# NCTA 2021 - Paper #5

**Paper Title:** *Working for the Apple Company and LinkedIn Profile Pictures*

**Reviewer** *#1*

## General Assessment (P) (Please assign scores using the following criteria (1=weakest; 6=strongest))

Relevance (Paper fits one or more of the topic areas?): 4 Originality (Newness of the ideas expressed): 4 Meaningfulness and R&D potential of the proposed position 2 Significance (Is the problem worth the given attention?): 2 Presentation (Structure/Length/English): 1

Overall Rating (Weighted value of above items): 3

## Improvement Suggestions (for authors to consider in the camera-ready version. Additional detail in "Obser- vations" )

Abstract and Introduction are adequate? No

Figures are adequate ? (in number and quality): Yes Conclusions/Future Work are convincing? Yes References are up-to-date and appropriate? Yes Paper formatting needs adjustment? Yes

Improve English? Yes

The position proposed needs further clarification Yes

## Detailed comments to authors, including aspects that must be improved in the camera-ready version of the paper:

The authors in this paper propose a relatively simple machine learning approach for emotion recognition between Apple and non Apple employees. The paper might be of interest, yet there are several issues that need to be addressed.

*−*

1. I would suggest to improve the title : the claim and the objective of the paper must be stressed. For example

, something like "Apple Employee recognition via Linkedin profile pictures" or since this is a position paper, the authors may also want to consider a question oriented title , e.g. "Is it possible to recognize Apple employees by their LinkedIn profile picture?"

*−*

Leslie

He suggests to make it really a position paper, like “we think this is possible, look what we have done”, instead of using it really as a research paper “we are sure we found this”. I have no problem with extra stress and maybe indeed the position paper approach.

1. The paper needs a thorough spell check and grammar check. For example, "searching for current employment

at Apple or not Apple and Bay Arena" should be \*at Bay Arena\*. Or "The research uses Machine Learning and neural network technology" should be \*technologies\*.

*−*

We need to find a native speaker who is also technical for that.

If you need, I can get a few friends to review it for us. After we review it ourself of course.

1. In the Introduction, the authors mention "A number of ML related statistical software algorithms is trained

[...]". This is not true . In the experimental part of the paper, the authors just show a Random Forest classifier.

There are two options. Either get the other ones in, as described by your thesis, or remove the offending sentence.  
  
  
Actually we did train bunch of different algorithms but i guess we just didnt include the details in this short paper.

1. At the end of page 2 ( left column) the sentence "This package is used in this work." can be omitted. At the beginning of section 4 the authors already mention that they use EmoPy.

Leslie

1. On page 2 (right column) the authors mention that they had to stack 2 models in order to predict 7 emotions instead of 4. Yet, at the beginning of right column, the authors seem to argue that EmoPy is already able to recognize presence/strength of 7 emotions ... so why the stack? The authors must discuss and present this aspect in more details .

We did this because the model that is able to recognize 7 emotions was not available at the time (due to a bug). Now its probably available though. So I used these 9 models, and take the average output of each emotion output. I think in my MBA paper I might have mistakenly said average of 2 models :(. But anyway I can confirm that this is actually what I did. here is the full code (https://colab.research.google.com/drive/16IGOlmyA2GHCALqKct39E\_PR3pKOd1-E?usp=sharing). Please see the diagram below (6) for further clarification.

Text

Description automatically generated

Nic

1. The role of the four face recognition methods is unclear. From page 3 (left column) it seems that 28 features are used (7 emotions x 4 recognition methods), whereas on right column the authors seem to suggest that only 7 features (7 emotions) are extracted from each image. The authors must discuss and present this aspect in more details .

Nic. I assumed that this was part of the Emopy package. Is there more documentation about this?

Would a diagram and explanation like this help

Diagram

Description automatically generated

1. raw image data

2. emotion output set 1 ('anger', 'fear', 'surprise', 'calm')

3. emotion output set 2-8 (...)

4. emotion output set 9 ('anger', 'happiness')

5. average emotion output for each algorithm, normalized to sum to 1

6. output from machine learning model to forecast Apple (eg Random Forest), we did try it with other ML algorithm (including Tensorflow Neural Network) as stated int he MBA thesis

Training algorithm : this is used during training of the Apple forcasting(6) model, This is not used during testing

1. The scikit links on page 3 (right column) seem to be dead

Leslie checks

here is the correct link

https://scikit-learn.org/stable/modules/generated/sklearn.inspection.permutation\_importance.html#rd9e56ef97513-bre

1. The sentence "Due to the balance sample requirement for the K Nearest Neighbour algorithm" on page 3 ( right column) is incorrect. K NN does not require class balancing: it might be a desiderata, but not a requirement.

*−*

*−*

Nic rewrites, maybe something like “we followed the common usage” or “we followed the preferred method” or something like that. We can also mention that analysis is more straightforward this way, and that it make the statistics more convincing via the 1000 times redrawing of training and sample.

I think we can say, "for simplicity of the analysis, we balance the samples by ....."

1. On page 3 (right column), the authors seem to suggest that a Random Forest can be equipped with many classifiers (Naiive estimate, Logistic Regression, Support Vector Machines, and so on) and this is not correct

. Whilst it is correct that any classifier can be stacked in an ensample (like Random Forests), Random Forests as such are ensemble of decision trees.

Nic

He is right, actually we should say that random-forest can be stacked (not equipped.)

1. What to the "True" and "False" mean on Table 1?

Leslie explains/rewrites

1. The distribution of the accuracies in Figure 3 refer to the accuracy on the test set? It ’ s worth specifying.

Nic

Answer is yes, but can you think of a nice formulation?

Each count is basically the accuracy of each test run (training a model from scratch). I think this is already the most appropriate visualization but I guess we just need to add more description at the bottom

1. On page 4 (left column) the authors highlight a negative relationship between sadness and surprise (and I quote) "is visible in the plots ". There are no such plots in the paper.

Leslie It was not meant to refer to the plots in the paper. But you have made them. As he is pushing more details (very unusual) we can include it few. The message is “use those 8 proceedings pages”

I think there is a negative relation for every emotion in general due to the output being a probability (sum to 1). This does not signify a real relationship but it is a result of normalization using weighted average

1. How the run in Figure 2 and Figure 5 has been selected? At random? If so, wouldn’t it be better to show the best confusion matrix/ROC curve or the average confusion matrix/ROC? Same reasoning holds for Figure 4.

I think I disagree. There is no way ahead of time to find the best one. And the best one is a fluke. On the other hand, it makes the paper more convincing, but for me that is just a statistical fluke. A fishing expedition for answer. The value of the 1000 runs is just to show that how robust the effect is.

1. Screenshots of parameters and confusion matrix are very inelegant. Can the authors replace screenshots with tables (confusion matrix) or lists (parameters of EmoPy)?

Leslie will do

If you need help with it let me know

1. In the Introduction, the authors must stress the claim of the paper. First of all , why Apple is the targeted company? And also, what do we expect from such experiments? What is the take home message? That Apple users show a "signature expression" conversely to other employees?

*−*

Leslie can do

1. How the hyperparameters of the random forest have been tuned?

Actually it has not been tuned, This is the default hyperparameter recommended by ScikitLearn website. Tuning will only make the model perform better however, it requires a lot more work. My assumption is that if we have time to tune the models, Deep Neural Network will perform better than random forest as is the case with most problems.

Here is the referenced website

https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

Nic

All in all , the paper can be of interest , but the manuscript is poorly written. I strongly invite the authors to better describe the feature generation step and to show the comparison of different algorithms (as allegedly claimed in the Introduction).

You can also read this differently “throw caution out of the window and state your claim forcefully” It is nice that a really good guy or girl says that.

If the authors need more space to better describe these steps , then Appendix A can be removed.

Yes agree. But I would keep the confusion matrix, personally.

Indeed, the readership of NCTA knows very well what a confusion matrix is, so there is no need to explain. Not to mention that a lot of space has been dedicated in introducing performance measures (e.g., Cohen’s Kappa, True Positive Rate, True Negative Rate, Prevalence, ...) that have not been used for assessing the performance of the classifier , hence that’s some free space that the authors might use for a better description of their approach and a deeper performance analysis.

# NCTA 2021 - Paper #5

**Paper Title:** *Working for the Apple Company and LinkedIn Profile Pictures*

**Reviewer** *#2*

## General Assessment (P) (Please assign scores using the following criteria (1=weakest; 6=strongest))

Relevance (Paper fits one or more of the topic areas?): 4 Originality (Newness of the ideas expressed): 4 Meaningfulness and R&D potential of the proposed position 3 Significance (Is the problem worth the given attention?): 3 Presentation (Structure/Length/English): 4

Overall Rating (Weighted value of above items): 3

Very different assessment from 1… that is common. It shows how subjective those things are.

## Improvement Suggestions (for authors to consider in the camera-ready version. Additional detail in "Obser- vations" )

Abstract and Introduction are adequate? Yes Figures are adequate ? (in number and quality): Yes Conclusions/Future Work are convincing? Yes References are up-to-date and appropriate? Yes Paper formatting needs adjustment? No

Improve English? No

The position proposed needs further clarification Yes

Agrees with 1.

## Detailed comments to authors, including aspects that must be improved in the camera-ready version of the paper:

This paper presents a classification problem considering the case of apple and non apple employees. This paper is globally well written and clear .

😉 well number 1 disagrees

my main remarks:

table 1 is unclear. What is control ? To my understanding, this corresponds to the three samples listed pp1. So why not to used it directly ?

Leslie clarifies

Figure 4 is interesting and must be more explained. I think the meand value 100/7 must be presented in order to see which emotion as a greater impact than the mean.

Nic can add something

The authors have transformed their 3 classes problem (apple, bay area no apple, and other) in a 2 classes problem (apple, not apple). I thik it is more interesting to keep the problem as a 3 classes problem. I suggest the authors to replace the appendi part which has no real interest and may easily replaced by a reference by the study of the 3 classes problem including confusion matrix.

It was now done as a 2 x 2 problem. And control now very well worked out in the paper. If a 1 x 3 problem is needed, probably random forest has to be rerun. That does not seem to worth the trouble. So I would keep it as a 2 x 2 design.

It will be interesting too to perfrom statistical test on data presented table 1 in order to see if one of these emotion is sufficent to classify da

This is also extra runs… how difficult is to run through the 7 emotions and see one by one if they are relevant?

My main fear is that it shows again that disgust is the most determining one and that is the one I trust the least. (nearly not prexent, very badly determined… not normally distributed, with a tail chopped off at the zero side)