# 7CS997 Independent Studies Project Plan

## Student/supervisor details

**Submission deadline for signed and completed project plan: 19/01/2024 12:00 Midday**

Project Title: **XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX**

Name: **XXXXXXXXXXXXXX**

Student Id: XXXXXXXXXX

MSc: **XXXXXXXXXX**

Supervisor: **Dr Oluwarotimi W. Samuel**

## Brief description

Cancer remains a formidable global health challenge, affecting millions of lives each year [1]. As researchers and healthcare professionals tirelessly seek innovative approaches to understand and combat this disease, the integration of advanced technologies, particularly enhanced machine learning, has emerged as a promising frontier [2]. Early detection is pivotal in improving cancer outcomes. Conventional diagnostic methods have long been the cornerstone of early detection[3][4]. However, the advent of data-driven technologies has ushered in a new era of precision and efficiency in cancer diagnostics [5]. The rich landscape of data, particularly in the form of comprehensive datasets like the one in this project, has become instrumental in unravelling the intricacies of cancer. These datasets, featuring a myriad of clinical and molecular attributes, empower researchers to discern patterns and biomarkers that might elude traditional analyses.

The primary aim of this project is to advance cancer prediction and prognosis through the application of enhanced machine learning (ML) techniques. By harnessing the power of data-driven methodologies, this project seeks to develop a robust predictive model that can accurately classify instances of cancer, providing valuable insights for prognosis. The analysis will delve into the intricate relationships between various features and the diagnosed outcomes. The project will further explore patterns, correlations, and potential biomarkers that contribute to accurate cancer prediction. Advanced statistical tests will be employed to validate findings and ensure the reliability of the developed models.

Ethical considerations will be prioritized due to the sensitivity of health-related data in the project [6]. Aligned with the broader mission of leveraging technological advancements, the project aims to enhance medical decision-making for more accurate and timely interventions in cancer-related healthcare scenarios. The findings hold promise for improving predictive capabilities and enabling more informed clinical decisions in cancer care through cutting-edge ML methodologies.

**References**

*[1] Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. Computational and structural biotechnology journal, 13, 8-17.*

*[2] Rasool, S., Husnain, A., Saeed, A., Gill, A. Y., & Hussain, H. K. (2023). Harnessing Predictive Power: Exploring the Crucial Role of Machine Learning in Early Disease Detection. JURIHUM: Jurnal Inovasi dan Humaniora, 1(2), 302-315.*

*[3] Shaikh, T. A., & Ali, R. (2019). Applying machine learning algorithms for early diagnosis and prediction of breast cancer risk. In Proceedings of 2nd International Conference on Communication, Computing and Networking: ICCCN 2018, NITTTR Chandigarh, India (pp. 589-598). Springer Singapore.*

*[4] Asri, H., Mousannif, H., Al Moatassime, H., & Noel, T. (2016). Using machine learning algorithms for breast cancer risk prediction and diagnosis. Procedia Computer Science, 83, 1064-1069.*

*[5] Bekisz, S., & Geris, L. (2020). Cancer modeling: From mechanistic to data-driven approaches, and from fundamental insights to clinical applications. Journal of Computational Science, 46, 101198.*

*[6] Vollmer, S., Mateen, B. A., Bohner, G., Király, F. J., Ghani, R., Jonsson, P., ... & Hemingway, H. (2020). Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. bmj, 368.*

## Project aims and objectives.

*List as clearly as possible up to 5 aims/objectives of your project.*

1. To highlight the importance of addressing cancer, a disease responsible for a significant number of deaths globally.

2. To explore the effectiveness of classification and data mining methods, particularly in the medical field, for diagnosing and analysing cancer.

3. To conduct a comparative analysis of machine learning algorithms and assessing the efficiency and effectiveness of each algorithm in cancer prediction and prognosis based on key evaluation metrics.

4. To identify and determine the algorithm that demonstrates the highest accuracy, lowest error rate, and overall efficiency for cancer risk prediction and diagnosis.

5. To contribute valuable insights to the field of cancer diagnosis by providing a comparative analysis of machine learning algorithms, aiding in the identification of the most effective approach.

## Plan of work

**Literature Review**

A comprehensive review of the existing literature will be undertaken to establish a robust foundation for the project, aiming to comprehend the historical context and identify any gaps or advancements in the field. This involves leveraging academic databases and journals to scrutinize research studies focusing on machine learning applications in cancer diagnosis. Special emphasis will be placed on exploring methodologies employed, addressing challenges encountered, and identifying innovative approaches within the domain.

**Methodology**

1. **Data Collection and Preparation**

Acquiring and preparing Cancer datasets is vital for ensuring data quality in machine learning tasks. Initial stages involve advanced preprocessing techniques, addressing missing values, normalizing features, and managing outliers. This aligns with a commitment to meticulous and personalized data management to extract meaningful insights in cancer risk prediction and diagnosis.

**(b)Selection of Machine Learning Algorithms**

Various machine learning algorithms, including Random Forest, Naive Bayes, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Decision Tree, will be chosen for a comprehensive evaluation in cancer risk prediction. The selection is based on relevance, diversity, and established usage in medical data classification. This decision considers individual algorithm strengths and their collective diversity for addressing the complexities of cancer risk prediction. Acknowledging the project's dynamic nature, openness to exploring additional algorithms is maintained for a refined analysis.

**Algorithm Implementation and Performance Evaluation Metrics**

The selected machine learning algorithms will be implemented using a suitable programming language and libraries to enable precise cancer prediction. Python will be employed as the programming language.

The performance of each algorithm will be rigorously assessed, utilizing a variety of metrics to offer a comprehensive insight into their strengths and limitations. Standard evaluation metrics such as accuracy, precision, sensitivity, and specificity will be applied.

**Analysis and Comparison**

The Analysis and Comparison phase involves evaluating machine learning algorithms' performance in cancer prediction, considering metrics like accuracy and sensitivity. It aims to identify the most effective algorithm by scrutinizing their strengths, weaknesses, and adaptability to diverse cancer data. The documentation will include visualizations, code snippets, and a concise narrative of challenges and successes for transparency and replicability.

**Documentation and Reporting**

The entire project will be thoroughly documented to guarantee transparency and replicability. A comprehensive report will be compiled, covering literature review discoveries, procedures for dataset preparation, implementations of algorithms, details of experimental setups, and in-depth analyses of results. The documentation will encompass visualizations, snippets of code, and any encountered challenges throughout the project.

**Future Recommendations**

The project aims to provide insights for future research and improvements in cancer risk prediction through machine learning, proposing enhancements and exploring collaboration opportunities. It strives to offer a meticulously documented examination of machine learning algorithms in cancer risk prediction and diagnosis.

## Gantt Chart

*In this section, include a Gantt Chart that shows the major tasks, milestones, and deliverables for your project. This is what you will use to measure your progress, so do make sure it is an accurate reflection of your plan of work.*

A screenshot of a computer

Description automatically generated

## Meeting plan

*In the table below, the project week number and dates correspond to the week numbers that were presented in the introductory lecture that are designated meeting weeks. You should fill in the date, time, and venue for the meeting schedule that you have agreed with your supervisor as part of planning your project.*

|  |  |  |  |
| --- | --- | --- | --- |
| Project week | Meeting date | Meeting time | Venue |
| 1: 05/02/2024 |  |  |  |
| 3: 19/02/2024 |  |  |  |
| 5: 04/03/2024 |  |  |  |
| 7: 18/03/2024 |  |  |  |
| 9: 15/04/2024 |  |  |  |
| 11: 29/04/2024 |  |  |  |

## Ethical approval

I confirm that I have discussed and agreed the project plan with my supervisor and have completed the ethical approval process. The PDF confirming ethical approval has been completed and uploaded to the relevant submission point on course resources.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Student Name Supervisor Name

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Student Signature Supervisor Signature

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Date Date