# Explainable Multimodal Deep Learning for Skin Lesion Classification Using Image and Metadata Fusion

**Ophelia Frimpong**, *Computational Science Program, University of Texas at El Paso*
**Maria C. Mariani**, *Department of Mathematical Sciences, University of Texas at El Paso*
**Osei K. Tweneboah,** *School of Theoretical and Applied Science, Ramapo College of New Jersey*

In the United States, diagnostic errors affect approximately 12 million adults each year, with nearly half potentially leading to harm (Newman-Toker et al., 2014; Johns Hopkins Medicine, 2014). According to the National Cancer Institute (2020), skin cancer treatment costs the U.S. about $8.1 billion annually, $4.8 billion for non-melanoma and $3.3 billion for melanoma.

This study aims to develop an explainable multimodal deep learning model that optimizes the accuracy of classifying skin lesions. The objective is to enhance patient care and outcomes by bridging the gap between the human understanding and the machine in terms of trust and accountability. We are using the Human Against Machine (HAM 10000) dataset from the Department of Dermatology at the Medical University of Vienna in Austria, which consists of 10,015 images of pigmented skin lesions and metadata describing patient's age, sex and lesion location.

In this work, image features are learned through a Convolutional Neural Network (CNN), while clinical metadata are modeled with a Multilayer Perceptron (MLP). The two feature sets are then merged to form a single, integrated representation. Explainable Artificial Intelligence (XAI) techniques, including Gradient-weighted Class Activation Mapping (Grad-CAM) and Shapley Additive explanations (SHAP), are used to visualize and interpret model predictions. This work aims to support the development of transparent and reliable diagnostic tools that can assist clinicians in early skin-cancer detection.