Machine Learning (COS6026-B)
First Coursework

Prof Rami Qahwaji
rsrqahwa@bradford.ac.uk

Background

Diabetic retinopathy (DR) is a consequence of diabetes which manifests itself in the retina. DR is a major cause of visual impairment globally. In 2011, 366 million people were diagnosed with diabetes and a further 280 million people were at risk of developing it. Approximately 40% of diabetic patients suffer from DR.
The Diabetic Retinopathy Debrecen Data Set Data Set, available at
http://archive.ics.uci.edu/ml/datasets/Diabetic+Retinopathy+Debrecen+Data+Set
This dataset contains features extracted from retinal images to predict whether an image contains signs of diabetic retinopathy or not. The detailed description of the attributes can be found at
http://archive.ics.uci.edu/ml/datasets/Diabetic+Retinopathy+Debrecen+Data+Set

Aim

-To produce a machine learning system for the detection of signs of DR. The system will be trained and tested on ***The Diabetic Retinopathy Debrecen Data Set Data Set***.
- To write a research report on the background, rationale, development, implementation, evaluation and conclusions of your machine learning system.

Design Problem

Problem statement

Machine Learning has the potential to support the development of innovative solutions for improved real-life and real-time healthcare delivery. For this coursework, your aim is to investigate the problem of detection of DR using machine learning. In doing so, you would be expected to perform the following tasks:

- Investigate the Diabetic Retinopathy Debrecen Dataset, and develop basic understanding of the significance and meaning of data attributes and features
- Develop pre-processing and/or feature extraction techniques (if needed)
- Develop and Implement (using Python or Matlab) a machine learning system that can be implemented to demonstrate your ideas
- Demonstrate the feasibility of your design by evaluating your system using machine learning metrics
- Discuss your findings making recommendations for further investigation and improved implementation.
- Write a 1000 words report to present your work and findings.

Dataset

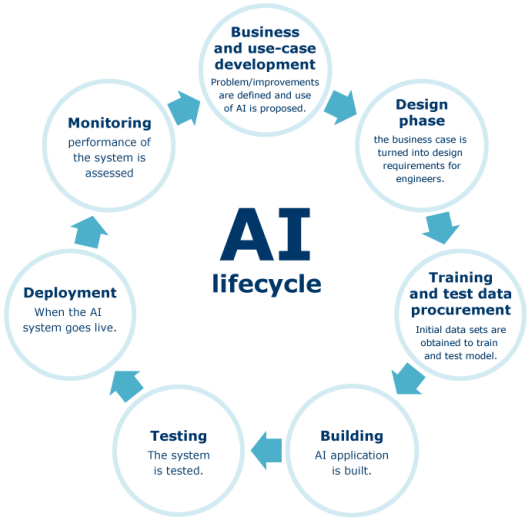
The Diabetic Retinopathy Debrecen dataset is available publically at:
http://archive.ics.uci.edu/ml/datasets/Diabetic+Retinopathy+Debrecen+Data+Set

Useful Reading

Balint Antal, Andras Hajdu: An ensemble-based system for automatic screening of diabetic retinopathy, Knowledge-Based Systems 60 (April 2014), 20-27.

Effort required

The coursework is worth 50% of the module marks. You are expected to demonstrate good effort in exploring the background of this problem, understanding the datasets, processing the data, developing meaningful machine learning system and implementing some basic aspects, and writing up your report. The coursework is worth 50% of the marks for this 20-credit module. A 20-credit module is nominally 200 hours of your time, some of which is spent in lectures, labs and private study. You should expect to be spending around 70 hours on this coursework.
The development of machine learning system goes through different phases, as discussed in lectures. The AI life Cycle is shown below, as a quick reminder. You should consider these phases in your report when demonstrating the design and development of your system. As there are many phases in the AI life Cycle, you need to describe the goals and objectives of your system, the feasibility of selected data, different users and their requirements, the basic design of the system, training and testing concepts, deployment challenges, etc. You are encouraged to implement one of these phases in a programming language (Python, Matlab, Java, etc).



Deadline for submission of report

The report must be submitted through Canvas before **4.00 pm Friday 19th March 2021**
Late submissions (without acceptable extenuating circumstances) will receive a mark of *zero*.

Report structure

You report (about 1000 words) should contain the information listed below.

- **Cover Page**: Title, author, affiliation of the author, date, and abstract
- **Introduction**: A rough overview of the diabetic care challenge you are tackling and its importance, frequently formulated to attract the reader's interest to the report.

- **Background**: Here you are required to have an overview about Diabetic care and AI technologies. Then, you can describe relevant approaches and systems, or you can introduce basic concepts that are necessary for understanding the later material.
- **Methodology and Data**: This section contains an explanation, demonstration, description of the system architecture, or some interesting implementation techniques. Discussion of some targeted methods for solving the Diabetic care problem will be mostly encouraged. Description of the knowledge representation (datasets used and data processing issues) are discussed in this part. The development of the AI system is discussed in this part as well
- **Analysis and Discussions**: You need to present your findings and discuss, analyse, evaluate and/or criticise what you implemented and described in the previous part.
- **Conclusions and suggestions for future work:** the major findings from doing this coursework will presented in this section. You can comment on the lessons learnt from doing this coursework, advantages/limitations of your AI, resources needed to implement the full system, what would you differently if you had more time, etc
- **Bibliography and Citations**: It is imperative that whenever you make reference to a fact of some sort, you cite an authoritative source for that fact; most frequently, these sources will be scientific articles.

Specific Learning outcomes

This coursework will enable you to apply basic principles of machine learning and to exercises on the development and creation of machine learning systems.

Marking Scheme:

Criteria	>80%	70-80%	60-70%	50-60%	40-50%	39-25%	<25%
Presentation of the report (25%)	Fully adheres to student guidelines. Complete lack of spelling and grammatical errors. Excellent use of appropriate and scientific language. Excellent structure and organisation. Excellent readability.	Good degree of adherence to student guidelines. Lack of spelling and grammatical errors. Good use of appropriate language. Excellent structure and organisation. Good readability	Mainly adheres to student guidelines. Minor spelling and grammatical errors. Good use of appropriate language. Well-structured with logical organisation.	Some adherence to student guidelines. Some spelling and grammatical errors. Inconsistent use of appropriate language. Organisation and progression evident.	Little adherence to student guidelines. Many spelling and grammatical errors. Minimal use of appropriate language. Inadequate attention to structure and organisation	Does not adhere to student guidelines. Major deficiencies in spelling and grammar. Lack of appropriate language. A disorganised report with lack of evident structure.	No report submitted. Paperwork submitted does not constitute a meaningful report.
Description and achievement (25%)	Novel or innovative solution to AI-based healthcare. Exceptional amount of high quality work, which could include comprehensive testing, or critical reviews, or results incorporated into prototype.	The AI system presented is meaningful, innovative challenging and possibly complex. Problem explicitly stated with precise explanation of all research objectives. Excellent analysis and recommendations.	A relevant and original topic which is effectively translated into project aims and objectives which are clearly stated. Good analysis and recommendations.	Appropriate problem area chosen. Objectives outlines with the main areas of investigation identified. Some analysis and recommendations.	Limited topic choice with the problem area poorly defined. Objectives vague or insufficient. Very limited analysis and recommendations. .	Simple or unoriginal problem showing lack of imagination. No analysis and recommendations	No software/project completed. No working software/project (does not apply to theoretical submission).
The demonstration of development capability (25%)	Signs of professionally developed AI concepts, excellent understanding of user requirements, excellent specifications and system	Fully developed AI concepts, fully-developed user requirement , fully developed specification and System requirement. Evidence of	Evidence of developed AI concepts, developed user requirements, developed specifications and System requirements. Good	Partial in-depth development of AI concepts, Partial in-depth developed user requirements, Partial in-depth developed specifications and System	Limited development of AI concepts, limited developed user requirements, limited developed specifications and System requirements.	No or very limited development of AI concepts, no or very limited developed user requirements, no or very limited developed specifications and	No demonstration given. System failed to work. Student could not explain working of software or technical aspects of project (does not apply to theoretical

	requirements. Specification effectively interpreted with signs of original thinking and/or research.	excellent understanding of AI-development process and applications.	understanding of the AI-development process and applications.	requirements. Evidence of a fair understanding of the AI-development process and applications.	Evidence of a limited understanding of the AI-development process and applications.	System requirements. No evidence of understanding of the AI-development process and applications.	submission).
Quality of work and understanding (25%)	Wide range of material investigated. Excellent evidence of critical evaluation and original thinking. Substantial number of appropriate references. Evaluation and recommendations fully and appropriately reviewed and presented. Excellent link to implementation. Research findings fully consider broader issues.	Good range of material. Good evidence of critical evaluation. Good number of appropriate references. Recommendations fully and appropriately reviewed and presented. Good discussion of implementation. Research findings discusses broader issues.	Focus on key areas using relevant sources. Good range of references with a varied bibliography. Clear recommendations identifying key issues. Research findings apparent with some consideration of broader issues.	Adequate information survey with some evidence of investigation of a key area. Appropriate range of references. Good bibliography. Recommendations identify some key issues. Research findings lack consideration of broader issues.	Limited sources of information used. Limited range of references. Limited bibliography. Limited recommendations. Lacking clarification of research findings.	Review of existing literature not evident. No references provided with limited or omitted bibliography. No recommendations. Conclusions do not link to research findings.	No understanding demonstrated. No conclusions. No future work suggested.