Life hacking of human body at a molecular level and understanding them is the research topic that our SURF project have investigated. However, although such advancing biotechnology and accumulating knowledge of human biology helps refine clinical descriptions of disease and contribute to generating new clinical therapies, multiple aspects of disease is yet limited by lack of multidisciplinary, interoperable framework to capture, integrate and disseminate knowledge. To improve such limitations, this project aims to build the framework necessary to construct the tools, resources, and cell atlases needed to determine how the relationships between cells can affect the health of an individual. These tissues include: (1) discrete, complex organs (kidney, ureter, bladder, lung, breast, small intestine and colon); (2) distributed organ systems (vasculature); and (3) systems comprising dynamic or motile cell types with distinct microenvironments (lymphatic organs: spleen, thymus, and lymph nodes) [citation]. Specifically in this research, the cell data has five classes: kidney, prostate, large intestine, spleen and lung.

However, there are problems arising with these cell images. First of all, as (author) stated [citation], the provided dataset from Human Protein Atlas (HPA) and Human BioMolecular Atlas Program (HuBMAP) is different in terms of many aspects, such as image resolution, color render, cell size and organ types. Furthermore, such multiclass of images causes domain adaptation problem. For instance, the segmentation of kidney cell alone is poorly adapted to other cells. Lastly, minute amount of dataset may not give meaningful performance estimation [citation]. As referred to figure 1, the total value of the dataset is 351, and the smallest number of organ kind is lung. This resulted that lung has the lowest accuracy of prediction.

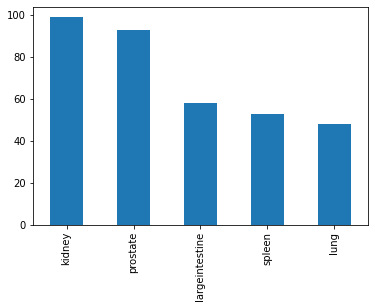
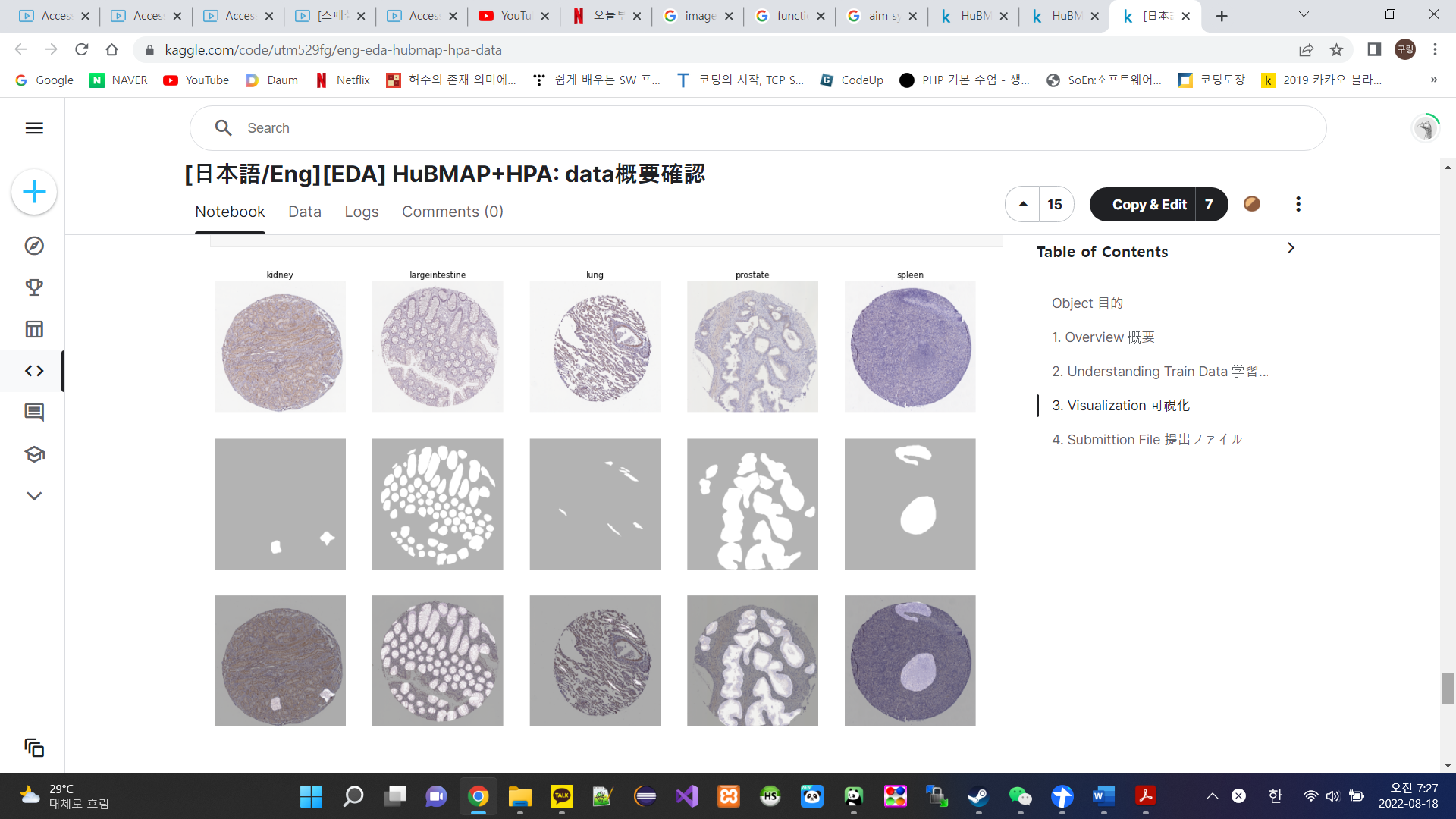


Fig. 1, The Amount of Each Organ Type

Medical artificial intelligence (AI) focuses on annotating organs on cell image dataset to perform clinical diagnoses and to suggest suitable treatments. In SURF 2022, our team attempted new methods of data preprocessing and models. Our study has shown a progress with such implementation, proposed model obtained an accuracy of 0.83 and a dice coefficient of 0.76.

Multiple aspects of disease analysis is a crucial issue in medical AI. It is addressed in many Kaggle Competitions, and this project targeted the Human BioMolecular Atlas Program (HuBMAP) that aims to create human reference atlas at a molecular level. The project worked based on Functional Tissue Units (FTUs) segmentation method. As it can be notified from the figure, it refers to the process of characterizing the number of FTUs per unit area in dependence on location in the human body []. However, it only performed well on certain datasets and organ samples []. Therefore, to improve such limitation, this project aims to push the boundaries by building algorithms that generalize across different organs and are robust across different dataset differences.

Multiple aspects of disease analysis is a crucial issue in medical AI. This is limited by lack of multidisciplinary, interoperable framework to capture, integrate and disseminate knowledge. In fact, Functional Tissue Units (FTUs) segmentation method is significant for modeling and understanding disease progression, thus widely utilized especially in Kaggle Competitions, where it can be notified in the figure. Specifically, the Human BioMolecular Atlas Program (HuBMAP) + Human Protein Atlas (HPA) is one of the current competition that this project targeted.



The Human BioMolecular Atlas Program (HuBMAP) + Human Protein Atlas (HPA) is one of the current Kaggle Competition that this project aimed for. It targets improvement of the problems in medical AI such as limited multiple aspects of disease. Functional tissue units (FTU) are pathobiological significance and relevant for modeling and understanding disease progression.