G-Tech : Jurnal Teknologi Terapan
Vol. 9, No. 1, 2025, p. 520-529
E-ISSN: 2623-064X | P-ISSN: 2580-8737

Unleashing the Power of Deep Learning: Revolutionizing Facial Recognition with GhostFaceNets

Ariana Tulus Purnomo[1]✉, Edrick Hansel Limantoro[2], Muhammad Nafis Aimanurrohman[3]

[1, 2, 3] Computer Science Study Program, Faculty of Engineering and Technology, Sampoerna University, Indonesia

✉ **Corresponding Author:** Ariana Tulus Purnomo (e-mail: ariana.purnomo@sampoernauniversity.ac.id)

Article Information

Article History

Received : January 12, 2025
Revised : January 25, 2025
Published : January 30, 2025

***Keywords*:**

Face recognition;
Face detection;
Deep learning;
GhostFaceNets;
Ghost Module .

ABSTRACT

Facial recognition technology has advanced significantly due to the development of deep learning algorithms. This paper explores deep learning, a branch of machine learning that employs multi-layered neural networks to simulate human decision-making processes in facial recognition. It provides a brief literature review of significant works in various deep learning architectures, including Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks. The core of the study is the implementation of the GhostFaceNets model, an enhancement of GhostNets, which is specifically designed for efficient and accurate facial recognition. By using Ghost Modules, this model reduces computational redundancy in generating additional feature maps through linear operations. An integrated attention mechanism is used in this study to emphasize critical facial features. Additionally, this study also employs the ArcFace loss function to improve class separation accuracy. The VGG2-FP dataset was used to train and evaluate this model and achieved an accuracy of 94.45%. This study contributes to the evolution of facial recognition systems, particularly in constrained computational environments.

INTRODUCTION

In the digital transformation era, deep learning has become the foundation for various technological innovations that revolutionize how humans interact with the digital world (Gohil, 2023). As a branch of machine learning, deep learning uses multi-layered neural networks to mimic the information-processing mechanisms of the human brain (Taye, 2023). This technology has proven reliable in various applications, ranging from voice recognition and natural language processing to image classification and object detection (Dargan et al., 2020). Its main advantage lies in its ability to process unstructured data, such as images and text, and automate the feature extraction process, which previously required human expert intervention (Navarro-Almanza et al., 2020).

Facial recognition, as one of the main applications of deep learning, has seen significant advancements in recent decades. This technology is widely used in various sectors, including security, financial services, and social media (Andrejevic & Selwyn, 2022). By relying on advanced algorithms, facial recognition systems can identify individuals with high accuracy by analyzing facial features from images or videos. Currently, the number of people wearing glasses is increasing, whether for vision aid or fashion purposes. For facial recognition, people will feel

This is an open access article under the CC BY license

uncomfortable if they have to remove their glasses or other facial accessories. Therefore, a system that can identify faces while retaining these properties will greatly enhance the comfort and usability of facial recognition. However, the use of glasses has introduced a new challenge: the need for reliable facial recognition even when glasses partially cover the face. This situation has driven researchers to develop innovative models to overcome these obstacles. In this context, facial recognition algorithms must be evaluated and updated to ensure their performance remains optimal amid changing societal needs. Various deep learning architectures, such as Convolutional Neural Networks (CNN) (Coşkun et al., 2017) (Lawrence et al., 1997), Long Short-Term Memory (LSTM) (Shukla et al., 2024) (Kankanamge et al., 2017. Each of these architectures brings a unique approach to addressing challenges in feature extraction and representation, which are fundamental to the success of facial recognition technology.

This paper explores the latest developments in deep learning and its applications in facial recognition. Specifically, we introduce a GhostFaceNets model, designed to enhance efficiency and accuracy in recognizing masked faces. This model leverages the Ghost Module to produce more efficient feature representations and uses attention mechanisms to improve accuracy. By evaluating the performance of this model using the VGG2-FP dataset, we demonstrate the great potential of this approach to address the challenges of facial recognition in the modern era.

LITERATURE REVIEW

For the last 30 years, different architectures has been introduced for deep learning, these include CNN, LSTM, AlexNet, GoogLeNet, and ResNet. In face recognition, Convolutional Neural Networks (CNN) serve as the primary foundation for extracting spatial features from faces (Alzubaidi et al., 2021). Architectures like AlexNet, GoogLeNet, and ResNet enhance accuracy and efficiency in recognizing faces from images (Almabdy & Elrefaei, 2019) (Krizhevsky, Sutskever, & Hinton, 2012) (Szegedy et al., 2015) Hasanah et al., 2023). AlexNet paved the way by introducing CNN for face classification (Krizhevsky, Sutskever, & Hinton, 2012). At the same time, GoogLeNet and ResNet improved performance through deeper and more efficient architectures, such as Inception modules and residual connections, which allow for training more complex networks without losing information (Szegedy et al., 2015) (Hasanah et al., 2023). For sequential data, such as facial video analysis, LSTM is used alongside CNN to understand temporal relationships between frames, thereby improving accuracy in recognizing expressions or identities from video sequences (Bellantonio, 2016).

Here is a brief explanation of CNN, LSTM, AlexNet, GoogLeNet, and ResNet in the context of face recognition

1. *Convolutional Neural Networks (CNN)*
   CNNs are essential to face recognition systems, designed to learn spatial hierarchies of features from input images. They employ layers of convolutional filters to detect patterns like edges, textures, and shapes, which are vital for identifying facial features (Alzubaidi et al., 2021).
2. *Long Short-Term Memory (LSTM)*
   LSTMs, a type of recurrent neural network (RNN), excel in learning and retaining long-term dependencies. In face recognition, LSTMs are paired with CNNs to analyze sequential data, such as video frames, capturing temporal relationships and enhancing recognition of facial expressions and identities over time (Staudemeyer & Morris, 2019; Bellantonio, 2016).
3. *AlexNet*
   AlexNet is a groundbreaking CNN architecture that greatly advanced image classification. It introduced ReLU activation functions, dropout for regularization, and data augmentation (Krizhevsky, Sutskever, & Hinton, 2012). In face recognition, AlexNet set the stage for using deep learning models to accurately classify and recognize faces.
4. *GoogLeNet (Inception Network)*

GoogLeNet introduced the Inception module, allowing for more efficient computation by using multiple convolutional filters of varying sizes in parallel (Szegedy et al., 2015). This architecture enhances the depth and width of the network without significantly increasing computational costs. In face recognition, GoogLeNet improves the ability to capture diverse facial features.

5. *ResNet (Residual Network)*
   ResNet introduced residual learning, which mitigates the vanishing gradient issue in training deep networks. Residual connections facilitate the learning of identity mappings, thus simplifying the training of deeper models (Hasanah et al., 2023). In face recognition, the deep architecture of ResNet enhances the model's capacity to learn intricate facial features.

Based on the explanation above, **Table 1** provides a comprehensive comparison of the advancements made in deep learning, especially in face recognition, highlighting the significant progress across various architectures and applications starting from CNN, LSTM,AlexNet, GoogLeNet, and ResNet.

Table 1. Comparison of advancements made in the field of deep learning.

Architecture	*Key Advancement*	*Common Implementations*
CNN	Presented the idea of local receptive fields, weight sharing, and spatial subsampling (Alzubaidi et al., 2021).	Image classification, object detection
LSTM	Solved the vanishing gradient problem in RNNs, enabling deep network training to emphasize temporal data (Staudemeyer & Morris, 2019).	Time series prediction, natural language processing
AlexNet	Utilized ReLU for faster convergence, dropout for controlling overfitting, and efficient GPU implementation for feasible training (Krizhevsky, Sutskever, & Hinton, 2012).	Image classification, transfer learning
GoogLeNet	Introduced the Inception module, significantly reducing the number of parameters (Szegedy et al., 2015).	Image classification, object detection
ResNet	Introduced residual learning to train extremely deep networks (Hasanah et al., 2023).	Image classification, object detection

RELATED WORK

Research in deep learning for facial recognition has experienced rapid development in recent decades. One significant milestone is the introduction of Convolutional Neural Networks (CNN), which introduced the concepts of local receptive fields and weight sharing to identify patterns in images (Khan et al., 2020). LeCun et al. pioneered the use of CNNs for character recognition, which later evolved into widely used techniques in image classification and object detection (LeCun et al., 2015). These innovations have made facial recognition more reliable and efficient, especially with architectures like AlexNet, GoogLeNet, and ResNet, which improve neural network performance with new techniques such as residual learning and Inception modules (Bhaidasna et al.)

Another advancement comes from the development of Long Short-Term Memory (LSTM), designed to address the vanishing gradient problem in recurrent neural networks (RNNs) (Staudemeyer & Morris, 2019). LSTM allows models to retain temporal information, making it an effective tool for processing sequential data, such as facial recognition in videos (Orozco et al., 2020). This technology enables facial recognition systems to handle changes in facial expressions and diverse lighting conditions, enhancing their reliability in real-world applications.

More recently, research has focused on developing lightweight models that still maintain high accuracy. Models like MobileFaceNets (Chen et al., 2018)6 and ShuffleFaceNet (Martindez-Diaz et al., 2019) are designed for resource-constrained devices, such as smartphones. These models adopt efficient architectures, including techniques like Depthwise Convolution and channel shuffle, to reduce computational complexity. This work paves the way for the application

of facial recognition across various platforms, expanding the accessibility of this technology to the broader public.

The landscape of face recognition technology is rich with algorithms that have pushed the boundaries of what's possible, the most popular being listed in **Table 2**. Light CNN, for instance, stands out for its ability to learn discriminative features by emphasizing the most significant feature map, resulting in a compact yet robust representation of facial features. This is achieved through a bootstrapping technique that iteratively refines the training process.

FaceNet (Schroff et al., 2015) uses a triplet loss function to ensure that images of the same person are closer together in the embedding space compared to images of different people. This is accomplished with a 128-dimensional embedding and utilizes the Inception-ResNet V1 architecture to process the information.

MobileFaceNets (Chen et al., 2018) and ShuffleFaceNet (Martindez-Diaz et al., 2019) both utilize the ArcFace loss to improve class separability. They represent facial features with a 512-dimensional embedding, but while MobileFaceNets use the MobileNetV2 (Chen et al., 2018) architecture and a Global Depthwise Convolution layer, ShuffleFaceNet (Martindez-Diaz et al., 2019) opts for the ShuffleNetV2 architecture, introducing channel shuffle operations to enhance feature channel communication.

VarGFaceNet introduces a novel approach with Recursive Knowledge Distillation, transferring knowledge from larger models to more compact ones (Yan et al., 2019). It also uses a 512-dimensional embedding and processes information through the VarGNet architecture, augmented with a Squeeze-and-Excitation module to emphasize important features.

Table 2. Comparison of face recognition methodologies.

Algorithm	*Loss Function*	*Representation Method*	*Processing Method*
Light CNN (Chen et al., 2024)	Max-Feature-Map	Compact embedding	Bootstrapping technique
FaceNet (Schroff et al., 2015)	Triplet loss	128-d embedding	Inception-ResNet V1 architecture
MobileFaceNet (Chen et al., 2018)	ArcFace	512-d embedding	MobileNetV2, GDC layer
ShuffleFaceNet (Martindez-Diaz et al., 2019)	ArcFace	512-d embedding	ShuffleNetV2, GDC layer
VarGFaceNet (Yan et al., 2019)	Recursive KD	512-d embedding	VarGNet, SE module

RESEARCH METHOD

Building upon the aforementioned algorithms, we turn our focus to an enhancement of GhostNets, known as GhostFaceNets. GhostNets, introduced by Han et al. in 2020, are based on the principle of creating more feature maps with fewer computations. They incorporate the concept of Ghost Modules, which apply a series of linear operations to generate "ghost" feature maps from the existing ones (Han et al., 2020). This method minimizes redundancy in feature maps, resulting in a more efficient model.

The algorithm can be outlined in the following steps.

1) Begin with a collection of intrinsic feature maps.
2) Perform a series of efficient linear operations to produce additional feature maps.
3) Concatenate the original and generated feature maps to form the final output.

Building upon GhostNets, GhostFaceNets by (Alansari et al., 2023) are designed specifically for face recognition tasks. They incorporate an attention mechanism to focus on important features and are trained using the ArcFace loss function, enhancing the learned features' discriminative power.

This methodology involves two main steps. First, information representation: GhostFaceNets uses Ghost Modules to represent facial features efficiently. The intrinsic feature maps are expanded through linear transformations, capturing various facial characteristics with minimal computational overhead. Second, information processing: The model employs a modified Global Depthwise Convolution (GDC) layer, replacing the traditional Global Average Pooling (GAP). Along with the ArcFace loss function, this layer processes the information to produce a highly discriminative feature vector robust to facial appearance variations.

The algorithmic process begins by inputting preprocessed facial images into the model. The Ghost Module then extracts and expands feature maps. Next, an attention mechanism is applied to focus on the most critical features. The Global Depthwise Convolution (GDC) layer combines the feature maps into a compact vector. During training, the ArcFace loss function is used to optimize the model, ensuring that the learned features are highly discriminative for accurate face recognition.

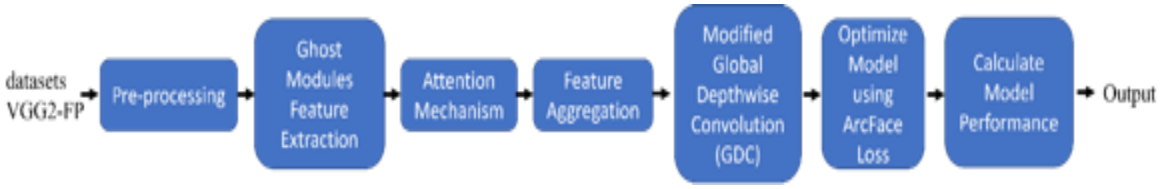

Figure 1. Diagram Block of Proposed System.

In this study, the proposed system can be seen in **Figure 1** with the following explanations.

A. Dataset

This study started by selecting a relevant dataset for training and evaluating the model. We use the VGG2-FP dataset, which is well-known in facial recognition for its variety, including different facial expressions, lighting conditions, and orientations. This dataset is chosen to ensure the model can handle real-world challenges.

B. Pre-processing

This step involves preparing raw data before it enters the model. This process includes image normalization, rotation, cropping, lighting, and resizing. Pre-processing aims to improve the quality of input data so the model can learn better without overfitting

C. Ghost Modules Feature Extraction

The model, GhostFaceNets, is designed using the Ghost module for more efficient feature representation. This module generates additional feature maps through cheap linear operations, reducing redundancy in feature maps and improving computational efficiency. GhostFaceNets also integrates an attention mechanism to focus on important facial features, enhancing recognition accuracy. To improve performance in recognizing faces with property, we use the ArcFace loss function during training. This function increases class separability by adding an angular margin, allowing the model to distinguish faces more precisely, even in difficult conditions.

D. Attention Mechanism

The attention mechanism enhances certain parts of the feature map relevant to a specific task, such as face recognition. This technique allows the model to "focus" attention on important features while ignoring less significant ones.

E. Feature Aggregation

This step combines features extracted from various layers or modules. Feature aggregation helps the model capture richer information and better representation of input data.

F. Modified Global Depthwise Convolution (GDC)

GDC (Global Depthwise Convolution) is a type of convolution that performs operations on each feature channel individually, without inter-channel merging (Martindez-Diaz et al., 2019). The modified version of GDC aims to enhance efficiency in the global feature extraction process by combining depthwise convolution principles with additional optimizations. This technique reduces the number of model parameters while still maintaining the ability to capture global information in the feature map. This is crucial in face recognition tasks, especially in resource-constrained conditions.

Several enhancements ensure that global features can be extracted more efficiently without sacrificing accuracy. In the modified version, global pooling operations are optimized by applying global average pooling before the depthwise convolution operation to capture information from the entire image. Adaptive pooling kernels are used to adjust better the global information captured from various input resolutions. Feature redundancy reduction is addressed by utilizing cheap linear operations in the Ghost module to generate additional features from the same data, thereby reducing redundancy and parameters. Advanced normalization is applied to the results from depthwise convolution to maintain the stability of the feature data distribution, which is crucial to prevent unstable gradients during training. Lastly, efficiency on small devices is enhanced by designing these modifications to speed up processing while maintaining accuracy on resource-constrained devices like mobile phones.

G. Optimize Model using ArcFace Loss

ArcFace Loss is a loss function designed to enhance class discrimination in the embedding space by projecting features onto a unit hypersphere and adding an angular margin (Alansari et al., 2023). This function helps the model produce more separated feature representations between different classes (e.g., different individuals' faces), thereby improving face recognition accuracy. ArcFace Loss value for the batch is described as the following equation (Alansari et al., 2023).

$$\mathcal{L}_{ArcFace} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s(\cos(\theta_{y_i}+m))}}{e^{s(\cos(\theta_{y_i}+m))}+\sum_{j=1,j\neq y_i}^{C}e^{s\cos(\theta_j)}} \quad (1)$$

where N represents the Number of samples in the batch, and s is the Feature scale (usually a large value, such as 64), m is the Angular margin to increase class separation (e.g., 0.5 radians), θ_{yi} is the angle between the feature (x_i) and the correct class weight (y_i), θ_j is the angle between the feature (x_i) and the class weight (j), and C is the number of classes in the dataset. ArcFace Loss enhances feature representation by adding an angular margin (m). Thus, it will make the embeddings of each class more separated from one another. This makes the model more accurate in distinguishing faces, even with similar attributes.

The combination of Modified GDC and ArcFace Loss in the Ghost Module is designed to address computational efficiency and high accuracy. Modified GDC ensures parameter reduction and computational load without compromising feature quality and ArcFace Loss enhances face class discrimination through angular margin. Modified Global Depthwise Convolution (GDC)

H. Calculate Model Performance

In this study, accuracy is used to assess the performance of the GhostFaceNets model in recognizing faces. The VGG2-FP dataset used in this research provides facial images with expression, lighting, and pose variations, allowing for a comprehensive evaluation of the model's accuracy. Accuracy is defined as

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ Predictions}\ x\ 100\% \quad (2)$$

This accuracy measures how often the model produces correct predictions compared to the total number of predictions made. In facial recognition, a correct prediction means the model successfully matches the given face with the corresponding identity (True Positive and True Negative). In contrast, total predictions include all outcomes, both correct and incorrect.

RESULTS AND DISCUSSION

Model Performance

In our exploration of face recognition methodologies, we have implemented a model that has demonstrated remarkable accuracy. Trained on the VGG2-FP dataset, a comprehensive and challenging collection of face images, our model has achieved an impressive accuracy rate of 94.45%, as shown in **Table 3**. This high percentage underscores the effectiveness of our proposed methodology in the complex task of face recognition.

Table 3. Accuracy of GhostFaceNetV1-1 using VGG2-FP dataset.

Model	Training datasets VGG2-FP (%)
GhostFaceNetV1-1	94.45

The VGG2-FP dataset, renowned for facial recognition, offers a wide variety of facial images, making it an ideal training ground for our model. This dataset includes over 3 million facial images and around 9,000 identities. The images are taken under various conditions, including age variations, facial expressions, lighting, and poses.

In this study, GhostFaceNets achieved a high accuracy of 94.45% on the VGG2-FP dataset. This indicates that the model can handle the challenges of facial recognition under various conditions, including different lighting and faces covered with properties. This accuracy was achieved by integrating the Ghost module, which significantly enhances feature representation efficiency by reducing redundancy.

Using the ArcFace loss function also plays a crucial role in improving the model's performance. With the angular margin introduced by this function, GhostFaceNets can better separate classes, thereby increasing face recognition accuracy. These results demonstrate that a loss function specifically designed for facial recognition offers significant advantages over conventional approaches. This accuracy rate holds significant potential for various applications, ranging from security and surveillance systems to social media platforms, where accurate facial recognition is paramount.

However, it's important to recognize that while the high accuracy rate is encouraging, it isn't the only measure of the model's effectiveness. Other performance indicators, such as the model's ability to generalize to unseen data, its robustness under various lighting conditions, angles, and facial expressions, and its resilience to adversarial attacks, are also crucial in evaluating the overall performance of a facial recognition system.

In summary, this facial recognition model's 94.45% accuracy rate underscores the validity of our proposed methodology. It emphasizes the transformative potential of deep learning in advancing facial recognition technology. As we continue to refine our model and methodology, we expect further enhancements in accuracy and robustness, paving the way for more reliable and secure facial recognition systems.

Checking The Same Face Between Two Images

As additional validation, we integrate GhostFaceNets into a Python-based facial recognition system for direct testing. Using libraries like DeepFace, we compare individual faces from two different images to evaluate accuracy. The implementation is used for verifying if two inputs of images has the same face or not. For example, the picture of Barack Obama is compared with Joe Biden as shown in **Figure 2** and check if the algorithms successfully detect it.

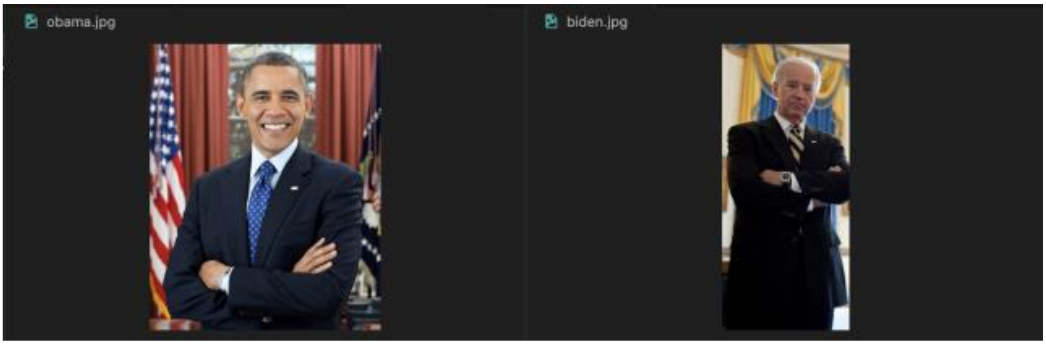

Figure 2. Obama and Biden Images

The output is as follow: {'verified': False} . The 'verified' component here marks the decision of the deep learning algorithm holding a Boolean value whether it has the same faces between the two inputs. As it is run, the 'verified' value is showing False value means that those two images have different faces. Next, we tried to do checking for the same faces between two images as shown in **Figure 3.** The output is as follow: {'verified': True}. Means that the deep learning successfully identify the face of Barack Obama even with different face expression.

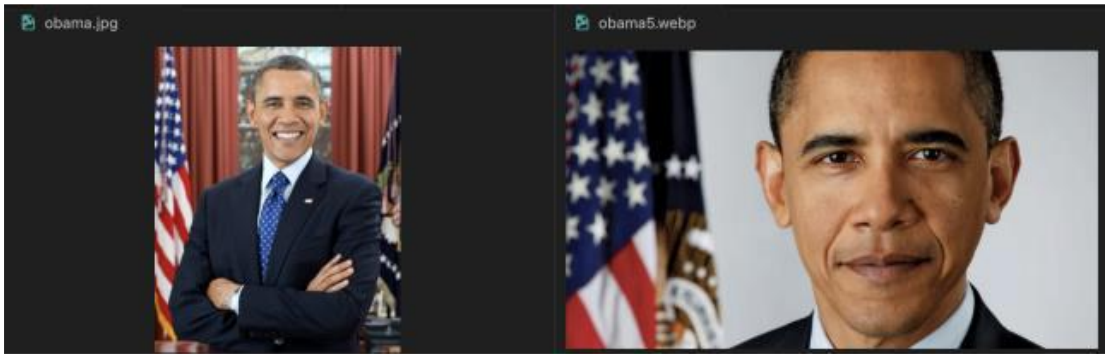

Figure 3. Two Obama Images.

Group Facial Recognition

In this research, we also investigate the potential of GhostFaceNets for group facial recognition. For group facial recognition, we attempted to implement our model and test it using other photo using glasses for the system's input. This implementation uses DCNN (Deep Convolutional Neural Network), a library available in Python. It successfully performs face detection and feature extraction, which involves DCNN generating the face encodings, face comparison, and recognition. **Figure 4** shows the input image for the algorithm to extract face encodings for two people, Edrick (left) and Nafis (right).

Figure 4. Input Image

Figure 5. Target Image

Figure 6. Output Image.

Figure 7. Zoomed version of Output Image.

The target image in **Figure 5** is then given to the program so that it can detect and recognize each face. If a face is known, the program creates a border around the face and shows the person's name below it. The result is seen in **Figure 6.** It successfully detects the faces with and without glasses and labels them as Edrick and Nafis. **Figure 7** is a close-up look of **Figure 6**.

CONCLUSION

This study discusses the application of the GhostFaceNets model to enhance facial recognition performance. By utilizing the Ghost Module for efficient feature extraction, attention mechanisms to strengthen the focus on important features, and a combination of Modified Global Depthwise Convolution (GDC) with ArcFace Loss, this model delivers accurate and efficient results. This approach demonstrates that reliable facial recognition can be achieved even under limited computational resources. The VGG2-FP dataset was used to evaluate the performance of this model. The GhostFaceNets model achieved an accuracy of 94.45%, indicating that integrating such as the Ghost Module and attention mechanisms can enhance accuracy. Although the results are promising, there are opportunities for further research. For instance, exploring more diverse datasets could provide a broader evaluation of this model. Additionally, developing more adaptive algorithms to handle extreme variations in facial attributes will be an important step forward. Thus, this study provides a strong foundation for further innovation in efficient and reliable facial recognition technology.

REFERENCES

Almabdy, S., & Elrefaei, L. (2019). Deep convolutional neural network-based approaches for face recognition. *Applied Sciences, 9*(20), 4397.

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., ... & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data, 8*, 1-74.

Andrejevic, M., & Selwyn, N. (2022). *Facial recognition.* John Wiley & Sons.

Bellantonio, M. (2016). Hybrid CNN+ LSTM for face recognition in videos (Master's thesis, Universitat Politècnica de Catalunya).

Bhaidasna, M. Z. C., Swaminarayan, P. R., & Bhaidasna, M. H. Z. Enhancing face recognition with deep learning architectures: A comprehensive review.

Chen, F., Li, S., Han, J., Ren, F., & Yang, Z. (2024). Review of lightweight deep convolutional neural networks. *Archives of Computational Methods in Engineering, 31*(4), 1915-1937.

Chen, S., Liu, Y., Gao, X., & Han, Z. (2018). Mobilefacenets: Efficient CNNs for accurate real-time face verification on mobile devices. In *Biometric Recognition: 13th Chinese Conference,*

CCBR 2018, Urumqi, China, August 11-12, 2018, Proceedings 13 (pp. 428-438). Springer International Publishing.

Coşkun, M., Uçar, A., Yildirim, Ö., & Demir, Y. (2017, November). Face recognition based on convolutional neural network. In *2017 International Conference on Modern Electrical and Energy Systems (MEES)* (pp. 376-379). IEEE.

Dargan, S., Kumar, M., Ayyagari, M. R., & Kumar, G. (2020). A survey of deep learning and its applications: A new paradigm to machine learning. *Archives of Computational Methods in Engineering, 27*, 1071-1092.

Gohil, M. G. (2023). A study on the impact of artificial intelligence (AI), automation, and the digital transformation on society. *Revista Review Index Journal of Multidisciplinary, 3*(4), 22-27.

Han, K., Wang, Y., Tian, Q., Guo, J., Xu, C., & Xu, C. (2020). GhostNet: More features from cheap operations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 1580-1589).

Hasanah, S. A., Pravitasari, A. A., Abdullah, A. S., Yulita, I. N., & Asnawi, M. H. (2023). A deep learning review of ResNet architecture for lung disease identification in CXR images. *Applied Sciences, 13*(24), 13111.

Kankanamge, S., Fookes, C., & Sridharan, S. (2017, September). Facial analysis in the wild with LSTM networks. In *2017 IEEE International Conference on Image Processing (ICIP)* (pp. 1052-1056). IEEE.

Khan, A., Sohail, A., Zahoora, U., & Qureshi, A. S. (2020). A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review, 53*, 5455-5516.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems, 25*.

Lawrence, S., Giles, C. L., Tsoi, A. C., & Back, A. D. (1997). Face recognition: A convolutional neural-network approach. *IEEE Transactions on Neural Networks, 8*(1), 98-113.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436-444.

Martindez-Diaz, Y., Luevano, L. S., Mendez-Vazquez, H., Nicolas-Diaz, M., Chang, L., & Gonzalez-Mendoza, M. (2019). Shufflefacenet: A lightweight face architecture for efficient and highly-accurate face recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops* (pp. 0-0).

Navarro-Almanza, R., Juárez-Ramírez, R., Licea, G., & Castro, J. R. (2020). Automated ontology extraction from unstructured texts using deep learning. In *Intuitionistic and Type-2 Fuzzy Logic Enhancements in Neural and Optimization Algorithms: Theory and Applications*, 727-755.

Orozco, C. I., Buemi, M. E., & Berlles, J. J. (2020, December). Towards an attention mechanism LSTM framework for human action recognition in videos. In *2020 IEEE Congreso Bienal de Argentina (ARGENCON)* (pp. 1-6). IEEE.

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 815-823).

Shukla, A. K., Shukla, A., & Singh, R. (2024). Automatic attendance system based on CNN–LSTM and face recognition. *International Journal of Information Technology, 16*(3), 1293-1301.

Staudemeyer, R. C., & Morris, E. R. (2019). Understanding LSTM--a tutorial into long short-term memory recurrent neural networks. *arXiv preprint arXiv:1909.09586*.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1-9).

Taye, M. M. (2023). Understanding machine learning with deep learning: Architectures, workflow, applications, and future directions. *Computers, 12*(5), 91.

Yan, M., Zhao, M., Xu, Z., Zhang, Q., Wang, G., & Su, Z. (2019). Vargfacenet: An efficient variable group convolutional neural network for lightweight face recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops* (pp. 0-0).