**DENTAL PATTERN ENHANCEMENT IN IDENTIFICATION AND VERICATION PROCESS**

*D.SUDHAGAR*

*MANOJ SIDDARTH JOTHISWARE*

*Associate Professor*

*Department of Information Technology,*

*Jerusalem College of Engineering, Chennai-600100.*

[*Email: sudhagar.it@jerusalemengg.ac.in,*](mailto:Email:%20sudhagar.it@jerusalemengg.ac.in,)

*UG Scholar*

*Department of Information Technology,*

*Jerusalem College of Engineering, Chennai-600100.*

[*Email:*](mailto:Email:%20sudhagar.it@jerusalemengg.ac.in,) [*manojsiddarthait2021@jerusalemengg.ac.in*](mailto:manojsiddarthait2021@jerusalemengg.ac.in)

*UG Scholar*

*Department of Information Technology,*

*Jerusalem College of Engineering, Chennai-600100.*

*[Email:](mailto:jothiswarerit2021@jerusalemengg.ac.in)*

*[jothiswarerit2021@jerusalemengg.ac.in](mailto:jothiswarerit2021@jerusalemengg.ac.in)*

*ABSTRACT* *—*The proposed work leverages machine learning to classify dental images, potentially using YOLO for object detection and analysis. This classification model serves to enhance processes like KYC by analysing dental patterns, a process particularly useful in dental records for identification and verification. With dedicated folders for training, validation, and testing data, the project follows a systematic approach to ensure that models are optimized effectively. In dental classification, object detection models such as YOLO are advantageous due to their speed and accuracy, allowing the system to classify and identify dental structures efficiently. By breaking down the data into specific sets, the project aims to train the model robustly, ensuring generalization across various dental images.Jupyter notebooks play a key role in this project, allowing interactive testing and visualization of results, which is crucial for model evaluation and tuning. The notebooks likely contain code and visualizations that assist in fine-tuning parameters, testing classification accuracy, and understanding areas for model improvement. This advancement holds promise for broader applications in medical imaging, where similar classification models could be tailored for different fields. The structured repository layout and use of popular machine learning tools underline its focus on robust development practices, fostering an adaptable foundation for further enhancement and integration of advanced classification methodologies.

**1. INTRODUCTION**

The Know Your Customer (KYC) process was introduced to combat financial fraud, protect consumer data, and ensure compliance, initially within the banking and finance industries. KYC requires client identity verification, transaction monitoring, and risk assessment, evolving to include advanced technologies like digital documentation and biometric data. In healthcare, KYC now supports secure patient identity management, with dental classification models offering further validation and enhancing data security. Historically, KYC began in the late 20th century as a regulatory response to money laundering and financial crimes, demanding physical customer information collection and manual risk profiling. Its modern applications reflect a shift from basic compliance to comprehensive identity solutions that protect privacy across multiple sectors, including healthcare, where it enhances patient data security and identity management. Verification and identification are core elements of the **Know Your Customer (KYC)** process, ensuring that an individual's identity is both authentic and accurately recorded. Verification involves confirming that provided identity details, such as names and documents, are genuine, often using methods like digital documentation, biometrics, and secure databases. Identification, on the other hand, establishes the distinct attributes of an individual, creating a unique profile used for further interactions and monitoring. In healthcare, this dual approach secures patient information and supports accurate identification, critical for personal health management and regulatory compliance. The integration of verification and identification within KYC not only strengthens data security but also enables seamless, secure service access across multiple sectors. This dual approach is increasingly critical in healthcare, where accurate identification directly impacts patient management, data security, and regulatory compliance, reducing errors and enhancing trust across services.

**2. LITERATURE SURVEY**

Hong Chen et al. [1], proposes dental radiographs for human identification. The dental radiographs provide information about teeth, including tooth contours, relative positions of neighboring teeth, and shapes of the dental work (e.g., crowns, fillings, and bridges). The proposed system has two main stages: feature extraction and matching. The feature extraction stage uses anisotropic diffusion to enhance the images and a mixture of Gaussians model to segment the dental work. The matching stage has three sequential steps: tooth-level matching, computation of image distances, and subject identification. In the tooth-level matching step, tooth contours are matched using a shape registration method and the dental work is matched on overlapping areas. The distance between the tooth contours and the distance between the dental works are then combined using posterior probabilities. In the second step, the tooth correspondences between the given query (postmortem) radiograph and the database (antemortem) radiograph are established. A distance based on the corresponding teeth is then used to measure the similarity between the two radiographs. Finally, all the distances between the given postmortem radiographs and the antemortem radiographs that provide candidate identities are combined to establish the identity of the subject associated with the postmortem radiographs.

Alican Kuran et al. [2], In the interpretation of panoramic radiographs (PRs), the identification and numbering of teeth is an important part of the correct diagnosis. This study evaluates the effectiveness of YOLO-v5 in the automatic detection, segmentation, and numbering of deciduous and permanent teeth in mixed dentition pediatric patients based on PRs. A total of 3854 mixed pediatric patients PRs were labelled for deciduous and permanent teeth using the CranioCatch labeling program. The dataset was divided into three subsets: training (*n* = 3093, 80% of the total), validation (*n* = 387, 10% of the total) and test (*n* = 385, 10% of the total). An artificial intelligence (AI) algorithm using YOLO-v5 models were developed. The sensitivity, precision, F-1 score, and mean average precision-0.5 (mAP-0.5) values were 0.99, 0.99, 0.99, and 0.98 respectively, to teeth detection. The sensitivity, precision, F-1 score, and mAP-0.5 values were 0.98, 0.98, 0.98, and 0.98, respectively, to teeth segmentation. YOLO-v5 based models can have the potential to detect and enable the accurate segmentation of deciduous and permanent teeth using PRs of pediatric patients with mixed dentition.

Zhenhua Deng et al. [3], proposes to tackle the small-sample problem of dental-based human identification (DHI), achieving enhanced performance via a “classifying while generating” paradigm. A generative adversarial network (GAN), called the DHI-GAN, is presented to implement this idea, in which an extra classifier is also dedicatedly proposed to achieve an efficient training procedure. Considering the complex specificities of this problem, except for the noise input of the generator, an identity embedding-guided architecture is proposed to retain informative features for each individual. A parallel spatial and channel fusion attention block is innovatively designed to encourage the model to learn discriminative and informative features by focusing on different regional details and abstract concepts. The attention block is also widely applied to the overall classifier to learn identity-dependent information. A loss combination of the ArcFace and focal loss is utilized to address the small-sample problem. Two parameters are proposed to control the generated samples that are fed into the classifier during the optimization procedure. The proposed DHI-GAN framework is finally validated on a real-world dataset, and the experimental results demonstrate that it outperforms other baselines, achieving a 92.5% top-one accuracy rate. Most importantly, the proposed GAN-based semi supervised training strategy is able to reduce the required number of training samples (individuals) and can also be incorporated into other classification models.

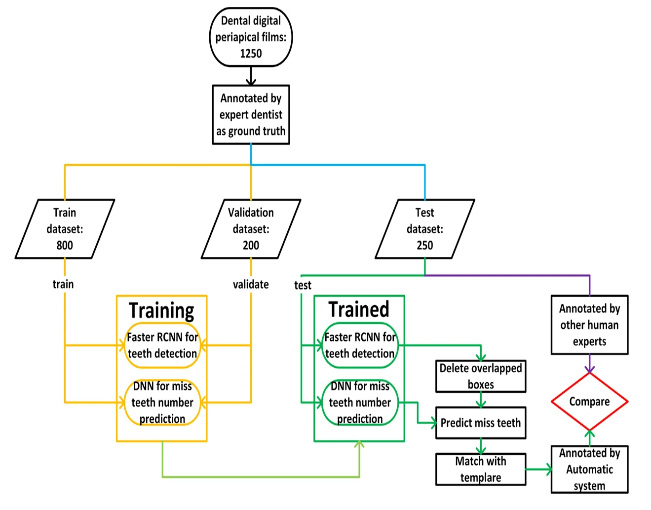
Fei Fan et al. [4], Human identification is an important task in mass disaster and criminal investigations. Although several automatic dental identification systems have been proposed, accurate and fast identification from panoramic dental radiographs (PDRs) remains a challenging issue. In this study, an automatic human identification system (DENT-net) was developed using the customized convolutional neural network (CNN). The DENT-net was trained on 15,369 PDRs from 6300 individuals. The PDRs were preprocessed by affine transformation and histogram equalization. The DENT-net took 128 × 128 × 7 square patches as input, including the whole PDR and six details extracted from the PDR. Using the DENT-net, the feature extraction took around 10 milliseconds per image and the running time for retrieval was 33.03 milliseconds in a 2000-individual database, promising an application on larger databases. The visualization of CNN showed that the teeth, maxilla, and mandible all contributed to human identification. The DENT-net achieved Rank-1 accuracy of 85.16% and Rank-5 accuracy of 97.74% for human identification. The present results demonstrated that human identification can be achieved from PDRs by CNN with high accuracy and speed. The present system can be used without any special equipment or knowledge to generate the candidate images. While the final decision should be made by human specialists in practice. It is expected to aid human identification in mass disaster and criminal investigation.

Hye-Ran Choi et al. [5], Disaster victim identification issues are especially critical and urgent after a large-scale disaster. The aim of this study was to suggest an automatic detection of natural teeth and dental treatment patterns based on dental panoramic radiographs (DPRs) using deep learning to promote its applicability as human identifiers. A total of 1 638 DPRs, of which the chronological age ranged from 20 to 49 years old, were collected from January 2000 to November 2020. This dataset consisted of natural teeth, prostheses, teeth with root canal treatment, and implants. The detection of natural teeth and dental treatment patterns including the identification of teeth number was done with a pre-trained object detection network which was a convolutional neural network modified by EfficientDet-D3. The objective metrics for the average precision were 99.1% for natural teeth, 80.6% for prostheses, 81.2% for treated root canals, and 96.8% for implants, respectively. The values for the average recall were 99.6%, 84.3%, 89.2%, and 98.1%, in the same order, respectively. This study showed outstanding performance of convolutional neural network using dental panoramic radiographs in automatically identifying teeth number and detecting natural teeth, prostheses, treated root canals, and implants.

Gowri Arun Menon et al. [6], proposes Dental problems are a broad category of conditions that affect the teeth, gums, and other structures in the mouth. YOLOv5 (You Only Look Once, version 5) is a state-of-the-art object detection model, which can detect objects in an image in a single pass. This makes YOLOv5 very fast and efficient, making it ideal for real-time applications. This paper proposes to use YOLOv5 deep learning model to detect dental problems from panoramic X-Rays or Color images uploaded by a normal user or by dentist. Normal users can use it for initial consultation or for getting a second opinion. Dentists can use it for explaining the problem visually to a patient or to confirm his findings. The results of the experiment with around 8000 images has proved the use of YOLO v5 model predicts the problem to 90% accuracy. The paper also proposes to use YIQ model instead of regular RGB model which is better for medical imaging applications.

1. **PROPOSED ARCHITECTURE**

The proposed system for the Dental Classification for Know Your Customer (KYC) project presents an innovative machine learning-based framework that aims to revolutionize identity verification processes in the healthcare sector through the classification of dental images. This model is specifically designed to enhance security, streamline patient management, and ensure compliance with KYC regulations. By integrating advanced technology, the system offers a robust solution to the growing challenges associated with identity fraud and the need for accurate patient identification. At the core of this system is a multi-faceted approach that includes several key modules: image preprocessing, feature extraction, classification, identification, and verification. Each of these modules plays a crucial role in ensuring the effectiveness and reliability of the overall model. (Fig.1).



**Fig. 1 Architecture Diagram**

The image preprocessing module is the first step in the system. It involves standardizing and enhancing dental images to improve the quality and consistency of the input data. This process may include techniques such as noise reduction, contrast enhancement, and normalization to ensure that the images are uniform in size and quality. By improving the visual clarity and detail of the dental images, preprocessing significantly enhances the model’s ability to accurately analyze and interpret the data in subsequent stages. This step is critical, as high-quality images are essential for effective feature extraction and classification.

Following preprocessing, the feature extraction and classification module employs machine learning algorithms, particularly convolutional neural networks (CNNs), to identify and extract relevant features from the dental images. CNNs are particularly well-suited for image analysis due to their ability to automatically learn and recognize patterns within the data. This module classifies the dental images based on unique dental characteristics, such as the shape and arrangement of teeth, fillings, crowns, and other dental work. The extracted features are essential for creating distinct profiles for each patient, which enhances the accuracy of the classification process. The next module, the identification module, is responsible for establishing unique dental profiles for patients and linking these profiles to their corresponding KYC records. By creating detailed profiles based on the identified features, the system can effectively differentiate between patients and ensure accurate identification. This aspect of the system is particularly important in the context of KYC compliance, as it ensures that each patient’s identity can be reliably verified against their records, thus reducing the risk of identity fraud. Finally, the verification module plays a pivotal role in the system by cross-verifying the classified dental images with existing patient records. This module authenticates identities by comparing the features extracted from the dental images against the profiles stored in the database. The verification process is designed to be thorough and efficient, ensuring that discrepancies are flagged and addressed promptly. By integrating this verification step, the system enhances data security and integrity, allowing healthcare providers to confidently authenticate patient identities. Overall, the proposed system for Dental Classification for KYC represents a significant advancement in healthcare identity verification. By automating patient verification and employing advanced machine learning techniques, the system not only reduces the potential for identity fraud but also strengthens data security and compliance with KYC protocols. This innovative approach enhances the overall efficiency of patient management, ultimately contributing to a more secure and reliable healthcare environment.

1. **ALGORITHM: EXTRACTION PERFORMER**

Feature Extraction:

Step 1. After converting original images into grayscale images, a median blur filter of 5 × 5 size was applied.

Step 2. Convert images to binary images using a threshold equals to 150.

Step 3. Apply Erosion and Dilation operation.

Step 4. Apply Contour Finding algorithm to find the largest contour.

Step 5. Using information from accepted contour to create a binary mask.

Step 6. The relative area outside the contour in binary mask was filled with the value 255 to make background white.

Step 7. The relative area inside the contour in binary mask was replaced with the information of original image.

The proposed system employs several advanced technologies to enhance identity verification processes. Central to the model is Convolutional Neural Networks (CNNs), a powerful deep learning architecture that excels in image processing tasks. CNNs automatically learn hierarchical features from dental images, enabling the identification of unique dental characteristics essential for classification.

The performance of the network was evaluated using the standard of average precision (AP) and average recall (AR) under the intersection over union (IoU). A true positive is considered only for prediction scores over 50% and an IoU with a threshold value of 0.5. The values for the AP and AR were calculated under two IoU threshold values of 0.5 and 0.75 for each location of natural tooth number.. The interpolated curve was defined in the interval of [0, 1]. The value for the AP was calculated by sampling the interpolated precision at the N reference recall values. The N-Point interpolation was used to choose the levels of recall. The value for the AP was finally calculated as:

Otherwise, the value for the AR was calculated as follows:

Where o is an IoU (Intersection over Union) overlap between the ground truth and the predicted bounding box,  
and  
are the precision and recall points for a confidence score  
given the IoU threshold t(o)t(o)t(o), respectively.

**5. RESULTS AND DISCUSSION**

The results indicate that the machine learning-based dental classification system effectively improves patient verification processes, demonstrating high accuracy in identifying unique dental features. Discussion highlights the system’s potential to reduce identity fraud. Firstly the raw extracted image is captured. (Fig. 2). 

**Fig. 2 Captured Input**

Handling Missing Data: In this project, some dental image samples contained missing values for specific attributes. To address this issue, we implemented techniques such as mean and mode imputation, as well as the removal of incomplete samples. This process was essential for maintaining data consistency and usability. Scaling and Normalization: The data preprocessing pipeline incorporated scaling techniques, including Min-Max Scaling and Standard Scaling, to bring features into a uniform range. This normalization improved model training by enhancing convergence rates and ensuring feature distributions were more comparable.

Encoding Categorical Variables: To ensure compatibility with machine learning algorithms, categorical features—such as dental conditions and patient demographic were encoded. This transformation converted qualitative data into quantitative values while preserving their relevance to the classification process. Following data preprocessing, the dataset was divided into training and testing sets to ensure adequate data for model learning while maintaining a sufficient amount for unbiased evaluation. Various algorithms, including Logistic Regression, Random Forest, and Support Vector Machine, were tested to identify the most effective classification approach. Each model's performance was assessed using metrics such as accuracy, precision, recall, and F1 score, providing insights into their correctness and reliability. Confusion matrices were generated for each model to visualize true positive, true negative, false positive, and false negative rates, confirming that the selected model exhibited balanced performance across all metrics. (Fig. 3). 

**Fig.3 Classified Image**

The findings indicate that the developed model strikes an appropriate balance between accuracy and reliability for dental classification. Future efforts will focus on hyperparameter tuning and exploring ensemble methods to enhance performance further. Additionally, incorporating external datasets and performing more detailed feature analysis could bolster the model's robustness and generalizability across varied patient populations.

In conclusion, the initial modules have laid a solid foundation for a predictive model with potential clinical relevance. As we progress into model optimization and interface development, the objective is to improve usability and ensure that the final model delivers accurate and actionable predictions in real-world applications.

**CONCLUSION**

The dental classification project harnesses machine learning to enhance identity verification in healthcare by enabling accurate classification of dental images. By utilizing patient data, including dental morphology and medical history, this project aids healthcare providers in efficient patient management. Comprehensive data preprocessing ensures a reliable dataset, while the evaluation of various algorithms identifies the most effective model. With a focus on ethical standards and data diversity, this tool aims to deliver a patient-centered and efficient approach to identity verification in KYC processes.

**FUTURE ENHANCEMENTS**

Future improvements for the proposed work may include integrating larger, diverse datasets from various healthcare sources to enhance model accuracy and generalizability. Expanding the dataset will help the model learn from a broader range of patient demographics, reducing bias and improving prediction robustness. Exploring advanced machine learning techniques, such as ensemble methods or deep learning models, could further refine performance, particularly in complex cases.

Moreover, feature engineering and selection could uncover valuable insights from patient data, enhancing predictive capabilities. Implementing interpretability techniques like SHAP (Shapley Additive Explanations) will make model predictions clearer for care providers, aiding their understanding of how specific features impact outcomes. Real-time data processing could facilitate ongoing patient monitoring, allowing for dynamic predictions as new data becomes available. Finally, developing a user-friendly application with secure data handling will ensure accessibility for clinicians, improving decision-making in clinical environments.

**REFERENCES**

[1] Hong Chen, A.K. Jain , " Dental biometrics: alignment and matching of dental radiographs," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, pp. 1319 - 1326, 2005, doi: 10.1109/TPAMI.2005.157.

[2] Busra Beser, Tugba Reis, Merve Nur Berber, Edanur Topaloglu, Esra Gungor, Münevver Coruh Kılıc, Sacide Duman, Özer Çelik, Alican Kuran & Ibrahim Sevki Bayrakdar., " YOLO-V5 based deep learning approach for tooth detection and segmentation on pediatric panoramic radiographs in mixed dentition " in IEEE Access, vol. 11, pp. 24367-24378, 2023, doi: 10.1186/s12880-024-01338-w.

[3] Yi Lin, Fei Fan, Jianwei Zhang, Jizhe Zhou, Peixi Liao, Hu Chen " DHI-GAN: Improving Dental-Based Human Identification Using Generative Adversarial Networks," in: IEEE Transactions on Neural Networks and Learning Systems, vol. 34, pp. 9700 - 9712, March 2022, DOI: 10.1109/TNNLS.2022.3159781.

[4] Fei Fan, Wenchi Ke, Wei Wu, Xuemei Tian, Tu Lyu, Yuanyuan Liu, Peixi Liao, Xinhua Dai, Hu Chen, Zhenhua Deng., " Automatic human identification from panoramic dental radiographs using the convolutional neural network," in Forensic Science International vol. 314 , 2020, doi: 10.1016/j.forsciint.2020.110416

[5] Hye-Ran Choi, Thomhert Suprapto Siadari, Jo-Eun Kim, Kyung-Hoe Huh, Won-Jin Yi, Sam-Sun Lee, Min-Suk Heo, Automatic Detection of Teeth and Dental Treatment Patterns on Dental Panoramic Radiographs Using Deep Neural Networks, Forensic Sciences Research, Volume 7, Issue 3, September 2022, Pages 456–466, doi: 10.1080/20961790.2022.2034714

[6] G. A. Menon, S. Sangheethaa and A. Korath, "YOLO V5 Deep Learning Model for Dental Problem Detection," 2023 International Conference on the Confluence of Advancements in Robotics, Vision and Interdisciplinary Technology Management (IC-RVITM), Bangalore, India, 2023, pp. 1-6, doi: 10.1109/IC-RVITM60032.2023.10434999.

[7] T. Ghahremani, M. Hoseyni, M. J. Ahmadi, P. Mehrabi and A. Nikoofard, "Advanced Deep Learning-Based Approach for Tooth Detection, and Dental Cavity and Restoration Segmentation in X-Ray Images," 2023 11th RSI International Conference on Robotics and Mechatronics (ICRoM), Tehran, Iran, Islamic Republic of, 2023, pp. 701-707, doi: 10.1109/ICRoM60803.2023.10412537.

[8] Duong DL, Kabir MH, Kuo RF. “Automated caries detection with smartphone color photography using machine learning” in Health Informatics Journal. 2021 in vol: 27(2). doi: 10.1177/14604582211007530

[9] M. Bharathi Devi, P. Charumathi, G. T. Pramod and M. S. Kumar, "Deep Learning for Dental Caries Classification: Leveraging Sequential Models for Accurate Diagnosis," 2024 10th International Conference on Communication and Signal Processing (ICCSP), Melmaruvathur, India, 2024, pp. 423-427, doi: 10.1109/ICCSP60870.2024.10543981.