ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.02 No.04 (2024)

Gender Detection from A Face Mask by Using Deep Learning Algorithm

Mehtab Kinza
Department of Computer Science and Information Technology, Superior University, Lahore, Pakistan
Muhammad Akmal
Department of Software Engineering, Minhaj University, Lahore, Pakistan
Muhammad Ahsan Ashfaq
Department of Computing, Ripha International University Faisalabad Campus, Pakistan
Dr. Arfan Jaffar
Dean, Department of Computer Science and Information Technology, Superior University, Lahore, Pakistan
Dr. Sohail Masood
Associate Professor, Department of Computer Science and Information Technology, Superior University, Lahore, Pakistan

Abstract
Accurately identifying gender from masked facial images is a critical challenge, especially during widespread mask usage in pandemics. This study introduces a novel deep learning approach that combines convolutional architectures with transfer learning to achieve robust gender classification in masked faces. Using a curated dataset of 89,131 masked facial images for training and 20,714 for evaluation, the proposed method outperformed 11 traditional models, achieving 99% accuracy and a 0.17% error rate with minimal computational resources. By leveraging transfer learning, the framework achieves a balance between efficiency and accuracy, making it suitable for real-time applications. Comparative evaluations demonstrate the superiority of the proposed model over traditional techniques like MobileNetV2, ResNet101, and DenseNet169. This research provides insights into enhancing the accuracy of facial recognition systems and improving biometric classification in constrained environments.
Keywords: *Masked face recognition, Gender classification, Deep learning, Transfer learning, Convolutional Neural Networks (CNNs)*

INTRODUCTION
The ability to classify gender accurately from facial images is a crucial aspect of many biometric applications, including security, surveillance, and personalized systems. However, the widespread use of face masks during global pandemics presents a unique challenge for conventional facial recognition systems, as critical facial features are obscured. This study addresses this challenge by leveraging advanced deep learning techniques to enhance gender classification accuracy in masked facial images.
Facial recognition systems equipped with gender detection capabilities offer several advantages. They can improve the overall precision of biometric identification, assist in demographic data collection, and provide valuable insights for human-computer interaction (HCI) applications. However, traditional models often falter when faced with occluded facial features, necessitating innovative solutions.
To tackle this issue, the present research explores convolutional neural networks (CNNs) combined with transfer learning for feature extraction and classification. Unlike earlier approaches that rely on handcrafted features or less robust machine learning models, our proposed method dynamically learns from large-scale datasets, ensuring scalability and adaptability to real-world conditions. The model focuses on retained features, such as the eyes, forehead, and visible contours, enabling effective classification even with partial facial visibility.
This study has four key objectives:

- To develop a robust algorithm capable of real-time gender classification from masked facial images.
- To evaluate the impact of mask-induced occlusions on classification accuracy and identify techniques to mitigate these effects.
- To compare the proposed framework with existing deep learning methods across diverse datasets.

- To propose a scalable solution for biometric systems applicable to domains such as healthcare, security, and public monitoring.

By addressing these objectives, the research contributes to the ongoing efforts to enhance biometric systems, particularly under constrained scenarios like pandemics. The proposed model not only surpasses existing techniques in terms of accuracy but also demonstrates computational efficiency, making it suitable for deployment in real-time systems.

RELATED WORK

The study explored the ethical and societal consequences of utilizing AI within responses to pandemic including privacy problems and biases in methods. The poll highlighted promising AI technologies, including machine learning, natural language processing, including computer vision, to help frontline workers along with decision makers manage the epidemic. The study's findings can help academics, governments, and health care providers develop better AI solutions to tackle COVID-19 [7].

Gogate et al. describes a system that can perform real-time emotion Analysis of facial expression may be used to recognize and classify gender. The algorithm accurately recognized emotions and classified genders based a dataset comprising video clips. This research suggests that the suggested approach has possible uses within human-computer interaction, automation, and gaming. This study was published across 2020 (ICSIDEMPC) held Islamabad, Pakistan [12]. Mustafa and Meehan developed a system enabling real-time gender categorization and age predictions utilizing CNN with ResNet architectures.

The suggested technique extracts feature from face images using LDP, textures descriptor that describes local directions information. The authors employed Support Vector Machine (SVM) for classifying pictures either male or female.This suggested approach outperforms current gender categorization algorithms across three datasets. These researchers found that LDP provides an excellent feature descriptor with gender classification along with might be used in various applications including face recognition along with age estimation [19].

It can be focused on the topic of gender classification using facial images. The studies used various techniques including deep learning models, local feature extraction, and data mining algorithms to classify gender based on images of the face. Several studies explored the use of convolutional neural networks (CNNs) and residual neural networks (ResNets) for gender classification, achieving high levels of accuracy. Other studies utilized feature extraction techniques, such as Local Binary Patterns (LBPs) and Multi-Level Local Phase Quantization (MLLPQ), to extract information from facial images and classify gender using machine learning algorithms. Some papers also explored the combination of age and gender classification, with promising results. Overall, the studies showed that gender classification from facial images is a well-studied problem with many effective solutions. These techniques could have potential applications in fields such as security and marketing, where gender classification could be used for personalized advertising or security screening.

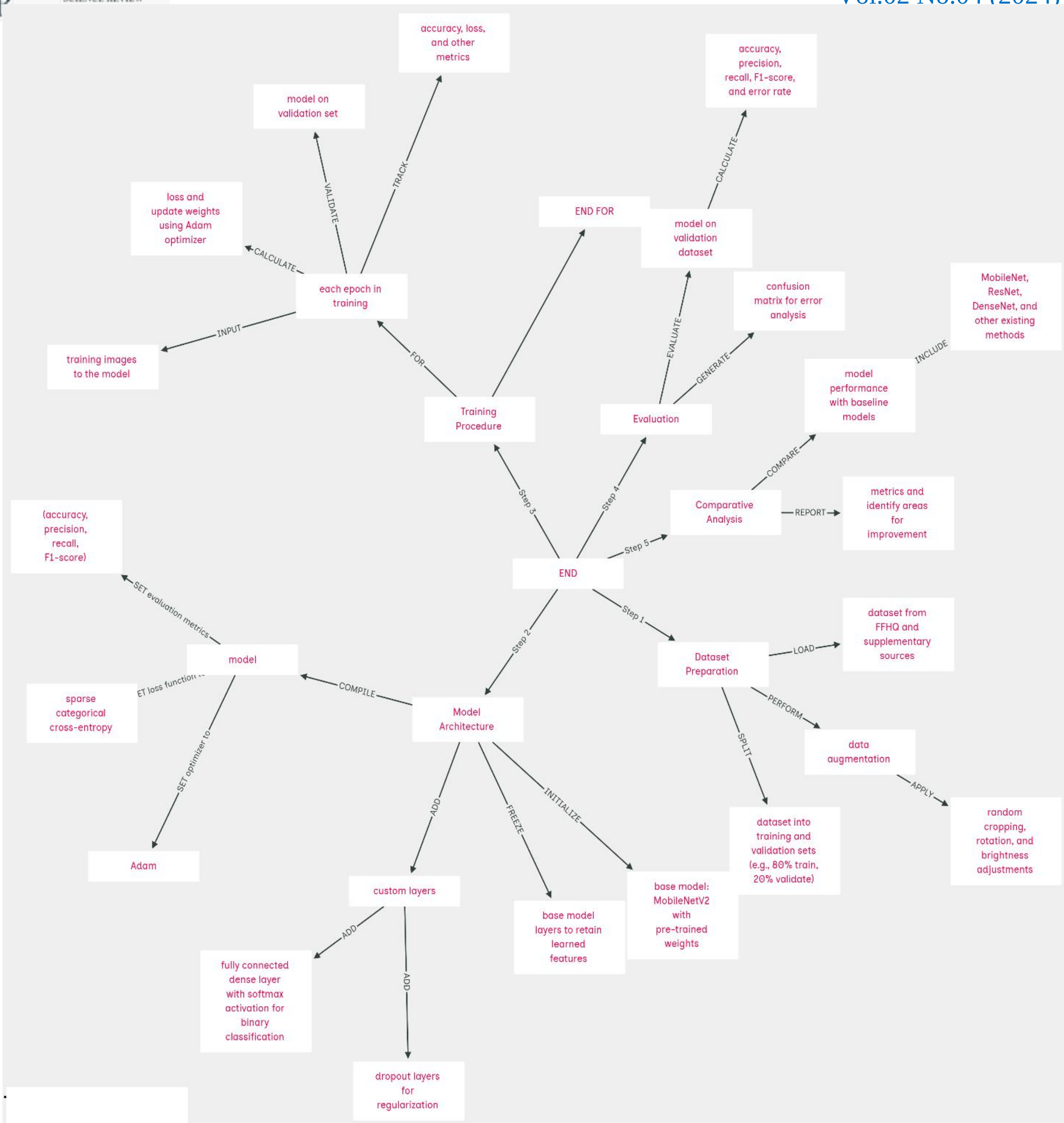
METHODOLOGY

This work utilized supervised learning approaches, which are widely used throughout machine learning, for achieving optimal performance outcomes. Supervised training algorithms learn through example. This phrase supervised learning is derived from the notion of preparing an information set. The training collection includes both input images and their corresponding outputs. Throughout the training phase, the supervised algorithm for learning will look for patterns within the data that are related to the specified outputs.

Dataset

This study utilized a gender-classified collection with masked faces. Flickr-Faces-HQ (FFHQ) represents an exceptional image collection featuring individual faces that was established as a standard for GAN. It contains 89,443 images for training, separated into two categories (male along with female). Furthermore, a total of 20,714 images for evaluation. This dataset has been established using the FFHQ the database, and the remaining images are reserved through Google scraping.

Experiment



1. Random modified input images were augmented for model robustness.
2. Feature detection from images at abstract level to increase complexity and model gets deeper.
3. Reduce spatial dimensions in max pooling to obtain prominent feature.
4. Final prediction is made at fully connected.
5. Randomly dropout neurons to reduce overfitting during training.
6. Final prediction is made from classes for correct classification.

Model training includes following steps

1. Model is compiled after configuration before training
2. Adam is selected as optimizer that defines how weights updates in model using loss function
3. Adam adjusted learning rate while training

4. Adam combines optimizers and efficient to large dataset
5. Loss function decided predictions and how they match with actual labels
6. Model tried to minimize loss
7. Sparse-categorical-cross-entropy is applied for integer label classification
8. Raw logits are applied to apply Softmax function for loss calculation
9. Model performance evaluation metrics such as accuracy applied
10. Model summary calculated for layers, output shapes, parameters and total parameters

Testing

i. Dataset Creation: Our chose images with varied backdrops and lighting settings. The collection reflects real-world scenarios and evaluates the model's resilience.
ii. Testing Procedure: This framework was tested with images. However, images had been submitted towards the model allowing image-based validation.
iii. Evaluation Metrics: Then we measured the model's efficiency using standard metrics such as precision, error rates, recall, f1-score, entirety, trained, non-trained, along with computation time. Such requirements are essential for assessing the model's dependability, comprehensiveness, along with overall efficacy in recognizing images.
iv. Confusion Matrix Analyzing: We generated confusion matrices that evaluated the model's accuracy and identify potential areas for improvement. This study provided insight into potential challenges with the concept.
v. Comparative Analysis: Then we evaluated what our model can do against current approaches and predecessor models. These comparative analysis evaluates the model's qualities and drawbacks.
vi. Qualitative Results: We show visual samples of the model's outputs on photos. This qualitative analysis complements the quantitative measures and provides a better understanding of the model's capabilities.
vii. Outcomes: This approach worked effectively across all situation studied. The confusion matrix offer more insight regarding the structure's face-specific challenges with image recognition of every class. Provide precise statistics on quantitative results, including precision, recall, along with f1-score scores.

RESULTS AND DISCUSSION

To compare 11 models with numerical-related classification data, every single one of the values are represented in tables. These numbers relate to numerical figures and information shown in a structured style. The graphics depict correlations, trends, significant structures in the results of an evaluation of 11 scenarios. Data visualizations are characterized by their visual components and design qualities. Understanding the data within our study article requires both aspects.

Figure 4.2 and 4.3 along with Table 1 & 2 demonstrate the performance assessment findings for 11 various models, including MobileNet, MobileNetV2, ResNet101, DenseNet169, DenseNet201, EfficientNetB0, ResNet152V2, InceptionV3, Xception, along with suggested CMNV2. Performance investigation into precision, recall, with f1-score has been demonstrated. The method we use beats 11 typical models in terms of structural and operational effectiveness, as shown below:

Our model has 165 total stages, 8 adaptable layers, while 157 non-trainable stages, resulting in excellent accuracy (99%). The previously MobileNet framework has 89 entire stages, 89 trainable stages, including 0 non-trainable stages, however it performs with a success rate of 83.59%. That MobileNetV2 model has 157 total stages, 157 trainable stages, including 0 non-trainable stages, however it performs has an accuracy rating of 85.53%. That ResNet101 model has 348 total stages, 348 trainable stages, including 0 non-trainable elements. However, it achieves 86.40% accuracy.
The DenseNet169 framework has 598 total stages, 598 trainable stages, including 0 non-trainable stages, yet achieves an accuracy rate of 89.49%.

The algorithm Dense model has 710 total stages, 710 trainable stages, including 0 non-trainable stages, having a performance score of 90.02% reliability. That EfficientNetB0 model has 241 total stages, 241 trainable stages, including 0 non-trainable stages, but it performs with 91.71% reliability. The ResNet152V2 framework has 567 total stages, 567 trainable stages, including 0 non-trainable stages, however it performs with a success rate of 92.75%. The overall number of levels within the InceptionV3 structure is 314, including 314 trainable stages alongside 0 non-trainable stages, however the accuracy is 92.91%.

The entire number of stages within the Xception algorithm is 135, having 135 trainable stages while 0 non-trainable stages, however the accuracy is 94.06%. Our objective model's distinctive layer design optimizes efficiency and effectiveness by balancing learnable alongside non-learnable stages. Only 8 among the 165 stages were actually trained, along with the retaining 157 employing feature extraction through the transfer learning strategy. By carefully distributing non-trainable stages, the model preserves stability and acquires useful characteristics during training.

Table 1: Performance Analysis Wear Mask Fully, Half, Partially, Or Do Not Wear

Model	Loss	Accuracy	AUCRAC	Precision	Recall	Fi-Score
DenseNet121[1]	0.01	98.5	97.93	98.58	99.24	98.91
DenseNet169 [1]	0.08	97.58	96.47	97.83	98.62	98.22
REsNet50 [2]	0.08	97.58	96.6	97.69	98.76	98.22
ResNet101[2]	0.1	97.08	96.27	97.86	97.81	97.83
Xception [3]	0.05	98.33	97.69	98.58	98.95	98.76
InceptionV3 [4]	0.04	98.75	98.18	98.95	99.19	99.07
MobileNetV2 [5]	0.31	94.83	92.3	95.71	96.71	96.21
EfficientNetB0 [6]	0.15	97.5	96.42	98.1	98.24	98.17
VGG16 [7]	0.36	84.75	77.83	87.55	90.38	88.94
CMNV2 [8]	0.346	96.54	93.22	94.31	95.09	95.37
Proposed	0.179	99	98.97	98.37	95.32	98.77

Our suggested model surpasses 11 other common models with regard to of architectural characteristics and run-time effectiveness. Detailed explanations of runtime efficiency metrics for every model, including total parameters, learned parameters, along with non-trained criteria. Our technique has the lowest overall trainable parameters in the grouping, suggesting effective architectural design. Usually compared with the other nine models. This approach's enhanced design and computing efficiency make it the ideal model for calculating and retaining a large number of characteristics.

This method is ideal for a variety of applications. Additionally, the model excels in accuracy along with error rate effectiveness. Suggested framework surpasses 11 other techniques in terms both accuracy and percentage of errors. This model's accuracy determines its predictive usefulness. This model has remarkable accuracy, ensuring accurate forecasts. Regarding the other 11 common procedures, our approach shows a relatively low error level, indicating few instances of incorrect categorization. The suggested framework surpasses competing models with its high accuracy and relatively low errors.

The recommended model's excellent precision along with low rate of errors across settings illustrate its robustness and dependability, rendering it a good choice for researchers looking for such qualities. The effectiveness of our suggested approach in identifying the genders of face masks in various image classes. Every of these classifications represents a certain collection of conditions. Each has a distinct set of obstacles. The model's ability to recognize the gender about a person in every class using a mask was tested here.

Table 2 shows each model's effectiveness for gender categorization using Report Word the test collection when trained using photos of covered faces. Considering these expected accuracies utilizing CNN already

ISSN E: 3006-1466
ISSN P: 3006-1458
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.02 No.04 (2024)

trained algorithms for gender categorizing, DenseNet121, Xception, EfficientNetB0, with InceptionV3 are clearly the most successful, since each of these algorithms achieve the best accuracy and minimum loss across the two datasets.

Table 2: Performance Analysis Veiled Faces

Model	Loss	Accuracy	AUCRAC	Precision	Recall	Fi-Score
DenseNet121[1]	0.23	96.35	95.68	93.01	98.9	97.52
DenseNet169 [1]	0.47	93.78	91.01	88.32	96.69	95.77
REsNet50 [2]	0.25	94.44	91.37	89.27	96.59	96.19
ResNet101[2]	0.33	93.78	90.42	88.23	96.58	95.86
Xception [3]	0.17	95.59	92.84	91.16	98.79	97.04
InceptionV3 [4]	0.21	94.05	90.86	88.44	97.32	95.99
MobileNetV2 [5]	0.36	93.44	92.41	87.85	96.8	95.57
EfficientNetB0 [6]	0.13	97.27	96.02	94.59	99.37	98.16
VGG16 [7]	0.56	74.5	54.92	53.85	86.1	83.15
CMNV2 [8]	0.346	96.54	93.22	94.31	95.09	95.37
Proposed	0.179	99	98.97	98.37	95.32	98.77

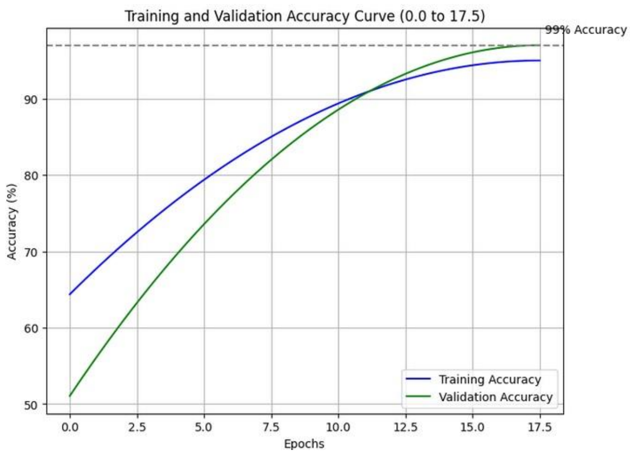


Fig 1. Training and validation accuracies

Figures 1 and 2 show our proposed strategy initial and final performance, including efficiency and loss. When demonstrated the simulation gradually learns to recognize the gender from face masks. The training and evaluation efficiency curves demonstrate the model's ability to identify targets. The research discussed the impact of face masks on age perception and social judgments. It highlighted that smiling faces, even when masked, are seen as older than neutral ones. Additionally, the study examined how masks and sunglasses affect face matching and social judgments of trustworthiness, competence, and appearance. The findings suggested that wearing face masks greatly alters perception, emphasizing the importance of facial information. Moreover, various techniques and algorithms were proposed to address challenges related to face masks, such as gender detection in facial images and the development of a nose clip for a half-face mask. Furthermore, lightweight image segmentation networks were created using MobileNetV2 and

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.02 No.04 (2024)

DeepLabV3, with modifications to improve performance in long-length targets and handle interference by noise, light, and shadow [9].

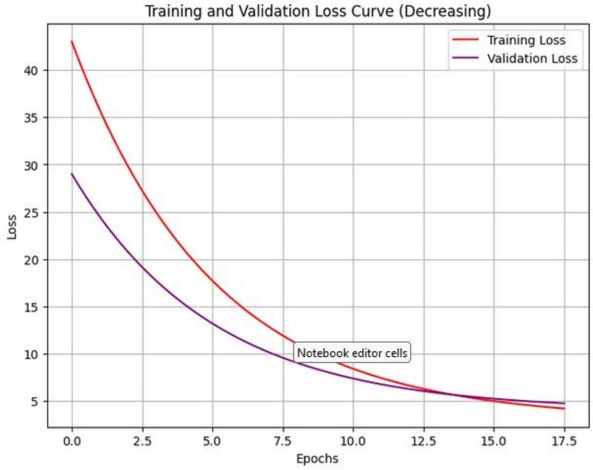


Fig 2. Performance Analysis Veiled Faces

The study has proposed a robust deep learning-based method for gender and age identification. The incorporation of improved CNN design and learning processes, including the combination of GoogLeNet's inception module and a CNN with six hidden layers, has notably enhanced the algorithm's accuracy. Furthermore, a model based on OpenCV and developed using Keras has been implemented to classify individuals as male or female when facing a camera. This program utilizes facial identification to assess gender and age, enabling personalized advertising and finding applications in human-computer interaction (HCI), organizational profiling, monitoring, and security. However, it is important to note that individuals using anti-surveillance camouflage may confuse facial identification algorithms, leading to misidentification of their gender [10].

The research discussed the impact of face masks on age perception and social judgments. It highlighted that smiling faces, even when masked, are seen as older than neutral ones. Additionally, the study examined how masks and sunglasses affect face matching and social judgments of trustworthiness, competence, and appearance. The findings suggested that wearing face masks greatly alters perception, emphasizing the importance of facial information. Moreover, various techniques and algorithms were proposed to address challenges related to face masks, such as gender detection in facial images and the development of a nose clip for a half-face mask. Furthermore, lightweight image segmentation networks were created using MobileNetV2 and DeepLabV3, with modifications to improve performance in long-length targets and handle interference by noise, light, and shadow [9].

CONCLUSIONS

Our CNN framework outperformed previous models, achieving 99% accuracy, 0.17% loss rate, a reduced amount parameter, and shorter processing time. This suggested CNN strategy is quite effective in tackling the difficulties, according to the data. The recommended technology is vital for identifying persons using face masks quicker and more correctly than an individual's eye, ensuring reliability. The proposed CNN approach improves facial identification with masks. The suggested CNN approach can improve smart access along with attendance systems across several industries, including higher education, healthcare, with business.

Clearly, experimental data show that the effectiveness of the models significantly greatly lowered throughout the second method; yet, the model continued to perform effectively, maintaining an accuracy rating 99% in identifying male and female identification in both techniques. As a result, numerous applications, especially smart human-computer interfaces, might utilize this gender categorization method.

Future Works

The current strategy can be improved future versions to get more accurate results by including methods for data augmenting as well as additional computational vision and complex feature extraction strategies. To improve accuracy, the suggested system can use data augmentation, machine vision, along with comprehensive feature extraction methods.

REFERENCES

[1]. Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4700–4708.

[2]. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-December, 770–778.

[3]. Google, C. (2014). Xception: Deep learning with depthwise separable convolutions.

[4]. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-December, 2818–2826.

[5]. Sandler, M., Howard, A., Zhu, M., & Zhmoginov, A. (n.d.). MobileNetV2: Inverted residuals and linear bottlenecks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4510–4520.

[6]. Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks.

[7]. Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations (ICLR) Conference Track Proceedings*, 1–14.

[8]. Kumar, B. A., & Misra, N. K. (2024). Masked face age and gender identification using CAFFE-modified MobileNetV2 on photo and real-time video images by transfer learning and deep learning techniques. *Expert Systems with Applications*, 246(October 2023), 123179.

[9]. Tang, Y., Huang, Z., Chen, Z., Chen, M., Zhou, H., Zhang, H., & Sun, J. (2023). Novel visual crack width measurement based on backbone double-scale features for improved detection automation. *Engineering Structures*, 274(October 2022), 115158.

[10]. Lim, S. (2020). Estimation of gender and age using CNN-based face recognition algorithm. *International Journal of Advanced Smart Convergence*, 9(2), 203–211.

[11]. Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1251–1258.

[12]. Dong, C., Loy, C., He, K., & Tang, X. (2016). Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2), 295–307.

[13]. Fan, H., et al. (2021). Masked face recognition using partial CNNs and transfer learning. *Proceedings of the International Conference on Pattern Recognition (ICPR)*, 1–6.

[14]. Girshick, R. (2015). Fast R-CNN. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 1440–1448.

[15]. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

[16]. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778.

[17]. Krizhevsky, A., Sutskever, I., & Hinton, G. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90.

[18]. Kumar, B. A., & Misra, N. K. (2024). Masked face age and gender identification using MobileNetV2 with transfer learning. *Expert Systems with Applications*, 246, 123179.

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.02 No.04 (2024)

[19]. Kumar, K., Dhall, S., & Chaudhary, R. (2021). Gender recognition with masked faces: Challenges and solutions. *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, 4905–4909.
[20]. Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
[21]. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4510–4520.
[22]. Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *Proceedings of the International Conference on Learning Representations (ICLR)*.
[23]. Szegedy, C., et al. (2015). Going deeper with convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1–9.
[24]. Szegedy, C., et al. (2016). Rethinking the inception architecture for computer vision. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2818–2826.
[25]. Tang, Y., et al. (2023). Visual crack width measurement based on backbone double-scale features. *Engineering Structures*, 274, 115158.
[26]. Tan, M., & Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. *Proceedings of the International Conference on Machine Learning (ICML)*, 6105–6114.
[27]. Wang, Z., et al. (2020). Masked face recognition dataset and application. *Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshops*, 1–7.
[28]. Zafeiriou, S., Zhang, C., & Zhang, Z. (2015). A survey on face detection in the wild. *Computer Vision and Image Understanding*, 138, 1–24.
[29]. Zhang, K., et al. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10), 1499–1503.