School of Computing & Communications | Lancaster University

SCC413 – Applied Data Mining

Coursework Assignment

1 Introduction

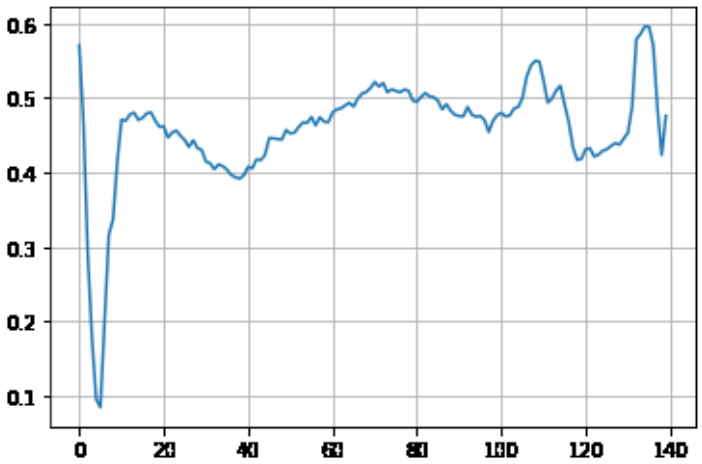
The objective of the assignment is to carry out a comparatives study of various neural network methods over a given real-world dataset. The assessment is expected to demonstrate your learning outcomes on:

- Be capable of the analysis of an applied data mining task over a given topic including its background, motivation and challenges, and can bring out your own methods/solution for this topic.
- Be able to carry out literature review on existing publication resources and justify your proposed/selected methods.
- Be able to demonstrate your understanding of your selected/proposed methods at a depth.
- Be able to implement and test your methods on the given dataset.
- Be able to carry out discussion and reflection on your research

In addition to your report, please submit your source code via GitHub, and share the link within the report. We expect your codes (Jupyter + Python) can run on Google Colab environment.

2 Data-Set

You are expected to use the ECG dataset provided in the coursework folder on the Moodle. The dataset consists of abnormal and normal subjects for classification. In total there are nearly 5000 samples.



The dataset is split into a train dataset of nearly 4000 samples and a test dataset of 1000 samples.

3 Tasks description

You are handed out the above ECG dataset. Your task is to classify these data sample into two classes: normal and abnormal, using 2-3 neural network methods (including MLP, autoencoder, CNN, LSTM, etc.).

You can use any deep neural network methods for your task, including,

1. MLP, Autoencoders, CNNs and LSTMs introduced in the labs and lectures. You are welcome to re-use your lab scripts;
2. Other standard deep learning classifiers you can find on GitHub or any other online resources, such as GoogleNet, ResNet, VGG, and Inception Models, etc.;
3. Your own developed novel Deep Learning architecture (you are welcome to come up with your own novel model and prove it can work better on ECG data).

Your core task is to compare your selected/proposed methods with each other with respect to their training process and test results, on the given dataset.

At the end, you need to deliver a report on your work, which is expected to include key sections:

1. An abstract summarizing your work
2. An introduction to address the needs of ECG data classification in the applied areas, and its motivation. Also highlight your work to do.
3. Literature review, where you can include a review on recent journal papers on ECG classification. What others have done? And why you are going to carry out this comparative study?
4. Your methodology, where you details on the selected/developed methods in your comparative study
5. Experimental results on training and test.
6. Discussions on pros and cons of each method, and reflection on your research process.
7. Conclusion, quite a response to abstract and introduction sections.
8. References – likely corresponding to your literature review and methodologies

Please submit your report in both PDF and Word versions.

4 Marking Scheme

For the overall report, marks are allocated as follows:

- Abstract (5%)
- Introduction (8%)
- Literature Review (10%)
- Methodologies (25%)
- Experiments (35%)
- Discussions (7%)
- Conclusion (5%)
- References (5%)

The total 100 marks is for the coursework 01, which will be converted to 50% for the whole module. The marking will consider key aspects including

- Structure and presentation
- Language and style

- Level of understanding
- Depth of analysis
- Use of literature and references
- Working, well annotated code and results
- Justification of selected methods
- Independent research and use of methods not given in the lectures

The above aspects will be considered in the marking of relevant sections.

5 Report Template

Please see the attached template (IEEE format). You need to complete all corresponding sections in the template. You need to submit the word version (or LaTex sources) of your report beside a PDF version.

6 Code Submission

It is recommended you can include all your Jupyter codes on your GitHub folder and allows click &run on colab. Commonly each method will be in one Jupyter demo script. You need to include the link to the code folder in your report.

You can alternatively submit your report with your scripts in one zipped file, to avoid any issues with GitHub, in case you are not sure.

Submissions without codes (GitHub links or zipped codes) will be marked as zero.

What a Mark Means in Lancaster University

70 + (Distinction)

Critical Understanding of Topic

Excellent understanding and exposition of relevant issues; insightful and well informed, clear evidence of independent thought; good awareness of nuances and complexities; appropriate use of theory.

Structure of Research

Substantial evidence of well implemented independent research and / or Substantial evidence of well selected evidence to support argument.

Use of Literature

Excellent use of literature to support argument /points.

Conclusion

Excellent; clear implications for theory and/or practice.

Language

Excellent; a delight to read.

Structure and Presentation

Arguments clearly structured and logically developed; sensible weighting of parts; meaningful diagrams; properly formatted references.

65 – 69% (Very Good Pass)

Critical Understanding of Topic

Clear awareness and exposition of relevant issues; some awareness of nuances and complexities but tendency to simplify matters; based on appropriate choice and use of theory.

Structure of Research

Some evidence of independent research reasonably well implemented and / or some evidence of identification of suitable evidence to support argument.

Use of Literature

Good use of literature to support arguments.

Conclusion

Very good; draws together main points; some implications for theory and/or practice

Language

Carefully written; negligible errors.

Structure and Presentation

Arguments clearly structured and logically developed; good weighting of parts; meaningful diagrams; properly formatted references.

60 – 65% (Good Pass)

Critical Understanding of Topic

Shows awareness of issues and theories; attempts at analysis but tendency to lapse into description

Structure of Research

Some evidence of independent research reasonably well implemented and / or some evidence of identification of suitable evidence to support argument.

Use of Literature

Use of standard literature to support arguments.

Conclusion

Reasonable conclusion that summarises essay; a few implications for theory and/or practice.

Language

A few errors; generally satisfactory.

Structure and Presentation

Arguments reasonably clear but undeveloped; some meaningless diagrams or poor structure.

50 – 59% (Pass)

Critical Understanding of Topic

Work shows understanding of topic but at superficial level; no more than expected from attendance at lectures; some irrelevant material; too descriptive.

Structure of Research

Insufficient evidence of independent research and / or very limited evidence used to support argument.

Use of Literature

Use of secondary literature to support arguments.

Conclusion

Conclusion does not do justice to body of essay; too short; no implications.

Language

Some errors; grammar and syntax need attention.

Structure and Presentation

Arguments not very clear; poor organisation of material; poor use of diagrams; poor referencing.

45 – 49% (Marginal Fail)

Critical Understanding of Topic

Establishes a few relevant points but superficial and confused; much irrelevant material; very little or no understanding of the issues raised by the topic or topic misunderstood; content largely irrelevant; no choice or use of theory; essay almost wholly descriptive; no grasp of analysis with many errors and/or omissions.

Structure of Research

No evidence of independent research and / or No attempt to identify suitable evidence to support argument.

Use of Literature

Relies on a superficial repeat of class notes.

Conclusion

No recognisable conclusion.

Language

Frequent errors; needs urgent attention.

Structure and Presentation

Arguments often confused and undeveloped; no logical structure; very poor organisation of material; many meaningless diagrams; negligible referencing.

0 – 44% (Clear Fail)

Critical Understanding of Topic

Establishes a few relevant points but superficial and confused; much irrelevant material; very little or no understanding of the issues raised by the topic or topic misunderstood; content largely irrelevant; no choice or use of theory; essay almost wholly descriptive; no grasp of analysis with many errors and/or omissions.

Structure of Research

No evidence of independent research and / or No attempt to identify suitable evidence to support argument.

Use of Literature

No significant reference to literature.

Conclusion

No recognisable conclusion.

Language

Frequent errors; needs urgent attention.

Structure and Presentation

Arguments often confused and undeveloped; no logical structure; very poor organisation of material; many meaningless diagrams; negligible referencing.