The application of automated image analysis has significant potential benefits for the interpretation of slide staining, particularly in immunohistochemical protocols where interpretation is based on the more subjective estimation of an overall staining pattern rather than quantitation (such as FISH spot counting) where automation can achieve more accurate results than visual inspection, while eliminating some of the factors that can influence visual interpretation.

HER2 status is clinically important because of its role in breast cancer, and the prevalence of the condition enables studies on large cohorts of patients that should provide a high level of confidence in the results. Therefore the present study, in which the authors seek to determine whether automated image analysis can provide comparable accuracy to visual inspection by a pathologist, has the potential to advance our understanding of the role of automated interpretation in assessing HER2 status. However, as presented, this manuscript has a number of serious errors and omissions that must be addressed in order for its value to be realized.

The paper presents both negative and positive results, and does not unequivocally demonstrate a clinical benefit to the use of automation. Therefore, its principal value lies in identifying as fully as possible the factors in the reported procedures and in the system under study that are responsible for the significant results.

Major compulsory revisions:
1. Table 1: Results for FISH non-amplified cases (n=1084) contains errors:
   a. Visual 1, row (0, 1+, 2+, 3+ and unscorable) adds up to only 1074, not 1084. Should “unscorable” have been 110?
   b. Machine 1, row (0, 1+, 2+, 3+ and unscorable) also adds up to only 1074, not 1084.
   c. Machine 2, row (0, 1+, 2+, 3+ and unscorable) has only 4 entered for every value, total 20.

   This calls into question the integrity of the other statistics reported, and the calculations should be included with any revision.
(2) The level of agreement between the two instruments and between the instruments and the pathologists requires explanation in terms of the observed tissue morphological and staining features, in addition to the statistical discussion. In particular, there is a wide variation in the levels of agreement between the scores from the four different sets of IHC scores and the FISH results. What features were responsible for this? For example, how many specimens exhibited chromosome 17 polysomy? How many showed heterogeneous gene amplification or protein expression in which evaluation of limited regions of the core could have played a role? In their discussion of this, the authors need to separate differences in the actual staining and interpretation (level of staining, numbers of spots, etc.) and underlying mechanisms.

(3) Details of how the auto analysis system was “trained” (parameters used) are necessary, preferably with a visual example or graphic explanation.

Minor essential revisions:

(1) Control for color / white balance is cited as a possible factor in the difference between the two machine results: the authors state “The descriptors of the color and shape of the positive and negative tumor cells were transferred from one system to another, therefore variations in the image analysis results depended only on the scanner settings, i.e. positioning and white balance, but not on the image analysis settings.” Therefore, the procedure for setting the white balance and positioning needs to be given in order to assess the extent to which these did vary and how this variation relates to the results.

Discretionary Revisions:

(1) While not essential for the results, an explanation for the large differences between the 0 and 1+ scores between the visual inspection and the instrument scores would be of general interest since it points to an underlying difference between visual and digital analysis.

Language:

(1) There are some missing definite (“the”) and indefinite (“a”) articles, especially in the final two paragraphs of the discussion, e.g. “between experienced operator and pathologists,” and “with further refinement Ariol system could be used for clinical service.”

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Needs some language corrections before being published

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:
I declare that I have no competing interests.