Supplementary Material for ”Machine learning from crowds using candidate set-based labelling”

Iker Beñaran-Muñoz
Basque Center for Applied Mathematics, Bilbao, Spain

Jerónimo Hernández-González
University of Barcelona, Mathematics and Computer Science Dpt., Barcelona, Spain

Aritz Pérez
Basque Center for Applied Mathematics, Bilbao, Spain

X	Descriptive variable
d	Dimension of the descriptive variable
x	Instance of X
Ω_X	Feature space
C	Class variable
c	Class label
Ω_C	Set of possible class labels
A	Set of available annotators
a	Annotator from A
l_x^a	Label provided by annotator a for instance x (full labelling context)
L_x^a	Candidate set provided by annotator a for instance x (candidate labelling contex
$\mathcal{L}_x$	Labelling for instance x
$\mathcal{L}$	Set of labellings for the whole training set
w_x	Candidate voting estimate
ω	Candidate voting function
α_{ck}^a	(Parameter) Probability that annotator a includes class label k in the candidate s class c
$\boldsymbol{\alpha}$	Set of all α_{ck}^a parameters
$q_{\hat{\boldsymbol{\alpha}}}(c\|x)$	Estimate of the probability that instance x belongs to class c, based on the set SL-C method)
h	Probabilistic classifier
θ	Parameter set of the probabilistic classifier
$q_{\hat{\boldsymbol{\alpha}},\hat{\theta}}(c\|x)$	Estimate of the probability that instance x belongs to class c, based on the set of JL-C method)
n	Total number of instances
r	Number of classes
m	Number of annotators
β	Label generation parameter that represents annotator expertise
$prop$	Proportion of sampled labels over the possible class labels when generating cand
g_a	Probability distribution that represents annotator a in the label generation proces
$\mathcal{C}_x$	Set of partial labels associated to instance x

Table 1. Notation used in the work.

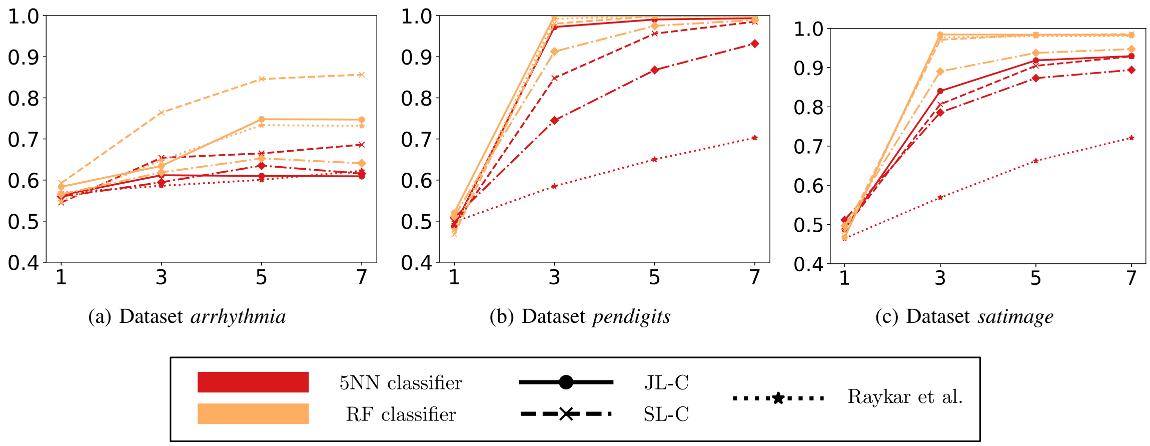


Figure 1. Experimental results throughout different values of the parameter β (annotator expertise), in terms of AUC metric, within different datasets (subplots). Results with classifiers RF and 5NN are displayed in orange and red colors, respectively. A different line style and marker is used for each method (SL-C, JL-C, RAY , DS). The rest of generative parameters are fixed to $m = 5$ and $prop = 0.5$.

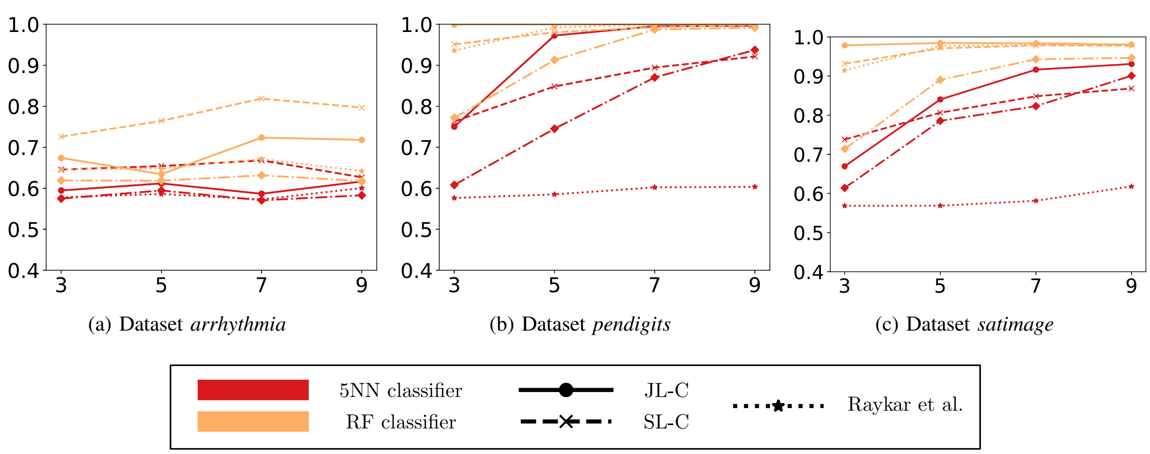


Figure 2. Experimental results throughout different values of the parameter m (number of annotators), in terms of AUC metric, within different datasets (subplots). Results with classifiers RF and 5NN are displayed in orange and red colors, respectively. A different line style and marker is used for each method (SL-C, JL-C, RAY , DS). The rest of generative parameters are fixed to $\beta = 3$ and $prop = 0.5$.

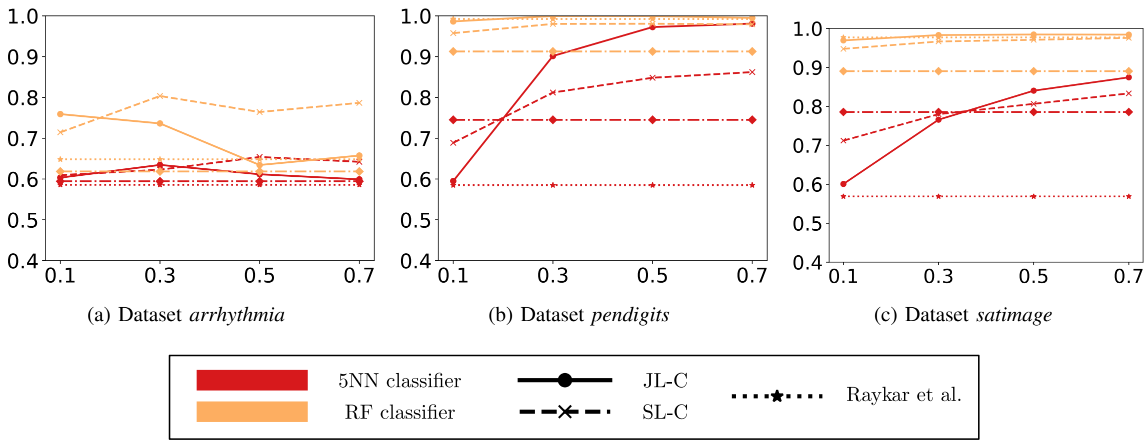
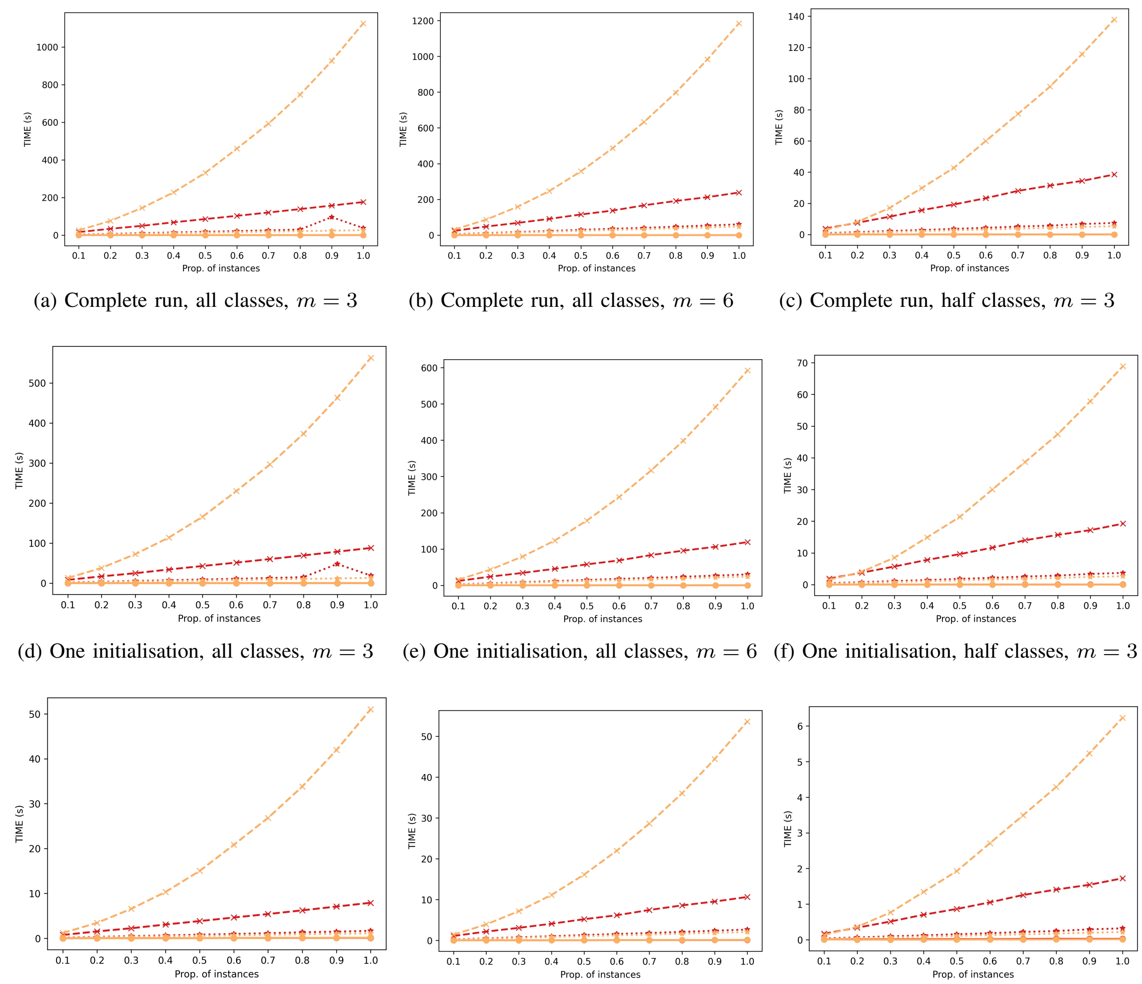


Figure 3. Experimental results throughout different values of the parameter $prop$ (flexibility of the annotators), in terms of AUC metric, within different datasets (subplots). Results with classifiers RF and 5NN are displayed in orange and red colours, respectively. A different line style and marker is used for each method (SL-C, JL-C, RAY , DS). The rest of generative parameters are fixed to $\beta = 3$ and $m = 5$.



(a) Complete run, all classes, $m = 3$ (b) Complete run, all classes, $m = 6$ (c) Complete run, half classes, $m = 3$

(d) One initialisation, all classes, $m = 3$ (e) One initialisation, all classes, $m = 6$ (f) One initialisation, half classes, $m = 3$

(g) One iteration, all classes, $m = 3$ (h) One iteration, all classes, $m = 6$ (i) One iteration, half classes, $m = 3$

Figure 4. Scalability test.