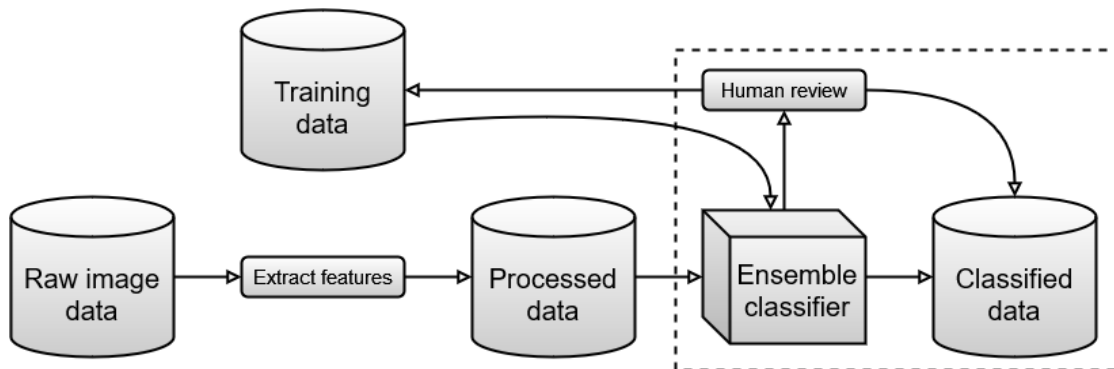

Figure 1. Flow chart showing the ensemble classifier system at a high level. Dotted rectangle corresponds to Figure 2
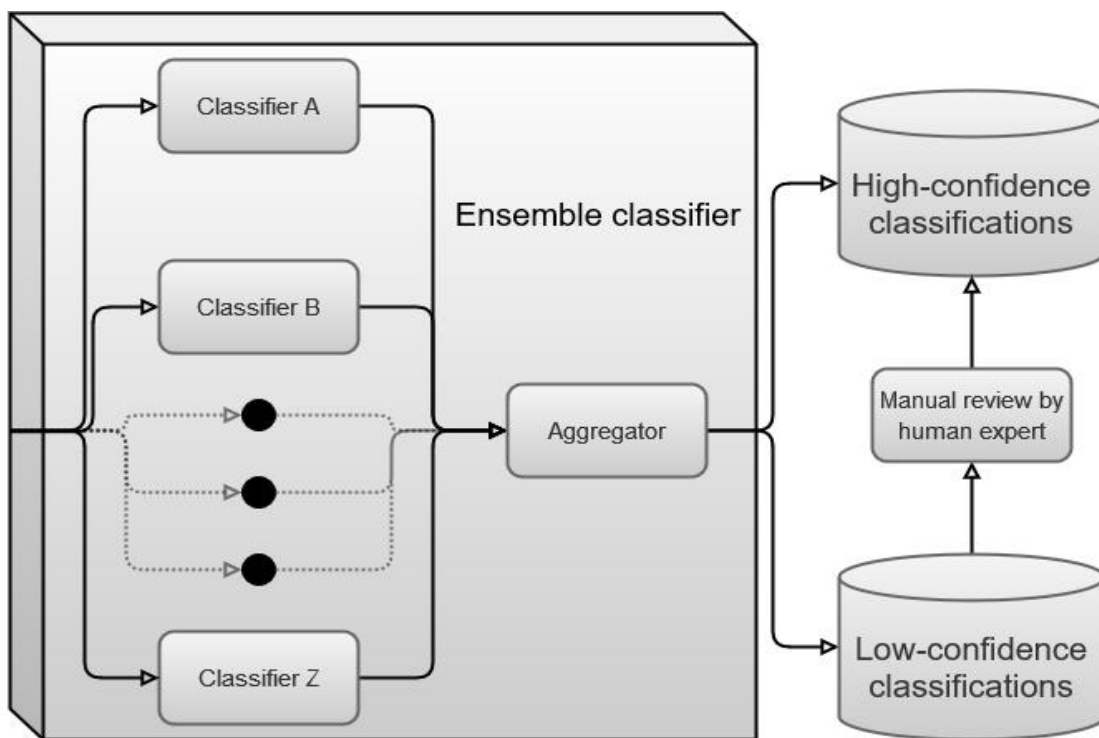

Figure 2. Flow chart showing of an ensemble classifier within the proposed system in greater detail.

References:

[1] Z Russell et al., Appalachian Regional Microscopy Society (AReMS) Annual Meeting (2019).

[2] Y Zhang et al., Sensors **19** (2019), p. 3914. doi:10.3390/s19183914

[3] S Mohapatra et al., Neural Computing and Applications **24** (2014), p. 1887. doi:10.1007/s00521-013-1438-3

[4] M Rashidi and R Wolkow, ACS Nano **12** (2018), p. 5185. doi:10.1021/acsnano.8b02208

[5] L Rokach in "Data Mining and Knowledge Discovery Handbook", eds O Maimon, L Rokach, (Springer, Boston), p. 957.

[6] I Gandhi and M Pandey, International Conference on Green Computing and Internet of Things (2015), p. 399.

[7] E Bauer and R Kohavi, Machine Learning **36** (1999), p.105. doi:10.1023/A:1007515423169

[8] K Tumer and J Ghosh, Connection Science **8** (1996), p. 385. doi:10.1080/095400996116839

[9] L Rokach, Artificial Intelligence Review **33** (2010), p.1. doi:10.1007/s10462-009-9124-7