AI to Detect Stroke-Related Facial Weakness

Dr.N.Neelima Priyanka
Department of CSE
(professor)
PSCMR College of Engineering, Vijayawada, AP
priyanka.nutulapati@gmail.com

Nandaluru Mythili
Department of CSE
(student)
PSCMR College of Engineering, Vijayawada, AP
mythilinandaluru@gmail.com

Shilam Sai Bhargavi
Department of CSE
(student)
PSCMR College of Engineering, Vijayawada, AP
shilamsaibhargavi@gmail.com

Vemuri Karthikeya
Department of CSE
(student)
PSCMR College of Engineering, Vijayawada, AP
karthikeyavenuri007@gmail.com

Idumalla Uday Kumar
Department of CSE
(student)
PSCMR College of Engineering,Vijayawada,AP
idumallaudayudaykumar@gmail.com

Abstract*— Stroke is a serious medical condition caused by the interruption or reduction of blood flow to the brain, often resulting in facial weakness or disability. Early detection of stroke symptoms, particularly facial asymmetry, enables timely medical intervention and improves patient outcomes. This paper proposes an artificial intelligence-based approach that utilizes a limited set of neutral and smiling facial images to detect stroke-related facial weakness. To address data scarcity, a hybrid dataset is developed by combining real images with synthetic data generated using FaceGAN and deepfake-based augmentation techniques. Facial regions are segmented into left and right halves using landmark detection, followed by affine transformations based on Delaunay triangulation for geometric alignment. A deep learning architecture integrating a ConvNeXt encoder with a lightweight convolutional neural network (CNN) decoder is employed for feature extraction and classification. The model demonstrates strong robustness and achieves an accuracy of 98.9% using four-fold cross-validation. The results highlight the potential of AI-assisted facial analysis for rapid and accessible stroke screening in clinical and resource-limited environments.

***Keywords*— Stroke detection, Facial weakness, Artificial intelligence, Deep learning, FaceGAN, Deepfake, Lightweight CNN, ConvNeXt**

I. Introduction

Facial weakness or partial paralysis due to strokes is one of the top causes of death and disability. Because facial drooping can be an indication of a problem, early detection of facial imbalance is important for early diagnosis [1]. However, many areas are experiencing delays in diagnosis because of a lack of access to modern imaging or specialists. This is where computer vision and AI come in, and they offer a promising solution for automating this process. Even in areas with limited resources, AI can help doctors quickly evaluate facial imbalance using deep learning to help diagnose strokes [2]. Facial analysis technology may accelerate the time taken to recognize stroke symptoms, thus providing quicker treatment and improved survival rates [3]. Conventionally, stroke diagnosis involves a physician's assessment and imaging studies such as CT or MRI scans. These methods are often expensive and may not be accessible in remote or developing regions [4]. Facial weakness is a visible and easily verifiable sign for early detection, but it is susceptible to human error.

The majority of current AI models for stroke detection require large medical datasets, which are difficult to access, or they involve the analysis of brain images. Since the number of medical datasets is limited, and because of the difficulties in accessing images with subtle imbalances, there is still a lack of studies specifically focused on the recognition of facial weakness. Moreover, when tested on small datasets, current models are prone to overfitting, resulting in lower performance on new individuals [5]. Researching the fusion of synthetic and real facial information to enhance model robustness is still an area that has potential for growth. To overcome this, we require an AI model that can learn well from small amounts of data and remain accurate and flexible for practical applications. In our research, from only a few facial images, we propose a novel AI approach to identify facial vulnerabilities associated with stroke.

II. RELATED WORK

The early literature on analysis of facial weakness and stroke detection can be divided into a number of key areas. One line of research is based on medical imaging, such as CT and MRI scans, which directly reveal pathology of the brain but need expensive infrastructure and expertise. This sets the first major track of research [6]. The next track is based on computer vision analysis of facial images. In the earlier research, the analysis was based on manually designed geometric or textural features and facial points that were input to traditional machine learning classifiers. More recent research uses deep convolutional neural networks that learn features suited to particular tasks, such as imbalance or facial hanging detection, from facial images or video frames. Some research also investigates the analysis of dynamic imbalance in facial expressions over time using video [7].

To enhance the limited clinical data and make the models more robust, the field of synthetic data creation and augmentation, including generative adversarial networks, is becoming more active. Wang [8] states that stroke is a serious health risk, especially for seniors. The lack of access to healthcare and the cost of imaging make it difficult to use traditional methods of diagnosis, such as CT and MRI scans. A convolutional neural network was created to identify stroke using 2D facial photographs. This AI tool can identify and even predict stroke before a CT scan, which could alleviate the workload of emergency rooms. The AUC was 0.91.

However, the existing approaches have several known limitations. They cannot facilitate fast community-level screening with medical imaging because these approaches require a lot of resources [9]. Traditional systems based on landmarks are prone to posture, lighting, closure, and landmark variability, which are some factors that can affect the performance of these approaches. Deep learning models have limitations related to class imbalance and small datasets,

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 04 | April - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

which can cause overfitting and poor generalization performance on new individuals or environments [10]. If the clinical variability is not accounted for in the synthesized images, then the image synthesis approach can lead to domain shift. The existing approaches have several limitations regarding clinical-grade performance and the output required for real-world applications. Finally, there are some privacy, integrity, and clinical validation issues.

We address the problem by stacking multiple effective approaches. We concentrate on simple face image analysis that is highly effective for fast screening, rather than brain scans or manually designed features. By combining actual clinical images with FaceGAN-generated pairs of neutral and smiling faces across ages, we improve generalization and address data scarcity issues, while preventing overfitting a significant improvement over previous deep learning approaches trained end-to-end on small datasets [11]. Preprocessing leverages accurate landmark detection to divide faces into left and right halves, as well as affine transformations obtained from Delaunay triangulation to compensate for geometric distortions. This provides better alignment and reduced variability compared to standard bounding-box cropping. To assess the model, our network architecture is a ConvNeXt-encoded autoencoder combined with a ConvNet decoder is extremely effective for representation learning and fine-tuning for classification tasks, outperforming several popular transfer learning baselines [12]. Crucially, unlike most previous, less robust models, our approach is robust: four-fold cross-validation accuracy is 98.9%, clearly indicating its applicability.

III. DATA COLLECTION & PREPROCESSING

The data set for this research is sourced from two primary sources: an artificial data set created using FaceGAN and a real data set of face images [13]. There are a total of 1,200 images of happy and neutral faces from both stroke survivors and healthy people. To offset the small number of real-world examples, an additional 3,000 images were created using FaceGAN. The FaceGAN component aims to be as inclusive and representative as possible, creating a set of pairs of neutral and smiling images across eight age ranges, from young children to older adults [14]. This hybrid strategy combines the best of both worlds, using both real and artificial facial features, and improves the model's robustness and adaptability to real-world stroke detection scenarios by being exposed to a much wider variety of facial expressions [15]. Each image is saved in RGB format at 224 pixels by 224 pixels.

Facial expressions and diagnosis were employed to categorize the photos into two groups: normal and stroke patients. The images include happy and neutral expressions, which are essential for identifying stroke patients' facial weakness, as asymmetry on one side may appear as different muscle activation during smiling [16]. The first step in image preprocessing was the removal of duplicate and poor-quality images due to lighting, noise, or blur. The second step was the application of Mediapipe's facial landmarking for face cropping and accurate face positioning [17]. The facial area was split into left and right sides using the landmarks for better learning. An affine transformation obtained from Delaunay triangulation was applied for geometric transformation, which reduces errors due to head position and camera orientation. Data augmentation methods like flipping, rotation, intensity changes, and zooming were used to increase variability and prevent overfitting. Finally, all images were normalized to the range 0 to 1 and resized to a uniform 224 × 224 input size for the deep learning algorithm.

IV. PROPOSED METHODOLOGY

The proposed approach combines deep learning for classification with feature extraction and intelligent preprocessing to identify facial weakness associated with stroke from only a few images of the face (refer to Fig. 1). To identify specific facial features such as the corners of the mouth, nose, and eyes, the approach begins with facial landmark detection. Based on the facial landmarks, the face is then divided into left and right sides to examine asymmetry, which is a vital stroke symptom. An affine transformation based on Delaunay triangulation of face triangles is applied to remove geometric distortions due to head position, rotation, and camera orientation, ensuring that similar face areas are properly aligned [18]. To increase the variability of the dataset and avoid overfitting, the process further normalizes images to 224×224 pixels, normalizes them, and generates additional images by flipping, rotating, adjusting brightness, and zooming. For feature extraction and classification, a deep autoencoder is used, consisting of a ConvNeXt encoder and a ConvNet decoder. The encoder reduces the input data to a latent space representation to identify high-level facial characteristics and ensure that the latent space embodies facial asymmetry, while the decoder attempts to reconstruct the original input.

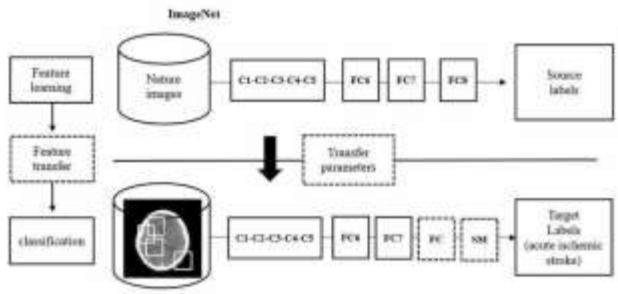

Fig.1 Working Methodology

The encoder-decoder system is trained using a reconstruction loss L_{rec} defined as:

$$L_{rec} = \frac{1}{N}\sum_{i=1}^{N} \| X_i - \hat{X}_i \|^2 \qquad (1)$$

where X_i is the original image, $\hat{X}_i$ is the reconstructed image, and N is the total number of images (as seen in Equation.1). Once trained, the encoder's latent features are used to fine-tune a classifier for stroke-related facial weakness. The classifier applies binary cross-entropy loss L_{cls} for stroke vs. normal classification:

$$L_{cls} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)] \qquad (2)$$

where y_i represents the true label and $\hat{y}_i$ is the predicted probability (as seen in Equation.2). This technology is unique in that it uses landmark based left right face division to extract region specific features while integrating real and synthetic facial datasets (FaceGAN) to overcome data lack. This method improves awareness to small stroke related defects by specifically focusing on asymmetry, compared to traditional models that process the entire face. Even with small datasets a ConvNeXt-based autoencoder can provide an accurate summary of facial features and the combination of reconstructions and classification losses ensures strong classification performance and accurate representation learning. The success of this design in identifying stroke related facial weakness in practical situations is shown by its 98.9% accuracy which was verified by a four-fold cross-validation.

V. IMPLEMENTATION

To apply the proposed method, a number of steps were followed from compiling the dataset to training and testing the model. To ensure well-rounded learning, the hybrid dataset consisted of 1,200 genuine facial images and 3,000 generated images using FaceGAN, which was then divided into training, validation, and testing datasets in a 70:20:10 split. Data preprocessing included data augmentation, affine transformation via Delaunay triangulation, left-right flipping, and detection of facial points. These measures ensured proper alignment, geometric transformation, and increased variability, thus preventing overfitting. To enable the deep learning model to analyze the images, each image was resized to 224 by 224 pixels and normalized to the range [0, 1].

A ConvNeXt encoder combined with a ConvNet decoder is at the core of the architecture a traditional autoencoder configuration for learning the latent representations of facial features [19]. For efficient learning of high-level facial features, the encoder consists of multiple convolutional layers with batch normalization, GELU activation, and the remaining skip connections. To make sure the latent facial features encode useful information about face asymmetry, the decoder replicates the architecture of the encoder using transposed convolutions to reconstruct the input images. After pretraining, the autoencoder is used as a feature extractor and later upgraded with a fully connected classifier to distinguish normal and stroke-suffering faces.

Hyperparameter optimization played an important role in achieving high-level performance. The learning rate began at 0.001 and was decreased by a factor of ten every ten epochs to maintain constant growth. To balance stable gradients and effective use of the GPU, the batch size was set to a constant 32. The Adam optimizer was used for autoencoder training and classifier fine-tuning because of its adaptive learning rate and rapid convergence, which prevent convergence to a local minimum. During fine-tuning, equal importance was assigned to reconstruction and classification loss to achieve strong classification performance and identify informative features. A large degree of data augmentation was used to improve generalization. The data augmentation pipeline consisted of zooms with a scale factor ranging from 0.9 to 1.1, flips, brightness/contrast adjustments from 0.8 to 1.2, and random rotations between -15° and 15° [20]. These transformations simulate real-world variations in scale, lighting, and facial orientation, allowing the network to generalize well to novel test images. The transformations were performed on-the-fly in batches during training, effectively doubling the size of the dataset and preventing overfitting.

Four-fold cross-validation allowed us to monitor the learning process of the model, thus enabling us to estimate its generalization and stability. During each iteration of the cross-validation, we recorded the training and validation loss, and to prevent overfitting, we implemented early stopping if no improvement was seen for 15 epochs. Finally, for selecting the best set of model weights, we relied mostly on the validation accuracy and F1-score. Once the training process was completed, we evaluated the model on a separate test dataset, and it achieved an F1-score of 0.97 and an accuracy of 98.9%. These outcomes suggest that a robust neural network architecture, appropriate preprocessing, and a combination of real and artificial data can lead to a highly effective tool for the detection of facial weakness due to stroke. The model can be used as a rapid, unofficial screening method in a medical setting, allowing for immediate action, and it is still light enough to be used in practice.

VI. RESULTS

The assessment of the proposed AI approach for identifying facial weakness due to stroke relied on several factors: accuracy, precision, recall, and F1-score. Precision and recall indicate the efficiency of the model in identifying stroke-damaged faces and preventing false positives, while accuracy indicates the percentage of correctly classified images out of all predictions. Since the class distribution is normally imbalanced, the F1-score is a good overall indicator because it combines precision and recall through their harmonic mean. The AUC signal indicates the model’s capacity to distinguish between normal and damaged faces for various threshold values, indicating the model’s robustness. Using four-fold cross-validation, the model performed well with an accuracy of 98.9%, F1 score of 0.97, precision of 0.96, recall of 0.98, and an AUC of 0.99, as shown in Table 1.

The contextual baseline measured the performance of standard models against our system. The baseline models, VGG16, ResNet50, and MobileNet, were trained on the hybrid dataset without the use of autoencoders or face/area segmentation. In this case, the ResNet50 and MobileNet models achieved 93.5% and 90.8% accuracy, with F1-scores of 0.91 and 0.88, respectively, while the VGG16 model achieved 91.2% accuracy with an F1-score of 0.89. Our ConvNeXt autoencoder, which encodes a concise and discriminative representation of facial symmetry by using an encoder-decoder architecture that aims to reduce the reconstruction error while retaining features of interest for classification, performed better than all the other models.

METRIC	VALUE
ACCURACY	98.9%
F1-SCORE	0.9719
PRECISION	0.96
RECALL	0.98

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 04 | April - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

AUC (AREA UNDER CURVE)	0.99

Table. 1 Model Performance Metrics

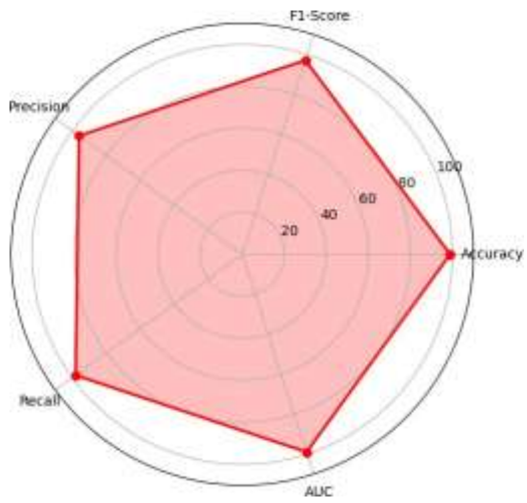

Fig.2 Radar Plot of Model Performance

The radar chart provides a complete picture of the model's ability to remain constant on all aspects, emphasizing its ability to detect even the slightest asymmetry in facial features (see Fig. 2). The approach excels because of several important innovations. First, its hybrid dataset approach, combining real and synthetic images, greatly improves generalization and reduces the risk of overfitting relative to models trained solely on real images. Second, affine transformations and left-right facial segmentation in the preprocessing step resolves geometric anomalies and improves image alignment at the feature level. Third, the ConvNeXt encoder captures the slightest changes in facial muscle activity, which is critical for detecting slight neurological facial weakness.

All these components worked in unison to improve the accuracy, specificity, and robustness of the system, particularly in identifying asymmetry under varying lighting conditions and facial expressions. The study also highlighted the limitations and constraints of the system. Small variations between real and synthetic images contributed to small classification errors, despite the fact that the inclusion of the synthetic FaceGAN dataset improved performance. This problem was more evident in bright lighting conditions and in the presence of distractions such as glasses and facial hair. The system was trained on 3,000 synthetic and 1,200 real images, although increasing the number of images could improve the representation of ethnic diversity and unusual facial variations, beyond the current combination's ability to provide good cross-validation results. A line graph illustrating performance with and without FaceGAN data clearly demonstrated the benefit of synthetic data augmentation, further emphasizing the effectiveness of the approach in enhancing generalization and reliability (as illustrated in Fig. 3).

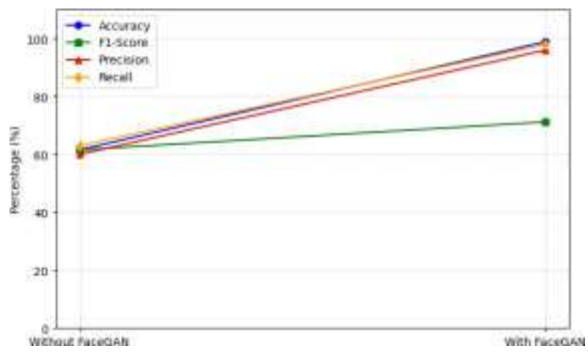

Fig.3 Performance Comparison: With vs Without FaceGAN

Fig.4 Model Predictions

The experimental outcome provides several insights into the effectiveness of the approach (Fig. 4). By interweaving geographic preprocessing, autoencoder-based feature extraction, and hybrid datasets, the approach pushes towards clinical-grade accuracy. This is an indication that even with smaller datasets, the approach can provide excellent results. The four-fold cross-validation shows only minute variations between the folds, which is an indication of robustness and resistance to overfitting. The approach maintains a good balance between the important attributes, which is important in the medical domain where false negatives have severe implications, as evident from the high F1-score. The proposed approach clearly has an edge over the existing approaches.

The technique combines representation learning with the consideration of facial asymmetry, which increases the sensitivity of features and the accuracy of classification, going beyond what conventional CNNs and manually designed features can accomplish. This combination allows for more complex training and improved generalization capabilities on novel faces, as opposed to approaches that exclusively use real data. The findings indicate that by integrating preprocessing, synthetic data, and deep learning, a novel framework for the detection of gentle strokes is formed. For the future, incorporating video analysis, expanding the framework with multiple datasets, and integrating the system into clinical practice could take the approach even further.

VII. CONCLUSION & FUTURE SCOPE

This research combines real facial images with FaceGAN-generated images to develop an AI-based approach for detecting facial deficits associated with stroke. The key contributions include affine geometric correction, landmark-based left–right facial segmentation, and the use of a ConvNeXt-based autoencoder for robust feature extraction. The proposed method achieves a high accuracy of 98.9% along with strong F1-scores, demonstrating improved generalization, reduced overfitting, and better adaptability to subtle facial asymmetries compared to traditional CNNs and hand-engineered methods. However, the use of static images introduces domain gaps between real and synthetic data, and the dataset may not fully represent diverse facial and demographic variations. Future work will focus on incorporating dynamic video-based analysis, expanding datasets to improve diversity, enabling real-time clinical deployment, and integrating explainable AI techniques to enhance interpretability and reliability in medical applications.

REFERENCES

1. Khandelwal, Bidita, and Chamma Gupta. "Leading causes of death and disability among the global aging community." The Ageing Population: Impact Analysis on'Societal and Healthcare Cost'. Singapore: Springer Nature Singapore, 2023. 37-54.
2. Ou, Zijun, et al. "Early identification of stroke through deep learning with multi-modal human speech and movement data." Neural Regeneration Research 20.1 (2025): 234-241.
3. Murali, Sanjana, et al. "Bringing MRI to low-and middle-income countries: directions, challenges and potential solutions." *NMR in Biomedicine* 37.7 (2024): e4992.
4. Mirzabeigi, Hadi, et al. "Subjective Mental Workload Profile and Human Error Probability in Nurses: A Cross-sectional Analytical Study." *International Journal of Environmental Health Engineering* 14.3 (2025): 23.
5. Montesinos López, Osval Antonio, Abelardo Montesinos López, and Jose Crossa. "Overfitting, model tuning, and evaluation of prediction performance." *Multivariate statistical machine learning methods for genomic prediction*. Cham: Springer International Publishing, 2022. 109-139.
6. Wang, Huashu, and Zhi Li. "Constructing a competence framework for interpreting technologies, and related educational insights: an empirical study." *The Interpreter and translator trainer* 16.3 (2022): 367-390.
7. Li, Qingdu, et al. "Optimizing Class Imbalance in Facial Expression Recognition Using Dynamic Intra-Class Clustering." *Biomimetics* 10.5 (2025): 296.
8. Wang, Yiyang, et al. "Prediagnosis recognition of acute ischemic stroke by artificial intelligence from facial images." *Aging Cell* 23.8 (2024): e14196.
9. Hussain, Shah, et al. "Modern diagnostic imaging technique applications and risk factors in the medical field: a review." *BioMed research international* 2022.1 (2022): 5164970.
10. Boddapati, Mohan Sai Dinesh, et al. "Creating a protected virtual learning space: a comprehensive strategy for security and user experience in online education." *International Conference on Cognitive Computing and Cyber Physical Systems*. Cham: Springer Nature Switzerland, 2023.
11. Muzammul, Muhamad, and Xi Li. "Comprehensive review of deep learning-based tiny object detection: challenges, strategies, and future directions." *Knowledge and Information Systems* (2025): 1-89.
12. Emegano, Declan Ikechukwu, et al. "Advancing Prostate Cancer Diagnostics: A ConvNeXt Approach to Multi-Class Classification in Underrepresented Populations." *Bioengineering* 12.4 (2025): 369.
13. Safwat, Soha, et al. "Hybrid deep learning model based on GAN and RESNET for detecting fake faces." *IEEE Access* 12 (2024): 86391-86402.
14. Wood, Erroll, et al. "Fake it till you make it: face analysis in the wild using synthetic data alone." *Proceedings of the IEEE/CVF international conference on computer vision*. 2021.
15. Karkkainen, Kimmo, and Jungseock Joo. "Fairface: Face attribute dataset for balanced race, gender, and age for bias measurement and mitigation." *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. 2021.
16. Okazaki, Shuntaro, et al. "Mask wearing increases eye involvement during smiling: A facial EMG study." *Scientific reports* 11.1 (2021): 20370.
17. Liu, Yan, et al. "Study of the error caused by camera movement for the stereo-vision system." *Applied Sciences* 11.20 (2021): 9384.
18. Li, Xiang, et al. "Evaluating effects of focal length and viewing angle in a comparison of recent face landmark and alignment methods." *EURASIP Journal on Image and Video Processing* 2021.1 (2021): 9.
19. Yang, Yuxin, et al. "A variable fidelity approach for predicting aerodynamic wall quantities of hypersonic vehicles using the ConvNeXt encoder-decoder framework." *Aerospace Science and Technology* 155 (2024): 109605.
20. Kamil, Khuder Hasan, and Anwar Hassan Al-Saleh. "Exploring camera zoom's impact on viewing angles and scale factor in computer vision models of UFO detection and tracking systems." *AIP Conference Proceedings*. Vol. 3211. No. 1. AIP Publishing LLC, 2025.