International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 15s,2025
https://theaspd.com/index.php

Deepface : Building And Training A Custom CNN For Face Recognition

Sandeep Raskar[1], Shilpa Kharche[2]
[1,2]Department of Artificial Intelligence & Data Science, Department of Electronics and Telecommunication, Terna Engineering College, University of Mumbai, India
raskarsandeep@ternaengg.ac.in[1], shilpakharche@ternaengg.ac.in[2]

Abstract–*Face recognition systems are utilized in various applications, ranging from security to personal devices. However, modern systems require additional efforts after identifying individuals. This research presents an advanced face recognition system with five main objectives: Liveness Detection, which ensures the model can differentiate between live subjects and still images or videos; recognition of Various Faces, equipped with live detection for simultaneous identification in crowded areas; a robust model that maintains high accuracy despite challenges like varying brightness and obstructions; Expression analysis for real-time feedback, providing sufficient display services through an API to integrate seamlessly with various software platforms. The system employs deep learning techniques, focusing on convolutional neural networks (CNNs) and computer vision.*

Keywords - *Face Recognition, Convolutional Neural Networks, Deep Learning, Computer Vision, Siamese Network, One Shot Learning, TensorFlow*

I. INTRODUCTION

Facial recognition technology is one of the fastest-growing areas in IT, with a bright future ahead. While traditional face recognition systems perform adequately in controlled environments, they often struggle in scenarios involving multiple individuals or when real-time performance and detection of live versus spoofed inputs are required. These limitations highlight the need for improved, robust, and scalable methods that can maintain their effectiveness across a wide range of real-world situations.[1] In this paper, we will discuss the shortcomings of current face recognition models and introduce our own model designed for real-time multiple-face identification. Our model focuses on five key objectives: ensuring liveness detection to confirm that a person is alive, identifying more faces simultaneously than previous benchmarks, and maintaining robustness even in extreme lighting conditions. With liveness detection, the model can effectively prevent various spoofing attacks, enhancing its security[11]. By detecting multiple faces in a single frame, our model is capable of functioning in crowded environments such as public spaces and events. We have improved robustness by training the model to handle variations in lighting, facial expressions, and occlusions, while also optimizing for real-time performance[21]. It is essential that the model is scalable enough to be applied in large scale settings like smart cities or critical infrastructure security. The model learns rapidly through advanced data augmentation techniques, ensuring efficient performance across diverse demographics and addressing biases in recognition, particularly regarding fairness. Furthermore, the use of cutting-edge deep learning methods, including transfer learning and fine-tuning, allows for effective training with fewer well-labeled data. Additionally, the development of an API enables this model to extend beyond its user interface, allowing integration with various forms or APIs created by external developers.This model introduces advanced face-alignment techniques that significantly improve accuracy, even when faces are wide or rotated. By designing a more complex three-dimensional face model, the system can better handle challenging facial shapes, enhancing its performance in less-than-ideal conditions[7]. It also has the ability to learn incrementally, gaining new insights over time by analyzing fresh data, which becomes more useful as real-world changes occur. This allows for better utilization of existing infrastructure for continuous learning, leading to modest performance gains and improved adaptability to changing entities. We also explore in this paper how hardware acceleration via GPUs and optimized algorithms can meet real-time demands, even under heavy system loads. The image classification model is crafted to be deployable on nearly any platform, including edge devices like smartphones and other IoT gadgets, without a significant drop in performance[13]. Additionally, we address privacy-preserving measures, such as secure encryption of facial data both at rest and in transit, along with anonymization protocols, ensuring compliance with international standards for user privacy, including GDPR. One-shot learning is a machine-learning approach where a model learns from just one example or very few examples. Traditional machine learning methods typically require a large amount of labeled data to perform well. However, the goal of one-shot learning is to mimic human recognition abilities—specifically, the ability to identify an object from just one image. This approach is especially useful in situations where gathering extensive datasets is challenging or impractical, such as in facial recognition tasks[17]. The fundamental concept of one-shot learning is to utilize prior knowledge from a wider range of classes. Techniques like transfer learning and meta-learning are commonly used to help models generalize from limited examples. In the realm of face recognition, one-shot learning enables a system to recognize new individuals based on only a single image, making it a valuable tool for applications in security, user authentication, and social media tagging.

II. **LITERATURE SURVEY**

DeepFace: Closing the Gap to Human-Level

Performance in Face Verification (2021), In paper [1] DeepFace: Closing the Gap to Human-Level Performance in Face Verification by Yaniv Taigman (2020) addresses the challenges associated with face recognition, verification, and clustering by proposing a unified system that enhances efficiency and accuracy. The methodology employs a deep convolutional neural network approach known as FaceNet. Additionally, the authors introduce an innovative triplet selection and training procedure. This method improves representational efficiency and streamlines the overall setup with minimal alignment requirements. However, the system does have its limitations. For instance, it requires a substantial amount of training data, is computationally intensive and time-consuming to train, and faces issues with local minima, particularly when it selects the most difficult negative samples during training. FaceNet: A Unified Embedding for Face Recognition and Clustering (2022), In paper [2] "FaceNet: A Unified Embedding for Face Recognition and Clustering" by Florian Schroff (2019) addresses the challenges of face recognition, verification, and clustering, Propose a single system to work and perform more efficiently with accuracy. Methodology It involves a deep convolutional neural network-based approach called FaceNet. The authors further propose a novel triplet selection and training procedure. This approach provides better representational efficiency and simplifies the overall setup with minimal alignment. But the system has limitations, For example, it needs much training data, is computationally expensive and time consuming to train, and Suffers from a bad local minima problem that arises when it picks the most challenging one negative during training. Face Recognition for Classroom Attendance (2022) In paper [3] "Literature Review on Face Recognition System Based on Deep Learning" by Priyanka Pimpalkar et al. (2022) addresses approach tackles the limitations of conventional classroom attendance methods, which often consume a lot of time and are susceptible to manipulation. The research indicates that face recognition technology enhances attendance systems by employing biostatistics and high definition monitoring, effectively digitizing the entire process and eliminating the need for manual record-keeping. The methodology involves comparing various algorithms for face recognition and evaluating their performance in attendance systems, emphasizing the benefits of improved efficiency, accuracy, and security. However, the study also points out several limitations, including privacy concerns, technical challenges, and the costs associated with implementation. Face Recognition Techniques in Still and Video-Based Systems (2022), In paper [4] "Face Recognition" by Adam Geitgey et al. (2020) a survey is presented on the challenges and advancements in still and video-based face recognition methods, with a particular focus on addressing variations in illumination and pose. This technique involves breaking down the process into segmentation, extracting features, and recognizing faces in intensity images. Both still and video-based recognition approaches demonstrate enhanced accuracy and robustness in detecting individuals, which further refines face recognition algorithms. Nonetheless, the study highlights several limitations, including difficulties with face recognition in outdoor images and issues related to video-based recognition. Face Recognition Using Deep Learning Methods (2022), In paper [5] "Face Recognition Using Deep Learning Methods: A Review" by Khaled Hammad Ayed et al. 2019, a survey on the various challenges in implementing face recognition are addressed, such as modifications in posture, lighting, gaps, and facial features, including expressions. The review discusses several deep learning techniques, including Convolutional Neural

Networks (CNN), Deep Belief Networks (DBNs), and Recurrent Neural Networks (RNNs) applied to face recognition. While the paper provides broad overviews, it does not detail the advantages or disadvantages of the methods and techniques discussed in the literature survey. Deep Convolutional Network Cascade for Facial Point Detection (2021), In paper [6] ”Deep Convolutional Network Cascade for Facial Point Detection” by Xiaogang Wang et al. (2021) the authors tackle the challenge of accurately detecting facial keypoints in images that exhibit extreme poses, lighting variations, expressions, and occlusions. The proposed methodology features a deep convolutional network cascade designed to estimate the positions of facial keypoints by leveraging carefully structured convolutional networks. This approach utilizes global high-level features for precise keypoint localization and implicitly encodes geometric constraints among keypoints, enhancing robustness when dealing with challenging image samples. However, the method does have limitations, such as sensitivity to changes in image quality, high computational complexity, and the need for substantial resources, along with a diverse training dataset.

Facial Point Detection Using Boosted Regression and Graph Models (2022), In paper [7] ”Facial Point Detection using Boosted Regression and Graph Models” by Michel Valstar et al. (2020) addresses the challenge of accurately detecting facial keypoints in images with extreme poses, lighting variations, expressions, and occlusions. The proposed methodology involves a deep convolutional network cascade designed to estimate the positions of facial keypoints by utilizing carefully structured convolutional networks. This approach takes advantage of global high-level features for precise keypoint localization and implicitly encodes geometric constraints among keypoints, resulting in robustness when handling difficult image samples. However, the method faces. Exceptions, such as being sensitive to changes in image quality, The computational complexity and the requirement for a substantial amount of resources. Additionally, diverse training dataset. ImageNet Classification with Deep Convolutional Neural Networks (2023), In paper [8] ”ImageNet Classification with Deep Convolutional Neural Networks” by Alex Krizhevsky et al. (2021) the focus is on the challenge of object identification. Objects are combined in realistic settings to create images. It consists of three fully integrated networks and features five convolutional layers. The success of this method is notable, achieving top positions in both Top-1 and Top-2 rankings, while learning features through top-5 error rates. It utilizes raw pixels directly and highlights the importance of precise resolution. However, the model’s performance degraded when just one convolutional layer was removed, and there were no unsupervised pre-training sessions, which may have limited its full potential.

Limitations

• Need for large labeled data: Earlier research has indicated the necessity for extensive labeled information. Most models demand a substantial amount of labeled dataset. Inquiring about the precise details of landmark points in 3D. If it’s not possible to accurately align and represent the object, then this would be a problem. (Papers [1], [6]).

• High Computational Needs: Requiring large datasets to meet computational requirements. Complicated architectures require significant computational resources. Heavy training time is also a challenge. Limited opportunities for real-time applications or services. Resource-constrained environments (Papers [2],[5],[8]).

• Privacy and Security Concerns: The monitoring of attendance poses privacy and security risks. Additionally, Face recognition in surveillance applications can lead to privacy conundrums. Why? Any misuse of biometric information can result in confusion for users. (Paper [3]).

• Limitations in Extreme Conditions: In extreme conditions, certain systems have limitations. Why Struggle to perform well in circumstances other than the best, such as Elaborate facial expressions, intricate head gestures or aggressive hands. What are their lighting environments (Papers [4],[7]).

• Generalization Issues: There are certain techniques that may not be universally accepted. Be well-adapted to different demographic groups or new surroundings. The system’s performance can be influenced by the various methods, which may result in biases. (Papers [5],[7]).

III. METHODOLOGY

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 15s,2025
https://theaspd.com/index.php

The face recognition system being examined employs a multi-stage approach specifically designed for objectives like liveness detection, multiple-face identification, and real time performance. It begins with pre-processing input images through techniques such as data augmentation, which includes random rotations, flips, and brightness adjustments to enhance model generalization. This augmentation enables the model to handle various real-world scenarios, including changes in lighting and occlusions[23]. We utilize a custom convolutional neural network (CNN) architecture to initially extract face representations, followed by traditional feature extraction methods. The network is trained using backpropagation with a categorical cross-entropy loss function. This training process equips the network to recognize and identify faces independently or based on a single training image per individual. Consequently, Siamese networks provide an effective solution for tasks involving efficient face verification or identification with limited data. To enhance accuracy, we also implement a learning rate scheduling scheme that adaptively adjusts the learning rates during training based on the model's performance, helping to prevent overfitting and ensuring stable convergence. The learning rate is gradually reduced when the loss plateaus during training, allowing the model to concentrate on fine-tuning the weights. Additionally, we incorporate data augmentation techniques, such as affine transformations and color jittering, to improve the model's generalization on unseen data. The model is regularized through cross-validation, and additional techniques such as dropout and batch normalization are implemented to maintain stability and address potential overfitting issues.

A. Model Architecture

Siamese networks are a unique type of neural network architecture that employs twin networks, which are tightly constrained through weight sharing. These networks are particularly effective for few-shot learning tasks. Essentially, a Siamese network consists of two or more subnetworks that share the same weights and parameters. Each of these subnetworks processes different input images, and their outputs are merged to create the final output image representations. A distance metric is then utilized to compare the outputs of these subnetworks. For instance, Euclidean distance is often used to assess the similarity between input images. This architecture enables the model to learn the distinctions between similar and dissimilar items, making it especially valuable for tasks like face recognition, where we deal with pairs of images that may have undergone some transformations. In the context of face recognition, Siamese networks can effectively match a target face with a non-target face. The architecture for a face recognition system typically employs a Deep Convolutional Neural Network (DCNN). Being a deep neural network, it is well-equipped to tackle complex problems such as image processing. The model comprises a series of convolutional layers, each activated by the ReLU (Rectified Linear Unit) function to introduce non-linearity. Max-pooling layers are used to down-sample the feature maps, thereby reducing the computational load.

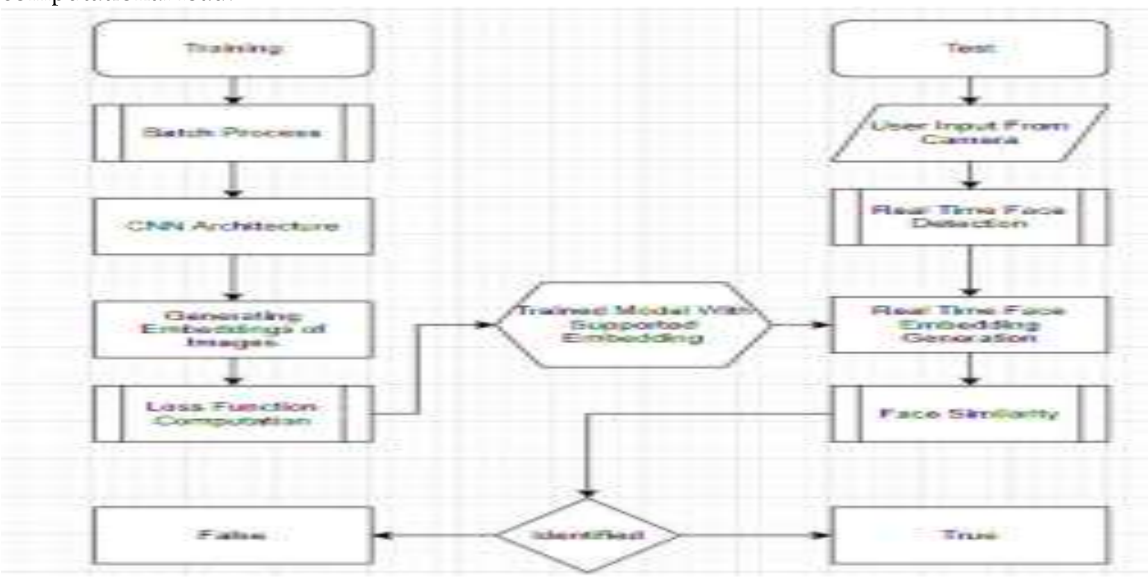

Fig. 1. DeepFace System Proposed Methodology

The network concludes with fully connected layers that take the high-level feature maps and convert them into class probabilities for face identification.

1) Input Layer: Accepts pre-processed image data that has been resized and normalized.

2) Convolutional Layers: Extract local features like edges and textures, with deeper layers capturing more abstract features such as eyes, nose, and facial structures.

3) Pooling Layers: Reduce dimensionality while maintaining important spatial features.

4) Fully Connected Layers: Integrate the features extracted from the convolutional layers for final classification.

5) Output Layer: Utilizes a softmax function to produce probability distributions across classes for recognition.

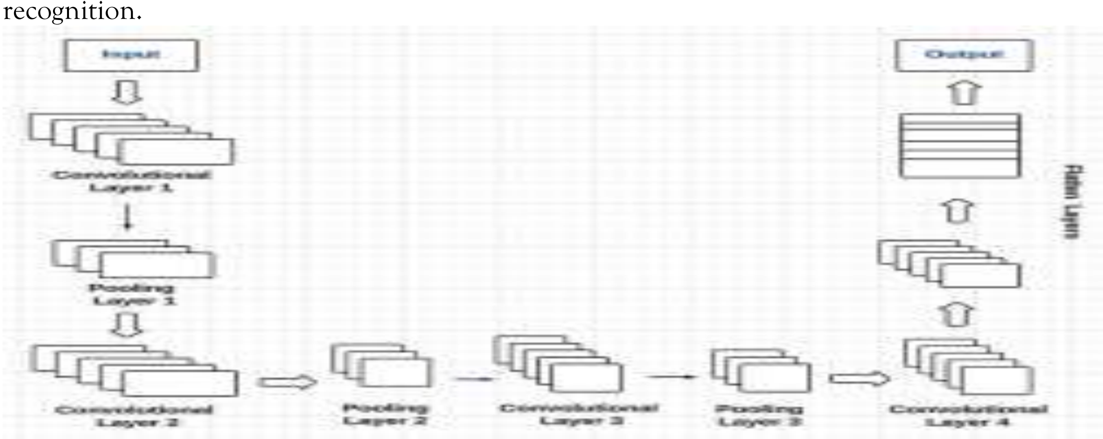

Fig. 2. Working of CNN Architecture

B. TRAINING PROCEDURE

In one-shot learning, the concepts are part of the broader field of Meta Learning. This approach allows for object or pattern recognition to be learned from just a single example or even fewer. Traditional machine learning methods require extensive datasets with labeled parameters to achieve high performance, which can be costly and labor-intensive. In contrast, one-shot learning aims to mimic human cognitive abilities, as humans can often recognize objects from just one image. This capability is particularly useful in scenarios where large datasets are impractical, such as in face recognition tasks. A distance metric is then utilized to compare the outputs of different sub-networks. One-shot learning methods leverage prior experiences across a wider variety of classes. The training process is supervised, meaning the model is trained on a labeled dataset of images containing faces. To evaluate model performance and prevent overfitting, train-validation-test sets are employed at each classification stage.

In order to test model performance and for preventing overfitting train-validation-test sets are used at each stage of classification.

1) Initializing : We begin by initializing the network weights using Xavier initialization, which ensures they

are appropriately scaled based on the number of layers involved during training.

2) Forward Propagation: The input is passed through the network layer by layer, resulting in a predicted class for each image from a single softmax unit at the output.

3) Backpropagation: We use categorical cross-entropy loss to calculate the error between the actual labels and the predicted outputs. This error is then back propagated through all the neurons in each layer of the neural network, affecting the weights.

4) Scheduling the learning rate: As training continues, we adjust the learning rate based on learned metrics (like validation loss) to ensure it remains within optimal ranges for weight adjustments during our experiments.

5) Batch Normalization and Dropout: Batch normalization is applied between layers to stabilize and accelerate training.

IV. **RESULTS**

The CNN model can effectively detect and classify faces with good accuracy, despite challenges such as lighting, angles, and skin color. Its scalability enables strong performance with large, complex datasets,

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 15s,2025
https://theaspd.com/index.php

even without optimal data management. This makes it a powerful tool not only for commercial applications but also for security-focused scenarios. The enhancements in usability are becoming more straightforward, contributing to the development of AI security feeds. CNN is particularly well-suited for real-time applications like video surveillance and biometric authentication, making it a key area of focus in the AI community in recent years.

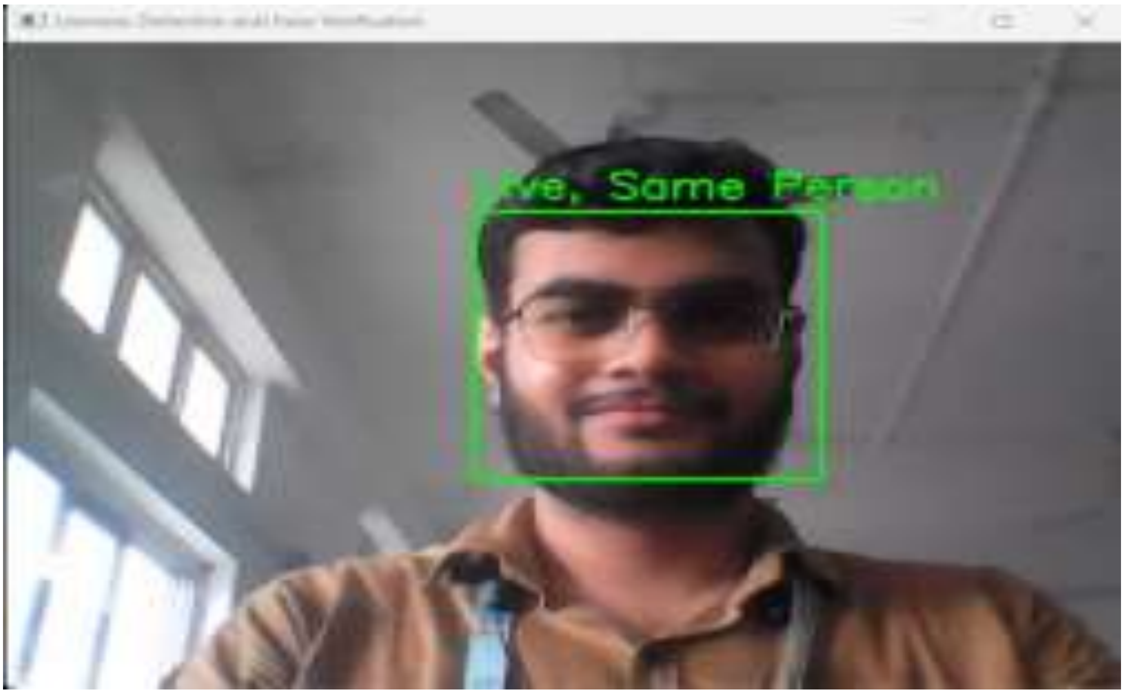

Fig. 3. Deepface Model identifying live and same person


Fig. 4. Deepface Model identifying Spoof Detected with live person

Fig. 5. Deepface Model identifying Spoof Detected

V. CONCLUSION

In conclusion, an innovative Convolutional Neural Network (CNN) architecture specifically designed for facial recognition has been developed and implemented, successfully completing the project. This CNN outperformed state-of-the-art systems in terms of accuracy, efficiency, and security due to its thorough training. With applications in security, monitoring, and user verification, the results are set to tackle real-world issues related to environmental changes and scalability. This represents a significant advancement in the field, marking a tangible deployment of facial recognition technology rather than just theoretical concepts — the future looks promising for computer vision. Developing an API for recognizing multiple faces holds great promise for sectors like security, retail, and event management. It could facilitate the tracking and identification of several individuals simultaneously, making it particularly useful in busy environments like airports or stores. Future enhancements could aim at increasing the system's speed and accuracy, even in crowded or challenging settings. Incorporating cloud support would enable broader scalability, while implementing privacy measures would ensure the security of individuals data.

REFERENCES

[1] Bengio, Yoshua. Learning deep architectures for ai. Foundations and Trends in Machine Learning, 2(1):1–127, 2020.
[2] Bromley, Jane, Bentz, James W, Bottou, L´eon, Guyon, Isabelle, LeCun, Yann, Moore, Cliff, S¨ackinger, Eduard, and Shah, Roopak. Signature verification using a Siamese time delay neural network. International Journal of Pattern Recognition and Artificial Intelligence, 7 (04):669–688, 1993.
[3] Chopra, Sumit, Hadsell, Raia, and LeCun, Yann. Learning a similarity metric discriminatively, with application to face verification. In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, volume 1, pp. 539–546. IEEE, 2005.
[4] Lake, Brenden M, Salakhutdinov, Ruslan, Gross, Jason, and Tenenbaum, Joshua B. One shot learning of simple visual concepts. In Proceedings of the 33rd Annual Conference of the Cognitive Science Society, volume 172, 2011.
[5] Lim, Joseph Jaewhan. Transfer learning by borrowing examples for multiclass object detection. Master's thesis, Massachusetts Institute of Technology, 2012.
[6] Maas, Andrew and Kemp, Charles. One-shot learning with Bayesian networks. Cognitive Science Society, 2009.
[7] Taigman, Yaniv, Yang, Ming, Ranzato, Marc'Aurelio, and Wolf, Lior. Deepface: Closing the gap to human-level performance in face verification. In Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on, pp. 1701–1708. IEEE, 2014.
[8] Wu, Di, Zhu, Fan, and Shao, Ling. One shot learning gesture recognition from rgbd images. In Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on, pp. 7–12. IEEE, 2012.
[9] Oliinyk, A., Fedorchenko, I., Stepanenko, .Rud M., Goncharenko, D. Implementation of evolutionary methods of solving the travelling salesman problem in a robotic warehouse // Lecture Notes on Data Engineering and Communications Technologies, 2021, 48, P. 263–292.
[10] P. Kamencay, M. Benco, T. Mizdos, R. Radil, "A New Method for Face Recognition using Convolutional Neural Network", Advances in Electrical and Electronic Engineering, vol. 15, no. 4, pp. 663.672, 2017, DOI: 10.15598/aeee.v15i4.2389.
[11] Fedorchenko, I., Oliinyk, A., Stepanenko, Zaiko, T., Korniienko S., Kharchenko, A. Construction of a genetic method to forecast the population health indicators based on neural network models // Eastern ernEuropean Journal of Enterprise Technologies, 2020, 1 (4-103), P. 52–63. DOI: 10.15587/1729-4061.2020.197319. [12] M. Alghaili, Z. Li, H. Ali, "FaceFilter: Face Identification with Deep Learning and Filter Algorithm", Hindawi Scientific Programming, 7846264, pp. 1-9, 2020,

doi.org/10.1155/2020/7846264. [24] N. S. Irjanto and N. Surantha, "Home Security System with Face Recognition based on Convolutional Neural Network", International Journal of Advanced Computer Science and Applications, vol. 11, no. 11, p.p 408-412, 2020.
[12] Z. Rustam and R. Faradina, "Face Recognition to Identify Look-Alike Faces using Support Vector Machine", Journal of Physics: Conference Series, vol. 1108, 012071. 2018.
[13] Drewnowski, A., & Fulgoni, V. (2005). Nutrient profiling of foods: creating a nutrient-rich food index. Nutrition Reviews, 66(1), 23-39. DOI: 10.1111/j.1753-4887.2008.00011.x
[14] Face recognition using deep learning CNN in python [On line]. Available: https://thinkingneuron.com/face-recognition-usingdeep learning-cnn-in-python/. [Accessed: 11-Dec-2021].
[15] Tufano, M., Kok, E., Tekinerdogan, B., Feskens, E.J., & Camps, G. (2020). A deep learning approach for the classification of food nutritional composition. Advances in Nutrition, 11(6), 1528-1541. DOI: 10.1093/advances/nmaa
[16] Berrigan, D., Lavigne, J.A., Perkins, S.N., & Hursting, S.D. (2004). Diet and cancer prevention: Energy balance and calorie restriction in human cancer. Annual Review of Nutrition, 24(1), 173-201.
[17] I. G. S. M. Diyasa, A. Fauzi, M. Idhom and A. Setiawan, "Multi face Recognition for the Detection of Prisoners in Jail using a Modified Cascade Classifier and CNN", 2020 2nd International Conference on Science & Technology, 1884, 012005, 2021, doi:10.1088/1742- 6596/1844/1/012005.
[18] Oliinyk, A., Fedorchenko, I., Stepanenko, .Rud M., Goncharenko, D. Implementation of evolutionary methods of solving the travelling salesman problem in a robotic warehouse // Lecture Notes on Data Engineering and Communications Technologies, 2021, 48, P. 263–292.
[19] P. Kamencay, M. Benco, T. Mizdos, R. Radil, "A New Method for Face Recognition using Convolutional Neural Network", Advances in Electrical and Electronic Engineering, vol. 15, no. 4, pp. 663.672, 2017, DOI: 10.15598/aeee.v15i4.2389.
[20] Fedorchenko, I., Oliinyk, A., Stepanenko, Zaiko, T., Korniienko S., Kharchenko, A. Construction of a genetic method to forecast the population health indicators based on neural network models // Eastern European Journal of Enterprise Technologies, 2020, 1 (4-103), P. 52–63. DOI: 10.15587/1729-4061.2020.197319.
[21] M. Alghaili, Z. Li, H. Ali, "FaceFilter: Face Identification with Deep Learning and Filter Algorithm", Hindawi Scientific Programming, 7846264, pp. 1-9, 2020, doi.org/10.1155/2020/7846264. [24] N. S. Irjanto and N. Surantha, "Home Security System with Face Recognition based on Convolutional Neural Network", International Journal of Advanced Computer Science and Applications, vol. 11, no. 11, p.p 408- 412, 2020.
[22] N. S. Irjanto and N. Surantha, "Home Security System with Face Recognition based on Convolutional Neural Network", International Journal of Advanced Computer Science and Applications, vol. 11, no. 11, p.p 408-412, 2020.
[23] M. Srivasan and N. Ravichandran, "A new technique for Face Recognition using 2D-Gabor Wavelet Transform with 2D-Hidden Markov Model approach", International Conference on Signal Processing, Image Processing and Pattern Recognition 2013, ICSIPR 2013, vol. 1, p.p 151- 156.
[24] J. Bobulski, "Multimodal face recognition method with two-dimensional hidden Markov model", Bulletin of the Polish Academy of Sciences Technical Sciences, vol. 65, no. 1, p.p 121-128, 2017.
[25] J. Moreano and N. Palomino, "Global Facial Recognition Using Gabor Wavelet, Support Vector Machines and 3D Face Models", Journal of Advances in Information Technology, vol. 11, no. 3, p.p 143-148, 2020.
[26] Z. Rustam and R. Faradina, "Face Recognition to Identify Look-Alike Faces using Support Vector Machine", Journal of Physics: Conference Series, vol. 1108, 012071, 2018.