**A close up of a sign

Description generated with very high confidence**

**“Black-box Tests for Algorithmic Stability”  
Regression & Coverage Stability  
Implemented on IVIM**

Kfir Levy & Priel Salomon

Technion - Electrical Engineering

048100 - Reliability in ML course

Final project

# Abstract –

In our work we our based on a [paper](https://arxiv.org/pdf/2111.15546.pdf) from December 23, 2022 "Black-box tests for algorithmic stability" by Byol Kim et al. Algorithmic stability is a concept from learning theory that expresses the degree to which changes to the input data (i.e. removal of a single data point) may affect the outputs of a regression algorithm. Knowing an algorithm’s stability properties is often useful for many downstream applications—for example, stability is known to lead to desirable generalization properties and predictive inference guarantees. However, many modern algorithms currently used in practice are too complex for a theoretical analysis of their stability properties, and thus we can only attempt to establish these properties through an empirical exploration of the algorithm’s behavior on various datasets. In this work, we lay out a formal statistical framework for this kind of black-box testing without any assumptions on the algorithm or the data distribution, and establish fundamental bounds on the ability of any black-box test to identify algorithmic stability.

לקחנו מאמר

לקחנו משהו מעולם הרפואה

לקחנו משהו מהקורס

הרצנו את הדבר מהקורס על הדבר ברפואה

ובסוף הרצנו את הדבר המאמר על הדבר מהקורס שהרצנו על הדבר מהרפואה.

* Section 1: Background and problem setup.

## Conformal prediction -

In this work, we implemented the Conformal prediction algorithm on a specific medical imaging estimation task. Conformal prediction (CP) is a statistical technique for producing prediction intervals without assumptions on the predictive algorithm (often a machine learning system) and only assuming exchangeability of the data. CP works by computing a nonconformity measure, often called a score function, on previously labeled data, and using these to create prediction sets (or intervals for a regression estimation) on a new (unlabeled) test data point. Conformal prediction requires a user-specified *significance level* for which the algorithm should produce its predictions. This significance level restricts the frequency of errors that the algorithm is allowed to make. For example, a significance level of 0.1 means that the algorithm can make *at most* 10% erroneous predictions. To meet this requirement, the output is an **interval prediction**, instead of a**point prediction**produced by standard supervised machine learning models. For regression tasks, this means that predictions are not a specific value, for example 34.768, but instead an interval of 31.56 – 37.67. Depending on how good the underlying model is (how well it can estimate the interval) and the specified significance level, these intervals can be smaller or larger. The output is prediction intervals, where a smaller significance level (fewer allowed errors) produces wider intervals which are less specific, and vice versa – more allowed errors produce tighter prediction intervals.

## IVIM Estimation –

This imaging technique has been developed with the objective of obtaining not only a functional analysis of different organs but also different types of lesions. Among many accessible tools in diagnostic imaging, IVIM MRI aroused the interest of many researchers in terms of studying its applicability in the evaluation of abdominal organs and diseases. The major conclusion of this article is that IVIM MRI seems to be a very auspicious method to investigate the human body, and that nowadays the most promising clinical application for IVIM perfusion MRI is oncology.

The intravoxel incoherent motion (IVIM) diffusion-weighted (DW) model as a possible imaging technique, using multiple b values and bi-exponential fitting for the concurrent estimation of the pure molecular water diffusion and microcirculation of blood water in randomly oriented capillaries (perfusion) was first introduced in the late 1980s by Le Bihan *et al*. [[1](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7757509/#ref1)]. The idea to use diffusion and IVIM magnetic resonance imaging (MRI) to acquire perfusion parameter maps was considered revolutionary but technically difficult, and, as a result, it was more than 20 years before the method started being used in clinical practice [[2](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7757509/#ref2)]. IVIM reflects the random microscopic motion of water molecules that occurs in each voxel on MR images not only in intra- or extracellular space but also in microcirculation of blood [[3](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7757509/#ref3)]. According to IVIM theory, diffusion and perfusion are affected by several tissue characteristics, including the presence of restrictive barriers within tissue, the viscosity of the fluid in which the spins are diffusing, and the velocity and fractional volume of perfusing spins [[4](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7757509/#ref4)]. Formerly, due to degradation of images caused by cardiac, respiratory, and other motion artifacts, IVIM imaging was restricted to neuroradiologic applications. Nowadays, it is pursued to apply IVIM MRI to evaluate almost entire human body. Over the last few years there has been a revival of interest in IVIM MRI and its applications in many fields, particularly in oncology.

רקע על האלגוריתם של conformal prediction ורקע על IVIM ורקע על המודל NN של ברביארי שנותן את השיערוך. להסביר על הסביבה שבה הרצנו את המודל

• Section 2: The chosen paper. Here you are expected to clearly describe and analyze the work you build upon (“base method”) and discuss its limitations.

להסביר על המאמר, על הרעיון שעומד מאחורי היציבות ולתת את ההגדרות שיש במאמר ולפרט על השימוש בהן בהרצת המודל.

• Section 3: Creative extension. This section is the core of the project. Discuss the modifications you suggest for improving the base method.

When observing the proposed method in [5], and also the known disadvantages of working with a model prediction score only (i.e. without a certainty coverage), it became interesting to examine a theory which would combine the black box stability tests with certainty coverages, in hope it would increase our understanding of such test and it’s meanings, in the same way that estimating a certainty coverage for a prediction model increases our understanding of the model’s prediction.

Our proposed method is to basically replace the way is calculated, while all other elements of the process remain the same. In our proposed method we use:

Function can be, for example,

or

when represent the upper limit and lower limit of the algorithm’s prediction interval, respectively. In the second proposal of , we calculate the average of the percentages of the intersection in relative to each of the intervals. It satisfies when there is an intersection between the two intervals and only if the intervals are identical. Negative values of are the result of no intersection between the intervals, e.g. .

The goal for this proposal is to test the intersection of the prediction intervals of trained models , and like in the original test, high intersection (the models produced similar intervals) means a stable solution, and not stable otherwise.

Working with intervals instead of with output alone could, as the means for evaluating stability, takes into account the model’s level of certainty in the prediction, since, just like we expect similar outputs, we expect similar level of certainty (which translates to similar confidence interal).

It is also important to note that one of the most important properties of these tests are that they can be applied as black box tests, and expanding the algorithm to confidence interval does not impact this property.

בחלק הזה נפרט על השימוש ברעיון היציבות והמימוש שלו על הcoverage.

• Section 4: Results. Provide a detailed performance analysis of your proposal and compare it to the base method, if relevant.

תמונות..... הרבה תמונות.. של האימון, של הסטייה מהשיערוך, של האלגוריתם conformal prediction על הדאטא, של היציבות של השיערוך ושל היציבות של התחום כיסוי.

• Section 5: Conclusion and future work. Discuss the limitations of your proposal and suggest future extensions.

הסבר מה ניתן לעשות עם זה הלאה.

* להרחיב לחישובי Interval אחרים, למשל גם לקחת בחשבון רוחב ממוצע של אינטרוול אם האינטרוול המחושב אינו ברוחב קבוע, כלומר לעשות את הstability test על כמה דוגמאות.

שאפשר לבחון את הנכונות המתמטית מאחורי השערת בדיקת יציבות הכיסוי.

לפרט על המבחן למדד הnon-conformity עבור תחומי הכיסוי, המרחק האוקלידי בין הגבולות וכו'.

* References. This section should include all the papers you cited throughout the report.

[1] Le Bihan, D., Breton, E., Lallemand, D., Grenier, P., Cabanis, E., & Laval-Jeantet, M. (1986). MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. Radiology, 161(2), 401-407.‏

[2]

[3]

[4]

[5] Kim, B., & Barber, R. F. (2021). Black box tests for algorithmic stability. *arXiv preprint arXiv:2111.15546*.‏

[6]