

A SYSTEMATIC REVIEW OF CONVOLUTIONAL NEURAL NETWORK ARCHITECTURES FOR COMPUTER VISION APPLICATIONS

Maitrakkumar Patel¹, Jaiprakash Narain Dwivedi²

^{1,2}Department of Information Technology, Parul Institute of Engineering and Technology, Parul University, Vadodara, Gujarat, India

¹maitrak2001@gmail.com

Keywords: CONVOLUTIONAL NEURAL NETWORKS, COMPUTER VISION, DEEP LEARNING, CNN ARCHITECTURES, IMAGE CLASSIFICATION

Abstract

Convolutional Neural Networks (CNNs) have contributed to significant progress in computer vision activities like image classification, object detection, segmentation, and autonomous navigation. This paper is a systematic review of CNN architectures that were posted between 2012 and 2025, examining their development, uses, effectiveness, restrictions and future tendencies. The results also emphasize the move to efficient, scalable, and deployable frameworks, where the emphasis is on lightweight and hybrid frameworks in real-time and edge applications, which can guide the development of intelligent vision systems in the future.

1. Introduction

Computer vision is one of the main areas of artificial intelligence through which machines are able to read visual information to be used in healthcare, surveillance, robotics, and smart transportation [1]. The escalating development of image and video data in industries has raised the need to have the correct and scalable visual processing solutions [2]. This has been especially based on the technique of deep learning, specifically Convolutional Neural Networks (CNNs), which learns hierarchical visual features automatically and performs better than manually created methods [3]. The CNN architectures over the last decade have been developed to be a

sophisticated architecture, which is highly optimized to be able to accomplish image classification, object detection, and segmentation with high accuracy [4]. Nonetheless, due to the increasing variety of architectures, choosing a model becomes a difficult task, as they differ in terms of complexity, performance, and applicability in different activities [5]. To resolve this problem, the current research paper performs a systematic literature review of CNN architectures in computer vision, focusing on the architectural trends, their performance trends, and future research direction [6-8].

Table 1: Summary of Existing Review Studies on CNN Architectures in Computer Vision

Authors	Year	Focus Area	Application Domain	Key Contribution	Reported Limitation
Yoo, H. J. [9]	2015	CNN in Computer Vision	General computer vision tasks	Early overview of deep CNN adoption in vision systems	Limited coverage due to early-stage developments
Zhao et al. [10]	2024	CNN developments in vision systems	Broad computer vision applications	Comprehensive review covering modern CNN architectures and trends	Rapid model evolution may outdate comparisons quickly
Sarraf et al. [11]	2021	Deep learning architectures	General vision applications	Detailed survey of deep learning models for image processing	Focus broader than CNN-specific analysis

Valaboju [12]	2015	Neural network architectures	General neural network studies	Overview of neural network architectures and learning models	Limited focus on computer vision applications
Li et al. [13]	2021	CNN-based vision systems in farming	Animal farming and agricultural monitoring	Demonstrates CNN use in precision farming and livestock monitoring	Domain-specific focus limits generalization
Aloysius & Geetha [14]	2017	Deep CNN models	General image processing tasks	Review of CNN learning mechanisms and architecture performance	Limited coverage of modern architectures

1.1. Background of CNN in Computer Vision

Convolutional Neural Networks are one of the most significant innovations in computer vision as they allow automatic hierarchical learning of features and eliminate the shortcomings of conventional handcrafted approaches [15-18]. With time, CNNs have been extended to other applications such as image classification, medical diagnosis, autonomous navigation and surveillance [19]. This part gives a brief overview of the essential elements, history of development, and uses of CNNs in computer vision [20].

1.1.1. Fundamentals of Convolutional Neural Networks: The CNNs are deep learning algorithms, which are image-specific to utilize the spatial correlations of the images [21]. They consist of multiple layers that are involved in processing the raw pixel data in a sequential manner to create understandable forms that can be utilized to decide [22]. The significant components of CNNs include convolution layers, pooling layers, activation functions and fully connected layers [23].

- Convolution Layers

The key component of CNNs is the convolution layers which apply learnable filters depending on features like edges, textures and object patterns [24]. The higher levels obtain more abstract representations whereas the sharing of parameters provides the efficiency to compute without spatial detail loss.

- Pooling Layers

The pooling layers also decrease the dimension of feature maps, which reduces the cost of computing, and enhances generalization [25]. Max and average pooling and other operations are operations that ensure robustness to small change in images and reduce overfitting by concentrating on the stronger features.

- Activation Functions

Non-linearity is achieved by activation functions and allows CNNs to capture more complicated patterns [26]. ReLU is popular because it is computationally efficient and also overcomes the vanishing gradient problem, with some variants, or extensions, like Leaky ReLU and ELU, providing even better training stability.

- Fully Connected Layers

The full connection layers, which are normally located close to the network output, fuse learnt features to final classification or regression [27]. Contemporary CNN designs tend to decrease the use of massive fully connected layers to decrease the number of parameters and computation cost.

1.2. Evolution of CNN Architectures

The CNN architectures have changed a lot with each generation that overcame the limitations of the previous ones and enhanced the performance, scalability, and computational efficiency [28].

1. **LeNet:** Among the earliest CNNs, which were also used successfully in handwritten digit recognition but were not scalable because of shallow architecture [29].
2. **Alex Net:** In the 2012 ImageNet challenge, breakthrough success was obtained, with the introduction of deeper networks, training on GPUs, ReLU activation, dropout, and data augmentation, speeding up the use of CNNs.

3. VGGNet: It was shown that more profound networks with small convolution filters are better learners of features, but at high computational cost.
4. Google Net (Inception Network): Added multi-scale feature extraction in layers, which minimizes the parameters and maximizes the accuracy.
5. ResNet: Limit degradation in deep networks by adding residual connections, in order to train very deep architectures [30].
6. Dense Net: Gradient flow is improved and reuse of features by linking layers closely, which allows a high level of performance with a low number of parameters.
7. Efficient Net: Scaling of proposed compound in terms of network depth, width, and scale to be able to achieve high performance with better efficiency.
8. Vision Transformers (Optional Extension): Substitute convolutions with attention mechanisms, which learn about the world globally and perform competitively with training on large datasets.

2. Methodology

The paper adopts the systematic review research methodology as it evaluates the developments of Convolutional Neural Networks architectures in computer vision and provides exhaustive coverage, clear screening, and systematic synthesis of previous research results.

2.1 Literature Search Strategy and Study Selection

Keywords were searched in major databases such as Scopus, Web of Science, IEEE Xplore, and ScienceDirect with one of the keywords being CNN architectures and computer vision tasks. The studies that were found to be published in the last five years (2012-2025) were taken to be reflective of recent changes. English-language articles in journals only were chosen and the duplicate or irrelevant records were eliminated in initial screening. The first workflow in the search and filtering process is shown in Figure 1.

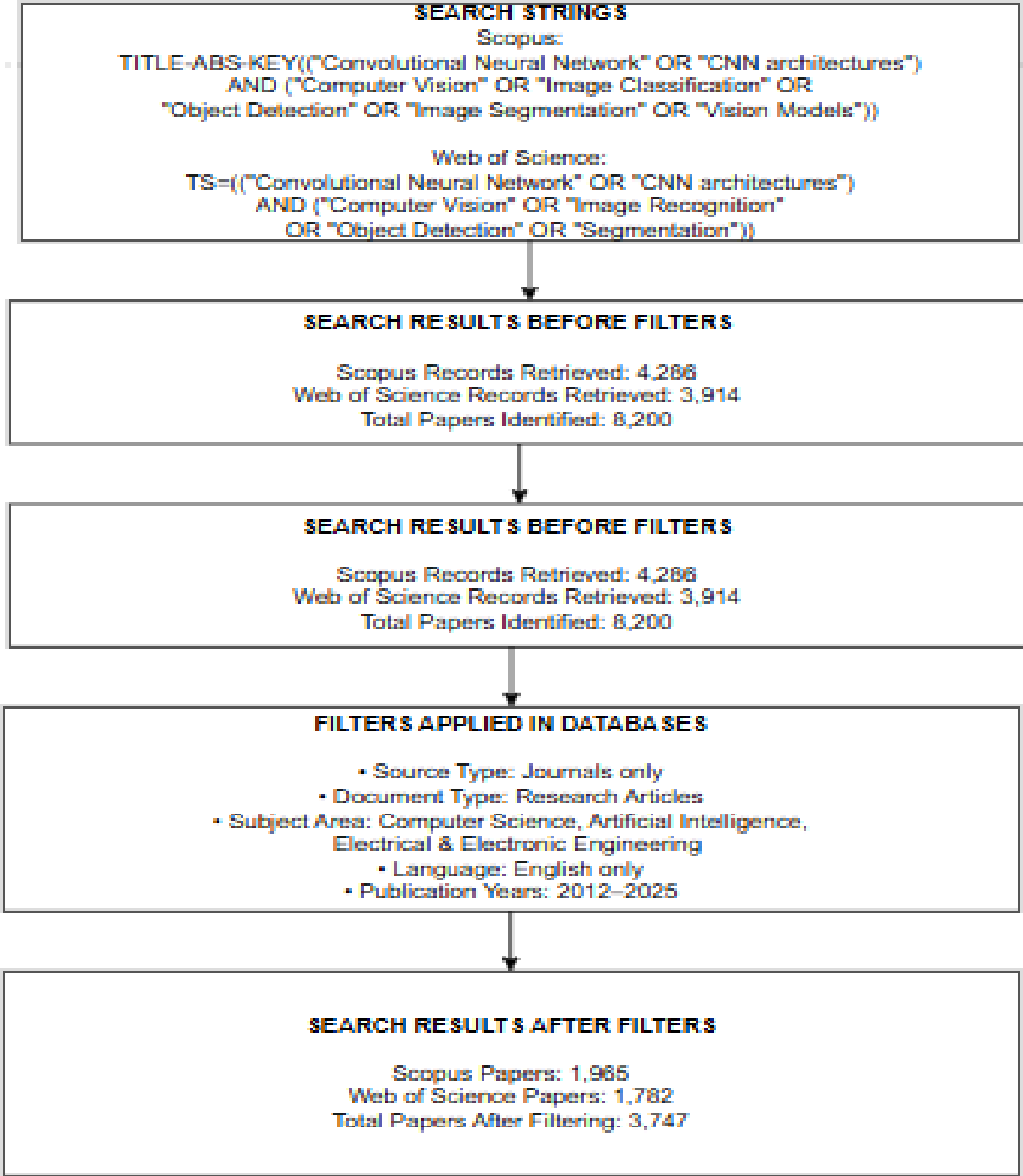


Fig 1: Search string depicting initial screening flowchar

.2.2 Screening and Eligibility Assessment

Titles and abstracts followed after a removal of duplicates to select irrelevant literature. Full-text analysis made sure that the studies that covered CNN architecture construction or computer vision were included. The ultimate selection was done as per PRISMA, and the workflow is presented in Figure 2.

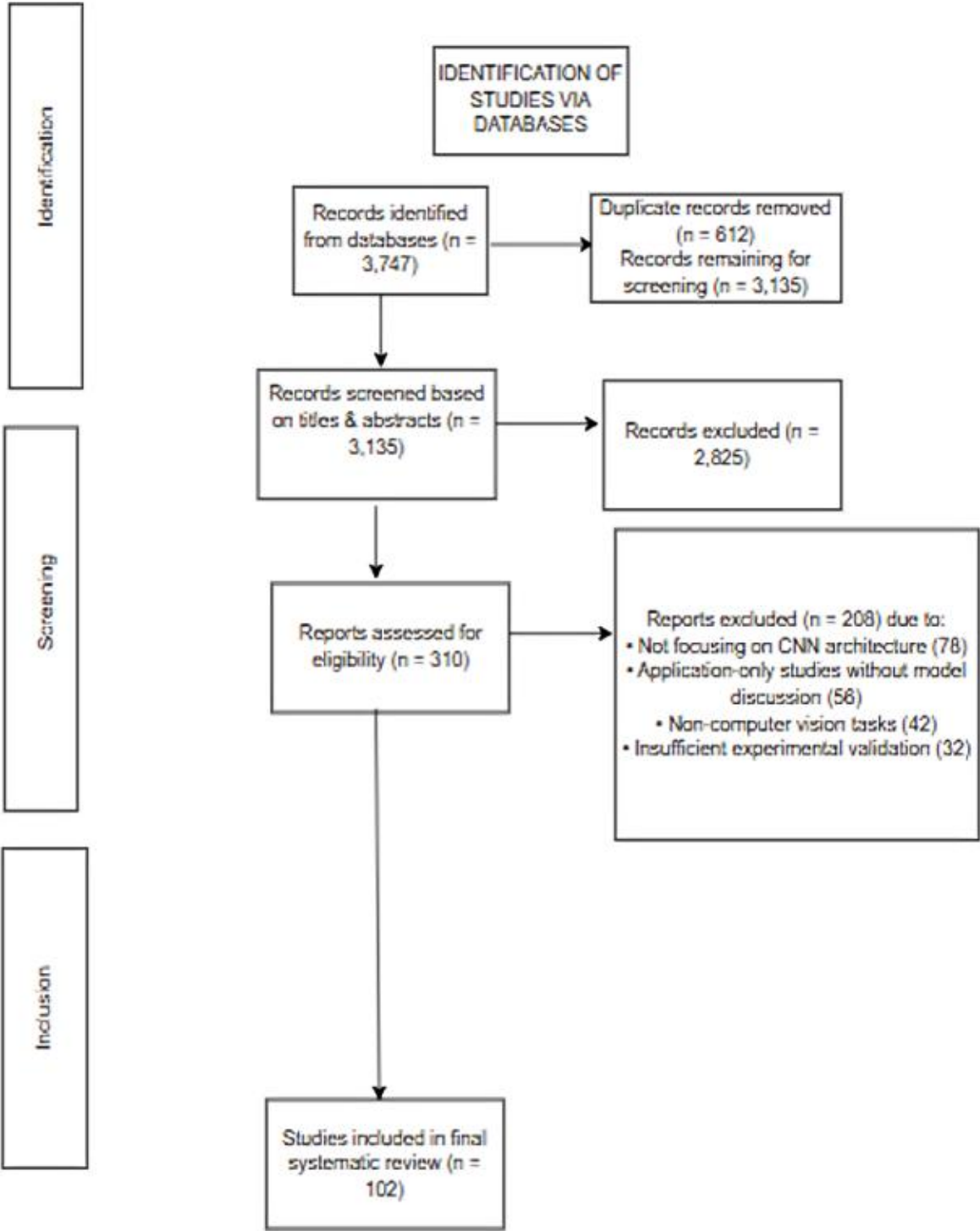


Fig 2 PRISMA flowchart depicting final screening

2.3 Data Extraction and Study Classification

The important information, such as the type of architecture, field of application, data sets, contributions, and limitations, have been obtained by means of the chosen studies. Research has been categorized into fields of image classification, object detection, medical imaging, autonomous driving and face recognition to allow systematic synthesis.

2.4 Architecture Comparison and Performance Synthesis

Past research findings were reported and synthesized to determine comparative results of architectures based on accuracy, computational efficiency, scalability, and deployment suitability which reveal trade-offs between CNN models.

2.5 Analysis of Strengths, Limitations, and Research Trends

The shortlisted studies were also subject to further analysis to find out the prevalent strengths, technical difficulties, and research trends, which served to synthesize limitation and future directions presented in the results.

2.6 Quality Assessment of Selected Studies

In order to be consistent, studies were considered according to the clarity of architectural description, the transparency of datasets, and experimental validation and only those studies that met the quality criteria were used in the final analysis.

3. Results And Discussion

This paragraph highlights CNN developments, applications, performance, and limitations that were found in the systematic review. It also points out the major challenges and future research prospects of computer vision.

3.1 Summary of Reviewed CNN Architecture Studies

Table 2 presents the development of CNN architectures, where initial networks such as AlexNet and VGGNet had high performance but need a significant amount of computational resources, with later networks such as ResNet and DenseNet having a more stable training and feature reuse.

Table 2: Summary of Representative CNN Architecture Studies for Computer Vision

Ref.	Year	CNN Architecture	Application	Dataset Used	Key Contribution	Limitation
Study A	2015	AlexNet	Image Classification	ImageNet	Demonstrated deep CNN superiority	Large parameter size
Study B	2016	VGGNet	Object Recognition	ImageNet	Improved feature learning depth	High computational cost
Study C	2017	ResNet	Image Classification	ImageNet	Introduced residual connections	Increased training complexity
Study D	2018	DenseNet	Object Detection	COCO	Feature reuse improves accuracy	Memory consumption
Study E	2019	MobileNet	Mobile Vision Tasks	ImageNet	Lightweight architecture	Slight accuracy reduction
Study F	2020	EfficientNet	Image Classification	ImageNet	Balanced scaling strategy	Tuning complexity
Study G	2022	CNN Hybrid Models	Medical Imaging	Chest X-ray Dataset	Improved disease detection	Requires large training data

More recent architectures, such as MobileNet, EfficientNet, and hybrid CNN models, emphasize on computational efficiency and domain-specific applications, but the issues associated with data requirements and model size persist.

3.2 CNN Architectures Across Vision Applications

Table 3 indicates the ubiquity of CNN architectures in large vision problems, with ResNet and EfficientNet models showing substantial progress in classifying, detecting, medical imaging, autonomous driving, and face recognition.

Table 3 CNN Usage Across Computer Vision Domains

Application Domain	Frequently Used Architectures	Major Outcome Reported in Literature	Research Challenge
Image Classification	ResNet, EfficientNet	High classification accuracy	Dataset bias
Object Detection	Faster R-CNN, YOLO, SSD	Real-time detection improvement	Hardware dependency
Medical Imaging	DenseNet, ResNet	Improved diagnostic support	Limited labeled data
Autonomous Driving	CNN + Detection Models	Enhanced environment perception	Real-time reliability
Face Recognition	CNN variants	High identity recognition accuracy	Privacy issues

Nevertheless, there are still challenges such as bias in the dataset, hardware requirements, scarcity of labeled medical data, real-time system reliability, and privacy issues in the recognition systems.

3.3 Reported Strengths and Limitations of CNN Architectures

Table 4 indicates that such architectures as VGGNet and ResNet have high performance and can do deep training, but they are also very demanding in terms of computational resources.

Table 4 Strengths and Limitations Observed in Literature

Architecture	Strengths Reported	Limitations Reported
VGGNet	Strong baseline performance	Very large model size
ResNet	Enables deep networks	Higher training complexity
DenseNet	Efficient gradient propagation	High memory requirement
MobileNet	Lightweight and fast	Accuracy trade-off
EfficientNet	Good performance-efficiency balance	Complex model scaling

DenseNet enhances feature reuse but is larger in terms of memory requirements, MobileNet allows efficient implementation with accuracy trade-offs and EfficientNet is a balance between performance and efficiency, emphasizing that all models are trade-offs among accuracy, efficiency and computational cost.

3.4 Emerging Research Directions in CNN Literature

Table 5 suggests that the recent trends of the CNN research are centered around lightweight models to deploy on mobile and edges, integrate with transformer, and self-supervised learning to minimize reliance on labeled data.

Table 5: Emerging Trends Identified in Literature

Trend	Literature Observation	Expected Impact
Lightweight CNN models	Increased mobile deployment research	Edge AI growth
CNN-Transformer hybrids	Improved feature learning	Next-gen architectures
Self-supervised learning	Reduced data labeling effort	Faster adoption
Real-time vision systems	Growth in autonomous systems	Industry applications

The trend in the future will be towards efficiency, scalability, and real-time implementation of intelligent vision systems.

4. Conclusion

In this review, the development of CNN architectures in advancing computer vision performance is emphasized, and the way newer models are more efficient, scalable, and capable of deployment has been criticized. Nevertheless, issues of computational cost, training complexity, data constraints, and real-time execution still exist, and possible future workforce is in lightweight, hybrid and data efficient models that will allow practical and scalable vision systems.

Acknowledgement

The authors sincerely thank the Department of Information Technology, Parul Institute of Engineering and Technology, Parul University, Vadodara, for providing the necessary academic support and research environment. The authors also express their gratitude to faculty members and researchers whose contributions and guidance supported the successful completion of this study.

References

- [1] Bhatt, D., Patel, C., Talsania, H., Patel, J., Vaghela, R., Pandya, S., ... & Ghayvat, H. (2021). CNN variants for computer vision: History, architecture, application, challenges and future scope. *Electronics*, 10(20), 2470.
- [2] Khamparia, A., & Singh, K. M. (2019). A systematic review on deep learning architectures and applications. *Expert Systems*, 36(3), e12400.
- [3] 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(8), 5455-5516.
- [4] Krichen, M. (2023). Convolutional neural networks: A survey. *Computers*, 12(8), 151.
- [5] Mauricio, J., Domingues, I., & Bernardino, J. (2023). Comparing vision transformers and convolutional neural networks for image classification: A literature review. *Applied Sciences*, 13(9), 5521.
- [6] Alsajri, A., & Hacimahmud, A. V. (2023). Review of deep learning: Convolutional neural network algorithm. *Babylonian Journal of Machine Learning*, 2023, 19-25.
- [7] Khan, S., Rahmani, H., Shah, S. A. A., Bennamoun, M., Medioni, G., & Dickinson, S. (2018). A guide to convolutional neural networks for computer vision.
- [8] Taye, M. M. (2023). Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions. *Computation*, 11(3), 52.
- [9] Yoo, H. J. (2015). Deep Convolution Neural Networks in Computer Vision: a Review: a Review. *IEIE Transactions on Smart Processing & Computing*, 4(1), 35-43.
- [10] Zhao, X., Wang, L., Zhang, Y., Han, X., Deveci, M., & Parmar, M. (2024). A review of convolutional neural networks in computer vision. *Artificial Intelligence Review*, 57(4), 99.
- [11] Sarraf, A., Azhdari, M., & Sarraf, S. (2021). A comprehensive review of deep learning architectures for computer vision applications. *American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS)*, 77(1), 1-29.
- [12] Valaboju, Y. (2015). A LITERATURE REVIEW ON NEURAL NETWORK ARCHITECTURES. *Journal of Interdisciplinary Cycle Research*, 7(2).
- [13] Li, G., Huang, Y., Chen, Z., Chesser Jr, G. D., Purswell, J. L., Linhoss, J., & Zhao, Y. (2021). Practices and applications of convolutional neural network-based computer vision systems in animal farming: A review. *Sensors*, 21(4), 1492.
- [14] Aloysius, N., & Geetha, M. (2017, April). A review on deep convolutional neural networks. In *2017 international conference on communication and signal processing (ICCSP)* (pp. 0588-0592). IEEE.
- [15] Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). A review of the use of convolutional neural networks in agriculture. *The Journal of Agricultural Science*, 156(3), 312-322.
- [16] E. T. Tan, L. Marinelli, Z. W. Slavens, K. F. King, and C. J. Hardy, "Improved correction for gradient nonlinearity effects in diffusion-weighted imaging," *Journal of Magnetic Resonance Imaging*, vol. 38, no. 2, pp. 448–453, 2013.
- [17] J. G. Sled, A. P. Zijdenbos, and A. C. Evans, "A nonparametric method for automatic correction of intensity nonuniformity in MRI data," *IEEE Transactions on Medical Imaging*, vol. 17, no. 1, pp. 87–97, 1998.
- [18] L. Khedher, J. Ramírez, J. M. Górriz, A. Brahim, and F. Segovia, "Alzheimer's Disease Neuroimaging Initiative, et al., Early diagnosis of Alzheimer's disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images," *Neuro-computing*, vol. 151, pp. 139–150, 2015.
- [19] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," 2017, <https://arxiv.org/abs/1710.09829>
- [20] P. Priyadarshini and B. S. E. Zoraida, "Computed tomography imaging lung cancer segmentation using TGVMFCM and classification using deep learning techniques," *Journal of Green Engineering*, vol. 10, no. 10, pp. 9184–9203, 2020.

- [21] J. E. Galvin, "Prevention of Alzheimer's disease: lessons learned and applied," *Journal of the American Geriatrics Society*, vol. 65, pp. 2128–2133, 2017. *Mathematical Problems in Engineering* 11
- [22] Emadi, M. Ozen, and A. Abdi, "A hybrid model to study how late long-term potentiation is affected by faulty molecules in an intraneuronal signaling network regulating transcription factor CREB," *Integrative Biology*, vol. 14, pp. 111–125, 2022
- [23] Association, "2018 Alzheimer's disease facts and figures," *Alzheimer's and Dementia*, vol. 14, pp. 367–429, 2018.
- [24] Naranjo-Torres, J., Mora, M., Hernández-García, R., Barrientos, R. J., Fredes, C., & Valenzuela, A. (2020). A review of convolutional neural network applied to fruit image processing. *Applied Sciences*, 10(10), 3443.
- [25] Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. *Computational intelligence and neuroscience*, 2018(1), 7068349.
- [26] Rawat, W., & Wang, Z. (2017). Deep convolutional neural networks for image classification: A comprehensive review. *Neural computation*, 29(9), 2352-2449.
- [27] Chai, J., Zeng, H., Li, A., & Ngai, E. W. (2021). Deep learning in computer vision: A critical review of emerging techniques and application scenarios. *Machine Learning with Applications*, 6, 100134.
- [28] Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., & Alsaadi, F. E. (2017). A survey of deep neural network architectures and their applications. *Neurocomputing*, 234, 11-26.
- [29] Chen, L., Li, S., Bai, Q., Yang, J., Jiang, S., & Miao, Y. (2021). Review of image classification algorithms based on convolutional neural networks. *Remote Sensing*, 13(22), 4712.
- [30] Nassif, A. B., Shahin, I., Attali, I., Azzeh, M., & Shaalan, K. (2019). Speech recognition using deep neural networks: A systematic review. *IEEE access*, 7, 19143-19165.