



WASHINGTON STATE
UNIVERSITY

CREATING THE WORLD'S LARGEST MAMMALIAN HAIR DATABASE TO CREATE A NEW HEALTH DIAGNOSTIC TOOL

Michael G. Fasci (1); Jasson Makkar (1); Tommy Duong (1); Liam Broughton-Neiswanger (2), DVM; Iwona M. Driskell (1), Ph.D.; and Ryan R. Driskell (1), Ph.D.
(1) Washington State University, Pullman, WA and (2) Washington Animal Disease Diagnostic Laboratory, Pullman, WA

**Esch
LEAD**

Learn, Experience,
Apply, and Deliver

ABSTRACT:

Humans have hairs which grow dynamically based on the position and health of the body, affecting their length, width, color, curls, and kinks. The human hair phenome is a combination of all these unique hair fibers as they collectively change over time. Quantifying these characteristics in large amounts without machine learning is incredibly time consuming and inaccurate. Importantly, hair fibers without the root are easy to collect, store, transport, and lacks DNA making hair fiber analysis a promising mechanism for running diagnostic tests. I have expanded a novel deep hair phenomics pipeline to digitally detect, extract, and sort any hair fiber from a high-resolution image. The adaptation of the pipeline can find color value variation of the hairs from root to tip and detect changes in thickness throughout the length of the hair, useful for applications such as determining the number and location of bends. By collecting these characteristics over the lifespan of a mammal and with different diseases and genotypes, we can create a database to compare samples for diagnosis. In conclusion, the continued development of deep hair phenomics has the potential to use hair fibers as a non-invasive diagnostic tool, which is easy to access, use and contribute to.

HAIR SLIDES:

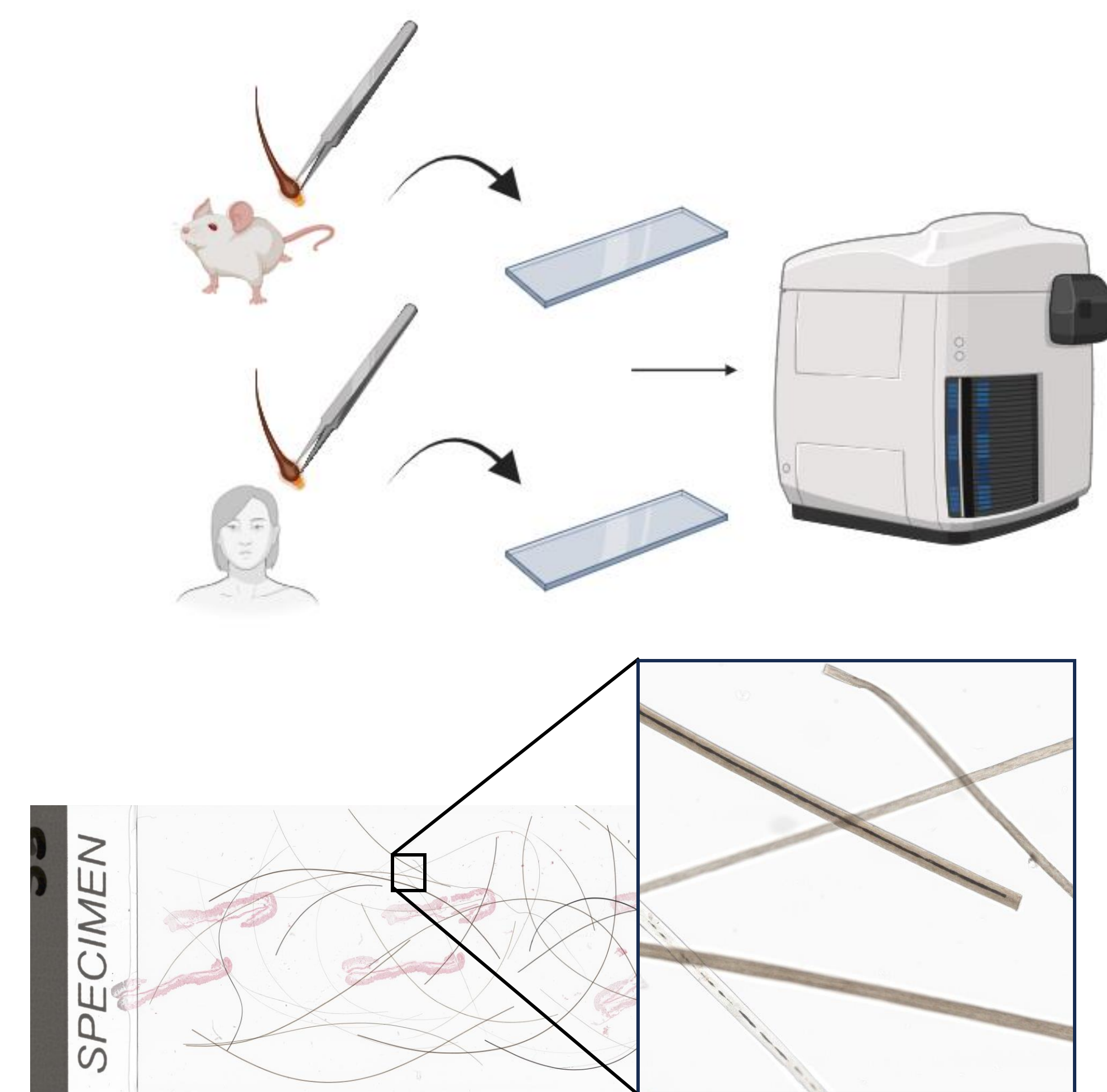
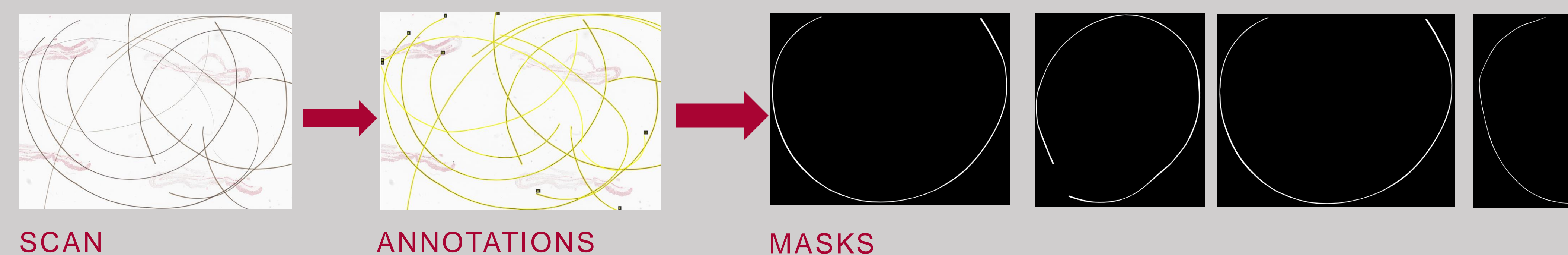


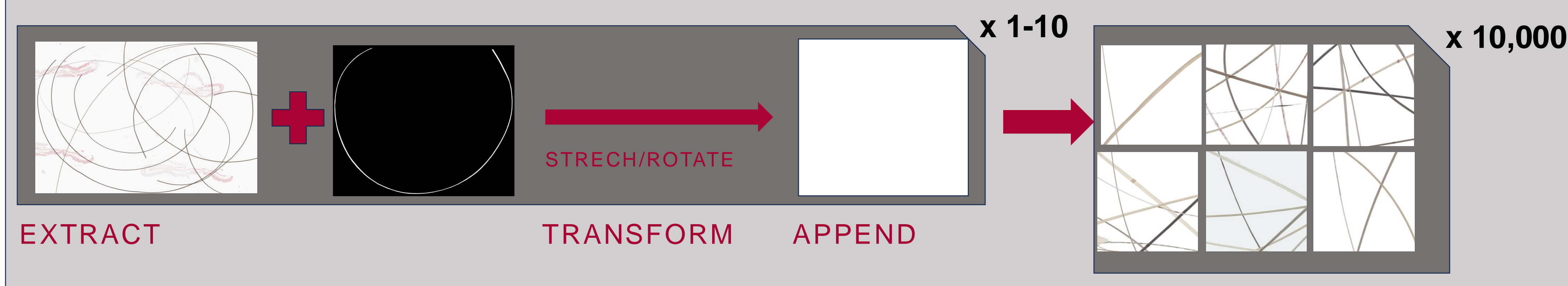
Figure 1.0: Hair was pulled from mice from various time points and parts of the body and placed on microscope slides. The hair was fixed to the slides and imaged at high resolution using an Aperio Scan Scope. Human hair clippings were placed and fixed to slides and imaged in the same process. Made with Biorender.com

CREATING THE MODEL:

HUMAN ANNOTATIONS:



SYNTHETIC DATA:



TRAINING:

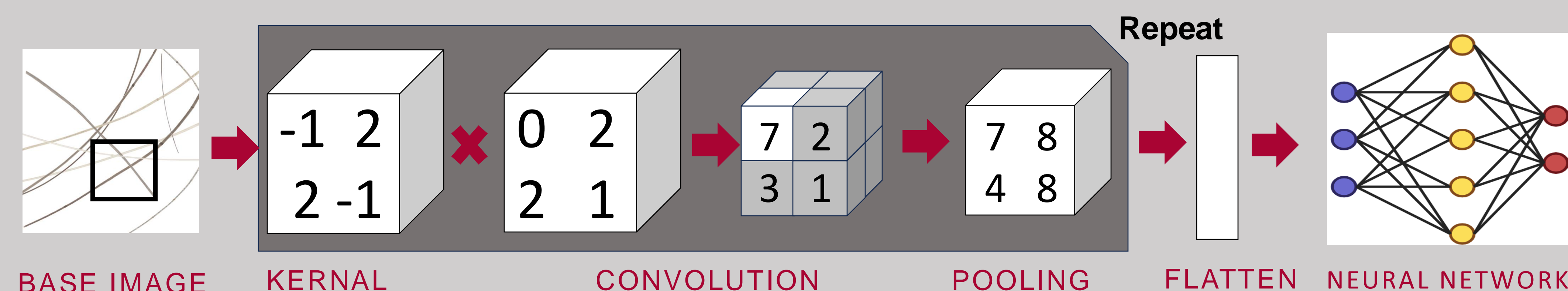


Figure 2.0: Scanned hair slides were then segment annotated by hand. Using the segmentation annotations, masks of each hair were generated. Using the masks and the original images, the hairs were extracted, transformed, and appended to background images to create synthetic data. This synthetic data was then used to train a computer vision model with segmentation and orientated bounding box capabilities.

USING THE MODEL:

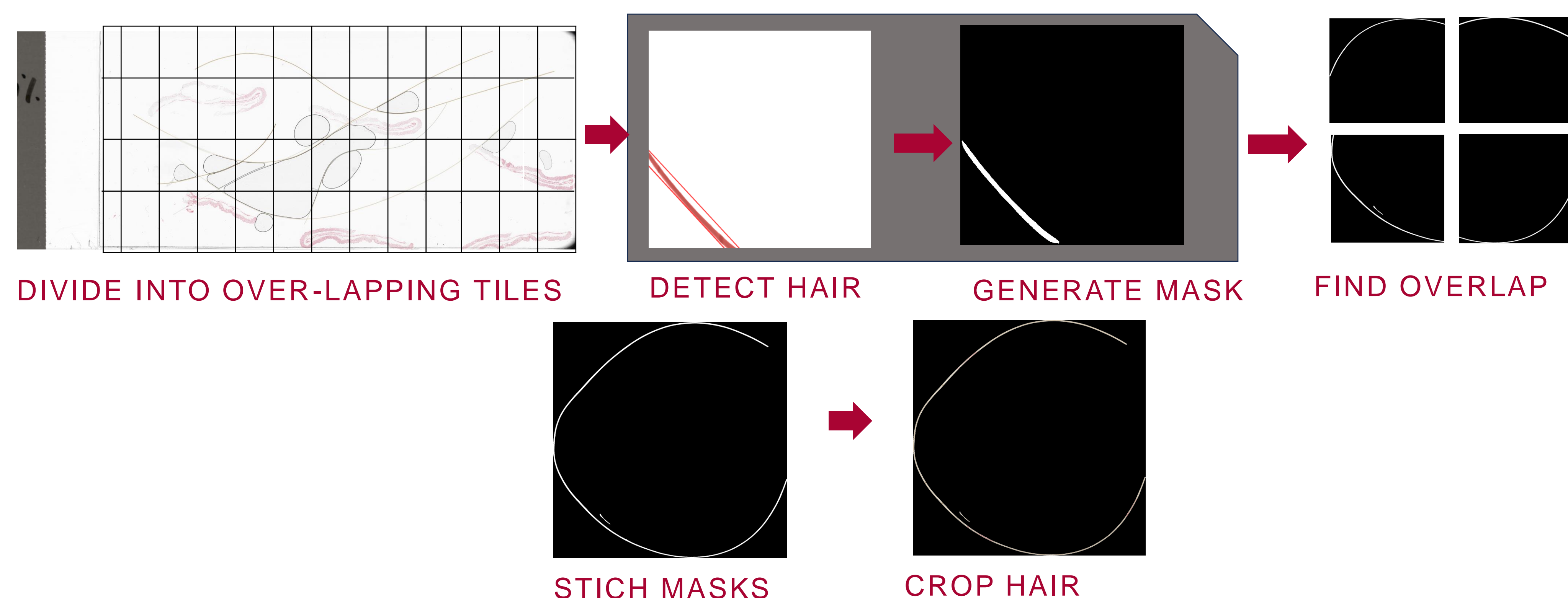


Figure 3.0: Scanned hair images are cut into overlapping tiles to reduce computational load of processing the full image. Each tile is run through the computer vision model to detect the hairs and return a mask cutout. Over-lapping masks are then stitched together so a full hair mask is returned. The stitched mask is used to crop the hair from the original image. This full mask and the cropped image are then used for analysis.

ANALYZING HAIRS:

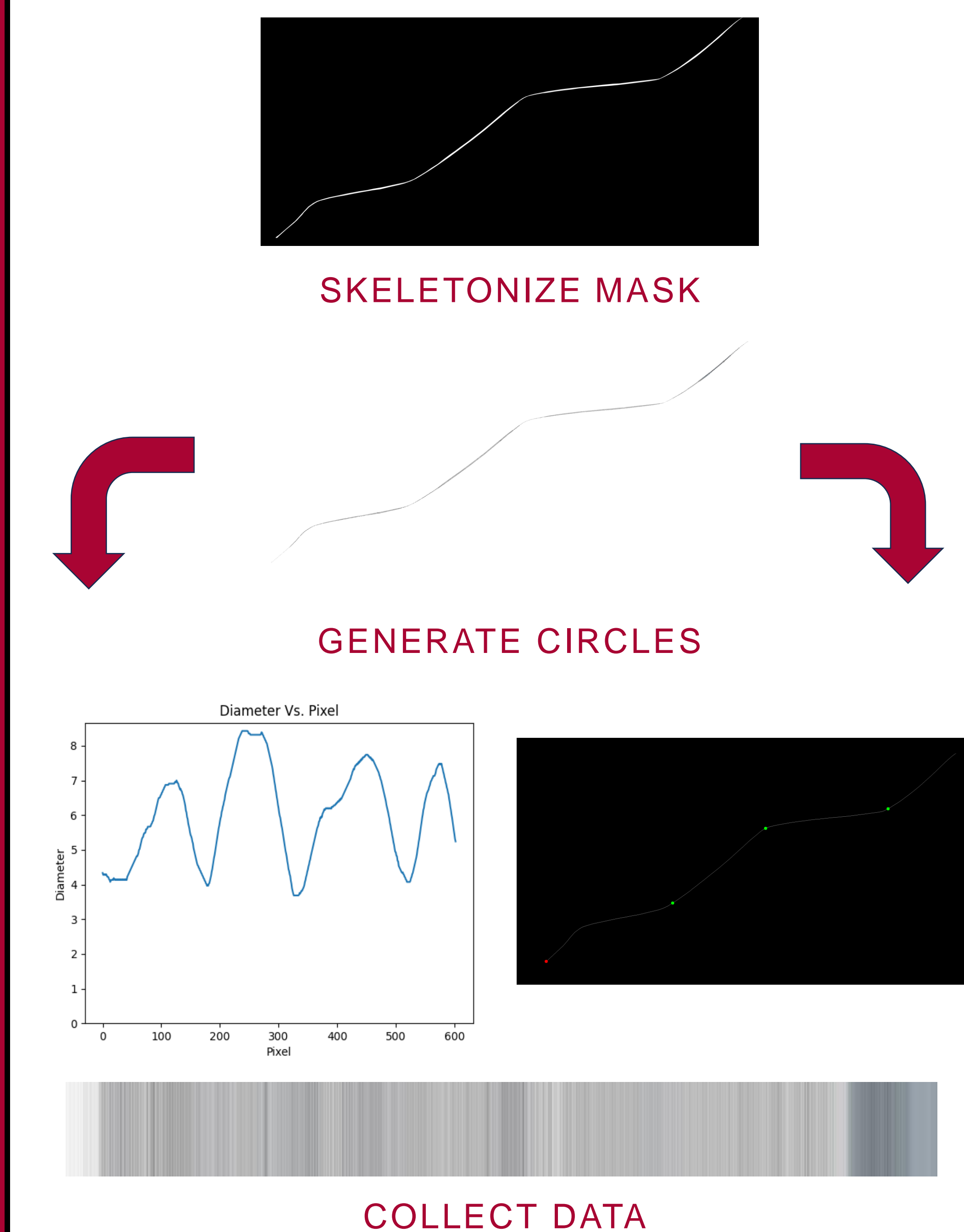


Figure 4.0: Hair masks are skeletonized for processing, creating a pixel wide line down the center of the hair. At each point along the skeleton, a circle is generated with a diameter equal to the mask. The radius is then recorded, and a color average is collected in the circle. The radius data is then smoothed with a moving average. Peaks are identified to locate bends on hairs.

CONCLUSION

Our current fiber pipeline can quickly analyze human and mouse hair, opening up analysis of hair phenomics that was limited before by labor and time. Contributing findings to a database would promote the exploration of hair phenotypes and may be a future diagnostic tool for medicine.

In the future, the process can be refined by improving sample preparation, increasing the recognized species, and adding more detectable traits.

ACKNOWLEDGEMENTS

Images were taken with the Aperio GT450 owned and operated by the Washington Animal Disease Diagnostic Laboratory.

Model training and data generation were performed on the Washington State University high-performance computing cluster, Kamiak.

Funded by a grant from the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health (R25EB027606)