

## Recent Advances in the Temple University Digital Pathology Corpus

*J. Smith<sup>1</sup>, S. Jones<sup>2</sup>, M. Smith<sup>1</sup>, J. Doe<sup>3</sup>, B. Smith<sup>1</sup> and J. Jones<sup>1</sup>*

1. Neural Engineering Data Consortium, Temple University, Philadelphia, Pennsylvania, USA
2. Center for Corporate Research, Acme Corporation, Los Angeles, California, USA
3. Advanced Technology Laboratory, XYZ Research, New York, New York, USA  
{jsmith, bsmithjjones}@temple.edu, sjones@acme.com, jdoe@xyzr.com

Authors using MS Word should download this template and start editing it directly. It contains most of the document formats that you will need. Paragraphs are justified. We use an 11 pt. Times New Roman font. The space after a paragraph is 12 pt. Line spacing is set to “Single.”

[... some sample text follows ...] The Neural Engineering Data Consortium (NEDC) is developing a large open source database of high-resolution digital pathology images known as the Temple University Digital Pathology Corpus (TUDP) [1]. Our long-term goal is to release one million images. We expect to release the first 100,000 image corpus by December 2020. The data is being acquired at the Department of Pathology at Temple University Hospital (TUH) using a Leica Biosystems Aperio AT2 scanner [2] and consists entirely of clinical pathology images. More information about the data and the project can be found in Shawki et al. [3]. We currently have a National Science Foundation (NSF) planning grant [4] to explore how best the community can leverage this resource. One goal of this poster presentation is to stimulate community-wide discussions about this project and determine how this valuable resource can best meet the needs of the public. [... some sample text follows ...]

Indented text should use a 10 pt. Times New Roman font. A 6 pt. spacing should be used before and after the paragraph:

Filenames:

tudp/v1.0.0/svs/gastro/000001/00123456/2015\_03\_05/0s15\_12345/0s15\_12345\_0a001\_00123456\_lvl0001\_s000.svs

Explanation:

*tudp*: root directory of the corpus

...

*0s15\_12345\_00123456.docx*: the filename for the corresponding case report

Tables and figures and tables are inserted in text boxes and allow text to flow around them. An example of a table is shown below in Table 1. A cross-reference should be used when you refer to the table in the abstract. The table title appears at the top. If the title is one line long, the line is centered. If it is multiple lines long, it is justified. The table should use a Times New Roman 9 pt. font.

Figures are done in a similar fashion. An example of a figure is shown on the following page. References to the figure, such as “please see Figure 1 on the following page,” should be cited as shown, and should be cross-referenced but not hyperlinked [5][6].

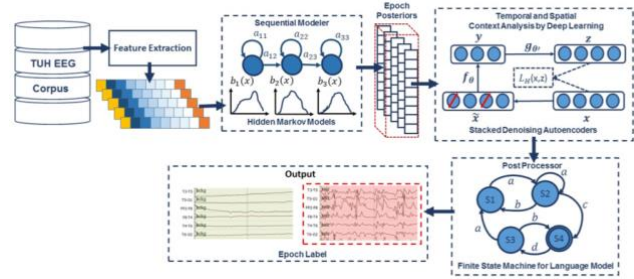
[... some sample text follows ...] We currently recognize fifteen tissue types in the first installment of the corpus. The raw image data is stored in Aperio’s “.svs” format, which is a multi-layered compressed JPEG format [7]. Pathology reports containing a summary of how a pathologist interpreted the slide are also provided in a flat text file format. A more complete summary of the demographics of this pilot corpus will be presented at the

**Table 1.** The TUSZ Corpus (v1.1.1)

Description	Train	Eval
Patients	196	50
Sessions	456	230
Files	1,505	984
No. Seizure Events	870	614
Seizure (secs)	51,140	53,930
Non-Seizure (secs)	877,821	547,728
Total (secs)	928,962	601,659

conference. [... some sample text follows ...] [... some sample text follows ...][... some sample text follows ...] [... some sample text follows ...] [... some sample text follows ...]

[... some sample text follows ...] Another goal of this poster presentation is to share our experiences with the larger community since many of these details have not been adequately documented in scientific publications. There are quite a few obstacles in collecting this data that have slowed down the process and need to be discussed publicly. Our backlog of slides dates back to 1997, meaning there are a lot that need to be sifted through and discarded for peeling or cracking. Additionally, during scanning a slide can get stuck, stalling a scan session for hours, resulting in a significant loss of productivity. Over the past two years, we have accumulated significant experience with how to scan a diverse inventory of slides using the Aperio AT2 high-volume scanner. We have been working closely with the vendor to resolve many problems associated with the use of this scanner for research purposes. [... some sample text follows ...]



**Figure 1.** A hybrid deep learning architecture that integrates temporal and spatial context

Here is an example of how equations should look:

$$Sensitivity (Recall) = (TP / (TP + FN)) \tag{1}$$

$$Specificity (Selectivity) = (TN / (TN + FP)) \tag{2}$$

$$Accuracy = ((TP + TN) / (TP + FN + TN + FP)) \tag{3}$$

$$Precision = (TP / (TP + FP)) \tag{4}$$

References use IEEE format. Examples are listed at the end. A citation in the text should be cross-referenced and enclosed in brackets. References should be numbered according to the order they appear in the text.

The source of funding, if appropriate, must be acknowledged at the end of the document in a section titled Acknowledgements. Please check with your supervisor or research sponsor so that the acknowledgement is written properly. Most funding agencies have very specific requirements for the wording of the funding acknowledgement.

Once you complete the development of your poster, you will submit a final version of your abstract with a copy of the poster attached. Details on that will follow as we proceed through the publication process.

At the end of your abstract, the sections titled Acknowledgement and References should appear. The heading style must follow the format shown below.

**ACKNOWLEDGMENTS**

[... example ...] This material is supported by the National Science Foundation under grants nos. CNS-1726188 and 1925494. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

[... example ...] This research would not be possible without the help of Dr. Nirag Jhala, MD, Professor of Pathology in the Lewis Katz School of Medicine and Laboratory Medicine Director, Anatomic Pathology/Cytology at Temple Hospital. He has been patiently teaching our undergraduates how to annotate pathology slides.

## REFERENCES

- [1] D. Houser et al., “The Temple University Hospital Digital Pathology Corpus,” in *Proceedings of the IEEE Signal Processing in Medicine and Biology Symposium*, 2018, pp. 1–7.
- [2] Leica Biosystems, “Leica Biosystems Aperio AT2 – High Volume, Digital Whole Slide Scanning,” *Leica Biosystems*, 2018. [Online]. Available: <https://www.leicabiosystems.com/digital-pathology/scan/aperio-at2/>.
- [3] N. Shawki et al., “The Temple University Digital Pathology Corpus,” in *Signal Processing in Medicine and Biology: Emerging Trends in Research and Applications*, 1st ed., I. Obied, I. Selesnick, and J. Picone, Eds. New York City, New York, USA: Springer, 2020, pp. 67–104.
- [4] I. Obeid and J. Picone, “CCRI: Planning: Digital Pathology Research Consortium.” CISE Community Research Infrastructure (CCRI), National Science Foundation, January 1, 2020 – June 1, 2021.
- [5] C. Campbell, N. Mecca, T. Duong, I. Obeid, and J. Picone, “Expanding an HPC Cluster to Support the Computational Demands of Digital Pathology,” *Proceedings of the IEEE Signal Processing in Medicine and Biology Symposium*, 2018, pp. 1–2.
- [6] “Aperio ImageScope - Pathology Slide Viewing Software,” *Leica Biosystems*, 2018. [Online]. Available: <https://www.leicabiosystems.com/digital-pathology/manage/aperio-imagescope/>.
- [7] “Aperio Format,” *OpenSlide*, 2018. [Online]. Available: <https://openslide.org/formats/aperio/>.