SCHOOL OF COMPUTING
Faculty of Engineering

SECB3203-01

Programming For Bioinformatics

23/24 - 1

Prediction of Ovarian Cancer Prognosis

Using Deep Neural Network (DNN)

Project Proposal

Submitted to:

Dr. Nies Hui Wen

NAME	MATRIC NUMBER
AISYAH BINTI MOHD NADZRI	A21EC0011
THUVAARITHA A/P SIVARAJAH	A21EC0137

Table Of Content

1.0 Introduction

Ovarian cancer, a formidable adversary in the realm of women's health, is known for its silent invasion and late-stage diagnosis, often making it challenging to treat effectively. This cancer is called forgotten cancer and is sometimes misdiagnosed (Tan JH et al,. 2020). It has been a constant concern for both patients and medical professionals. Despite advances in therapy, ovarian cancer remains the most deadly of the gynecological cancers where less than 30% of women with advanced stage disease survive long term (Badgwell D & Bast RC Jr., 2007). Improving the prognosis of ovarian cancer patients has become an urgent and essential mission in the field of oncology,

Predicting the prognosis of ovarian cancer involves not just estimating survival but also understanding the effectiveness of treatments and the long-term quality of life for survivors. Recent decades have seen remarkable advancements in medical research and technology, offering a glimmer of hope in the fight against this relentless disease. Innovative predictive models, fueled by diverse biomedical and clinical data, have emerged as a promising tool to transform ovarian cancer prognosis.

Early detection and accurate diagnosis are vital aspects of improving the prognosis for ovarian cancer patients. In the context of ovarian cancer prognosis, an accurate diagnosis of the disease is a critical first step. Once diagnosed, it becomes possible to make informed predictions about the course of the disease and the patient's likely outcome. In this study, a deep learning approach is employed to demonstrate its capability and effectiveness in the prediction of ovarian cancer prognosis. Accurate predictions have the potential to guide treatment strategies, enhance survival rates, and ultimately contribute to better outcomes for ovarian cancer patients.

2.0 Problem Background

Cancer is a disease in which some of the body's cells grow uncontrollably and spread to other parts of the body. The fundamental abnormality resulting in the development of cancer is the continual unregulated proliferation of cancer cells. Likewise, ovarian cancer is a type of cancer that begins in the ovaries, considering that this cancer is considered as one of the most deadly causes of death among women. Ovarian cancer is a highly malignant gynecological tumor that has a poor prognosis, reducing the chances of successful treatments and survival as this disease is often diagnosed at an advanced stage. Forasmuch as, due to inadequate screening instruments and difficulties in timely detection, 80% of patients with ovarian cancer are diagnosed at an advanced stage and 50% - 70% patients will experience recurrence within two years of finishing therapy with a poor 5-year survival rate of 30%. Currently, the main treatments for ovarian cancer are platinum-based chemotherapies and surgery. The 5 - year survival rate has been slowly improving despite recent advancements in therapy. Considering the limitations of ovarian cancer treatments, reliable innovative prognostic models are desperately needed to increase the viability of targeted therapies since the limits of ovarian cancer treatments need the development of new therapeutic targets in order to enhance the clinical outcome of ovarian cancer.

3.0 Problem Statement

Predicting the prognosis of ovarian cancer accurately, typically involves analyzing large scale biological data and leveraging various computational methods. The insufficiency of accurate prognostic information may lead to less optimal treatment decisions and outcomes for ovarian cancer patients. This problem is significant because accurate prognosis prediction can help healthcare providers make informed treatment decisions and potentially improve patient outcomes.

4.0 Aim & Objectives

In accordance with the accomplishment of this project, these are the aims and objectives that we wish to achieve.

1. To gain overall understanding of ovarian cancer disease through literature review and provide a highly accurate prediction model.

2. To implement and develop a prediction model by applying deep learning for early prognosis identification of ovarian cancer, allowing for timely and appropriate planning for treatment.

3. To evaluate the effectiveness of the proposed prediction model in terms of accuracy and reliability.

Despite that, the aim and objectives for predicting ovarian cancer prognosis aim to improve patient outcomes, enhance survival rates and contribute to ongoing research and treatment advancements in the field of technology.

5.0 Scopes

5.1 Domain of the data

On top of overall data we came across, we have decided to utilize a dataset retrieved from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE12418 with 54 samples of ovarian cancer disease. The background of the dataset chosen is as below. The raw data dataset contains 30, 000 rows and 55 columns. However it is yet to undergo data preprocessing nor cleaning.

- The time frame of the data retrieved was in the range of August 2008 until March 2012 which ideally is 4 years.
- After a few considerations from the datasets we have chosen and observed, we decided to use this dataset with Gene Expression Profiling by Array data type.
- From the datasets, there are several suitable attributes to put to use such as an instance of mRNA expression, DNA methylation, miRNA expression, CNA of genes/miRNA.
- Area of study covered in the dataset is comparing fresh frozen tumors and tumors from deceased patients of ovarian cancer disease to find genes where the expression differed between the two groups.

5.2 Techniques to be used

1. Deep Neural Networks (DNN)

   The primary technique that would be used is DNN as it is used for feature extraction and prediction. It plays a central role in our research,

2. Gene Expression Profiling

   The dataset we chose involves gene expression profiling. This technique is essential for analyzing how genes are expressed and can be associated with cancer prognosis.

3. Biomedical Data Analysis

   This project involves the analysis of biomedical and clinical data. This likely includes multi omics data.

4. Machine Learning for Predictive Modeling

   While DNN is highlighted, this project is likely to include machine learning techniques for building predictive models. These could include various algorithms such as decision trees, Support Vector Machines (SVM) or Random Forest (RF)

5. Evaluation of Predictive Model

   This project aims to evaluate the effectiveness of the predictive model in terms of accuracy and reliability. This evaluation will involve techniques for model validation and performance assessment.

6. Literature Review

   This technique is used to gather existing knowledge on our topic.

7. Data Retrieval and Preprocessing

In order to clean, format and prepare the dataset for analysis, we will be using the Principal Component Analysis (PCA). It is a dimensionality reduction technique that helps in transforming data into a new system to capture the most significant variance in the data while reducing the dimensionality.

5.3 Methodology

Notwithstanding what has just been said, in order to analyze the dataset we have chosen, we should break it down into several methods of analyzation in which as below.

- Approach: Put Deep Neural Networks (DNN) to use as DNN models have the potential to provide new insights in the study of cognitive processes in which it imitates human decision making as it is high in capacity and data -driven design.

- Testing/experimental tools/measurement: Principal Component Analysis (PCA) for data preprocessing. PCA can be applied to reduce the dimensionality of metadata while retaining most of the relevant information. PCA can be a valuable method for data preprocessing.

- Test data: Utilize Scikit-Learn Python for testing predictive models in Python. The processes involved are data splitting, model training, model evaluation, etc.

5.4 Limitation of the research

1. Late Diagnosis

    Ovarian cancer is often diagnosed at an advanced stage due to the lack of specific early symptoms and effective screening methods. This late diagnosis significantly reduces the chances of successful treatment.

2. Heterogeneity

    Ovarian cancer is not a single disease but a group of distinct subtypes each with different characteristics and responses to treatment. This complicates research efforts to find a one-size-fits-all solution.

3. Limited biomarkers

    The lack of highly specific and sensitive biomarkers for early detection and monitoring of ovarian cancer is a significant challenge as the existing ones lack the accuracy needed for screening.

4. Data Source

    The data source mentioned is from a specific time frame (August 2008 to March 2012), and ovarian cancer research and treatments may have evolved since then. Using older data might not reflect the current state of the disease.

6.0 Conclusion

To sum up, an essential component of enhancing patient outcomes and customizing treatment regimens is predicting the prognosis of ovarian cancer. In this work, researchers used Deep Neural Networks (DNN) to create a predictive model based on a dataset that included 54 samples from patients with ovarian cancer which it retrieved from Gene Expression Omnibus (GEO). DNNs are a subclass of deep learning methods that are especially useful for managing high-dimensional, complicated datasets, such as those used in medical research.

In addition to that, to guarantee its quality, we will be starting by preprocessing and cleaning the dataset. We then used deep neural networks (DNNs) to mine the data for complex patterns and associations that would help us develop a prediction model for the prognosis of ovarian cancer. The model was able to evaluate and forecast the probability of distinct prognostic outcomes by feeding the DNNs with a variety of clinical and molecular data. This allowed the model to offer important insights into the course of the disease and support the development of individualized treatment plans.

Last but foremost, by utilizing deep learning techniques, the findings of this study present a viable option for improving the care of patients with ovarian cancer. This study should assist in advancing the continuing endeavor to enhance early detection and treatment approaches for ovarian cancer by utilizing the predictive powers of DNNs, which will eventually improve patient outcomes and raise survival rates.

7.0 References

1. Sorayaie Azar, A., Babaei Rikan, S., Naemi, A., Bagherzadeh Mohasefi, J., Pirnejad, H., Bagherzadeh Mohasefi, M., & Wiil, U. K. (2022). Application of machine learning techniques for predicting survival in ovarian cancer. *BMC medical informatics and decision making*, *22*(1), 345. https://doi.org/10.1186/s12911-022-02087-y

2. Badgwell, D., & Bast, R. C., Jr (2007). Early detection of ovarian cancer. *Disease markers*, *23*(5-6), 397–410. https://doi.org/10.1155/2007/309382

3. Jingyang Zhou, Weiwei Cao, Lan Wang, Zezheng Pan & Ying Fu (2022) Application of artificial intelligence in the diagnosis and prognostic prediction of ovarian cancer. Retrieved from https://www.sciencedirect.com/science/article/pii/S0010482522004000

4. Ying Ye, Qinjin Dai & Hongbo Qi (2021). A novel defined pyroptosis-related gene signature for predicting the prognosis of ovarian cancer.
   Retrieved from https://www.nature.com/articles/s41420-021-00451-x

5. Mingyang Bao, Lihua Zhang & Yueqing Hu (2020). Novel gene signatures for prognosis prediction in ovarian cancer.
   Retrieved from https://onlinelibrary.wiley.com/doi/full/10.1111/jcmm.15601

6. E Sun Paik, Jeong-Won Lee, Jeong-Yeol Park & Ju-Hyun Kim (2019) Prediction of survival outcomes in patients with epithelial ovarian cancer using machine learning methods. Retrieved from https://synapse.koreamed.org/articles/1125153