How to Write a Machine Learning Paper for (not so) Dummies

Preprint · December 2019

DOI: 10.13140/RG.2.2.27708.18562

1 author:

Anselmo Ferreira
European Commission

How to Write a Machine Learning Paper for (not so) Dummies

Anselmo Ferreira[1]

[1]University of Cagliari, Department of Informatics and Mathematics. Via Ospedale 72, 09124, Cagliari, Sardinia - Italy

Abstract: Hello folks! I decided to write this paper to help you, someone who is like I was before, to finally learn how to write a machine learning paper. For this, I will report here all my knowledge acquired in about 10 years of research, with a best thesis and best journal paper awards on my back, reviewing machine learning papers for five journals and writing and researching for three different countries in three different continents. So lets start with the Abstract! the abstract is the most important part of your scientific paper, as it is the first text the readers will read about your work when they search for it in digital libraries. In the abstract, you summarize your work by talking about (i) the problem you want to solve; (ii) why it is important; (iii) how the others deal with the same problem; (iv) what are the limitations of other works; (v) how do you solve such limitations and what is the novelty of your work; and (vi) are the results promising? what did you learn from the experiments? please be aware that the abstract has a limitation of the number of words. You need to check it in the journal/conference website.

Here you put the keywords, which are words that best describe your work and will also be used to find your work in digital libraries.

Correspondence: ***A paper has a corresponding author, who is the person that will deal with all procedures related to submission and publication of the paper. Usually, the corresponding author is your advisor or the guy who has financial support to pay for the publication (if that is the case), or sometimes, this can also be the first author. Here you put the name and e-mail address of the corresponding author.***

1. Introduction. Hello world! I wrote this paper in order to define a "standard" to the papers written by my students and, this way, minimize rejection of papers that are considered bad written. So, I really hope it can help you too, specially if you are beginning now in the scientific life. My intention here is not to define a *silver bullet* approach that can work all the times and must be followed forever, but to make the reviewers job of rejecting papers (sic) as difficult as possible. So, with experience, you will find your own paper format, but I am sure that most of what is written here must remain in your paper. Please be aware that such steps I am presenting here are only for research papers that present a novel approach, so *Tutorials* and *Overview* papers should not follow the steps here. I chose a paper format that is a standard for a given journal style, but the rules I describe here can be applied to any journal or conference format. Please be aware that journals and conferences have limitations of the number of pages and number of references, so you don't want to have a paper rejected for such a stupid reason, correct?

Tip: be careful with the Introduction!!!!. Ok, now your reader/reviewer will start to know what you are really up to in your paper. The Introduction is, besides the Abstract, the most *dangerous* and *important* section of any paper. One simple English mistake here can cost you a fast rejection. So, be perfectionist in this section!!!

What to write in the Introduction. Your introduction is an extended version of the Abstract, where you can discuss with further details about the problem you want to solve and how you want to solve. So, be aware of the following discussions that MUST be present in the Introduction:

1. What is the problem you want to solve?
2. Why such a problem is important?
3. Is there any important impact for the community if solving such a problem? do the media discuss about its importance? it would be very nice to cite something from a news media company about this problem to make it closest to the reality of the reader/reviewer.
4. how does the literature touch such a problem? what are the solutions proposed so far? what are their limitations in terms of proposed solution, datasets and difficulty of the experiments?
5. how is your proposed approach supposed to defeat such limitations? what are you proposing? what is the novelty? how does it work? are the results promising? is there any difficulties in the experiments you are considering here but were not considered previously?

After discussing these points, it is highly recommendable to make it CLEAR, in a new paragraph, the contributions of the paper in a numbered list. For example, check the text below: "In summary, the contributions of this paper are"

1. Contribution one
2. Contribution two
3. *etc.*

About this last issue, I really like to use odd numbers (at least three) to do such a contribution list. Maybe I am superstitious about it (LOL).
So, after presenting the problem, confirming its importance, telling the limitations of the related works and showing your novelties, it is time to finish the Introduction by telling what is coming next in the paper. You can start it in the following way: "The remaining of this paper is organized as follows. Section II discusses BLABLABLA, Section III presents BLABLABLA and so on."

Related Work in the Introduction?. This is something I really don't like to do, but it happens in some papers I have been reviewing in my life. Some people do a very exhaustive related work discussion in the Introduction, which can make the sections of the paper unbalanced in terms of size (as a reviewer, I don't expect that the Introduction Section is bigger than the Proposed Method section, for example). So, my advice here is citing the related work in a batch depending on their subdivision (if the related work can be subdivided in different branches), telling about their limitations in general in a very brief manner. Then, in a separated section, you discuss about them with more details. Anyway, it is your option, but we aware of how big will be your Introduction and how unbalanced your paper will be in the end.

Identifying important features of a paper. Figures, tables, sections and equations are written with the first letter in capital or not, depending on different situations, as the examples below show:

- Figure/Table/Section/Equation 1 shows that...
- From this figure/table/section/equation we can see that...

Use of numbers in a paper. This is probably a myth, but one day somebody told me that, if I want to talk about a number that is less than 10, I must use letters to write about it. Otherwise, I use numbers. Look at the examples below.

- Five-fold cross validation
- 10-fold cross validation

Is that true? I am curious about it too. Check with your advisor about it and tell me what you found!

2. Background and Related Work (optional). I truly recommend you to write a section like this. Sometimes, the reviewers are PhD students who are using the opportunity to review to learn better the concepts of the area they want to act, or sometimes they are forced by their supervisors to do it for the journal they are editors (sic). Additionally, making your work as auto-contained as possible will maintain the attention of the reader to your paper, without stop reading it to find basic concepts anywhere. So, your acceptance chances for the reviewers and interest by the readers will be higher. It is also good for you, because now YOU WILL EXPLAIN THE PROBLEM with your own words, so you will learn better by teaching (BTW, teaching something is the best way of learning it, It's proven by science ;-)).
You can use both of them in the same section or divide them in two different sections (background comes first). I did both of these options for my papers.

Background. Here you will discuss about the basic concepts of the environment you are aimed to act. For example, in the financial market, there are a series of procedures that the Artificial Intelligence (AI) developer must know before applying an AI solution to it. For example, what are buy and

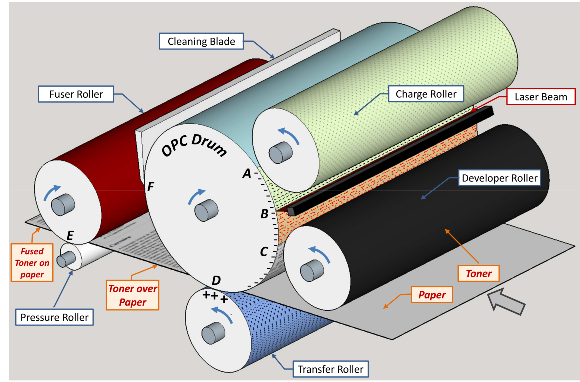

Fig. 1. This is a very nice picture a co-author of mine did to explain how laser printers work. It is based on a Wikipedia article we saw before. Please notice that this is not plagiarism, saying that is the same as Apple trying to patent rectangles because Apple's smartphones are rectangular (sic). Don't forget to explain, in a very succinct way, what is the picture you are showing about.

hold operations? what are long and short operations? how is the stocked market data available to the general public? what are time series? so, different environments contain different concepts that must be explained to the reader, so she/he can understand the environment you want to act. One very nice thing to do is, depending on the environment you want to talk about, putting a very nice picture to explain it.

About Pictures in a Scientific Article. Please, be super aware about using somebody else's pictures from another paper in your paper. These pictures are usually copyrighted by the publisher, so SOMETIMES you need to ask permission from the authors AND the publisher to use the figure in your paper. Another solution is YOU BUILDING THE PICTURE YOURSELF, based on the one you wanted to use in a given image editor. Be aware that this is not plagiarism, as a picture explaining an environment can be drawn the same way by different people. You just need to build it yourself in a given image editor and, of course, doing some modifications in order to don't make them the same picture. Take a look at Figure 1, where my co-author based his drawing on another one present in the literature.

What About Different Backgrounds?. Sometimes, besides the background about the application, you also have the background of the mathematical model you are using as a solution for the problem (*e.g.,* you want to talk about basic concepts of deep learning). I don't like to put such different backgrounds in the same section, but you can do it. When its the case for me, I usually talk about the basic concepts of the environment in the Introduction and the basic concepts of the solution in this section. Other option would be talking about the background of the environment in this section and the background of the solution in the Proposed Method Section, which shall be discussed later.

The Related Work. Ok, so here we start talking about solutions that acted in a problem similar to ours (or something similar if you are solving a new problem never tackled before). Please, pay attention to the fact that I have been witnessed lots of papers rejected by a very stupid reason, which is an outdated related work discussion. So, my algorithm to eliminate such a problem is:

- try to discuss something like 20 related works;
- consider the two last years, plus the current year;
- balance your references in conference and journal papers;
- consider old references only if they are *classical solutions*, with a super high number of citations; and
- use a diverse number of publishers with high reputation in your related works (*e.g.,* IEEE, Elsevier, ACM, Springer, among others).

Sometimes, in your review process, the reviewers will suggest you to consider SOME SPECIFIC ARTICLES EVEN IF THEY ARE OLD OR LESS CITED. If that happens, BE SURE THAT THE AUTHOR OF THE PAPER IS THE REVIEWER OF YOUR PAPER. So, do what they say and try to understand their approach, explain it as best as possible and try to be kind with the limitations of their approach (LOL). By the way, they work for free so they want their papers been cited and usually use the review process for that (#SadButTrue). The same can happen when sometimes the reviewers or editors ask you to include more references from their journal. Just do what they say and they will be happy (#SadButTrue2.0).

Abbreviations in Machine Learning Papers. If you are a machine learning enthusiast, you might heard about terms such as SVMs (Support Vector Machines), PCA (Principal Component Analysis) among others. However, how do you know that the reader of your paper is an enthusiast like you? so, I really don't like to see what I call *undeclared variables* in machine learning papers. I know that some terms become boring to repeat in the text, but every time you think about using an abbreviation, don't forget to declare what that abbreviation means before in your text and use it as you want. In My Humble Opinion (IMHO), this will grab the reader's attention to your text, and he/she will not need to look what that ****** abbreviation means anywhere. So, IMHO, keep your abbreviations declared in your text (you just need to do it once before you start using them).

The Use of Latin words and Identifying a List of Authors. Maybe you saw in your life the use of the Latin word *etc.*, but when you read related works papers you will find other words, such as:

1. *e.g.,* this is a Latin word acronym that means *exempli gratia*, which translates to, literally, "for example." Periods come after each letter and a comma normally follows unless the example is a single word and no pause is natural. I like to use it in parenthesis to give examples of something (*e.g.,* like this)
2. *i.e.,* this means *id est*, which translation to English is "that is". Loosely, "i.e." is used to mean "therefore" or "in other words." Periods come after each letter and a comma normally follows or no, depending on whether the wording following the abbreviation dictates a natural pause. I also like to use them in parenthesis (*i.e.,* to say the same thing in other words).
3. *et al.* from the Latin *et alii*, which literally means "and others". It must always be typed with a space between the two words, and with a period after the "l" (since the "al." is an abbreviation). A comma does not follow the abbreviation unless the sentence's grammar requires it.

Be aware that some journals italicize these terms because they come from Latin, but most do not. I usually put them in Italic.
About the last word *et al.*, it is used to identify a list of authors with three or more authors. So, one and two authors are usually identified by their last names in capital letters, and three and more authors are identified by the first author's surname followed by the *et al.* word. Check the examples below:

- RONALDO [1] proposed to celebrate his goals by spinning in the air.
- SIMON and GAFUNKEL [2] composed a very nice song about the sounds of silence.
- GIBB *et al.* [3], also known as the *Bee Gees*, wrote something about staying alive.

Please, be aware that sometimes the authors names are not cited in papers, just their reference. Check what is the case with the journal you want to submit.

How to search for related work. There are several tools from different publishers to help you finding papers related to the application you are working on. In the list that follows I will indicate some websites that were active at the time I was writing this paper:

1. Elsevier: https://www.sciencedirect.com/
2. IEEE: https://ieeexplore.ieee.org/Xplore/home.jsp
3. ACM: https://dl.acm.org/
4. Springer: https://link.springer.com/

Basically, you can use the same keywords as yours (do you remember them?) in these tools to find similar papers, or just putting a phrase that can be easily found on them. To describe the related work, here are my tips:

1. The Introduction and Abstract of a paper are the places most probable to help you understanding somebody's method. So, read them and write with your own words what you understood from them. What does that approach has that differentiates it from the others?
2. If you can divide the related works into branches, do it. For example, what are active and passive approaches? what are the proposed approaches for each of them?

3. Look for a limitation that you are exploring in your paper, but was not touched by the literature before. Its easy to find, even your approach has limitations! so what are the limitations of the literature solutions that will be faced by your approach?

I really like to finish the related works with this last issue. I can use it as a link to the next section, where I explain my approach and tell how I will deal with the literature limitations.

3. Proposed Method. If the reviewer got here without finding any problem in your paper, congratulations! you have now about 20% of chances of having your paper accepted! so, as I told you in the previous section, I like to finish the related work with limitations of the existing methods, and I start the proposed method section by telling HOW I AM TACKLING EXISTING SOLUTIONS LIMITATIONS. So, basically, I start this section by giving an overview of my approach, telling how it works in a general way.
Ok, pay attention to this. There are two ways of grabbing the reviewer's attention to your paper:

1. what the reviewer READS.
2. what the reviewer SEES.

Can you realize that both things are done with the eyes? which one do you think is the best? yes, me too, so, I will teach you several *tricks* to grab the reviewer's attention to your text without using regular text. I call them *attention tools* and they are the following:

1. Figures
2. Subsections
3. Equations
4. Algorithms

Let's check them one by one in the next subsections.

Figures. You know that a picture worth a thousand words, correct? so, what if you could motivate and explain your approach with beautiful pictures about the problem and your solution? reviewers and readers will surely like it. Although I understand that not all applications will allow us to do that (*e.g.,* it is very hard to understand the different stock market behaviors at the same time and put in just one graph), in some cases we can do that. Suppose that the problem you have has a Gaussian distribution, so you can plot such a Gaussian distribution to show the reviewer that your approach has a reason to be, and then you show a picture of how your approach will deal with such a problem. Nice, don't you think? for example, in Figure 1, I showed how a printer works. Now, I will show how a printer prints characters in order to motivate my texture descriptor developed in order to identify the source of a printer. Check Figure 2, it's a nice way to convince the reviewer that a printer can be identified before telling about my printer source attribution approach, don't you think?

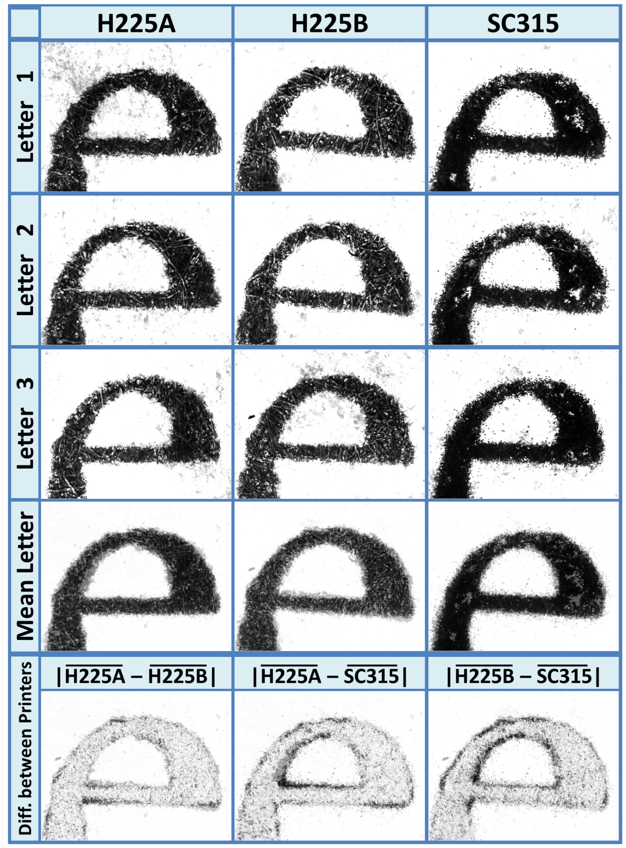

Fig. 2. Here is an example picture to motivate your approach. These are microscope pictures of letters printed by different printers. It can be seen from this figure that there are differences of textures between the same letter printed by these different printers. So, with this picture I motivate my solution without showing it to the reviewer.

Ok, now you have done so far the following steps: (i) identified the limitations of existing approaches; (ii) explained your approach in general terms; and (iii) motivated your solution. So, its time to finally presenting your approach to the reviewer. So, how to do that? WITH ANOTHER FIGURE, OF COURSE! I show you in Figure 3 a figure for another application I wrote a paper, check from the figure that it is possible to understand the approach just seeing that picture.

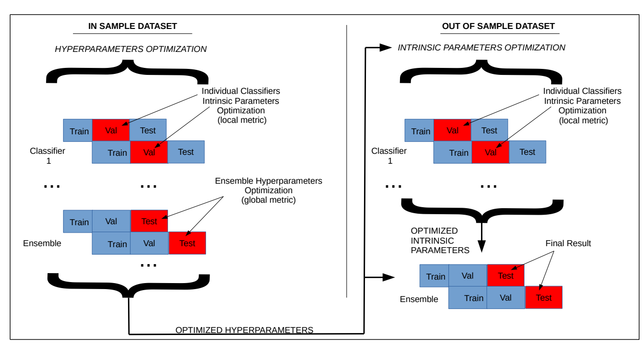

Fig. 3. A pipeline of how to generate auto-configurable ensembles to time series data classification. You can see that the approach has two steps, and different parameters are optimized in different kinds of data. This kind of figure is interesting, because it has several steps, so I can create subsections, algorithms, formulas and even more figures about each of them.

Ok, together with this figure you should also discuss, with

other words, how the approach works because, in the case the figure is not clear enough, you are discussing about it. You can also add more figures to show details of your approach, or some piece of your figure that requires more details. I usually do this in subsections of my papers.
So, in summary, the types of figures you can use at the beginning of your proposed method section are the following:

- Motivation Pictures: to show which kind of behavior have you noticed about the data from the problem you want to solve

- Approach (or pipeline) Pictures: will show how your approach works in general, with a given number of procedures dealing with the input and generating an output.

Don't forget to always explain these figures without specific details in the beginning of your method's section.
Now you can start to use the other *attention tools* to make your reviewer and reader interested in your approach. We will do this in the following.

Subsections. This is an awesome trick. If the reviewer has doubts about a piece of your pipeline picture he/she read at the beginning of the section, you are already making available further details of it in an easy to find section of your paper. I usually split pieces of my approach into subsections to detail how they work. Another solution I do is using one subsection to give basic concepts of the approach, or the mathematical foundation behind it. I usually do it in the *Background* section if the technique is very famous (*e.g., deep learning*) and use the *Proposed Method* Section to only show how I designed and dispose it in the proposed pipeline. Then, of course, I give details about the pipeline in other subsections.

Equations. You approach, even if its not totally new, has some mathematical foundation. So yes, you need to put equations in your text. There are lots of possibilities to do that: do you transform your data? which calculations do you do after pre-processing? is there any normalization of the output? how does your classifier work? so, answering all these questions will help you to find the right equations to your text.
Please be aware that all the Equations of your paper must have their symbols defined after they are presented. Look at one example to do that below:

$$g_e(x,y) = \exp\left[-\frac{1}{2}\left(\frac{(x\cos\theta + y\sin\theta)^2}{\sigma^2} + \right.\right. \tag{1}$$

$$\left.\left.\frac{(-x\sin\theta + y\cos\theta)^2}{(\gamma\sigma)^2}\right)\right] \times$$

$$\cos\left(\frac{2\pi}{\lambda}x\cos\theta + y\sin\theta\right), \tag{2}$$

where λ is the frequency of the sinusoid factor, θ is the orientation of the normal to the parallel stripes of the Gabor function, σ is the variance of the smooth curve (envelope) that outlines the extremes of the signal, and γ is the spatial aspect ratio which specifies the Gaussian ellipticity. Finally, parameters λ and σ specify the resolution of the descriptor. By varying the parameters λ and σ, the Gabor filters act considering multiple resolutions.
You can notice two things about the equation above: (i) it is part of the same paragraph that calls it; and (ii) it is finished with a comma or period, which will indicate if the following text is part of the same paragraph or not. Remember again: all symbols must be defined.
Although there is not a rule of thumb, I don't expect a machine learning paper with less than five equations (of course, that can be alleviated for conference papers, but try to follow this rule as much as you can).

Algorithms. Here we arrived to the *grand finale*, or the apotheosis of your proposed approach section, where you will finally give as much details as possible of your approach. Here you will really say how to transform your idea into code. Here, you will do what the scientific community in machine learning begs for researchers to do: MAKE YOUR WORK REPRODUCIBLE. I show you in Algorithm 1 one example of doing that.

Algorithm 1 Proposed hyperparameter search approach

```
Require:
  IS=time series from in sample data
  I= list of intra-parameters
  H=list of hyperparameters
  C=list of classifiers from the ensemble
Ensure:
  h'= Optimized hyperparameters
  procedure RETURN_HYPERPARAMETERS(IS, I, H, C)
      MAX_FINAL_METRIC ← 0
      ENS_METRIC ← 0
      for h in H do                    ▷ for each hyperparameter combination
          W[h] ← buildWalks(IS, h(window_size))   ▷ Starts non-anchored WF0
          for w in W[h] do                             ▷ for each walk
              F ← buildFeatures(w, h(lags))            ▷ get features
              F' ← icaTransform(h(ica_comp), F)        ▷ transform features
              MAX_WALK_METRIC ← 0
              for c in C do                            ▷ for each classifier
                  for i in I do   ▷ for each intrinsic parameter, train and validate
                      M[i] ← trainClassifier(F', h(train_size), c[i])
                      METRIC ← testClassifier(M[i], F'[h(train_size) * 0.3])
                      if METRIC > MAX_WALK_METRIC then
                          E[c, w] ← M[i]
                          MAX_WALK_METRIC ← METRIC
                      end if
                  end for
              end for
              test_data ← F'[h(window_size) − h(train_size) − h(train_size) * 0.3]
              ENS_METRIC ← ENS_METRIC + testClassifier(E[C, w], test_data)
          end for
          if ENS_METRIC > MAX_FINAL_METRIC then
              h' ← h
              MAX_FINAL_METRIC ← ENS_METRIC
          end if
      end for
      return h'
  end procedure
```

The Algorithm you show here is just a confirmation about what the reader or reviewer already knows about your approach from the previous *attention tools*, but now you discuss in totally specific details the approach. So, it is good to explain what the approach does line by line of the algorithm. To finish the proposed method section, a recommendable action to do is telling the *complexity* of the algorithm, in order also to tell the reviewer about your approach in terms of running time. Doing this, I believe this section is done.

4.Experimental Setup. I always like to tell about the information of the experiments in a specific section, even though I know that some people do this inside the *Experiments* section. Although its not mandatory in general, here are some topics I like to discuss in this section

- Datasets built, considering the same level or different levels of difficulty.
- Metrics used to assess the performance of the proposed approach.
- Methodology used for experiments (cross dataset, cross validation, *etc.*)
- Experimental scenarios, telling the difficulty of each experiment.
- Other approaches (baselines) proposed before in the literature that will compete with my proposed approach in the experiments. I usually start declaring each competitor with an abbreviation (*e.g.*, Principal Component Analysis, or PCA). **DON'T IMPLEMENT THEM YOURSELF! Send a message to the authors of the machine learning paper you are interested, and ask for the code.** I witnessed a case that an author implemented a baseline method via reverse engineering and got a paper rejected because, according to the reviewer's opinion, the way he coded was wrong. This is totally subjective, so avoid this kind of problem.
- Statistical tests performed, in order to declare that my approach is not winning by luck.
- Implementation aspects of the approaches considered for the experiments, including the programming language used, hardware, parameters of the classifiers, *etc.*

5.Experiments. Here we are, we will finally convince the reviewer to accept our paper in this section. Here, you will report the experiments done in order to prove that your approach deserves respect. However, I've been noticing that most of the papers just focus on showing that their approach simply perform their tasks better than the others, and let me tell you, from my experience this is not enough. There are a series of diverse experiments to be done in order to make the reviewer better convinced about the benefits of your approach. I will discuss them in the following:

Set of experiments #1: Preliminary Experiments. This part is usually fused with the *Proposed Method* section, but I really like to do it in this section in order to make the reviewer curious before and surprised later. The preliminary experiments are experiments related to your approach only, and are used to show the reviewer that what you are proposing really works. Some kinds of preliminary experiments are:

1. **Varying parameters of the proposed approach:** suppose that one part of your approach deals with a different input than everything that was done before. So, you do an experiment (you can use training and validation data for that) to show the difference of performances your approach gets with these different inputs. I show in Table 1 one preliminary experiment I've done that proves that given CNNs architectures work better with pre-specified inputs for remote sensing image classification. You can also do here an experiment to tell what happens with and without your proposed feature selection, feature transform or pre-processed approach.
2. **Importance of features:** you can use a Random Forest classifier in order to tell the importance of features, plot it and discuss about it. Check Figure 4 I did to identify the best features in my proposed feature set to detect blurred images.
3. **T-SNE plot:** you can use the T-SNE approach to plot your features and also from the competitors in different subfigures. Then, you can discuss how good is your approach to generate features that are easy separable by a machine learning classifier. I show in Figure 5 how I did that to compare my approach with another.
4. **Hyper-plane Plot**: this one I have seen in several machine learning tutorials and I am sure this would be a very nice thing to do in a paper (I never done that before). If you show that your data is very separable by the separating hyper-plane of a classifier like an SVM, the reviewer will certainly like it.

Rank	INPUT	Metrics calculated considering PROPOSED_SUBMODEL1 after classifying bag_2 images			
		F	NACC (%)	TPR (%)	FPR (%)
1	NEAR INFRARED CHANNEL	0.47	84.10	94.26	26.06
2	FALSE COLOR	0.34	74.52	92.90	43.86
3	GREEN CHANNEL	0.29	67.00	81.91	47.90
4	RED CHANNEL	0.24	60.47	96.89	75.15
5	BLUE CHANNEL	0.20	50.02	100.00	99.96

Table 1. This is a preliminary experiment I did to show which is the best input for my CNN approach to be applied in a specific kind of remote sensing image.

Main Experiments. NOW IT'S TIME TO SHINE! here you will report the results considering your approach and the competitors from the literature in a real-world application. Basically, you can use the following tools to do that:

- Tables, to report metrics of your approach and literature solutions, ranked by a given metric. I like to build tables in Excel and convert the to pdf, using figures as tables. However, as I like to put citations by the side of the literature approaches abbreviations on that table, sometimes I need to change the table if the references change, so, choose what is the best for you. Look at my example in Table 2.

Rank	Method	Statistics Calculated on CMEN Dataset after 5X2 Cross-Validation Experiments				
		F-MEASURE (%)	ACC (%)	TPR (%)	FPR (%)	PRECISION (%)
1	***MULTISCALE BKS-RF- LVT***	**89.96**	**90.76 ± 14.36**	**82.95 ± 28.41**	**1.43 ± 3.15**	**98.31**
2	***BKS-RF-LVT***	**88.72**	**89.76 ± 14.51**	**81.10 ± 28.73**	**1.57 ± 2.85**	**98.10**
3	***MULTISCALE BKS-SVR-LVT***	**87.12**	**88.52 ± 15.67**	**78.03 ± 31.27**	0.98 ± 2.27	**100.00**
4	***BKS-SVR- OTSU***	**86.49**	**87.95 ± 13.06**	**77.35 ± 25.38**	**1.44 ± 3.34**	**98.17**
5	***BKS-SVR- LVT***	**83.64**	**85.91 ± 16.28**	**72.55 ± 32.55**	0.72 ± 1.60	**100.00**
6	*SURF [33]*	80.40	83.50 ± 19.97	67.89 ± 39.41	0.87 ± 2.42	**100.00**
7	*BKS [5]*	79.52	82.76 ± 17.57	66.00 ± 35.18	0.48 ± 1.32	**100.00**
8	*SIFT [33]*	75.49	80.03 ± 21.30	60.63 ± 42.96	0.56 ± 1.19	**100.00**
9	*Multiscale Voting [34]*	74.56	78.33 ± 17.77	59.45 ± 37.24	**2.79 ± 5.23**	**99.97**
10	*THRESHOLD VOTING (T=4)*	68.15	75.73 ± 19.42	51.69 ± 38.82	0.22 ± 1.01	100.00
11	*Zernike2 [48]*	66.35	73.98 ± 20.74	49.65 ± 40.36	**1.69 ± 3.30**	**99.98**
12	*Zernike [14]*	62.38	72.56 ± 16.79	45.33 ± 33.39	**0.20 ± 1.35**	100.00
13	*DCT [6]*	**54.22**	**68.53 ± 16.79**	**37.19 ± 33.55**	**0.13 ± 0.53**	100.00
14	*KPCA [13]*	**48.51**	**65.93 ± 15.66**	**32.02 ± 31.32**	**0.14 ± 0.77**	100.00
15	*THRESHOLD VOTING (T=6)*	**42.48**	**63.46 ± 17.02**	**26.97 ± 34.00**	**0.04 ± 0.43**	100.00
16	*Hierarch-SIFT [31]*	**39.51**	**61.99 ± 17.72**	**24.62 ± 35.70**	0.64 ± 2.56	100.00
17	*BAYESIAN FUSION [4]*	**8.48**	**52.20 ± 2.01**	**4.43 ± 4.02**	**0.03 ± 0.16**	100.00

Legend:
xx.xx = Five best methods in the column metric
xx.xx = Five worst methods in the column metric

Table 2. Example of Table of results I did in Excel and converted to pdf, using it as an image in my paper. I usually consider abbreviations to call my approach and the competitors and I leave my approaches in bold. I also highlight the best metrics in bold and use different colors to show the top-5 or bottom-5 metrics.

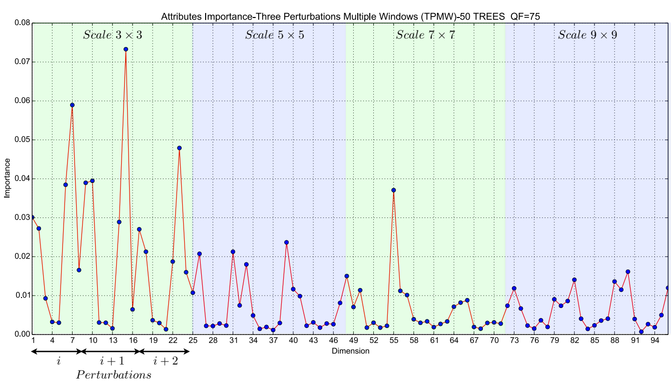

Fig. 4. I use in this figure the Random Forests output that shows the importance of each feature. As different parts of my proposed feature set are created by different steps of the algorithm, I can say what are the steps that are more important in my approach.

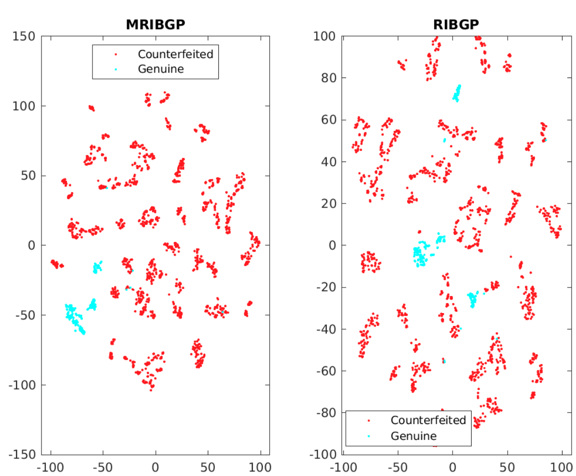

Fig. 5. I used T-SNE approach to show that my feature set (left) generates two classes of features (red and blue) in the N-dimensional space that are in clusters far away from each other. The same does not happen with the baseline approach (right). This facilitates the job of binary classification techniques if applied to my proposed approach.

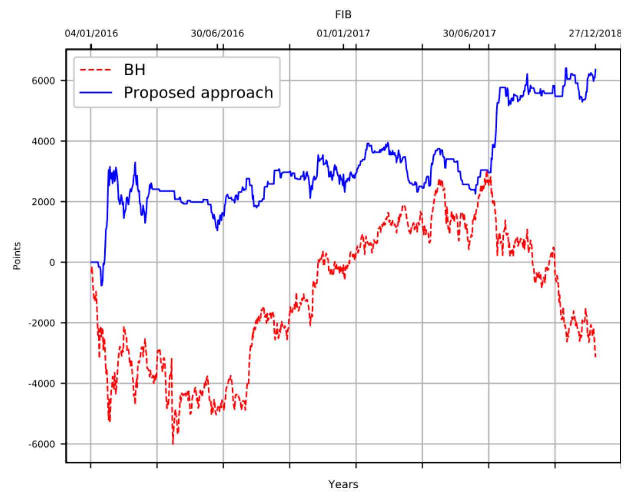

Fig. 6. This is an Equity Curve, or Profit and Loss curve done by my co-author. This curve is used to compare how much money our proposed machine approach earns through time in the stock market, compared with a baseline approach (BH).

- Figures, such as the Receiver Operational Curve and many others. Look at Figure 6 to see one example.

In this part of experiments, different difficulties can be considered in different subsections (for example, results of the approach with and without attacks). Its a good practice to also show samples that were misclassified by the baselines, but classified perfectly by your approach and explain why. I do this as an example in Figure 7.

Statistical Tests. In this kind of experiments, you use the chosen statistical test that you discussed in the *Experimental Setup* section, do you remember? then, you discuss here the results of such a test. I usually put a beautiful table here, to show if my approach wins the others in terms of confidence level and sum up the scores, making a kind of leader board in the end. If you decide to do this and is performing several experiments, I suggest you to do this for all experiments OR the most important and difficult one OR for the experiments that your approach wins with small metrics differences. Take

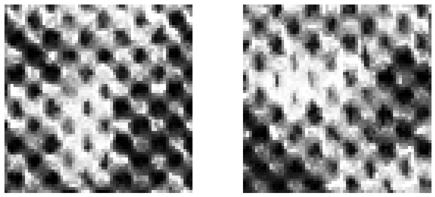

Fig. 7. Some 64×64 patches representing regions from two different genuine 2D barcodes misclassified by the baseline approach, but correctly classified by the proposed approach. The difficulty of such a problem is highlighted by the fact that the halftones in the genuine class can assume multiple sizes, without any fixed pattern, which can be confused with irregular and additional edges from counterfeited barcodes if the descriptor used is not scale invariant.

Rank	Method	PROPOSED_FUSION	PROPOSED_SUBMODEL2	TMH3 [66]	GLCM-MD [60]	VGG-19 [16]	TOTAL
1	***PROPOSED_FUSION***	0	1	1	1	1	4
2	***PROPOSED_SUBMODEL2***	-1	0	1	1	1	2
3	*TMH3 [66]*	-1	-1	0	1	1	0
4	*GLCM-MD [60]*	-1	-1	-1	0	1	-2
5	*VGG-19 [16]*	-1	-1	-1	-1	0	-4

1 = Line method is better than column method
0 = Line method is equivalent to column method
-1 = Line method is worse than column method

Table 3. A statistical test table that show who wins, who loses and if there is any statistical tie between each pair of approaches I considered for remote sensing image classification.

a look at Table 3 to see one example. Don't forget to inform the p-value or specific parameters of the statistical tests calculated.

Running Time Experiments. This is the last and less important experiment to be done in my opinion. With the technology being evolved every time and with the Moore's law, what is considered slow today can be fast tomorrow. But anyway, reviewers usually like this kind of experiment.
So here, you calculate the running time of your approach and compare against the baselines. You can use it for training step, testing step or both. Of course, its nice to discuss why your approach is slower or faster, but I don't believe that your paper will be rejected because precision is what matters in most of the applications and as I said before, what is slow today can be accelerated tomorrow according to Moore's law. Take a look at Table 4, where I do such a running time comparison that shows how slow is my approach.

6. Conclusion. Wow! its finally done! we finished our six-sections paper and now all we need to do is conclude our work! although there is not a rule of thumb of concluding something, here is a step-by-step approach I do in this sec-

APPROACH	Mean Running Time Per Image (s)
SIFT [33]	2.43
ZERNIKE [14]	42.08
ZERNIKE2 [48]	2025.85
BKS [5]	2893.75
MULTISCALE BKS-RF-LVT	2955.91

Table 4. Running time of my proposed approach (in bold) and competitors. Please notice that my approach is composed of running eight approaches in a sequence that could be run in parallel. So, such a result can be misleading.

tion:

1. I motivate the problem and discuss existing solutions limitations again, but in a very short paragraph.

2. I summarize what I am proposing in the paper and discuss what I discovered from experiments results, difficulties and even discuss drawbacks of my approach. Additionally, I let it clear what are the contributions my paper is giving to new researchers of the area

3. I discuss about future work that can be inspired from my research reported in the paper, such as modifying other parameters, trying different classifiers, making the datasets more difficult and even studying the possibility of transferring my solution to another problem.

That's all folks! my intention here is, again, to try to help beginners to learn an initial format for their papers. I truly believe that other formats can be considered, but similar ideas can be found in this tutorial. Additionally, I believe that other research areas can also use some of my concepts here to help them. Finally, the drawback of this tutorial is that it is not recommended for other paper formats, such as tutorials and overview papers, which follow other specific rules.
Did you like it? or don't you? is there something missing? did you find typos or grammar mistakes? anyway, I am hungry for feedback and for new ideas to evolve this tutorial and help other machine learning researchers like you. Can you help me on this? send me a message to anselmo.ferreira@gmail.com and let's talk about it!
All the figures used in this paper are from my published works or works under review, so, I am not citing them here. You can find my publications at http://www.ic.unicamp.br/ anselmoferreira. You can also find me on ResearchGate, GitHub, Google, Scopus and Web of Science.

ACKNOWLEDGEMENTS
Here you will thank everybody that helped in your work. This can be (i) somebody that helped you in collecting datasets; (ii) somebody who sent to you the source code of their approach; (iii) somebody who helped you to make English proofread; and more importantly (iv) the funding institutions that supported financially your research. Don't forget to inform the grant number of such a support.

Bibliography

Different journals use different bibliography styles. Check with your journal of interest about it.