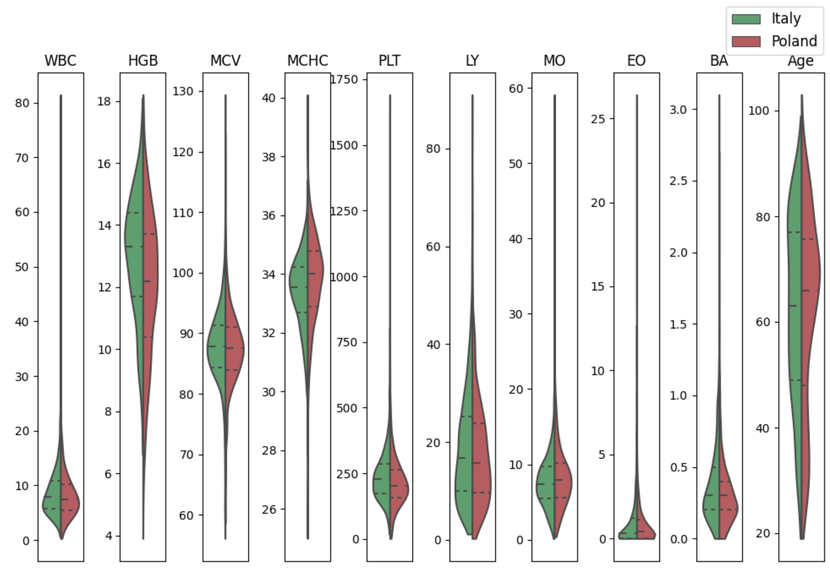
Challenges of domain adaptation in medical domain

Barbara Klaudel, Piotr Frąckowski, Wasyl Badyra, Sebastian Krajna

Gdańsk University of Technology, Lion.ai (prev. Radiato.ai)

1 Introduction

It is generally believed as a rule of thumb that the more training data, the better, and transfer learning always improves the results. In the medical domain, access to data is limited and annotation is costly. To tackle this problem, researchers often extend the data from their local institutions with open-source data. However, the population characteristics of the datasets related to the same task may vary significantly, and merging them may hurt the performance. In our work, we test a variety of different approaches related to domain adaptation and model fusion. For this purpose, we use datasets for SARS-CoV-2 detection from complete blood count from 6 countries.



2 Results - external validation

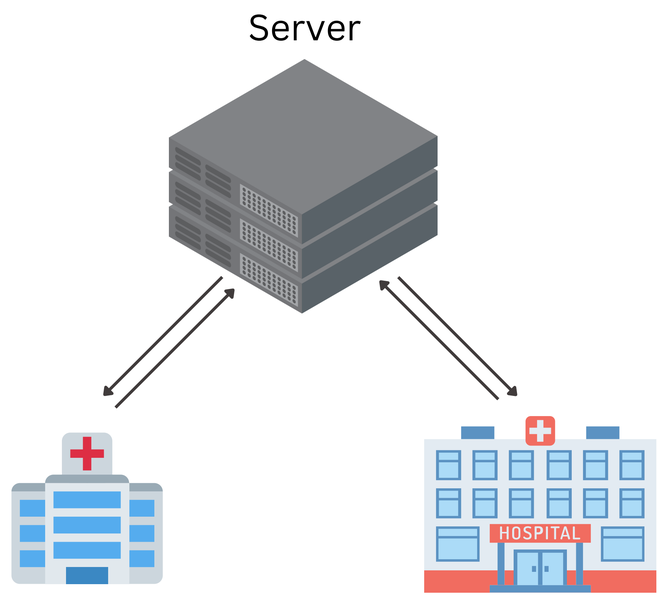
We tested the effectiveness of developing a model on training data from 6 institutions and external validation on data from 2 other institutions. We used 5 methods of domain adaptation and model fusion. The purpose of the experiment was to check the usability of the models developed in several countries during different stages of the pandemic on the data from new countries. The experiment shows that the development of such a model is challenging and the availability of data from the same country as the test data gives an advantage. Kernel Mean Matching produced bast results for both datasets.

1 dataset

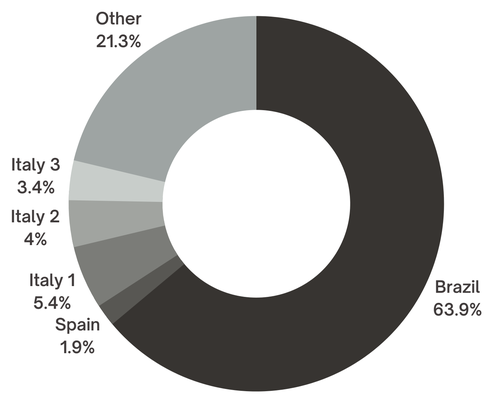
other datasets

1. Transform source based on the target
2. Pick a different dataset as a source and use the other datasets as a target.
3. Repeat for each dataset. Use datasets before transformation as a target.

The strategy of domain adaptation. We evaluated 5 methods of instance-based domain adaptation.



We trained each model on data from a single institution and sent the parameters to a server after each epoch.



Method	PL	IT
Model fusion	50%	37%
Balanced Weighting	45%	56%
TrAdaBoost	46%	54%
WANN	36%	50%
KMM	54%	63%
KLIEP	45%	55%

F1 scores for different domain adaptation strategies evaluated on data from Polish and Italian hospitals.

3 Results - 2 countries

Model	Baseline F1	After DA
Logistic Regression	48%	52% (+4pp)
kNN	55%	56% (+1pp)
Decision Tree	62%	59% (-3pp)
Random Forest	58%	77% (+19pp)
Naive Bayes	43%	60% (+17pp)
XGBoost	60%	67% (+7pp)
CatBoost	63%	65% (+2pp)
ANN	37%	69% (+32pp)
Mean	53%	63% (+10pp)

For this experiment, we limited our data to the Polish and Italian datasets. We compared the performance of the models trained on the data from 1 institution with the training on data from both the target institution and a supplementary institution. The supplementary dataset was transformed to resemble the distribution of the training data. The performance of the models improved after adding a transformed supplementary dataset. The training on the target dataset and a supplementary dataset in its original form decreased the performance compared with training only on the target dataset.

Model	Baseline F1	After DA
Logistic Regression	33%	63% (+30pp)
kNN	45%	62% (+17pp)
Decision Tree	59%	60% (+1pp)
Random Forest	48%	62% (+14pp)
Naive Bayes	39%	63% (+24pp)
XGBoost	33%	64% (+31pp)
CatBoost	69%	67% (-2pp)
ANN	33%	64% (+31pp)
Mean	45%	63 (+18pp)

4 Conclusion

We can improve the performance of a model by adding external datasets, however, such an approach, requires careful inspection of the underlying distribution of the data and might necessitate the use of domain adaptation. The proximity of supplementary datasets to the target dataset improves knowledge transfer.

Related literature

Klaudel, Barbara, et al. "Machine-aided detection of SARS-CoV-2 from complete blood count." International Conference on Diagnostics of Processes and Systems. Springer, Cham, 2023.

Cabitza, Federico, et al. "The importance of being external. methodological insights for the external validation of machine learning models in medicine." Computer Methods and Programs in Biomedicine 208 (2021): 106288.

Cabitza, Federico, et al. "Development, evaluation, and validation of machine learning models for COVID-19 detection based on routine blood tests." Clinical Chemistry and Laboratory Medicine (CCLM) 59.2 (2021): 421-431.