SMART BRIDGE INTERNSHIP

2019

Project Name:

Survival Analysis Cancer Prediction

Team Members:

V.Manasa

A.Bhavana

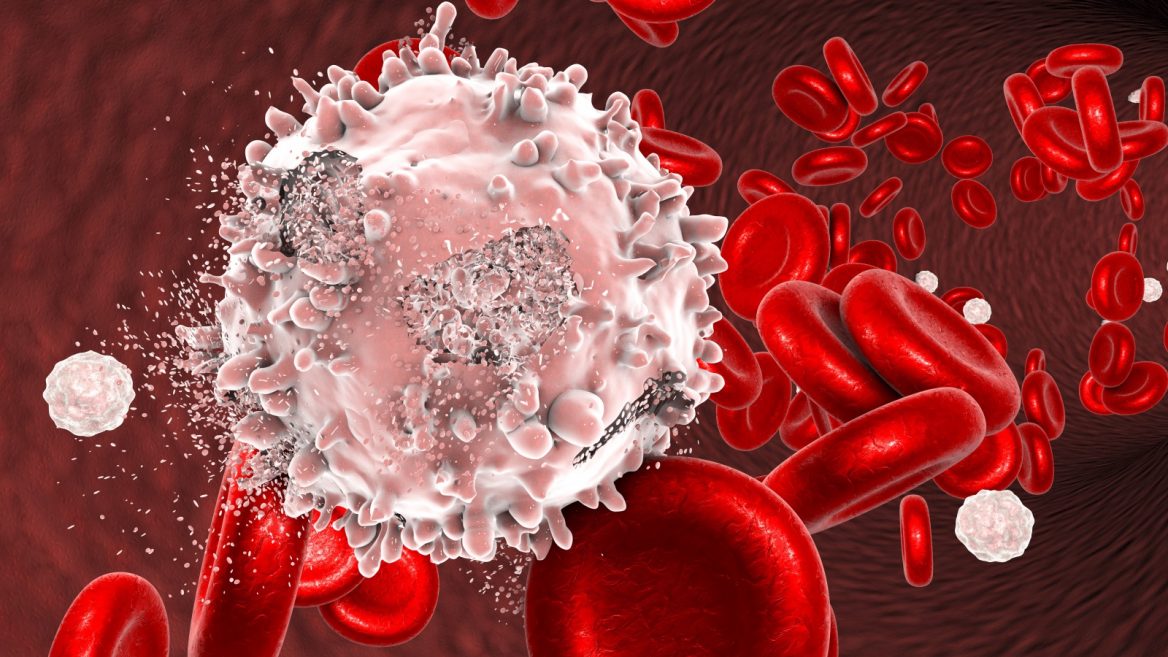
K.Sai Rithika

S.Venkata Bhavani Ramya

* **Introduction**
* **Objectives of Research**
* **Problem Statement**
* **Review of Literature**
* **Data Collection**
* **Methodology**
* **Figures and Tables**
* **Statistical Techniques and Data Visualization**
* **Data modeling using supervised ML techniques**
* **Conclusion**

1.Survival Analysis Cancer Prediction

**1.1 Introduction:**



Over the past decades, a continuous evolution related to cancer re-search has been performed. Scientists applied different methods such as screening in early stage, in order to find types of cancer before they cause symptoms. Moreover, they have developed new strategies for the early prediction of cancer treatment outcome. With the advent of new technologies in the field of medicine, large amounts of cancer data have been collected and are available to the medical research community. However, the accurate prediction of a disease outcome is one of the most interesting and challenging tasks for physicians. As a result, ML methods have become a popular tool for medical researchers .These techniques can discover and identify patterns and relationship between them, from complex datasets, while they are able to effectively predict future outcomes of a cancer type .Given the significance of personalized medicine and the growing trend on the application of ML techniques, we here present a review of studies that make use of these methods regarding the cancer prediction and prognosis. In these study prognostic and predictive features are considered which may be independent of a certain treatment or are integrated in order to guide therapy for cancer patients, respectively. In addition, we discuss the types of ML methods being used, the types of data they integrate, the overall performance of each proposed scheme while we also discuss their pros and cons .An obvious trend in the proposed works includes the integration of mixed data, such as clinical and genomic. It is that the application of ML methods could improve the accuracy of cancer susceptibility, recurrence and survival prediction. Based on the accuracy of cancer prediction outcome has significantly improved by15%–20% the last years, with the application of ML techniques. These methods suffer from low sensitivity regarding their use in screening at early stages and their difficulty to discriminate benign from malignant tumors. Various aspects regarding the prediction of cancer outcome based on gene expression signatures are discussed in the present work only studies that employed ML techniques for modeling cancer diagnosis and prognosis are presented.

**1.2. Objectives of Research**

1. Provide an integrated training program in cancer biology and treatment.
2. Provide training in development of partnerships between basic and clinical researchers.
3. Provide training in development of industrial partnerships.
4. Provide stipends for MSc and PhD trainees and postdoctoral Fellows in basic biological and physical sciences.
5. Create "meet-and-greet" programs for research sponsors, researchers-in-training, and senior researchers in basic and clinical sciences to speed the transfer of information from basic laboratories to the private sector, the clinic, and the community.

6.Molecular and Physical

7.Cancer Imaging

8.Basic Cancer Biology

9.Basic Cancer Physics

10.Development of Novel Therapies and Diagnostic Tools

**1.3 Problem Statement**

1. Falls are especially important when pain, fatigue and

other symptoms of cancer treatment may enhance the

risk in older cancer patients (Holley, 2002).

2. Approximately 50% of adults with advanced cancer will

fall while hospitalized (Stone et al 2012).

3.17% of cancer patients fall within 6 months of a cancer

diagnosis (Puts et al 2012).

Some of the purpose statements are given below

The purpose of this research was to:

understand the extent to which falls occurred,

how falls related to depression, age, functional status and

cognition

to develop a model for predicting falls in older cancer

patients

4.narratives associated with falls.

**2.Review of Literature**

The primary PubMed search produced 900 citations, of which 263 met the inclusion criteria. The first published study that met our inclusion criteria was published in 1984. The secondary PubMed search for the key words “cancer survivorship” in the title or abstract resulted in 391 articles, 72 of which met the inclusion criteria and were not duplicated by the initial search. The Medline search resulted in 719 articles. Of these, 170 articles met the inclusion criteria and were not duplicated by the initial searches. An additional 169 studies were identified in reference lists from articles that were identified in the searches conducted. The final result of the literature review was 674 articles.

The number of survivorship studies overall has increased since 1984, with an observable increase around 1996, the year the

A close up of a pencil

Description automatically generated

Office of Cancer Survivorship (OCS) was created. Of the 674 studies that met the inclusion criteria, 618 or 91.7% were quantitative. As depicted in the use of descriptive and case control studies has been relatively unchanged over time, while the use of descriptive and case control studies has been relatively unchanged over time, while the use of qualitative studies has fluctuated. Since 1996, randomized controlled trials and cross sectional designs were used with increasing frequency, especially in the last decade.

Artificial Intelligence includes three areas of specialization and it contains machine learning, natural language processing ,speech,

Export systems planning, scheduling and optimization and robotics and vision.

Machine Learning contains two types. They are

1.Deep Learning

2.Predictive Analytics

Natural language Processing contains three types. They are

1.Translation

2.Classification and Clustering

3.Information extraction

Speech has two types. They are

1.Speech to text

2.Text to speech

Vision contains two types. They are

1.Image Recognition

2.Machine Vision

Artificial Intelligence:

Any technique which enables computers to mimic human behavior.

Machine Learning:

Subset of AI technique which use statistical methods to

enable machines to improve the experience.

Deep Learning:

Subset of ML technique which make the computation of multi linear neural networks feasible.

Python:

Python is a high level programming language:

Interpreted: python is processed at runtime by the interpreter

Interactive: you can use a python prompt and interact with the interpreter directly to write your programs

Object-oriented: python supports object-oriented technique of programming

Beginner’s Language: python is a great language for the beginner-level programmers and supports the development of a wide range of applications.

Python Basics:

String: To identify the range of word using array

List: A list in python is an ordered group of items or elements and these list elements don’t have to be of the same type

Python lists are “mutable” objects that can change their Values

Tuples: python tuples are “Immutable” that cannot be

Changed.

Dictionaries: python dictionaries are kind of hash table which type consist of key-values pairs of unordered elements

Sets: sets are similar to lists but duplicates are not allowed.

Data Collection

The NHS collects a wide range of data and information about cancer patients and their treatment. We work with all our partners, including the health service, specialist clinical teams and cancer registries, to understand what information is required for the delivery of cancer services for effective patient care within the NHS, to improve data quality, timeliness and outcomes.

[Cancer](https://api.seer.cancer.gov/rest/glossary/latest/id/546cedfae4b0d965832a8c22) [data](https://api.seer.cancer.gov/rest/glossary/latest/id/5545176be4b0426fced41c0f) collection begins by identifying people with cancer who have been diagnosed or received medical care in hospitals, [outpatient](https://api.seer.cancer.gov/rest/glossary/latest/id/554419a5e4b0426fced3ed6d) clinics, radiology departments, doctors' offices, laboratories, surgical centers, or from other providers who diagnose or treat cancer patients.

In this we are taking the Haberman dataset .It contains Age, Operation years and Axillary Nodes as input parameters and Survival Status as output parameters.

Input parameters are independent variables and output parameters are dependent variables as output parameters depending on the input parameters and we are getting two outputs in this it may 1 or 2.

A screenshot of a social media post

Description automatically generated

Methodology

Methodologies should present a new experimental or computational method, test or procedure. The method described may either be completely new or may offer a better version of an existing method. The article must describe a demonstrable advance on what is currently available. The method needs to have been well tested and ideally, but not necessarily, used in a way that proves its value.

*Cancer & Metabolism* strongly encourages that all datasets on which the conclusions of the paper rely should be available to readers. We encourage authors to ensure that their datasets are either deposited in publicly available repositories (where available and appropriate) or presented in the main manuscript or additional supporting files whenever possible. Please see Springer Nature’s[information on recommended repositories](http://www.springernature.com/gp/group/data-policy/repositories). Where a widely established research community expectation for data archiving in public repositories exists, submission to a community-endorsed, public repository is mandatory. A list of data where deposition is required, with the appropriate repositories, can be found on the[Editorial Policies Page](http://www.biomedcentral.com/submissions/editorial-policies#availability+of+data+and+materials).

4.1.1. Figures and Tables

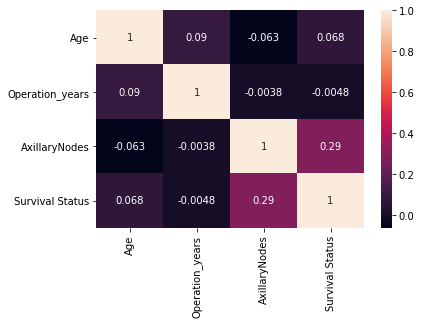
In this we are taking the Logistic Regression of Classification.

In this we will get the fixed output and It also uses the sigmoid function.

In the Logistic Regression we are using Standard Scalar and accuracy score.

A screenshot of a cell phone

Description automatically generated



4.2 Statistical techniques and data visualization

Data visualization needs extraordinary techniques to efficiently process large volume of data within limited run times. Data visualization techniques involve a number of disciplines, including statistics, data mining, machine learning, neural networks, social network analysis, signal processing, pattern recognition, optimization methods and visualization approaches. There are many specific techniques in these disciplines, and they overlap with each other hourly. Optimization Methods have been applied to solve quantitative problems in a lot of fields, such as physics, biology, engineering, and economics. Several computational strategies for addressing global optimization problems are discussed, such as simulated annealing, adaptive simulated annealing, quantum annealing, as well as genetic algorithm which naturally lend itself to parallelism and therefore can be highly efficient. Stochastic optimization, including genetic programming, evolutionary programming, and particle swarm optimization are useful and specific optimization techniques inspired by the process of nature.

4.3. Data modeling using supervised ML techniques

In ML there are of three types. They are

Supervised Learning

Unsupervised Learning

Reinforcement Learning

Now we consider supervised ML techniques

It contains input and output parameters.

In this we have of two types

1.Regression

2.Classification

To build a machine learning model we use these steps as followed.

-Data gathering

-Data Cleaning

-Feature Extraction

-Model Training

-Prediction

6.Conclusion

In this review we have attempted to explain, compare and assess the performance of different machine learning that are being applied to cancer prediction and prognosis. Specifically we identified a number of trends with respect to the types of machine learning methods being used, the types of training data being integrated, the kinds of endpoint predictions being made, the types of cancers being studied and the overall performance of these methods in predicting cancer susceptibility or outcomes.