Response to Reviewer 2 Comments

This study aimed to analyze the cytological features of benign and malignant thyroid nodules with the use of computerized cytomorphometry for the differential diagnosis of FNAC.


Response 1: Thank you for your suggestion! We have reviewed the above papers and cited them and explain the differences between our study and these two in the last two paragraph of Discussion section (page 8, line 211-220) in the revised manuscript: “In our study, the Red-Green-Blue (RGB) color space of image pixels were first converted into the color space of hue (H), saturation (S), and value (V). To the best of our knowledge, this is the first study to evaluate the thyroid FNA with indeterminate categories using color space of HSV. In the study conducted by Celik ZE et al. morphological and chromatic variables were used to establish their computerized cytomorphometry system which included five variables. Among these five variables, only one was chromatic variable and was in grey scale. In another study, Gilshtein H et al. applied 50 variables in their computerized cytomorphometry but the texture of nuclear chromatin was also in gray scale. In addition, our study had collected a largest dataset, 240 cases, so far in the literature including 125 cases of indeterminate cytological categories while the previous studies had only 40 and 58 cases, respectively.” In the Introduction section (page 2, line 53-55), we have also clarified the possible role of the computerized cytomorphometry as follows: “It serves as a sequential reader to the conventional cytological evaluation. We aimed to analyze the cytological features of benign and malignant thyroid nodules using computerized cytomorphometry to aid in the differential diagnosis of FNAC.”

2. Lines 212-216, would the color value of the pixels of the image be varied due to potential variations in the staining techniques? Please clarify the consistency of the staining techniques and how these would affect the image analysis.
Response 2: Thank you for your comment! The Riu’s staining was performed in a standard procedure. We have added this standard procedure in the Materials and Methods section (page 9, line 251-256) “Riu's staining was performed in the following standard procedure: After air dry, the smears were covered with solution A (0.05% methylene blue and 0.17% eosin Y in methanol) for 30 seconds, followed by placement of solution B (0.12% azure I, 0.14% methylene blue, 2.52% Na₂HPO₄·12H₂O and 1.25% KH₂PO₄ in distilled water) for 90 seconds. The stain mixtures were then washed off and the slides were observed under a light microscope.” With this standard procedure, the variations of the color value would be small. Besides, the chromatic features regarding color space were not based on the absolute values of hue, saturation, or value. These features were relative measures comparing color space values of the nuclei and cytoplasm in the same specimen. This should also decrease the variation caused by staining technique. The calculating formulae were moved to the Supplemental material according to the other referee’s suggestion:

\[ \text{NCHR} = \frac{\sum_{(i,j) \in N} H_{ij} / n_N}{\sum_{(i,j) \in C} H_{ij} / n_C}; \text{NCSR} = \frac{\sum_{(i,j) \in N} S_{ij} / n_N}{\sum_{(i,j) \in C} S_{ij} / n_C}; \text{and NCVR} = \frac{\sum_{(i,j) \in N} V_{ij} / n_N}{\sum_{(i,j) \in C} V_{ij} / n_C} \]

NCHR: nuclear-to-cytoplasmic hue ratio
NCSR: nuclear-to-cytoplasmic saturation ratio
NCVR: nuclear-to-cytoplasmic value ratio
nN: number of pixels in N. nC: number of pixels in C.
(i,j): pixel position

Since the formulae have been moved to the supplementary material, we have added a simpler description about this inner comparison in the text of Materials and Methods section (page 10, line 276-279): “The H of nuclei and cytoplasm in the same cytologic specimen was compared and the difference was calculated in a nuclear-to-cytoplasmic hue ratio (NCHR) using a formula presented in the supplementary material. Nuclear-cytoplasmic saturation ratio (NCSR) and nuclear-cytoplasmic value ratio (NCVR) were generated in similar fashion.” However, different staining methods may result in different NCHR, NCSR, and NCVR values. We will re-evaluate these parameters in different staining methods such as Papanicolaou stain in the future study. We have added in Discussion section (page 8, line 233-235): “However, different staining methods may result in different NCHR, NCSR, and NCVR values. We should collect and analyze specimens of other staining methods such as Papanicolaou stain in the future.”
3. It is unclear that which software was used for the image analysis. MATLAB? More information should be provided for the segmentation of nuclei. How the Canny edge detection method incorporated in the image processing software? Is any deep learning method used in the segmentation? To increase the transparency of the method, more information about the image analysis should be included.

Response 3: The computerized analysis was performed using AmCAD-CA (AmCad BioMed Corp., Taipei, Taiwan). The Canny edge detection algorithm was incorporated in the AmCAD-CA software without using deep learning method. The image processing algorithm used in the software can be better seen in the patent filed by the software company and has been added in the Materials and Methods section (page 10, line 269-274) of the revised manuscript: “The image processing and analysis were performed using AmCAD-CA (AmCad BioMed Corp., Taipei, Taiwan). The digital cytological images were segmented into three clusters, including background, cytoplasm, and nucleus, based on pixel value in hue and saturation with Otsu’s method. After segmentation, the background cluster was set to zero, and the Canny edge detection was performed on the remaining image to find the boundary of nuclear area. The detailed algorithm used by the software can be seen in the patent filed by the software company.”

4. According to the footnote of Table 2, statistical tests were performed to evaluate whether the data were normally distributed or not. However, the statistical tests used were not described in the methodology. Please include this information in the section “Statistical Analysis”.

Response 4: The ANOVA was performed and D'Agostino-Pearson test was used to test for normal distribution of each feature. The testing methods used have been added in the revised manuscript in the Materials and Methods section (page 11, line 294).

5. This study did not compare the diagnostic performance between computer-based method with the pathologist-based method in the assessment of thyroid FNAC specimens. In the study, the result showed that the computerized quantification method yielded 100% sensitivity and 17.6-33.3% specificity. However, if the pathologist-based method achieves better result than the computerized quantification method, the clinical application of the computerized method is questionable. The authors should consider comparing the result of pathologist assessment with the computerized method, and discuss the additive value of using computerized method.
Response 5: Thank you for your precious comment! This computerized diagnostic score was developed for the purpose assisting cytopathologists and clinicians in challenging situations of indeterminate thyroid cytology, such as atypia, follicular neoplasm, and suspicious of malignancy.

According to the American Thyroid Association (ATA) guidelines, in the situation of cytological diagnosis of atypia, clinicians should discuss with the patients to decide further management, including repeated fine needle aspiration (FNA), genetic study, or thyroidectomy. In our dataset, we can set the cut-off point of the diagnostic score at 0.523 to get the best overall diagnostic power (sensitivity: 87%; specificity: 70.6%) (Table 4). Less aggressive management such as repeated FNA may be preferred for a diagnostic score less than 0.523. In contrast, more aggressive management such as thyroidectomy may be preferred for a diagnostic score more than 0.523. Or we can set the cut-off point at 0.209 to get the best sensitivity (100%) for not missing any malignant cases while avoiding unnecessary surgical operation, i.e., 17.6% (3 out of 17) benign atypia cases in our dataset could have been prevented from aggressive treatment such as thyroidectomy.

This scenario may also be used in the situation with cytological diagnosis result of suspicious for malignancy (SUSP). We can set the diagnostic score of computerized cytomorphometry at 0.733 to get the best overall diagnostic power (sensitivity: 64%; specificity: 100%) (Table 4). Or we could set the cut-off point at 0.426 to get the best sensitivity (100%) while avoiding extensive thyroidectomy, i.e., 33% (1 out of 3) benign SUSP cases in our dataset, could have been exempted from aggressive treatment.

In the situation of follicular neoplasm (FN)/suspicious of FN (SFN), the overall diagnostic power for differentiation of follicular adenoma (FA) and follicular thyroid carcinoma (FTC) is poor (Table 4). The comparison of the diagnostic score of FA and FTC was shown in figure 2C. The ROC curve of the diagnostic score to differentiate benign and malignant lesions in FN/SFN was shown in figure 1C with the AUC of 0.568 and p value of 0.52. However, if we set the cut-off point of the diagnostic score at 0.075 to get the best sensitivity (100%), we could still avoid 13.6% (3 out of 22) benign FN/SFN cases in our dataset from aggressive treatment.

Briefly, the development of this computerized morphometry is based on the cytologic parameters of PTC and benign follicular lesions (other than follicular adenoma), and was aimed to offer some reference to help further management decision of the doctors and patients in situations of indeterminate cytologic diagnosis. In other words, the prediction model was built based on the histologic diagnostic results and then the cases of indeterminate cytologic diagnosis were used to test the model to see how the model can help the cytologist. Therefore, there was no comparison between the diagnostic
performance of pathologists and the computer in this study. It is supposed that the computer will be used as an assistance (or a sequential reader) when pathologists had indeterminate cytologic diagnosis. (Figure 1B, 1C, 1D, Figure 2B, 2C, 2D, and Table 4). We totally agree with the reviewer that we should discuss the additive value of using computerized method more clearly. So we added the related discussion in the Discussion section (page 7, line 178-201): “This computerized diagnostic score was developed for the purpose of assisting cytopathologists and clinicians in challenging situations of indeterminate thyroid cytology…from aggressive treatment.”

6. The authors should discuss the value of using computerized method in routine clinical practice. Should the computerized method be used for initial screening of the specimens before the pathologist’s assessment, or it is used after the pathologist’s assessment?

Response 6: Thank you for your suggestion! This computerized diagnostic score was developed for assisting cytopathologists and clinicians in challenging situations of indeterminate thyroid cytology, such as atypia, follicular neoplasm, and suspicious of malignancy. According to the American Thyroid Association guideline, in the situation of cytological results of indeterminate thyroid cytology, clinicians should discuss with the patients to decide the next step, including repeated FNAC, genetic study, or thyroidectomy. This computer morphometry was aimed to offer some reference to help further decision of the doctors and patients in situation of indeterminate cytologic diagnosis. Therefore, it is designed to be used after the pathologist’s assessment and when the pathologists’ cytologic diagnosis is indeterminate. We added this discussion in the Discussion section (page 7, line 178-201): “This computerized diagnostic score was developed for the purpose of assisting cytopathologists and clinicians in challenging situations of indeterminate thyroid cytology…from aggressive treatment.”