Computing&AI Connect
ISSN: 3104-4719
2024, Article ID. 2004.0002, Cite as: https://doi.org/10.69709/CAIC.2024.194138

SCIFINITI PUBLISHING

Research Article

OPEN ACCESS

Mirror, Mirror on the Wall: Automating Dental Smile Analysis with AI in Smart Mirrors

Mariia Baidachna[✉,1,*] **Haneen Fatima**[✉,2] **Rahaf Omran**[✉,3] **Nour Ghadban**[✉,2] **Muhammad Ali Imran**[✉,2] **Ahmad Taha**[✉,2] **Lina Mohjazi**[✉,2,*]

[1] School of Computing Science, University of Glasgow, 18 Lilybank Gardens, Glasgow G12 8RZ, UK
[2] James Watt School of Engineering, University of Glasgow, Glasgow G12 8QQ, UK
[3] Align Technology, 2800 The Crescent, Solihull Parkway, Birmingham Business Park, Solihull B37 7YL, UK

Article History
Received: May 14, 2024 Accepted: July 15, 2024 Published: July 29, 2024

Abstract
This paper presents a smart diagnostic framework for dental smile analysis. To accurately and efficiently identify esthetic issues from a single image of a smile, a convolutional neural network (CNN) was trained. To overcome the limitations of scarce data, a diffusion model was employed to generate dental smile images in addition to manually curated data. The CNN was trained and evaluated on three datasets: all real images, all generated images, and a hybrid dataset comprising of equal proportions of real-to-generated images. All three models demonstrate accuracy significantly above the baseline in detecting excessive gingival display, unlocking a novel diagnostic method in smile analysis. Notably, the hybrid model achieved the highest accuracy of 81.61% (p - value < 0.01), highlighting the effectiveness of generative data augmentation for machine learning. The proposed solution could be part of a standalone home-deployed smart mirror or connected to a network of innovative Internet-of-Mirrors to facilitate patient-dentist communication.

Keywords:
smile analysis; machine learning; computer vision; convolutional neural networks; generative artificial intelligence; Internet-of-Things; connected healthcare

1. Introduction

The rapid growth in the number of connected devices manifested by the Internet-of-Things (IoT) and the expanding capabilities of artificial intelligence (AI) has set the stage for accelerated growth in digital beauty and healthcare [1,2]. However, a significant gap is evident in the landscape of the dental sector, particularly in aesthetic smile analysis. Closing this gap is imperative, given the escalating demand for smile rectification [3]. The importance of aesthetic smile treatment for many patients arises from the impact a person's smile has on various interpersonal aspects, including attractiveness, self-confidence, perceived success and intelligence, social acceptance, and psychological well-being [3,4]. In some cases, an undesirable smile may lead to neglect in oral self-care, increasing the prevalence of periodontal problems [5]. Orthodontic treatment of smile aesthetic issues, sometimes in the form of surgical intervention, results in an improvement of many of these variables [6,7]. However, before a patient can undergo treatment, the specific elements of the smile that require correction must be analyzed. At present, the evaluation of esthetic appeal is confined to human eye assessment of smile elements, which have been heavily researched [8]. One such element, the lip line, pertains to the extent of maxillary teeth displayed upon smiling [9]. An ideal lip line exposes a range of three-quarters to the full clinical crown of maxillary teeth and only the interdental gingiva. A low lip line exposes less than three-quarters of the clinical crown height, while a high lip line exposes more than the full clinical crown with a continuous band of gingiva [10]. Figure 1 shows a high lip line before and after treatment. Among aesthetic issues, gingival display was selected as the focus of this study due to

* **Corresponding Author:**
Mariia Baidachna, School of Computing Science, University of Glasgow, 18 Lilybank Gardens, Glasgow G12 8RZ, UK, 2828197b@student.gla.ac.uk
Lina Mohjazi, James Watt School of Engineering, University of Glasgow, Glasgow G12 8QQ, UK, lina.mohjazi@glasgow.ac.uk

© 2024 Copyright by the Authors.
Licensed as an open access article using a CC BY 4.0 license.

2024, Vol. 1, Article ID. 2004.0002
https://doi.org/10.69709/CAIC.2024.194138

its widespread occurrence, significant influence on smile attractiveness, and recognition as the most distracting elements by dental students [3,11,12].

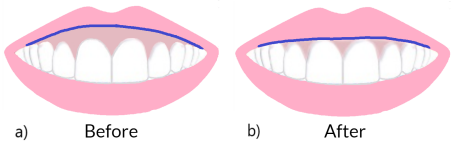


Figure 1: Presence of excessive gingival display in (a) before orthodontic treatment. An ideal lip line after treatment shown in (b). The blue line shows the location of the lip line.

As clinical dentistry progresses towards being computerized, digital documentation of photos has become a standard procedure [8]. This shift is significant as paper-based photographs can lead to a substantial overestimation of a smile's perceived attractiveness [13]. Additionally, biases such as subjective esthetic appreciation and the speciality of the assessing dentist can affect the assessment [8]. Different aspects of the smile are noticed by dentists of different specialties [14]. This discrepancy in clinical observations raises the question of whether general dentists or newly graduated clinicians without substantial experience might benefit from having "another set of eyes" when making complex clinical decisions.

Another drawback of the conventional methods is the delay between patients initiating appointments and receiving rectification, sometimes taking months before redirection to a professional [15]. Long waiting times are one of the major factors negatively affecting patient experience, patient-provider relationship, and the patient return rates to the same provider [16]. While review appointments play a crucial role in tracking treatment progress, a more efficient system connecting patients with healthcare providers could be developed to minimize prolonged waiting times. Similar challenges arise in the broader medical sector due to the workforce deficit in the UK National Health Service, which has been struggling to recover post the COVID-19 pandemic, resulting in the worst-recorded waiting times for cancer care [17]. This trend is reflected globally, with the shortage of healthcare workers projected to be higher in future post-pandemic calculations [18]. To counter this, there has been a drastic shift towards remote consultation and digital healthcare [19,20]. However, digital smile analysis is lagging behind other healthcare sectors. To the best of our knowledge, there are no reliable, non-invasive, time-efficient diagnostic tools or post-treatment trackers available for smile esthetics evaluation.

The limitations of existing dental aesthetic diagnosis methods prompt consideration of novel solutions. For instance, AI and machine learning (ML) are currently employed in numerous healthcare fields and have the potential to improve accuracy, increase efficiency, and aid in decision-making processes. This could revolutionize daily practices and have positive societal and environmental impacts [21]. Moreover, wireless communication networks pave the way for enhanced connectivity, enabling a multitude of applications that can revolutionize diagnostic tools by ensuring efficiency and timely delivery of results through the IoT [2].

Various areas have benefited from connected technology, where IoT systems serve as the backbone of smart devices. Integrating the IoT into healthcare has the power to improve the quality of life [22,23]. We hypothesize that combining the IoT with ML and big data analysis can non-invasively, or without the need for intrusive and surgical procedures, resolve many limitations of traditional healthcare, particularly within smile analysis. ML drives the IoT in two main aspects: network communication and application-specific analysis. Network problems, such as routing, traffic, and resource control, have been successfully solved with ML models [24]. Additionally, ML can be used to detect patterns in data received from the IoT sensors for diagnosing patients. For example, ML has been applied to analyze heart pulse, blood pressure, and temperature sensor input to identify high blood pressure, elevated heart rate, or even heart attacks [25]. The full integration of ML into an IoT system equipped with a camera would allow for an automatic non-invasive smile analysis tool.

1.1. Related Work

Before ML can be safely integrated into diagnostic IoT devices, robust model training and analysis must be conducted to ensure accurate inference. Regarding dental smile analysis, previous research efforts have focused on automating the recognition of dental issues through advanced computer vision (CV) techniques and ML algorithms. There is a multitude of existing ML techniques, and it is imperative that the most suitable ones are applied to particular problems. Recent years have seen a surge of a subfield of ML called deep learning (DL). DL is empowered by interconnected layers of neural networks (NN) and has been adapted to many applications in the scope of healthcare and CV [26]. There is a range of model architectures within DL for CV with their respective trade-offs in performance, resources, and latency.

Previous studies that employed DL in dental analysis have adapted variable architectures with varying degrees of success. One notable study [27] chose a combination of detection and segmentation, employing an architecture within DL called Mask Region-based Convo-

2024, Vol. 1, Article ID. 2004.0002
https://doi.org/10.69709/CAIC.2024.194138

lutional Neural Network (R-CNN). Mask R-CNN builds upon Faster R-CNN by introducing an additional branch dedicated to predicting masks at the pixel level, which are calculated based on rigorous labeling shown in Figure 2, for discerning and effectively segmenting relevant objects [28]. Due to the absence of publicly available datasets, [27] the study collected and annotated only 100 images, achieving a pixel accuracy of tooth segmentation between 90.1% and 97.4% for natural teeth and lower for dentures. The observed high accuracy is consistent with the nature of Mask R-CNN methods, which prioritize performance over speed. An analysis of a single frame using Mask R-CNN typically takes a few seconds, depending on the data to be analyzed and the number of inferences to be made [29]. This delay will add up with image preprocessing and routing, and it will be multiplied by the number of frames to be analyzed. For example, if three images are taken for averaged results or different angles for other applications, the delay can be 10-30 seconds. In pursuit of scalable real-time results and a smooth user experience, another study [30] employed an extension of You Only Look Once (YOLO), the YOLACT++ instance segmentation model, analyzing a dataset comprising 5500 images of faces distributed across diverse classes. While the model exhibited high accuracy in detecting facial features like the nose, eyebrows, and eyes, a significant decline in precision was observed when segmenting the gingiva and buccal corridor. This decrease in precision may be attributed to a relatively lower number of annotated instances per class for these specific regions in the training dataset, highlighting the importance of a well-balanced dataset in DL. Another study [31] employed routine clinical practices to acquire 1250 X-ray images for their objective of tooth detection and identification with faster R-CNN, citing the high cost of instance segmentation labeling as their rationale.

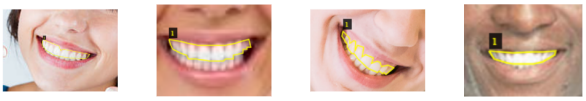


Figure 2: A sample of images with labeled maxillary teeth required to create an attribute-value pair JSON file of pixels to labels for instance segmentation using an annotation tool called VGG Image Annotator.

The common thread in all these studies is the use of relatively small datasets, which often prevents unlocking the full potential of NN and leads to poor generalization [32]. While the growth of big data has undeniably improved the capabilities of NN, particularly in the realm of CV, notable data gaps persist. In our case, this is evident due to the absence of an adequate dataset capturing dental smiles. Models trained on small datasets have a tendency to result in overfitting or not being able to generalize on unseen testing data, performing with high accuracy during training but low accuracy during testing. There is a multitude of techniques under data augmentation that compensate for the lack of data by artificially expanding the dataset. Data augmentation in CV makes for better DL models by making multiple copies of the same image with diverse geometric transformations, color augmentations, kernel filters, random erasing, and the application of generative artificial intelligence (GenAI) [33]. The latter, unlike the other listed methods, synthetically generates new data points by capturing and changing distinctive objects within the image and generating new images with additional noise to increase performance on unseen images [34]. GenAI, in particular, is promising in CV, especially with the advancements in text-to-image generation capabilities in recent years [33,35].

Numerous studies have been taking advantage of generative adversarial networks (GANs), a subclass of GenAI to generate plausible data and augment small or imbalanced datasets. One study used 13 small medical datasets to show that the optimal proportion of GAN augmentation significantly enhances the performance of all ML classifiers [36]. Similarly, another study built an emotion classification model and used GAN augmentation to increase samples of less common classes such as disgust, which increased classification accuracy by 5%–10% [37]. The study [32] observed a similar trend, employing GAN data augmentation to amplify an X-ray dataset and improving the CNN model for pneumonia and COVID-19 detection. Other generative techniques have been used to solve similar problems in different areas. For instance, [38] used autoencoders to successfully generate multivariate data for fire scenarios, taking slope, vegetation, and other factors into account. This shows the wide range of problems generative models could be leveraged in to aid classical ML in analyzing complex data.

While there is notable progress in data augmentation in the field of medical image analysis for tooth identification and segmentation, the issue of data scarcity remains a challenge for smile classification. This challenge is even more pronounced given that a comprehensive dental smile dataset is not yet available, limiting advancements in dental smile analysis. To address this challenge and fill this research gap, this study presents a novel and accurate framework for dental smile analysis empowered by the combination of DL methods and GenAI-empowered data augmentation. To the best of our knowledge, this is the first study to explore the application of ML in dental smile analysis, which has the potential to be applied not only in dig-

ital dental esthetic technology in clinics and hospitals but also in standalone home devices for accurate and efficient smile analysis.

1.2. Contribution and Vision

In this paper, we propose a novel diagnostic smile analysis tool for excessive gingiva detection, leveraging a CNN model and GenAI data augmentation. We employ a sequential topology architecture for its simplicity and adaptability. As with previously outlined studies, a persistent challenge in our research stems from the insufficient number of data available to generalize to unseen content. To resolve this issue, we curate a dataset of a total of 512 dental smile images from available open-source images and text-to-image AI-generated samples. The inclusion of AI-generated data from Adobe Firefly's diffusion model [39] significantly enhances classification accuracy, demonstrating the efficacy of our methods. The usage of text-to-image generative models in data augmentation is still an early practice, but we show that it has potential in dentistry. It enables us to attain higher quality results, for example, detecting excessive gingival display correctly on previously unseen images with 81.607% accuracy, without incurring the costs associated with annotating segmentation data.

Furthermore, we integrated the trained CNN model into a minimalistic user interface (UI) that guides the user to the correct position and captures a single image. The image is subsequently preprocessed according to specifications and passed through the pre-trained CNN model. The image is then analyzed in the backend, and the UI displays the result of the smile analysis obtained from the model in real-time, forming an end-to-end application.

This application could be stand alone or integrated into an innovative IoT smart mirror technology, termed as the Internet-of-Mirrors (IoM) [40]. This visionary ecosystem of interconnected smart mirrors could be used to integrate the smile analysis application, as well as other digital health and beauty applications. Telehealth services, such as teledentistry, are time and cost-effective ways of health-related communication [41]. The motivation lies in connecting patients with healthcare professionals or products in a timely manner, as depicted in Figure 3. In this system, each mirror would be equipped with sensors, such as a camera, to acquire user input. In the context of smile analysis, the mirror would take the dental smile image, crop it according to the specifications, and use pre-trained CNN model to produce an output and inform the patient of their smile condition. The results of potential dental issues and personalized suggestions would be displayed on the UI. Smart mirrors placed in homes would serve users by promptly delivering tailored products or healthcare recommendations to align with their individual needs, and effectively facilitating connections between patients and professionals. Furthermore, smart mirrors placed in clinics could resolve the smile analysis subjectivity issue by providing a consistent basis for diagnosis.

To summarize, the main contributions of the paper are the following:

- **Novel Diagnostic Smile Analysis Tool:** We introduce a CNN model trained on real and AI-generated images of dental smiles to detect excessive gingival display, providing a non-invasive digital diagnostic tool.
- **Dental Smile Dataset:** A dataset of 512 dental smile images with a mix of real and AI-generated images from Adobe Firefly's diffusion model, leading to a significant improvement of classification accuracy and addressing the challenge of limited data points for dental image analysis.
- **Application Integration:** Integration of the proposed solution into an application, offering a smooth human-computer interaction experience and instantaneously displaying the results of the diagnosis and suggestions.

2. Materials and Methods

In this section, we outline the methodology employed for our research, detailing the data collection process, preprocessing steps, and the architecture of the NN used for smile classification.

2.1. Dataset

To train CNN models to an acceptable degree, a large dataset is required. There are many resources of fully curated and often readily preprocessed datasets of images for various CV uses online. In the context of dental smiles, the absence of a suitable dataset in open-source repositories posed a significant challenge. Consequently, the ethical collection of a diverse array of dental images encompassing various ethnicities, genders, and degrees of gingival display became a persistent endeavor.

In total, two sets of 256 (counting 512 combined) images were collected and evenly distributed into "gummy" and "normal" classes. The first set was manually curated from publicly available images scrapped from online frontal face images of people smiling. This set included samples showing highly positive excessive gingival displays as well as more discrete displays. The second set was obtained from text-to-image AI-generated images using Adobe Firefly's diffusion model. The text prompts for

2024, Vol. 1, Article ID. 2004.0002
https://doi.org/10.69709/CAIC.2024.194138

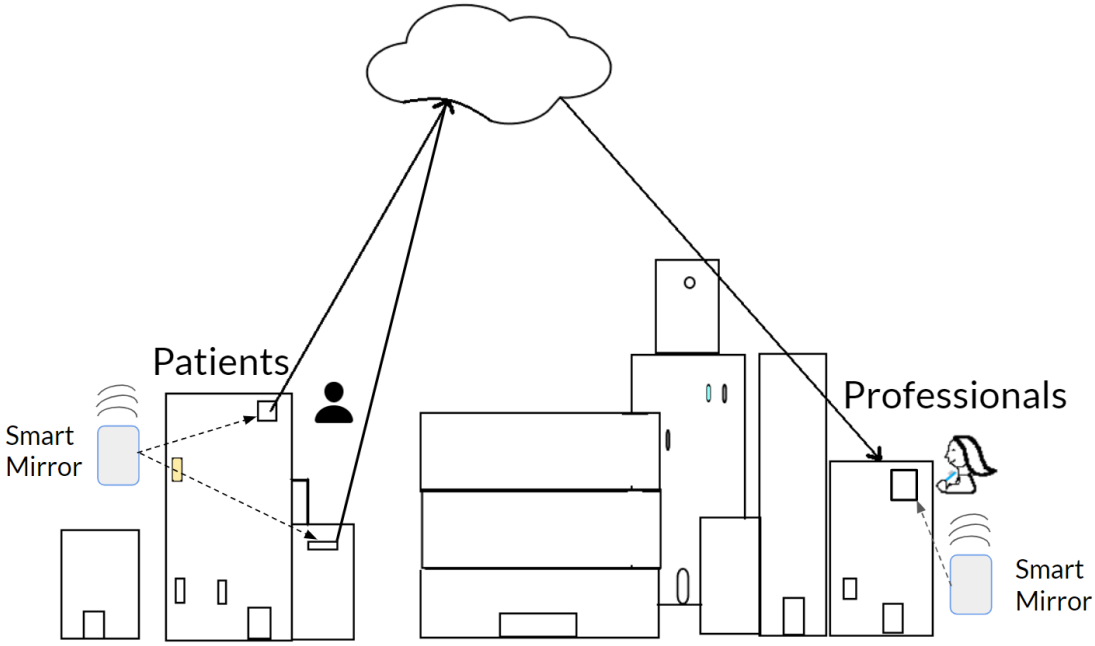


Figure 3: Interaction possibilities between professional dentists and patients as proposed in [40]. The diagram depicts each patient equipped with a smart mirror that delivers information to dentists over a wireless network. The patient can then be redirected to a professional through booking an appointment. The professionals can also rely on smart mirrors for providing a basis in subjective diagnosis. This vision of interconnected dentistry would unlock the potential of building smart cities through a similar system.

the "normal" class ranged from "frontal portrait of a person smiling with teeth; this person does not have a gummy smile or braces" to simply selecting relevant images in "frontal portrait of a person smiling". It is important to note that smile analysis with braces was excluded from the training dataset, as individuals undergoing treatment do not typically undergo aesthetic interventions until treatment is completed.

Generating excessive gingival display proved to be more challenging. The text-to-image diffusion model excels in comprehending daily human speech, but its grasp of medical terminology remains somewhat restricted from our observations. Changing terms like "excessive gingival display" to more commonly used phrases like "gummy smile" resulted in images of people chewing gum for several images. Providing special tags such as "three millimeters between the teeth and the upper lip" worked better, though not all candidates qualified. In the end, the prompts varied throughout the data collection, ensuring only relevant images were added. Despite requiring some filtering, this approach still consumed less time and resources compared to manually annotating teeth with tools like LabelMe or VGG Annotator. This is due to the fact that segmentation labels require 20–30 points to precisely trace the pixels of the teeth to be applied to the image. On the contrary, labeling classes only requires one identifying label per image.

Both sets contributed 56 randomly chosen images each to the testing folder, setting aside 112 of the original 512 images for testing. The remaining 400 images (200 from real, 200 AI-generated by the diffusion model) were randomly split with 80% assigned to training and 20% to validation. In total, three separate datasets were curated: real images (256 images), AI-generated images (256 images), and combined images (512 images). Compared to other CV classification datasets, the model trained on this dataset was more susceptible to overfitting. Therefore, it was vital that the dataset's quality was superlative for the model to classify unseen images accurately. This was ensured by balancing a diverse range of individual frontal smile images evenly between the two classes and validating the correctness of the labels with a professional dentist.

2.2. Preprocessing

Proactive measures were taken in the preprocessing step to mitigate the risk of overfitting. Initially, all images were normalized and converted to grayscale using the OpenCV library. Following this, a pre-trained facial landmark detector from the Dlib library was employed to identify the coordinates of the mouth in each image. Subsequently, every image was cropped to a size of 28 by 28 pixels, centered around the mouth. This systematic preprocessing step was consistently applied to both training, validation,

and testing data, ensuring the overall quality and consistency of the dataset. Due to the nature of this preprocessing approach, images of users' faces were not saved but were funneled through the pipeline up to the cropping stage. This approach mitigated many privacy and ethical concerns related to data retention and destruction.

Before feeding these images into the model, multiple augmentation layers were applied to artificially expand the dataset. In addition to the generated images, each training image underwent a horizontal flip, three random zoom factors between -0.2 and 0.2, random contrast with a factor between 0.25 and 0.75, and random brightness with a factor between -0.5 and 0.5. This approach increased the size of each dataset by a factor of seven (the original image and six augmentation layers), resulting in a total of 3584 images.

The entire process pipeline is illustrated in Figure 4. which presents a flowchart from receiving raw images to preprocessing and training, using an example Adobe Firefly image.

2.3. Model Architecture

A CNN classifier model was built using Tensorflow and Keras to classify previously unseen images into either "gummy" or "normal" smile classes. A sequential model architecture was chosen for its simplicity, offering a linear topology that enables the customization of layer stacking while maintaining one input tensor and one output tensor. This choice was driven by the primary goal of minimizing the risk of overtraining, as small datasets tend not to generalize well to unseen testing data. The overall architecture of the model follows a standard layer sequence. The pattern of a 2D convolution layer followed by a 2D max pooling is used twice. The rectified linear unit (ReLU) function, represented by R(z) = max(0, z), served as the activation parameter. The ReLU activation function was also used in the first of the three densely-connected NN layers. The second dense layer used the softmax s(xi) function, which takes a vector xi and out- puts the probability distribution of each class for N classes, denoted by:

$$s(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{N} e^{x_j}}. \quad (1)$$

The final dense layer converted the output into a probability distribution across the classes. Each model, distinguished by real, AI-generated, and combined datasets, was trained using the Adam optimizer. Through hyperparameter tuning, the optimal number of epochs was determined to be 150, and the batch size was set to 32. All other parameters remained constant throughout training sessions.

2.4. Mirror User Interface

The envisaged IoM calls for the integration of an interactive dashboard. The interface of the dashboard was originally designed in-house, as shown in Figure 5 by the CSI group at JWSE at the University of Glasgow. This served as a base for a prototype UI to connect to the pre-trained smile analysis model. A basic layout of applications with corresponding functionalities was created using modular programming, allowing for flexible addition of buttons with corresponding function calls to accommodate the evolving functionalities of the application. For instance, ML models trained to detect and classify skin diseases could be paired with the Skin Analysis button in the backend function call without altering the overall structure of the code.

Upon pressing the Smile Analysis button, the system initiates video capture from the Intel RealSense Depth Camera D435, accompanied by real-time red bounding box tracking of the user's face on the screen. Simultaneously, pre-trained Haar cascades for face and smile detection operate in the background. The bounding box changes color to green when the user maintains a smile for two seconds, at which point an image is captured. A demo of the entire process can be viewed here https://youtu.be/yj_FavXCL2I. The captured image then enters the standardized preprocessing pipeline and undergoes analysis within the classification model, yielding the resulting class instantaneously on the screen.

3. Results

In the results section, we first present the statistics of the best-performing model, which consists of real and AI-generated data. We then compare these results to those of the uniformly trained models. Lastly, we briefly explain the results of the segmentation model.

3.1. Gingival Display CNN Model

The research yielded several key findings. First, we evaluated the performance of the CNN model for detection of excessive gingival display. The model employing diffusion model data augmentation in addition to geometric transformations achieved an average accuracy of 81.607% with a standard deviation of 1.938. This statistic reflects the average accuracy across 10 trials, each involving training the model over 150 epochs and evaluating its performance on 56 previously unseen images to ascertain the accuracy of label predictions. The standard deviation was then calculated using the standard formula, where the sum of the squared differences is divided by the number of data points denoted by n:

2024, Vol. 1, Article ID. 2004.0002
https://doi.org/10.69709/CAIC.2024.194138

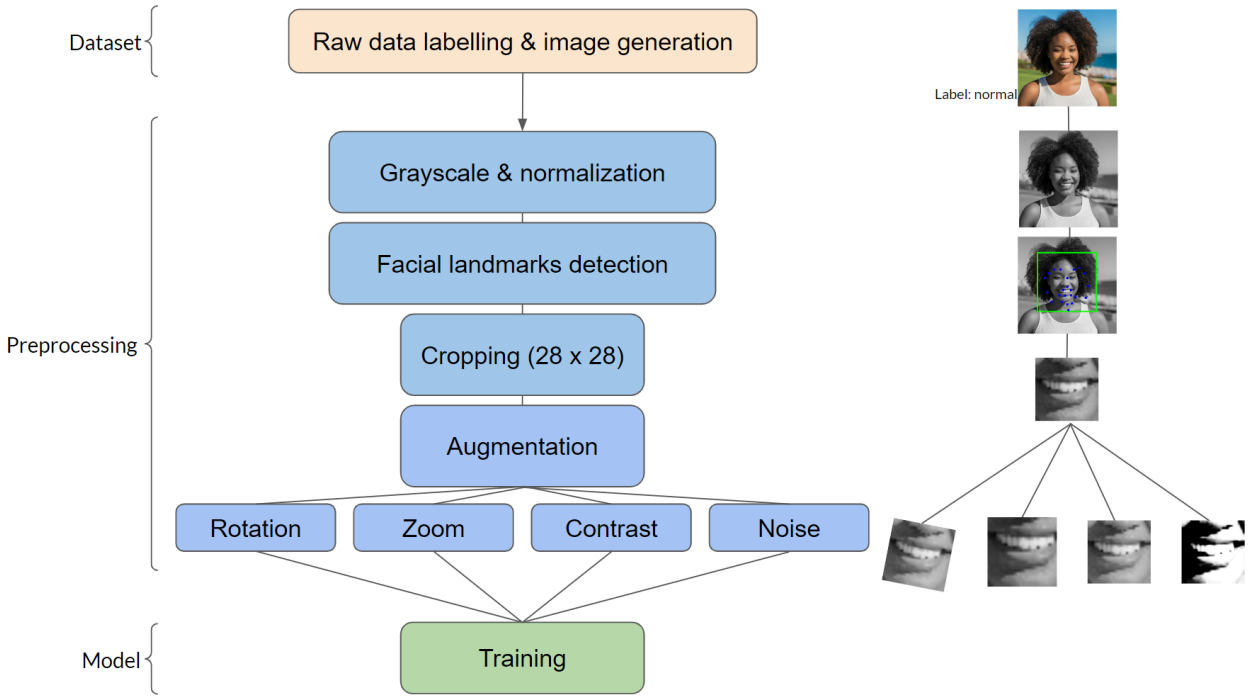


Figure 4: Diagram of the pipeline from data collection to preprocessing to CNN training. On the right is a visualization of the process a single image goes through. The image used is a random sample image from the Adobe Firefly AI-generated dataset with some of the included data augmentations.

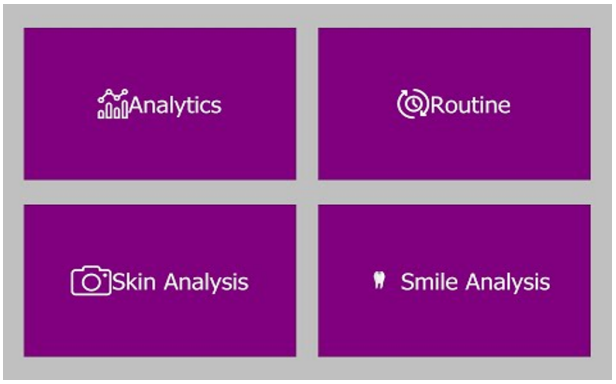


Figure 5: UI layout for an assistive diagnostic application. This study focuses on the Smile Analysis feature.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}} \tag{2}$$

Figure 6 illustrates the fluctuation in accuracy percentage throughout the progression of epochs for a representative model, showing an upward logarithmic trajectory of training accuracy. The validation accuracy starts around 0.50 and closely mirrors the training accuracy trend, though slightly lower. The training and validation loss of the same session follows the opposite trend, as depicted in Figure 6. Training beyond 150 epochs proved ineffective, as validation accuracy and loss diverged from training resulting in a plateauing training accuracy and decreasing validation accuracy.

The GenAI-empowered CNN model was compared to the two byproduct models, one trained on 200 real images and the other on 200 AI-generated images. For consistency, each model's testing accuracy was averaged over 10 trials. The box plots of each model are shown in Figure 7. As indicated by the error bars on the graph and the p value greater than 0.1, there was no significant difference between the performance of the model trained on real human images and the model trained solely on generative images. However, the median accuracy of the model trained on both real and generated data was significantly higher than that of the models trained on either real or GenAI images alone (p value < 0.01) and higher than the baseline of 50% (p value $<< 0.01$).

Additionally, the confusion matrix of a representative model with an accuracy of 82.14% and an F1 score of 0.82 highlights the importance of a balanced dataset in achieving roughly even results in both 'gummy' and 'normal' classes. Figure 8 shows the squares on the right diagonal represent mismatches between true labels and predicted labels, indicating false positives and false negatives. These mismatches are generally outnumbered by the true positives in the first row and column and true negatives in

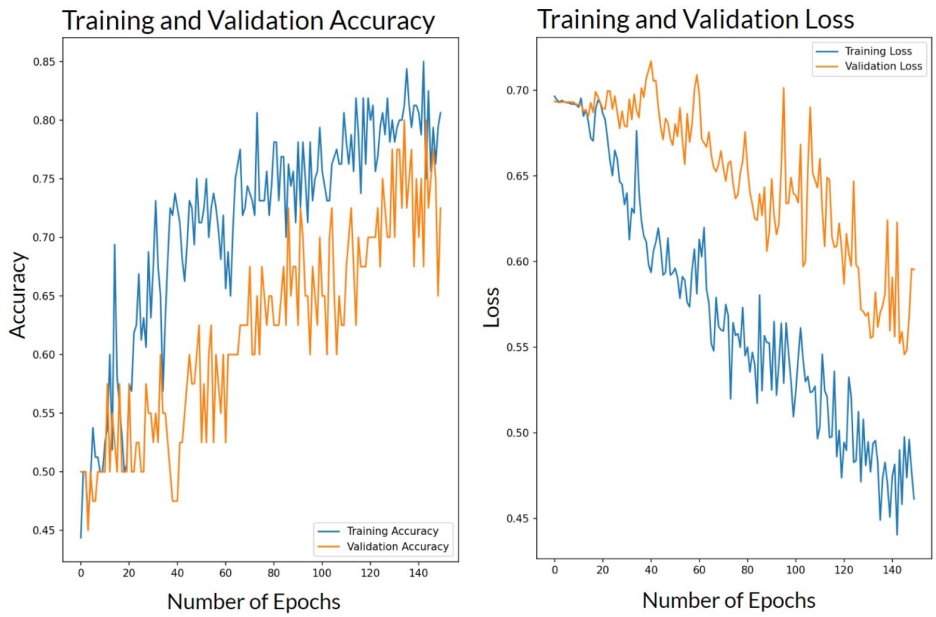


Figure 6: A representative sample of training and validation accuracy and loss graphs of the CNN model throughout training.

the second row and column, representing a high numbers of correctly labeled testing data points.

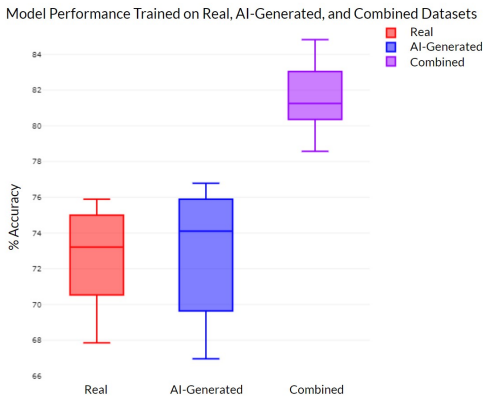


Figure 7: Box plots of the testing accuracy of models trained on real, AI-generated, and combined datasets.

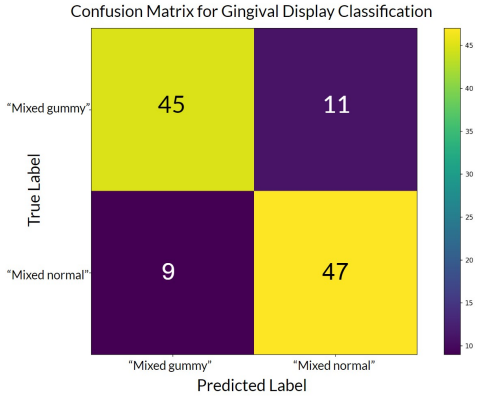


Figure 8: The confusion matrix of the CNN model results.

4. Discussion

The research represents a significant step forward in applying DL to dental smile analysis. It underscores the potential for utilizing CV and ML techniques in real-time healthcare applications. The mean testing accuracy of 81.607% indicates strong model performance, considering the baseline value is 50%. While the small dataset presents limitations, it also offers a notable advantage: computational efficiency. Training models on a small dataset and a large batch size is computationally inexpensive, allowing for the averaging of results from multiple trials—10 in this case—for each model.

We opted to evaluate each model using a combined dataset comprising both real and AI-generated images, rather than relying solely on real images. This decision stems from the recognition that both types of images may harbor biases, albeit potentially in divergent directions. The manually curated 'real' dataset exhibits significant variation in image quality but may lack proportional ethnic representation, potentially resulting in the model's underperformance for minority groups. Conversely, the diffusion model was trained on a diverse dataset, thus yielding diverse outcomes, but it may lack certain nuanced human characteristics, the absence of which could disrupt the model's accuracy. Lacking a precise quantitative method

2024, Vol. 1, Article ID. 2004.0002
https://doi.org/10.69709/CAIC.2024.194138

to gauge these biases, it would be imprudent to assume that either dataset thoroughly captures reality. While an ideal testing dataset would mirror reality in terms of camera quality, lighting, and angle, our current approach involves making estimations based on the combined dataset, relying on data augmentation to compensate for potential inconsistencies.

The integration of pre-trained CV and Dlib models for user image capture and analysis proved effective in guiding the user towards capturing a smile. However, capturing the ideal dental smile poses a significant challenge, and any form of smile analysis, whether digital or traditional, will inherently lack precision without the acquisition of a proper dental smile. Consequently, we must rely on a combination of on-screen instructions and guidance from the CNN to ensure accurate identification and analysis of smiles for our purposes.

Additionally, it is noteworthy that the model demonstrated nearly instantaneous classification of a smile when executed locally with a checkpoint. The only bottleneck identified pertained to the duration required to initialize the camera for the initial smile capture. Mitigating this bottleneck could be achieved through the utilization of more suitable tools for interfacing with device cameras; however, such optimizations fall beyond the scope of this study. It is important to acknowledge that this efficient performance may not necessarily translate seamlessly when deploying over a network. Therefore, the selection of an appropriate model for network-based operation necessitates careful consideration of varying performance characteristics, particularly with regard to latency. Given the diverse array of available models, it is imperative to prioritize the identification of a model that achieves optimal balance between speed and quality.

5. Conclusion and Future Work

The purpose of this study was to prove a concept and serve as a pilot study in detecting smile aesthetic issues with NN. We accomplished this goal by demonstrating an accuracy of 81.607% in detecting excessive gingival display using a GenAI-empowered CNN model. As the potential of DL and GenAI continue to expand, their applications grow exponentially. However, several areas warrant further investigation and a cautious approach is necessary. We reiterate the importance of testing these models on images representative of patients to be analyzed in a functional application. The optimal proportion of AI-generated training images for a realistic testing set may not be a 1:1 ratio. Exploring the impact of varying proportions of real and AI-generated training images on CNN model performance could yield valuable insights. To eliminate biases, employing multiple GenAI frameworks to augment the dataset would be beneficial. Furthermore, having reliable testing data would provide an opportunity to experiment with other model architectures, such as Tensorflow's Functional API model or a version of YOLO.

Alternatively, an outside-the-box approach could be developed: rather than investing resources in adapting a model to a dataset, the dataset could be adapted to the model. In other words, a diffusion model could be built to generate a dataset in the style of realistic images and then train an NN purely on generated images. This innovative method bypasses the need for perfecting the model architecture through meticulous hyperparameter tuning. Instead, it prioritizes improving the start of the pipeline, fortifying the model against overfitting. Success in this endeavor could mark a significant leap forward, not only in digital smile analysis but also in broader medical fields. It would revolutionize the diagnosis of rare or infrequently photographed diseases, launching a new era of precision and effectiveness in medical diagnostics.

Regardless of the approach taken, the ultimate goal within the scope of digital smile analysis remains steadfast: to develop an assistive diagnostic tool capable of tracking changes over time, enabling healthcare professionals to deliver more accurate and timely treatment. This concept holds promise for further development into a fully functional application compatible with multiple operating systems or integration within a novel ecosystem of smart mirrors interconnected through the IoM. The application can be extended to assess other smile analysis components, such as the smile arc, to provide patient with a more in-depth overview from the same image. Such advancements could facilitate the interaction between patients and healthcare professionals providing reliable, accurate, and automated smile analysis to be utilized by clinicians. Embracing smarter medical guidance and living is a catalyst for fostering deeper engagement and collaboration between patients and professionals in the digital landscape. The realization of this vision is rapidly approaching, as the expanding capabilities of DL, generative AI, and networks continue to unfold exponential applications. The goal is to refine this technology to be a standard procedure in the process of digital smile analysis.

List of Abbreviations

AI	Artificial Intelligence
CNN	Convolutional Neural Network
DL	Deep Learning
GenAI	Generative Artificial Intelligence
IoM	Internet-of-Mirrors
IoT	Internet-of-Things

ML	Machine Learning
NN	Neural Networks
R-CNN	Region-based Convolutional Neural Network
ReLU	Rectified Linear Unit
UI	User Interface

Author Contributions

Conceptualization, M.B., H.F., R.O., A.T., M.A.I. and L.M.; Methodology, M.B., H.F., R.O., A.T., and L.M.; Software, M.B.; Validation, M.B. and H.F.; Formal Analysis, M.B.; Investigation, M.B., H.F., R.O., A.T., and L.M.; Resources, M.A.I., A.T., and L.M.; Data Curation, M.B. and R.O.; Writing – Original Draft Preparation, M.B.; Writing – Review & Editing, R.O., N.G., A.T., M.A.I. and L.M.; Visualization, M.B. and H.F.; Supervision, R.O., A.T., M.A.I., and L.M.; Project Administration, R.O., A.T., M.A.I., and L.M.; Funding Acquisition, M.A.I., A.T., and L.M. All authors have reviewed and approved the published version of the article.

Availability of Data and Materials

The data that support the findings of this study are available on request from the corresponding author, M. Baidachna.

Consent for Publication

Not applicable, as no personal data or images were used. The images presented in Figures 2 and 4 are randomly selected samples generated using the Adobe Firefly AI dataset, including some of the applied data augmentation techniques.

Conflicts of Interest

The authors have no conflicts of interest to declare.

Funding

This work is supported in parts by EPSRC grant EP/T517896/1 for Summer 2023, University of Glasgow Chancellor's fund, and the Autonomous Systems and Connectivity research division ECDP fund.

Acknowledgments

The authors would like to thank the EPSRC, the University of Glasgow Chancellor's fund, and the Autonomous Systems and Connectivity research division ECDP fund for their support. The authors also declare that no AI tools were used to generate the text of this manuscript. However, the images presented in Figures 2 and 4 are random sample images from the Adobe Firefly AI-generated dataset.

References

[1] M. Senbekov et al., "The recent progress and applications of digital technologies in healthcare: A review," *Int. J. Telemed. Appl.*, vol. 2020, 2020, Art. no. 8830200. [CrossRef]

[2] M. Al-Quraan, A. Khan, A. Centeno, A. Zoha, M. A. Imran, and L. Mohjazi, "FedTrees: A novel computation-communication efficient federated learning framework investigated in smart grids," *arXiv*, 2022. [CrossRef]

[3] J. Armalaite et al., "Smile aesthetics as perceived by dental students: A cross-sectional study," *BMC Oral Health*, vol. 18, no. 1, 2018, Art. no. 225. [CrossRef]

[4] A. Lukez, A. Pavlic, M. Trinajstic Zrinski, and S. Spalj, "The unique contribution of elements of smile aesthetics to psychosocial well-being," *J. Oral Rehabil.*, vol. 42, no. 4, pp. 275–281, 2015. [CrossRef] [PubMed]

[5] M. Luca, A. Luca, C. M. A. V. Grasso, and C. Calandra, "Nothing to smile about," *Neuropsychiatr. Dis. Treat.*, vol. 10, pp. 1999–2008, 2014. [CrossRef] [PubMed]

[6] J. May, P. V. Bussen, and D. M. Steinbacher, "Smile aesthetics," in *Aesthetic Orthognathic Surgery and Rhinoplasty*, D. M. Steinbacher, Ed., New York, NY, USA: Wiley, 2019, pp. 253–287. [Online]. Available: https://www.wiley.com/en-ae/Aesthetic+Orthognathic+Surgery+and+Rhinoplasty-p-9781119186977

[7] R. Saini, N. Thakur, R. Jindal Goyal, K. Sharma Rai, H. Bagde, and A. Dhopte, "Analysis of smile aesthetic changes with fixed orthodontic treatment," *Cureus*, vol. 14, no. 12, 2022, Art. no. e32612. [CrossRef]

[8] S. H. Sajjadi, B. Khosravanifard, F. Moazzami, V. Rakhshan, and M. Esmaeilpour, "Effects of three types of digital camera sensors on dental specialists' perception of smile esthetics: a preliminary double-blind clinical trial," *J. Prosthodont.*, vol. 25, no. 8, pp. 675–681, 2016. [CrossRef] [PubMed]

[9] M. Khan, S. M. R. Kazmi, F. R. Khan, and I. Samejo, "Analysis of different characteristics of smile," *BDJ Open*, vol. 6, no. 1, 2020, Art. no. 6. [CrossRef]

[10] S. Desai, M. Upadhyay, and R. Nanda, "Dynamic smile analysis: changes with age," *Am. J. Orthod. Dentofac. Orthop.*, vol. 136, no. 3, pp. 310.e1–310.e10, 2009. [CrossRef]

[11] S. Akyalcin, L. K. Frels, J. D. English, and S. Laman, "Analysis of smile esthetics in American Board of Orthodontic patients," *Angle Orthod.*, vol. 84, no. 3, pp. 486–491, 2014. [CrossRef]

[12] L. K. Ríos, "Laypeople's perceptions of smile esthetics: Why is it important and what do we need

to know?" *J. Oral Res.*, vol. 1, pp. 27–29, 2020. [CrossRef]

[13] S. H. Agou, "Comparison of digital and paper assessment of smile aesthetics perception," *J. Int. Soc. Prev. Community Dent.*, vol. 10, no. 5, 2020, Art. no. 659. [CrossRef]

[14] S. H. Sajjadi, B. Khosravanifard, M. Esmaeilpour, V. Rakhshan, and F. Moazzami, "The effects of camera lenses and dental specialties on the perception of smile esthetics," *J. Orthod. Sci.*, vol. 4, no. 4, 2015, Art. no. 97. [CrossRef]

[15] S. V. Prasanna and R. Vignesh, "Evaluation of waiting period, recall period, and appointment scheduling of outpatients in a dental hospital," *Drug Invent. Today*, vol. 11, no. 7, pp. 1580–1583, 2019. [View Online]

[16] M. R. Inglehart, A. H. Lee, K. G. Koltuniak, T. A. Morton, and J. M. Wheaton, "Do waiting times in dental offices affect patient satisfaction and evaluations of patient-provider relationships? A quasi-experimental study," *J. Dent. Hyg.*, vol. 90, no. 3, pp. 203–211, 2016. [PubMed]

[17] A. Aggarwal et al. "The future of cancer care in the UK-time for a radical and sustainable national cancer plan," *Lancet Oncol.*, vol. 25, no. 1, pp. e6–e17, 2024. [CrossRef] [PubMed]

[18] E. Downey, H. S. Fokeladeh, and H. Catton, *What the COVID-19 Pandemic Has Exposed: The Findings of Five Global Health Workforce Professions*. Geneva, Switzerland: World Health Organization, 2023. [Online]. Available: https://www.who.int/publications/i/item/9789240070189

[19] C. B. Susilo, I. Jayanto, and I. Kusumawaty, "Understanding digital technology trends in healthcare and preventive strategy," *Int. J. Health Med. Sci.*, vol. 4, no. 3, pp. 347–354, 2021. [CrossRef]

[20] Shilpa and T. Kaur, "Digital healthcare: Current trends, challenges and future perspectives," in *Proceedings of the Future Technologies Conference (FTC) 2021*, Berlin, Germany: Springer, 2022, vol. 2, pp. 645–661. [CrossRef]

[21] A. Dhopte and H. Bagde, "Smart smile: Revolutionizing dentistry with artificial intelligence," *Cureus*, vol. 15, no. 6, 2023, Art. no. e41227. [CrossRef]

[22] S. P. Mohanty, U. Choppali, and E. Kougianos, "Everything you wanted to know about smart cities: The Internet of things is the backbone," *IEEE Consum. Electron. Mag.*, vol. 5, no. 3, pp. 60–70, 2016. [CrossRef]

[23] Z. N. Aghdam, A. M. Rahmani, and M. Hosseinzadeh, "The role of the Internet of Things in healthcare: Future trends and challenges," *Comput. Methods Programs Biomed.*, vol. 199, 2021, Art. no. 105903. [CrossRef]

[24] W. Li et al. "A comprehensive survey on machine learning-based big data analytics for IoT-enabled smart healthcare system," *Mob. Netw. Appl.*, vol. 26, pp. 234–252, 2021. [CrossRef]

[25] H. K. Bharadwaj et al. "A review on the role of machine learning in enabling IoT based healthcare applications," *IEEE Access*, vol. 9, pp. 38859–38890, 2021. [CrossRef]

[26] J. Walsh et al. "Deep learning vs. traditional computer vision," in *Proc. Int. Conf. Comput. Vis. Workshops (ICCVW)*, Seoul, South Korea, 2019. [CrossRef]

[27] G. Zhu, Z. Piao, and S. C. Kim, "Tooth detection and segmentation with mask R-CNN," in *Proc. 2020 Int. Conf. Artif. Intell. Inf. Commun. (ICAIIC)*, Fukuoka, Japan, 2020, pp. 070–072. [CrossRef]

[28] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *arXiv*, 2018. [CrossRef]

[29] J. C. V. Soares, M. Gattass, and M. A. Meggiolaro, "Visual SLAM in human populated environments: Exploring the trade-off between accuracy and speed of YOLO and mask R-CNN," in *Proc. 2019 19th Int. Conf. Adv. Robot. (ICAR)*, Belo Horizonte, Brazil, 2019, pp. 135–140. [CrossRef]

[30] S. Lee and J. Kim. "Evaluating the precision of automatic segmentation of teeth, gingiva and facial landmarks for 2D digital smile design using real-time instance segmentation network," *J. Clin. Med.*, vol. 11, no. 3, 2022, Art. no. 852. [CrossRef]

[31] H. Chen et al., "A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films," *Sci. Rep.*, vol. 9, no. 1, 2019, Art. no. 3840. [CrossRef]

[32] S. Motamed, P. Rogalla, and F. Khalvati, "Data augmentation using Generative Adversarial Networks (GANs) for GAN-based detection of Pneumonia and COVID-19 in chest X-ray images," *Inform. Med. Unlocked*, vol. 27, 2021, Art. no. 100779. [CrossRef]

[33] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, pp. 1–48, 2019. [CrossRef]

[34] A. V. Nesen, "Deep neural networks for detection of rare events, novelties, and data augmentation in multimodal data streams," Ph.D. thesis, Purdue Univ., West Lafayette, IN, USA, 2022. [Online]. Available: https://docs.lib.purdue.edu/dissertations/AAI30506112/

[35] S. Frolov, T. Hinz, F. Raue, J. Hees, and A. Dengel, "Adversarial text-to-image synthesis: A review," *Neural Netw.*, vol. 144, pp. 187–209, 2021. [CrossRef] [PubMed]

[36] M. Alauthman et al., "Enhancing small medical dataset classification performance using GAN," *Informatics*, vol. 10, no. 1, 2023, Art. no. 28. [CrossRef]

[37] X. Zhu, Y. Liu, Z. Qin, and J. Li, "Data augmentation in emotion classification using generative adversarial networks," *arXiv*, 2017. [CrossRef]

[38] S. Cheng, Y. Guo, and R. Arcucci, "A generative model for surrogates of spatial-temporal wildfire nowcasting," *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 7, no. 5, pp. 1420–1430, 2023. [CrossRef]

[39] "Adobe Firefly," [Online]. Available: https://firefly.adobe.com/

[40] H. Fatima, M. A. Imran, A. Taha, and L. Mohjazi. "Internet of mirrors for connected healthcare and beauty: A prospective vision," *arXiv*, 2023. [CrossRef]

[41] S. Ghai, "Teledentistry during COVID-19 pandemic," *Diabetes Metab. Syndr. Clin. Res. Rev.*, vol. 14, no. 5, pp. 933–935, 2020. [CrossRef]

Disclaimer/Publisher's Note: The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the publisher or editors. The publisher and editors assume no responsibility for any injury or damage resulting from the use of information contained herein.