|  |  |  |  |
| --- | --- | --- | --- |
| **Project Title** | Mass Spectrometry Imaging in Detecting Tumor Heterogeneity | | |
| **Track** | **Mass Spectrometry Imaging - Graduation Project** | | |
| **Supervisor** | Dr. Ahmed Morsy | **Supervisor** | Dr. Walid Abdelmoula |
| **Team Name** | Saviors | | |
| **Team Members** | Ibrahim Elsayed | Donia Abd Elsalam | Renad Taher |
|  | Mariem Ahmed | Mustafa Yehia |  |
| **Problem Summary** | Tumor subpopulations have molecular phenotypes that drive tumor progression and determine disease outcome which is essential for a more personalized therapy. Mass spectrometry imaging has proven its ability to identify diagnostic and prognostic biomarkers. In this research, we seek to determine tumor subpopulations that affect patient outcomes and the statistically associated subpopulations with poor survival and tumor metastasis. Here we introduce spatially mapped t-distributed stochastic neighbor embedding (t-SNE), a nonlinear visualization of the data that can better resolve the biomolecular intratumor heterogeneity. The outcomes will allow us to uncover subpopulations statistically associated with patient survival in primary tumors of gastric cancer and with metastasis in primary tumors of breast cancer. | | |
| **Methodology** | Before doing any work, we need to understand the Mass spectrometry imaging technique and different methods to solve our problem. Then we split our problem to some small problems and start solving it sequentially according to the objective of each small problems. First, we represent the original microscopic image in Gastric and Breast parts to know what they consist of. In gastric data, we found that it consists of 82 proteins in each sample, on the other side breast data consists of 62 proteins in each sample. From that, we get each data contains millions of data (High dimensional data), which make the problem (Curse of dimensionality) in analysis. The first objective we face is to found out the type of tumors in samples, so we apply dimensionality reduction using t-SNE to avoid the previous problem in breast and gastric data. Then, the second objective is to segment the tumors in each sample in discrete space because that we apply the K-means clustering in the result of t-SNE. Using this result to solve the third objective that is the survival analysis of each tumor. Then build the classification model to use in identifying the type of tumors in new sample. | | |
| **Achievements and Skills Gained** | 1. Learn some information about biotechnology techniques Like MSI, MALDI, etc.… 2. How to solve the curse of dimensionality. 3. How to apply machine Learning algorithms in computer vision. 4. How to segment the inter/intratumor heterogeneity in discrete and continuous Space. 5. How to apply the Survival Models in medical data. 6. Learn some algorithms about medical prognosis. 7. How to apply the SAM analysis of medical data. | | |
| **References** | [1] Schober, Patrick, and Thomas R. Vetter. "Survival analysis and interpretation of time-to-event data: the tortoise and the hare." Anesthesia and analgesia 127.3 (2018): 792.  [2] J. Cadima, "Principal component analysis: a review and recent developments," THE ROYAL SOCIETY PUBLISHING, vol. 16, 2016.  [3] L. v. d. Maaten, "Visualizing non-metric similarities in multiple maps," Springerlink, vol. 23, 2011.  [4] Abdelmoula, Walid & Balluff, Benjamin & Englert, Sonja & Dijkstra, Jouke & Reinders, Marcel & Walch, Axel & Mcdonnell, Liam & Lelieveldt, Boudewijn. (2016). Data-driven identification of prognostic tumor subpopulations using spatially mapped t-SNE of mass spectrometry imaging data. Proceedings of the National Academy of Sciences. 113. 10.1073/pnas.1510227113. | | |
| **Project Title** | Mass Spectrometry Imaging in Detecting Tumor Heterogeneity | | |
| **Main Results** | Chart, bar chart  Description automatically generated | | |
| **Discussion and Conclusion** | Intratumor heterogeneity is a key factor in tumor progression, affecting patient outcomes and treatment. Tumor subpopulations can be histologically indistinguishable but still have molecular phenotypes that drive tumor progression and determine disease outcome. The identification of these clinically relevant tumor subpopulations is of utmost importance for understanding cancer development and the management of cancer patients. | | |
| **Future Work and Suggestions** | We shall now establish a technique to distinguish between biomarkers we got from SAM analysis, and to do that we have first to classify them using any classification technique (e.g., KNN) and then we will have to differ our data (As explained previously, our data is relatively small, but significant) so, we will use a technique known as Leave-One-Patient-Out (LOPO) and apply the same procedures on data extracted. That will give us a strong evidence for our result and solve the problem of small data. | | |
| **Group Photo** |  | | |