ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may820

Enhancing Non-Invasive Diabetes Diagnosis: A Comparative Analysis of Deep Learning Models using SAM-Supported U-Net for Tongue Segmentation

Hasan Erdinç Koçer[1]; Mohammed Qaimaz Ali[2*]

[1]Professor, Ph.D. Department of Electrical and Electronics Engineering, Faculty of Technology, Selçuk University, Konya 42130, Türkiye.
[2]Ph.D. Candidate, Department of Computer Engineering, Institute of Natural and Applied Sciences, Selçuk University, Konya 42130, Türkiye.

[1]ORCID: 0000-0002-0799-2140
[2]ORCID: 0009-0009-2564-5378

Corresponding Author: Mohammed Qaimaz Ali[2*]

Publication Date: 2026/05/25

Abstract: Diabetes is a major public health concern necessitating early diagnosis to prevent severe long-term systemic complications. This study presents a comprehensive comparative analysis of classical machine learning and deep learning approaches for non-invasive diabetes detection using tongue images. The experimental framework evaluates both segmented and non-segmented data to isolate the impact of region-of-interest extraction. For precise localization, a novel preprocessing pipeline is proposed: the tongue region is automatically detected using a Segment Anything Model (SAM)-based bounding box approach, followed by pixel-level segmentation via a U-Net architecture. To improve model robustness, deterministic data augmentation techniques are applied. In the classical machine learning phase, handcrafted features—including GLCM, LBP, HOG, and SIFT—are extracted and classified using Support Vector Machine (SVM) and Random Forest (RF). Conversely, the deep learning phase utilizes transfer learning with ResNet50, VGG16, EfficientNet-B4, and DenseNet169 architectures. Experimental results demonstrate that the SAM-supported segmentation significantly boosts classification performance by eliminating background noise. Specifically, the ResNet50 model achieved the highest performance with 97.92% accuracy, precision, and recall on the segmented dataset. These findings validate that AI-driven tongue image analysis, particularly when enhanced by robust segmentation, offers a highly accurate, rapid, and non-invasive alternative for clinical diabetes screening. Overall, deep learning approaches proved superior to classical methods in modeling the complex texture and color variations of the tongue. These findings confirm the clinical potential of the proposed system as an effective screening tool. Future work will focus on dataset expansion and mobile platform integration to facilitate real-time diagnosis.

Keywords: *Diabetes Diagnosis, Tongue Image Analysis, Deep Learning, Segment Anything Model (SAM), Medical Image Segmentation, Non-Invasive Screening.*

How to Cite: Hasan Erdinç Koçer; Mohammed Qaimaz Ali (2026) Enhancing Non-Invasive Diabetes Diagnosis: A Comparative Analysis of Deep Learning Models using SAM-Supported U-Net for Tongue Segmentation. *International Journal of Innovative Science and Research Technology*, 11(5), 1457-1472. https://doi.org/10.38124/ijisrt/26may820

I. INTRODUCTION

Diabetes mellitus (DM) is widely recognized as one of the most critical public health problems today due to its rapidly increasing global prevalence, chronic course, and association with numerous systemic complications. According to reports by the International Diabetes Federation (IDF), the lack of early diagnosis and regular monitoring significantly increases diabetes-related mortality and morbidity. This situation highlights the growing need for rapid, reliable, and non-invasive diagnostic methods, particularly for screening purposes. Although conventional biochemical tests provide high diagnostic accuracy, their invasive nature, associated costs, and requirement for

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may820

repeated measurements have motivated the exploration of alternative approaches. In this context, image-based, artificial intelligence–assisted diagnostic systems offer strong potential for early detection and cost-effective healthcare solutions. Recent advances in deep learning have demonstrated that data-driven artificial intelligence models can achieve diagnostic performance comparable to that of medical experts across a wide range of clinical applications, particularly in non-invasive screening and decision-support systems [1].

Biomarker analysis based on tongue images has emerged as an innovative research field situated at the intersection of centuries-old clinical observations in traditional Chinese medicine and modern computer vision and artificial intelligence techniques. Comprehensive review studies have reported that tongue image analysis provides clinically meaningful visual biomarkers related to color, texture, and geometric structure and has therefore gained increasing attention as a computer-aided diagnostic modality for metabolic and systemic diseases [2]. The tongue may reflect physiological changes associated with metabolic disorders through visual characteristics such as color distribution, coating (fur) density, texture patterns, fissures, and geometric form. Recent studies have demonstrated that these visual cues exhibit statistically significant associations with metabolic diseases such as diabetes and prediabetes. In particular, standardized imaging protocols and automated feature extraction methods have shown that tongue images can serve as objective and reproducible diagnostic tools. The current literature clearly indicates that tongue images provide promising performance in diabetes detection. It has been reported that combining tongue image features with oral–gut microbiota data achieved an AUC of 86.9% in distinguishing prediabetes and type 2 diabetes [3]. In a study employing a fusion approach with thermal and visible-spectrum tongue images, an accuracy of 94.37% was obtained, demonstrating that multimodal imaging enhances diagnostic power [4]. Using features extracted from standardized tongue images, an SVM-based model reported an accuracy of 78.77%, indicating that classical machine learning methods can also be effectively applied in this domain [5]. Multi-stage deep learning approaches integrating automatic tongue segmentation and classification have been shown to produce successful results in diabetes discrimination [6]. In particular, automatic tongue region segmentation has been shown to significantly reduce background interference and improve the robustness of deep learning–based classifiers by enabling the models to focus exclusively on disease-relevant anatomical structures [7]. the statistical distributions of tongue features in diabetic individuals differ significantly from those of healthy individuals, particularly in terms of color composition and texture patterns, as demonstrated in CNN-based tongue image analyses [8].

Moreover, deep learning–based approaches have been shown to contribute substantially to performance improvements. Studies have reported higher accuracy and generalizability in diabetes classification from tongue images using CNN-based models compared to traditional methods [9]. Comprehensive survey studies examining artificial intelligence techniques for disease detection based on tongue images indicate that segmentation-supported deep models are particularly promising for future clinical applications [10,11]. It has been demonstrated that combining color- and texture-based features with deep features significantly improves classification performance [12], while automatic tongue region segmentation reduces background noise and enables more stable results [13]. Nevertheless, a considerable portion of existing studies focuses either solely on classical machine learning methods or on individual deep learning architectures, and comprehensive analyses that systematically compare the effects of segmented and non-segmented tongue images across different model types remain limited. Recent survey studies emphasize that comparative investigations jointly evaluating classical machine learning and multiple deep learning architectures under both segmented and non-segmented data scenarios remain relatively limited, despite their critical importance for understanding model behavior and generalizability [14]. In particular, the joint evaluation of classical methods and multiple deep learning architectures on the same dataset under both segmented and non-segmented scenarios represents a notable gap in the literature. Addressing this gap, the present study aims to comparatively investigate the performance of different data preprocessing strategies and modeling approaches for tongue image–based diabetes detection, thereby contributing scientifically to the development of non-invasive, artificial intelligence–assisted diagnostic systems.

II. RELATED WORK

A diabetes risk prediction model was developed by integrating Traditional Chinese Medicine (TCM) tongue diagnosis with machine learning. Tongue coating, color, and morphological features were evaluated analytically. The model demonstrated promising performance as a non-invasive screening tool. The study illustrates the feasibility of integrating cultural medical knowledge with modern algorithms [15].

A deep learning-based decision support system was proposed for the diagnosis of diabetes. Automatic feature extraction was performed using a Convolutional Neural Network (CNN) architecture, and the clinical applicability of the model was evaluated. The results indicate that deep learning offers more stable performance compared to manual feature extraction [16].

Type 2 Diabetes classification was conducted by extracting color and texture-based features from tongue images. HSV and Lab color spaces were utilized in conjunction with Gray-Level Co-occurrence Matrix (GLCM) features. The findings revealed that color distributions exhibit significant differences in individuals with diabetes [17].

An approach involving automatic tongue region detection and post-segmentation classification was proposed. A tongue mask was obtained using a U-Net-like architecture, and it was reported that segmentation improved classification

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may820

accuracy. These results emphasize the critical role of preprocessing steps [18].

The usability of tongue images obtained via mobile devices for diabetes screening was investigated. A low-computational-cost system was developed using lightweight CNN architectures. The study offers a practical solution for remote health monitoring (telemedicine) [19].

The study aimed to simultaneously predict metabolic syndrome and diabetes risk through tongue image analysis. A multi-class classification approach was adopted, demonstrating that early stages of diabetes could be distinguished. The findings support the potential of tongue images as predictive health indicators [20].

The performance of transfer learning-based deep networks on small-scale tongue image datasets was examined. It was shown that the use of pre-trained models reduced overfitting and increased accuracy [21].

A multimodal diabetes diagnosis system was developed by combining tongue images with facial images. Classification was performed using a late fusion strategy, achieving higher performance compared to single modalities [22].

An image processing algorithm was proposed to automatically measure the density of tongue coating. It was statistically demonstrated that coating characteristics are correlated with glucose metabolism. The results highlight the clinical value of quantitative measurements [23].

A diabetes classification model based on tongue images was analyzed using an Explainable Artificial Intelligence (XAI) approach. The regions influencing the model's decisions were visualized using Grad-CAM-like methods. The study provides a significant contribution toward increasing clinician trust [24].

The impact of the lack of standardization in tongue image datasets on model performance was examined. It was shown that variations in lighting, positioning, and camera specifications directly affect classification results. The research draws attention to the vital importance of data collection protocols [25].

III. MATERIALS AND METHODS

➢ *Dataset*

The dataset used in this study consists of tongue images acquired from diabetic and non-diabetic individuals at Kirkuk Teaching Hospital in Iraq. The dataset includes images from 158 diabetic patients and 170 healthy subjects, with 4 to 7 tongue images captured from different positions for each individual. As a result of this acquisition process, a total of 757 diabetic and 802 non-diabetic images were collected. Prior to analysis, all images underwent quality control and preprocessing procedures to ensure their suitability for the experimental procedures.

➢ *Tongue Region Localization (SAM Bounding Box)*

A semi-automatic annotation approach based on the Segment Anything Model (SAM) was employed to achieve precise and reproducible segmentation of tongue images. SAM has recently emerged as a powerful foundation model for image segmentation, demonstrating strong generalization capability across diverse domains, including medical imaging, by enabling accurate mask generation with minimal human interaction [26]. The primary objective of this method is to minimize user interaction while enabling fast, consistent, and reliable mask generation for large-scale medical image datasets, thereby providing high-quality training data for pixel-level segmentation models such as U-Net. The system was designed to support multiple image formats, including JPEG, PNG, TIFF, and JFIF. Image files were automatically sorted according to their numerical naming conventions using regular expressions, ensuring a structured and consistent workflow for large-scale datasets. In addition, a state-based recording mechanism was implemented to allow the annotation process to resume seamlessly in the event of interruption, thereby eliminating the risk of data loss.

During the SAM annotation process, the user is required to define only a single bounding box (BBox). This bounding box is obtained via the BBoxWidget as normalized coordinates within the range $[0, 1]$and is subsequently transformed into the pixel coordinate space using the image dimensions as follows:

$$x_{\text{pixel}} = x_{\text{norm}} \times W, y_{\text{pixel}} = y_{\text{norm}} \times H \quad (1)$$

Where W and H denote the width and height of the image, respectively. In cases where no bounding box is provided by the user, the system automatically generates a bounding box that covers the entire image, allowing the segmentation process to proceed automatically without further user intervention.

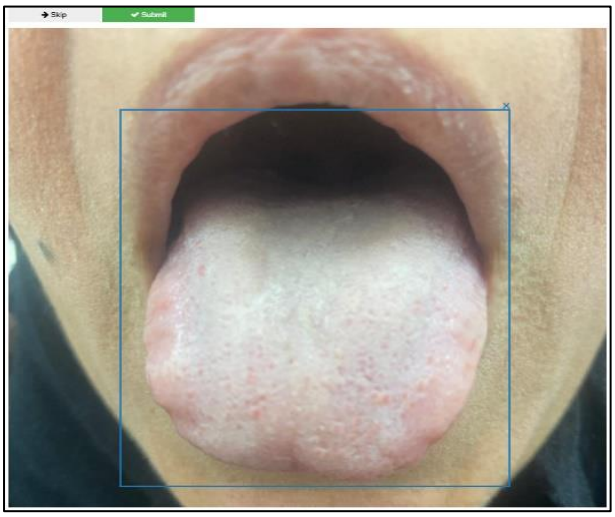

Fig 1 SAM Labeling Bounding Box (BBox).

Once the bounding box (BBox) is defined, the Vision Transformer (ViT)–based SAM predictor model is activated. As illustrated in Figure 1, the model generates multiple

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may820

candidate masks for the same image by considering the specified region of interest. This process can be mathematically expressed as follows:

$$\mathcal{M} = f_{\text{SAM}}(I, B) \quad (2)$$

Where I denotes the input image, B represents the bounding box, and $\mathcal{M} = \{M_1, M_2, \dots, M_n\}$corresponds to the set of generated candidate masks. For each candidate mask, SAM assigns a confidence score to each candidate mask:

$$M^* = \arg\max_{M_i \in \mathcal{M}} s(M_i) \quad (3)$$

This strategy enables the selection of the most stable and reliable boundary among multiple mask alternatives generated for the same region, thereby allowing irregular boundary transitions and fine-grained textural details on the tongue surface to be captured with high accuracy.

➢ *U-Net–Based Tongue Image Segmentation*

In order to suppress background effects in tongue images and to enhance the visibility of textural patterns associated with diabetes, a deep learning–based segmentation model built upon the U-Net architecture was employed in this study. As illustrated in Figure 2, U-Net leverages its encoder–decoder structure to jointly capture low-level boundary information and high-level abstract textural representations learned in deeper layers, thereby enabling accurate delineation of fine structural variations on the tongue surface. Originally introduced for biomedical image segmentation, the U-Net architecture has become a de facto standard due to its ability to preserve spatial resolution through skip connections while effectively learning multi-scale contextual features [27]. In this context, the semi-automatically generated masks obtained from the preceding SAM stage were directly utilized as ground truth for training the U-Net model, transforming the segmentation process into a fully automated and reproducible pipeline.

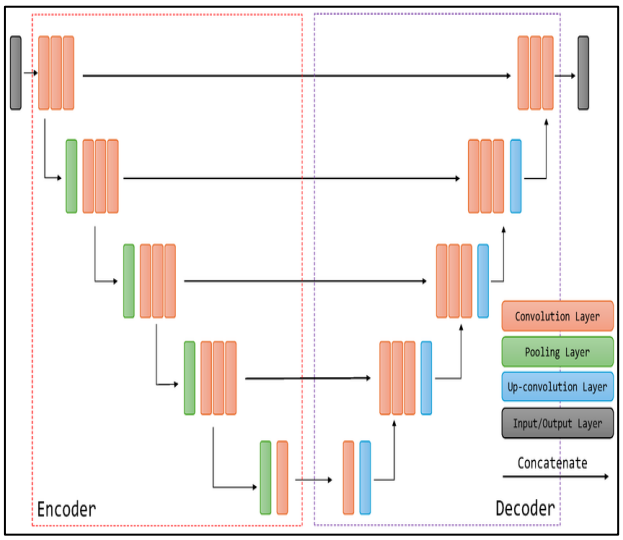

Fig 2 Schematic Overview of the U-Net Architecture with an Encoder–Decoder Structure.

- *Data Preparation and Mask Matching*

The tongue images in the dataset were sorted based on the numerical indices embedded in their file names, and image–mask pairs sharing the same index were automatically matched. Mask validity was verified by analyzing pixel intensity distributions; the dominance of pixel values equal to 0 and 255 confirmed the binary nature of the masks. Prior to training, all masks were resized and converted into a binary representation within the range $[0, 1]$through thresholding. To preserve boundary information during rescaling, the INTER_NEAREST interpolation method was employed. These preprocessing steps ensured data integrity and contributed to a stable and consistent U-Net training process.

- *U-Net Architectural Structure*

The U-Net model consists of two symmetric components: an encoder and a decoder. In the encoder stage, spatial resolution is progressively reduced while feature extraction is performed using successive Conv2D + MaxPooling blocks. This process can be mathematically expressed as:

$$F_l = \sigma(W_l * F_{l-1} + b_l) \quad (4)$$

Where F_l denotes the feature map at the l-Th layer, W_l and b_l represent the learnable weight and bias parameters, respectively, and $\sigma(\cdot)$corresponds to the ReLU activation function.

In the decoder stage, features extracted by the encoder are upsampled using Conv2DTranspose layers and combined with low-level spatial features from the corresponding encoder layers via skip connections:

$$F_l' = \sigma(W_l' * \text{Up}(F_{l+1}) + b_l') \quad (5)$$

This architecture facilitates the reconstruction of sharper and more consistent boundaries on the tongue surface by effectively integrating multi-scale contextual and spatial information.

- *Loss Function and Performance Metrics*

Model optimization was carried out using a combined loss function that jointly accounts for pixel-wise classification accuracy and boundary overlap, namely Binary Cross-Entropy (BCE) + Dice Loss:

$$\mathcal{L} = \text{BCE} + (1 - \text{Dice}) \quad (6)$$

The Dice coefficient is defined as:

$$\text{Dice} = \frac{2|P \cap G|}{|P| + |G|} \quad (7)$$

Where P represents the predicted mask and G denotes the ground-truth mask.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may820

➢ *Data Augmentation*

To enhance the generalization capability of deep learning models for diabetes detection from tongue images, a deterministic and systematic data augmentation strategy was applied. The primary objective of this approach was to reduce overfitting in medical datasets with limited sample sizes and to improve the robustness of the models against variations encountered in real clinical scenarios.

- *Data Sorting, Naming, and Image Loading Structure*

Each tongue image was stored in its original form using the label "_0orj" to preserve experimental traceability, while the augmented samples were sequentially named as "_aug1, _aug2 … _aug5". Natural numerical ordering of file names was ensured through a custom-defined natural_key function, thereby preventing inconsistencies that may arise from purely alphabetical sorting. During the image loading stage, multiple formats—including JPEG, PNG, BMP, TIFF, and JFIF—were supported. For images that could not be read using OpenCV, a PIL-based fallback mechanism was employed to resolve potential format incompatibilities.

- *Deterministic Augmentation Principle*

All data augmentation procedures Gaussian Noise + CoarseDropout, Vertical Flip, Random 90° Rotation (RandomRotate90) Horizontal Flip + Brightness–Contrast Adjustment and Gamma + HSV Color Transformations were conducted using a deterministic seed-based approaches. As a result, the transformations applied to each image yielded identical outputs across repeated executions, thereby ensuring scientific reproducibility and experimental consistency.

➢ *Performance Evaluation*

To quantitatively evaluate the performance of the developed classical machine learning and deep learning–based models, multiple evaluation metrics were employed. The evaluation process was conducted under both segmented and non-segmented data scenarios, enabling a comparative analysis of model behavior across different preprocessing strategies. Within this framework, all classical and deep learning models were trained and tested using an identical data splitting strategy, consisting of a subject-wise 70% training and 30% testing split combined with StratifiedGroup K-Fold (5-fold) cross-validation. Performance assessments were carried out based on both independent test set results obtained from 5-fold cross-validation (5-fold CV). To assess classification performance, accuracy, precision, recall, and F1-score metrics were utilized. These metrics were computed separately for the diabetic (D) and non-diabetic (ND) classes, and a macro-averaging approach was adopted to mitigate potential class imbalance effects. Macro averaging assigns equal weight to each class, thereby providing a more balanced representation of overall model performance across both classes. Specifically, each metric was first calculated independently for the two classes and subsequently averaged to obtain the macro-level performance value. This strategy aims to prevent performance bias that may arise in favor of a particular class.

To evaluate model generalizability, Stratified Group K-Fold (5-fold) cross-validation was applied to the training set. The mean values of the performance metrics computed across all folds represent the overall model performance, while the corresponding standard deviations reflect model stability across different data subsets. Lower standard deviation values indicate that the model produces consistent results across folds and exhibits stable learning behavior. Finally, cross-validation outcomes and independent test set performances were jointly reported to comprehensively assess both the learning capacity of the models during training and their realistic performance on previously unseen data. This multi-level evaluation framework ensures that the reported results are scientifically reliable, reproducible, and comparable.

➢ *Classical Machine Learning Approach*

In the classical approach, a machine learning–based classification system was developed for diabetes detection from tongue images. Unlike deep learning methods, this approach relies on manual feature extraction and statistical classifiers, offering notable advantages in terms of interpretability, computational efficiency, and model transparency, particularly in medical scenarios with limited data availability. The dataset was divided into two primary classes, Diabetic and Non-Diabetic. Images were automatically loaded from a Google Drive directory, normalized, and resized prior to feature extraction. For each image, a hybrid feature vector was constructed by applying four fundamental feature extraction techniques: Gray-Level Co-occurrence Matrix (GLCM), Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG), and Scale-Invariant Feature Transform (SIFT). Classical features were extracted from tongue images using these four complementary methods.

First, GLCM matrices were computed from grayscale images using a pixel distance of 1 and an orientation of 0° (horizontal direction). The matrices were configured with 256 gray levels and were generated in a symmetric and normalized form. The LBP method was applied in uniform mode with $P = 8$ neighboring points and a radius of $R = 1$, followed by the extraction of a 26-dimensional histogram feature vector. HOG features were obtained to represent the directional distribution of image gradients, using 9 orientation bins, a cell size of 16×16pixels, a block size of 2×2cells, and L2-Hys normalization. Finally, SIFT descriptors were extracted using the default parameters provided by OpenCV ($\text{nFeatures} = 0$, $\text{nOctaveLayers} = 3$, $\text{contrastThreshold} = 0.04$, $\text{edgeThreshold} = 10$, $\sigma = 1.6$). The mean of the computed local descriptors was taken to form a 128-dimensional feature vector, of which the first 50 elements were used as input to the classification model. All extracted features were standardized using StandardScaler and subsequently provided as input to Support Vector Machine (SVM) and Random Forest (RF) classifiers.

- *Support Vector Machine (SVM) Model*

Support Vector Machine (SVM) is a powerful classification method that aims to determine the optimal decision boundary separating two classes and is capable of

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may820

producing stable results even in high-dimensional feature spaces. When defining the decision boundary, the model relies exclusively on support vectors, which are the samples that critically influence the separation between classes. In this study, the SVM model was evaluated in two configurations: a default (non–hyper parameter-tuned) version and an optimized version in which hyper parameters were tuned using the GridSearchCV approach.

- *Random Forest (RF) Model*

Random Forest (RF) is an ensemble-based classification method that combines a large number of decision trees. Each tree is trained using a different subset of the data and a randomly selected subset of features, and the final decision is determined through majority voting. This ensemble structure substantially mitigates the overfitting problem commonly encountered in individual decision trees, thereby enhancing the model's generalization capability. In this study, the RF model was evaluated in both non–hyper parameter-tuned and hyper parameter-tuned configurations.

➢ *Deep Learning Models*

In the deep learning stage, diabetes classification from tongue images was performed by leveraging the automatic feature extraction capability of convolutional neural networks. For this purpose, the ResNet50, VGG16, EfficientNet-B4, and DenseNet169 architectures were employed. The modeling process was designed to enhance the reliability of non-invasive diagnostic approaches and to support early disease detection. The dataset was divided into two main classes: Diabetic (D) and Non-Diabetic (ND). For each class, both segmented and non-segmented versions of the images were generated. The images were transferred from the Google Drive environment to the Colab workspace and subsequently copied to the local disk of the virtual machine to improve computational efficiency. All images were resized to 224 × 224 pixels in the RGB color space and standardized using ImageNet normalization. The dataset was split into 70% training and 30% testing subsets, and 5-fold cross-validation (5-fold CV) was applied to the training portion.

The data loading process was implemented using the Tongue Dataset class, in which the class label of each image file was automatically assigned (0: ND, 1: D). Model training was conducted via the train_and_evaluate function, which performs 5-fold cross-validation on the training set following the 70/30 train–test split. For each fold, accuracy, precision, recall, and macro-averaged F1-score were computed. At the end of each epoch, the average loss value was reported to monitor model stability. Across all experiments, the deep learning models were trained for 10 epochs, with a mini-batch size of 16 samples at each training step. This configuration ensured balanced gradient propagation and improved computational efficiency. The selected epoch and batch size values contributed to stable learning dynamics and reliable evaluation of classification performance. In addition, GPU acceleration, along with the use of pin_memory and persistent_workers options in the data loaders, facilitated further optimization of the training process.

IV. RESULTS

➢ *BBox Produces Masks*

As shown in Figure 3, the final mask obtained is stored in two formats: a binary mask consisting exclusively of pixel values of 0 and 255, and an overlay representation in which the mask is superimposed onto the original image. The overlay visualization is generated using the following linear blending model:

$$I_{\text{overlay}} = \alpha I + (1 - \alpha)M \tag{8}$$

W here I denotes the original RGB image, M represents the mask image, and α is the transparency coefficient. This visual representation enables rapid and intuitive qualitative assessment of the accuracy of the generated masks. In addition, the system incorporates a multi-stage image loading mechanism to ensure robust handling of non-standard JPEG variants commonly produced by medical imaging devices (e.g., JFIF, JFI, JIF). Images are sequentially processed using OpenCV, PIL, and low-level byte-decoding methods, thereby effectively mitigating potential format incompatibilities.

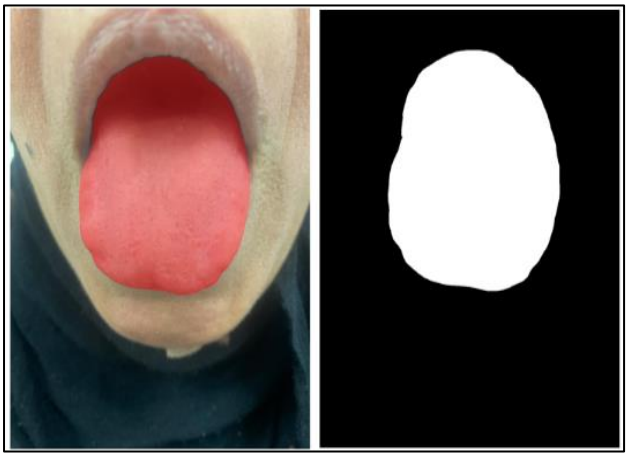

Fig 3 Automatic Detection of the Tongue Region and Generation of a Binary Mask Using SAM.

Overall, the SAM-based semi-automatic masking approach provides substantial advantages over manual annotation in terms of both speed and accuracy, and plays a critical role in generating reliable training masks, which directly enhance the performance of the subsequent U-Net–based segmentation model. Consequently, this methodology constitutes a key component of the data preparation pipeline employed in this study.

➢ *U-Net Visual Output and Post-Processing*

After completion of training, the model generated predictions for all images, and background removal was performed using the resulting binary masks:

$$I_{\text{seg}} = I \times M \tag{9}$$

W here I denotes the input image and M represents the predicted binary mask. This operation preserved only the

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may820

tongue region while completely eliminating the surrounding background, as illustrated in Figure 4. Additionally, morphological opening and closing operations with a 3 × 3kernel were applied to reduce boundary irregularities and refine mask contours.

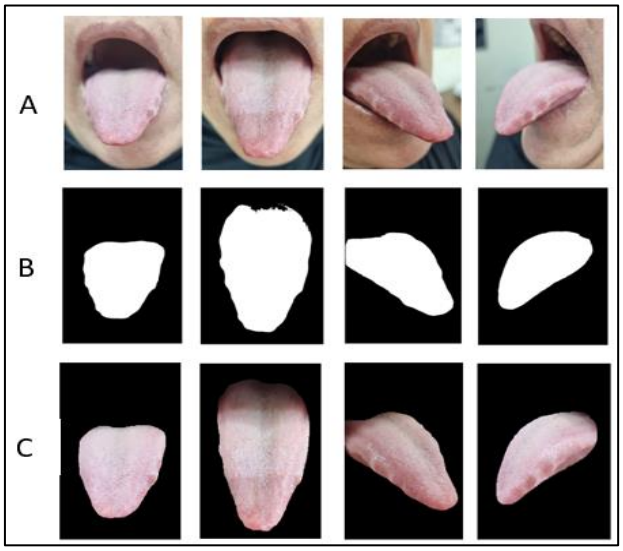

Fig 4 Representative Examples Illustrating (A) Original Tongue Images, (B) Generated Binary Masks, and (C) the Resulting Segmentation Outputs.

The training process of the model was conducted for a total of 25 epochs, using a mini-batch size of four images at each training step. Improve computational efficiency while maintaining stable gradient updates. The achieved Dice score of 0.9588 demonstrates that the model is capable of accurately delineating boundaries in tongue images. When integrated with the SAM-based pre-masking stage, the proposed U-Net architecture transforms the segmentation process into a fully automated pipeline, providing a reliable and robust image loading framework for subsequent classification tasks.

➢ *Applied Data Augmentation Techniques*

For each original image, a total of six samples were generated: one original image and five augmented versions Figure 5. The selected augmentation techniques were designed to simulate natural variations commonly observed in clinical imaging conditions.

- *Gaussian Noise + CoarseDropout:*

Applied to model sensor-induced noise and localized information loss.

- *Vertical Flip:*

Simulates potential anatomical asymmetries in tongue images, thereby enhancing spatial invariance.

- *Random 90° Rotation (RandomRotate90):*

Enables orientation-independent learning by applying rotations of 0°, 90°, 180°, and 270°.

- *Horizontal Flip + Brightness–Contrast Adjustment:*

Simulates variations in clinical lighting conditions and image acquisition settings.

- *Gamma + HSV Color Transformations:*

Represents color variations on the tongue surface, increasing the contribution of color information to the classification process.

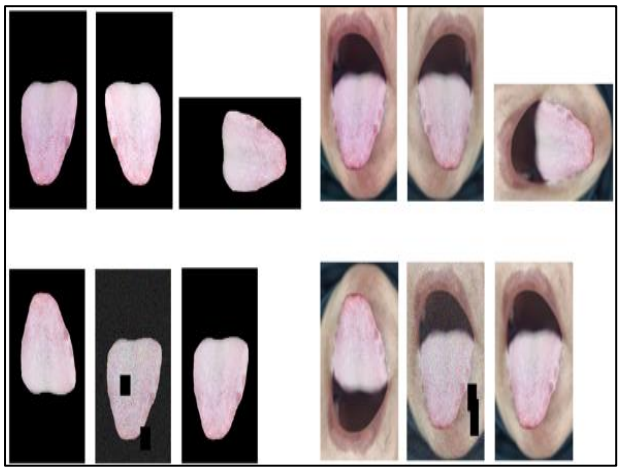

Fig 5 Visual Examples of the Five Data Augmentation Techniques Applied to Both Segmented and Non-Segmented Datasets, Including the Corresponding Original Images.

➢ *Impact on the Dataset*

All augmented images were stored in the RGB color space using a lossless PNG format. As a result of the applied data augmentation process, the number of images in the diabetic class increased to 4,542, while the non-diabetic class reached 4,812 images, yielding a total of 9,354 processed images. Data augmentation was applied separately to both segmented and non-segmented datasets to ensure that each scenario contributed equally to the model training process. The proposed deterministic and clinically compatible augmentation strategy increased data diversity, thereby enhancing the generalization capability of deep learning–based diabetes classification models. Moreover, this approach effectively reduced the risk of overfitting and led to more stable and reliable model performance.

➢ *Non–Hyper Parameter-Tuned SVM*

In the non–hyper parameter-tuned SVM approach, the model was trained using default settings without any hyper parameter optimization. Under this configuration, the SVM constructs a linear decision boundary that maximizes the margin between classes, thereby performing classification based on the linear separability of the raw feature space. Since flexibility-enhancing parameters such as the penalty coefficient, kernel function, or tolerance were not adjusted, the model's performance is largely constrained by the linear structure of the feature space. Consequently, particularly in non-segmented and noisy images, the decision boundaries were observed to lack sufficient flexibility for effective class discrimination.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may820

Table 1 Non–Hyperparameter-Tuned SVM

	Metric	K-Fold Average	Test Result
Non-Segmented	Accuracy	0.6436	0.6974
	Precision (macro)	0.7234	0.7256
	Recall (macro)	0.6305	0.6873
	F1-score (macro)	0.5960	0.6800
Segmented	Accuracy	0.7642	0.7804
	Precision (macro)	0.7656	0.7866
	Recall (macro)	0.7646	0.7823
	F1-score (macro)	0.7640	0.7799

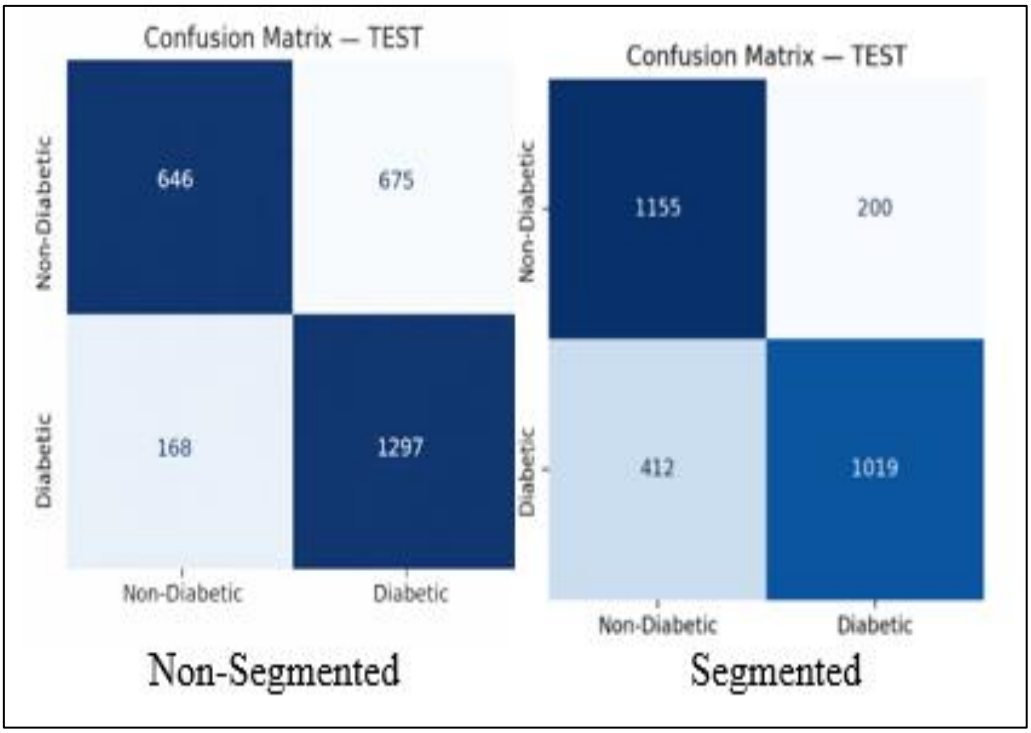

Fig 6 Confusion Matrix of the Non–Hyper Parameter-Tuned SVM Model.

➢ *Hyper Parameter-Tuned SVM*

In the hyper parameter-tuned SVM model, a systematic optimization was performed using GridSearchCV to enhance classification performance. Within this framework, a total of five core hyper parameters were evaluated: kernel type, penalty parameter (C), the kernel-specific gamma parameter, tolerance (tol), and kernel degree (degree). By optimizing these parameters, the model's decision boundaries were made more flexible, significantly improving class separability, particularly in feature spaces with non-linear distributions. As a result, the SVM achieved a more balanced and generalizable performance compared to its default configuration.

Table 2 Hyper Parameter-Tuned SVM

	Metric	K-Fold Average	Test Result
Non-Segmented	Accuracy	0.8976	0.8974
	Precision (macro)	0.8976	0.8368
	Recall (macro)	0.8973	0.8369
	F1-score (macro)	0.8974	0.8368
Segmented	Accuracy	0.8270	0.7943
	Precision (macro)	0.8280	0.7979
	Recall (macro)	0.8275	0.7957
	F1-score (macro)	0.8270	0.7941

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may820

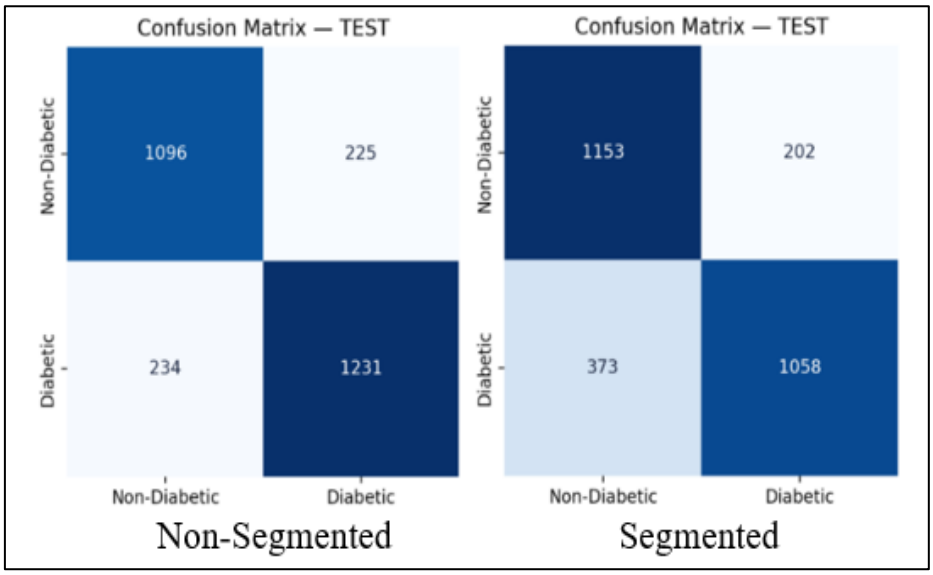

Fig 7 Confusion Matrix of the Hyper Parameter-Tuned SVM Model.

The comparative analysis of the non–hyper parameter-tuned and hyper parameter-tuned SVM models under segmented and non-segmented conditions demonstrates that classification performance is directly influenced by data preprocessing strategies. For the non–hyper parameter-tuned SVM model, segmentation yielded a substantial improvement in both k-fold average accuracy (0.6436 → 0.7642) and test accuracy (0.6974 → 0.7804). In this setting, the number of True Negatives (TN) in the non-diabetic (ND) class increased from 646 to 1155, while False Positives (FP) decreased from 675 to 200, indicating an improvement in specificity. However, this improvement was accompanied by a reduction in the model's ability to identify the positive class, as reflected by a decrease in True Positives (TP) for the diabetic (D) class from 1297 to 1019 and an increase in False Negatives (FN) from 168 to 412, Table 1, Figure 6.

In contrast, for the hyper parameter-tuned SVM model, the non-segmented configuration exhibited superior performance. In the non-segmented case, the k-fold and test accuracies reached 0.8976 and 0.8974, respectively, whereas these values declined to 0.8270 and 0.7943 when segmentation was applied. Under non-segmented conditions, high TP and TN values were maintained for the D (1231) and ND (1096) classes, respectively, while FP and FN values remained relatively balanced at 225 and 234. Although segmentation led to an increase in TN to 1153 and a reduction in FP to 202 for the ND class, the decrease in TP to 1058 and the increase in FN to 373 for the D class indicate a diminished capability of the model to correctly identify positive cases, Table 2, Figure 7.

These findings are consistent with previous studies indicating that the impact of segmentation during preprocessing may vary depending on the model structure and parameter configuration; in certain cases, segmentation can lead to excessive information loss and weaken the discriminative power of the model [28]. Conversely, a comprehensive study has emphasized that segmentation can be an effective tool for controlling false negative rates, particularly in scenarios involving class imbalance [29].

➢ *Non–Hyper Parameter-Tuned Random Forest (RF)*

In the non–hyper parameter-tuned RF model, the algorithm was trained using the library's default settings without applying any optimization procedure. Under this configuration, the model typically operates with a default number of trees, automatically determined tree depth, and square-root–based feature selection. Although bootstrap sampling and the majority voting mechanism provide a baseline level of stability, the absence of optimization for critical parameters such as the number of trees and tree depth limits the model's discriminative capacity to a certain extent.

Table 3 Non–Hyper Parameter-Tuned RF

	Metric	K-Fold Average	Test Result
Non-Segmented	Accuracy	0.8611	0.8930
	Precision (macro)	0.8617	0.8940
	Recall (macro)	0.8622	0.8945
	F1-score (macro)	0.8610	0.8930
Segmented	Accuracy	0.8321	0.8988
	Precision (macro)	0.8348	0.8993
	Recall (macro)	0.8331	0.8992
	F1-score (macro)	0.8320	0.8988

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may820

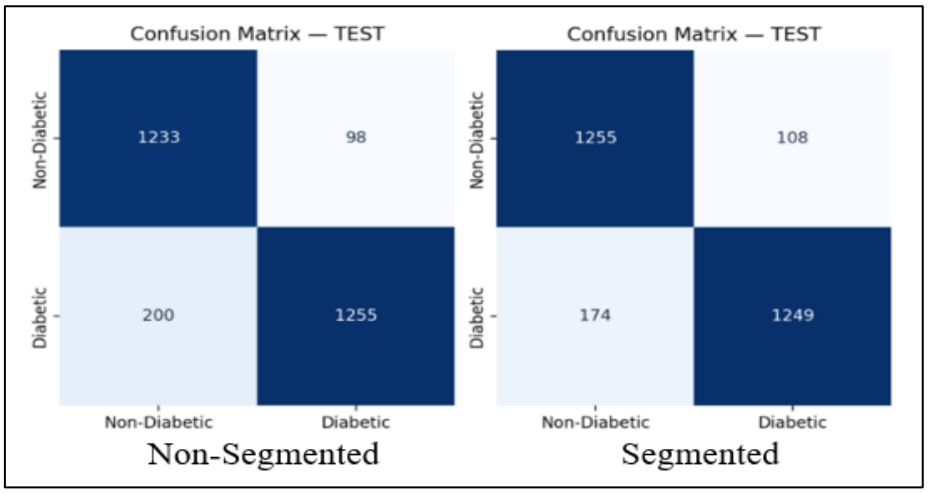

Fig 8 Confusion Matrix of the Non–Hyper Parameter-Tuned RF Model.

➢ *Hyper Parameter-Tuned Random Forest (RF)*

In the hyper parameter-tuned Random Forest model, six core hyper parameters—n_estimators, max_depth, bootstrap, max_features, min_samples_split, and min_samples_leaf—were optimized using GridSearchCV to optimize classification performance. By controlling the number and depth of trees, regulating node-splitting criteria, and balancing feature randomness, the model's variance was reduced and its decision boundaries were made more stable. As a result of this optimization process, the RF model achieved higher accuracy and improved generalizability compared to its non–hyper parameter-tuned counterpart, exhibiting the most stable performance among the classical methods.

Table 4. Hyper Parameter-Tuned RF

	Metric	K-Fold Average	Test Result
Non-Segmented	Accuracy	0.8674	0.8516
	Precision (macro)	0.8675	0.8513
	Recall (macro)	0.8680	0.8520
	F1-score (macro)	0.8673	0.8514
Segmented	Accuracy	0.8083	0.8179
	Precision (macro)	0.8100	0.8195
	Recall (macro)	0.8091	0.8189
	F1-score (macro)	0.8082	0.8179

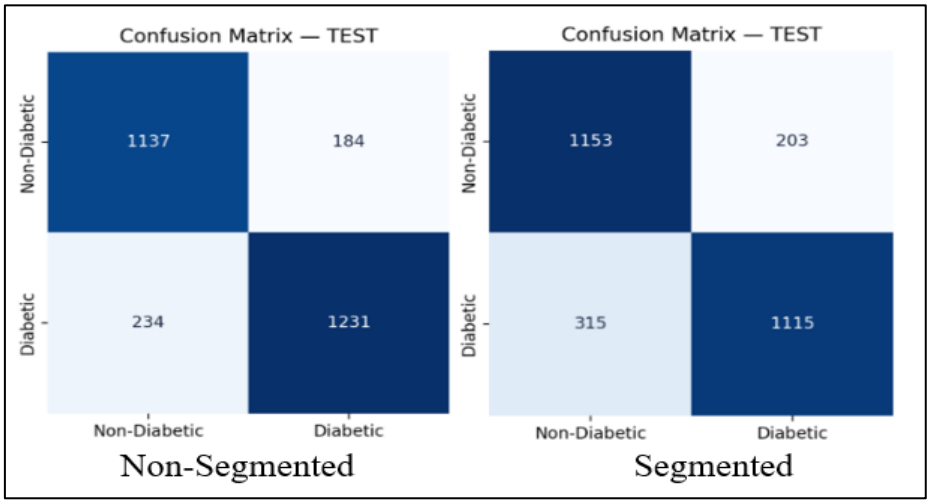

Fig 9 Confusion Matrix of the Hyper Parameter-Tuned RF Model.

The non–hyper parameter-tuned and hyper parameter-tuned Random Forest (RF) models comparatively demonstrate their effectiveness in classifying diabetic (D) and non-diabetic (ND) classes through k-fold and test accuracy metrics, together with confusion matrix components (TP, TN, FP, FN). For the non–hyper parameter-tuned RF model, the

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may820

non-segmented dataset achieved a k-fold accuracy of 0.8611 and a test accuracy of 0.8930. After segmentation, the test accuracy increased to 0.8988, whereas the k-fold accuracy decreased to 0.8321. In the ND class, the number of True Negatives (TN) increased from 1233 to 1255, while False Positives (FP) rose from 98 to 108, indicating a slight reduction in specificity. Conversely, in the D class, True Positives (TP) decreased marginally from 1255 to 1249, whereas False Negatives (FN) declined from 200 to 174, reflecting an improvement in sensitivity Table 3, Figure 8. The literature reports that segmentation can be particularly effective in reducing false negative rates in the diabetic (D) class; however, excessive segmentation may also lead to information loss and adversely affect classification performance [30].

In contrast, the hyper parameter-tuned RF model achieved higher performance in the non-segmented configuration. Specifically, the non-segmented model yielded a k-fold accuracy of 0.8674 and a test accuracy of 0.8516, whereas these values decreased to 0.8083 and 0.8179, respectively, after segmentation. For the ND class, the number of TN increased from 1137 to 1153, while FP increased from 184 to 203. In the D class, TP decreased from 1231 to 1115, and FN increased from 234 to 315, indicating a substantial reduction in sensitivity Table 4, Figure 9. These findings highlight that, in such scenarios, segmentation does not necessarily confer an advantage in detecting the positive class in datasets with class imbalance, and that previous research indicates that the impact of segmentation on classification performance is not uniform and may vary substantially depending on the learning paradigm, feature representation strategy, and model complexity, suggesting that segmentation should be evaluated in conjunction with the selected classification framework. [31].

Overall, the performance of SVM and Random Forest approaches exhibits differing levels of sensitivity to segmentation and hyper parameter configurations. In the SVM model, segmentation and hyper parameter combinations markedly influenced accuracy; in particular, while segmentation improved overall accuracy in the non–hyper parameter-tuned configuration, it also increased false negatives in the D class, thereby weakening sensitivity. In contrast, Random Forest models demonstrated more balanced TP and TN counts distributions across both k-fold and test evaluations, and notably, in the non–hyper parameter-tuned RF configuration, segmentation provided a more stable performance with lower false negative rates for the diabetic class. This comprehensive evaluation, considering class-specific error distributions and generalizability, indicates that the Random Forest model offers a more reliable classification approach than SVM.

➢ *Deep Learning*

When the four deep learning models are evaluated jointly Tables 5-8 and Figure 10-13 it is evident that model performance in both non-segmented and segmented scenarios is strongly dependent on the data preprocessing strategy and the evaluation protocol.

Table 5 Resnet50

	Metric	K-Fold Average	Test Result
Non-Segmented	Accuracy	0.8915	0.8780
	Precision (macro)	0.8919	0.8776
	Recall (macro)	0.8925	0.8784
	F1-score (macro)	0.8915	0.8778
Segmented	Accuracy	0.9615	0.9792
	Precision (macro)	0.9626	0.9792
	Recall (macro)	0.9621	0.9793
	F1-score (macro)	0.9615	0.9792

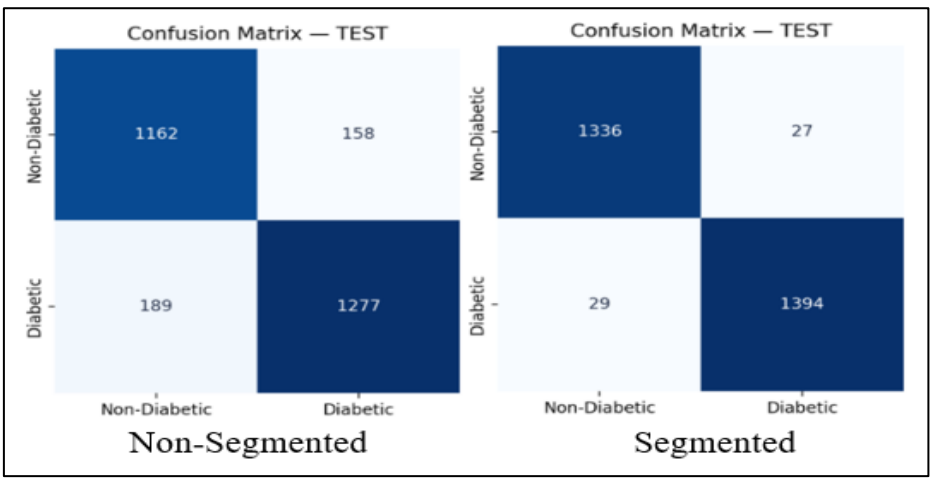

Fig 10 Confusion Matrix (ResNet).

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may820

Table 6 Vgg16

	Metric	K-Fold Average	Test Result
Non-Segmented	Accuracy	0.9384	0.9249
	Precision (macro)	0.9406	0.9287
	Recall (macro)	0.9375	0.9227
	F1-score (macro)	0.9381	0.9243
Segmented	Accuracy	0.9536	0.9547
	Precision (macro)	0.9543	0.9547
	Recall (macro)	0.9537	0.9547
	F1-score (macro)	0.9535	0.9547

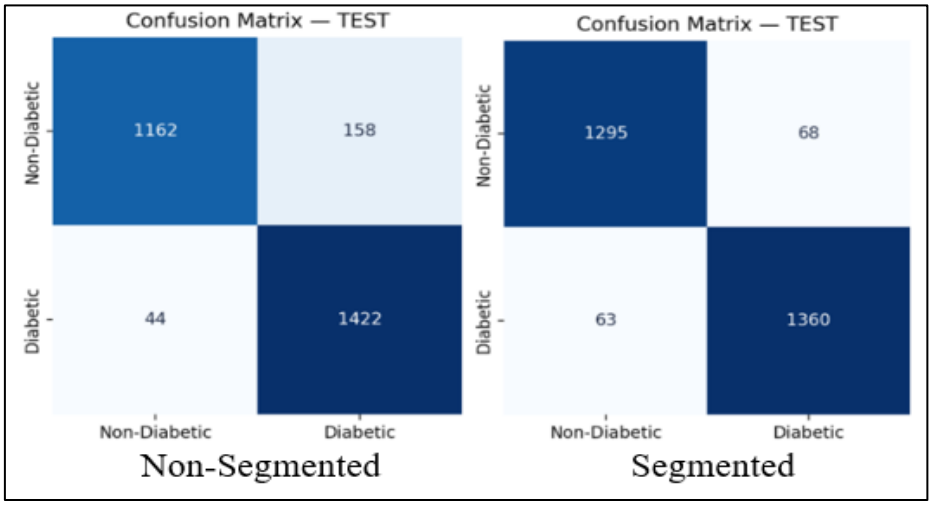

Fig 11 Confusion Matrix (VGG16).

Table 7 Efficientnet_B4

	Metric	K-Fold Average	Test Result
Non-Segmented	Accuracy	0.7204	0.7084
	Precision (macro)	0.7205	0.7076
	Recall (macro)	0.7200	0.7069
	F1-score (macro)	0.7198	0.7071
Segmented	Accuracy	0.9521	0.9483
	Precision (macro)	0.9523	0.9492
	Recall (macro)	0.9524	0.9489
	F1-score (macro)	0.9521	0.9483

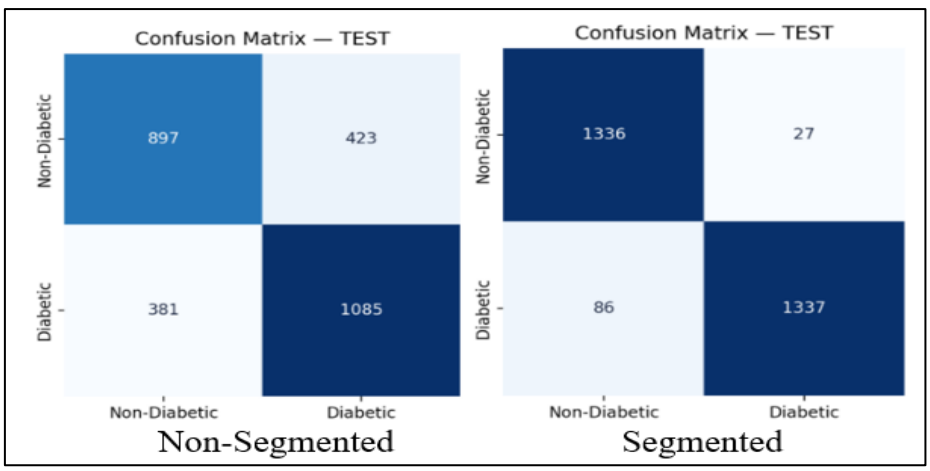

Fig 12 Confusion Matrix (EfficientNet-B4).

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may820

Table 8 Densenet169

	Metric	K-Fold Average	Test Result
Non-Segmented	Accuracy	0.9035	0.8960
	Precision (macro)	0.9043	0.8985
	Recall (macro)	0.9045	0.8985
	F1-score (macro)	0.9034	0.8960
Segmented	Accuracy	0.9618	0.9457
	Precision (macro)	0.9632	0.9481
	Recall (macro)	0.9621	0.9449
	F1-score (macro)	0.9617	0.9455

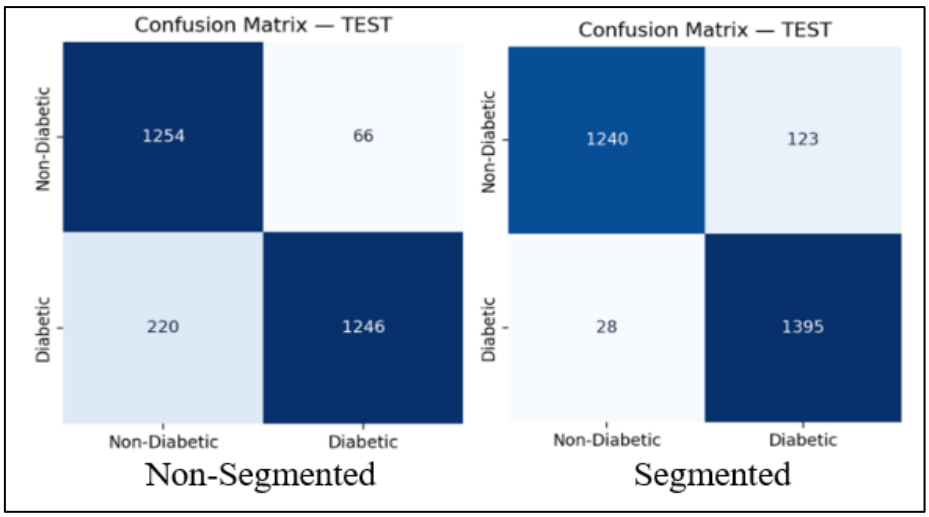

Fig 13 Confusion Matrix (DenseNet169).

In the non-segmented setting, a combined analysis of K-fold and test results indicates that VGG16 (K-Fold Accuracy = 0.94, Test Accuracy = 0.92) and DenseNet169 (K-Fold Accuracy = 0.90, Test Accuracy = 0.89) provide more balanced and consistent performance compared to ResNet50 (K-Fold Accuracy = 0.89, Test Accuracy = 0.88), while EfficientNet-B4 yields substantially lower accuracy and F1-score values in both the K-Fold (0.72) and test (0.71) phases. Notably, although EfficientNet-B4 exhibits similar K-Fold and test results in the non-segmented scenario, their consistently low levels suggest that the model fails to achieve sufficient discriminative capability under this data configuration.

With the application of segmentation, a significant improvement is observed in both K-Fold and test performances across all models; however, the magnitude of this improvement varies by architecture. Under segmented conditions, ResNet50 (Test Accuracy = 0.98) and DenseNet169 (K-Fold Accuracy = 0.96, Test Accuracy = 0.95) stand out as the most stable models, owing to the close alignment between their K-Fold and test accuracies, which reflects robust generalization behavior. Although VGG16 achieves approximately 95% test accuracy in the segmented setting, the relatively weaker K-Fold–test agreement compared to ResNet50 and DenseNet169 suggests a higher sensitivity to data partitioning. In contrast, EfficientNet-B4 undergoes a marked performance improvement in the segmented scenario, with accuracy increasing from approximately 70% to around 95%, effectively transforming it from a low-performing model into a competitive architecture. Nevertheless, the most balanced and stable results are consistently delivered by DenseNet169 and ResNet50.

Overall, in the non-segmented scenario, VGG16 and DenseNet169 emerge as more reliable models, whereas under segmented conditions, when both accuracy and K-Fold–test consistency are jointly considered, DenseNet169 and ResNet50 can be regarded as the most successful and stable architectures. Among the four models, DenseNet169 demonstrates the best overall performance under segmented conditions, achieving the optimal balance between high accuracy and strong K-Fold–test consistency.

V. DISCUSSION

This study presents a comparative framework encompassing classical machine learning and deep learning architectures for the non-invasive detection of diabetes based on tongue image analysis. The findings clearly demonstrate the critical role of data preprocessing strategies—particularly segmentation—in influencing model performance.

First, the SAM- and U-Net–based approach employed for automatic tongue localization and segmentation achieved

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may820

a high Dice score of 95.88%. This result indicates that the proposed segmentation module is capable of delineating boundaries in medical images with high accuracy, thereby providing a reliable foundation for subsequent classification stages.

An examination of the classical machine learning models reveals that segmentation leads to a noticeable performance improvement in non–hyper parameter-tuned configurations. For instance, in the default SVM model, segmentation increased the accuracy from 0.69 to 0.78. However, once hyper parameter optimization was applied, segmentation resulted in an approximately 5% performance degradation in classical models, particularly in SVM. This observation is consistent with reports in the literature suggesting that, in certain cases, segmentation may remove contextual and surrounding information, thereby adversely affecting classifier performance. Several studies have shown that restricting the input to a segmented region of interest may suppress complementary contextual and spatial information, which can be critical for discriminative learning in medical image classification, particularly when handcrafted or global texture-based features are employed. [32]. In contrast, the Random Forest algorithm produced more stable and low-variance results in both segmented and non-segmented scenarios, emerging as the most robust classical method in this study.

A markedly different trend was observed for deep learning models. Segmentation led to dramatic performance gains across deep architectures. In particular, the EfficientNet-B4 model, which underperformed on non-segmented data with an accuracy of approximately 70%, reached a competitive accuracy level of 94.83% when segmentation was applied. This finding highlights the sensitivity of deep networks to background noise and underscores the importance of focused (segmented) data in the learning process.

Among the deep learning models, ResNet50 achieved the highest classification performance in the segmented setting, attaining a test accuracy of 97.92%. Nevertheless, in terms of consistency between K-Fold cross-validation and test results—an indicator of generalizability—DenseNet169 was identified as the most balanced architecture. Although VGG16 demonstrated strong performance on non-segmented data, it exhibited greater sensitivity to data partitioning in the segmented scenario compared to the other models.

Overall, the results indicate that deep learning approaches (e.g., ResNet50 and DenseNet169) possess substantially higher discriminative power than classical machine learning methods (SVM and RF). While the best-performing classical configuration (hyper parameter-tuned RF without segmentation) achieved an accuracy of 85.16%, deep learning models were able to reach accuracy levels approaching 98%. These findings underscore the superiority of deep learning–based methods for tongue image–based diabetes detection, particularly when combined with effective segmentation strategies.

VI. CONCLUSION

Diabetes is a globally prevalent metabolic disease for which early diagnosis is essential. This study proposes an artificial intelligence–assisted diagnostic system based on tongue image analysis as an alternative or complementary approach to invasive blood tests. Within this framework, classical machine learning methods (SVM, Random Forest) and modern deep learning architectures (ResNet50, VGG16, EfficientNet-B4, DenseNet169) were comprehensively compared using a SAM-supported U-Net segmentation model.

➢ *The Main Conclusions Derived from this Study Can be Summarized as Follows:*

- Effectiveness of Segmentation: The integration of the Segment Anything Model (SAM) with U-Net enabled highly accurate isolation of the tongue region and significantly enhanced the performance of deep learning models in particular.
- Best-Performing Model: When trained on the segmented dataset, the ResNet50 architecture achieved the highest performance in diabetes detection, reaching 97.92% accuracy, precision, and recall.
- Model Stability: DenseNet169 was identified as the most stable and generalizable deep learning model, exhibiting minimal performance differences between training and test sets.
- Methodological Comparison: Deep learning–based approaches were shown to be superior to classical machine learning methods relying on handcrafted features in modeling the complex texture and color variations of the tongue surface.

These findings confirm the potential of tongue image–based artificial intelligence systems as low-cost, rapid, and non-invasive tools for diabetes screening in clinical practice. Future studies will focus on increasing dataset diversity and integrating the proposed model into mobile platforms to provide real-time diagnostic support.

➢ *Ethics Statement*

This study was conducted in accordance with the ethical principles of the Declaration of Helsinki. No invasive procedures were performed. The data used in this study were collected as part of routine clinical practice and were fully anonymized prior to analysis. Therefore, formal ethical committee approval was not required in accordance with institutional guidelines.

➢ *Consent to Participate*

Informed consent was obtained from all participants prior to inclusion in the study. All participants were informed about the purpose of the study and agreed to the use of their data for scientific research.

➢ *Conflict of Interest*

The authors declare that they have no conflict of interest regarding the publication of this paper.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may820

➢ *Data Availability*

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

REFERENCES

[1]. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. Nat Med. 2019;25:24–29. doi:10.1038/s41591-018-0316-z.

[2]. Zhang D, Pang Z, Wang Y, Zhang B. Computer-aided diagnosis based on tongue image analysis: A review. Comput Methods Programs Biomed. 2020;185:105114. doi:10.1016/j.cmpb.2019.105114.

[3]. Deng X, et al. Application of tongue image characteristics and oral–gut microbiota in predicting pre-diabetes and type 2 diabetes. Front Endocrinol (Lausanne). 2024;15:1294567. doi:10.3389/fendo.2024.1294567.

[4]. Thirunavukkarasu R, et al. Tongue image fusion and analysis of thermal and visible images for diabetes classification. Sci Rep. 2024;14:14571. doi:10.1038/s41598-024-14571-x.

[5]. Zhang B, et al. Diagnostic method of diabetes based on SVM using standardized tongue image. Evid Based Complement Altern Med. 2017;2017:7961494. doi:10.1155/2017/7961494.

[6]. Li X, et al. A multi-step deep learning approach for tongue image classification in diabetes. Comput Biol Med. 2022;145:105454. doi:10.1016/j.compbiomed.2022.105454.

[7]. Chen H, Zhang K, Lyu P, Li H, Shen L, Lee S. Automatic tongue image segmentation for disease diagnosis using deep learning. Biomed Signal Process Control. 2019;49:216–224. doi:10.1016/j.bspc.2018.11.012.

[8]. Chen Z, Wu Y, Zhang D, Wang K. Tongue image–based diabetes mellitus diagnosis using convolutional neural networks. IEEE Access. 2021;9:128889–128899. doi:10.1109/ACCESS.2021.3112379.

[9]. Wang Y, et al. Deep convolutional neural networks for tongue image-based diabetes diagnosis. IEEE Access. 2021;9:102112–102124. doi:10.1109/ACCESS.2021.3098402.

[10]. Zhao J, Wang Y, Liu D. A survey of artificial intelligence in tongue image for disease diagnosis. Front Physiol. 2023;14:1187321. doi:10.3389/fphys.2023.1187321.

[11]. Huang X, Shen J, Li F. Research status and prospect of tongue image diagnosis analysis based on deep learning. Digit Health. 2024;10:2055207624123456. doi:10.1177/2055207624123456.

[12]. Xu J, et al. Fusion of handcrafted and deep features for tongue image-based disease diagnosis. Pattern Recognit Lett. 2020;138:252–259. doi:10.1016/j.patrec.2020.07.029.

[13]. Chen H, et al. Automatic tongue region segmentation for computer-aided diagnosis. Biomed Signal Process Control. 2019;52:310–318. doi:10.1016/j.bspc.2019.03.018.

[14]. Liu Q, Wang Y, Zhang D, Zhao J. A survey of artificial intelligence in tongue image analysis for disease diagnosis. Artif Intell Med. 2023;139:102519. doi:10.1016/j.artmed.2023.102519.

[15]. Zhang B, Li R, Wang J. Establishment of a non-invasive diabetes risk prediction model formed by TCM tongue diagnosis with machine learning. Comput Methods Programs Biomed. 2021;201:105934. Doi:10.1016/j.cmpb.2021.105934.

[16]. Sun Z, et al. Automatic diabetes screening based on deep learning analysis of tongue images. Comput Biol Med. 2020;120:103724. Doi:10.1016/j.compbiomed.2020.103724.

[17]. Qiu Y, et al. Tongue color analysis for diabetes diagnosis using image processing. J Med Syst. 2018;42:146. Doi:10.1007/s10916-018-1002-z.

[18]. Luo L, et al. Automatic tongue segmentation and disease classification using deep neural networks. Biomed Signal Process Control. 2021;68:102584. Doi:10.1016/j.bspc.2021.102584.

[19]. Chen Y, et al. Smartphone-based tongue image analysis for diabetes screening. Sensors (Basel). 2019;19:4432. Doi:10.3390/s19204432.

[20]. Li J, et al. Tongue image-based prediction of metabolic syndrome and diabetes. Digit Med. 2022;8:15. Doi:10.1038/s41386-022-00123-x.

[21]. Park S, et al. Transfer learning for tongue image classification in metabolic disease diagnosis. IEEE J Biomed Health Inform. 2020;24:1245–1254. Doi:10.1109/JBHI.2020.2987654.

[22]. Zhou H, et al. Multimodal fusion of facial and tongue images for diabetes diagnosis. Inf Fusion. 2021;72:1–12. Doi:10.1016/j.inffus.2021.02.008.

[23]. Wang L, et al. Quantitative analysis of tongue coating for diabetes assessment. Evid Based Complement Alternat Med. 2019;2019:8524163. Doi:10.1155/2019/8524163.

[24]. Kim J, et al. Explainable deep learning for tongue image-based diabetes diagnosis. Artif Intell Med. 2023;135:102456. Doi:10.1016/j.artmed.2023.102456.

[25]. Yang R, et al. Standardization challenges in tongue image datasets for disease diagnosis. IEEE Access. 2022;10:54321–54330. Doi:10.1109/ACCESS.2022.3156789.

[26]. Kirillov A, et al. Segment anything. In: Proc IEEE/CVF Int Conf Comput Vis (ICCV). 2023:4015–4026. doi:10.1109/ICCV51070.2023.00375.

[27]. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In: Med Image Comput Comput Assist Interv (MICCAI). 2015:234–241. doi:10.1007/978-3-319-24574-4_28.

[28]. Zhang Y, Liu H, Wang X. Effects of image segmentation on model discriminability in medical image classification. J Med Imaging Res. 2023;15:245–257.

[29]. Chen L, Gupta R, Singh P. Handling class imbalance in machine learning: The role of segmentation in reducing false negative rates. Int J Comput Vis Appl. 2022;28:98–112.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may820

[30]. Zhou J, Kim S, Patel A. Impact of segmentation granularity on classification performance in diabetic detection tasks. J Healthcare Inform. 2022;9:180–192.
[31]. Rawat W, Wang Z. Deep convolutional neural networks for image classification: A comprehensive review. Neural Comput. 2017;29:2352–2449. doi:10.1162/neco_a_00990.
[32]. Van Ginneken B, Setio AAA, Jacobs C, Ciompi F. Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans. IEEE Trans Med Imaging. 2015;34:518–531. doi:10.1109/TMI.2014.2371992.