COMP1804	**Applied Machine Learning**	**Faculty Header ID:**	**Contribution 100% of course**
Course Leader **Dr. Dimitrios Kollias**	**COMP1804 Coursework**		**Deadline Date:** **16/04/2021**
This coursework should take an average student who is up-to-date with the lectures and the labs approximately 50 hours Feedback and grades are normally made available within 15 working days of the coursework deadline			
Learning Outcomes: 1. Demonstrate understanding and ability to building and pre-processing data for Machine Learning (ML) algorithms. 2. Critically evaluate the merits and limitations of machine learning approaches and algorithms and evaluate the choice of algorithms for specific real-world contexts and requirements. 3. Apply ML algorithms to a selected real-word problem in practice and understand the processes involved in their deployment. 4. Demonstrate ability to evaluate the performance of ML algorithms in the context of the specific real-world problem.			

Plagiarism is presenting somebody else's work as your own. It includes: copying information directly from the Web or books without referencing the material; submitting joint coursework as an individual effort; copying another student's coursework; stealing coursework from another student and submitting it as your own work. Suspected plagiarism will be investigated and if found to have occurred will be dealt with according to the procedures set down by the University. Please see your student handbook for further details of what is / isn't plagiarism.

All material copied or amended from any source (e.g. internet, books) must be referenced correctly according to the reference style you are using.

Your work will be submitted for plagiarism checking. Any attempt to bypass our plagiarism detection systems will be treated as a severe Assessment Offence.

Coursework Submission Requirements

- An electronic copy of your work for this coursework must be fully uploaded on the Deadline Date using the link on the coursework Moodle page for COMP1804.
- For this coursework you must submit 7(+2) separate files:
  - A single pdf file named ‘report.pdf’ which will be the written report; the written report must have a maximum limit of 3500 words
  - A single zip file named ‘model.zip’ which will be the trained model weights in TensorFlow format
  - A single csv file named ‘annotations.csv’ which will include the annotations for the given unlabelled data. Regarding the format:
    - the first line should be:
      image_name,wrinkles,freakles,glasses,hair_color,hair_top
    - next lines should list the image name & corresponding annotations, e.g.:
      flick_23.jpg,1,1,2,5,2
      photobucket_123.jpg,0,1,0,3,1
      pexels_0.jpg,0,0,1,6,0
  - A single ipython notebook file named ‘test_code.ipynb’ which will be the source code for testing the developed model on new images (this code should take as input new images’ file locations and output a file with the predictions of the model in the same format as the above described annotations.csv file)
  - A single ipython notebook file named ‘training_code.ipynb’ which will be the source code for training the model on images (this code should read the images and their annotations from a file having the same format as the above described annotations.csv file)
  - A text file named ‘requirements.txt’ containing the libraries required to run the above described test_code.ipynb, if there are any
  - A single video file named ‘demo.mp4’ which will be the demo showing the developed model predicting the facial attributes on a few (3-10) unseen images (not belonging to the provided dataset)
  - If the provided annotation tool was used: the video that you created for the annotation tool named ‘video.mov’ (.mp4/.avi or any other extension)
  - If the provided annotation tool was used: the text file created by the provided script (‘merge_images_to_video.py’); the text file should be named ‘image_names.txt’

  In general, any text in the document must not be an image (i.e. must not be scanned) and would normally be generated from other documents (e.g. MS Office using "Save As .. PDF"). An exception to this is hand written mathematical notation, but when scanning do ensure the file size is not excessive.
- There are limits on the file size (see the relevant course Moodle page).
- Make sure that any files you upload are virus-free and not protected by a password or corrupted otherwise they will be treated as null submissions.
- Your work will not be printed in colour. Please ensure that any pages with

colour are acceptable when printed in Black and White.
- You must NOT submit a paper copy of this coursework.
- All courseworks must be submitted as above. Under no circumstances can they be accepted by academic staff

The University website has details of the current Coursework Regulations, including details of penalties for late submission, procedures for Extenuating Circumstances, and penalties for Assessment Offences.

See http://www2.gre.ac.uk/current-students/regs

Detailed Coursework Specification

Designing a machine learning solution requires considering several aspects of the problem, the availability of data and corresponding annotations, nature of the problem addressed, methodology choice, evaluation among others. It is important for our students to be up to date with current practices and Machine Learning techniques used in the modern software that drives many computers and devices today and be familiar with their strengths and limitations. It is of equal importance for our students to familiarize with the whole data processing and evaluation pipeline enabling successful implementation of Machine Learning techniques. Adding these skills to their portfolio will increase the employability of our graduates and will help them to aim for higher paying jobs in industry, as well as academia.

The task is to implement a ML solution for facial attribute recognition/classification.
The facial attributes are:

- wrinkles (binary: has/does_not_have), class 0: does_not_have, class 1: has
- freakles (binary: has/does_not_have), class 0: does_not_have, class 1: has
- glasses (3 values: do_not_wear/wear_normal/wear_sunglasses), class 0: does_not_wear, class 1: wear_ normal, class 2: wear_sunglasses
- hair_color (9 values: brown/black/gray/blond/red/white/mixed/other), class 0: brown, class 1: black, class 2: gray, class 3: blond, class 4: red, class 5: white, class 6: mixed, class 7: other, class 8: not_visible
- hair_top (4 values: bald or shaved, has_few_hair, has_thick_hair), class 0: bald or shaved, class 1: has_few_hair, class 2: has_thick_hair, class 3: not_visible

A related unlabelled dataset will be distributed to each student. The student should implement the whole procedure for designing a ML approach for solving the problem.

Tasks:

1. **Practical Assignment (65 Marks)** (complete training and testing code, trained model weights and annotations to replicate the results). The source code must be **error free** (i.e. no debugging necessary to run). The assignment includes:
    - Annotation and pre-processing: this should include data labelling, data splitting,

generating statistics and data pre-processing (20 marks; it will be possible to obtain 10 more marks if an additionally provided dataset is also labelled and used; this additional dataset to be provided upon request)
    - ML methodology: an appropriate ML method should be used that has a coherent implementation and a sound pipeline, without any errors (25 marks)
    - Experimental results: this should include evaluation of the ML algorithm performance with metrics and figures/tables. The method's performance will additionally be evaluated by the teaching staff on a held-out test set (20 marks)

2. **Written Report (35 Marks)**:
    - Document in IEEE conference format (Use template available online: https://www.ieee.org/conferences/publishing/templates.html)
    - Should include references (citing other work) where appropriate (when images, data, code, or any other resources have been used from other sources)
    - Document structure:
        - **Introduction**: Introduce the problem to be solved (motivation, expectances, goals, implications)
        - **Related work**: Short survey of other existing/state-of-the-art work on the studied problem
        - **Dataset preparation**: Describe data collection, processing and partitioning (sources, augmentation techniques, training setup, …)
        - **ML method**: Motivate method-related choices, explain how the method works (motivate decisions during training, …)
        - **Evaluation**: Evaluate, present, analyse and explain method performance (highlight the pros and cons of it; focus on both good and bad: an unsatisfactory result, if well explained, can help others focus on working solutions; include a demo)
        - **Future work**: Reflections on how your work could be extended in the future; what addition can be made to it
        - **References**: All existing works and resources (code/images/etc) you used or talked about in your report must be cited properly

Deliverables:

An admissible coursework submission needs to include:

- All coursework submission requirements as specified in the Coursework Submission Requirements section above should be uploaded by the Deadline Date using the link on the coursework Moodle page for COMP1804.

Grading Criteria

For a distinction (mark over 70) the following is required:

- An excellent implementation, showing a whole system with all requirements

implemented; all components are working and provide a very good result.
- An excellent report and demonstrating a good understanding of motivating, building and evaluating a working machine learning application.

Note: In order to be eligible for very high marks (**80 & over**) you will need to have:
- An innovative implementation, showing all requirements are implemented to a higher standard. The components should be working properly and provide an excellent result; potentially extra credit features implemented.
- An outstanding report with excellent portfolio, demonstrating a thorough understanding of motivating, building and evaluating a working machine learning application.

For a mark in the range 60 to 69 the following are required:

- A good implementation, showing a system with no errors and with a sound training pipeline and annotation leading to good evaluation-results.
- A good report demonstrating a good understanding of motivating, building and evaluating a working machine learning application.

For a mark in the range 50 to 59 the following are required:

- An implementation showing a reasonable system with at least following minimum requirements implemented: basic annotation, flawless ML implementation, reasonable evaluation-results.
- An adequate report showing some understanding of motivating, building and evaluating a working machine learning application.

For a mark below 50:
- A system that fails to implement the minimum requirements, including basic annotation, flawless ML implementation, reasonable evaluation-results.
- An unsatisfactory report showing little understanding of motivating, building and evaluating a working machine learning application.

Assessment Criteria

The practical assessment = 65 Marks
Marks will be given for:

- Features implemented.
    - The extent to which a successful annotation, pre-processing, training and evaluation pipeline was implemented will have an important effect on your overall mark.

- The quality of the system you produce.
    - Credit will be given for excellent robust design, complexity of the implementation, components that are working without giving any errors and reliably producing good result; possible enhancements to the system.

The report = 35 Marks
Marks will be given for:

- Critical understanding of relevant concepts, reading and referencing related papers, appropriate explanation and discussion.
- Quality of the report:
  Are all the required sections included and completed properly? Is the report clear, well formatted and easy to read? Does it have a logical structure? Does it have a discussion on design decisions? Is the evaluation realistic, does it show that you have really thought about your system and how you went about developing it.