Response to Reviewer 1 Comments

The paper is interesting however major clinical weaknesses are easily found, probably due to some problems in the material and methods section that should be clarified.

1. The model is designed to differentiate not only PTC nodules but also benign nodular adenomatous goiter and hyperplasia. Moreover other histotypes are not included: MTC, rare neoplasms. In particular what about NIFTP subgroup?

Response 1: Thank you for your important question!

(1) We aimed to analyze the cytological features of benign and malignant thyroid nodules using computerized cytomorphometry to aid in the differential diagnosis of FNAC. In this pilot research, we focused on common follicular cell-derived lesions only, including papillary thyroid cancer, follicular thyroid cancer, and benign follicular cell-derived lesions. We added this explanation of patient selection in the Introduction section (page 2, line 55-59): “Papillary thyroid cancer (PTC) is the most common thyroid malignancy and constitutes up to 90% of differentiated follicular cell–derived thyroid cancers (FCTC) worldwide. In this preliminary study, we focused on the cytological features of PTC and the other type of FCTC, namely, follicular thyroid cancer (FTC), in comparison with those benign follicular cell-derived lesions.” We also changed the title into “Computerized Cytological Features for Papillary Thyroid Cancer Diagnosis – Preliminary Report” to reflect our limited but worthwhile focus on the PTC. Other histotypes such as MTC and anaplastic thyroid cancer will be analyzed in the future once a sufficient sample of specimens is collected. We added this pitfall and future perspective in the Discussion section (page 8, line 225-229):”For a more comprehensive clinical use, we should also collect a sufficient sample of other types of thyroid cancers, such as anaplastic thyroid cancer, poorly differentiated thyroid cancer and medullary thyroid cancers, to be included in the analysis. However, the findings of the current study are still of value in differential diagnosis of PTC, which constitutes a majority of thyroid cancers.”

(2) Following your suggestion, we have also analyzed in the revised manuscript the subtypes of PTC in the sample collection of our study. We found that only five cases were notated by the histologists with follicular variants of PTC (FVPTC). Three of them had non-encapsulated subtype of FVPTC (NFVPTC) and two had mixed classical type of PTC (cPTC) and NFVPTC. The other 105 cases had cPTC. None of the FVPTC was encapsulated. So there was no non-invasive follicular thyroid neoplasm with papillary like nuclear features (NIFTP) in our study. Thus, we are not able to determine how
this NIFTP variant subtype, with a low malignant potential, affects the model in this study. We have added the subtype of PTC in our study in the Results section (page 2, line 65-67): “In patients with pathologic diagnosis of PTC, five cases had follicular variants of PTC (FVPTC). Three of them had non-encapsulated subtype of FVPTC (NFVPTC) and two had mixed classical type of PTC (cPTC) and NFVPTC. The other 105 PTCs were cPTC.”

2. Please show to the readers a direct comparison between particularly challenging situations like follicular adenoma Vs FTC (does the model work?). Or hyperplastic nodules Vs NIFTP.

Response 2: Thank you for your suggestion! This computerized diagnostic score was developed for the purpose of “assisting” cytopathologists and clinicians in challenging situations with cytological results of atypia, follicular neoplasm, or suspicious for 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.

We have shown the above results in the Results section of the previously submitted paper: Figure 1, Figure 2, Table 4, and the text (page 6, line 125-144). But we think the reviewer’s concern is right. Our presentation may not be clear enough to the readers. So we have added a paragraph to more directly clarify how our study results could be applied to the challenging situations 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. According to the American Thyroid Association guideline…avoid 13.6% (3 out of 22) benign FN/SFN cases in our dataset from aggressive treatment.”

As to differentiation of hyperplastic nodules from NIFTP, we are not able to conduct the analysis in the current study due to the lack of the NIFTP cases in our dataset. We have added this limitation of our current study and possible future study in the Discussion section (page 8, line 222-225): “Besides, in our dataset, more than 95% of PTC (105 out of 110) were cPTC and only five cases were follicular variants of PTC. We have noted that the cPTC constitutes 80% of all PTC in the literature and further study is needed especially for the non-invasive follicular thyroid neoplasm with papillary like nuclear features (NIFTP) because of its different malignant potential.”

3. PTC is a dramatically heterogeneous diagnostic category: how many classic? How many follicular variant did they include?

Response 3: We have analyzed the subtypes of PTC in our study as suggested. We found that only five cases had FVPTC. Three of them had non-encapsulated subtype of FVPTC (NFVPTC) and two had mixed classical PTC (cPTC) and NFVPTC. The other 105 cases had cPTCs. None of the FVPTC was encapsulated. We added this description in the Result section (page 2, line 65-67): “In patients with pathologic diagnosis of PTC, five cases had follicular variants of PTC (FVPTC). Three of them had non-encapsulated subtype of FVPTC (NFVPTC) and two had mixed classical type of PTC (cPTC) and NFVPTC. The other 105 PTCs were cPTC.”

4. Please, avoid pathogenetic suppositions as “Whether the increased cytoplasmic saturation is due to an increase in organelles or cellular products of protein and thus correlates with malignancy needs further research. PTC cells had greater mean nuclear elongation than benign follicular cells in our study. This is probably because mitosis would result in elongated nuclear shape and PTC cells show higher rates of
mitosis.” It is very dangerous in our experience to put attention to mitoses in thyroid pathology, except from the Turin criteria for undifferentiated thyroid carcinoma.

**Response 4:** Thank you for your comment! We have removed these supposition as suggested.

5. Figure 4 is quite intriguing, however the reader should be helped in understanding better the various computational morphological criteria (elongation, Nuclear-to-cytoplasmic saturation ratio, polarity)

**Response 5:** Thank you for your suggestion! We have changed Figure 4 and the legend. By including feature visualization of a case of hyperplasia and a case of papillary thyroid carcinoma, we hope the readers are able to understand better the various computational morphological criteria.

**Figure 4.** Visual comparison for clinically important features. The images from a case of hyperplasia (A) and a case of papillary thyroid carcinoma (B) acquired from a thyroid fine-needle aspiration cytology smear (Riu's stain, 400X). The image (red rectangle region of interest)
was cropped for computerized analysis. NCR (nuclear-cytoplasmic ratio visualized with nuclei in purple; cytoplasm in yellow); Intranuclear inclusion (pink); Nuclear size (nuclei segmented by red contours); Elongation (elongated nuclei in red); Polarity (alignment of red major axes of elongated nuclei); NCSR (nuclear-cytoplasmic saturation ratio visualized with saturation contrast between nuclei and cytoplasm).

6. The paper is direct to clinicians, so the mathematical model should be included as supplementary data, since it is very difficult to understand.

Response 6: Thank you for your suggestion! We have moved the mathematical model to the supplementary. The computerized analysis of cytologic features in the Materials and Methods section of the main text was shortened as the following (page 10-11, line 269-289): “The image processing and analysis were performed using AmCAD-CA …calculating formulae were listed in the supplementary material.”

7. On the contrary, the discussion should include the limitations or the possible “plus” in applying the model to really controversial entities like AUS-FLUS and SFN.

Response 7: Thank you for your suggestion! The diagnostic score was helpful in AUS category but not performed well in FN/SFN category. We had already described this in the Results and Discussion sections (page 7, line 204-206), for example: “The cytological features that can differentiate benign from malignant thyroid tumors in AUS/FLUS and SUSP categories of cytological specimens do not perform well for the category of FN/SFN.” To make it clearer to the readers, we have added the limitations and possible values when applying the model to indeterminate entities like AUS and FN/SFN in the Discussion section (page 7, line 178-201): “This computerized diagnostic score was developed for the purpose of assisting…..avoid 13.6% (3 out of 22) benign FN/SFN cases in our dataset from aggressive treatment.”, and in the Discussion section (page 8, line 222-225): “Besides, in our dataset, more than 95% of PTC (105 out of 110) were cPTC and only five cases were follicular variants of PTC. We have noted that the cPTC constitutes 80% of all PTC in the literature and further study is needed especially for the non-invasive follicular thyroid neoplasm with papillary like nuclear features (NIFTP) because of its different malignant potential.”