Detecting and Learning from Unknown by Extremely Weak Supervision: eXploratory Classifier (xClass)

Plamen Angelov
Lancaster University

Eduardo Soares (e.almeidasoares@lancaster.ac.uk)
Lancaster University https://orcid.org/0000-0002-2634-8270

Research Article

Keywords: Extremely weak supervision, Novelty detection, Interpretability, Autonomously self-learning model

Posted Date: April 12th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-393581/v1

Abstract

In this paper, we break with the traditional approach to classification, which is regarded as a form of supervised learning. We offer a method and algorithm, which make possible fully autonomous (unsupervised) detection of new classes, and learning following a very parsimonious training priming (few labeled data samples only). Moreover, new unknown classes may appear at a later stage and the proposed xClass method and algorithm are able to successfully discover this and learn from the data autonomously. Furthermore, the features (inputs to the classifier) are automatically sub-selected by the algorithm based on the accumulated data density per feature per class. In addition, the automatically generated model is easy to interpret, is locally generative and based on prototypes which define the modes of the data distribution. As a result, a highly efficient, lean, human-understandable, autonomously self-learning model (which only needs an extremely parsimonious priming) emerges from the data. To validate our proposal we approbated it on four challenging problems, including imbalanced Faces-1999 data base, Caltech-101 data set, vehicles dataset, and iRoads dataset, which is a dataset of images of autonomous driving scenarios. Not only we achieved higher precision (in one of the problems outperforming by 25% all other methods), but, more significantly, we only used a single class beforehand, while other methods used all the available classes) and we generated interpretable models with smaller number of features used, through extremely weak and weak supervision. We demonstrated the ability to detect and learn new classes for both, images and numerical examples.

Full Text

This preprint is available for download as a PDF.

Figures

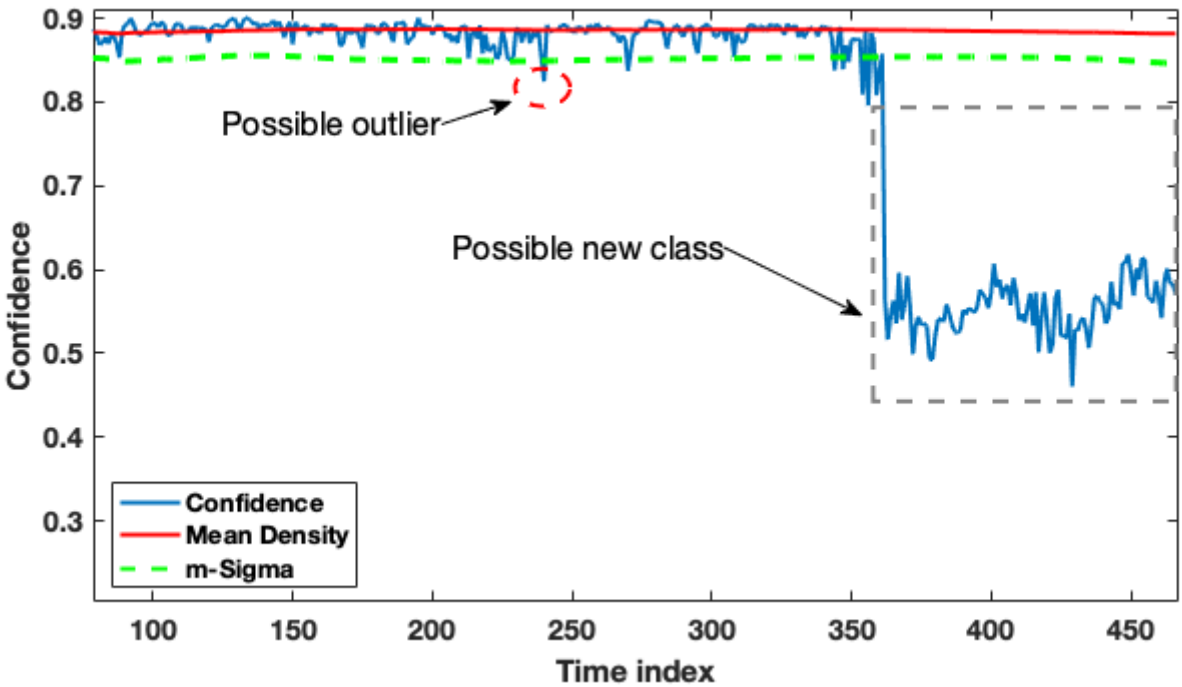

Figure 1

Drop of confidence of the proposed method when a new a unseen class arrives

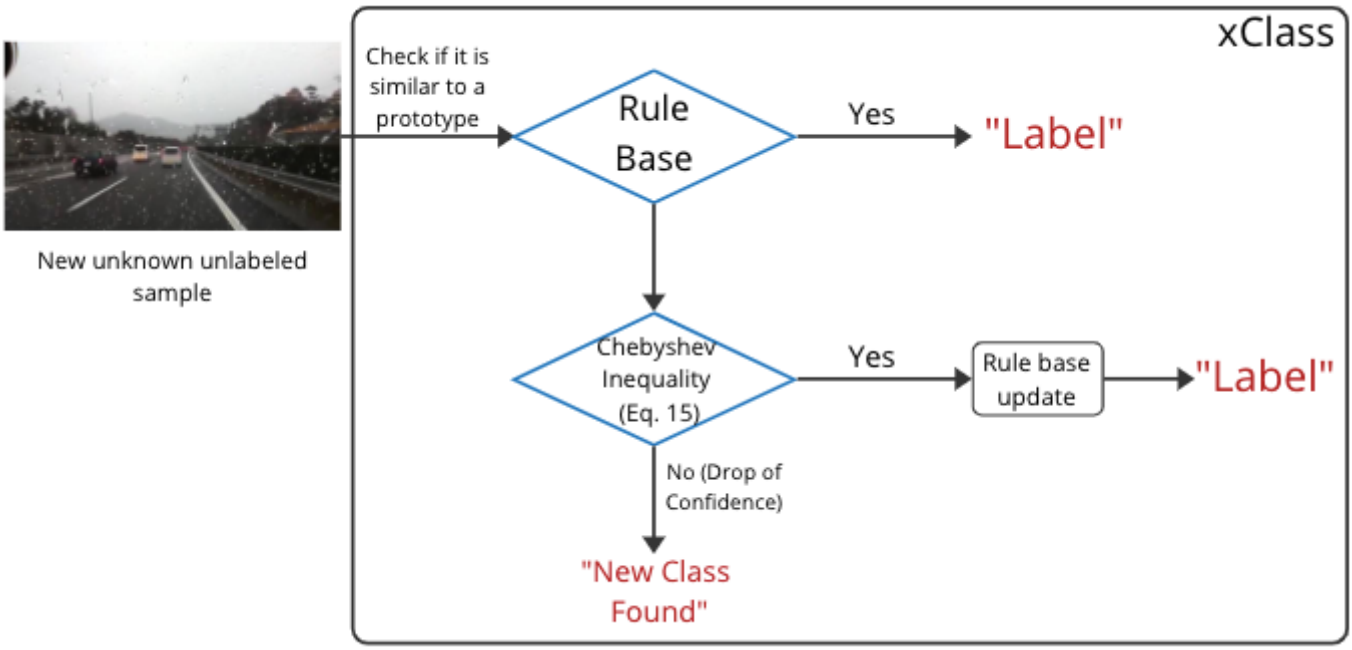

Figure 2

General structure of xClass – block diagram

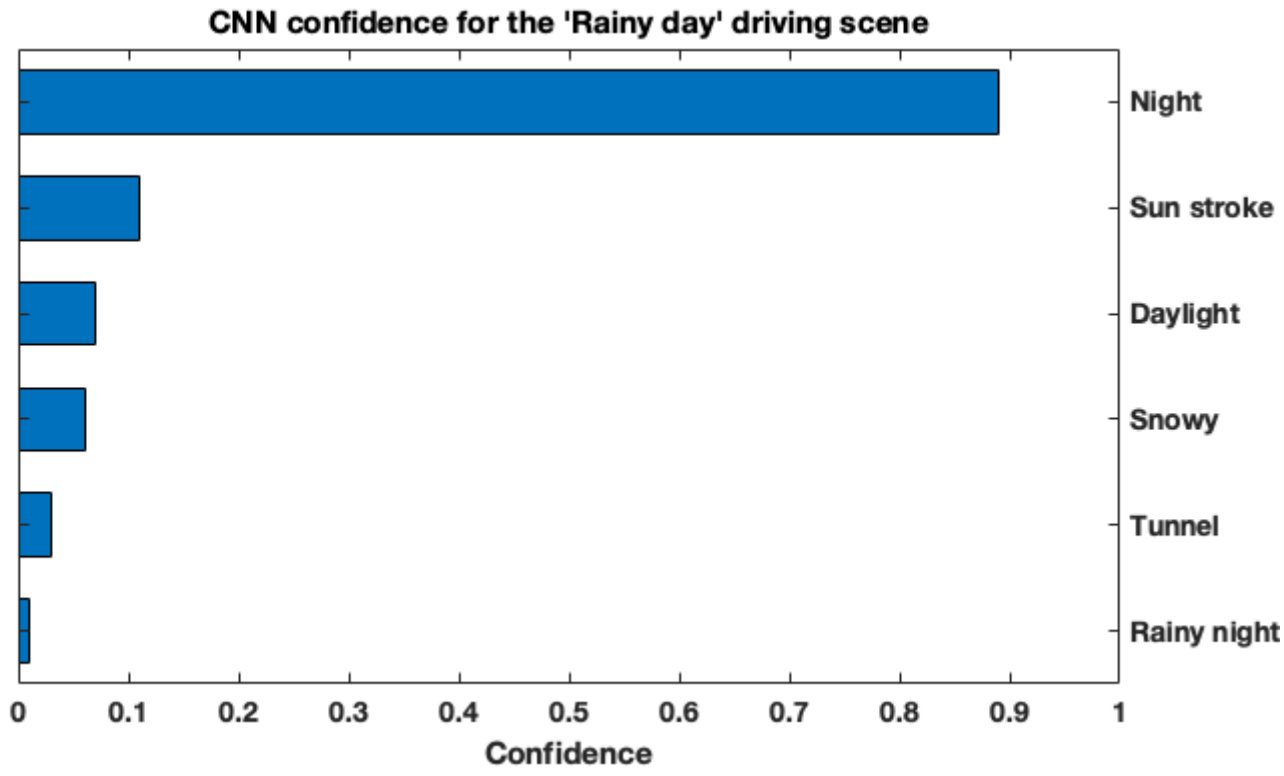

Figure 3

Wrong classification given by VGG−16 for a new unknown class (Rainy Day).

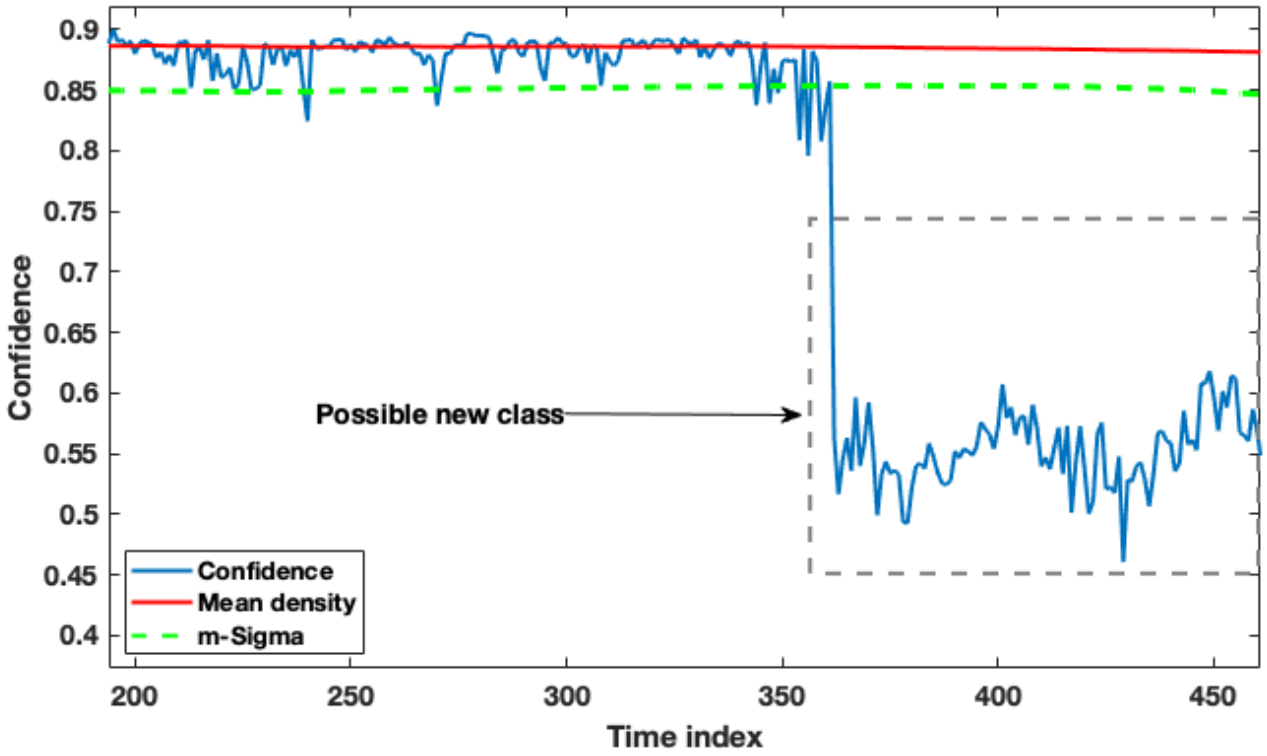

Figure 4

Sudden drop of confidence due the presentation of new unknown classes.

R_{new}: IF $(Image \sim$  $)$ THEN 'New class'

Figure 5

A new rule is proactively created when a sudden fall in the confidence is detected through the inequality (13). The proposed xClass classifier is highly interpretable due to its rule-based nature. This advantage favors human experts analysis as it provides a transparent structure, differently from the 'black box' approaches such as deep neural networks.

R_7: IF $(Image \sim$  $)$ OR $(Image \sim$  $)$ OR

... OR $(Image \sim$  $)$ THEN 'Rainy day scene'

Figure 6

Final rule given by the xClass classifer for the new detected class. Label is attached during the validation phase. Differently from ‘black box’ approaches as deep neural networks, xClass provides highly interpretable rules which can be used by human experts for different analysis as necessary.

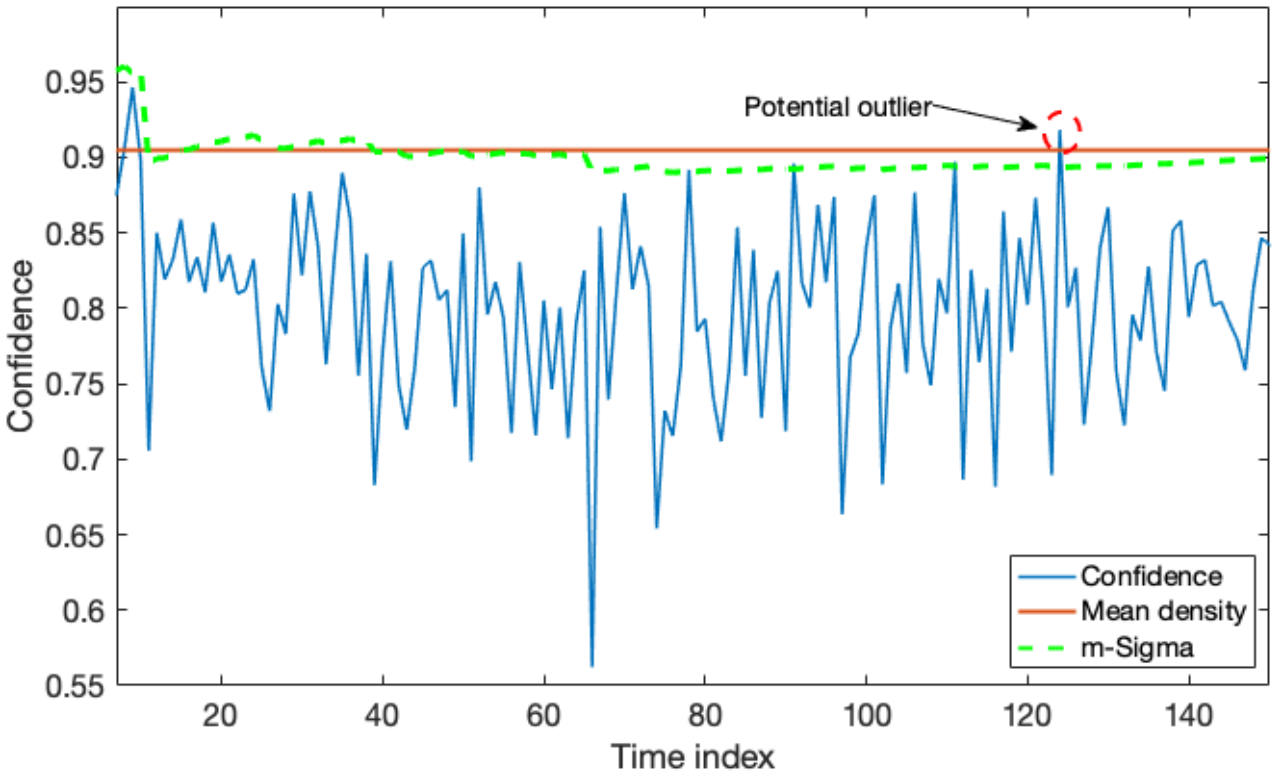

Figure 7

Sudden drop of confidence due the presentation of new unknown classes for the Faces-1999 dataset.

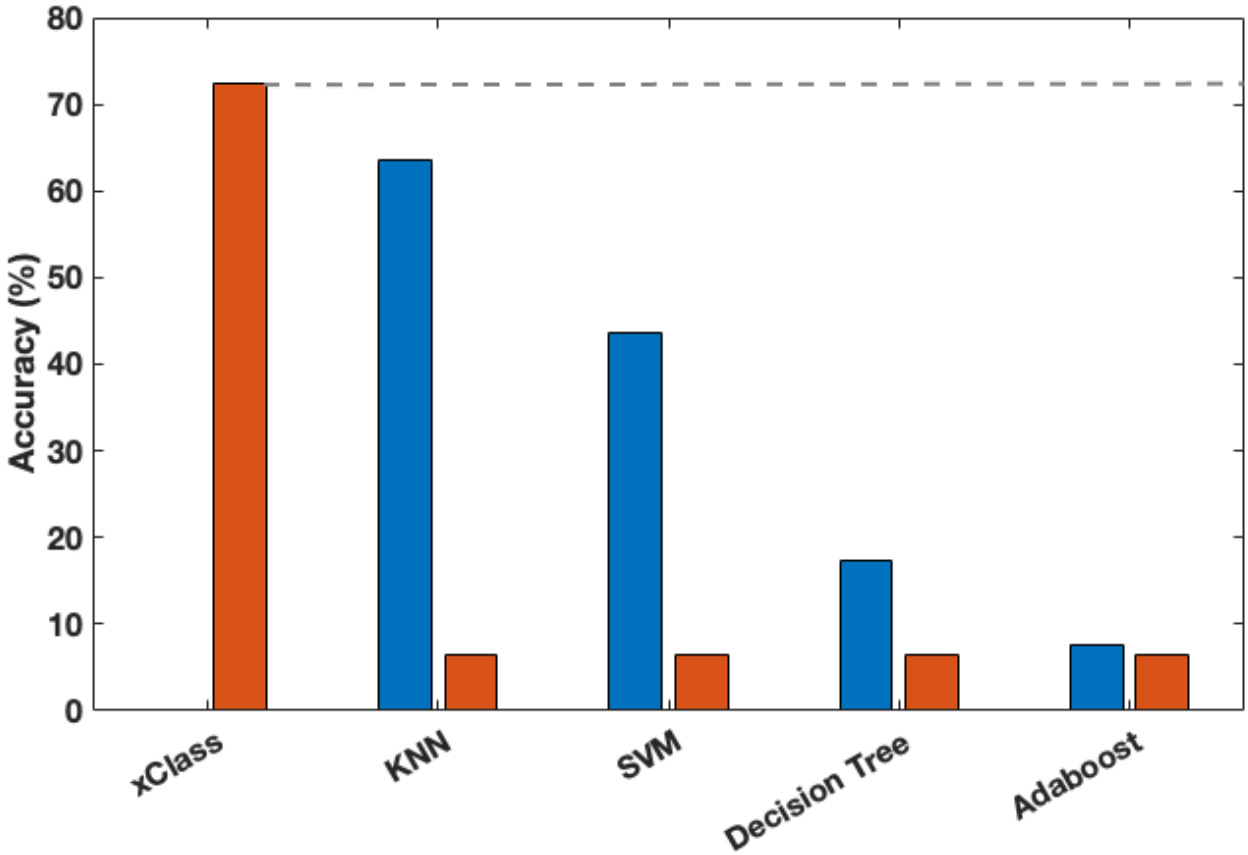

Figure 8

Accuracy for extremely weak supervision classification for the Faces-1999 dataset. red bars illustrate the results obtained by state-of-the-art approaches when just one class is provided during the training phase. The blue bars indicate the results when all the classes are provided.

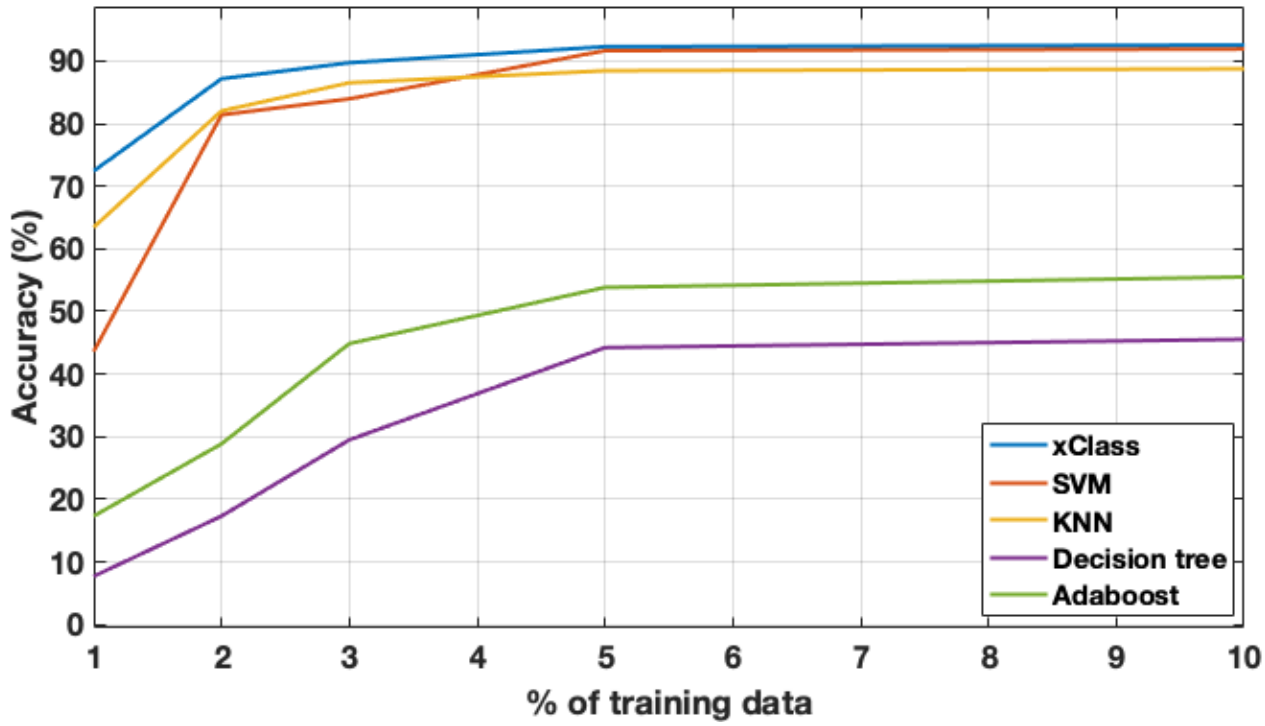

Figure 9

Classification curve for different number of training samples for the Faces-1999 dataset.

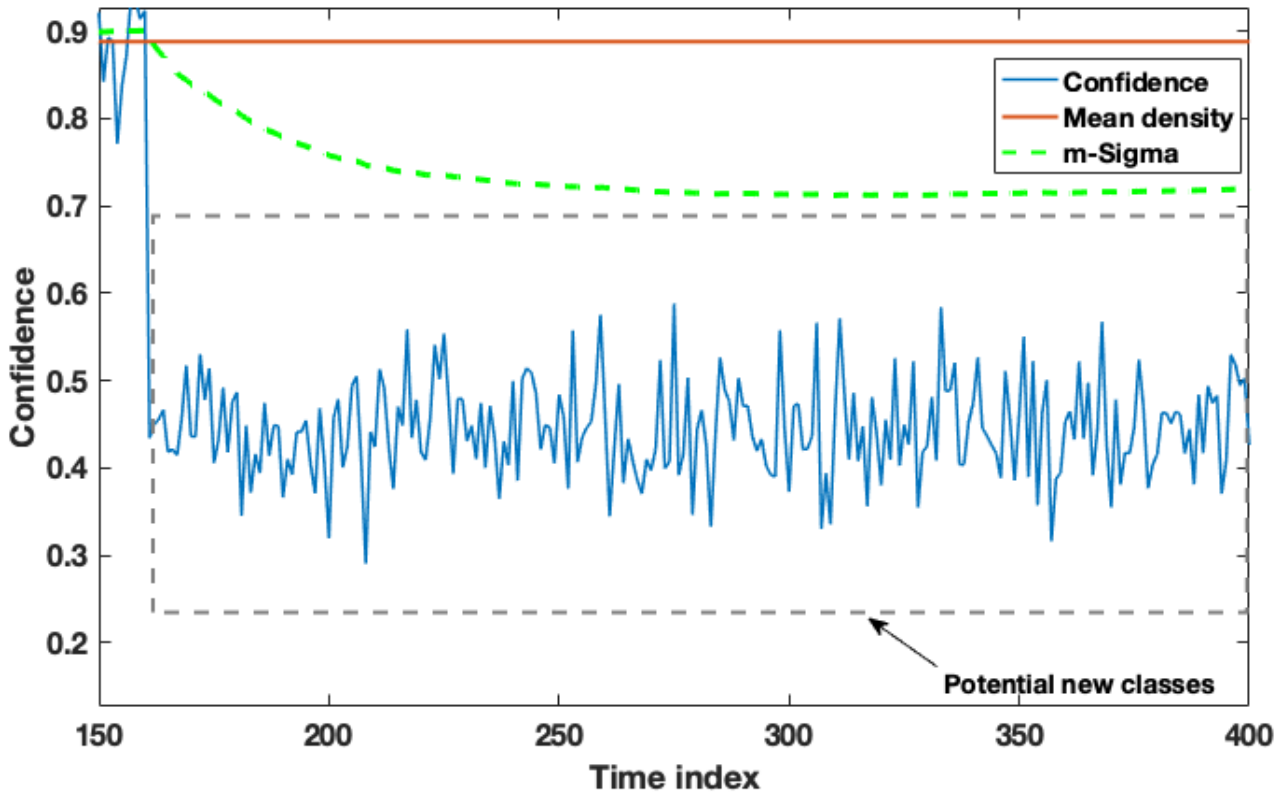

Figure 10

Sudden drop when new unknown are classes are presented to the xClass method – Caltech-101 dataset.

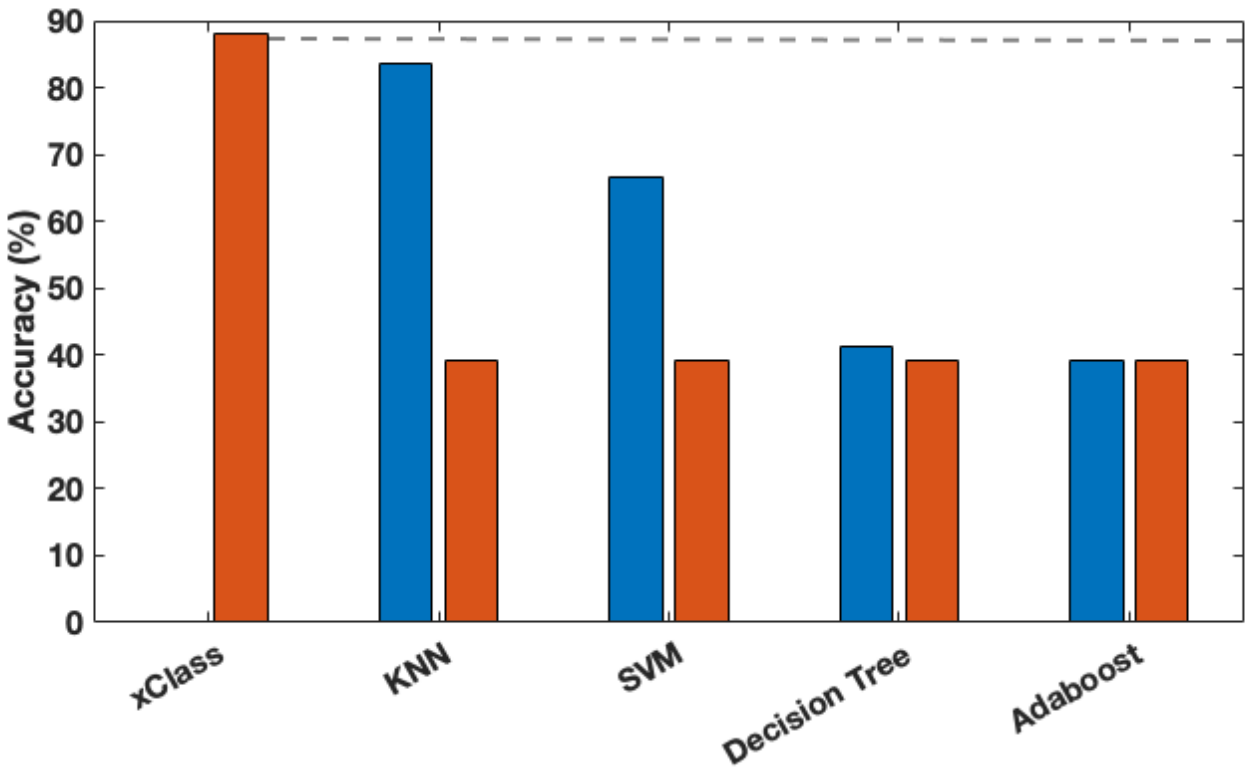

Figure 11

Accuracy for extremely weak supervision classification for the Caltech-101 dataset.

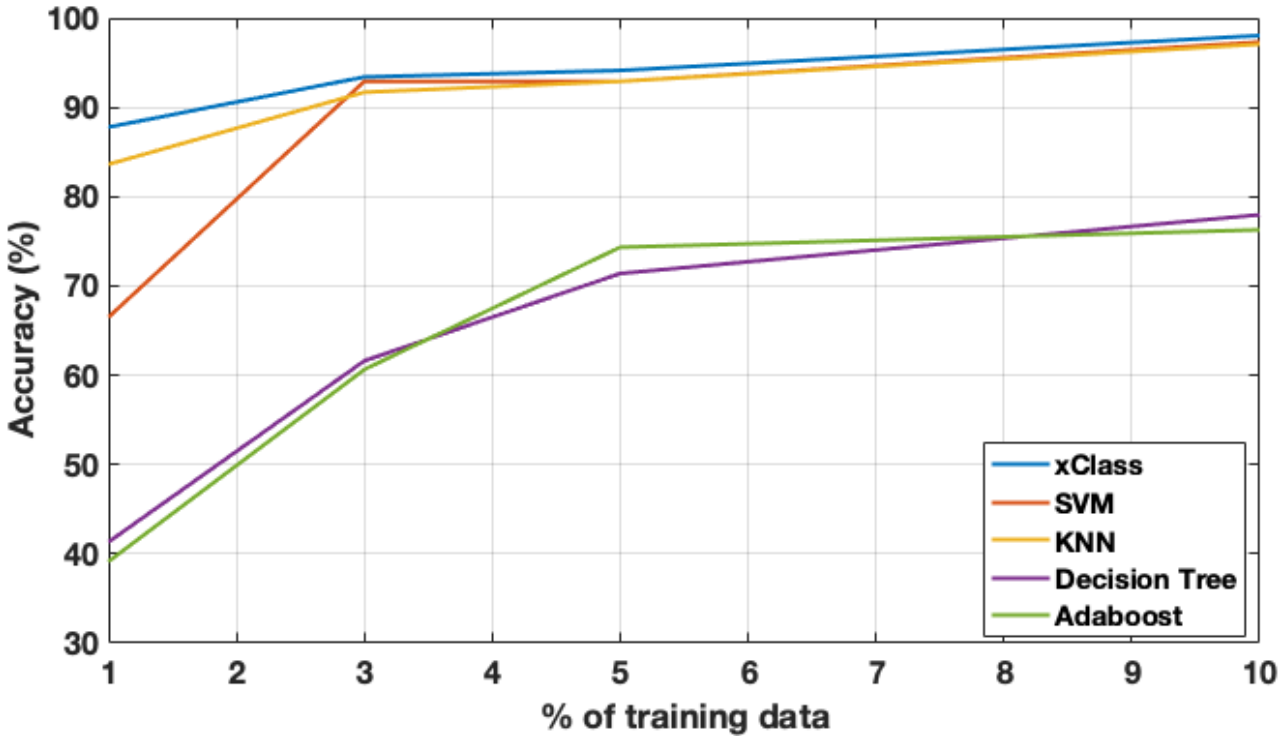

Figure 12

Classification curve for different number of training samples for the Caltech-101 dataset.

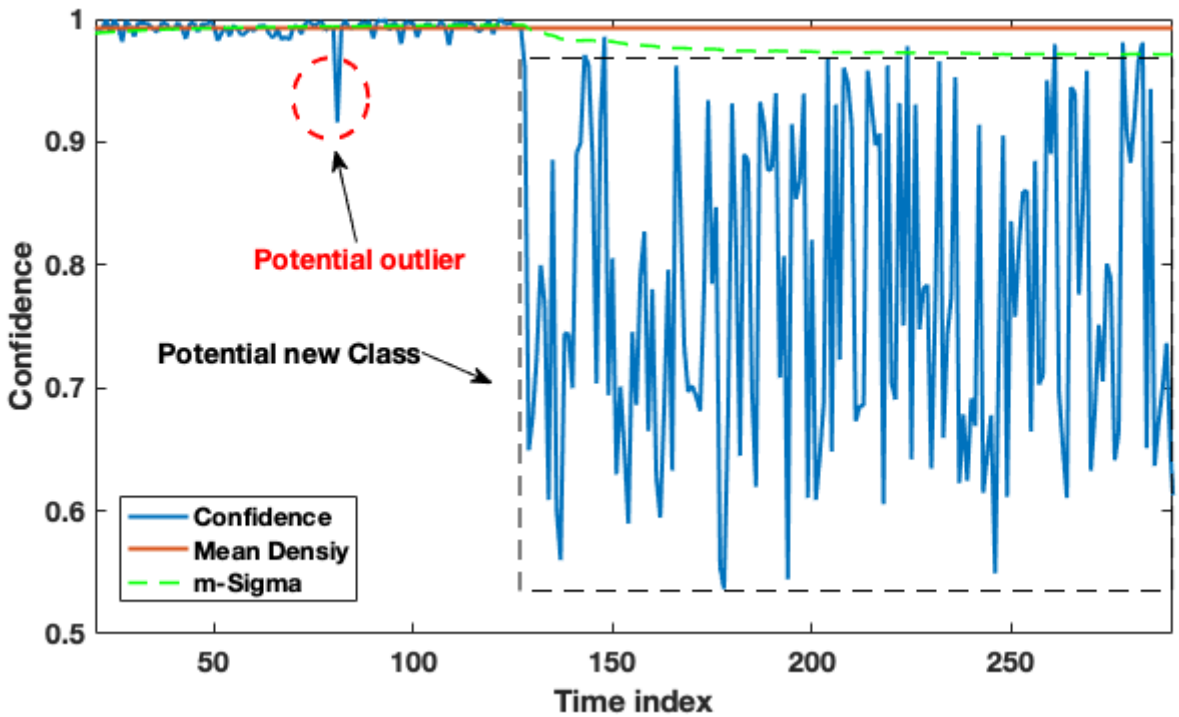

Figure 13

Sudden drop of confidence due the presentation of new unknown classes – Cars dataset.

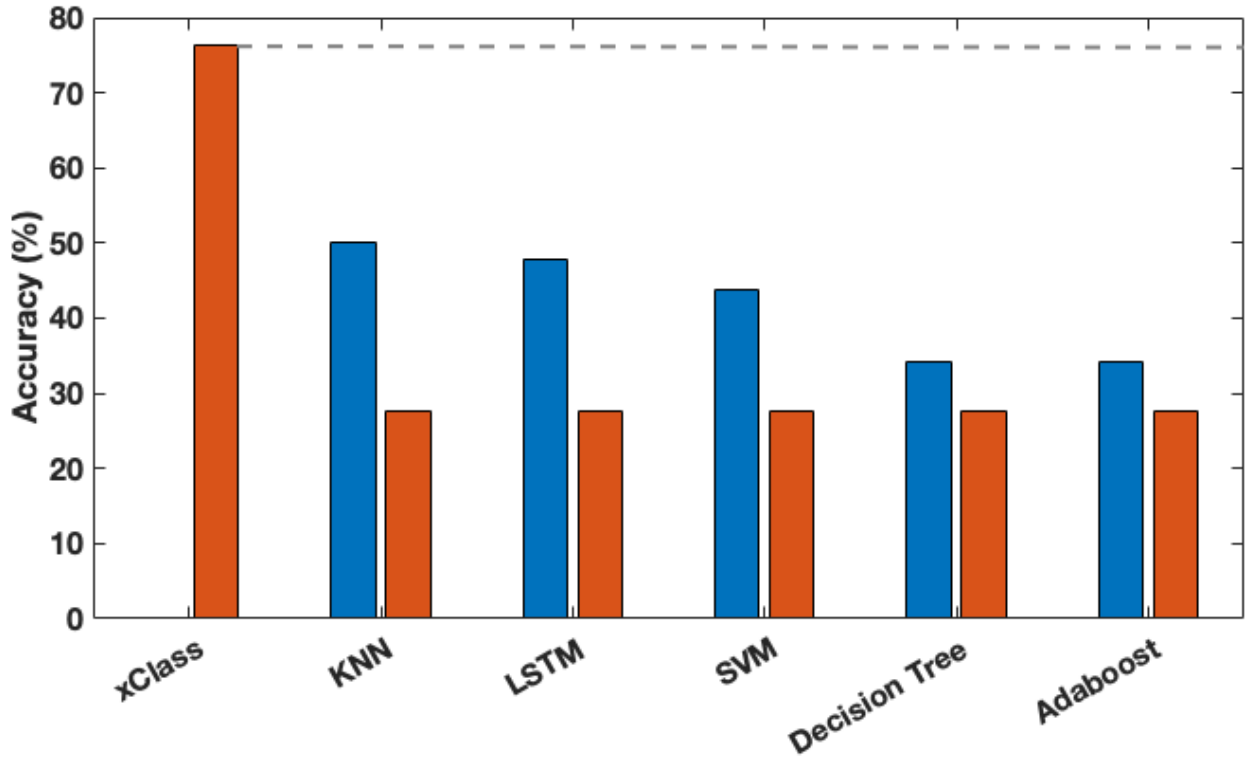

Figure 14

Accuracy for extremely weak supervision classification for the Cars dataset.

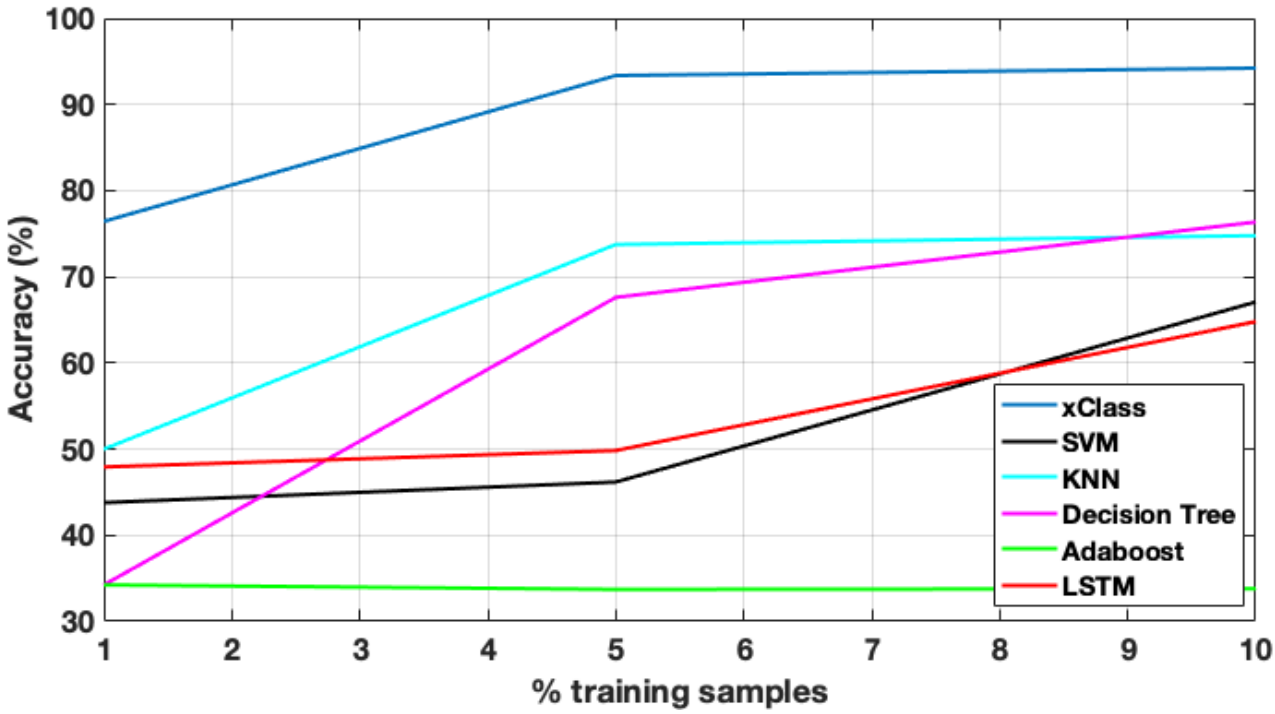

Figure 15

Classification curve for different number of training samples for the Cars dataset.