Assignment 4: Diagnostic Systems

Group Number: 59
Module Number: 4
Members:
Alexander Boström (Personal Number: 19970816-2210, Program: MPMOB)
Anton Levinsson (Personal Number: 19981123-2470, Program: MPALG)
Email: alexander.bostroem@gmail.com, anton.levinsson@gmail.com
Group Configuration: Two-person group
Declaration: We hereby declare that we have both actively participated in solving every exercise. All solutions are entirely our own work, without having taken part of other solutions.
Hours Spent: Alexander - 18 hours, Anton - 18 hours (excluding lectures)

February 13, 2024

1 Reading and reflection

1.1 Machine learning techniques to diagnose breast cancer from image-processed nuclear features of fine needle aspirates

The paper focuses on employing machine learning techniques for diagnosing breast cancer using image-processed nuclear features obtained from fine needle aspirates, as the title suggests. Fine needle aspirates are samples collected from breast tissue, and the study utilizes advanced image processing to extract nuclear features. Machine learning algorithms are then applied to analyze these features and make predictions about the presence or absence of breast cancer. The goal is to improve diagnostic accuracy and efficiency through the integration of computational methods with medical imaging. The study demonstrates the extensive preprocessing required, as well as the numerous characteristics of the nuclear features that are measured in order to achieve such a high level of accuracy.

1.2 The Mythos of Model Interpretability

This paper explores the challenges and complexities associated with interpreting machine learning models. While there is a growing demand for building models that are interpretable and understandable, the paper argues that achieving complete interpretability is often a formidable task, encompassing numerous factors, some of which are sometimes overlooked. The author contends that interpretability is not a monolithic concept but rather a collection of diverse and sometimes conflicting ideas that require clarification and evaluation.

The complexities of modern machine learning models make it challenging to fully interpret their decision-making processes. The paper discusses various aspects of interpretability, one example being transparency, and more specifically, simulatability. Simulatability is considered achieved if the model is simple enough for a person to comprehend its entire workings at once. For instance, a human should be able to follow every calculation required to produce a prediction within a reasonable timeframe.

Another example provided is text explanations, a post hoc interpretation technique that employs natural language to explain the predictions or behavior of a model. For instance, a model diagnosing diseases might offer a verbal justification for its diagnosis based on the patient's symptoms and medical history.

2 Implementation

Text and code-description in separate PDF.

3 Discussion

Interpretability, particularly in healthcare, refers to the ability to understand and explain the decisions or predictions made by machine learning models in a way that is meaningful to medical professionals. It involves translating complex model outputs into understandable terms, providing insights into the features that influence predictions, and allows for trust in the model's recommendations.

Interpretability is essential for addressing ethical concerns. In healthcare, decisions often have profound consequences, and understanding how a model arrives at a certain prediction helps ensure accountability and fairness in the decision-making process. Having models with lower interpretability might result in biases if people of similar features are dominant in testing data. Moreover, interpretable models can build trust among medical professionals and patients. If the decision-making process is transparent and can be easily explained, individuals are more likely to trust the model's recommendations, leading to better acceptance and adoption in real-world clinical settings.

Additionally, a more obvious scenario where interpretability plays a crucial role is in making potentially life-critical healthcare decisions. While high accuracy is undoubtedly preferred, lacking interpretability means predictions could be based on unknown metrics. This could lead to several adverse effects in both the short and long run, especially if the input data would turn out to be incorrect.

However, having a interpetable model might also have some drawbacks. As in our case with the rule-based classifier, it was almost a certainty that it would results in too many false negatives. This is often the case because you need to simplify your model which can lead to a loss of predictive performance.