Amrita Vishwa Vidyapeetham

Amrita School of Computing

Technical Report

# Explainable deep learning models for Classification of Ventriculomegaly using Fetal Brain MRI Images

Team : CB1

Roll No : AM.EN.U4CSE20232 HARDIK KUMAR SINGH
Roll No : AM.EN.U4CSE20222 DEVU PAWAN ASHUTOSH
Roll No : AM.EN.U4CSE20225 YESHWANTH REDDY
Roll No : AM.EN.U4CSE20163 SHOAIB AKHTAR

Project Guide : Dr. SIMI S.
Project Coordinator : NAMITHA K

Signature : Signature :

December 22, 2023

# Contents

**Abstract**

The problem addressed in this project is the identification of ventriculomegaly, a condition where brain ventricles are enlarged. The aim is to create an AI model that not only classifies whether a brain has this condition but also explains its reasoning, making it explainable AI. This matters because ventriculomegaly can have serious health implications, and early detection is crucial for effective treatment.

The motivation behind this project is to improve medical diagnosis accuracy and provide understandable insights into the AI's decision-making process. The persistent challenge lies in creating AI models that not only classify accurately but also provide transparent explanations for their decisions, especially in complex medical scenarios. Achieving both accuracy and explainability remains a significant hurdle in AI-assisted medical diagnostics.

# 1 Introduction

In this project centered on identifying ventriculomegaly, we've employed sophisticated eXplainable Artificial Intelligence (XAI) techniques, specifically Lime and GradCAM models. These models not only classify the presence of this brain condition but also provide transparent explanations for their diagnostic decisions. Ventriculomegaly poses significant health risks, necessitating early detection for effective intervention. By integrating Lime and GradCAM into our AI framework, we aim to enhance the interpretability of our model's classifications. This innovative approach promises to improve medical diagnostics, offering clear insights into the AI's decision-making process for improved patient care strategies and treatment outcomes.

This research is driven by the need for accurate and transparent medical diagnoses. Detecting ventriculomegaly, an enlargement of brain ventricles, early is critical for timely treatment. By employing AI methods like Lime and GradCAM, we aim to not only identify this condition but also explain why our AI makes its diagnosis. This dual approach enhances trust in AI-based diagnostics and improve medical understanding. The project strives to bridge the gap between accuracy and interpretability, offering clearer insights into the AI's decision-making process. Ultimately, this work seeks to transform how we diagnose conditions, leading to improved patient care and treatment strategies.
Consider a patient's brain scan revealing ventriculomegaly detected by Lime and GradCAM models. Instead of a simple "yes," these AI models explain where the enlargement occurs. This precise detail guides better treatment plans, showcasing AI's vital role in early identification and informed medical decisions for improved patient care.

Detecting ventriculomegaly faces challenges due to limited high-quality data, complexities in brain anatomy, and the need for AI models to be both accurate and explainable.

Clinical validation, ethical concerns, and the scarcity of relevant data hinder the seamless integration of AI into medical practice for ventriculomegaly detection.
In this project, our approach involves leveraging Lime and GradCAM models within AI frameworks to detect ventriculomegaly while providing understandable explanations for the AI's diagnoses. Our research objectives are to:

1. Develop an Accurate Classifier: Create an AI model that accurately identifies ventriculomegaly from brain scans.
2. Ensure Interpretability: Enable the AI model to explain its decision-making process transparently using Lime and GradCAM.
3. Validate Clinical Applicability: Validate the model's accuracy and usefulness in real clinical scenarios to ensure its reliability.

This approach is promising because it combines accurate classification with the ability to offer clear explanations. By focusing on both accuracy and interpretability, our method enhances trust in AI-based diagnoses, crucial for medical integration and improved patient care. The validation in clinical settings ensures practical applicability, aiming for a robust and reliable diagnostic tool.

This work contributes by developing an AI model that accurately detects ventriculomegaly, incorporating Lime and GradCAM models for clear explanations. Validated in real clinical settings, it aims to revolutionize diagnostics, offering both accuracy and transparency, potentially enhancing patient care outcomes significantly.

- 

## 2 Literature Survey

1.Using deep-learning algorithms to classify fetal brain ultrasound images as normal or abnormal.(2020)
The authors propose a deep learning approach employing a convolutional neural network (CNN) to classify fetal brain ultrasound images as normal or abnormal. Their CNN was trained on a dataset exceeding 29,000 images, achieving a high accuracy of 96.3 percent. Notably, this study was conducted using data from a single medical center, suggesting potential performance variation when applied to datasets from different centers. Moreover, the method is currently limited to specific diagnostic planes, lacking support for others. Future efforts should prioritize creating multicenter datasets to enhance the model's generalizability. Additionally, expanding the algorithm to encompass various diagnostic planes is crucial. Exploring transfer learning methods could further optimize the model's performance across diverse datasets from multiple medical centers, potentially improving its adaptability and reliability in clinical settings.

2.Explainable Deep Learning for Fetal Brain Ventriculomegaly Classification Year - 2021
The authors assessed their model using a separate test set of 2,500 images, achieving an accuracy of 97.3 percent. They also confirmed the model's ability to offer accurate explanations for its predictions. However, being trained on a single-center dataset implies potential performance variation across different centers. Additionally, the model's scope is limited to specific fetal brain abnormalities. To enhance its performance on diverse datasets, exploring transfer learning is recommended. Future efforts should focus on evaluating the model's explainability with real clinical data and developing a practical system for clinician use in clinical settings.

3.Fetal Ventriculomegaly and Hydrocephalus What Shouldn't be Missed on Imaging Year - 2021
The authors highlight the importance of using a standardized approach to measuring the lateral ventricles on ultrasound images. The paper is focused on imaging findings, so it does not discuss the clinical aspects of fetal ventriculomegaly Develop new imaging techniques for the diagnosis and assessment of fetal ventriculomegaly, such as diffusion tensor imaging (DTI) and tractography.

4.U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications Year - 2021
U-Net models exhibit strong generalization capabilities to new data, adept at segmenting small and rare lesions challenging for conventional methods. They excel in simultaneous segmentation of multiple anatomical structures. As a review article, it lacks novel research and overlooks U-Net limitations, including sensitivity to hyperparameters and reliance on extensive training data. To address this, there's a call to develop more explainable U-Net variants, enabling clinicians to comprehend the model's prediction rationale. These variants should enhance interpretability while retaining the ability to segment multiple anatomical structures concurrently.

5.Detecting and Classifying Fetal Brain Abnormalities Using Machine Learning Techniques Year - 2018
The paper introduces a convolutional neural network (CNN)-based machine learning approach for detecting and categorizing fetal brain abnormalities using ultrasound images. Trained on a dataset of nearly 30,000 images from a single medical center, encompassing normal and abnormal cases, the method's performance might vary when applied to data from diverse medical centers. Presently, it only supports the axial plane, while fetal brain ultrasounds are also obtained in coronal and sagittal planes. Future research could explore transfer learning to enhance model performance across multiple centers. Additionally, evaluating the method on a broader range of abnormalities like microcephaly, encephalocele, and agenesis of the corpus callosum would be beneficial for its comprehensive applicability in clinical settings.

# 3 Proposed Methodology

An XAI model is like a friendly AI that explains how it makes decisions, unlike other complicated ones that keep their thinking hidden. It shows its work, so we can understand why it does what it does. It uses tricks to help us see which parts of the information it uses are most important, or it might give explanations that make sense to people. This helps us trust it more and even improve how it works.
Like any other model, XAI also requires data cleaning as well as preprocessing(mostly image data). Then the data is divided into training,testing as well as validation dataset which is trained on our basic CNN model over a sequence of iterations. The XAI model in integration with the CNN model will give us the required output.

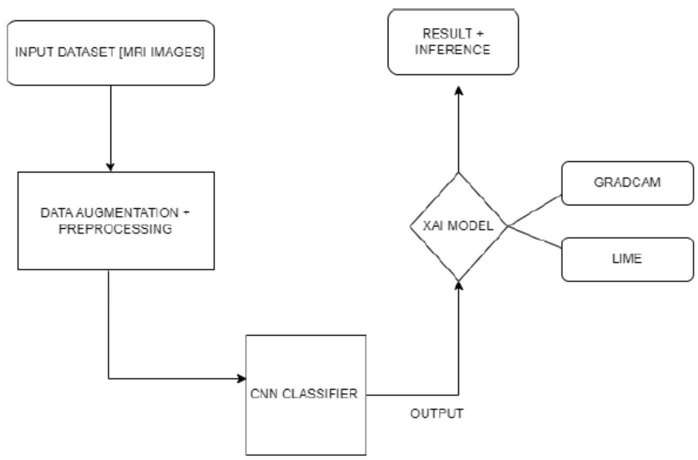


**Figure 1:** *methodology*

## 3.1 LIME

LIME (Local Interpretable Model-agnostic Explanations) is an eXplainable AI (XAI) technique designed to provide insights into the decision-making process of complex machine learning models. It operates by approximating the behavior of the black-box model by creating interpretable, locally faithful surrogate models for specific predictions. LIME works by perturbing input data samples and observing how the model's predictions change, generating explanations at the local level. It identifies relevant features impacting a specific prediction, allowing users to understand why the model made a particular decision. This high-level design enables LIME to offer transparent and understandable explanations for black-box model predictions, enhancing trust and interpretability in AI systems across various domains.

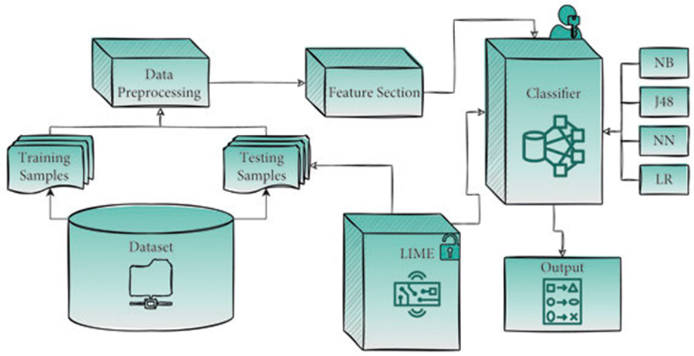


**Figure 2:** *high level design*

### 3.2 GRADCAM

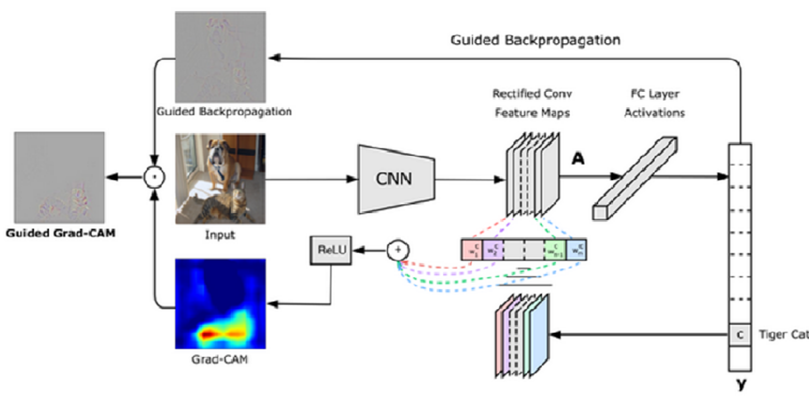


**Figure 3:** *dataflow diagram*

Grad-CAM (Gradient-weighted Class Activation Mapping) is an eXplainable AI (XAI) technique devised to elucidate the decision-making process of convolutional neural networks (CNNs). It functions by utilizing the gradient information flowing into the final convolutional layer to identify important regions within an image that heavily influence the model's prediction. Grad-CAM generates a heatmap that highlights these discriminative regions, showcasing where the model focused its attention while making a specific classification. By visualizing these areas, Grad-CAM provides insights into the CNN's decision, offering interpretable explanations regarding which parts of an input image were crucial for the model's classification, thus enhancing the transparency and interpretability of deep learning models.

### 3.3 Algorithms

Effectively using eXplainable AI (XAI) requires understanding machine learning, data, programming skills, knowledge of XAI techniques, awareness of interpretability-trade-offs, ethical considerations, and domain expertise for accurate model interpretation and decision-making.

**Algorithm 1** Pseudocode for LIME

```
Input x: Input data sample (instance to be explained), f: Black-box model to
  be explained, K: Number of perturbed samples for explanations, L: Interpretation
  complexity (e.g., number of features)
Output : Local surrogate model (interpretable representation of f)
Retrieve the coefficient values from the database for the given input subset of
  features
for each record in the features' coefficient values do
    f(x):prediction of black box model f for the instance x
    g(x'):Prediction of the local surrogate model g for interpretable representation
  x of instance x
    W: Weights in g representing feature importance
end for
```

**Algorithm 2** Pseudocode for GRADCAM

```
Input Input image I, Pre-trained convolutional neural network (CNN) model M,
  Target class index c (the class for which we want to visualize important regions)
Output Heatmap H highlighting important regions of the input image
Retrieve the coefficient values from the database for the given input subset of
  features
for each record in the features' coefficient values do
    Calculate the importance weights for each channel k in the final convolutional
  layer through Global Average Pooling (GAP)
    Generate the heatmap H by combining the importance weights and final con-
  volutional layer activations
end for
```

# 4 Experimental Results

## 4.1 Experimental Setup

Dataset used – Custom Medical Video MRI Dataset comprising of two folders of manually segmented images [I.E.- VM/NOT-VM]

The types of experiments done – LIME-XAI / GRADCAM-XAI

## 4.2 CNN and LIME

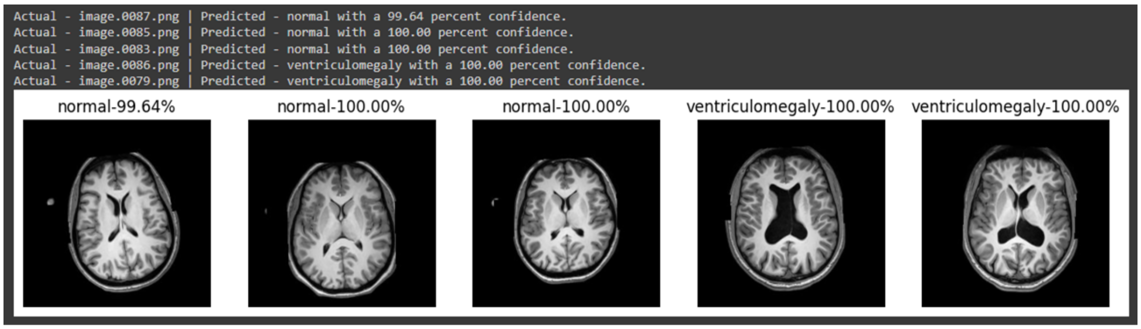


**Figure 4:** *CNN classification*

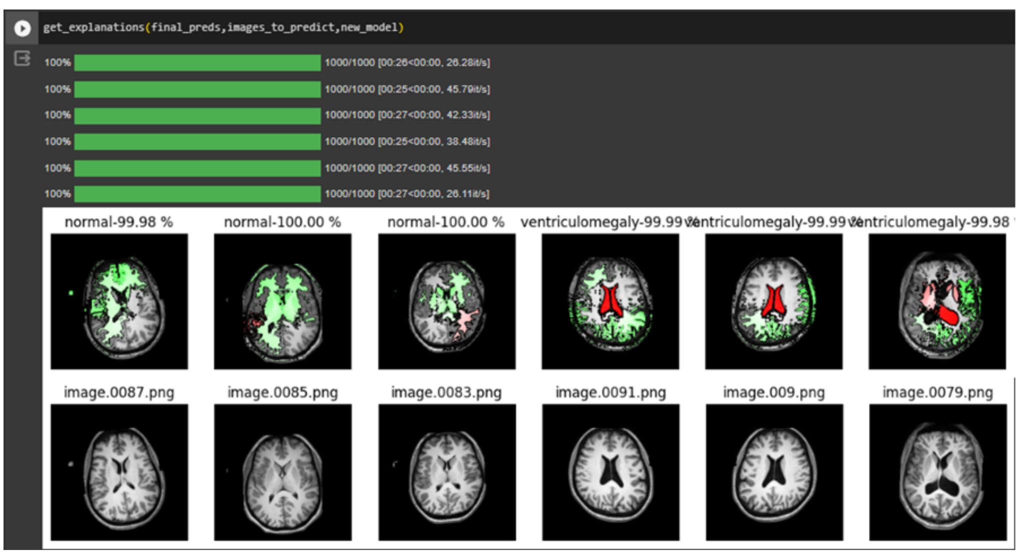


**Figure 5:** *LIME model*

The first, second and third images are classified as normal and correctly so because the ventricles are normal in size. In LIME, the green highlighted area indicates that the certain feature contributes positively towards certain class or a decision. Because there is no visible swollen behaviour shown by the ventricles, the cerebrospinal fluid is present in the correct amount, not less neither excess. Due to this the white matter around the ventricles cover most of the region around the ventricles and the white matter seems to contribute towards the normal classification of the image. As can be seen in the image that the white matter around the ventricles is highlighted in green.

This fourth, fifth, and sixth images are an example of a case of ventriculomegaly, due to the way the model has been trained, the white matter around the ventricles always contributes positively towards normal classification hence the white matter is by default highlighted in green. Apart from this, the model spots a difference in ventricle size and shape from that of the normally classified images. The ventricles have been swollen due to the excess cerebrospinal fluid present in the brain. The swollen ventricles are a clear indication that the given image is a case of ventriculomegaly and hence it is highlighted as red i.e. it negatively contributes towards classifying it as a normal condition.

## 4.3 GRADCAM

The heatmap generated by Grad-CAM is based on the importance weights and the activations of the final convolutional layer. The heatmap highlights specific regions in the brain MRI that the model identified as critical for its decision-making process regarding ventriculomegaly. These regions effectively illustrate where the model focused its attention, aiding in understanding the MRI features contributing to the classification.

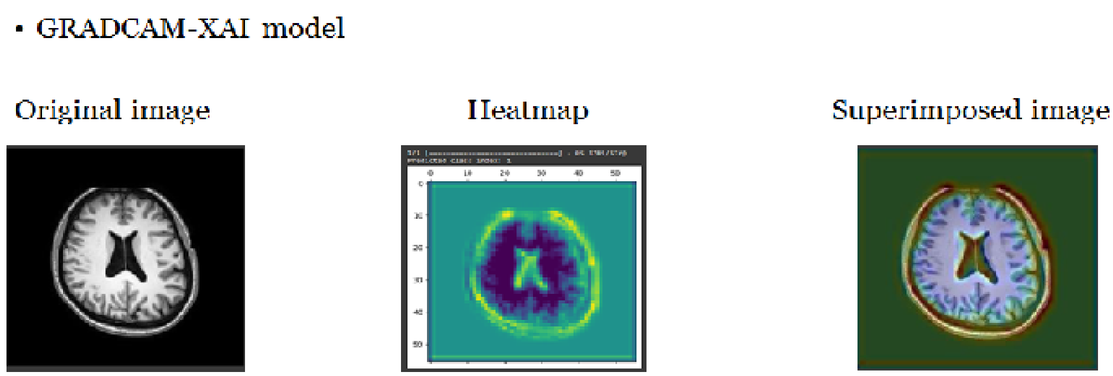


**Figure 6:** *GRADCAM 1*

The given example is a case of ventriculomegaly and we can clearly see that the white matter is mildly resposible for any of the decisions that is made(unless it is classified as normal). apart from that, we can see that the swollen ventricles in both the images are intensly highlighted in the heatmap, specially towards the edges. This is where the difference is present between the normal and the ventriculomegaly cases. Hence we see the output. A similar exampl is given below with a more dense case of ventriculomegaly.

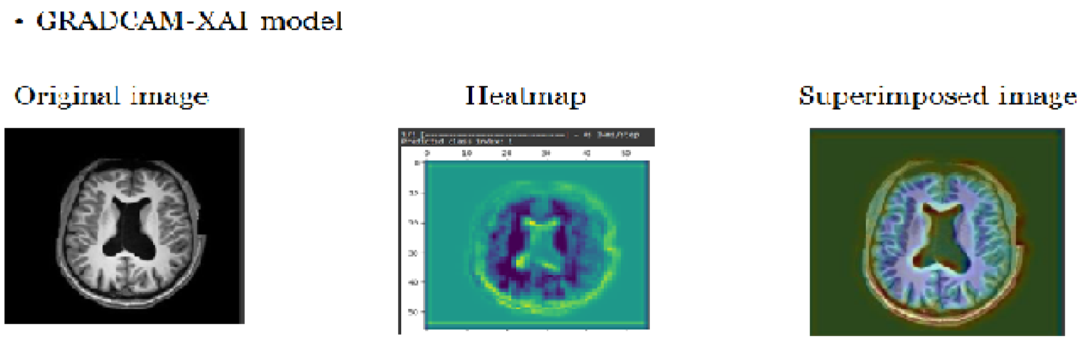


**Figure 7:** *GRADCAM 2*

# 5 Conclusions

The research focused on creating a Convolutional Neural Network (CNN) specifically tailored for fetal ventriculomegaly classification using ultrasound images. It involved extensive experimentation, encompassing model development, integration of explainable AI frameworks (LIME and Grad-CAM), hyperparameter optimization, transfer learning, and real-world implementation. Key findings included the successful development of a CNN model for accurate classification, establishing performance baselines, and integrating LIME and Grad-CAM to provide transparent insights into the model's decision-making process.
The contributions comprised a novel CNN model for precise classification and the integration of explainable AI to enhance interpretability. Future goals encompass enhancing model generalization, continuous improvement through feedback, multi-modal integration, and extending explainability through advanced XAI techniques. Overall, the research not only introduced a robust CNN model but also laid the groundwork for practical integration into prenatal care, outlining future paths for clinical enhancements and advancements.
future scope also involves getting improving our classification model by measuring the sizes of ventricles both laterally and vertically, which will result in higher accuracy while classification and a more explainable way to interpret the result using XAI.

# References

• Xie HN, Wang N, He M, Zhang LH, Cai HM, Xian JB, Lin MF, Zheng J, Yang YZ. Using deep-learning algorithms to classify fetal brain ultrasound images as normal or abnormal. Ultrasound Obstet Gynecol. 2020 Oct;56(4):579-587. doi: 10.1002/uog.21967. PMID: 31909548.
• Vahedifard, F.; Ai, H.A.; Supanich, M.P.; Marathu, K.K.; Liu, X.; Kocak, M.; Ansari, S.M.; Akyuz, M.; Adepoju, J.O.; Adler, S.; et al. Automatic Ventriculomegaly Detection in Fetal Brain MRI: A Step-by-Step Deep Learning Model for Novel 2D-3D Linear Measurements. Diagnostics 2023, 13, 2355. https://doi.org/10.3390/diagnostics13142355

• Sira LB, Kozyrev DA, Bashat DB, Constantini S, Roth J, Shiran SI. Fetal Ventriculomegaly and Hydrocephalus – What Shouldn't be Missed on Imaging?. Neurol India 2021;69, Suppl S2:298-304
• N. Siddique, S. Paheding, C. P. Elkin and V. Devabhaktuni, "U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications," in IEEE Access, vol. 9, pp. 82031-82057, 2021, doi: 10.1109/ACCESS.2021.3086020.
• O. Attallah, H. Gadelkarim and M. A. Sharkas, "Detecting and Classifying Fetal Brain Abnormalities Using Machine Learning Techniques," 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, USA, 2018, pp. 1371-1376, doi: 10.1109/ICMLA.2018.00223