Graduate School of Science and Technology Master's Thesis Abstract

Laboratory name (Supervisor)	Computational Systems Biologoy (Shigehiko Kanaya (Professor))		
Student ID	2111318	Submission date	2023 / 1 / 11
Name	RUMMAN MAHFUJUL ISLAM		
Thesis title	Information maximization-based clustering of histopathology images using deep learning method		

Abstract

Pancreatic cancer is one of the most pernicious types of cancers despite being very uncommon. The reason for it being irregular is that the symptoms don't really arise until the cancer is already in the late stages, making it almost impossible to remove by surgery as it has spread to the other parts of the body from pancreas by then. So, the detection of pancreatic cancer in the early stages is indispensable to ensure the surgical removal of tumor. Histopathological diagnosis is the gold standard approach for cancer diagnosis but it is a very tiring and time-consuming process. However, computer aided diagnosis (CAD) can support for meliorating the quality of diagnosis by enhancing the accuracy within a short time. In this work, we enacted a deep learning approach for pancreatic cancer pathology image processing in order to automatically separate cancerous tissue regions from the noncancerous ones. We used pancreatic cancer data of KPC mouse, which is a clinically relevant model of pancreatic ductal adenocarcinoma (PDAC). The pancreatic tumors observed in these mice develop similar histological features as humans. At first, we created random patches from whole-slide images (WSIs). The WSIs are stained with five different staining techniques. The patches have been created in such a way that each of them exhibits the same biological features with different staining. Then, we embedded these patches into an integrated latent space using a convolutional autoencoder. Subsequently, information maximization was performed on the lower level latent space with an aim of segregating different tissue features in non-identical clusters in an unsupervised manner due to the absence of labels in our dataset. Moreover, we leveraged uniform manifold approximation and projection (UMAP), a nonlinear dimension reduction technique to visualize the patterns of embedded features in the upper dimensional latent space as different samples are expected to have different embeddings. Finally, we calculated Dunn index (DI), a cluster validation technique for determining the optimal number of clusters utilizing the upper level latent features to make sure that the patches in an individual cluster is similar to each other and contrasting with unalike clusters.