**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

We based our work on an article "Black-box tests for algorithmic stability" (Byol & Kim, 2021). 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. Therefore, 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. 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*. The idea is to use diffusion and IVIM magnetic resonance imaging (MRI) to acquire perfusion parameter maps. 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. 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. Formerly, due to degradation of images caused by cardiac, respiratory, and other motion artifacts, IVIM imaging was restricted to neuroradiologic applications. Over the last few years there has been a revival of interest in IVIM MRI and its applications in many fields, particularly in oncology.

The basic IVIM diffusion and perfusion model for the signal intensity (per pixel):

In this model we have 3 different parameters to estimate for any pixel in MRI image:

which in simple words are the diffusion and perfusion factors and their proportion in the physical scanned voxel.

## IVIM DNN approach

A paper from 2020 by Barbieri, S., Gurney‐Champion, O. J., Klaassen, R., & Thoeny, H. C. proposed a Deep Learning approach to solve the IVIM model's parameters. A feed‐forward backward‐propagation DNN was trained to generate estimates of IVIM parameters (). Training is unsupervised and needs to be repeated for data sets with different distributions (e.g., because of different acquisition protocols or imaged anatomical regions). Given that the goal is to encode a given data set, separate training and testing data sets are not required and the network was trained directly on the data set of interest. The network is composed of an input layer, 3 hidden layers, and an output layer. The passthrough input layer is made of neurons, which take the normalized diffusion‐weighted signal sampled at each b‐value as input. The 3 hidden layers are fully connected, with a number of neurons equal to the number of b‐values of the data of interest and an exponential linear unit activation function. The output layer is made of 3 neurons, which hold the estimated parameter values. Initial network weights were set using He initialization or using a previously trained network. An Adam optimizer was used for training with the mean squared error between the observed input S(b) and the signal Ŝ(b), reconstructed based on the IVIM model and as loss function. Early stopping was implemented by terminating training after the loss function did not improve for 10 consecutive iterations. The proposed neural network architecture is essentially an autoencoder with the constraint that the input signal should be encoded by the 3 IVIM parameters. The network does not impose any restrictions on the range of fitted parameter values.

• 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.

# The chosen paper *“Black-box Tests for Algorithmic Stability”*

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.

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.

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.

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

Le Bihan, D., Breton, E., Lallemand, D., Aubin, M. L., Vignaud, J., & Laval-Jeantet, M. (1988). Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging. *Radiology*, *168*(2), 497-505.

|  |
| --- |
|  |
| Barbieri, S., Gurney‐Champion, O. J., Klaassen, R., & Thoeny, H. C. (2020). Deep learning how to fit an intravoxel incoherent motion model to diffusion‐weighted MRI. *Magnetic resonance in medicine*, *83*(1), 312-321. |
| Chicago |  |