MISCARRIAGE PREDICTION USING ENSEMBLE DEEP LEARNING MODELS

Project Outline

SUPERVISOR: MUHAMMAD ASLAM [mua19]

SOUNIC AKKARAJU [soa39]
STUDENT ID: 230072268
COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE [GG4R]

MODULE CODE: CS39440
MODULE NAME: MAJOR PROJECT
DATE: 9th FEBRUARY 2025

Department of Computer Science | Aberystwyth University

Project Description

The project “Miscarriage Prediction using Ensemble Deep Learning Models” seeks to create a predictive tool that can aid clinical staff in the early identification of miscarriage risks. In this project, explainable deep learning techniques are applied to improve the model’s performance and its interpretation.

Miscarriage is one of the most delicate issues in prenatal healthcare, but predicting it can help delivering care for at-risk pregnancies. Conventional statistical models often fail to address the intricacies and variabilities of medical data. As a result, this project will utilize deep learning techniques to predict clinical data corresponding to pregnancy health indicators and achieve higher accuracy and transparency in diagnosis.

This model will be implemented using a dataset that contains several health indicators concerning pregnancy. The quality of the data and its predictive accuracy will be enhanced by the application of feature extraction techniques such as **Least Absolute Shrinkage and Selection Operator (LASSO), Least Angle Regression (LARS) and Random Forest (RF),** as well as data balancing techniques like **Synthetic Minority Over-Sampling Technique (SMOTE)-Tomek and NearSMOTE**. The core models comprise **Echo Dense Inception-Blended (EDI-Blend)** and **Dense Reservoir Inception Modular Network (DRIM-Net)**; both are ensemble deep learning models that consist of **Echo State Networks (ESN) together with Densely Connected Convolutional Network (DenseNet)** and **Inception Network (InceptionNet)** for multi-scale feature extraction and classification improvement.

A key aspect of the project is the interpretability of the model, which provides the means through SHAP for the healthcare professionals to understand the most relevant factors leading to the risk of miscarriage. The component is crucial because it allows the model predictions to be utilized in real-life clinical scenarios. The final system will be tested on effectiveness by measuring its accuracy, precision, recall, F1 score and Area Under Curve-Receiver Operating Characteristic (AUC-ROC).

Proposed Tasks

1. Research and Planning:
    - Carry out an extensive analysis of current literature regarding the prediction of miscarriage, particularly concentrating on the deep learning based medical predictions.
    - Understand the EDI-Blend and DRIM-Net architectures through a detailed study to analyse their strengths and weaknesses.
    - Study and review **Shapley Additive Explanations (SHAP**) and **Local Interpretable Model-agnostic Explanations (LIME)** explainability methods to understand model transparency.
2. Data Preprocessing and Feature Engineering:
    - Perform feature ranking and selection using LASSO, LARS, and Random Forest, evaluating the impact of the chosen features on model performance.
    - Measure effectiveness between **SMOTE-Tomek** and **NearSMOTE**, implementing the one which performs better.

- Perform **exploratory data analysis (EDA)** to discover possible trends and biases within the dataset.

3. Baseline Model Development:
   - Create a **baseline Multilayer Perceptron (MLP)** model that includes **Batch Normalization (BatchNorm), Dropout**, and **L2 regularization** to reduce overfitting.
   - Test the model using **5-fold cross validation** to check whether the model learned a general concept.
   - Check the **accuracy**, **precision**, **recall**, **F1-score**, and **AUC-ROC** to assess model performance.
4. Implementation of the **EDI-Blend**:
   - Utilize Echo State Networks (ESN) to analyse sequential pregnancy related health data.
   - Incorporate DenseNet to enhance hierarchical feature extraction in medical records.
   - Integrate InceptionNet to collect multi-scale representations for better classification accuracy.
   - Create a blending mechanism which utilizes the advantages of these models for the final prediction.
5. Implementation of the **DRIM-Net Model**:
   - Adjust the deep learning pipeline for parallel feature extraction from ESN, DenseNet, and InceptionNet.
   - Create feature fusion layers, allowing the model to better combine the extracted features.
   - Train and test the optimised **DRIM-Net** and compare its performance against **EDI-Blend** and the **baseline model**.
6. Model Explainability and Interpretation:
   - Put in place **SHAP**-based interpretability on all the purchased models.
   - Make summary plots, dependence plots and force plots to show the most salient pregnancy risk factors.
   - Check SHAP feature rank against the **LASSO**, **LARS**, and **Random Forest** feature selection to confirm the use of the final selection features.
   - Capture observations to improve model explainability and decision-making.
7. Finalize Report & Mid-Project Demonstration:
   - Detailed and consolidate outcomes, processes, and progress for the mid-project demonstration.
   - Design and prepare visual aids depicting model architecture, comparisons, and explainability observations.

Project Deliverables

1. **Dataset**: A fully processed and balanced dataset that contains a critical set of pregnancy health indicators.
2. **Predictive Models**: A collection of deep learning models created with the objective of improving the miscarriage prediction accuracy: Baseline Model, EDI-Blend, DRIM-Net

3. **Documentation and Codebase**: A complete GitLab repository with all source code, including scripts for pre-processing, training, and evaluation, which is fully documented. This repository will be augmented with the documentation that allows the user to reproduce the results easily.
4. **Progress Report:** A report that captures the significant milestones reached, the models developed, the evaluations performed, and key takeaways from the project. The report is to be discussed with the project supervisor on a weekly basis.
5. **Final Project Report:** An explanation of the research beginning with methodology, implementation, model evaluations, challenges, and references. It also outlines the ethical implications and strategies for deployment.

Initial Annotated Bibliography:

1. **Aggarwal, C. C.** (2018). *Neural Networks and Deep Learning: A Textbook*. Springer.
    - Provides a structured introduction to deep learning, including ensemble methods and performance evaluation metrics.
2. **Bishop, C. M.** (2006). *Pattern Recognition and Machine Learning*. Springer.
    - Explores **statistical learning approaches, feature selection, and probabilistic models**.
3. **Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P.** (2002). "SMOTE: Synthetic Minority Over-sampling Technique." *Journal of Artificial Intelligence Research*.
    - Explains **SMOTE-based balancing methods**, used for handling class imbalance in medical datasets.
4. **Goodfellow, I., Bengio, Y., & Courville, A.** (2016). *Deep Learning*. MIT Press.
    - A comprehensive book on deep learning techniques, including **architectures, optimizations, and ensemble methods**.
5. **Google AI Blog.** (2021). "Explainability in AI: Using LIME and SHAP to Improve Trustworthiness." *Google AI Research*.
    - A practical guide comparing **SHAP vs. LIME** for making deep learning models explainable.
6. **Javed, A., Javaid, N., Hasnain, M., et al.** (2024). "Applying Advanced Data Analytics on Pregnancy Complications to Predict Miscarriage with eXplainable AI." *IEEE Access*. DOI: [10.1109/ACCESS.2024.3486058]
    - This paper introduces **EDI-Blend and DRIM-Net models** and explores their effectiveness in miscarriage prediction.
7. **Lundberg, S. M., & Lee, S. I.** (2017). "A Unified Approach to Interpreting Model Predictions." *Advances in Neural Information Processing Systems (NeurIPS)*.
    - A foundational paper on **SHAP (Shapley Additive Explanations)** for model interpretability.
8. **Hinton, G. E., Srivastava, N., & Swersky, K.** (2012). "Neural Networks for Machine Learning." *Coursera Lecture Notes*.
    - Covers **deep learning architectures, regularization techniques (Dropout, L2), and optimization methods**.
9. **Murphy, K. P.** (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press.
    - Covers **Bayesian methods, regression techniques, and advanced classification models** applicable to miscarriage prediction.