**Plagiarism Check:**

Though the plagiarism check came back with a 95% score, this was due to the fact that the paper was matched against a preprint made available on our web site. Hence, this is a non-issue.

**General Comments:**

We hope to submit an expanded version of this abstract as a book chapter. We want to provide a reference for people using the data, so we submitted this abstract to serve as documentation for the data. The abstract, if accepted, will be referenced in the release documentation.

**Reviewer 2:**

**===============**

*Section C*: Please state your recommendations and why.

I would suggest a more logical structure for the paper (SPMB paper structure format). Instead of describing FCDP in Part A and TUDP in Part B, followed by a summary in Part C, it would be more effective to integrate the information.

**===============**

We appreciate your suggestion. However, the organization of the data is based on the availability of metadata in the respective datasets. The dataset from Fox Chase Hospital (FCDP) contains detailed metadata, while most of the dataset from Temple Hospital (TUDP) lacks reports or additional information. Therefore, we had to rely on image labels and further research to gather information on tissue types and staining for TUDP. In contrast, FCDP contains abundant metadata, which facilitated our understanding of the data. Considering the significant difference in the available information between the two datasets, we opted to separate them into distinct sections. This choice was made to enhance the users' comprehension of the dataset contents.

**===============**

It seems that the labels of whether the tumors are malignant or benign are needed in this study because to a large extent, the treatment of the tumor depends on knowing this.

**===============**

There are good reasons to sort the data into four groups: low grade, intermediate grade, high grade, and unknown. The "unknown" category is clear – it's for data that doesn't have a grade. The FCDP data comes from a cancer hospital. Except for the "unknown" category, almost all cases in our dataset are cancerous.

Knowing the grade of a tumor is crucial when dealing with cancer. It affects how much treatment costs for both the hospital and the patient. Not all cancers require the same level of treatment. A simple surgery might be enough for a low-grade tumor, while a high-grade one might need chemotherapy. Simply saying if it's cancer or not isn't enough. It won't make things easier for pathologists or reduce costs for patients and hospitals.

In this field, many claim to have systems that diagnose cancer with performance better than humans. The reality, however, is that this determination needs to be a lot more nuanced to be useful to clinicians. Our pathology corpora have been informed by our collaborations with pathologists at Temple Hospital and Fox Chase Cancer Center.

We have rewritten the sections describing the grading systems in an attempt to clarify these points.

**===============**

*Section D*: Please state ways to improve the abstract.

The paper mentions cases where the grade couldn’t be determined and how these were categorized as “unknown”. It would be helpful to elaborate on how these cases might impact the usability of the dataset.

**===============**

We have added several sentences in the summary to address these issues.

**===============**

The paper briefly mentions the possibility of augmenting FCDP with TUDP. Actually, I could not find where they explained this augmenting process. Providing more details on how this augmentation can be done and the potential benefits would be valuable.

**===============**

Table 6 on page 5 explains how to merge datasets based on the same labeling methodology. The main issue is how to combine data with different labeling systems. Table 6 demonstrates how to do this.

**===============**

< *Section E*: Miscellaneous comments.

The first sentence would benefit from stronger support than what is provided by Reference1 and its manner of citation.2. Page 2 of 19 paragraph 2, line 3: The error in the sentence is the repetition of ”shown in shown in.” It should be corrected to: ”The data was classified into the same four cate-gories shown in Table 1 using the information found in columns AF (“Block Level TissueHistology”) and AG (“Gradeclin Desc”).”>

**===============**

It has been fixed.

**===============**

page 2, table 1, While they mentioned the existence of seven types of tumors (four categorized as high-risk and three as general), Table 1 does not reflect these seven types

**===============**

According to a universal tumor grading system [11], cancer grades are classified into four risk groups and three general classes which are presented in Table 1. The risk groups are determined based on the grades as follows: well-differentiated, moderately differentiated, poorly differentiated, and undifferentiated. The well-differentiated category is classified as "low grade," the moderately differentiated category is classified as "intermediate grade," and both poorly differentiated and undifferentiated categories are classified as "high grade." In cases where the grading cannot be determined, they are mentioned as "Grade X" based on the universal grading and considered a separate class called "unknown." We made the final labels for the data based on the class column in the table. However, this can be changed based on the users’ goal for classification. Sometimes users like to do classification based on the grades, or based on the risk group. Table 1 just introduces three ways that users can use the dataset for the classification task. We have changed this part of the paper and added some explanations to clarify it for the users (the last paragraph on page 1 and first full paragraph on page 2).

**===============**

and it raises questions about why the unknown type is denoted as ’X,’ corresponding to 10 in Roman numerals

**===============**

The information is based on the universal grading this is how they show non-gradable options :

<https://www.cancer.gov/about-cancer/diagnosis-staging/diagnosis/tumor-grade#how-tumor-grade-is-determined>

<https://www.mdanderson.org/patients-family/diagnosis-treatment/a-new-diagnosis/cancer-grade-vs--cancer-stage.html>

**===============**

4. Page 2 of 19 paragraph 2, line 5: the period shouldn’t be after “unknown”

**===============**

It has been fixed.

**===============**

5.Table 5 title style is different from other tables>

**===============**

It has been fixed.

**Reviewer 4:**

**===============**

< *Section D*: Please state ways to improve the abstract.

* The abstract mentions an imbalance in the data, but it could provide more details or statistics to quantify this imbalance.

**===============**

Table 3 on page 2 clearly explains the number of samples in each class.

**===============**

More information about the imaging process and any preprocessing steps performed on the data would be useful

**===============**

This has been explained completely in the second reference ( the first release of the TUDP dataset):

B. Doshna, Z. Wevodau, N. Jhala, I. Akhtar, I. Obeid, and J. Picone, “The Temple University Digital Pathology Corpus: The Breast Tissue Subset,” in *Proceedings of the IEEE Signal Processing in Medicine and Biology Symposium* (SPMB), I. Obeid, I. Selesnick, and J. Picone, Eds., Philadelphia, Pennsylvania, USA: IEEE, 2021, pp. 1–3. doi: *10.1109/SPMB52430.2021. 9672275*.

**===============**

The abstract introduces the classification criteria for tumors based on grading systems. Is it verified by an oncology researcher.

**===============**

This is based on the universal tumor grading system :

<https://www.cancer.gov/about-cancer/diagnosis-staging/diagnosis/tumor-grade#how-tumor-grade-is-determined>

<https://www.mdanderson.org/patients-family/diagnosis-treatment/a-new-diagnosis/cancer-grade-vs--cancer-stage.html>

Note also that many of the co-authors are clinical and research pathologists.

**===============**

Are there any potential limitations or challenges associated with the datasets to provide a balanced view for potential users?

**===============**

This biggest drawback to this data is that a large percentage of the data lacks annotations. We strengthened our comments about this in the summary.

**===============**

< *Section E*: Miscellaneous comments.

Providing more context and potential applications for these directions would be helpful. This could be provided if there is a room with respect to the word limit.

**===============**

We have modified the summary to address this. In this abstract, our goal was to introduce the data. The book chapter we plan to write if accepted will provide more background and context for this work.