

INTRODUCTION TO CONFORMAL PREDICTION WITH APPLICATION ON IMBALANCED DIABETES CLASSIFICATION

Mabrouka Salmi¹

¹Universidad de Cordoba, Cordoba, Spain. z12salsm@uco.es

Introduction

Traditional machine learning methods provide single value predictions without indicating the confidence or reliability of these predictions [1]. Conformal prediction (CP) addresses this limitation by producing prediction sets or intervals that come with a measure of their accuracy and reliability, thereby guaranteeing a given error rate [2]. CP methods rely on minimal assumptions, such as the test data being drawn from the same distribution as the training data and the exchangeability of the data points [3]. Non-conformity measures (NCMs) are used to assess how unusual a new example is compared to the training examples. Based on the NCM, the CP framework generates prediction sets by evaluating the randomness of hypothetical completions of test examples and including only those labels that meet a predefined significance level. The validity of CPs is ensured by calculating p-values for hypothetical completions and forming prediction sets that reflect a chosen confidence level. CPs aim for a balance between validity (ensuring error rates do not exceed the significance level) and efficiency (keeping prediction sets as small as possible without missing correct labels).

Background

Transductive and Inductive CP Algorithms
Inductive conformal prediction (ICP) splits the training data into training and calibration sets. After training a model, nonconformity scores are calculated for both sets. These scores are ordered, and the significance level α is chosen. The prediction intervals are constructed as $\hat{y} \pm \alpha_s$, where α_s is the $(1 - \alpha)$ percentile of the nonconformity scores. In classification, nonconformity scores, similar to p-values, indicate the probability of more extreme outcomes under the null hypothesis. Transductive conformal prediction (TCP) uses all available data without needing a separate calibration set, leveraging the entire dataset for efficient and precise interval predictions. Nonconformity measures, such as absolute error for regression tasks or Hinge loss for binary classification, are selected based on the problem:

Non-conformity score: $\alpha = |y - \hat{y}|$

Non-conformity score: $\alpha = 1 - P(y|x)$

The model is trained on the entire dataset, and nonconformity scores for hypothetical completions are calculated. P-values are computed to determine the statistical significance of these scores, guiding the inclusion of hypothetical completions in the prediction set based on the chosen significance level α .

Label-Conditional Conformal Prediction
Label-conditional conformal prediction (Mondrian CP) constructs prediction intervals based on the conditional probability of labels given features. Nonconformity scores are computed conditioned on the predicted label:

$\alpha(\mathbf{x}, y)$ = deviation between expected output and true label

P-values for test objects with hypothetical labels are calculated, considering only examples with the same label as the hypothetical label. This approach leads to more accurate and calibrated prediction sets, especially when the conditional distribution of labels varies significantly across instances.

Methodology

We want to experiment with CP for the classification learning task. We use the three common classical conformal prediction algorithms: the transductive or full conformal prediction, the inductive conformal prediction, and the label conditional conformal prediction. We interestingly experimented with conformal prediction for imbalanced data, which is not well explored yet, particularly in medical diagnosis applications. First, we select an imbalanced medical data set to expose the inductive and transductive algorithms for imbalanced classification. Then, to better evaluate the conformal prediction on imbalanced data, a simple random forest classifier is used as a baseline performance to compare the results of the classification of patient cases.

Experimental work and Results

0.1 Used Dataset and Evaluation

The Pima Indian diabetes dataset, obtained from the UCI Repository, is a popular benchmark [4]. It presents classification challenges due to its moderate imbalance and small sample size. Data was split for CP algorithms as follows:

- **Transductive CP**: 80-20 train-test split.
- **Inductive CP**: 70-15-15 train-calibration-test split.

The Random Forest classifier, capable of handling continuous and discrete variables, was used. CP models were evaluated for validity and efficiency:

- **Validity**: Ensured by CP framework.
- **Efficiency**: Measured by the informativeness of predictions.
- **Metrics**: Accuracy, mean error of classification, and number of single class predictions.

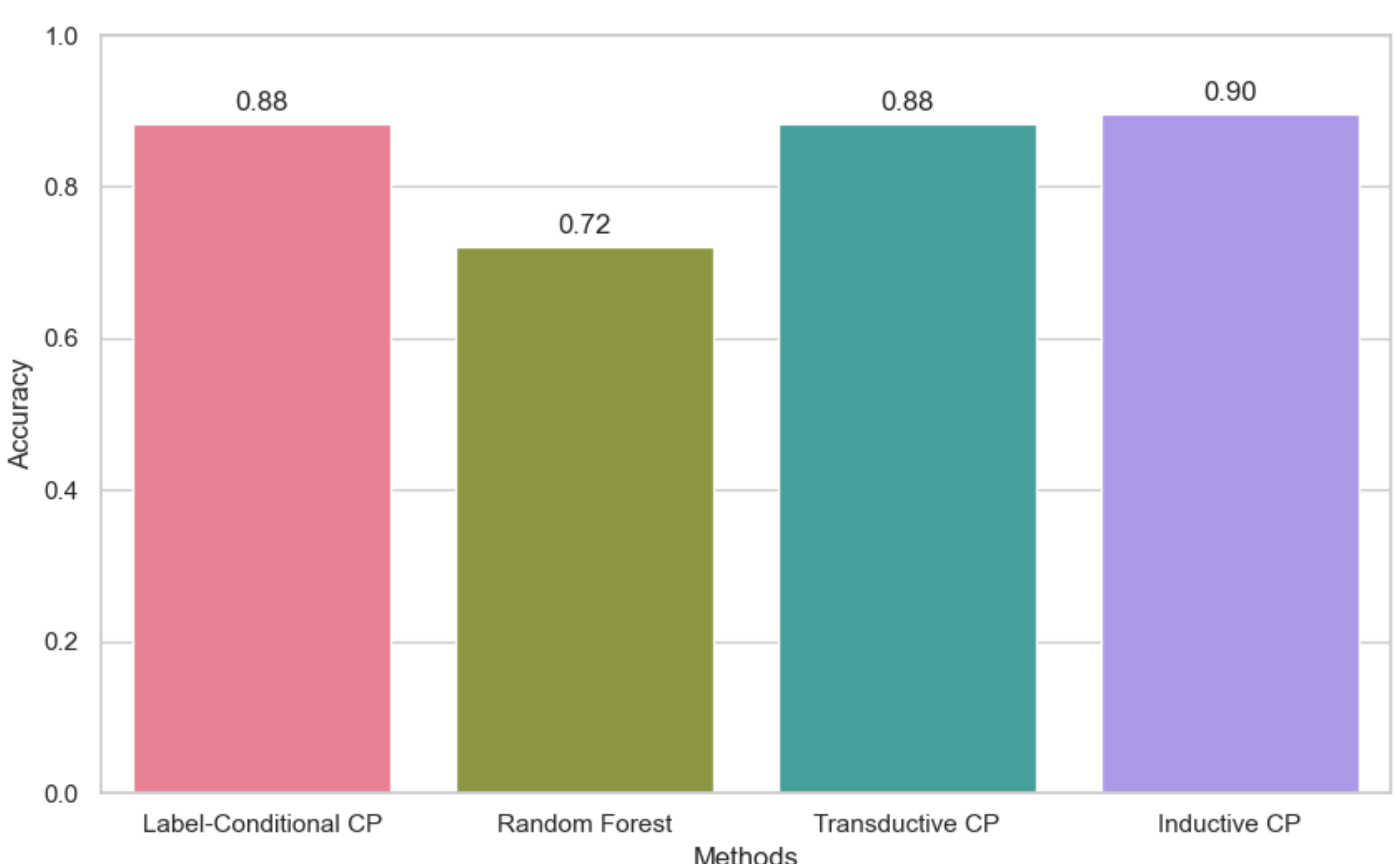
0.2 Results

We applied three conformal prediction algorithms based on the Random Forest classifier: transductive CP, inductive CP, and label-conditional CP. The Random Forest classifier was trained on the train data and used to produce test set predictions. The table below displays the CP models and the efficiency and validity metrics scores of their predictions. All related code is available on this GitHub repository [5] to share further related investigations.

CP Model\Metric	Err Rate	Avg C/pred
Inductive CP	0.1034	1.30
Transductive CP	0.1168	1.3311
Label-Conditional CP	0.1168	1.3831

Note: Avg C/pred: Average number of classes per prediction; Err Rate: Error Rate

To compare the performance of the CP and non-CP models, we refer to calculating the accuracy of the predictions of the test set. Figure 0.2 plots the accuracies of each CP and random forest model. Accordingly, we see the higher performance achieved by the CP models.



Discussion

The CP algorithms for the diabetes classification have surpassed the random forest classifier regarding prediction accuracy. However, the three employed CP algorithms present slight differences between them. The inductive CP has produced the most accurate diabetes classification results with the least error rate and the highest accuracy. Although the difference could be noticed concerning the error rate, the efficiency measure shows almost similar values referring to insignificant differences in CP model efficiency. Moreover, as seen in figure 0.2, the CP models not only showed good accuracy values and relatively low error rates, but the performance of the build CP models also exceeds that of the models proposed in the most recent experimental research with different ML algorithms for Indian diabetes classification [4]. The accuracy of the optimal classification model in that research is around 80%. Then, the CP models trained for CP trials on class imbalance data within this work have largely improved the performance of Pima diabetes classification without reference to class imbalance data level methods. Such preliminary results are promising for further experimentation with imbalanced data and validating the CP performance in different dataset configurations.

Conclusions

The evolving conformal prediction approach for soft computing helps provide informative predictions for future objects without relying on past object distribution or parametric methods. Such an agnostic model approach could be adapted to the problem and the data. Besides, its validity property is guaranteed regardless of the underlying machine learning algorithm or non-conformity measure. Recent research is focused on broadening the usage of such methods in different contexts and problems like image and time series data. Hence, even when data assumptions are violated, the CP could still be adapted to various scenarios. In our empirical work with Pima Indian diabetes, one of the challenging benchmark datasets, it is revealed that CP models considerably improved the classification results and significantly outperformed the state-of-the-art results. As was expected, the CP models based on random forest outperformed the non-CP random forest. Finally, these preliminary results should be completed with a thorough evaluation of the CP classifiers on various imbalanced datasets regarding data size and imbalance degree.

References

[1] Paolo Toccaceli. Introduction to conformal predictors. *Pattern Recognition*, 124:108507, 2022.

[2] Alexander Gammerman and Vladimir Vovk. Hedging predictions in machine learning. *The Computer Journal*, 50(2):151–163, 2007.

[3] Paolo Toccaceli and Alexander Gammerman. Combination of inductive mondrian conformal predictors. *Machine Learning*, 108:489–510, 2019.

[4] Victor Chang, Jozeene Bailey, Qianwen Ariel Xu, and Zhili Sun. Pima indians diabetes mellitus classification based on machine learning (ml) algorithms. *Neural Computing and Applications*, 35(22):16157–16173, 2023.

[5] Mabrouka Salmi. Conformal prediction for imbalanced data, 2023. URL <https://github.com/Bouka12/Conformal-Prediction-for-Imbalanced-data>. GitHub repository.