Multimodal for Stroke Detection

Michael Dang
School of Science and Engineering
University of Missouri - Kansas City
Email: ld8np@umsystem.edu

Chaitanya Krishna Sairam Padamata
School of Science and Engineering
University of Missouri - Kansas City
Email: cp3nr@umsystem.edu

Karthik Chellamuthu
School of Science and Engineering
University of Missouri - Kansas City
Email: karthikc.1729@gmail.com

***Abstract*—Stroke is a leading cause of morbidity and mortality worldwide, necessitating early and accurate detection for effective intervention. This project leverages advanced machine learning techniques to develop a robust stroke detection model. By integrating data preprocessing, feature engineering, and ensemble learning methods, the system aims to predict stroke risk with high precision and recall. The model is deployed using a full-stack architecture, providing an intuitive interface for clinicians and researchers. This work demonstrates the potential of AI in healthcare, showcasing how data-driven approaches can enhance diagnostic accuracy and decision-making.**

I. Introduction

Strokes represent a critical public health challenge, accounting for significant disability and mortality rates globally [1]. The timely detection of stroke risk is essential for initiating preventive measures and improving patient outcomes. Traditional diagnostic methods, while effective, are often time-consuming and require significant expertise.

Recent advancements in artificial intelligence and machine learning have opened new avenues for automating and augmenting medical diagnostics [2]. This project explores the application of supervised learning algorithms to predict stroke occurrence based on demographic and clinical data. The dataset, sourced from reliable medical repositories, undergoes rigorous preprocessing and feature selection to enhance the model's performance.

In addition to model training and evaluation, the project focuses on end-to-end deployment using a modern web-based interface. This ensures accessibility for healthcare professionals, enabling real-time predictions. By bridging the gap between AI research and practical application, this project aims to contribute meaningfully to the field of stroke prevention and healthcare analytics.

II. Body

In this section, we provide a detailed overview of our platform's foundational components. This includes the system architecture, the datasets utilized, and the machine learning models that power the stroke detection and response functionalities. Our platform leverages a multi-modal approach, combining various data sources and cutting-edge machine-learning techniques to deliver accurate predictions and actionable insights.

A. Dataset

- The dataset comprises **5,029** images evenly distributed across two classes: acute stroke and non-stroke cases. To ensure the model performs well in diverse and real-world scenarios, various data augmentation techniques were employed. These techniques included horizontal and vertical flipping, random rotation, scaling, cropping, and brightness adjustments. Such augmentations help mitigate over-fitting by exposing the model to a broader range of variations within the same class, effectively simulating the variability encountered in real-life medical imaging.

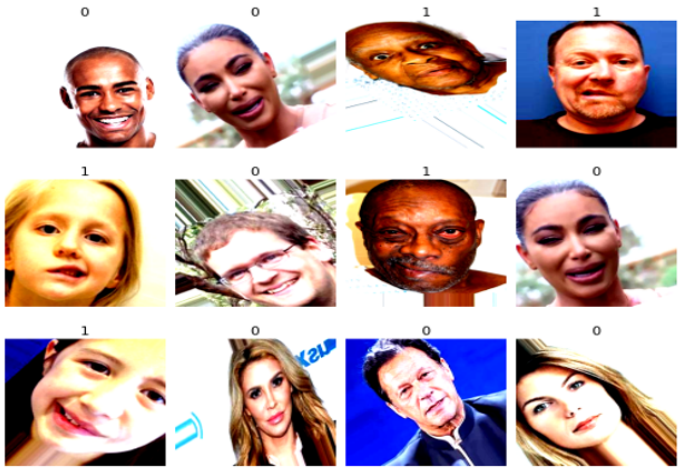

Fig. 1. Snapshot of the dataset.

- A second dataset contains **5,110** observations with 12 features, designed to predict whether a patient is likely to have a stroke. Input features include demographic and clinical parameters such as gender, age, history of hypertension or heart disease, smoking status, and other relevant variables. As Table 1 describes, this dataset has categorical variables and continuous variables. However, there is an imbalanced data distribution in the stroke case shown in Figure 2. To address this problem we applied SMOTE techniques to balance out the dataset shown in Figure 3.

Categorical variables	Continuous variables
1. Gender	1. Age
2. Hypertension	2. ID (drop)
3. Heart disease	3. BMI
4. Ever married (drop)	4. Average glucose level
5. Work type (drop)	
6. Residence type (drop)	
7. Smoking status	
8. Stroke	

TABLE I
CATEGORICAL AND CONTINUOUS VARIABLES.

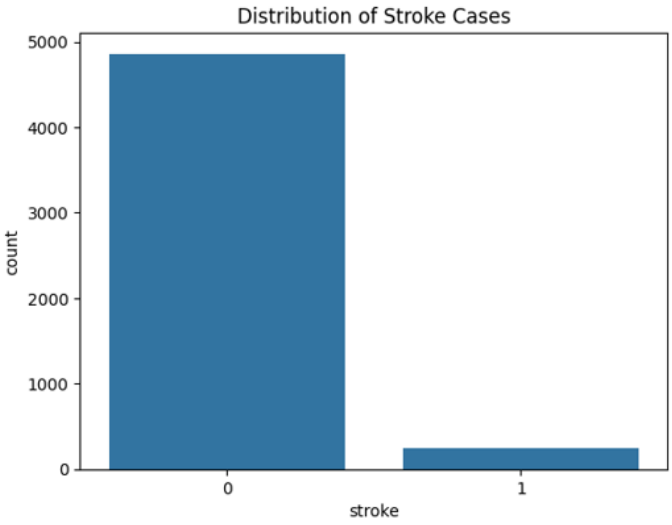

Fig. 2. Distribution of stroke cases.

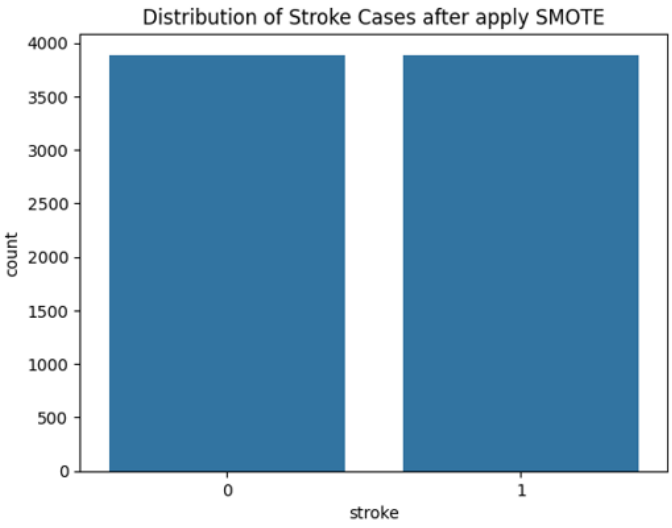

Fig. 3. Distribution of stroke case.

B. Architecture

- The architecture integrates various components to create a robust and efficient system. The frontend uses Gradio, an interactive tool for building user interfaces for machine learning models. Gradio simplifies deployment by providing a web-based platform where users can input data, such as images or medical history, and receive predictions in real-time. It bridges the gap between complex backend operations and the user experience.
- The backend, implemented in JavaScript with frameworks like Express.js, acts as the central hub connecting the frontend, database, and models. It handles user requests, processes data, and ensures smooth communication between components. The database, powered by MongoDB, stores relevant data such as user inputs, predictions, and historical records for further analysis or system improvement.
- Model training is conducted using PyTorch and Scikit-learn, two widely used frameworks for machine learning and deep learning. PyTorch is particularly suitable for training the EfficientNet and other deep learning models, while Scikit-learn supports traditional machine learning methods like Random Forest and Gradient Boosting. Pretrained models, such as Meta's Llama 3, enhance the system by providing a multi-modal large language model to interpret complex user queries and generate recommendations.
- This architecture supports a cohesive and scalable pipeline, where integrating advanced technologies ensures accurate predictions, efficient data processing, and an accessible user interface.

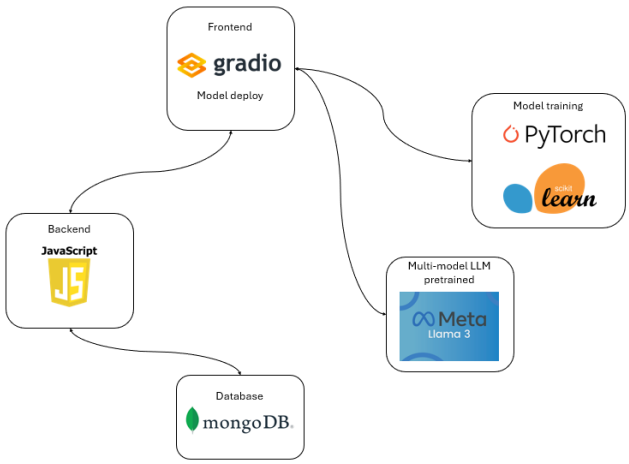

Fig. 4. Architecture of the platform.

C. Model

For facial stroke detection, the architecture incorporates advanced deep learning models including ResNet34, ResNet50, EfficientNet, and ConvNeXt. Each of these models offers unique strengths in feature extraction and classification.

- ResNet34 and ResNet50, part of the ResNet family, leverage residual connections to tackle the vanishing gradient problem, making them highly effective for deep neural networks. ResNet50, with its deeper architecture,

is particularly adept at capturing complex patterns in facial images. [3]
- EfficientNet focuses on scaling model size efficiently, balancing depth, width, and resolution to achieve high accuracy with fewer computational resources. [4]
- ConvNeXt, a modernized convolutional neural network, combines the simplicity of traditional CNNs with advanced architectural designs inspired by transformer models, making it highly efficient for image classification tasks. [5]

By using these models, the system ensures robust and accurate predictions, leveraging the strengths of diverse architectures for facial stroke detection.

The system incorporates machine learning algorithms such as Random Forest, Gradient Boosting, XGBoost, and Support Vector Machines (SVM) to predict strokes based on medical history.

- Random Forest, a robust ensemble learning method, excels in handling imbalanced datasets and capturing complex relationships between features by combining multiple decision trees. [6]
- Gradient Boosting and XGBoost, both popular boosting algorithms, iteratively refine predictions by minimizing errors, making them highly effective for structured medical history data. XGBoost, in particular, is known for its speed and efficiency, leveraging parallel processing and optimized tree-building techniques. [7] [8]
- Support Vector Machines (SVM) provide strong performance for classification tasks by identifying optimal decision boundaries, especially when working with high-dimensional feature spaces. [9]

These algorithms complement one another by offering diverse strengths, ensuring reliable and accurate predictions for medical history-based stroke risk assessment.

D. Result

The results from the experiments highlight the performance of various models for stroke prediction, leveraging both facial image data and medical history data. The first table focuses on the performance of CNN architectures, including ResNet34, ResNet50, EfficientNet, and ConvNeXt, for facial stroke prediction under two learning rates (0.001 and 0.01). Metrics such as accuracy and the number of iterations until convergence were evaluated to identify the most efficient and accurate architecture. The second table summarizes the results of machine learning algorithms, such as Random Forest, Gradient Boosting, XGBoost, and SVM, for medical history-based stroke prediction, showcasing their accuracy, precision, recall, F1-score, and ROC-AUC. These tables provide insights into the suitability of different models for each modality, revealing the potential of combining them for a comprehensive multimodal stroke prediction system.

These results highlight the strengths of using CNNs for image-based stroke prediction, with EfficientNet standing out for its efficiency and accuracy, and the superiority of Gradient Boosting for structured medical history data. Together, these models provide a comprehensive framework for accurate and robust multimodal stroke prediction.

TABLE II
PERFORMANCE OF DIFFERENT MODELS FOR FACIAL STROKE PREDICTION

Model	LR:0.001	LR:0.01	# iteration(0.001)	# iteration(0.01)
ResNet34	98.09%	97.74%	30	24
ResNet50	99.48%	93.02%	34	39
EfficientNet	**99.83%**	**90.8%**	**21**	29
ConvNeXt	99.48%	91.15%	43	**20**

For **facial stroke prediction**, EfficientNet outperformed all other CNN architectures with an accuracy of **99.83%** and converged in only **21 iterations** at a learning rate of 0.001. ResNet50 and ConvNeXt also achieved high accuracies of **99.48%**, but ResNet50 required more iterations (**34**) compared to ConvNeXt (**43**) at the same learning rate. Interestingly, when the learning rate was increased to 0.01, the performance of all models except ResNet34 dropped significantly, highlighting the sensitivity of these models to hyperparameter tuning. EfficientNet and ConvNeXt demonstrated the fastest convergence, requiring the fewest iterations overall.

TABLE III
PERFORMANCE OF DIFFERENT MODELS FOR MEDICAL HISTORY STROKE PREDICTION

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
RandomForest	89.56%	90.1%	89.58%	89.53%	89.57%
GradientBoosting	**96.5%**	**96.53%**	**96.5%**	**96.5%**	**96.5%**
XGBoosting	94.19%	94.21%	94.19%	94.19%	94.19%
SVM	79.38%	79.62%	79.39%	79.35%	87.75%

For **medical history-based stroke prediction**, Gradient Boosting achieved the highest performance across all metrics, with accuracy, precision, recall, F1-score, and ROC-AUC of **96.5%**, showcasing its effectiveness in handling structured data. XGBoost closely followed with an accuracy of **94.19%**, while Random Forest achieved a slightly lower accuracy of **89.56%**, reflecting its robustness but less optimization for this specific task. SVM, while achieving moderate precision and recall (79.4%), lagged behind the ensemble methods in overall performance, underscoring its limitations when compared to tree-based methods for imbalanced and complex datasets.

To enhance the accuracy and reliability of stroke prediction, we combined the two best-performing models: EfficientNet for facial stroke prediction and Gradient Boosting for medical history-based stroke prediction. This multimodal approach leverages the strengths of both models, assigning weighted importance of **60%** to the facial prediction and **40%** to the medical history prediction. The final prediction (P_{final}) is computed as a weighted sum of the predictions from these models, as shown in the equation:

$$P_{final} = 0.6 \cdot P_{EfficientNet} + 0.4 \cdot P_{GradientBoosting}$$

By prioritizing facial data, which demonstrated higher individual accuracy, and supplementing it with critical contextual insights from medical history, this hybrid strategy improves the overall robustness of the system. This integration allows the model to handle diverse scenarios effectively, ensuring a balanced and comprehensive approach to stroke detection

while minimizing the limitations of relying solely on a single data source.

We incorporated the Llama 3-8B Multimodal [10] model as an integral component of our stroke detection framework. The Llama 3-8B is a state-of-the-art large language model with multimodal capabilities, enabling it to process and analyze text and image data simultaneously. This feature made it highly suitable for integrating facial image predictions and medical history data, ensuring seamless communication between different modalities. By leveraging its advanced contextual understanding and inference capabilities, the model provided not only robust predictions but also valuable insights into the decision-making process. Furthermore, its pretraining on diverse datasets allowed it to generalize effectively, complementing specialized models like EfficientNet and Gradient Boosting. The inclusion of Llama 3-8B enhanced the interpretability of results, bridging the gap between data-driven predictions and actionable insights, and demonstrated the potential of multimodal LLMs in healthcare applications.

III. Discusion

This study explores the application of deep learning for stroke detection using standard camera images that analyze facial cues. Although not intended as a medical-grade diagnostic tool, the approach demonstrates the significant potential of multimodal integration to improve early stroke response. By combining facial analysis with additional contextual data, such as basic patient medical history, environmental context, or even real-time video inputs, the system can identify potential stroke markers and trigger immediate actions, such as alerts to healthcare providers or emergency responders.

This multimodal framework enables rapid screening, especially in settings where advanced medical equipment is unavailable. For example, early detection through facial analysis could prove vital in pre-hospital environments, community health initiatives, or rural regions lacking advanced diagnostic infrastructure. While not replacing comprehensive clinical evaluation, such methods act as effective proxies, bridging critical gaps in early stroke detection and ensuring timely intervention.

The benefits of this approach are profound. Swift identification of stroke-like symptoms facilitates faster access to emergency medical care, which is critical for conditions like stroke, where every minute of delay significantly impacts survival rates and long-term outcomes. By reducing the time to treatment, this framework can lower the risk of severe disability, improve recovery chances, and optimize resource allocation for healthcare systems.

While the study demonstrates feasibility, further work is necessary to refine the method. Enhancements such as diverse training datasets, improved interpretability of results, and integration with patient-centered data streams will be pivotal in optimizing performance and usability. Nevertheless, this system lays the groundwork for a low-cost, scalable tool that complements traditional diagnostic workflows and broadens access to life-saving stroke interventions.

IV. Conclusion

The integration of facial analysis and multimodal data presents a promising approach for early stroke detection in non-clinical settings. Though not diagnostic, such methods provide a valuable proxy that supports quicker emergency responses, significantly improving patient outcomes. By reducing time to treatment and offering a scalable solution for diverse settings, this technology has the potential to transform pre-hospital stroke care and enhance public health initiatives worldwide.

V. Acknowledgment

We extend our deepest gratitude to **Dr. Yugyung Lee** (leeyu@umkc.edu) for her invaluable guidance, supervision, and encouragement throughout the duration of this project. Her expertise and insights were instrumental in shaping the direction of our research and ensuring its successful completion.

VI. Contribution

This project was a collaborative effort with contributions from all team members, led primarily by Michael Dang, who contributed **75%** of the work. Michael was responsible for both the frontend and backend development, including connecting the Gradio-based user interface to the JavaScript-based backend, integrating the database, and deploying the models. He also took the lead in finding and preprocessing the datasets, training the machine learning models, and managing the database setup. Additionally, Michael prepared the PowerPoint presentation and contributed extensively to the paper writing. Chaitanya Krishna Sairam Padamata contributed **20%**, assisting in the training of machine learning models, tuning hyperparameters, and contributing significantly to writing the paper. Karthik Chellamuthu contributed **5%**, focusing on preparing the PowerPoint presentation and providing feedback for the presentation delivery. This distribution of contributions ensured a balanced team effort, with each member playing a key role in achieving the project goals.

References

[1] C. D. Wolfe, "The impact of stroke," *British medical bulletin*, vol. 56, no. 2, pp. 275–286, 2000.

[2] S. Rahman, S. Ibtisum, E. Bazgir, and T. Barai, "The significance of machine learning in clinical disease diagnosis: A review," *arXiv preprint arXiv:2310.16978*, 2023.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[4] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*. PMLR, 2019, pp. 6105–6114.

[5] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 11 976–11 986.

[6] L. Breiman, "Random forests," *Machine learning*, vol. 45, pp. 5–32, 2001.

[7] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp. 1189–1232, 2001.

[8] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[9] C. Cortes, "Support-vector networks," *Machine Learning*, 1995.
[10] A. Dubey, A. Jauhri, A. Pandey, A. Kadian, A. Al-Dahle, A. Letman, A. Mathur, A. Schelten, A. Yang, A. Fan *et al.*, "The llama 3 herd of models," *arXiv preprint arXiv:2407.21783*, 2024.
[11] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.

APPENDIX

This study utilized the Synthetic Minority Oversampling Technique (SMOTE) to address class imbalance in the medical history dataset. SMOTE generates synthetic samples for the minority class by interpolating between existing data points. Specifically, for each minority class instance, SMOTE identifies its nearest neighbors in the feature space and creates new synthetic data points along the line segments connecting these neighbors.

By balancing the dataset, SMOTE reduces bias in the predictions, improves recall for the minority class, and enhances the overall F1-score, ensuring better detection of stroke cases. The application of SMOTE significantly improved the model's performance, as reflected in increased recall and F1-scores for stroke prediction.

The following hyperparameters were optimized to enhance model performance:

A. EfficientNet

- Learning rate: 0.001
- Batch size: 32
- Number of epochs: 50

B. Gradient Boosting

- Number of estimators: 100
- Learning rate: 0.1
- Maximum depth: 3

The optimal hyperparameters were determined using grid search and cross-validation techniques.

C. SHAP (SHapley Additive exPlanations)

SHAP values were used to interpret the Gradient Boosting model. Key features such as `age` and `average glucose level` had the highest impact on predictions. SHAP visualizations provided feature importance values, enhancing the model's transparency and trustworthiness.

D. Grad-CAM (Gradient-weighted Class Activation Mapping)

For facial stroke detection, Grad-CAM was applied to convolutional layers of EfficientNet. Saliency maps generated by Grad-CAM highlighted areas such as the eyes and mouth, correlating with clinical symptoms like drooping or asymmetry. These visualizations improved the interpretability of the deep learning model. [11]

The project adheres to data privacy standards, including GDPR and HIPAA, as follows:

E. General Data Protection Regulation (GDPR)

- All patient data was anonymized to ensure no personal information was identifiable.
- Assumed user consent for datasets used in public repositories.

F. Health Insurance Portability and Accountability Act (HIPAA)

- Patient data was securely stored using encryption methods.
- Data minimization principles were followed to restrict the use of unnecessary patient attributes.
- Access to data was limited to authorized personnel.

By addressing these privacy and compliance standards, the project ensures ethical and secure handling of sensitive data.