A Comparative Study of ResNet50 and YOLOv9 for Face Detection and Gender Classification

Aseil Nadhim Kadhim

Faculty of Artificial Intelligence, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia
n-20@graduate.utm.my (corresponding author)

Syahid Anuar

Faculty of Artificial Intelligence, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia
syahid.anuar@utm.my

Saiful Adli Bin Ismail

Faculty of Artificial Intelligence, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia
saifuladli@utm.my

ABSTRACT

Gender classification based on facial features plays a central role in numerous intelligent applications such as surveillance cameras, interactive advertising, and human-computer interaction. This study aimed to compare two deep models, YOLOv9 and ResNet50, on face detection and gender classification, focusing on accuracy and inference speed. YOLOv9 performed well in terms of speed, with an inference time of 332 ms per image and a processing speed of 3 fps, and had a precision of 86.8%, 86.1% of recall, and 86.54% of F1-score. These performance characteristics make YOLOv9 suitable for real-time applications with high-speed response demands, even with moderately low classification accuracy. Conversely, ResNet50 was applied directly to gender classification after data preparation on images and had high classification accuracy, with a precision of 93.6%, 92% of recall, and 92.79% of F1-score. Its inference time was slower at 446.33 ms per image, with a 2.24 fps processing speed and a long training time of 9 hours and 18 minutes. These results show that YOLOv9 has high performance within a time scope of face detection, with reference to detecting enough faces within short timeframes with a limited number of computational resources, whereas ResNet50 has better classification accuracy. Depending on particular use case scenario demands, one corresponding model with a preferred feature can be selected: YOLOv9, if high-speed response is a concern during real-time applications, and ResNet50, if high classification accuracy is a concern.

Keywords-object detection; face detection; gender classification; YOLO;ResNet50

I. INTRODUCTION

State-of-the-art computer vision techniques have enabled intelligent face recognition and surveillance applications to spread rapidly, with CNN-based models at the forefront of enhanced object recognition performance [1]. Face detection and gender classification have been central activities in this context, but high accuracy with real-time performance remains a research dilemma. YOLO has been favored for real-time object detection, with a series of enhanced variations proposed with improved performance under different settings [2-6]. ResNet50 has shown high image classification performance due to deep residual learning [7-8], showing success even with applications such as face mask detection [9-10]. Numerous studies have investigated gender classification as a faceness

function with a variety of deep models proposed to address this task [11]. Face detection models can fundamentally be categorized into two groups: single-stage detectors, such as YOLO, which are common due to the high-speed inference for real-time applications, as well as their ability to compromise between precision and computational burden [12-13]. Other studies have shown that these models still require additional improvements in their detection performance [14].

In contrast, two-stage detectors, such as Faster R-CNN, tend to gain improved precision, particularly in applications that require reliable output, such as gender and age classification [15]. ResNet50 has been presented with particular aptness at extracting high-semantic visual information from face images, and consequently, it emerges as a viable candidate

for face-based classification applications [16-22]. Although previous studies have investigated gender classification with alternative biometric cues, such as speech or body gesture [23-24], face-based classification still stands out as by far the dominant and reliable paradigm applied to intelligent visual systems. Some studies have proposed fusions between ResNet50 and face detection systems that can facilitate classification performance against images and video streams [25-26]. Other studies have also investigated multimodal fusions of approaches devised to enhance classification robustness [12, 27, 28]. Although there is extensive research within this domain, there is a shortage of systematic comparative research to evaluate and compare the performance of detection models, such as YOLOv9, and classification models, such as ResNet50, within a common framework. This study aimed to fill this gap by comparatively experimenting and evaluating both models with respect to classification precision, inference speed, and deployment feasibility.

II. DATA COLLECTION

A dataset of 140 images of university student faces was collected under varying natural illumination conditions, from direct sunlight down to partial cloud cover, at varying hours of the day and viewing angles. The dataset includes only two classes (male and female), comprising a total of 140 images (72 male and 68 female). This diversity in image acquisition was intentionally introduced to enhance the dataset's representativeness and to promote better generalization during the training phase.
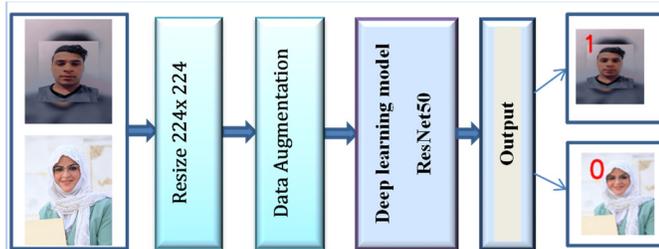

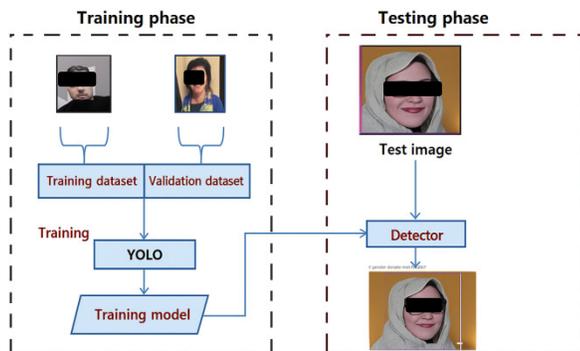Fig. 1.   A diagram showing the ResNet50 model working method.

Fig. 2.   A diagram showing the YOLOv9 model's working method.

III. METHODOLOGY

A. Image Preprocessing and Dataset Partitioning

The images were subjected to a preprocessing phase, in which their dimensions were adjusted to meet the input requirements of each model. As shown in Figure 1, the images were resized to 224×224 pixels for the ResNet50 model and to 640×640 pixels for the YOLOv9 model. Subsequently, the dataset was split into training (80%) and validation (20%) subsets to ensure an objective evaluation of model performance following the training phase, as shown in Figure 2.

B. Data Augmentation

A total of 140 original facial images were initially collected and preprocessed to ensure uniformity, including resizing to 640×640 pixels, grayscale conversion, and pixel normalization between 0 and 1. To enrich the dataset and improve the generalization of the models, a comprehensive data augmentation strategy was applied using the Roboflow platform. Each image was subjected to six augmentation techniques: rotation, translation, cropping, brightness adjustment, contrast variation, and spectral noise injection. This augmentation process expanded the dataset from 140 to 980 images. The resulting enhanced dataset was divided into 784 training images (80%) and 196 validation images (20%). Additionally, a separate set of 361 testing images, not seen during training, was reserved exclusively to evaluate the final performance of both ResNet50 and YOLOv9 models. Only the input size parameter was adjusted to fit each model's requirements, ensuring consistency and fairness during the comparative analysis and allowing for objective performance measurement.

C. ResNet50 Network Architecture

This study utilized the ResNet50 algorithm with its basic architecture as a 50-layer deep neural network based on Residual Learning [29] to train extremely deep networks without experiencing performance degradation, as well as overcoming the vanishing gradient problem. There are three basic phases in this model architecture: an input layer that handles images resized to 224×224 pixels, typical residual blocks, and an output layer with resulting classification probabilities [30].

Although the model's backbone was preserved without modification, a minor customization was applied to the classifier head, which included replacing the final Fully Connected (FC) layer, originally designed for 1,000-class classification (ImageNet), with a layer tailored for binary classification (male/female) and using the sigmoid activation function instead of softmax, to suit the binary nature of the task. This modification does not alter the core architecture of the model but is a necessary adaptation to align the model with the specific data and the classification objective. This architecture demonstrated high efficiency in extracting complex features from facial images and contributed to improved classification accuracy, particularly when applied to a relatively small dataset.

D. Architecture of YOLOv9

YOLOv9 is an advanced version of the YOLO (You Only Look Once) object detection family, offering a powerful balance between high detection accuracy and real-time processing speed, making it well-suited for practical applications and deployment on resource-constrained systems [32-34]. The architecture of YOLOv9 comprises four primary components:

- Backbone: Responsible for extracting deep visual features from the input images using convolutional layers and RepNCSPELAN modules, while gradually reducing the spatial resolution to retain essential semantic information.

- Neck: Facilitates multi-scale feature aggregation through upsampling and concatenation operations, enhancing spatial representation by integrating contextual cues across different levels using the SPPELAN module.

- Head: Consists of three parallel detection layers operating at different spatial resolutions (80×80, 40×40, 20×20), allowing the network to detect objects of varying sizes accurately.

- Auxiliary path: A complementary pathway that includes CBFuse and CBLinear modules, designed to improve the overall flow of information and optimize model efficiency during the training process.

E. Training Platform and Parameter Settings

Both the ResNet50 and YOLOv9 models were trained based on their original architectures using the Google Colab platform, using a Tesla T4 GPU to ensure high computational performance. The training was carried out over 300 epochs, with a batch size of 16 samples per iteration. The initial learning rate was set at 0.001 and dynamically adjusted during training to promote model stability and facilitate better convergence for both architectures. For the ResNet50 model, the input images were resized to 224×224 pixels, and the backbone layers were frozen to prevent weight updates. Only the classifier head was modified by replacing the original FC layer with a binary classification layer (male/female) with a sigmoid activation function. This adaptation retained the basic architecture unchanged but fine-tuned the model, accommodating the task's highly specific nature. The input images of the YOLOv9 model were resized to 640×640 pixels, and the entire original architecture remained intact with the Backbone (RepNCSPELAN), Neck (SPPELAN), Head, and the Auxiliary Path (CBFuse and CBLinear). In addition, a model made use of Programmable Gradient Information (PGI) and Generalized ELAN (GELAN) modules aimed at fostering enhanced gradient flows, reduced computational complexity, and increased inference speed and accuracy.

F. Performance Standards

Performance evaluation metrics were utilized as indicators of the gender classification models deployed (ResNet50 and YOLOv9), namely Accuracy, Precision, Recall, and F1-score. These performance metrics are based on confusion matrices, a fundamental assessment tool to characterize the relationship between the model's prediction results and the actual ground truths [36]. The confusion matrix consists of four key components:

- True Positive (TP): The case in which the model correctly predicts the presence of the positive class (e.g., the model predicts "male" and the actual label is "male").

- True Negative (TN): The case in which the model correctly predicts the presence of the negative class (e.g., the model predicts "female" and the actual label is "female").

- False Positive (FP): The case in which the model incorrectly predicts the positive class (e.g., the model predicts "male" but the actual label is "female").

- False Negative (FN): The case in which the model incorrectly predicts the negative class (e.g., the model predicts "female" but the actual label is "male").

Confusion matrices are commonly deployed to calculate performance metrics such as precision, precision, recall, and F1 score, which are important to measure the performance of a model and determine errors [35, 36].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (1)$$

$$Precision: P = \frac{TP}{TP+FP} \qquad (2)$$

$$Recall: R = \frac{TP}{TP+FN} \qquad (3)$$

$$F1 \ score: F1 = 2 \times \frac{P \times R}{P+R} \qquad (4)$$

Finally, inference time was considered a critical performance indicator, defined as the time taken to generate an output for a single input image. Measured in ms, inference time serves as a key metric for evaluating the model's suitability for real-time applications.

IV. RESULTS AND DISCUSSION

The performance of the ResNet50 and YOLOv9 models was evaluated based on gender classification using a cumulative test dataset consisting of 361 images, after data augmentation and preprocessing. The results indicate that both models had good results with considerable differences in their accuracy, inference speed, and training time.

The ResNet50 model achieved 93.6% classification precision, 92% recall, and 92.79% F1-score, indicating its excellent performance at effectively classifying male and female faces based on face features. Its good performance could be attributed to its deep architecture as well as fine-grained feature extraction properties. Its last training loss was 0.07 at 300 epochs, affirming steady convergence as well as efficient learning with time. However, the model had relatively slow inference with 446.33 ms per-image processing time and 2.24 fps frame rate. Its training took approximately 9 hours and 18 minutes, which is a disadvantage for its use on real-time systems or on low-resource devices.

Conversely, the YOLOv9 model had superb performance in terms of speed and execution efficacy. It had a 332 ms per-image processing time as well as 3.00 fps frame rate with a comparatively shorter training time of approximately 1 hour

and 45 minutes. Although its classification accuracy was lower compared to ResNet50's (86.8%), a steady trend of continuous learning was achieved after approximately 150 epochs. Figure 5 shows a gender classification accuracy comparison between ResNet50 and YOLOv9, with a visible performance difference in regard to ResNet50, as demonstrated in Table I.

This performance advantage is largely due to architectural improvements such as the integration of PGI (Programmable Gradient Information) and GELAN (Generalized ELAN), which contribute to faster loss reduction and enhanced detection of overlapping and small-scale objects. Figure 6 presents a comparison of the training loss curves for both models, showing YOLOv9's rapid loss reduction during early training stages, whereas ResNet50 demonstrates gradual and steady convergence. Table I shows the performance comparison between YOLOv9 and ResNet50 algorithms.

Regarding the confusion matrices, ResNet50 slightly outperformed YOLOv9 in terms of TP classifications, registering 170 TP compared to 157 for YOLOv9, while also showing fewer false positives and false negatives. This reinforces its strength in achieving precise classifications. On the other hand, the superior speed and computational efficiency of YOLOv9 make it more suitable for real-time or embedded applications. Figure 7 and Table II present the confusion matrix analysis for YOLOv9 and ResNet50.
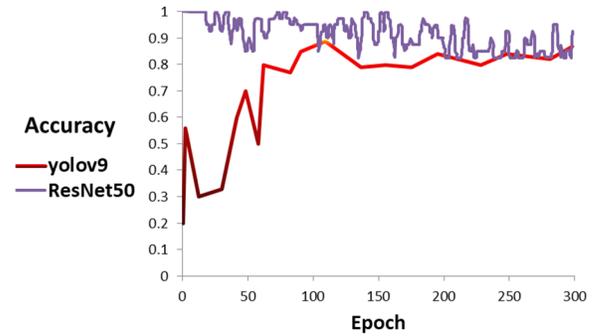

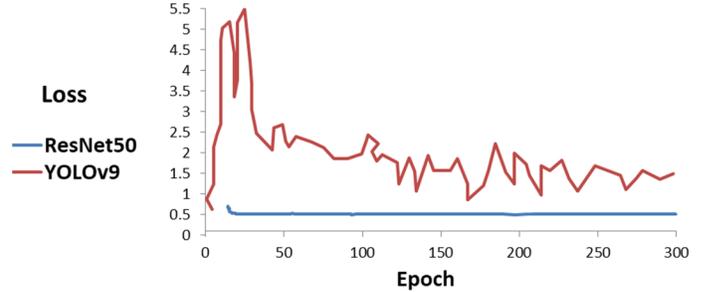Fig. 3.   Accuracy comparison for ResNet50 and YOLOv9 for gender detection.

Fig. 4.   Comparison of training loss for ResNet50 and YOLOv9 gender detection algorithms.

TABLE I.   PERFORMANCE COMPARISON OF RESNET50 AND YOLOV9 FOR GENDER DETECTION

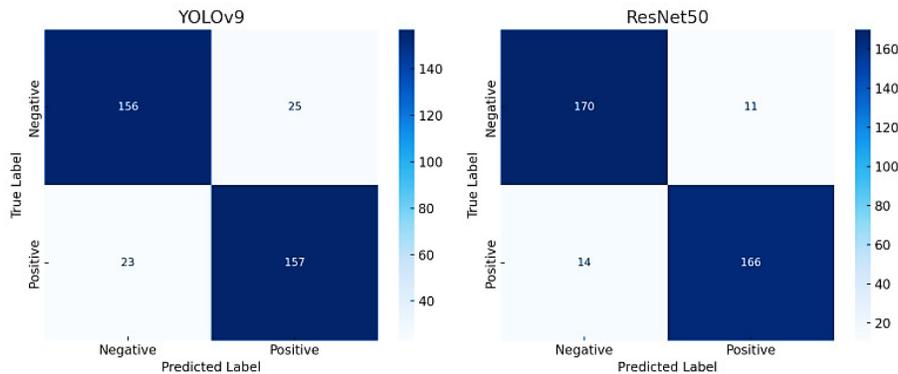| Model	Precision (%)	Recall (%)	F1 score (%)	Inference time per image (ms)	FPS	Training time	Parameters (Millions)
YOLOv9	86.8	86.1	86.54	332	3.00	1:45:00 s	65 M
ResNet50	93.60	92	92.79	446.33	2.24	9:18:41 s	26 M

Fig. 5.   Confusion matrices of YOLOv9 and ResNet50.

TABLE II.   CONFUSION MATRIX ANALYSIS FOR YOLOV9 AND RESNET50 MODELS

Model	True Positives (TP)	False Negatives (FN)	False Positives (FP)	True Negatives (TN)
YOLOv9	156	25	23	157
ResNet50	170	11	14	166

V.   CONCLUSION

This study presented a comparative evaluation between two deep learning models, ResNet50 and YOLOv9, for the task of binary gender classification using facial images. The two models were trained and tested on similar data, subjected to common data augmentation strategies, to ensure a fair comparison. The ResNet50 model achieved a classification

accuracy of 93.60%, demonstrating its remarkable ability to extract deep and discriminative features from face images. However, due to its deep architecture and high parameters, ResNet50 has a larger computational overhead and slower rates of inference. Against these drawbacks, it is less suitable when applied to real-time applications or running on limited-resource machines. In contrast, with a relatively lower accuracy of 86.8%, YOLOv9 surpassed ResNet50 with a much quicker processing speed and computational efficiency. With an inference time of just 332 ms per image, YOLOv9 brings advanced architectural upgrades such as PGI and GELAN into play, making it a suitable solution appropriate for real-time gender classification applications with high turnaround demands.

This comparison clearly highlights the significant difference in the number of parameters between the two models under study. The number of trainable parameters is a critical indicator of both model complexity and computational efficiency. The YOLOv9 model, configured with a CSP+ELAN backbone and PANet neck, consists of approximately 64.9 million parameters, reflecting its capacity to capture rich, multi-scale features, although it requires greater computational resources. In contrast, the ResNet50 model contains approximately 25.6 million parameters, representing a well-balanced architecture that offers high classification accuracy with relatively lower computational cost. This disparity in parameter counts reinforces the study's core assertion that model selection should consider not only accuracy and inference speed, but also architectural complexity and resource constraints in practical deployment environments.

VI.   FUTURE WORK AND RECOMMENDATIONS

These results show that there exists a clear imperative of developing hybrid models whose high accuracy of ResNet50 can be combined with YOLOv9's speed, such that both architectures' strengths can be accommodated within a balanced framework. This can be achieved by incorporating lightweight modules within the framework of a model to reduce computational complexity, as well as by incorporating model compression to maintain performance efficacy and classification accuracy.

REFERENCES

[1] X. Ming, X. Nanfeng, Z. Mengjun, and Y. Qunyong, "Optimized Convolutional Neural Network-Based Object Recognition for Humanoid Robot," *Journal of Robotics and Automation*, vol. 4, no. 1, Feb. 2020, https://doi.org/10.36959/673/363.

[2] K. Jiang *et al.*, "An Attention Mechanism-Improved YOLOv7 Object Detection Algorithm for Hemp Duck Count Estimation," *Agriculture*, vol. 12, no. 10, Oct. 2022, Art. no. 1659, https://doi.org/10.3390/agriculture12101659.

[3] Y. Xia, M. Nguyen, and W. Q. Yan, "A Real-Time Kiwifruit Detection Based on Improved YOLOv7," in *Image and Vision Computing*, vol. 13836, W. Q. Yan, M. Nguyen, and M. Stommel, Eds. Springer Nature Switzerland, 2023, pp. 48–61.

[4] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 6517–6525, https://doi.org/10.1109/CVPR.2017.690.

[5] U. Nepal and H. Eslamiat, "Comparing YOLOv3, YOLOv4 and YOLOv5 for Autonomous Landing Spot Detection in Faulty UAVs,"

[6] Y. Wang and L. Pan, "YOLOV5s-Face face detection algorithm," in *2022 China Automation Congress (CAC)*, Xiamen, China, Nov. 2022, pp. 1107–1112, https://doi.org/10.1109/CAC57257.2022.10054674.

[7] A. Dhillon and G. K. Verma, "Convolutional neural network: a review of models, methodologies and applications to object detection," *Progress in Artificial Intelligence*, vol. 9, no. 2, pp. 85–112, Jun. 2020, https://doi.org/10.1007/s13748-019-00203-0.

[8] H. Jiang and E. Learned-Miller, "Face Detection with the Faster R-CNN," in *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, Washington, DC, DC, USA, May 2017, pp. 650–657, https://doi.org/10.1109/FG.2017.82.

[9] S. D. Meena, C. S. Siri, P. S. Lakshmi, N. S. Doondı, and J. Sheela, "Real time DNN-based Face Mask Detection System using MobileNetV2 and ResNet50," in *2023 International Conference on Inventive Computation Technologies (ICICT)*, Lalitpur, Nepal, Apr. 2023, pp. 1007–1015, https://doi.org/10.1109/ICICT57646.2023.10133992.

[10] B. Mandal, A. Okeukwu, and Y. Theis, "Masked Face Recognition using ResNet-50." arXiv, Apr. 19, 2021, https://doi.org/10.48550/arXiv.2104.08997.

[11] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, "DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection." arXiv, Jun. 18, 2020, https://doi.org/10.48550/arXiv.2001.00179.

[12] Z. Yu, H. Huang, W. Chen, Y. Su, Y. Liu, and X. Wang, "YOLO-FaceV2: A Scale and Occlusion Aware Face Detector." arXiv, Aug. 04, 2022, https://doi.org/10.48550/arXiv.2208.02019.

[13] W. Chen, H. Huang, S. Peng, C. Zhou, and C. Zhang, "YOLO-face: a real-time face detector," *The Visual Computer*, vol. 37, no. 4, pp. 805–813, Apr. 2021, https://doi.org/10.1007/s00371-020-01831-7.

[14] T. W. Shen, D. Wang, K. W. K. Cheung, M. C. Chan, K. H. Chiu, and Y. K. Li, "A Real-Time Single-Shot Multi-Face Detection, Landmark Localization, and Gender Classification," in *2021 3rd International Conference on Image Processing and Machine Vision (IPMV)*, Hong Kong, China, May 2021, pp. 1–4, https://doi.org/10.1145/3469951.3469952.

[15] D. K. Srivastava, E. Gupta, S. Shrivastav, and R. Sharma, "Detection of Age and Gender from Facial Images Using CNN," in *Proceedings of 3rd International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications*, vol. 540, V. K. Gunjan and J. M. Zurada, Eds. Springer Nature Singapore, 2023, pp. 481–491.

[16] H. Guan, H. Li, R. Li, M. Qi, and V. Velmurugan, "Face Detection System Based on Deep Learning," in *Proceedings of the 2nd International Conference on Cognitive Based Information Processing and Applications (CIPA 2022)*, 2023, pp. 525–531, https://doi.org/10.1007/978-981-19-9376-3_59.

[17] M. DhivyaShree, K. R. Sarumathi, and R. S. V. Durai, "An Ensemble Model for Face Mask Detection Using Faster RCNN with ResNet50," in *Artificial Intelligence and Speech Technology*, vol. 1546, A. Dev, S. S. Agrawal, and A. Sharma, Eds. Springer International Publishing, 2022, pp. 593–603.

[18] B. Li and D. Lima, "Facial expression recognition via ResNet-50," *International Journal of Cognitive Computing in Engineering*, vol. 2, pp. 57–64, Jun. 2021, https://doi.org/10.1016/j.ijcce.2021.02.002.

[19] Dr. S. Gothane, "A Practice for Object Detection Using YOLO Algorithm," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, pp. 268–272, Apr. 2021, https://doi.org/10.32628/CSEIT217249.

[20] D. Qi, W. Tan, Q. Yao, and J. Liu, "YOLO5Face: Why Reinventing a Face Detector." arXiv, Jan. 27, 2022, https://doi.org/10.48550/arXiv.2105.12931.

[21] S. Ennaama, H. Silkan, A. Bentajer, and A. Tahiri, "Enhanced Real-Time Object Detection using YOLOv7 and MobileNetv3," *Engineering, Technology & Applied Science Research*, vol. 15, no. 1, pp. 19181–19187, Feb. 2025, https://doi.org/10.48084/etasr.8777.

[22] D. Fajalia, R. Satkar, A. Jejurkar, and S. Dhanake, "Video Based Face Mask Detection and Face Recognition using CNN, YOLO and Google Facenet," *International Research Journal of Modernization in Engineering Technology and Science*, vol. 2, no. 6, pp. 259–266, May 2023, https://doi.org/10.56726/IRJMETS38770.

[23] Nandita and B. Jain, "A Comprehensive Review of Machine Learning Techniques for Voice-Based Gender Recognition," in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Delhi, India, Jul. 2023, pp. 1–7, https://doi.org/10.1109/ICCCNT56998.2023.10306939.

[24] R. Bhatia and N. P. Singh, "Gender Recognition by Voice Using Machine Learning," in *Advanced Network Technologies and Intelligent Computing*, 2022, pp. 307–318, https://doi.org/10.1007/978-3-030-96040-7_25.

[25] K. Jain, M. Chawla, A. Gadhwal, R. Jain, and P. Nagrath, "Age and Gender Prediction Using Convolutional Neural Network," in *Proceedings of First International Conference on Computing, Communications, and Cyber-Security (IC4S 2019)*, vol. 121, P. K. Singh, W. Pawłowski, S. Tanwar, N. Kumar, J. J. P. C. Rodrigues, and M. S. Obaidat, Eds. Springer Singapore, 2020, pp. 247–259.

[26] M. J. Awan, A. Raza, A. Yasin, H. M. F. Shehzad, and I. Butt, "The Customized Convolutional Neural Network of Face Emotion Expression Classification," *Annals of R.S.C.B.*, vol. 25, no. 6, pp. 5296–5304, 2021.

[27] S. Mittal, K. Thakral, P. Majumdar, M. Vatsa, and R. Singh, "Are Face Detection Models Biased?," in *2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG)*, Waikoloa Beach, HI, USA, Jan. 2023, pp. 1–7, https://doi.org/10.1109/FG57933.2023.10042564.

[28] R. Li and J. Yang, "Improved YOLOv2 Object Detection Model," in *2018 6th International Conference on Multimedia Computing and Systems (ICMCS)*, Rabat, Morocco, May 2018, pp. 1–6, https://doi.org/10.1109/ICMCS.2018.8525895.

[29] E. Suherman, B. Rahman, D. Hindarto, and H. Santoso, "Implementation of ResNet-50 on End-to-End Object Detection (DETR) on Objects," *SinkrOn*, vol. 8, no. 2, pp. 1085–1096, Apr. 2023, https://doi.org/10.33395/sinkron.v8i2.12378.

[30] T. N. V. S. Praveen, D. Sivathmika, G. Jahnavi, and J. Bolledu, "An In-depth Exploration of ResNet-50 for Complex Emotion Recognition to Unraveling Emotional States," in *2023 International Conference on Advancement in Computation & Computer Technologies (InCACCT)*, Gharuan, India, May 2023, pp. 1–5, https://doi.org/10.1109/InCACCT57535.2023.10141774.

[31] S. Shivadekar, B. Kataria, S. Hundekari, K. Wanjale, V. P. Balpande, and R. Suryawanshi, "Deep Learning Based Image Classification of Lungs Radiography for Detecting COVID-19 using a Deep CNN and ResNet 50," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 1s, pp. 241–250, Jan. 2023.

[32] G. Boesch, "YOLOv9: Advancements in Real-time Object Detection," *viso.ai*, Sep. 27, 2024. https://viso.ai/computer-vision/yolov9/.

[33] C. Y. Wang, I. H. Yeh, and H. Y. M. Liao, "YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information." arXiv, Feb. 29, 2024, https://doi.org/10.48550/arXiv.2402.13616.

[34] S. Chaudhuri *et al.*, "Infrared Thermography of Turbulence Patterns of Operational Wind Turbine Rotor Blades Supported With High-Resolution Photography: KI-VISIR Dataset," *Wind Energy*, vol. 28, no. 1, Jan. 2025, Art. no. e2958, https://doi.org/10.1002/we.2958.

[35] X. Xue *et al.*, "Design and Analysis of a Deep Learning Ensemble Framework Model for the Detection of COVID-19 and Pneumonia Using Large-Scale CT Scan and X-ray Image Datasets," *Bioengineering*, vol. 10, no. 3, Mar. 2023, Art. no. 363, https://doi.org/10.3390/bioengineering10030363.

[36] E. Yildirim, "ResNet-based Gender Recognition on Hand Images," *Engineering, Technology & Applied Science Research*, vol. 14, no. 6, pp. 17969–17972, Dec. 2024, https://doi.org/10.48084/etasr.8922.