CCT College Dublin

Assessment Cover Page

|  |  |
| --- | --- |
| Module Title: | Advanced Data Analytics  Big Data Storage and Processing |
| Assessment Title: | Integrated CA1 Sem 2 MSc in Data Analytics |
| Lecturer Name: | David McQuaid  Muhammad Iqbal |
| Student Full Name: | Mohamed Noor |
| Student Number: | SBS22070 |
| Assessment Due Date: | 03/10/2022 |
| Date of Submission: | 02/10/2022 |

Declaration

By submitting this assessment, I confirm that I have read the CCT policy on Academic Misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source. I declare it to be my own work and that all material from third parties has been appropriately referenced. I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution.

Image Recognition for Learning Arabic Words

Mohamed Noor   
*Information and Communications Technology (ICT) Faculty   
CCT College Dublin*Dubln, Ireland  
sbs22070@student.cct.ie

*Abstract*—IEEE format requires you to include an abstract at the start of your paper, followed by a list of keywords. In the “Styles” section in Microsoft Word, you can find the appropriate styles for all the different sections and headings in the paper, which are already applied here. For example, the “abstract” style is applied to this text, the “keywords” style to the next section. Note that the titles “Abstract” and “Keyword” should remain as they are written here: italicized and followed by an em dash.

Keywords—template, Scribbr, IEEE, format

# INTRODUCTION

Level 1 and 2 headings (as well as the paper title) should be written with title case capitalization, while level 3 and 4 headings are written in sentence case.

Unlike English, Arabic is an unusual language. For example, there are 200 words that refer to lion in Arabic [1]. Moreover, the definite article ‘the’ in Arabic (‘al’, ال) never stands alone. Yet, when joined with another word will change its morphology (for example, lion (أسد) will become الأسد). There are also regional variations, which is why learners of Arabic usually tend to follow the Modern Standard Arabic (cite).

Given the complexity of Arabic compared to English, it would be helpful for learners to have the ability to instantaneously convert images to Arabic words, especially through smartphones.

Image recognition through deep learning is a particularly useful method for classifying images. There are two potential solutions: one is multi-class problem where each image is assumed to only take one label and the entire dataset consists of multiple labels, or a multi-label problem where each image can take multiple labels. For instance, an image of a woman walking with a dog could be classified potentially as a ‘woman’ or a ‘dog’ (multi-class) but never both, with a multi-label algorithm will identify both elements.

As a proof-of-concept, this paper discusses the use of pre-trained image classification algorithms for identifying images in Arabic. Pre-trained algorithms take advantage of the transfer learning concept in deep learning, where a model trained on a problem can be used in another problem if these are relatively similar. For instance, a model training on recognizing XXX can be used for YYY (give examples peer-reviewed).

Rationale for selecting the MMID – why this dataset and not the others???

MMID reference paper: [2]

## This Is a Level 2 Heading

### And this is a level 3 heading: Equations should be typed in either Times New Roman or Symbol font, or, if the equation is multileveled, inserted into your text as a graphic instead. On the far right of the line containing the equation, number it in parentheses, and use this number to refer to it in the text (1).

*a**b* 

### This is another level 3 heading: Lorem ipsum…

| Table Head | Table Column Head | | |
| --- | --- | --- | --- |
| Table column subhead | Subhead | Subhead |
| text | Texta |  |  |

#### And this is a level 4 heading: Make sure that the

1. This is a figure caption. It appears directly underneath the figure.

appropriate style is still applied to each section, reapplying styles if necessary.

#### This is another level 4 heading: It’s also possible to add bullet points when appropriate, using the “bullet list” style:

* Treat the word “data” as plural, not singular.
* For example, “the data indicate that …”

## This Is Another Level 2 Heading

Place any figures or tables you use at the top or bottom of a column. Don’t place them in the middle of a column. If particularly wide, a table or figure can span across both columns. Insert a table or figure after the point where it is first cited in the text.

When inserting a figure, such as a photograph or infographic, use 8 pt. Times New Roman for any labeling text within the image and for the figure caption. You can see an example of a figure caption in Fig. 1, above. Refer to figures like that, using the abbreviation “Fig.” and the figure’s number.

A table heading (using the “table head” style) appears above a table. This will automatically number the table for you. Any footnotes appear below the table, using the “table footnote” style. Footnotes are indicated by superscript lowercase letters within the table. An example of a table can be seen in Table I, below.

# RESULTS

All the headings in the main body of your paper are numbered (automatically).

## Image selection

For this project, Python XX was used together with YYY.

## Programming environment

For this project, Python XX was used together with YYY.

## Exploratory data analysis

YYY.

1. This Is the Heading for a Table
2. This is a table footnote.

Lorem ipsum….

##### Acknowledgments

This study used data from MMID ….

“Acknowledgment(s)” is spelled without an “e” after the “g” in American English.

As you can see, the formatting ensures that the text ends in two equal-sized columns rather than only displaying one column on the last page.

This template was adapted from those provided by the IEEE on their own website.

##### References

[1] P. Zemánek, “Two Hundred Ways to Call a Lion in Arabic: Names or Epithets?,” *Z. Dtsch. Morgenländischen Ges.*, vol. 171, no. 2, pp. 343–374, 2021, doi: 10.13173/zeitdeutmorggese.171.2.0343.

[2] J. Hewitt, D. Ippolito, B. Callahan, R. Kriz, D. T. Wijaya, and C. Callison-Burch, “Learning Translations via Images with a Massively Multilingual Image Dataset,” in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Melbourne, Australia, Jul. 2018, pp. 2566–2576. doi: 10.18653/v1/P18-1239.