Which Explanation Makes Sense? A Critical Evaluation of Local Explanations for Assessing Cervical Cancer Risk Factors

Project Proposal submitted for
CS598: Deep Learning For Healthcare, SP24

Instructors:

Jimeng Sun
Siddhartha Laghuvarapu

Prepared by:

Himangshu Das	hdas4@illinois.edu
Jeremy Samuel	sjeremy3@illinois.edu
Mahesh Matta	maheshm3@illinois.edu

Which Explanation Makes Sense? A Critical Evaluation of Local Explanations for Assessing Cervical Cancer Risk Factors

1. Citations to the original paper:

Citation: Ayad, Wafa & Bonnier, Thomas & Bosch, Benjamin & Read, Jesse & Parbhoo, Sonali. (2023). Which Explanation Makes Sense? A Critical Evaluation of Local Explanations for Assessing Cervical Cancer Risk Factors Ecole polytechnique. 1-50.

Additional Citations:

1. Singh, Tonjam & Balaguru, Karthik & Wahengbam, Monita. (2023). An Ensemble Approach for Uterine Pathology Classification in MRI Imaging Using Deep Learning. 1672-1677. 10.1109/ICSCNA58489.2023.10370073. Link
2. Chauhan, Nitin & Singh, Krishna & Kumar, Amit & Kolambakar, Swapnil. (2023). HDFCN: A Robust Hybrid Deep Network Based on Feature Concatenation for Cervical Cancer Diagnosis on WSI Pap Smear Slides. BioMed Research International. 2023. 10.1155/2023/4214817. Link.
3. Lalasa, Mukku & Thomas, Jyothi. (2023). A Review of Deep Learning Methods in Cervical Cancer Detection. 10.1007/978-3-031-27524-1_60. Link.

2. General Problem:

Cervical cancer is a life-threatening disease and one of the most prevalent types of cancer affecting women worldwide. Being able to adequately identify and assess factors that elevate risk of cervical cancer is crucial for early detection and treatment. Advances in machine learning have produced new methods for predicting cervical cancer risk, however their complex black-box behavior remains a key barrier to their adoption in clinical practice. Recently, there has been substantial rise in the development of local explainability techniques aimed at breaking down a model's predictions for particular instances in terms of, for example, meaningful concepts, important features, decision tree or rule-based logic, among others. While these techniques can help users better understand key factors driving a model's decisions in some situations, they may not always be consistent or provide faithful predictions, particularly in applications with heterogeneous outcomes. With this project, we present a critical analysis of several existing local interpretability methods for explaining risk factors associated with cervical cancer. Our goal is to help clinicians who use AI to better understand which types of explanations to use in particular contexts. We present a framework for studying the quality of different explanations for cervical cancer risk and contextualize how different explanations might be appropriate for different patient scenarios through an empirical analysis. Finally, we provide practical advice for practitioners as to how to use different types of explanations for assessing and determining key factors driving cervical cancer risk.

3. Specific Approach:

To investigate the effectiveness of local explanations in interpreting deep learning models' decisions regarding cervical cancer risk factors. To compare and contrast various methods of generating local explanations, including LIME (Local Interpretable Model-agnostic Explanations), SHAP (SHapley Additive exPlanations), Integrated Gradients etc. The project focuses on assessing the quality of different explanations for cervical cancer risk using a multistage analysis pipeline. The approach involves: 1. Testing supervised learning models to predict cervical cancer risk. 2. Selecting the best-performing model. 3. Applying local explainability techniques to interpret the selected model. 4. Computing metrics to evaluate the plausibility and coherence of the explanations. 5. Providing domain experts with insights on factors contributing to cervical cancer risk and when certain explanations may be preferable. 6. Empirical analysis to determine the most suitable explanations for assessing cervical cancer risk factors. 7. Offering practical advice on using different types of explanations for evaluating key factors driving cervical cancer risk.

4. Hypotheses to be tested:

The paper presents a critical analysis of several existing local interpretability methods for explaining risk factors associated with cervical cancer. The goal is to help clinicians who use AI to better understand which types of explanations to use in particular contexts. We shall test the framework provided in the paper for studying the quality of different explanations for cervical cancer risk and contextualize how different explanations might be appropriate for different patient scenarios. It will provide an empirical method to compute the faithfulness metric across various machine learning interpretability methods by using the feature and rank agreement and ensuring that removing the top N features predicted by the local explanations don’t drastically affect model performance.

5. Ablations planned:

The following ablations are planned-

1. Training the models using only the data from UCI repository without using synthetic samples from ADASYN. This will ensure that we are able to check how the different machine learning models are able to handle the undersampled data and whether their accuracies are affected by the synthetic samples introduced in the dataset which may have added unrealistic data for the presence of cancer.
2. Removing a subset of top features from the dataset, retraining the model on the reduced dataset and then evaluating the model accuracy or feature importance - this will help us identify the Faithfulness metric: Remove and Retrain (ROAR) based on the feature and rank agreement for the various explainability methods and to choose the one that provides the most faithful explanations across all the different interpretability methods like SHAP, LIME, DiCE, Local Surrogate etc

6. Accessing Data:

To gather data for our project on assessing cervical cancer risk factors, we will employ a multi-step approach:

1. ***Data Sources***: Our primary data sources will include publicly available datasets obtained from reputable sources. Specifically, we will leverage datasets from

Kaggle, available here and the UCI repository available here. We also plan to use ADASYN (Adaptive Synthetic Sampling) to create synthetic samples to balance the dataset by oversampling the minority class.

2. ***Data Collection Strategy***: We will compile a comprehensive dataset encompassing various cervical cancer risk factors, including demographic details, medical history, and screening outcomes.
3. ***Data Preprocessing***: Before model training, we will conduct preprocessing tasks to clean the data, normalize features, and perform any necessary feature engineering. We plan to use additional python scripts to process/cleanse/transform the data.
4. ***Data Privacy and Ethics***: We are committed to upholding data privacy and ethical standards throughout the project. To this end, we will obtain all necessary approvals and permissions for accessing and utilizing sensitive medical information, prioritizing patient confidentiality and regulatory compliance.

7. Feasibility of computation:

Ensuring the feasibility of computation is vital for the successful execution of our project. Here's how we plan to address this aspect:

1. ***Assessment of Computing Resources***: We will assess the availability of computing resources, including both hardware infrastructure and software tools essential for model training and explanation generation.
2. ***Utilization of High-Performance Computing (HPC)***: To handle the computational demands of training deep learning models and generating local explanations, we will explore the use of high-performance computing clusters or cloud-based services.
3. ***Software Dependency Management***: We will carefully manage software dependencies, including Python programming language, TensorFlow or PyTorch for deep learning, scikit-learn for data preprocessing, and LIME and SHAP libraries for explanation generation.
4. ***Resource Allocation and Scalability***: We will allocate sufficient time and resources for conducting experiments, tuning model hyperparameters, and analyzing results.

8. Use the existing code or not:

We will be using the existing code[2] but also try to test their hypotheses and framework provided by assessing the quality of local explanations for not only random forest but also for other models for cervical cancer risk using several relevant explainability methods.

9. References

1. https://www.researchgate.net/publication/374061335_Which_Explanation_Makes_Sense_A_Critical_Evaluation_of_Local_Explanations_for_Assessing_Cervical_Cancer_Risk_Factors_Ecole_polytechnique
2. https://github.com/cwayad/Local-Explanations-for-Cervical-Cancer