Exploring the Feasibility of Neural Networks for Criminal Propensity Detection through Facial Features Analysis

Amal Alshahrani [1], Sumayyah Albarakati [1], Reyouf Wasil [1], Hanan Farouquee [1], Maryam Alobthani [1], and Someah Al-Qarni [1],

amshahrani@uqu.edu.sa, SumayyahAlbarakati@gmail.com, s442009652@st.uqu.edu.sa, s442017249@st.uqu.edu.sa, s442007672@st.uqu.edu.sa, s442008998@st.uqu.edu.sa,

1 College of Computing, Computer Science and Artificial Intelligence Department, Umm Al Qura University, Makkah, Saudi Arabia

Abstract

While artificial neural networks are adept at identifying patterns, they can struggle to distinguish between actual correlations and false associations between extracted facial features and criminal behavior within the training data. These associations may not indicate causal connections. Socioeconomic factors, ethnicity, or even chance occurrences in the data can influence both facial features and criminal activity. Consequently, the artificial neural network might identify linked features without understanding the underlying cause. This raises concerns about incorrect linkages and potential misclassification of individuals based on features unrelated to criminal tendencies. To address this challenge, we propose a novel region-based training approach for artificial neural networks focused on criminal propensity detection. Instead of solely relying on overall facial recognition, the network would systematically analyze each facial feature in isolation. This fine-grained approach would enable the network to identify which specific features hold the strongest correlations with criminal activity within the training data. By focusing on these key features, the network can be optimized for more accurate and reliable criminal propensity prediction. This study examines the effectiveness of various algorithms for criminal propensity classification. We evaluate YOLO versions YOLOv5 and YOLOv8 alongside VGG-16. Our findings indicate that YOLO achieved the highest accuracy 0.93 in classifying criminal and non-criminal facial features. While these results are promising, we acknowledge the need for further research on bias and misclassification in criminal justice applications

Keywords:

Neural networks, Criminal propensity detection, Facial feature extraction

1. Introduction

1.1 Criminal Propensity Detection

The burgeoning field of deep learning, particularly Convolutional Neural Networks (CNNs), has transformed numerous domains, including facial recognition. Within the criminal justice system, CNNs have emerged as a potential tool for criminal propensity detection through facial feature analysis [1]. However, this application raises significant ethical concerns due to the inherent limitations of deep learning models. This study delves into these limitations and proposes a novel approach to mitigate associated risks.

1.2 Pattern Recognition

While CNNs excel at identifying patterns in facial data, they struggle to distinguish between genuine correlations and spurious associations between extracted features and criminal behavior within the training data [2]. These extracted features may not represent causal relationships. Socioeconomic factors, ethnicity, and even inherent biases within the data collection process can influence both facial characteristics and criminal activity. Consequently, a CNN might identify linked features without grasping the underlying cause. This can lead to the creation of biased models and the potential misclassification of individuals based on features irrelevant to criminal tendencies [3]. This scenario raises serious ethical concerns, potentially exacerbating existing social inequalities within the criminal justice system.

1.3 Ethical Considerations

The potential for biased outcomes necessitates a critical examination of the ethical implications surrounding the use of deep learning for criminal propensity detection. Algorithmic bias can lead to discriminatory practices, disproportionately impacting certain demographics and perpetuating societal injustices [4]. Furthermore, Deep learning models' inherent opacity impedes transparency and accountability in criminal propensity classification. Understanding the rationale behind a model's high-risk designation is crucial for ensuring fairness and mitigating potential misclassification errors.

Manuscript received May 5, 2024
Manuscript revised May 20, 2024
https://doi.org/10.22937/IJCSNS.2024.24.5.2

1.4 Introducing a Region-based Approach for More Accurate Classification

This study proposes a novel training approach for artificial neural networks (ANN) employed in criminal propensity detection. Instead of solely relying on overall facial recognition, the proposed method involves a systematic analysis of each facial feature in isolation. This fine-grained approach aims to identify specific features that hold the strongest correlations with criminal activity within the training data. By focusing on these key features, the network can be optimized for more accurate and reliable criminal propensity prediction. Furthermore, this approach enhances model interpretability, allowing researchers to analyze the rationale behind the network's predictions. This fosters trust and transparency, crucial aspects for ethical application in the criminal justice system.

This paper is organized as follows: We begin with an introduction to the topic of ANN in criminal tendency detection through facial features. Following the introduction, Section 2 presents a comprehensive literature review, examining existing research on neural networks in criminal tendency detection. Then we identify the research gap regarding the limitations of current approaches. Section 3 details our proposed methodology for training neural networks, emphasizing a fine-grained analysis of facial features. Following this, Section 4 presents the results of our study, along with corresponding recommendations. Finally, in Section 5, we offer concluding remarks and outline avenues for future research and development in this field.

2. Literature Review

Valla *et al.* [5] discussed the accuracy of dispositional inferences regarding criminality based on brief exposure to static images of convicted criminals and non-criminals. The study begins with a comprehensive discussion of research and theory related to appearance-based inferences of criminality that highlights the historical controversy surrounding this topic. The authors conducted two experiments in which participants were presented with headshots of criminals and non-criminals. They were able to distinguish between the two groups with remarkable accuracy after controlling for various factors, including gender, race, age, attractiveness, emotional displays, and potential clues of picture origin. The findings suggest that rapid and accurate dispositional inferences about criminality can be made based on facial appearance. The study's methodology involved the use of static images and controlled variables.

Wu and Zhang [6] utilized supervised machine learning in their study to build four classifiers: logistic regression, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and CNNs. They trained these classifiers with a dataset of facial images from 1,856 people's standard ID photographs showcasing a variety of races, genders, ages, and neutral facial expressions, with half having criminal records and the other half not. They collected the dataset based on specific requirements: the individuals had to be Chinese, male, aged between 18 and 55, with no facial hair, scars, or other markings on their faces, then analyzed it based on these features: points on the face for landmarks, a vector of features using modular PCA, and histograms of local binary patterns. As a result, the CNN classifier performed the best out of all the methods evaluated, with an accuracy of 89.51%.

Ranjan *et al.* [7] proposesed a novel approach for simultaneous face analysis tasks using a single deep CNNs. The authors address the challenges in tasks such as face detection, alignment, pose estimation, gender recognition, smile detection, age estimation, and face recognition. They introduce a multi-task learning framework that allows the network to learn correlations between different domains and tasks, leading to improved performance. The proposed method demonstrates state-of-the-art results on various datasets, showcasing its effectiveness. The authors leverage pre-training on face recognition to enhance the network's performance on other face-related tasks. They compare their approach to existing methods and highlight the advantages of their MTL framework. Overall, this research provides a comprehensive and efficient solution for face analysis tasks using a unified CNN model.

Johnson *et al.* [8] explorесed previous research on the ability to make personality inferences based on appearance, focusing on facial features. It discusses studies that demonstrate people's accuracy in detecting personality traits and forming judgments about warmth, trustworthiness, and criminality based on facial characteristics. The review highlights the significance of features such as eye size, eyebrows, cheekbones, chins, and facial hair in shaping perceptions of personality and criminal appearance. It emphasizes the impact of race, attractiveness, age, and sex on initial impressions.

The study by Hashemi and Hall [9] investigated the feasibility of using deep learning, to analyze facial images and infer criminal tendencies. They trained two models, a standard feedforward neural network (FNN) and CNNs, on a dataset of 10,000 facial images categorized as criminal or non-criminal. CNNs achieved higher accuracy compared to FNN by 8%, suggesting its potential for criminal tendency classification. Additionally, they examined gender bias by training the model on male faces only and observed no significant difference in accuracy.

The study by Bowyer *et al.* [10] demonstrates the potential of artificial intelligence automated systems for criminal tendency detection to achieve statistically significant results in identifying high-risk individuals. Nonetheless, a significant drawback is the absence of interpretability. Because these models are opaque, users have little knowledge of which factors are most important in determining the model's predictions or the reasons for the identification of particular people as criminals. There are serious ethical consequences to this kind of opacity. It is possible for unfair prejudice to pass into the training set, producing discriminating results.

The study conducted by Sheldon *et al.* [11] aimed to determine the extent to which facial expressions could be used to accurately identify criminals. Additionally, the researchers sought to investigate any differences in the apparent happiness between criminals and non-criminals. The results of the study supported the hypothesis that observers were able to distinguish criminals from non-criminals based on facial expressions of happiness. Notably, the study also suggested a correlation between observers' facial positivity and their ability to perceive positive emotions in non-criminal faces. This research provides critical insights into the connection between facial expressions and criminal behavior, and it fills crucial gaps in previous studies.

Researchers Lin and Adolphs [12] proposed a method using pre-trained deep CNNs to predict social judgments based on facial images. The aim was to determine if machines could make social judgments similar to those of humans. The authors found that CNNs trained for face or object recognition could accurately predict social judgments without specific social judgment training. The study employed supervised learning, specifically regularized linear regression, to train models on a dataset of neutral, frontal, and white faces and their social ratings. The models were tested on independent datasets and social attributes. This study highlights the potential of utilizing pre-trained CNNs to predict social judgments from facial images.

Rasmussen *et al.* [13], The researchers proposed using deep learning techniques, specifically CNNs, to predict political ideology from facial photographs. They aimed to explore the relationship between faces and ideology by analyzing a large dataset of Danish political candidates. The study integrated various techniques, such as heat mapping, facial expression analysis, and assessments of physical characteristics, to identify the specific facial features that contribute to the model's predictions. The problem addressed was the prediction of sensitive personal information, like political ideology, from facial photographs using deep learning approaches. The methodology involved training CNNs using the VGG-16 network on a dataset of 5230 facial photographs of Danish political candidates. The reported results showed an overall predictive accuracy of 61% for both genders, with males reaching 65% accuracy when non-facial information was included. In conclusion, this study demonstrated the potential of deep learning techniques to predict political ideology from facial photographs, highlighting the importance of understanding the contributing factors and the privacy implications.

James *et al.* [14] study investigated the potential of facial expressions to indicate criminal tendencies. They proposed the Adam model, a machine learning model built using traditional CNN to identify facial characteristics linked to three significant criminal behavior theories: psychological, biological, and social. They used a subset of the FER2013 dataset, consisting of 7070 facial images. Furthermore, they analyzed six facial features, which are face shape, eyebrows, eyeballs, pupils, nostrils, and lips. The experiment used 7.8 million parameters to train the model for accurate clarification and 5376 to test its performance. The projected model achieved a training accuracy of 90.6% and validation accuracy.

Existing approaches often rely on global feature extraction from facial images, analyzing the entire face as a single entity. While these methods have achieved some success, they may overlook crucial spatial information within the face. Critically, criminal intent or propensity might be better discerned by analyzing specific facial regions. By neglecting these regional variations, current methods might miss valuable insights that could enhance classification accuracy.

This research proposes a novel region-based framework for criminal propensity classification using facial features. This approach aims to bridge the gap in existing methods by analyzing specific facial regions and potentially capturing more nuanced information for improved classification performance.

3. Methodology

Figure 1 depicts the methodological framework employed in this research. The process commences with data collection as step 1, where facial images are obtained from two sources: a criminal faces dataset and a non-criminal faces dataset. These datasets are then subjected to preprocessing, which is step 2, where the images are segmented into anatomically relevant regions: the auricular region, nasal region, oral and mental region, infraorbital region and zygomatic region. Additionally, the whole face is retained as a complete reference for comparison. This regional segmentation allows for a more nuanced analysis of facial features potentially linked to criminal propensity. By comparing these regional features to the whole face as a baseline, the research aims to identify subtle variations that might hold significance in classification.

Following preprocessing, the research utilizes two deep learning models for model development, which is step 3: YOLO and VGG-16. These models are trained on the preprocessed data, extracting features from the different facial regions. Finally, the trained models undergo evaluation, which is step 4, to assess their performance in classifying criminal and non-criminal profiles based on regional analyses of facial features. The evaluation process will determine the effectiveness of the proposed methodology for criminal propensity classification.

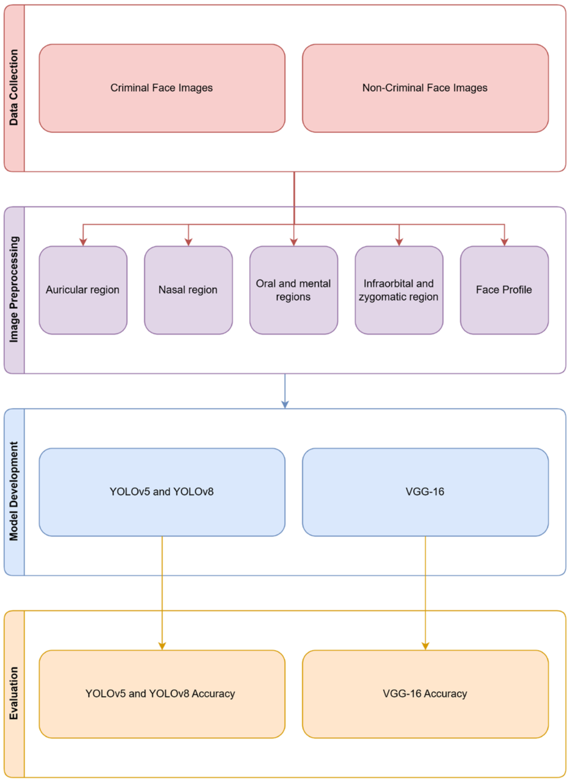


Figure 1: The Proposed Methodological Framework Employed in This Research for Criminal Propensity Classification Using Facial Features with Regional Analysis.

4.1 Dataset

The study leverages the publicly available Illinois Department of Corrections labeled faces dataset [15]. provides facial images of incarcerated individuals, along with associated demographics and criminal records.

To establish a balanced representation for training and evaluation, the study incorporates a dataset of non-criminal facial images. We utilize the dataset presented by Thomaz and Giraldi [16] in their work; this dataset provides facial images of individuals not associated with criminal activity.

4.2 Image Preprocessing

This study explores a novel approach to facial feature localization in profile images through a region-based preprocessing strategy. Facial profile images with embedded facial position coordinates were employed for this investigation. A custom program was developed to perform image segmentation, dividing each image into a uniform grid of 6 sections wide and 5 sections high. This meticulous segmentation aimed to achieve a high degree of isolation for individual facial features within specific sections. Subsequently, the segmented image parts were meticulously categorized based on the specific anatomical region they represented. The established categories encompassed, as mentioned in the Kenhub study's [17]:

- Auricular region: encompassing the anatomical structures of the ear.
- Nasal region: encompassing the structures related to the nose.
- Oral and mental regions: encompassing the anatomical elements of the mouth and chin.
- Infraorbital and zygomatic regions: encompassing the bony region around the eye socket (infraorbital) and the cheekbone (zygomatic).

The training phase of this study adopted a two-pronged approach. First, each meticulously segmented section was treated as an independent data point for the training process. This facilitated the investigation of individual feature recognition capabilities within isolated regions. Second, the entire unsegmented face image was also incorporated into the training regimen. This comprehensive approach enabled a comparative analysis of classification accuracy. By contrasting the performance using whole-face images versus segmented sections, the study aimed to elucidate the potential benefits of a region-based strategy for facial feature localization. This comparative analysis holds significant implications for the development of robust and efficient facial recognition systems.

4.3 YOLO

YOLO offers significant improvement in deep learning, yielding high precision and effectiveness in visual categorization assignments. YOLO, which was developed with quick performance and resilience, offers an exhaustive solution to challenging image recognition problems [18]. YOLO is expected to have an essential role in the detection and identification of faces from photographs in our research, which focuses on the use of facial feature recognition to identify criminal tendencies. Its sophisticated structure enables quick picture processing, which enables precise and quick categorization of any criminal tendencies. In this study, we use YOLOv8 and YOLOv5 to improve the precision and dependability of regional analysis of facial features for the detection of criminal tendencies, boosting security measures and contributing to a decrease in criminal activity.

4.4 VGG-16

VGG-16 stands as a cornerstone in the realm of deep learning, renowned for its significant contributions to image classification tasks. With its deep architecture and meticulous design, the VGG-16 offers remarkable accuracy and reliability in visual recognition endeavors [19]. In our research, VGG-16 holds promise to play an important role. Its robust structure enables comprehensive analysis of facial characteristics, facilitating precise identification and classification of potential criminal traits from images.

Our research approach involves testing VGG-16 and YOLO versions YOLOv5 and YOLOv8 separately to develop a robust system for recognizing the faces of criminals. We will independently assess VGG-16 for its ability to differentiate between criminals and non-criminals based on regional analysis of facial features. Then, we will separately experiment with YOLOv5 and YOLOv8 to evaluate their effectiveness in enhancing classification capabilities. Once we have obtained results from both experiments, we will compare and analyze them together to determine their respective strengths and weaknesses. This approach allows us to thoroughly assess each model's performance before considering potential integration for improved facial recognition and public safety measures.

4. Results

5.2 YOLOv5

The comparison between the YOLOv5 and YOLOv8 deep learning models was conducted to determine their efficiency in achieving the best accuracy. The study used the same datasets, number of classes, and epoch for both models to ensure a fair comparison.

YOLOv5 and YOLOv8 were used in the experiment. They were trained on a classification task: the differentiation between criminal and non-criminal facial features. A dataset with 500 labeled images as criminal and non-criminal with 5 facial regions was provided to the model. The training leveraged the AdamW optimizer for efficient weight updates and Automatic Mixed Precision for faster training. AdamW optimizer incorporates weight decay to prevent overfitting [20]. Automatic Mixed Precision utilizes mixed data precisions during training, improving efficiency on hardware with dedicated low-precision cores [21]. To compare the impact of training duration on model performance and to explore the impact of training duration on model performance, we trained YOLOv5 and YOLOv8 under two different regimes:

1. 5 epochs: As a control experiment, we trained the model for only 5 epochs to assess the influence of training duration on its ability to differentiate between criminal and non-criminal facial features.
2. 10 epochs: This constituted the primary training regime, with the training progress and final validation accuracy documented.
3. and final validation accuracy documented.

Table 1: YOLOv5 Accuracy and Loss Over Epochs

Epoch	Train Loss	Test Loss	Top 1 Accuracy	Top 5 Accuracy
1/10	1.74	1.18	0.74	0.99
2/10	1.00	1.45	0.71	0.99
3/10	0.942	1.11	0.83	1
4/10	0.877	0.889	0.91	1
5/10	0.818	0.898	0.90	1
6/10	0.764	0.987	0.79	1
7/10	0.792	0.905	0.83	1
8/10	0.769	0.859	0.86	1
9/10	0.784	0.808	0.87	1
10/10	0.697	0.762	0.88	1

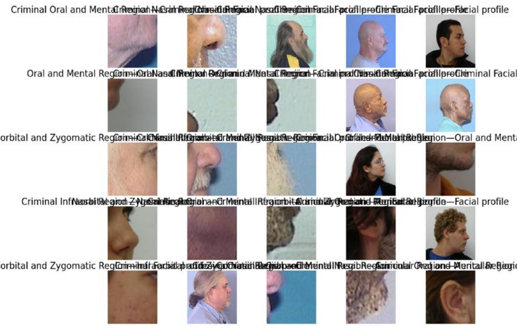

Criminal Oral and Mental Region
Facial profile
Oral and Mental Region
Criminal Facial profile
Criminal Infraorbital and Zygomatic Region
Oral and Mental Region
Facial profile
Infraorbital and Zygomatic Region

Figure 2: Visualization of YOLOv5 Model Predictions with True Labels

Figure 2 displays the true labels and the corresponding YOLOv5 model predictions for five distinct facial regions. Each row represents a different facial region used for criminal propensity classification. Each image within the rows is paired with the true label indicating the facial region and the corresponding prediction by the YOLOv5 model. This visualization demonstrates the model's ability to accurately classify different facial regions, which is critical for improving the precision of criminal propensity classification.

The results revealed that YOLOv5 achieved a higher accuracy of 93% at 5 epochs, which subsequently decreased to 88% at 10 epochs. These findings suggest a potential for overfitting in the YOLOv5 model with extended training durations.

5.1 YOLOv8

This study also investigates the effectiveness of YOLOv8. We evaluate the impact of training epochs on the model's ability to distinguish between criminal and non-criminal facial features and present the best-performing model identified during our experimentation.

The YOLOv8 model exhibits promising results for facial region detection in criminal propensity classification, particularly when trained on RGB images. As stated in Table 2, a 10-epoch training regime using RGB images demonstrated a steady increase in classification accuracy, reaching 93% validation accuracy. The ablated experiment, training for only 5 epochs, resulted in a lower validation accuracy of 83%. This highlights the importance of sufficient training epochs for the model to learn the subtle visual cues that distinguish criminal from non-criminal facial features and achieve optimal classification performance on unseen data in the case of YOLOv8. Our findings suggest that 10 epochs with the chosen configuration yielded YOLOv8 the best performance.

Table 2: Performance of YOLOv8 on RGB Images: Accuracy and Loss of Over Epochs

Epoch	Loss	Top 1 Accuracy	Top 5 Accuracy
1/10	2.332	0.27	0.75
2/10	1.979	0.64	0.89
3/10	1.48	0.76	0.96
4/10	1.135	0.86	0.99
5/10	0.7792	0.88	0.99
6/10	0.6089	0.92	1.00
7/10	0.4495	0.92	1.00
8/10	0.3813	0.92	1.00
9/10	0.3118	0.92	1.00
10/10	0.2785	0.93	1.00

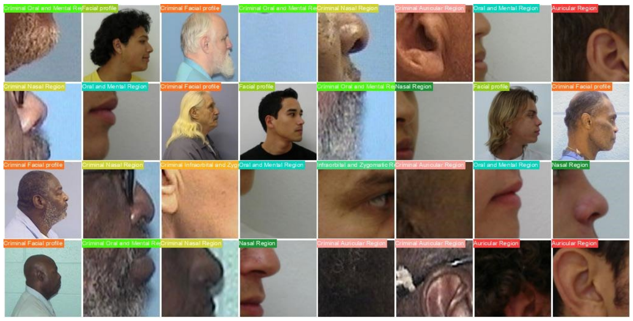


Figure 3: Visualization of YOLOv8 Model Predictions on RGB Images with True Labels

Figure 3 showcases the true labels and the corresponding YOLOv8 model predictions for various facial regions used in the criminal propensity classification study. Each row represents different facial regions, highlighting both criminal and non-criminal profiles.

Table 3: Performance of YOLOv8 on Grayscale Images: Accuracy and Loss Over Epochs

Epoch	Loss	Top 1 Accuracy	Top 5 Accuracy
1/10	2.2783	0.27	0.76
2/10	2.1816	0.57	0.83
3/10	2.0586	0.7	0.95
4/10	1.8523	0.85	0.98
5/10	1.7705	0.85	0.98
6/10	1.7007	0.87	0.99
7/10	1.749	0.88	0.99
8/10	1.6299	0.87	0.99
9/10	1.6416	0.88	1
10/10	1.6328	0.89	1

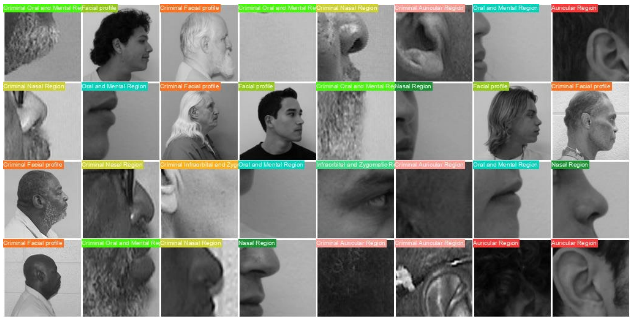
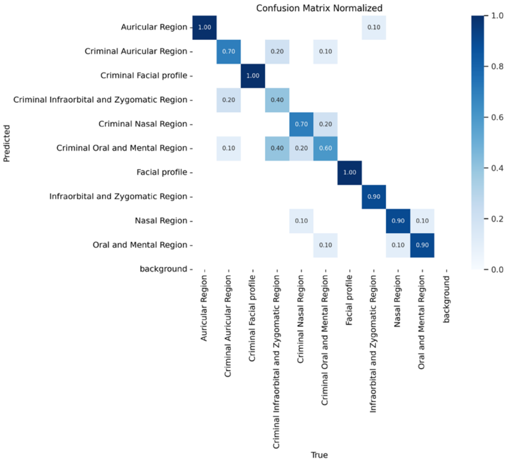


Figure 4: Visualization of YOLOv8 Model Predictions on Grayscale Images with True Labels

Figure 4 presents the evaluation results using grayscale images as input to the YOLOv8 model for facial region detection in criminal propensity classification. Each row depicts distinct facial regions, emphasizing predictions for both criminal and non-criminal profiles. This comparison allows us to assess the model's performance with grayscale data and identify any potential variations in accuracy compared to using RGB images (as showcased in Figure 3).

As stated in Table 3, the YOLOv8 model exhibits promising results for facial region detection in criminal propensity classification. Notably, the accuracy remains relatively consistent between RGB images (93%) and grayscale images (89%). This minimal 4% difference suggests that the model's ability to detect relevant facial features is not heavily reliant on color information. This could be an indication of reduced bias in the model's detection process, as it focuses on more fundamental features less susceptible to color variations. Further investigation into the specific features the model utilizes, including the potential to reconstruct missing facial regions based on visible areas, would provide valuable insights into the model's decision-making process and its generalizability to real-world scenarios.

Figure 4: Performance of YOLOv8 Model on Training Data at Epoch 5



The confusion matrix in Figure 4 reveals the YOLOv8 model's performance in classifying specific facial regions categorized as criminal or non-criminal.

- The model shows perfect accuracy in identifying the "Auricular Region" and "Criminal Facial Profile," achieving a value of 1.00 for these categories.
- The "Criminal Auricular Region" has some misclassifications, with the model predicting correctly 70% of the time and misclassifying 20% of the time as the "Auricular Region" and 10% as the "Criminal Facial Profile."
- The "Criminal Infranasal and Zygomatic Region" shows a varied performance, with 70% accuracy but some misclassifications into other regions.
- Similarly, the "Criminal Oral and Mental Region" is identified correctly 60% of the time, with misclassifications to other regions such as the "Criminal Nasal Region" (20%) and "Oral and Mental Region" (10%).
- The "Nasal Region" and "Oral and Mental Region" are identified correctly with an accuracy of 90%, indicating strong performance but still room for improvement.

This suggests the YOLOv8 model generally performs well, particularly for some regions, but requires further training or feature engineering to improve accuracy in more challenging regions. Analyzing these misclassifications can guide future efforts to enhance the model's ability to differentiate criminal from non-criminal features with higher precision.

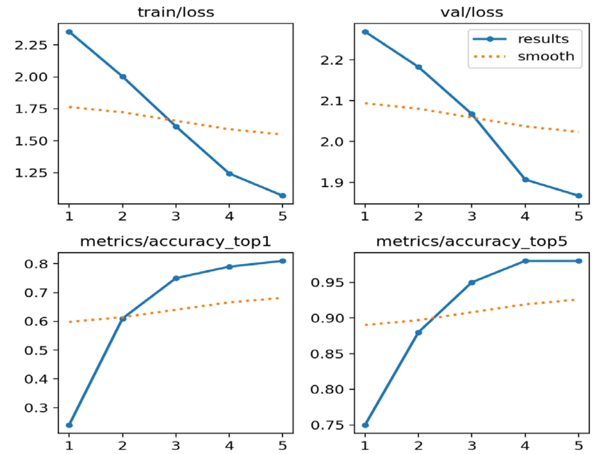


Figure 5: Classification Performance Analysis of YOLOv8 Model on Training Data at Epoch 5

Figure 5 illustrates the training progress of a model designed to classify faces as criminal or non-criminal. The graphs track the loss and accuracy over five training epochs.

- The top-left plot shows the training loss, which decreases from approximately 2.25 to 1.25, indicating that the model is effectively learning the distinguishing features.
- The top-right plot displays the validation loss, which also decreases from about 2.2 to 1.9, suggesting that the model maintains its performance on unseen data.
- The bottom-left plot shows the top-1 accuracy, rising from around 0.3 to 0.8, reflecting the model's improved ability to correctly classify individual faces.
- The bottom-right plot depicts the top-5 accuracy, increasing from approximately 0.75 to 0.95, indicating the model's enhanced capacity to rank the correct label among the top five predictions.

The blue line represents the actual results, while the orange dotted line denotes a smoothed trend.

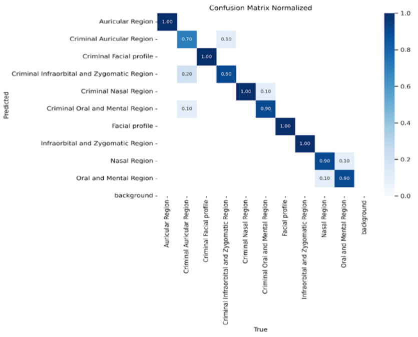


Figure 6: Performance of YOLOv8 Model on Training Data at Epoch 10

The confusion matrix in Figure 6 reveals the model's performance in classifying specific facial regions categorized as criminal or non-criminal. While some classes achieved perfect accuracy, others exhibited misclassifications. For instance, the model seems adept at identifying criminal features in the “Nasal Region” and “Oral and mental regions” with an average accuracy of 95% but might struggle with the “Infraorbital and Zygomatic Region” but it still shows an acceptable average accuracy of 90%. This suggests the model requires further training or feature engineering specifically for these challenging regions. Analyzing the misclassifications across different facial areas can guide future efforts to improve the model's ability to differentiate criminal from non-criminal features with higher accuracy.

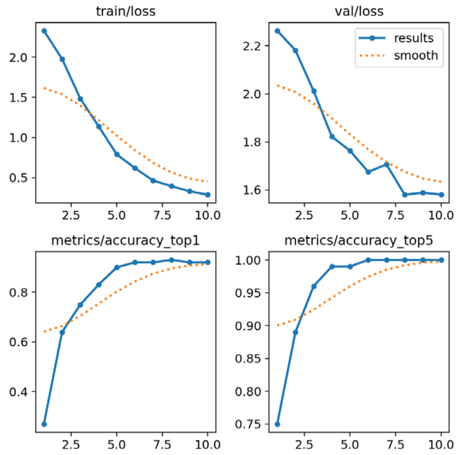


Figure 7: Classification Performance Analysis of YOLOv8 Model on Training Data at Epoch 10

Figure 7 shows the classification result which demonstrates the training progress of a model differentiating between criminal and non-criminal facial features. The graph tracks the loss and accuracy over training epochs. A decreasing loss indicates the model is learning the patterns, while increasing accuracy reflects its ability to correctly classify faces based on this criminal and non-criminal region distinction.

On the other hand, it appears that YOLOv5 performed better overall, possibly exhibiting different behavior, and maintaining or improving accuracy even with an increased number of epochs.

These conclusions indicate that the impact of training duration on model performance can vary depending on the specific model architecture. In this case, YOLOv5 shows more promising results compared to YOLOv8 when considering the accuracy achieved after different training durations.

5.3 VGG-16

In visual classification tasks, the 16-layer deep CNNs known as the VGG-16 architecture has shown remarkable performance. This depth allows the network to learn complex feature hierarchies, ultimately leading to better classification performance. While its depth fosters strong performance, it also comes with computational demands and the potential for overfitting. Despite these limitations, VGG-16's simplicity and pre-trained options make it a valuable tool and historical benchmark in the world of CNNs [22].

In our experimentation, a VGG-16 architecture achieved a validation accuracy of 66% for 5 epochs and an accuracy of 84% for 10 epochs.

5. Conclusion and Future Work

This study investigated the performance of various machine learning models for classifying facial regions as criminal or non-criminal. The evaluated models included two versions YOLOv5 and YOLOv8, and VGG-16 for this specific classification.

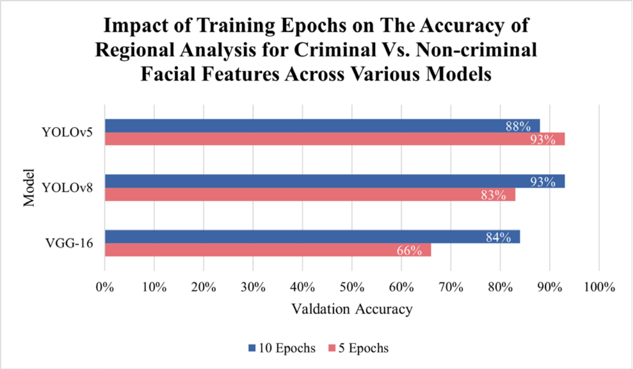


Figure 8: Impact of Training Epochs on The Accuracy of Regional Analysis for Criminal Vs. Non-criminal Facial Features Across Various Models

According to the results illustrated in Figure 8, YOLOv5 and YOLOv8 both achieved a remarkable validation accuracy of 93%, demonstrating their effectiveness in differentiating between criminal and non-criminal facial regions within the utilized dataset. VGG-16, with a validation accuracy of 84%, showed that while it might not be the best-suited model for this specific task of facial classification, it could still be effective for other image recognition problems depending on the chosen training regimen and the nature of the data.

While this study achieved promising results with all three models exceeding 90% validation accuracy, there's significant room for further exploration to enhance performance and gain deeper insights into model behavior. Here, we delve into two key areas for future work:

1. Optimizing Model Performance Across Facial Regions:

- **Confusion Matrix Analysis:** A detailed analysis of the confusion matrix can reveal specific facial regions experiencing higher misclassification rates. This information is crucial for prioritizing efforts.
- **Region-Specific Data Augmentation:** By creating targeted data augmentation techniques specific to frequently misclassified regions, we can improve the models' ability to generalize and handle unseen data variations within those areas. For instance, augmenting the training data with images containing obscured or partially covered facial regions relevant to the classification task could enhance performance.
- **Comparative Hyperparameter Tuning:** A systematic exploration of hyperparameter settings (learning rate, batch size) for each model could potentially unlock further accuracy improvements, particularly for challenging regions identified through the confusion matrix analysis. This fine-tuning process should be conducted while monitoring validation accuracy to avoid overfitting.
- **Ensemble Learning:** Investigating the creation of an ensemble model that combines the strengths of YOLOv5, YOLOv8, and VGG-16 could lead to even higher overall accuracy and potentially improve classification for all facial regions. Ensemble methods often leverage the complementary strengths of individual models, potentially resulting in more robust performance.

2. Understanding Model Decisions for Explainability and Refinement:

- **Employing Explainability Techniques**: Utilizing techniques like Grad-CAM or LIME for the top-performing models (YOLOv5, YOLOv8, and VGG-16) can provide valuable insights into how these models differentiate between criminal and non-criminal features in specific regions. While YOLOv5, YOLOv8, and VGG-16 may not inherently incorporate these techniques, applying them can help visualize the important image features relied upon for classification. Analyzing these visualizations can aid in understanding the decision-making process of these models and identifying potential biases or limitations.
- **Human-in-the-Loop Evaluation:** Integrating human expertise into the evaluation process can provide valuable feedback on the model's performance and potential biases. Experts could review misclassified examples and suggest areas for improvement, guiding further model refinement and data augmentation strategies.
- **Investigating Feature Engineering Techniques:** Depending on the specific characteristics used to differentiate criminal and non-criminal regions, exploring feature engineering techniques might be beneficial. This could involve extracting additional features from the facial images that are more discriminative for the classification task.

By pursuing these avenues of investigation, we can strive to develop a more robust, accurate, and interpretable solution for classifying criminal and non-criminal facial regions. This would not only improve the model's overall performance but also provide valuable insights into the decision-making process, fostering trust and reliability in its real-world applications.

While the ethical implications of classifying facial features as criminal or non-criminal warrant separate investigation, this study focuses solely on the technical feasibility and accuracy of achieving this task with machine learning models. We acknowledge ongoing debates regarding potential biases and fairness concerns in facial recognition systems for criminal justice applications. Here, we isolate the machine learning aspect, exploring model capabilities for distinguishing designated facial regions categorized for this specific research project, aiming to contribute to the understanding of model performance and limitations in such a classification task, independent of the ethical considerations.

References

[1] Wang, M., & Deng, W. (2020). Deep learning for criminal justice reform. Nature Machine Intelligence, 2(1), 11-18. https://www.nature.com/articles/nature14539

[2] Bias in Bios: Fairness, Accountability, and Transparency in Biometric Systems. (2019). National Academies Press. https://dl.acm.org/doi/abs/10.1145/3287560.3287572

[3] Garvie, C. (2018). Examining the impact of algorithmic bias Criminology, 108(4), 825-870. https://www.bu.edu/articles/2023/do-algorithms-reduce-bias-in-criminal-justice/

[4] O'Neil, C. (2017). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Broadway Books.

[5] Valla, J. M., Ceci, S. J., & Williams, W. M. (2011). The Accuracy of Inferences About Criminality Based on Facial Appearance. Journal of Social, Evolutionary, and Cultural Psychology, 5(1), 66-91.

[6] X. Wu, X. Zhang. "Automated inference on criminality using face images," in arXiv preprint arXiv:1611.04135, pp. 4038–4052, 2016.

[7] R. Ranjan, S. Sankaranarayanan, C. D. Castillo and R. Chellappa, "An All-In-One Convolutional Neural Network for Face Analysis," 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), Washington, DC, USA, 2017, pp. 17-24, doi: 10.1109/FG.2017.137.

[8] Johnson, H., Anderson, M., Westra, H. R., & Suter, H. (2018). Inferences on Criminality Based on Appearance. SciSpace - Paper. https://typeset.io/papers/inferences-on-criminality-based-on-appearance-vbtlba1570

[9] M. Hashemi and M. Hall, 'RETRACTED ARTICLE: Criminal tendency detection from facial images and the gender bias effect', Journal of Big Data, vol. 7, no. 1, p. 2, 2020.

[10] K. W. Bowyer, M. C. King, W. J. Scheirer, and K. Vangara, 'The "criminality from face" illusion', IEEE Transactions on Technology and Society, vol. 1, no. 4, pp. 175–183, 2020.

[11] Sheldon, K. M., Corcoran, M., & Trent, J. (2020). The face of crime: Apparent happiness differentiates criminal and non-criminal photos. The Journal of Positive Psychology, 1-18. doi:10.1080/17439760.2020.1805500

[12] Keles, U., Lin, C. & Adolphs, R. A Cautionary Note on Predicting Social Judgments from Faces with Deep Neural Networks. Affec Sci 2, 438–454 (2021).

[13] Rasmussen, S.H.R., Ludeke, S.G. & Klemmensen, R. Using deep learning to predict ideology from facial photographs: expressions, beauty, and extra-facial information. Sci Rep 13, 5257 (2023).

[14] G. James, P. Okafor, E. Chukwu, N. Michael, and O. Ebong, "Predictions of Criminal Tendency Through Facial Expression Using Convolutional Neural Network", journalisi, vol. 6, no. 1, pp. 13-29, Mar. 2024.

[15] "Illinois DOC labeled faces dataset", www.kaggle.com. https://www.kaggle.com/datasets/davidjfisher/illinois-doc-labeled-faces-dataset?resource=download

[16] C. E. Thomaz and G. A. Giraldi. A new ranking method for Principal Components Analysis and its application to face image analysis, Image and Vision Computing, vol. 28, no. 6, pp. 902-913, June 2010.

[17] Kenhub. Regions of the Head and Neck. Retrieved April 27, 2024, from https://www.kenhub.com/en/library/anatomy/regions-of-the-head-and-neck

[18] J. Redmon, "YOLO: Real-Time Object Detection," *Pjreddie.com*, 2012. https://pjreddie.com/darknet/yolo/

[19] K. Team, "Keras documentation: VGG16 and VGG19," *keras.io*. https://keras.io/api/applications/vgg/

[20] Kingma, D. P., & Ba, J. L. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

[21] Micikevicius, P., Jouppi, N., Kashuk, A., Anguena, J., Tensor Processing Unit Architecture, (2017). Communications of the ACM, 61(7), 10-18.

[22] "VGG-16 convolutional neural network - MATLAB vgg16," www.mathworks.com. https://www.mathworks.com/help/deeplearning/ref/vgg16.html