AI/machine learning to predict the personal needs for people living with cancers

October 2023

By

Al-mizaan jamal

Student number 202137942

Word count: 2858

Introduction

An important area of research and development is the use of AI and machine learning to predict the individual needs of cancer patients. It is a very important area of research which aims to improve patients outcomes and enhance individual patient care as cancer has many different subtypes and is a complicated, heterogeneous disease, it is crucial to customise treatments and supportive care. By using machine learning models, large amounts of data which include patient medical history, lifestyle factors, treatment responses and genomic profiles can be much more efficiently analysed and processed. Using this data, patient specific needs such as treatment efficacy, the effect on their quality of life and any possible adverse side effects. Through constant training, these machine learning models can help healthcare providers in optimising alternative therapies, developing personalised treatment plans and recommending reassuring measures that take into account the psychological, physical and social difficulties faced by cancer patients on a day to day basis. Such a personalised approach would not only improve the standard of care but also reduce time wasted on unnecessary procedures and their side effects.

Objectives

The objective of this project is to use AI and machine learning to develop models using datasets that can anticipate and predict the individual needs of people living different types of cancer. Apart from developing these useful models, this project will not address the medical needs but also the psychological, emotional and social requirements of these patients.

Primary SMART objectives

Specific

- To develop AI/ machine learning models using extensive analysis of diverse datasets which include medical records, genomic information, treatment history, patient

surveys in order to predict the individual medical, emotional, social and psychological needs of people living different types of cancer.

Measurable

- To try and achieve a prediction of a minimum of 85% of the identified cancer patient needs. This will be done through extensive testing and evaluation using other existing models and through metrics such as recall, F1-score and precision.

Achievable

- To Implement a user-friendly interface so that healthcare professionals can access the predictive models and use the generated insights to make informed decisions and offer individualised support to cancer patients
-

Relevant

- To deliver accurate predictions that will assist healthcare professionals in enhancing the support and quality of care by providing specific interventions ultimately improving patient well-being and satisfaction.

Time-bound

- To ensure the development, integration and testing of the predictive models by april, this will ensure that the models are operational and are providing patients care by the end of the timeline.

<u>Secondary SMART Objectives</u>

Specific

- To put together an extensive dataset consisting of up to 1000 different cancer patients records, I will make sure to include a diverse group of cancer types, treatment modalities and demographics which will help enhance the validity and accuracy of my predictive models.

Measurable

- In order to increase accuracy of the models, I will use a cross-validate approach whereby I will compare it to an independent dataset across different subsets of patient data, this will ensure a consistent prediction accuracy for universal patients datasets.

Achievable

- Develop and deploy user training programs for healthcare professionals to effectively utilise predictive models. This includes crucial documentation on reading output data and tutorials for integrating personalised use for each patient.

Relevant

- Through ethical consideration, collaboration with healthcare professionals and surveys so as to get feedback on the effectiveness and usability of the predictive methods.

Timebound

- To carry out regular performance reviews on the predictive models and update the systems every 6 months from the initial deployment, this will consist of adding any enhancements or adjustments deemed necessary after new datasets or feedback are incorporated.

Scope

Apart from the extensive collection and analysis of various datasets which include, medical records, treatment histories, genomic information and patient surveys. The project will also focus on non medical needs such as social and emotional needs. The project will also focus on building a comprehensive system for data collection and storage, development of a predictive model, formulating a model evaluation plan and lastly testing of the system.

It is important to note that the project will involve data collection directly from patients for clinical interventions, this project will also not substitute for the expertise of healthcare professionals. I will also not be collecting any sensitive data and I will be more sure to ensure strict observance of ethical guidelines and data privacy laws.

Data Analysis

This involves the comprehensive gathering and examination of various datasets mentioned above. Apart from this, the data will need to be cleaned to avoid any anomalies or inconsistencies. This Data analysis will help point out valuable insights which will result in an accurate predictive model which can help identify the specific needs of cancer patients.

Predictive Model Development

I will design and develop an AI/ML predictive model that will be able to anticipate and predict the individual requirements of various cancer patients using preset parameters. As mentioned right above, analysed datasets will be used to develop this model as it will yield higher accurate results of the medical and non-medical needs of the patients.

Model Evaluation Plan

This plan will define the metrics, such as accuracy, precision, recall. It will be used to assess the effectiveness of the predictive models in predicting the diverse needs of cancer patients. The output data will also be analysed and compared to already established datasets and actual patient requirements, resulting in the strengths and limitations of the predictive models be known. These insights will be used to refine and improve their accuracy.

Testing and deployment

I will conduct extensive rigorous validation and methods of testing throughout the deployment process. This will ensure that the predictive models are reliable and accurate, as already mentioned, i will use independent established datasets for the validation through a comparison analysis process which will help verify the effectiveness of the model.

Deliverables

- Data analysis and development of Machine learning models for personalised patient needs predictions
- Design basic UI for “patients” and healthcare professionals to accesses these predictive models
- Comprehensive documentation outlining the guidelines, methodology, data sources and the model development process.
- Project analysis report on the effectiveness of the predictive models performance and accuracy.

Constraints

Data Availability - the availability and access to diverse cancer datasets is the most crucial part of this project as it is needed for accurate predictions, the lack of these datasets could hugely impact my project deeply, preventing the analysis and model accuracy.

Ethical Compliance - confidentiality(NHS code of practice) must be followed at all times during this project, these policies could restrict access to some essential data which would reduce the scope of the amount of datasets therefore affecting the predictive model accuracy.

Time constraints - with such a limited project timeline(6 months), which could result in rushed research, development and testing of the predictive models.

Technological limitations - the availability of computing infrastructure and resources needed for these complex machine learning algorithms may be unavailable resulting in inefficient training models which may lead to time constraints.

Assumptions

Assumption of Access to Diverse Data and relevant information : so as to prepare for any missing datasets or information, I will presume already published data and research findings will show an extensive understanding of the current research done on these predictive methods.

Justification: the justification of this assumption is based on published literature that is readily available on open source sites such as(Keggle and Ieee xplore) .

Project rationale and operation

Improved Treatment Planning and Outcomes

The predictions made through AI-driven models will provide guidance to healthcare professionals when giving treatment plans for patients which will consequently improve patient outcomes and the overall quality of care for patients

Empowered Decision-Making for Healthcare Providers

Through predictive insights , healthcare professionals will be better equipped to make more informed decisions and carry out more proactive and focused interventions.This in turn will optimise resource allocation, improving the efficiency of healthcare delivery

Elevated Patient Satisfaction and Well-being

Through targeting the varied needs of cancer patients the project will enhance patient satisfaction and comfort and their overall wellbeing.This will foster a more patient oriented and supportive healthcare environment.

Project operation

The intended framework the project will use is Scrum methodology.The structure of scrum will encourage an incremental and iterative approach to the development process.This will ensure and facilitate the adaptability and flexibility when addressing any possible modifications that arise.
Weekly sprints will be used in conjunction with frequent meetings covering progress updates,issue resolution and planning for the upcoming sprint.Scrum methodology will foster better communication. This collectively will contribute to a more streamlined and effective project process.

To aid task management and tracking the project will utilise Trello boards as a software.Trellos customisable boards and accessible,friendly interface will facilitate prioritising tasks, identify bottlenecks and track/visualise the projects progress effectively. Gnatt charts will be implemented to add a comprehensive outlook of the project timetable, milestones and dependencies.
The use of Gnatt charts will aid in administering resources, scheduling tasks and ensuring the project is on course in accordance with the allocated time frame. The project aims to maintain an organised and effective development process through the implementation of these methodologies and tools, ultimately to aid the successful delivery of the intended outcomes.

Options

Considering these steps, the project will explore the following available tools, techniques, and design parameters:

Techniques

Research and analysis

Carry out detailed research on different techniques, tools and design and set parameters used in similar projects or other AI/ ML projects in healthcare. This research should help me make informed decisions throughout my project.

Requirements Assessment
I will conduct comprehensive research on the project's specific needs, whereby I will consider data scalability, complexity and user interface in order to deploy a predictive model with technical and functional requirements.

Prototyping and Testing
Design prototypes using various types of tools, and carry out extensive testing in order to evaluate, usability, accuracy, performance and general compatibility with project objectives.

Consultation and Expert Opinions
I will seek advice from AI/Ml experts(this could be tutors), domain experts and with the proper ethical consideration medical experts could be consulted in order to make informed decisions throughout my project that have been proven to work in practice.

Risk analysis

Time Management Risk

Due to the predefined 6 month timetable, effective time management is a crucial aspect to the success of the project.Issues can arise that can significantly impact or delay the project schedule, for instance an unforeseen complexity in the model development or prolonged data processing which can jeopardise a timely completion of key milestones.

Therefore, to mitigate this risk a thorough project timeline will be established to outline specific milestones and deliverables for each phase of the project.Frequent progress reviews and proactive identification of any potential bottlenecks will aid mitigation techniques such as resource reallocation and efficient allocation of sprint durations throughout the process of development.

Lack of Relevant Datasets:

The potential lack of comprehensive and varied datasets critical to the project's goals is a significant risk factor. Limited availability of crucial patient data such as lifestyles or specific demographics can hinder the development of precise and thorough prediction models.
To mitigate this risk a wide source of datasets must be acquired through engaging with multiple data providers and research institutes.This will aid the aim of ensuring comprehensive representation of the target population.Implementing robust data augmentation and integrating external data sources are crucial techniques that will overcome the limitations that arise from a lack of relevant data sets, improving the predictive models accuracy.

Resources required

The primary resources required for this project are large diverse datasets, machine learning frameworks such as TensorFlow or PyTorch, and an integrated development environment (IDE) such as Jupyter Notebook. My project will require these crucial resources for data analysis, experimentation and model development. Any delay in the availability of these machine learning frameworks and datasets could significantly hinder the project's progress in the development and testing of the system. I will solve this problem by creating contingency plans, which will include the use of alternative datasets, dummy data and machine learning frameworks.

Tools:

Data analysis: part i will utilise the following tools such as Python with libraries like Pandas, NumPy, and SciPy, R, or MATLAB for comprehensive data analysis and preprocessing.

Machine Learning Frameworks: such as TensorFlow, PyTorch, or scikit-learn for building and training AI-driven predictive models.

User Interface Design Tools: HTML, CSS, and for front-end development, along with different frameworks for creating interactive and user-friendly interfaces.

Project Management Tools: Jira, trello or Microsoft Project for efficient project planning, task tracking.

Version Control Systems: Git for effective version control and collaborative development.
Documentation Tools: Microsoft Word, Docs

Project methodology and outcomes

Initial project plan

<u>Tasks and milestones</u>

Data Collection and Preprocessing phase

- Identify relevant sources for cancer patient datasets
- collect diverse datasets including medical records, treatment history, lifestyle data and genomic data if required.
- Clean and preprocess the collected data to ensure consistency and quality
- Research and review established projects of similar magnitude.
-

Model Development and Training phase

- Design and develop an AI/ML predictive model prototype for individual cancer patients' needs.
- Segment the datasets into training, testing and validation sets for effective predictive models.
- Carry out extensive testing of these prototypes in order to evaluate their performance and accuracy.

- Fine tune the model for effective predictive performance
- Evaluate the predictive model using F1-score and recall
- Using comparative analysis(with other published models) adjust the model architecture if need be.

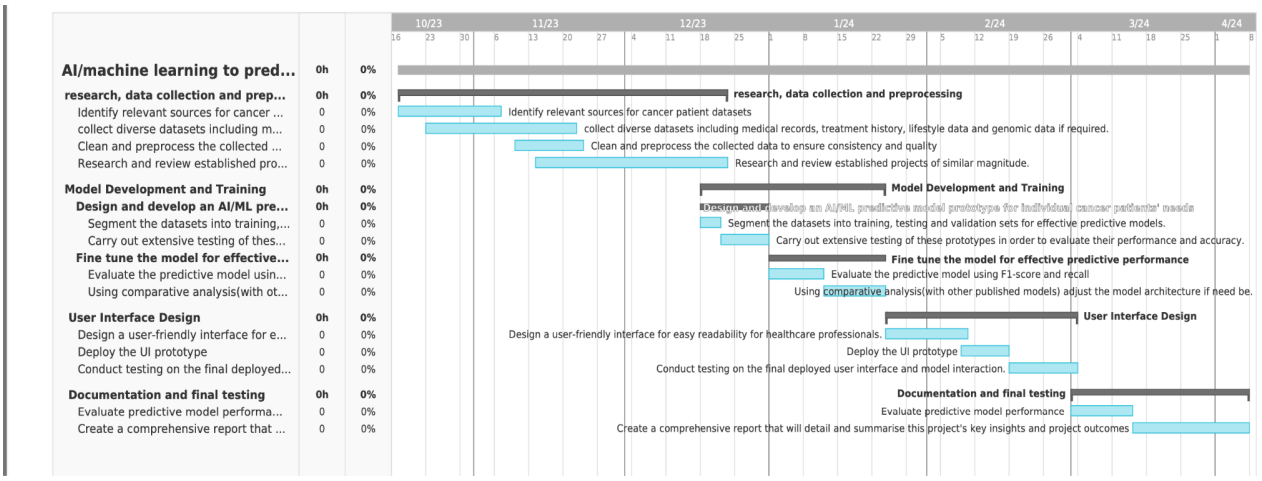
User Interface Design phase

- Design a user-friendly interface for easy readability for healthcare professionals.
- Deploy and the UI prototype
- Conduct testing on the final deployed user interface and model interaction.

Documentation and Reporting phase

- Create a comprehensive report that will detail and summarise this project's key insights and project outcomes.
- Generate a document which outlines the project's finding and methodology.

Schedule Gantt chart



Project control

The project will adhere to scrum methodology,tasks will be organised into weekly sprints to guarantee goals are manageable and targeted.Weekly evaluations will be carried out based on using sprint tasks outlined in the Trello board.A all encompassing project schedule will be maintained facilitating the adherence to predetermined deadlines and budget constraints.

The main focus of performance monitoring will be ongoing evaluations of key performance indicators, highlighting the predictive models accuracy and integration into the healthcare system.Regular evaluations of project development within the weekly sprints will provide an insight into how it is adhering to the defined project plan

The successfulness of the project will be determined based on the timely completion of the project, achievement of predefined prediction accuracy targets and the demonstration of the

project's capacity.Collectively the aim is to improve the quality of care and support for cancer patients through effective utilisation of datasets within the designated weekly sprint structure.

Project evaluation

I intend to evaluate the projects outcomes and artefacts by examining the performance of the predictive models.This involves analysing metrics such as precision,recall and F1-score comparing them with industry standards.Alongside this the user interface design will be assessed by testing its user friendliness and accessibility for healthcare professionals

In terms of user evaluation I plan to simulate realistic healthcare situations to test the functionality of the predictive models.In order to understand user experiences and satisfaction levels.I will also analyse historical patient data to measure the impact of the predictive models on improving the quality of care for cancer patients.Throughout the project I will continuously analyse the outcomes to identify any areas that can be improved and adjust the models accordingly.

Articles

Below are some articles I plan to use as a comparison and inspiration for my predictive model.

1. *Chien, S.-C.; Chang, Y.-H.; Yen, C.-M.; Chen, Y.-E.; Liu, C.-C.; Hsiao, Y.-P.; Yang, P.-Y.; Lin, H.-M.; Lu, X.-H.; Wu, I.-C.; et al. Predicting Long-Term Care Service Demands for Cancer Patients: A Machine Learning Approach. Cancers 2023, 15, 4598. https://doi.org/10.3390/cancers15184598*

2. *Zhang B, Shi H, Wang H. Machine Learning and AI in Cancer Prognosis, Prediction, and Treatment Selection: A Critical Approach. J Multidiscip Healthc. 2023 Jun 26;16:1779-1791. doi: 10.2147/JMDH.S410301. PMID: 37398894; PMCID: PMC10312208.*

3. *Xu L, Sanders L, Li K, Chow J*
Chatbot for Health Care and Oncology Applications Using Artificial Intelligence and Machine Learning: Systematic Review
JMIR Cancer 2021;7(4):e27850
URL: https://cancer.jmir.org/2021/4/e27850
DOI: 10.2196/27850

https://www.kaggle.com/

https://ieeexplore.ieee.org/Xplore/home.jsp

References

B, H.N. (2020) Confusion matrix, accuracy, precision, recall, F1 score, Medium. Available at: https://medium.com/analytics-vidhya/confusion-matrix-accuracy-precision-recall-f1-score-ade299cf63cd (Accessed: 12 October 2023).
https://neptune.ai/blog/f1-score-accuracy-roc-auc-pr-auc

Department of Health and Social Care (2003) *Confidentiality: NHS code of practice*, *GOV.UK*. Available at: https://www.gov.uk/government/publications/confidentiality-nhs-code-of-practice (Accessed: 15 October 2023).