**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 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.

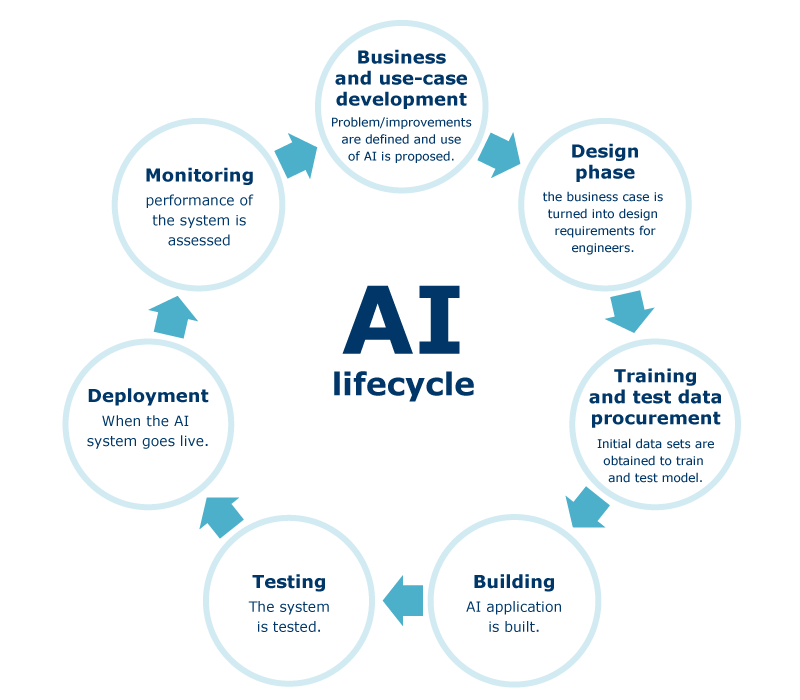
**Dataset**

The Diabetic Retinopathy Debrecen dataset is provided as an attachment with this document.

**Reading for the Project**

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

The AI life Cycle is shown below, as a quick reminder. 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.



**Report structure**

The report should be 1000 words

* **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 will be presented in this section. You can comment on the lessons learnt from doing this project, 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.