**Final Project Report**

**Denna Saji, Prem Almeida, Yogesh Chaudhari, Uma Dixit, Sheng Li**

CSCI 8950 Machine Learning

Department of Computer Science

University of Georgia

Athens, GA 30605

**Abstract**

Prognosis and prediction are extremely crucial steps in detecting breast cancer and can save a large number of patients from receiving any further treatments, time, or medical costs. In this study, we attempt to reassess a multimodal deep neural network method, improve the accuracy of the predication model (prognosis and diagnosis), and the inclusion of an autoencoder.

**1 Background**

* 1. **Literature Review**

With breast cancer being one of the most highly aggressive cancers, especially in females, it is crucial that a method to limit its spread is found. The American Cancer Society found case estimates to be near 250,000 of invasive breast cancer and about 40,000 death approximations in 2017. This disease is mainly quantified by varied molecular feature, clinical behavior, morphological appearance and disparate response to therapy. Thus, attempting to increase the accuracy of prognostication is critical in helping patients learn more about their life expectancy as well as helping clinicians make more informed decisions for the benefit of the patient.

With the recent developments in gene expression patterns analysis and microarrays there have been many contributions to the overall knowledgebase of molecular signatures of breast cancer that are based purely on gene expression patterns. One of the most influential studies in this field shows breast cancer prognosis through gene expression profiles in which they identify 70 gene signatures from 98 primary breast cancer patients by clustering the gene expression profile data and correlating them with the values they calculated from the prognostics.

The dataset which was used in the original paper consists of 25,000 genes per patient however for the purpose of this study we have chosen 15,000 genes

* PCA
* PGM

**2 Methods and Materials**

**2.1 Materials**

**-** Metabric now has 2500 but we are using 1,500

**2.2 Feature Selection**

- MRMR (still following the original paper, table 2)

- Variational autoencoder

- Denoising autoencoder

- Deep multimodal autoencoder

- CNA 900, clinical data is 19, gene expression 1,227

**2.3 Prediction Models for Single and Multidimensional Dataset**

- Original paper dataset and find the accuracy

- Adding a multimodal with dataset and finding the accuracy

- changed the parameters and added RNA

**2.4 Experimental Design**

- Make a block diagram (add autoencoder block separately, MRMR)

**2.5 Other**

**3 Results**

**3.1 Comparison**

**3.2 Validation**

**4 Discussion and Conclusion**

**References**

[1] Extracting High-Quality Features From Biomedical

Datasets Using Multimodal Autoencoders (Dmitry Kazhdan)

[2] A Multimodal Deep Neural Network for Human

Breast Cancer Prognosis Prediction by

Integrating Multi-Dimensional Data (Dongdong Sun, Minghui Wang, and Ao Li)