**Title: A comprehensive testing framework for deep learning models**

**Abstract:**

Deep learning models are now widely recognized as game-changers in critical fields such as autonomous vehicles and healthcare diagnostics. With their ability to provide accurate and reliable results, these models hold the potential to revolutionize these industries completely. As we head towards a more automated future, the significance of these models cannot be ignored. They represent a significant advancement in technology and have already proven their effectiveness. For their reliable deployment in real-world scenarios, these models must be robust on both a global and local scale. Traditional testing methods do not differentiate between various levels of robustness, leading to overlooked critical errors in model behavior. Therefore, adopting alternative testing approaches to identify such errors is crucial.

In my research, I developed a comprehensive testing approach for deep learning models that evaluates global and local robustness. This methodology investigates how models react to various perturbations, such as changes in visual inputs, to showcase their ability to handle challenges or weaknesses. For instance, a model may show excellent overall accuracy, but its performance may decline significantly when presented with rotated or partially hidden images, indicating a lack of local robustness.In addition, this framework includes a unique error summarization module that can detect and classify model errors, enabling the identification of patterns and forecasting of potential issues. This thorough error analysis is crucial in identifying particular categories or cases where the model's dependability declines, offering practical insights on how to improve these weaknesses.

The ability to analyze model performance at both local and global levels, along with the use of advanced error summaries, sets this approach apart and drives the field towards more fine and effective evaluations. The practical implications of this research are significant, as it enhances the safety, reliability, and trustworthiness of automated systems by ensuring that they are robust enough to handle the complexities of the real world and safeguard human lives across various applications. While the current methodology uses a uniform method for sampling data to produce test cases, future research will aim to improve the efficiency and effectiveness of test case generation by exploring strategic sampling strategies.

Arooj