A COMPARATIVE STUDY OF FACIAL FEATURE EXTRACTION USING MTCNN, RETINAFACE AND DLIB FACE DETECTOR FOR PERSONALITY TRAITS RECOGNITION

Nurrul Akma Mahamad Amin[1*], Nilam Nur Amir Sjarif[1], Siti Sophiayati Yuhaniz[1]

[1]Department of Intelligence Informatics, Faculty of Artificial Intelligence,
54100 Universiti Teknologi Malaysia, Kuala Lumpur

Emails: nurrulakma@graduate.utm.my[1*](Corresponding Author), nilamnur@utm.my[1], sophia@utm.my[1]

ABSTRACT

Facial feature extraction is a fundamental step in various computer vision tasks, including face recognition, emotion detection, and personality traits recognition. The efficiency of these tasks depends on choosing the right face detector model to extract facial features. As for personality traits recognition tasks, face detection is important in understanding the facial expressions that underlying personality traits. There are several face detector models like Multi-Task Cascaded Convolutional Neural Network (MTCNN), RetinaFace, and DLIB that can detect and extract facial features. However, the challenge arises in selecting the most effective face detector model, particularly when dealing with diverse facial expressions, orientations, and occlusion. There is a lack of comprehensive comparisons that have been made between MTCNN, RetinaFace, and DLIB for face detection ability, particularly in video-based personality traits recognition. Thus, this study presents a comparative analysis of MTCNN, RetinaFace, and DLIB models, focusing on their ability to detect human faces from key frames that are extracted from videos. This study used the ChaLearn dataset, which consists of 15-second videos of people speaking in front of a camera. MTCNN and RetinaFace were able to detect higher numbers of faces consistently, even in cases where the faces were not strictly frontal. In contrast, DLIB has problems detecting non-frontal faces and resulting in fewer face detections. We demonstrate that MTCNN and RetinaFace are more suitable for tasks that require robust face detection, especially across datasets that consist of a variety of facial poses. Additionally, using MTCNN and RetinaFace as face detector models gives prominent accuracy performance for video-based personality recognition.

Keywords: Computer Vision; Personality Traits Recognition; Face Detector Model; Facial Features; Facial Landmarks.

1.0 INTRODUCTION

Personality Traits Recognition (PTR) is a computer vision task designed to automatically detect individual personality traits based on their behavioral signals. Behavioral signals such as facial features, facial expressions, gestures, or body movements can be easily collected from user-generated data, including social media posts, comments, online reviews, blogs or forum posts, wearable devices, and more [1]. With advancements in computer vision technology, personality traits recognition has the potential to automate personality judgments, which can enhance social interactions, help business marketing, improve user profiling, enable product personalization [2], and support telemedicine services [3], [4]. Personality traits recognition can be developed using artificial intelligence, machine learning, and deep learning techniques to analyze various data modalities including text, audio, and video data to automate personality judgments and predict an individual's personality. These judgments commonly use personality models from the field of psychology as the foundation for final classification. A personality model such as the Big Five personality traits is widely employed, where it provides the criteria or benchmarks that the system uses to make the final decision about a person's personality traits. The extracted features are processed by a machine learning or deep learning algorithm to classify personality traits in line with these models. For instance, someone's high energy in speech and frequent smiling may correlate with extraversion as defined by the Big Five model. By grounding the classification with well-recognized frameworks in psychology, the judgments become scientifically informed and more consistent.

As for video-based personality traits recognition, the models' accuracy highly depends on effective face detection and facial features extraction. Extracting meaningful features assists models in learning and understanding the relationship between facial features and personality traits. Facial features and facial landmarks are key components in face detection and recognition tasks. They also play an important role in achieving accurate analysis in face detection and personality traits recognition. Several popular face detector models like MultiTask

Cascaded Convolutional Neural Network (MTCNN), RetinaFace, and DLIB are widely used to detect human faces and extract meaningful features. Even though there are several well-known face detection models available, choosing the best one is still difficult, especially when dealing with situations like occlusion, poor lighting, or non-frontal images. Face detection in MTCNN, RetinaFace, and DLIB follows distinct approaches. MTCNN detects faces in three steps using small neural networks called P-Net, R-Net, and O-Net. It gradually refines the face location and landmarks including eyes, nose, and mouth, at each stage to accurately detect and align faces, even with different sizes or angles. In contrast, RetinaFace leverages a single-shot CNN with Feature Pyramid Networks (FPN) to detect multi-scale faces and predict not only bounding boxes but also five facial landmarks. RetinaFace works well for faces in difficult conditions like non-frontal view or in poor lighting conditions. Meanwhile, DLIB offers two options for face detection which are a traditional Histogram of Oriented Gradients (HOG) method that extracts gradient-based features for frontal face detection and a more accurate deep learning method using a CNN with 68-point landmarks. Each of these models has its own algorithm, which provides various strengths and weaknesses in the detection and extraction operation. MTCNN is known for its speed and ability to handle multiple tasks like face alignment and key point localization [5]. On the other hand, RetinaFace excels in its accuracy, especially for detecting non-frontal faces. Whereas DLIB is well known for its efficiency in detecting frontal faces, which relies on a histogram of oriented gradients and linear classifiers. However, there is a lack of comprehensive comparisons that evaluate the robustness of MTCNN, RetinaFace, and DLIB for face detection in key frame images, particularly within the context of video-based personality traits recognition. The main challenge in developing an automatic personality recognition model is extracting and selecting relevant features from video data to provide a better classification score [6], [7], [8]. Due to the complex nature of video data, the number of frames or images may vary depending on the video's frame rate, or frames per second (FPS). According to Gharahbagh et al., [9], handling video processing for the recognition process is computationally expensive depending on the duration of the video. Thus, in this study, we implemented key frame extraction and selection methods to select the most significant frames for personality trait recognition.

Key frame extraction and selection is a novel process for identifying and extracting the best frames from a video input that significantly differ from each other [10]. Key frames are the best frames in a video that capture important visual features and represent significant content. The key frame usually represents the most relevant features of each video shot. Key frame selection typically involves an initial step of extracting candidate frames based on some criteria and then selecting key frames from these candidates. Clustering techniques are popular for key frame extraction and selection in video processing, such as K-means clustering [11], density clustering [12], fuzzy C-mean clustering [13], adaptive clustering [14] and HDBSCAN clustering [15]. HDBSCAN shows its robustness in terms of parameter selection, where the minimum cluster size is the only required primary parameter, which can be set in an intuitive manner [16]. The key frame selection methods aim to reduce computational resources, including storage, memory, and runtime spaces when extracting frames for video processing, making the processing of video data more efficient and faster. Several criteria have been used as the basis of key frame selection such as pixel-wise absolute frame differences, scene changes, visual quality, colour histogram, histogram difference, correlation, entropy difference, and etc., using algorithms or computer vision libraries. Key frames also encode the highest information compared to other frames in video input sets. These key frames are considered the best frames that give a significant overview of the content in the video [17]. Thus, accurately extracting and selecting key frames can effectively reduce processing time, required runtime space, and memory usage [18].

The main objective of this study is to evaluate and compare the performance of MTCNN, RetinaFace, and DLIB for face detection on key frames images from video data of ChaLearn dataset. By focusing on their robustness and accuracy, this study aims to understand how each model handles challenging conditions typically encountered in real-world video-based applications, such as varied facial orientations and occlusions. In the following, the study also looked at how effective face detection contributed to the performance of personality traits recognition using CNN-based approaches. This comparative study will provide further insights on how well each model performs in terms of face detection accuracy in video-based personality traits recognition tasks. It will help researchers in choosing the right face detection model for similar research and applications. The remainder of this paper is organized as follows: In Section 2, we review relevant literature on facial detection models and their use in facial features extraction. Section 3 explains the methodology used in this study. In Section 4, we discuss in detail the experimental results, comparing the performance of MTCNN, RetinaFace, and DLIB in terms of both face detection and personality traits prediction accuracy. Finally, Section 5 provides a brief conclusion and outlines potential future research directions.

2.0 LITERATURE REVIEW

In the field of psychological study, personality measurements serve as powerful tools for understanding an individual's personality and predicting outcomes such as personal preferences, academic achievement, job satisfaction, and job performance [19]. Personality measurements allow for more systematic approaches to measuring and identifying individuals' personality traits based on personality models. The Big Five (Big-5) model provides a structured way to assess personality trait dimensions, whether applied in recruitment, educational

development, or personal growth. Although various personality models are available, such as the Big-5, Myers-Briggs Type Indicator (MBTI), the Sixteen Personality Factor Questionnaire (16PF), the Eysenck Personality Questionnaire-Revised (EPQ-R) and the Three Traits Personality Model (PEN), the Big-5 model is the most widely used in personality recognition. This is due to the widely accepted status and popularity of the Big-5 model in psychological literature, as it has been proven to be highly reliable in describing human personality [20]. The interpretation of human personality represented by each of these models differs from each other. The Big-5 model consists of five personality dimensions, including openness, conscientiousness, extraversion, agreeableness, and neuroticism. The Big-5 Model also is one of the dominant taxonomies of personality that has been proven to predict professional performance across decades of research [21]. These five factors are also often used as predictors in personality recognition during employment screening [22], [23], [24] . The implementation of employment screening with the adaptation of artificial intelligence has leveraged digital-based tools and gamification approaches in making personality recognition more engaging yet effective [25]. The primary intention of using digital-based tools in employment screening is to make recruiting more efficient, convenient, and cost-savvy in selecting suitable candidates who fit the positions [26], [27]. Personality recognition tests are commonly used in employment screening as tools for assessing personality traits. They can measure a candidate's capabilities and reveal their personality or underlying abilities. These tests are often used to identify suitable candidates by eliminating unqualified applicants [28]. In addition, individuals' interaction styles, personality traits, interpersonal communication skills, competencies, job performance, preferences, and behavioral tendencies can also be discovered through personality testing [29], [30], [31].

Personality traits are subjective and may be perceived differently depending on the situation, culture, and environment. Personality traits recognition is a modern solution that tries to solve this subjective task by using machine-generated content, such as images, videos, text, and audio with computational approaches [32]. This modern solution aims to classify human personality into personality traits classes based on personality model dimensions or characteristics. Personality trait recognition also has a wide range of applications including recruitment, education, mental health assistance, user experience profiling and many more. Initially, personality traits recognition relied on conventional techniques like questionnaires and self-assessments, where individuals described their own characteristics, often using well-known models like the Big-5 inventory. However, due to the advancement of computer vision and machine learning technologies, there has been a transition from self-reporting tools to computational approaches that utilize machine generated data. This transition offers a more objective and scalable approach to personality recognition, reducing reliance on subjective self-reports and avoiding distortions in assessments. Computational approaches also enable the integration of multi-visual data, combining inputs like face appearance and the geometry of facial landmark features. This diversity of input enhances the ability of personality assessment to capture dynamic behaviors that static questionnaires cannot address. Furthermore, using questionnaires with closed-ended questions in personality tests to predict personality traits is inadequate and not comprehensive. Compared to traditional methods of personality assessment, computational approaches using image-based data are more natural, genuine, truthful, and language-insensitive [33]. Thus, automatic personality traits recognition has become the current solution to automate personality testing and mitigate issues in traditional approaches. This also marks a significant turning point in the integration of traditional psychology and modern technology, paving the way for more comprehensive personality assessments.

Detecting faces and their key points, such as lips, nose, eyes, and mouth, was previously a difficult task. However, deep learning algorithms have recently demonstrated their ability to address this challenge. According to Kachur et al., deep learning algorithms successfully reveal multidimensional personality profiles using facial features, which involve the shape and structure of the front of the head, from the chin to the top of the forehead [34]. Similarly, a study conducted by J. Li et al., found that that personality traits can be reliably predicted from faces and their key points using deep learning-based algorithms [35]. The baseline model for personality trait recognition developed by Kaya et al., also used a deep learning-based algorithm to extract facial features and achieved 91% accuracy in its final predictions [36]. Another study conducted by Cai and Liu discovered relationships between facial features and the Big Five personality model traits, finding that points from the right jawline to the chin contour showed a significant negative correlation with agreeableness [37]. Furthermore, several studies in personality traits recognition have utilized facial features from video data to automatically identify attributes of the Big-5 personality model [8], [38], [39]. Facial features are relevant for personality recognition because they provide valuable information about human expressions and behaviors. For example, individuals with higher scores in conscientiousness exhibit greater fluctuations in pupil size, while those who blink more frequently tend to be more neurotic [40]. The degree of mouth opening and the percentage of eyelid closure over the pupil over time are two metrics used to identify fatigue among drivers [41]. Thus, for successful personality traits recognition, an accurate and robust face detector model is essential, which will lead to an effective facial feature extraction process. Facial feature extraction is a key step in personality traits recognition tasks, which involve detecting faces and analyzing facial features on a face. Existing studies on personality traits recognition have used facial features extracted from random frames, such as selecting 30 random frames from the entire ChaLearn video [42], [43]. Another study by [8] extracted frames uniformly, taking 15 frames from the 15-second video, equivalent

to one frame per second. These features can be used to identify unique characteristics of an individual or to understand their emotions and facial expressions that underlie personality traits.

Numerous face detection models have been developed over the years to help in computer vision tasks, especially to detect faces and facial features in both image and video input. The most popular and widely used models are Multi-Task Cascaded Convolutional Networks (MTCNN), RetinaFace, and DLIB. Each of these models adopts unique methods and algorithms for detecting faces, extracting facial features, and calculating landmark points, which makes them suitable for different types of computer vision tasks. MTCNN was introduced by K. Zhang, et al., in 2016 to detect faces and five key points on the face [44]. MTCNN uses a cascade of three convolution networks that gradually improve detection results and ensure accurate recognition even in challenging conditions such as varying face orientations and occlusions. In the first stage, a fully convolutional network generates candidate windows and corresponding bounding box regression vectors. Next, the second stage processes these candidates by eliminating many false positives and improving the bounding box predictions. Finally, in the third stage, the model performs accurate facial landmark detection, identifying the five main facial points. This cascade structure, combining face detection with landmark alignment, allows MTCNN to deliver robust performance in a variety of scenarios. A previous study successfully proposed a classroom face detection method based on the improved MTCNN to detect faces under classroom scenarios with different angles of view, uneven distributions of face scales, and occlusion [45]. In general, the occurrence of occlusions significantly affects face detection and may reduce the overall accuracy of the model [46]. Another study developed a real-time vision system that performs face detection and transmits the detected face coordinates to a facial emotion classification model for further analysis [47]. A portable embedded device with face recognition capabilities using MTCNN was developed to facilitate visually impaired persons to recognize faces [48].

On the other hand, RetinaFace is another recent face detector model based on the RetinaNet object detection framework. RetinaFace uses a deep convolutional neural network to detect faces and important facial features. It works well in difficult situations like occlusion, poor lighting, or non-frontal images. It also leverages deep residual networks and applies a feature pyramid network with independent context modules to extract features at multiple scales. RetinaFace uses ResNet50 as its backbone, supplying feature vectors from multiple layers of ResNet50 to the detection stages [49]. This feature makes it effective for detecting faces in crowded scenes or images with various face sizes. Face mask detection was a popular research topic during COVID-19, aimed at developing automatic mask-wearing detection systems based on monitored images. Face mask detection using the RetinaFace algorithm has demonstrated better performance in quickly detecting people who are not wearing masks in crowded places [50]. The RetinaFace model was also used to study infant faces and address the closely related issue of estimating infant body posture. The authors, Wan et al., presented a collection of baby faces annotated with pose attributes and facial landmark points [51]. Rui Zhong proposed a method for multi-view face detection and expression recognition using RetinaFace, and the experimental results showed that the RetinaFace algorithm is highly robust, demonstrating impressive detection accuracy and processing time [52].

DLIB is another well-known face detection model that is popular due to its ease of use and speed in landmark feature extraction. A variety of machine learning techniques for face detection, facial landmark extraction, object detection, and other applications are available in this DLIB open-source library. DLIB was an older approach based more on traditional machine learning techniques [53]. For example, the DLIB frontal face detector is a specific component within the DLIB library, designed for detecting faces in images that uses a Histogram of Oriented Gradients (HOG) feature combined with a linear classifier. The DLIB face detector and the DLIB facial landmark are combined to design drowsiness detection applications using real-time video input captured through a webcam [54]. DLIB has proven effective at identifying frontal faces in images, but it struggles with occlusions and non-frontal faces. DLIB is also a popular choice for real-time applications with constrained computational resources since it is lightweight and efficient, even though it may not always match the performance of more complex models like MTCNN and RetinaFace. However, the limitations of DLIB became clearer as the demand for more advanced models increased. The use of hand-crafted features and traditional machine learning methods makes it less effective in dealing with complex facial poses, non-frontal faces, and other real-world challenges like personality traits recognition. Table 1 summarizes the context in which face detector models have been used and implemented in various applications in previous studies.

Hence, the advancement of deep learning models like MTCNN and RetinaFace has helped to overcome limitations in DLIB's frontal face detection. These deep learning models excel in extracting facial features more accurately, even under challenging conditions like varying facial orientations, occlusions, and poor lighting. The evolution from traditional machine learning to deep learning has significantly enhanced the ability of the face detection model to extract more meaningful facial features mainly for personality traits recognition which requires a complex interpretation of facial expressions. Several comparative studies have been done that focused on the performance of machine learning and deep learning algorithms in face detection systems [55], [56]. The performance of these algorithms is influenced by several factors, including the size and diversity of the dataset. A larger dataset typically provides more diverse samples, allowing the model to learn from a wide range of facial features and expressions. Some other challenges such as variability in angle, orientations, illumination, occlusion, and background also affect the performance of these systems [57].

Table 1: A summary context of used among face detector models

Author(s)	Context of Used / Application	Face Detector	Strength	Limitation/ Future Work Suggested
Baskar et al., (2023) [48]	Faces recognition using compact wearable device for visually impaired people	MTCNN	Experimental results show the MTCNN based LPB uses optimal CPU utilization and improve the accuracy of real-time face recognition	The improvements to the proposed system aim to enable it to function in various scenarios, such as capturing real-time data from people walking with wearable devices, and optimizing frames per second to enhance speed
Kumar et al., (2023) [60]	Face detection and recognition system for criminal identification	MTCNN	Different facial features are extracted using MTCNN classifiers. Grayscale images from this step are used to identify criminals and train the model.	Model execution can be improved by considering different qualities other than face images like the age and sex of an individual.
Huang et al., (2023) [50]	To detect the masked face (people wearing masks) by utilizing RetinaFace in crowded places	RetinaFace	Uses Res2Net as the backbone network, and enhances feature extraction by introducing a weighted bidirectional feature pyramid and CBAM (Convolutional Block Attention Module)	Future studies will further optimize the network topology and aim to apply it to real-world scenarios, provided that the accuracy of mask-wearing detection is ensured.
Phienphanich et al., (2023) [61]	Use of facial image dataset containing neutral and smiling expressions to classify facial weakness which is a common sign of stroke	RetinaFace	RetinaFace employs a multi-task learning deep convolutional neural network to detect and locate five key facial landmarks, including the eyes, nose, and mouth. It is capable of detecting faces even under challenging conditions, such as varying lighting, poses, and facial expressions.	Collecting more data in future work to increase the accuracy of facial weakness screening and incorporate progressive FGANs to enhance existing models so that it can be used on different face angles and 3D face models.
Gu et al., (2022) [45]	Classroom face detection under various angles, small scales images and occlusions	MTCNN	A deep residual feature generation module is introduced to improve the detection accuracy of small-scale faces. Experimental results demonstrate that the proposed method achieved superior accuracy results over some state-of-the-art approaches	MTCNNmodel has weak generalization ability, poor robustness, and poor performance for small scale face detection
Wan et al., (2022) [51]	Facial detection for infants especially in the early prediction of infants' developmental disorder	RetinaFace	Introduce the dataset of infant faces annotated with facial landmark coordinates and pose attributes. Performed tests on infant faces using RetinaFace model and tackles the closely related problem of infant body pose estimation.	Future work and further research are needed in infant face segmentation to improve the localization of infants faces and facial landmarks.

Table 1: *Continued*

Author(s)	Context of Used / Application	Face Detector	Strength	Limitation/ Future Work Suggested
Noor Reza et al., (2021) [54]	Drowsiness detection applications are designed based on face landmark recognition	DLIB	Several facial detection methods such as computer vision, dlib face detector, dlib facial landmark, and eye aspect ratio (EAR) are combined to design drowsiness detection applications.	The CPU and power consumption when the application is running is large enough to cause the laptop to heat up quickly, and the battery soon runs out.
Zhou et al., (2021) [47]	Face detection for facial emotions classification	MTCNN	Successfully eliminated the interference factors of the multiple faces in the image	Lot of noise found in the facial expressions captured in real life like blurred image, blocked face etc.
Ullah et al., (2021) [53]	Face detection for facial emotions classification	DLIB	Successfully detected frontal face on dataset and 68 landmark is used to predict facial features	Feature selection can aid in detecting facial expressions across cultures, but further research is needed to develop a generalized model.
Deng et al., (2020) [49]	Face detection in diverse datasets with varying lighting and facial orientations	RetinaFace	Unifies face box prediction, 2D facial landmark localisation and 3D vertices regression. Experimental results show that RetinaFace can simultaneously achieve consistent face detection, accurate 2D face alignment, and robust 3D face reconstruction	Future work to improve the robustness of proposed face detection in other datasets and various conditions
Zhao et al., (2020) [41]	Driver fatigue status detection	MTCNN	Efficiently detect driver fatigue status using driving images. The percentage of eyelid closure over the pupil over time and mouth opening degree are two parameters used for fatigue detection.	To further test the actual performance and robustness of the proposed method
Gyawali et al., (2020) [62]	Age range estimation based on face images	MTCNN	MTCNN helped to extract only the facial features from the image data which helps to determine the most relevant features from the face. The age range estimate performance was greatly enhanced by using MTCNN and fine-tuning the VGG-Face model.	There are a limited number of dataset available for age estimation, which could be enhanced in future efforts.

Hybrid models which combine the strengths of traditional machine learning approaches like DLIB and advanced deep learning techniques like MTCNN and RetinaFace may offer a better solution. By leveraging the best features of both approaches, hybrid models have demonstrated improved performance in facial expression recognition, emotion prediction, and even mental health detection such as depression from facial features [58], [59]. Additionally, hybrid models are often more adaptable and robust in real-world scenarios as they can balance the accuracy of deep learning with the efficiency and speed of traditional methods. Another key advantage of these hybrid systems is their ability to handle dynamic and vast data, making them ideal for real-time applications. They also have the potential to solve problems related to facial expression variability and complexity, which have typically been challenges in emotion and personality trait recognition systems.

3.0 METHODOLOGY

Generally, in the development of the personality trait recognition model, there are several main processes that are carried out consecutively, namely data preprocessing, feature extraction and selection, classification modelling, and final prediction. The initial step of data preprocessing is the process of extracting information from raw data sources, such as identifying key images or frames from video sequences. Key frame extraction is an important task in video processing that involves selecting the best frames to represent the content of a video. The key frame selection step is to choose highly relevant input data that can be used in the next step of feature extraction. Feature extraction is the step of extracting features from the modality input as representations, while feature selection is related to choosing the most relevant features to improve classification accuracy and reduce computational resources [63]. Following the extraction and selection of relevant features, the classification step utilizes the data to determine feature classes based on the characteristics of the features. The final step of the personality trait recognition model is to classify subjects into personality traits classes based on the chosen personality model traits. This study used the Big Five personality model which consists of five classes of traits, namely openness, conscientiousness, extraversion, agreeableness, and neuroticism. Our proposed method consists of several steps aimed at extracting the best frames from a video for the personality traits recognition task. The steps start with extracting key frames from videos and then are followed by applying face detection models to detect human faces and extract facial features. The extracted facial features are then fed into CNN layers, fused in fully connected layers and finally used a sigmoid layer is used to get the final score of the Big Five personality traits model. In the following section, each step for key frames selection, facial feature extraction with a face detector model and personality traits classification using CNN-based approaches is explained in detail.

3.1 Key Frame Selection

In the initial stage of key frame selection, video pre-processing is carried out by converting video data in MP4 format into a sequence of still images in JPEG format. This conversion is important to allow for the subsequent analysis of individual frames in each video. The pre-processing phase allows for extracting meaningful frames that effectively represent the content of the video. The overall process for key frame selection and extraction involves several sequential steps, which are frame differencing, smoothing the frame differences, finding local maxima, clustering similar candidate frames using HDBSCAN, and finally selecting key frames based on the Laplacian score. At the end of these steps, a set of key images is generated for each video. In detail, the process begins with frame differencing, where it starts with identifying and choosing possible frames that differ with each other as candidate frames. This helps identify candidate frames that capture significant differences or changes in the video. To calculate these differences, the cv2.absdiff function from the OpenCV library is used. The cv2.absdiff function highlights the areas of frame transitions by giving a measure of pixel-level changes and calculating the absolute difference between two consecutive frames. The degree of change between frames is reflected in these differences, providing important information for identifying candidates for key frames. A series of frame differences is generated by analyzing every pair of frames sequences in the video.

Following frame differences, the process moves to smoothing the frame differences. The smoothing process reduces the noise in the difference data and highlights the important changes between frames. By focusing on notable changes, this step ensures that only the most relevant frames are retained for further analysis. The smoothing process also eliminates inconsistencies and ensures that the dataset contains only high-quality candidate frames with meaningful features. This step enhances the reliability of key frame extraction and selection. Next, once the frame differences have been smoothed, the local maxima are identified. Local maxima are specific points within the data where the value of the frame difference is greater than the values immediately before and after it. These local maxima are used to detect frames that represent a significant peak in change. In the context of video key frame extraction, local maxima work as indicators of potential key points where critical changes or transitions occur. Identifying these peaks helps focus the analysis on frames with the most impactful changes, minimizing redundancy and improving the quality of candidate frames. Moving forward, the process involves the step of clustering the candidate key frames using the HDBSCAN clustering algorithm. HDBSCAN is a density-based clustering method that groups similar frames together while discarding noise or outliers. This algorithm is useful

for removing redundancy in the set of candidate frames by grouping visually similar frames into clusters. By ensuring that only distinct frames are retained, the clustering process further improves the selection of key frames. This step is also important for enhancing the efficiency of the model as it reduces the computational load and ensures that only the most representative frames are carried forward.

Finally, the key frames are selected based on their Laplacian score. The Laplacian score is a measure used to assess an image's level of texture in detail. The textures and edges are emphasized and computed using a Laplacian operator. Within each cluster in the previous step, the frame with the highest Laplacian score is selected as the key frame because it contains more detailed visual information compared to others. By calculating the Laplacian score for each candidate frame, it helps to identify the frame with the highest score in each cluster. In short, the higher the Laplacian scores, the more informative and significant the frames are. This step ensures that the selected key frames are rich in visual details and are able to provide a comprehensive summary of the video content. At the end of the entire process, a refined set of key frames is obtained for each video in JPEG format. These frames serve as the basis for further analysis of personality traits recognition. The combination of sequential steps including frame differencing, smoothing frame differences, local maxima identification, clustering candidate frames with HDBSCAN, and Laplacian scoring ensures that the selected key frames are both meaningful and relevant for the facial features extraction task. Figure 1 illustrates each step in the key frame extraction and selection process.

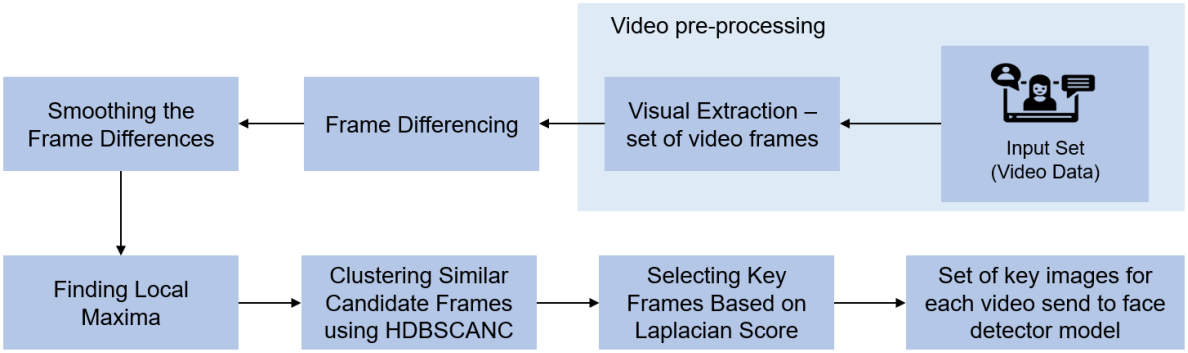

Fig. 1: Illustration of key frame extraction and selection

3.2 Facial Features Extraction

After selecting the key frames, the next step is to apply face detection using three selected models, which are MTCNN, RetinaFace, and DLIB, each applied independently. Employing multiple models allows for a comparative analysis of model performance in terms of face detection accuracy, efficiency, and robustness within the selected key frames. These face detector models are applied to each frame to identify whether a human face is present in the image. The use of multiple models is particularly important for understanding the strengths and weaknesses of each model in handling varied scenarios such as changes in facial orientation, lighting conditions, and occlusions. Once a human face is detected by the face detector algorithm, the model proceeds to extract facial landmark features. These features are used as a basis to compute geometric features or appearance-based features for further processing in the CNN layer. Geometric features are those based on the geometry or shape of an object. In the context of facial features, geometric features refer to attributes such as positions, angles, distances, and relationships between key landmark points on the face. Key landmark points include eyes, nose, and mouth. These features provide a structured representation of the face's spatial configuration.

On the other hand, appearance-based features involve the visual texture and pixel-level details of the face. These can include attributes like color histograms, edge orientations and features extracted by convolutional neural networks (CNNs). For example, VGG16 is a widely used CNN-based model for generating deep facial features that capture high-level abstract patterns present in the image. These appearance features complement geometric data, offering a richer representation of the facial structure and characteristics. To enhance the prediction capabilities of the personality traits recognition model, the extracted data combines both geometric and appearance features. The Euclidean distances between facial landmarks are calculated to represent geometric relationships quantitatively and then combined with CNN-based appearance features, enriching the model's input dataset. The result is a series of facial feature data that includes the raw image, VGG16-based deep features, facial landmarks extracted by the chosen detection models, and calculated geometric Euclidean distances. By integrating these various features, the process ensures a richer representation of facial attributes, enhancing the robustness and accuracy of the personality traits recognition model. This multi-visual approach improves the ability to classify personality traits effectively and makes the model adaptable to varying conditions in video datasets.

3.3 Personality Traits Classification using CNN-Based Techniques

The final step in our proposed method employs a customized CNN-based approach specifically designed for personality traits recognition. This approach is inspired by previous research on the development of Descriptor Aggregation Networks (DANs) for personality traits recognition [42]. This approach is tailored to process multiple types of inputs, ensuring that multi-visual facial feature data can be utilized effectively for accurate personality classification. Our custom CNN-based model takes four inputs, including raw key images sized 224 by 224 with 3 color channels (224X224X3), VGG16-based features with 4096 dimensions, facial landmarks, and Euclidean distances. The facial landmarks vary in representation depending on the face detection model used. For MTCNN and RetinaFace, they are represented as a 10-dimensional vector corresponding to five key points, which are left eye, right eye, nose tip, left corner of the mouth, and right corner of the mouth. In contrast, DLIB landmarks are represented as a 136-dimensional vector, capturing a vector set of 68 facial points. Similarly, the Euclidean distances between landmarks are a 10-dimensional vector for MTCNN and RetinaFace, while for DLIB, the distances span a 2278-dimensional vector, reflecting the greater number of detected landmark points.

All inputs are processed through distinct layers of the CNN model. The raw image input goes through several convolutional layers to detect key facial features and all related information. Simultaneously, the VGG16 features, facial landmarks, and Euclidean distances are processed independently through dense layers, which transform these numerical inputs into high-dimensional feature representations. These independent streams of data are then concatenated into a unified feature vector, which is passed through a fully connected layer. This integration allows the model to combine significant information from all inputs, creating a richer representation of facial features. A dropout layer is added to enhance generalization and prevent overfitting. The final layer of the model employs a sigmoid activation function to predict scores for the Big Five personality traits, which are openness, conscientiousness, extraversion, agreeableness, and neuroticism. Each trait is treated as an independent regression target, enabling the model to output continuous scores for multi-label classification. The training process utilizes the Adam optimizer for its adaptive learning rate, with Mean Squared Error (MSE) as the loss function, which is suitable for regression tasks. This approach effectively combines features from raw images and landmark features extracted using a face detector model. Figure 2 illustrates the steps for key frame selection, facial feature extraction with a face detector model, and personality traits classification using CNN-based approaches.

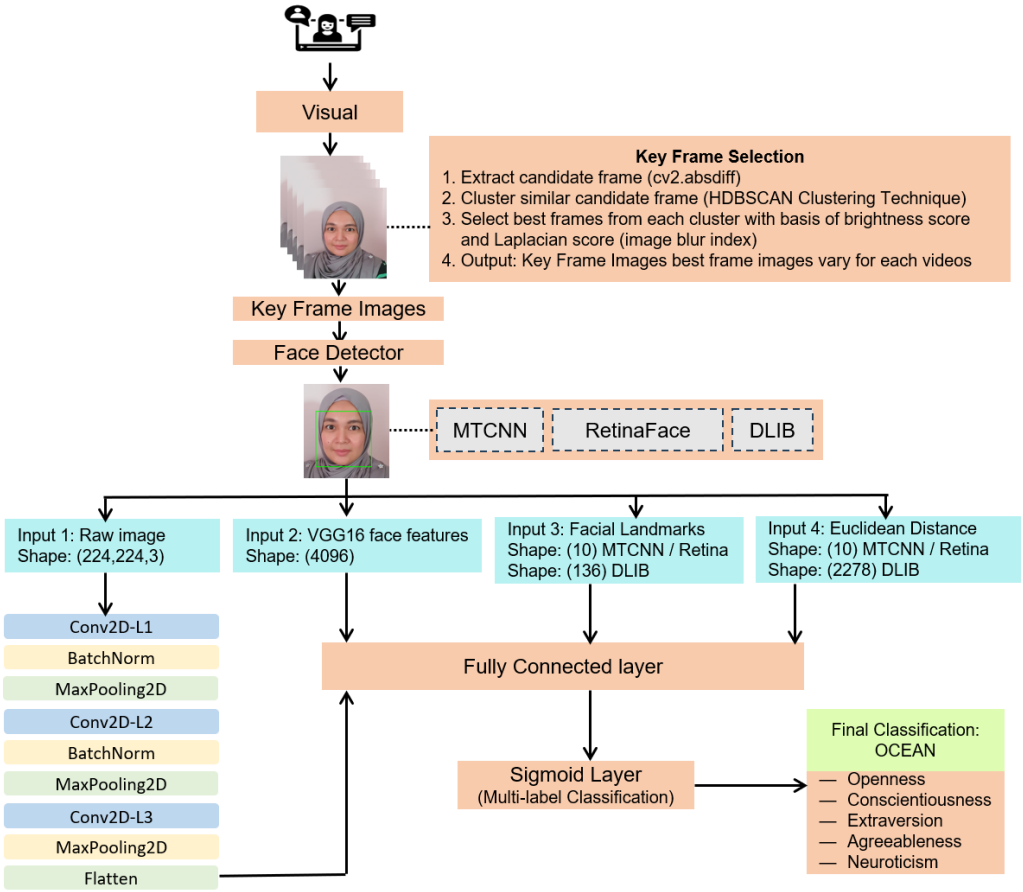

Fig. 2: Illustration of facial feature extraction steps with face detector model for PTR

4.0 RESULTS AND DISCUSSION

4.1 Dataset and Evaluation Metric

This study used the publicly available ChaLearn dataset which was previously developed for personality traits recognition research during the Job Candidate Screening Coopetition [36]. The ChaLearn dataset consists of 60,000 short videos for training and 20,000 videos for validation. The videos show people speaking in front of a camera, demonstrating a wide range of human behavior, facial orientations, expressions, and occlusions. The duration of each video is 15 seconds, and they were collected through the YouTube platform. The dataset is labelled with five classes of the Big Five model personality traits, which are openness, conscientiousness, extraversion, agreeableness, and neuroticism. The labelling process was done by human annotators using Amazon Mechanical Turk (AMT). The ChaLearn dataset has played a crucial role in helping personality traits recognition research using video data, as it provides valuable resources for researchers to conduct their studies. There are several personality traits recognition studies that used the ChaLearn dataset in their experiments [39], [64], [65].

To provide a more comprehensive evaluation of face detection performance, we incorporated standard detection metrics, including recall (true positive rate), precision, and F1-score, for each face detector. These metrics offer deeper insights into the consistency and reliability of each model, particularly under challenging conditions such as occlusion and variations in facial orientation. Next, the performance of the proposed personality traits recognition model is evaluated by closely monitoring both training and validation metrics throughout the learning process. To prevent overfitting and maintain optimal performance, two callbacks are employed, which are ReduceLROnPlateau to adjust the learning rate used and EarlyStopping to halt the training process if validation loss no longer improves. As for the evaluation metric, loss and Mean Absolute Error (MAE) are tracked for both the training and validation datasets. Lower loss values generally indicate better model performance, meaning the predicted values are closer to the true labels. Lower MAE indicates more accurate predictions, as it directly reflects the average absolute difference between predictions and true values. The prediction accuracy is also computed as one minus the MAE value, providing a straightforward measure of how closely the predicted values align with the ground truth. This approach allows for a clear assessment of the model's performance in accurately predicting personality traits from the video data, ensuring that the predictions closely reflected the actual observed traits. By employing this evaluation approach, we were able to quantify the performance of our proposed models and assess their effectiveness in predicting personality traits from video data. In the following section, we discuss our experimental results in detail.

4.2 Analysis of Experimental Result

In this study, the evaluation of the proposed approach is based on two key results, which are the face detection performance and the accuracy of the personality traits recognition model. The consistency and effectiveness of face detection on key frames are important, as they directly influence the quality of extracted features and, consequently, the recognition performance. To assess face detection reliability, three well-known face detector models, including MTCNN, RetinaFace, and DLIB, were compared based on their ability to detect faces under varying conditions, such as non-frontal poses, occlusion, and low-resolution imagery, which are common in the ChaLearn dataset. The aim is to identify the model that consistently provides the most accurate and robust detection across the ChaLearn dataset. A total of 76,658 key frames were extracted from the training videos, and 25,409 from the validation set. MTCNN successfully detected faces in 35,618 training and 12,142 validation images. RetinaFace achieved similar performance, detecting 35,513 and 11,802 faces, respectively. In contrast, DLIB detected significantly fewer faces, with only 17,601 in training and 6,188 in validation. To provide a comprehensive evaluation, performance metrics including detection rate, recall, precision, and F1-score, were calculated using formulas as listed below.

i. Detection Rate (True Positive Rate, Recall):

$$Recall = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP) + False\ Negative\ (FN)} = \frac{Detected\ Key\ Frames}{Total\ Key\ Frames} \quad (1)$$

ii. Precision

$$Precision = \frac{TP}{TP + FN} = 1 \quad (\text{since False Positive (FP)} = 0) \quad (2)$$

iii. F1-Score

$$F1 = 2 \text{ X } \frac{Precision \; X \; Recall}{Precision + \; Recall} = \frac{2 \text{ X } Recall}{1 + Recall} \tag{3}$$

Since all key frames are known to contain faces, false positives are absent, and precision remains at 1.0 for all models. Thus, recall and detection rate become the primary indicators of detection robustness. On the training set, MTCNN and RetinaFace achieved detection rates of 46.46% and 46.33%, respectively, while DLIB showed lower performance at 22.96%. Similar trends were observed on the validation set, with MTCNN and RetinaFace scoring 47.79% and 46.44%, compared to 24.36% for DLIB. These detection rates directly correspond to recall and F1-score, emphasizing the superior ability of MTCNN and RetinaFace to handle complex facial variations in the dataset. These results are also expected due to the challenges present in the ChaLearn dataset. Although all key frames are extracted from video frames containing faces, many of these faces appear under non-frontal angles, low resolutions, occlusions, or poor lighting conditions, which can significantly impact face detector performance. Importantly, MTCNN and RetinaFace demonstrate stronger robustness to such conditions than DLIB, as reflected in their higher recall and F1-score values. Table 2 summarizes the number of detected faces by each model, while Table 3 and Table 4 present the corresponding detection metrics across training and validation sets respectively. The results clearly indicate that MTCNN and RetinaFace are more robust and reliable face detectors in the context of personality trait recognition on the ChaLearn dataset.

Table 2: Total key frame images for each face detector model

Total Key Frames Images	Total Key Frames Images with MTCNN Face Detector	Total Key Frames Images with RetinaFace Face Detector	Total key Frames Images with DLIB Face Detector
Training 76,658	35,618	35,513	17,601
Validation 25,409	12,142	11,802	6,188

Table 3: Training Set (Total Key Frames = 76,658)

Detector	Detected Key Frames (TP)	Recall (Detection Rate)	Precision	F1-Score
MTCNN	35,618	46.46%	1	0.6345
RetinaFace	35,513	46.33%	1	0.6332
DLIB	17,601	22.96%	1	0.3735

Table 4: Validation Set (Total Key Frames = 25,409)

Detector	Detected Key Frames (TP)	Recall (Detection Rate)	Precision	F1-Score
MTCNN	12,142	47.79%	1	0.6467
RetinaFace	11,802	46.45%	1	0.6343
DLIB	6,188	24.35%	1	0.3917

The accuracy of the personality traits recognition model is used to evaluate the performance of the proposed approach. Figure 3 illustrates the training and validation accuracy using three different face detection algorithms on key images. MTCNN achieved a training accuracy of 0.95091, closely followed by RetinaFace at 0.94780, and then DLIB at 0.94483. Despite these differences in training accuracy, validation accuracy remained consistent across all models, with MTCNN at 0.89618, RetinaFace at 0.89555, and DLIB at 0.89622. This consistency in validation accuracy suggests that all three models generalize comparably well on unseen data. However, it's important to note that DLIB only analysed half of the dataset during face recognition, so that DLIB's accuracy results cannot be directly compared with MTCNN and RetinaFace. This limitation occurred because DLIB struggled to detect faces in certain challenging conditions such as non-frontal poses or when the face was partially

hidden or blocked (occlusion). In contrast, MTCNN and RetinaFace, were able to detect faces in more key frame images covering half of the dataset. Although DLIB achieved validation accuracy similar to MTCNN and RetinaFace, this result is based on a much smaller portion of the dataset. Since it missed many faces in key images, its accuracy is calculated on a smaller range of data, which may not fully represent the overall dataset. In simple words, DLIB's performance looks good, but it was tested on fewer frames, meaning it did not cover more diverse data like MTCNN and RetinaFace. This makes it unfair to directly compare DLIB's accuracy with the other two models. Nevertheless, the experiments demonstrated that the features extracted from detected faces are sufficiently reliable for personality recognition tasks. Since MTCNN and RetinaFace were able to detect more faces in the key image dataset, both models are more suitable face detectors for personality traits recognition, especially for the Chaleran dataset.

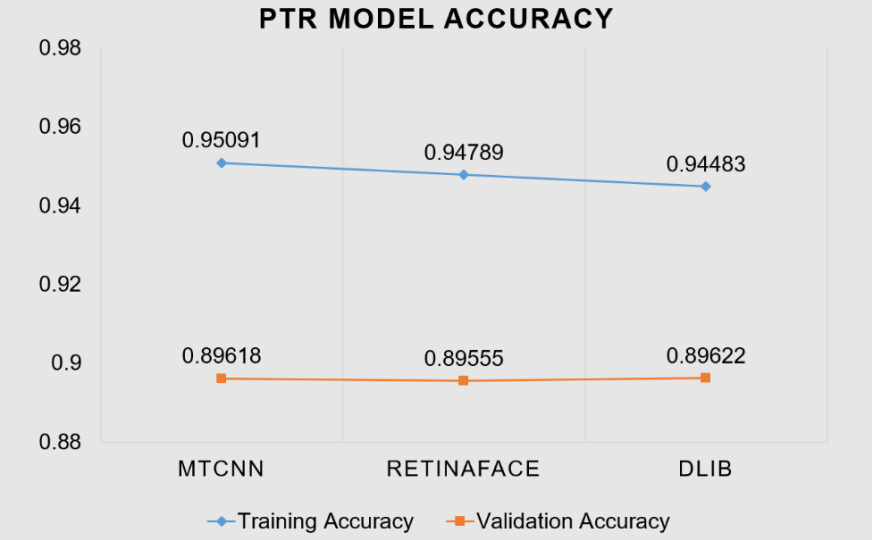

Fig. 3: Training and validation accuracy of PTR for each face detector model

5.0 CONCLUSION

In this study, we presented a comparative evaluation of MTCNN, RetinaFace, and DLIB face detection models for personality traits recognition from video-based key frames. Our experimental results demonstrated that MTCNN and RetinaFace were significantly more effective than DLIB for face detection in the ChaLearn dataset, which contains videos of individuals speaking directly to the camera. Both MTCNN and RetinaFace detected more than double the number of faces than DLIB, highlighting their capability in handling facial variations. MTCNN and RetinaFace have been developed with multi-stage processing layers that adapt well to various angles, making them more robust in detecting faces regardless of face orientations and occlusions. On the other hand, DLIB struggles with any non-frontal view, as seen in its lower detection rates. This study emphasizes the importance of robust face detection for accurate recognition of personality traits recognition. MTCNN and RetinaFace demonstrated consistency and flexibility in detecting faces and extracting facial features, which are essential for this task. Furthermore, in the personality recognition task, models utilizing features detected by MTCNN and RetinaFace achieved higher training accuracy rates compared to those relying on DLIB. However, in terms of validation accuracy, all three models performed similarly, indicating stable performance across varied conditions. Given this stability, any of these models could be a practical option for video-based personality trait recognition, with the choice depending on other factors such as processing time, resource availability, and dataset complexity. Recent studies have demonstrated that the MTCNN face recognition algorithm offers good response speed for intelligent education management systems [66], RetinaFace exhibits high performance in face counting tasks using pre-trained models [67], and DLIB has been successfully applied to detect fatigue while driving based on facial features [68].

The challenging aspect in processing video data also lies in the high computation costs due to the nature of video datasets [9], [69]. With the large number of frames, video processing requires large memory and runtime space. To address this challenge, we implemented key frame extraction and selection from a video dataset to make sure that only significant frames are selected for the recognition process. In this study, we extracted key frames from the dataset by leveraging the OpenCV library, smoothing, and clustering techniques. The key frame selection method was also developed to address issues in redundancy and computational inefficiency associated with processing entire video frames. The key frame selection method focuses on selecting frames that capture the most

significant content of each video. This method makes use of techniques for identifying key frames based on their visual content, which significantly reduces the number of frames that need to be processed while preserving the essential characteristics required for accurate personality traits recognition. This study also highlights the importance of a large and diverse dataset in improving model accuracy, as well as the need for handling challenges such as facial orientation, illumination, and background variability. This study also discovered that robust and efficient face detection ensures that the model can capture and analyze facial features more accurately. Facial features are an important input for understanding personality traits, making them popular for personality recognition tasks [70], [71]. In addition to facial features, other modalities such as audio signals and text also make significant contributions to the development of more comprehensive models for personality trait recognition [72], [73]. This kind of recognition system has promising applications in a range of fields, including digital business marketing, human-computer interaction, and psychological analysis. Future research could explore combining multiple feature extraction models to enhance detection accuracy across diverse datasets, such as facial expression and non-verbal signals from disabled people. Reliable personality traits recognition can advance and promote human-centered AI by enhancing psychological evaluation tools, improving user experience on interactive platforms, and strengthening adaptive systems in social robotics.

REFERENCES

[1] L. V. Phan and J. F. Rauthmann, "Personality computing: New frontiers in personality assessment," *Soc. Personal. Psychol. Compass*, vol. 15, no. 7, 2021, doi: 10.1111/spc3.12624.

[2] G. Srivastava and S. Bag, "Modern-day marketing concepts based on face recognition and neuro-marketing: a review and future research directions," *Benchmarking*, vol. 31, no. 2. 2024, doi: 10.1108/BIJ-09-2022-0588.

[3] M. S. Alam, Z. Tasneem, S. A. Khan, and M. M. Rashid, "Effect of Different Modalities of Facial Images on ASD Diagnosis using Deep Learning-Based Neural Network," *J. Adv. Res. Appl. Sci. Eng. Technol.*, vol. 32, no. 3, 2023, doi: 10.37934/araset.32.3.5974.

[4] R. Perez-Siguas, H. Matta-Solis, E. Matta-Solis, L. Perez-Siguas, H. Matta-Perez, and A. Cruzata-Martinez, "Emotion Analysis for Online Patient Care using Machine Learning," *J. Adv. Res. Appl. Sci. Eng. Technol.*, vol. 30, no. 2, 2023, doi: 10.37934/araset.30.2.314320.

[5] Chunming Wu and Ying Zhang, "MTCNN and FACENET Based Access Control System for Face Detection and Recognition," *Autom. Control Comput. Sci.*, vol. 55, no. 1, 2021, doi: 10.3103/S0146411621010090.

[6] S. Song, S. Jaiswal, E. Sanchez, G. Tzimiropoulos, L. Shen, and M. Valstar, "Self-supervised Learning of Person-specific Facial Dynamics for Automatic Personality Recognition," *IEEE Trans. Affect. Comput.*, 2021, doi: 10.1109/TAFFC.2021.3064601.

[7] C. Stachl *et al.*, "Personality Research and Assessment in the Era of Machine Learning," *Eur. J. Pers.*, vol. 34, no. 5, 2020, doi: 10.1002/per.2257.

[8] X. Sun, J. Huang, S. Zheng, X. Rao, and M. Wang, "Personality Assessment Based on Multimodal Attention Network Learning with Category-Based Mean Square Error," *IEEE Trans. Image Process.*, vol. 31, 2022, doi: 10.1109/TIP.2022.3152049.

[9] A. A. Gharahbagh, V. Hajihashemi, M. C. Ferreira, J. J. M. Machado, and J. M. R. S. Tavares, "Best Frame Selection to Enhance Training Step Efficiency in Video-Based Human Action Recognition," *Appl. Sci.*, vol. 12, no. 4, 2022, doi: 10.3390/app12041830.

[10] B. O. Sadiq, B. Muhammad, M. N. Abdullahi, G. Onuh, A. A. Muhammed, and A. E. Babatunde, "Keyframe Extraction Techniques: A Review," *Elektr. J. Electr. Eng.*, vol. 19, no. 3, pp. 54–60, Dec. 2020, doi: 10.11113/ELEKTRIKA.V19N3.221.

[11] T. Hachaj, "Key frames detection in motion capture recordings using machine learning approaches," in *Advances in Intelligent Systems and Computing*, 2017, vol. 525, doi: 10.1007/978-3-319-47274-4_9.

[12] H. Tang, H. Liu, W. Xiao, and N. Sebe, "Fast and robust dynamic hand gesture recognition via key frames extraction and feature fusion," *Neurocomputing*, vol. 331, 2019, doi:

10.1016/j.neucom.2018.11.038.

[13] E. T. Khalid, S. A. Jassim, and S. Saqaeeyan, “Fuzzy C-mean clustering technique based visual features fusion for automatic video summarization method,” *Multimed. Tools Appl.*, 2024, doi: 10.1007/s11042-024-18820-w.

[14] G. Man and X. Sun, “Interested Keyframe Extraction of Commodity Video Based on Adaptive Clustering Annotation,” *Appl. Sci.*, vol. 12, no. 3, 2022, doi: 10.3390/app12031502.

[15] M. Dhanushree, R. Priya, P. Aruna, and R. Bhavani, “A Keyframe Extraction Using HDBSCAN With Particle Swarm Optimization,” *Proc. 10th Int. Conf. Signal Process. Integr. Networks, SPIN 2023*, pp. 445–450, 2023, doi: 10.1109/SPIN57001.2023.10117200.

[16] J. W. Ma and F. Leite, “Performance boosting of conventional deep learning-based semantic segmentation leveraging unsupervised clustering,” *Autom. Constr.*, vol. 136, 2022, doi: 10.1016/j.autcon.2022.104167.

[17] J. Ren, X. Shen, Z. Lin, and R. Mech, “Best Frame selection in a short video,” in *Proceedings - 2020 IEEE Winter Conference on Applications of Computer Vision, WACV 2020*, 2020, pp. 32112–3221, doi: 10.1109/WACV45572.2020.9093615.

[18] Z. G. Jiang and X. T. Shi, “Application Research of Key Frames Extraction Technology Combined with Optimized Faster R-CNN Algorithm in Traffic Video Analysis,” *Complexity*, vol. 2021, 2021, doi: 10.1155/2021/6620425.

[19] L. Tisu, D. Lupșa, D. Vîrgă, and A. Rusu, “Personality characteristics, job performance and mental health the mediating role of work engagement,” *Pers. Individ. Dif.*, vol. 153, 2020, doi: 10.1016/j.paid.2019.109644.

[20] S. E. Babcock and C. A. Wilson, “Big Five Model of Personality,” *Wiley Encycl. Personal. Individ. Differ. Personal. Process. Individ. Differ.*, pp. 55–60, 2020, doi: 10.1002/9781119547174.ch186.

[21] P. J. Ramos-Villagrasa, J. R. Barrada, E. Fernández-Del-Río, and L. Koopmans, “Assessing job performance using brief self-report scales: The case of the individual work performance questionnaire,” *Rev. Psicol. del Trab. y las Organ.*, vol. 35, no. 3, 2019, doi: 10.5093/jwop2019a21.

[22] H. Kaya and A. A. Salah, “Multimodal Personality Trait Analysis for Explainable Modeling of Job Interview Decisions,” *Explain. Interpret. Model. Comput. Vis. Mach. Learn. Cham Springer Int. Publ.*, pp. 255–275, 2018, doi: 10.1007/978-3-319-98131-4_10.

[23] J. Anglim, K. Molloy, P. D. Dunlop, S. L. Albrecht, F. Lievens, and A. Marty, “Values assessment for personnel selection: comparing job applicants to non-applicants,” *Eur. J. Work Organ. Psychol.*, vol. 31, no. 4, pp. 524–536, Jul. 2022, doi: 10.1080/1359432X.2021.2008911.

[24] T. L. Ting and K. D. Varathan, “Job recommendation using facebook personality scores,” *Malaysian J. Comput. Sci.*, vol. 31, no. 4, 2018, doi: 10.22452/mjcs.vol31no4.5.

[25] J. R. Bodhe, “Gamification of personality tests for recruitment,” *Int. J. Res. Hum. Resour. Manag.*, vol. 3, no. 2, 2021, doi: 10.33545/26633213.2021.v3.i2a.66.

[26] M. Chugunova and A. Danilov, “Use of Digital Technologies for HR Management in Germany: Survey Evidence,” *SSRN Electron. J.*, 2022, doi: 10.2139/ssrn.4010539.

[27] S. T. Janetius, P. Varma, and S. Shilpa, “Projective tests in human resource management and hiring process: a challenge and a boon,” *Int. J. Indian Psychol.*, vol. 7, no. 4, 2019.

[28] M. M. Karim and W. Bin Latif, “Conceptual Framework of Recruitment and Selection Process,” *Artic. J. Bus. Soc. Sci. Res.*, vol. 11, no. 02, 2021.

[29] M. Collins, “Time to Make it Personal: How Personality Testing in Law Schools Can Improve Lawyer Well-Being,” *SSRN Electron. J.*, 2021, doi: 10.2139/ssrn.3772461.

[30] A. Remaida, A. Moumen, Y. El Bouzekri El Idrissi, B. Abdellaoui, and Y. Harraki, “The use of personality tests as a pre-employment tool: A comparative study,” *SHS Web Conf.*, vol. 119, 2021, doi: 10.1051/shsconf/202111905007.

[31] R. M. Spielman, K. Dumper, W. Jenkins, A. Lacombe, M. Lovett, and M. Perlmutter, “Personality Assessment,” *Psychology-H5P Edition*, 2020. https://pressbooks.bccampus.ca/psychologyh5p/chapter/personality-assessment/ (accessed April. 05, 2025).

[32] Y. Mehta, N. Majumder, A. Gelbukh, and E. Cambria, “Recent trends in deep learning based personality detection,” *Artif. Intell. Rev.*, 2019, doi: 10.1007/s10462-019-09770-z.

[33] X. Huang, J. Sang, and C. Xu, “Image-Based Personality Questionnaire Design,” *ACM Trans. Multimed. Comput. Commun. Appl.*, vol. 18, no. 4, 2022, doi: 10.1145/3503489.

[34] A. Kachur, E. Osin, D. Davydov, K. Shutilov, and A. Novokshonov, “Assessing the Big Five personality traits using real-life static facial images,” *Sci. Rep.*, vol. 10, no. 1, 2020, doi: 10.1038/s41598-020-65358-6.

[35] J. Li, A. Waleed, and H. Salam, “A Survey on Personalized Affective Computing in Human-Machine Interaction,” *arXiv Prepr. arXiv2304.00377*, 2023.

[36] H. Kaya, F. Gurpinar, and A. A. Salah, “Multi-modal Score Fusion and Decision Trees for Explainable Automatic Job Candidate Screening from Video CVs,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2017, vol. 2017-July, doi: 10.1109/CVPRW.2017.210.

[37] L. Cai and X. Liu, “Identifying Big Five personality traits based on facial behavior analysis,” *Front. Public Heal.*, vol. 10, 2022, doi: 10.3389/fpubh.2022.1001828.

[38] H. Y. Suen, K. E. Hung, and C. L. Lin, “TensorFlow-Based Automatic Personality Recognition Used in Asynchronous Video Interviews,” *IEEE Access*, vol. 7. 2019, doi: 10.1109/ACCESS.2019.2902863.

[39] C. Suman, S. Saha, A. Gupta, S. K. Pandey, and P. Bhattacharyya, “A multi-modal personality prediction system,” *Knowledge-Based Syst.*, vol. 236, 2022, doi: 10.1016/j.knosys.2021.107715.

[40] M. Saberi, S. DiPaola, and U. Bernardet, “Expressing Personality Through Non-verbal Behaviour in Real-Time Interaction,” *Front. Psychol.*, vol. 12, 2021, doi: 10.3389/fpsyg.2021.660895.

[41] Z. Zhao, N. Zhou, L. Zhang, H. Yan, Y. Xu, and Z. Zhang, “Driver Fatigue Detection Based on Convolutional Neural Networks Using EM-CNN,” *Comput. Intell. Neurosci.*, vol. 2020, 2020, doi: 10.1155/2020/7251280.

[42] C. L. Zhang, H. Zhang, X. S. Wei, and J. Wu, “Deep bimodal regression for apparent personality analysis,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016, vol. 9915 LNCS, doi: 10.1007/978-3-319-49409-8_25.

[43] X. Zhao *et al.*, “Integrating audio and visual modalities for multimodal personality trait recognition via hybrid deep learning,” *Front. Neurosci.*, vol. 16, 2023, doi: 10.3389/fnins.2022.1107284.

[44] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, “Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks,” *IEEE Signal Process. Lett.*, vol. 23, no. 10, 2016, doi: 10.1109/LSP.2016.2603342.

[45] M. Gu, X. Liu, and J. Feng, “Classroom face detection algorithm based on improved MTCNN,” *Signal, Image Video Process.*, vol. 16, no. 5, 2022, doi: 10.1007/s11760-021-02087-x.

[46] K. R. Jamaluddin and S. Ibrahim, “A Review on Occluded Object Detection and Deep Learning-Based Approach in Medical Imaging-Related Research,” *J. Adv. Res. Appl. Sci. Eng. Technol.*, vol. 34, no. 2,

2024, doi: 10.37934/araset.34.2.363373.

[47] N. Zhou, R. Liang, and W. Shi, “A Lightweight Convolutional Neural Network for Real-Time Facial Expression Detection,” *IEEE Access*, vol. 9, 2021, doi: 10.1109/ACCESS.2020.3046715.

[48] A. Baskar, T. G. Kumar, and S. Samiappan, “A vision system to assist visually challenged people for face recognition using multi-task cascaded convolutional neural network (MTCNN) and local binary pattern (LBP),” *J. Ambient Intell. Humaniz. Comput.*, vol. 14, no. 4, 2023, doi: 10.1007/s12652-023-04542-8.

[49] J. Deng, J. Guo, E. Ververas, I. Kotsia, and S. Zafeiriou, “Retinaface: Single-shot multi-level face localisation in the wild,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2020, pp. 5203–5212, doi: 10.1109/CVPR42600.2020.00525.

[50] Q. Huang, W. Pan, and X. Fan, “A Mask Detection Algorithm Based on RetinaFace,” in *Proceedings of the 2023 7th International Conference on Machine Learning and Soft Computing*, 2023, pp. 198–204, doi: 10.1145/3583788.3583818.

[51] M. Wan *et al.*, “InfAnFace: Bridging the Infant-Adult Domain Gap in Facial Landmark Estimation in the Wild,” in *Proceedings - International Conference on Pattern Recognition*, 2022, vol. 2022-August, doi: 10.1109/ICPR56361.2022.9956647.

[52] R. Zhong, B. Jiang, N. Li, Q. Wu, and H. Chang, “A multi-view face detection and expression recognition method with improved RetinaFace,” in *International Conference on Mechanisms and Robotics (ICMAR 2022)*, 2022, pp. 1141–1145, doi: 10.1117/12.2652194.

[53] S. Ullah, A. Jan, and G. M. Khan, “Facial Expression Recognition Using Machine Learning Techniques,” in *7th International Conference on Engineering and Emerging Technologies, ICEET 2021*, 2021, pp. 1–6, doi: 10.1109/ICEET53442.2021.9659631.

[54] M. A. Noor Reza, E. A. Zaki Hamidi, N. Ismail, M. R. Effendi, E. Mulyana, and W. Shalannanda, “Design a Landmark Facial-Based Drowsiness Detection Using Dlib And Opencv For Four-Wheeled Vehicle Drivers,” in *Proceeding of 15th International Conference on Telecommunication Systems, Services, and Applications, TSSA 2021*, 2021, pp. 1–5, doi: 10.1109/TSSA52866.2021.9768278.

[55] N. Singhal, V. Ganganwar, M. Yadav, A. Chauhan, M. Jakhar, and K. Sharma, “Comparative study of machine learning and deep learning algorithm for face recognition,” *Jordanian J. Comput. Inf. Technol.*, vol. 7, no. 3, 2021, doi: 10.5455/JJCIT.71-1624859356.

[56] H. Wang and L. Guo, “Research on Face Recognition Based on Deep Learning,” in *2021 3rd international conference on artificial intelligence and advanced manufacture (AIAM)*, 2021, pp. 540–546, doi: 10.1109/AIAM54119.2021.00113.

[57] W. Ali, W. Tian, S. U. Din, D. Iradukunda, and A. A. Khan, “Classical and modern face recognition approaches: a complete review,” *Multimed. Tools Appl.*, vol. 80, no. 3, pp. 4825–4880, 2021, doi: 10.1007/s11042-020-09850-1.

[58] G. Verma and H. Verma, “Hybrid-Deep Learning Model for Emotion Recognition Using Facial Expressions,” *Rev. Socionetwork Strateg.*, vol. 14, no. 2, 2020, doi: 10.1007/s12626-020-00061-6.

[59] Vandana, N. Marriwala, and D. Chaudhary, “A hybrid model for depression detection using deep learning,” *Meas. Sensors*, vol. 25, 2023, doi: 10.1016/j.measen.2022.100587.

[60] K. K. Kumar, Y. Kasiviswanadham, D. V. S. N. V. Indira, P. Priyanka palesetti, and C. V. Bhargavi, “Criminal face identification system using deep learning algorithm multi-task cascade neural network (MTCNN),” in *MaMaterials Today: Proceedings*, 2023, vol. 80, pp. 2406–2410, doi: 10.1016/j.matpr.2021.06.373.

[61] P. Phienphanich, N. Lerthirunvibul, E. Charnnarong, A. Munthuli, C. Tantibundhit, and N. C. Suwanwela, “Generalizing a Small Facial Image Dataset Using Facial Generative Adversarial Networks for Stroke’s Facial Weakness Screening,” *IEEE Access*, vol. 11, pp. 64886–64896, 2023, doi:

10.1109/ACCESS.2023.3287389.

[62] Di. Gyawali, P. Pokharel, A. Chauhan, and S. C. Shakya, “Age Range Estimation Using MTCNN and VGG-Face Model,” in *2020 11th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2020*, 2020, pp. 1–6, doi: 10.1109/ICCCNT49239.2020.9225443.

[63] F. Saleem *et al.*, “Human gait recognition: A single stream optimal deep learning features fusion,” *Sensors*, vol. 21, no. 22, 2021, doi: 10.3390/s21227584.

[64] S. Aslan, U. Güdükbay, and H. Dibeklioğlu, “Multimodal assessment of apparent personality using feature attention and error consistency constraint,” *Image Vis. Comput.*, vol. 110, 2021, doi: 10.1016/j.imavis.2021.104163.

[65] X. Zhao, Z. Tang, and S. Zhang, “Deep Personality Trait Recognition: A Survey,” *Front. Psychol.*, vol. 13, p. 2390, May 2022, doi: 10.3389/FPSYG.2022.839619/BIBTEX.

[66] X. Xiao, Z. Su, Q. Ye, Z. Qin, and L. Wu, “Intelligent education management system design for universities based on MTCNN face recognition algorithm,” *Appl. Math. Nonlinear Sci.*, vol. 9, no. 1, 2024, doi: 10.2478/amns.2023.2.01712.

[67] A. A. Bengeri *et al.*, “Face Counting Based on Pre-trained Machine Learning Models: A Brief Systematic Review,” in *Lecture Notes in Networks and Systems*, 2024, vol. 819, doi: 10.1007/978-981-99-7820-5_29.

[68] Y. Bao and W. Xu, “Design and Implementation of a Fatigue Detection System Based on Dlib for Driver Facial Features,” in *Frontiers in Artificial Intelligence and Applications*, 2024, vol. 381, doi: 10.3233/FAIA231266.

[69] P. Pareek and A. Thakkar, “A survey on video-based Human Action Recognition: recent updates, datasets, challenges, and applications,” *Artif. Intell. Rev.*, vol. 54, no. 3, 2021, doi: 10.1007/s10462-020-09904-8.

[70] A. Benlamoudi, “Frame-Difference & Multi-blocks for Job Candidate Screening Competition,” *ChaLearn LAP2017*, 2017. https://chalearnlap.cvc.uab.cat/challenge/23/track/22/result/ (accessed April. 28, 2025).

[71] S. E. Bekhouche, F. Dornaika, A. Ouafi, and A. Taleb-Ahmed, “Personality Traits and Job Candidate Screening via Analyzing Facial Videos,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2017, vol. 2017-July, doi: 10.1109/CVPRW.2017.211.

[72] F. Gürpinar, H. Kaya, and A. A. Salah, “Multimodal fusion of audio, scene, and face features for first impression estimation,” *Proc. - Int. Conf. Pattern Recognit.*, vol. 0, pp. 43–48, 2016, doi: 10.1109/ICPR.2016.7899605.

[73] Y. Li *et al.*, “CR-Net: A Deep Classification-Regression Network for Multimodal Apparent Personality Analysis,” *Int. J. Comput. Vis.*, vol. 128, no. 12, 2020, doi: 10.1007/s11263-020-01309-y.