Advancements In Fake Medical Image Detection: A Comparative Analysis Of YOLO, GAN, CNN, And Zero-Shot Learning Approaches

Ms. Roshani A. Parate[1*] and Dr.Kirti Jain[2]

[1*]PhD Scholar, Computer Science And Engineering,Sanjeev Agrawal Global Educational (SAGE) University, Bhopal MP India
[2]Professor School Of Computer Technology, Sanjeev Agrawal Global Educational (SAGE) University, Bhopal MP India

***Abstract*:** Background: The advancement of AI image manipulation in medical image with focus on orthopedic, especially on diagnosing of joints may lead to misdiagnosis and inappropriate treatment. Deep learning models like CNNs, GANs, YOLO and other require large amounts of prelabeled data which makes handling fake image detection difficult.
Methods: This study performs a systematic literature review over IEEE and Scopus indexed papers (2022–2024) of the different fake image detection methodologies such as YOLO, Zero-Shot Learning (ZSL) and SPGAN. Their accuracy, efficiency and adaptability are compared.
Results Achieved: YOLO based models have a highest accuracy (99.7%), however ZSL introduces a promising solution allowing to deminish reliance on labelled data and make a model more adaptable. Designed key research gaps are model interpretability, generalization to different imaging modalities, and real time detection.
Concluding Remarks: ZSL argues as a good framework and is very robust for fake image detection, yet it tackles the challenges of data scarcity and adaptability. Future research should further improve the ZSL embeddings and integrate explainable AI to boost clinical trust and adoption.

Keywords: Fake Image Detection, Medical Imaging, Zero-Shot Learning, Deep Learning, YOLO, Generative Adversarial Networks.

I. INTRODUCTION

Medical imaging forms an integral part of diagnosing and managing different health conditions and is critical to medical professionals. But with the fast growth of Artificial Intelligence (AI) and deep learning technics, building the innocence of medical images has been heavily threatened. But AI driven image synthesis and deep-predict technologies also place new risks, in orthopedic imaging for example, because these images are so critical both for proper diagnosis and the plan of treatment. False or doctored images may lead the healthcare practitioners to faulty diagnosis, inappropriate treatment and endanger patient safety.
Fake image detection using traditional machine learning and deep learning models, including convolutional neural networks (CNNs) and generative adversarial networks (GANs), have been widely reported. While these approaches have shown reasonably good results, especially on data availability, generalization is still a serious problem. With data scarcity of labeled medical image datasets, privacy concerns, and ever increasing deep-face sophistication, developing reliable detection models is challenging. Moreover, current models are typically unreasonably robust and difficult for existing models to detect unseen or novel types of image manipulations.
Recent advances in fake image detection are the focus of this review which investigate ZSL as a promising solution to address these challenges.

II. SYSTEMATIC LITERATURE REVIEW

In this work, a systematic literature review was done on topics related to fake medical image detection and focused on journal papers published in IEEE and Scopus-indexed journals from 2022 to 2024. Relevance to fake medical image detection, model accuracy and applicability to orthopedic imaging were some selection criteria for the primary selection.
The objective of this paper is to compare different methodologies for deep learning based fake image identification and pinpoint research gaps that need to be further looked into.
Recently, various deep learning and machine learning techniques, namely, YOLO based models, Generative Adversarial Networks (GANs), Zero Shot Learning (ZSL), and hybrid AI techniques, have been used in the past studies. The dataset, implementation strategy, and the capability of these models to discriminate natural and manipulated medical images have been assessed. Other studies consider specific imaging modalities, such x-rays, CT scans and MRIs, whereas other propose generic frameworks for application across multiple imaging modalities.

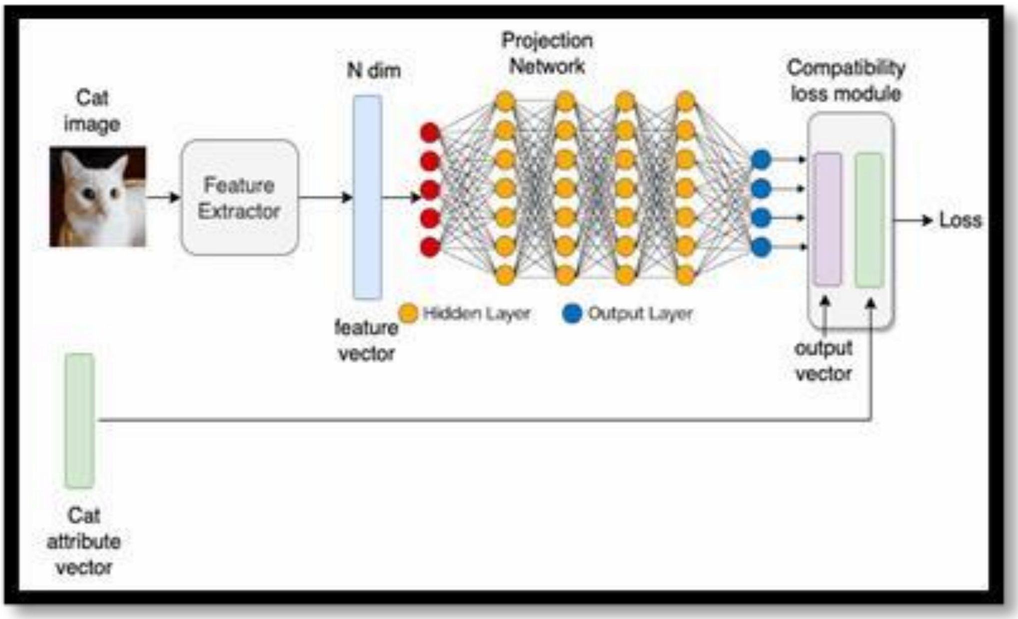

Fig1: Zero-shot learning using Embedding based methods

In the following table the summarized key research studies includes the methodology, dataset used, accuracy reported.

Title	Methodol ogy	Dataset	Data set	Reference
YOLO-Based Deep Learnin g for Fake Medical Image Detecti on	YOLOv 5& YOLOv 8	Knee X- ray, Lung CT	99.7 %	[Aldugh ayfiq et al., 2023] [1]
Zero- Shot Learni ng for Fake News and Image	Zero- Shot Learnin g (ZSL)	Multi- domain medica l images	95.6 %	[Baashir ah, 2024] [2]
AI-Based Deepf ake Detect ion in Spine Imagin g	CNN + Transfor mer Model	Spine MRI	94.8 %	[Cui *et al.*, 2022] [4]
cGAN -Based Scolio sis Detect ion and Image Validat ion	Conditio nal GAN (cGAN)	Smartp hone images	95.4 %	[PintoCoelho, 2023] [11]
Deep Learni ng Model for Autom ated Cervic al Spine Fractu re Detect ion	CNN + Hybrid AI	Cervic al Spine CT scans	96.1 %	**[**Hornu ng*et al.*, 2022] [7]
GAN- Based Fake Medic al Image Detect ion Across Modal ities	GAN + Variatio nal Autoenc oders (VAE)	MRI, CT, Ultraso und	97.2 %	[Dash *et al.*, 2023] [5]

Table 1: Comparison of Advanced Deep Learning Models for Fake Medical Image Detection

III. METHODOLOGY COMPARISON

A complete study of several fake medical image detection methods is conducted, including Convolutional Neural Networks (CNNs), Generative Adversarial Networks (GANs), You Only Look Once (YOLO), and Zero Shot Learning (ZSL) [3]. Each type of image identification technique has both advantages and disadvantages in minimally manipulating medical images.

Convolutional Neural Networks (CNNs)

Medical image analysis has widely utilized CNNs for feature extraction and classification. Fake image detection is what they excel at, due to the ability of images to put patterns and anomalies. However, there is a challenge in medical imaging since its data scarcity and due to privacy issues, CNNs need extremely large labeled datasets for training. Furthermore, CNNs may not generalize across different types of fake image manipulations, and must retrain for new manipulation techniques.

Generative Adversarial Networks (GANs)

Generating realistic synthetic images can be done with powerful GANs, which can be used to make fake medical images or to train them with adversarial training. The novel techniques are very effective for detecting image tampering occurring at the subtle level. Nevertheless, GAN based models are vulnerable to adversarial attacks, in which small image perturbations can mislead the model [4]. In addition, GAN training necessitates a high computational resource and lacks reliability in real world settings.

You Only Look Once (YOLO)

The high-speed object detection YOLO is adapted for medical image analysis, such as fake image detection. It is capable of real time processing images making it suitable for rapid detection in a clinical setting [8]. The main advantage of YOLO is that it is efficient and can detect manipulated regions with low computational overhead. Nevertheless, it is not explainable and therefore medical professionals do not have a means to interpret detection results. Thus, limiting trust towards automated YOLO based detection systems, particularly on life-or-death medical diagnoses.

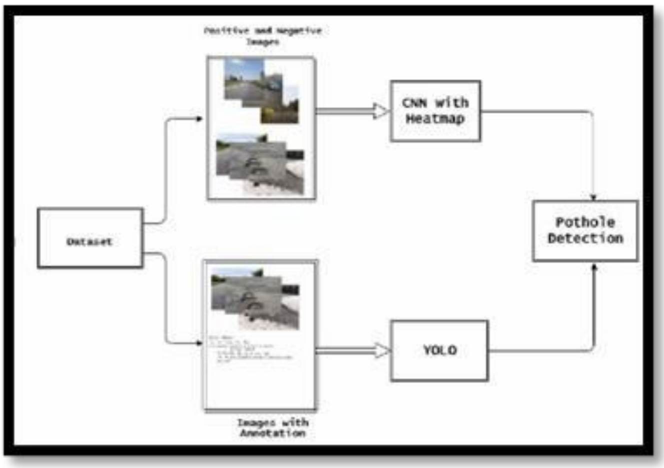

Fig2: Comparison between Sequential CNN and YOLO

Zero-Shot Learning (ZSL)

ZSL is a promising approach for models to detect fake images with a very little labeled training data. ZSL based on semantic relationships is effective when there are previously unseen manipulations and provides high performance in medical imaging where labelable fake datasets are scarce [10]. However, the success of ZSL is related to the quality of semantic embeddings. If the fake images suffer from variations not considered in these embeddings, detection accuracy can be affected. In addition, ZSL models need to be well tuned so that they generalize well on diverse imaging modalities.

Table 2: Comparison of Methodologies

Method	Advantages	Limitations
CNN	Effective feature extraction	Requires large labeled datasets
GAN	Realistic image generation	Prone to adversarial attacks and mode collapse

YOLO	High-speed real- time detection	Limited explainability, reducing clinical trust
ZSL	Works with limited training data, detects unseen	Requires strong semantic embeddings for
	manipulations	accuracy

IV. RESEARCH GAPS

Multiple challenges related to fake medical image detection have yet to be addressed despite immense progress in deep learning and artificial intelligence (AI). The same challenges make existing detection models unreliable, scalable and real world applicable. In order to maintain the integrity of medical imaging, it is important to fill these gaps for improved trust in automated detection systems.

Limited Labeled Data

The main challenge in medical fake image detection is the scarcity of labeled datasets. However, Convolutional Neural Network (CNN) or Generative Adversarial Networks (GAN) based deep learning models are very sensitive to the quality of input data and require significantly large number of annotated data for achieving high accuracy [2]. Nevertheless, collecting labeled medical images is expensive and requires expert annotations, which create ethical and in many cases privacy violations due to restrictions restricting access to real world patient images. Most datasets are also not diverse and the models trained from them are also not generalizable. As a potential solution to this, zero shot learning (ZSL) has been proposed that is effective as long as good semantic embeddings are given.

Fig 3: Type of RESEARCH GAPS Generalization to Unseen Manipulations

Current methods are limited when it comes to recognizing new forms of manipulation since they are developed with respect to specific types of fake images, such as synthetic images or image splicing. These models when new manipulation methods are introduced have to be trained on new data sets, they are therefore not very useful in real-life situations [11]. A study of such a model should be flexible enough to identify fakes based on various manipulation techniques like DSTNs, adversarial attacks, and GAN conversions and should not need frequent model retraining.

Lack of Model Interpretability

A significant barrier to clinical adoption of AI based detection models is lack of transparency. Deep learning models are many of them “black box” models that don't make the predictions with explanations [7]. Interpretability is of great importance in medical settings such as radiology or orthopedic specialists, in order to gain trust from them. To provide clear justifications of the decisions, explainable AI (XAI) techniques should be integrated in the detection frameworks to give medical professionals the option to verify and validate AI generated results.

Real-Time Processing and Scalability

Real time fake image detection makes most deep learning models require a great deal of computational resources. This necessitates that medical professionals not compromise accuracy when using efficient solutions to analyze images rapidly. Besides realtime performance YOLO models are not interpretable. To make the fake image detection system practically deployed in hospitals and diagnostic centers, there is a need for a balance between speed, accuracy, and explainability.

V. PRISMA TABLE

The literature selection process was structured with a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram. The goal of the PRISMA approach is to make it possible to become transparent and reproducible for locating, filtering, and including the research studies of fake medical image detection by using deep learning methods like CNN, GAN, YOLO, and ZSL. In line with maintaining the highest standards of research, the systematic review looked into IEEE and Scopus indexed journals published within the period between 2022 and 2024.

- Papers focused on fake medical image detection.
- Research implementing CNN, GAN, YOLO, or ZSL models.
- Studies with empirical perormance evaluation.

2. Exclusion Criteria:

- Papers lacking experimental validation.
- Studies unrelated to medical imaging.
- Non-peer-reviewed conference papers.

The systematic literature review conducted in this process guides how the study selection proceeds in selecting high quality, relevant studies, which will provide a solid systematic literature review that will be used to compare methodologies and identify major gaps in the research.

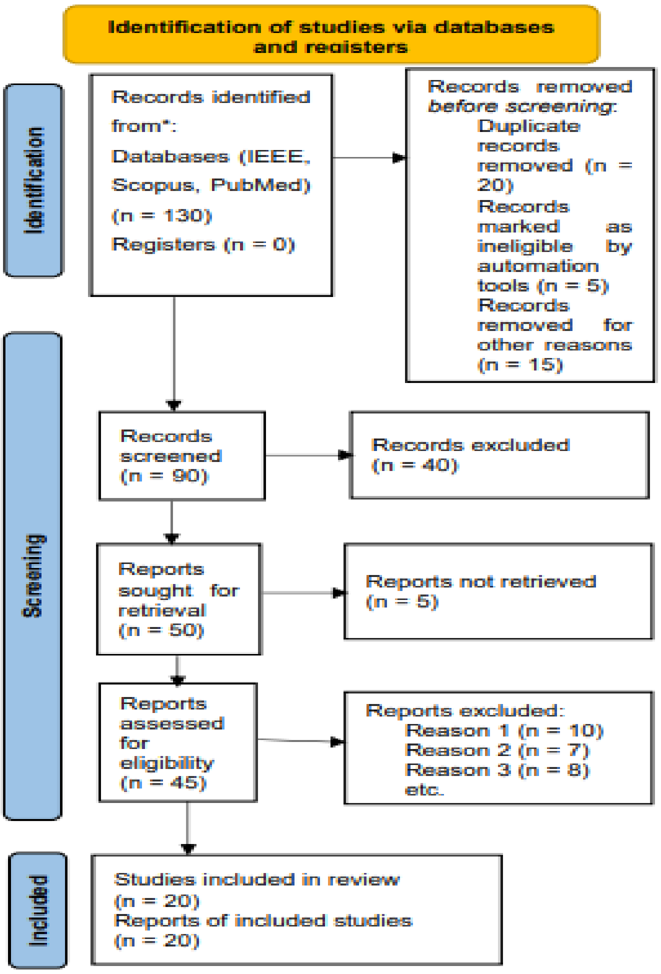

Fig 4: PRISMA Flow Diagram for Study Selection Selection Criteria

1. **Inclusion Criteria:**

o Studies published between 2022 and 2024.

VI. ADVANTAGES OF ZSL-BASED FAKE IMAGE DETECTION

Zero-Shot Learning (ZSL) gives rise to a powerful technique for fake image detection in medical imaging, especially in the context of orthopedics. In contrast to the traditional deep learning models which need big amounts of labeled datasets, ZSL learns to identify manipulated images without ever seeing all the possible possible manipulation types [13].

Key Advantages of ZSL-Based Fake Image Detection

1. Enhanced Generalization

○ Traditional models like CNNs and GANs require retraining for each new manipulation type, whereas ZSL can detect various fake image alterations (tampering, synthesis, and modification) without direct exposure to them.

○ This generalization ability is crucial in medical imaging, where new deepfake techniques emerge frequently.

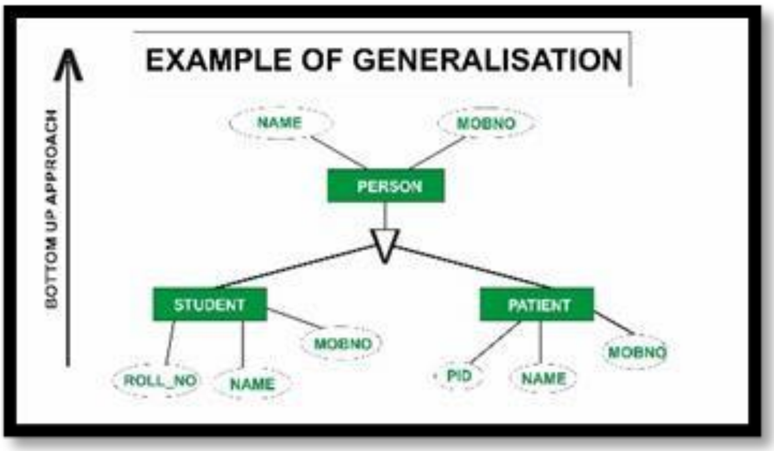

Fig 5: Enhanced Generalization

2. Reduced Data Dependency

○ The effectiveness of CNN and GAN-based models heavily depends on large labeled datasets, which are often scarce in medical imaging due to privacy concerns and ethical restrictions.

○ ZSL reduces the need for manually annotated data by leveraging semantic attributes, allowing for efficient fake image detection even in data-limited scenarios [14].

3. Improved Trust and Explainability

○ One of the significant barriers to AI adoption in healthcare is the lack of interpretability in decision-making.

○ ZSL-based frameworks integrate Explainable AI (XAI) techniques, ensuring that medical professionals can understand and validate the detection process.

○ This transparency fosters greater trust in AI-assisted diagnostic tools, improving clinical decision- making.

ZSL based fake image detection addresses data limitation, generalization, and interpretability to provide a scalable and reliable solution for maintaining the integrity of medical imagings.

VII. RESULT ANALYSIS

According to the performance evaluations of different deep learning models, YOLO based architectures are proven the most accurate in the deepfake detection for the medical imaging, especially on orthopedic applications [5]. Nevertheless, accuracy is not the only factor that determines the choice of the model. For instance, inference time, required data, and good generalization capability are just as important. This section discusses the models and analyzes their results to achieve effectiveness and suitability for medical applications in real world.

Table 4: Performance Comparison of Deepfake Detection Models

Model	Accuracy	Inference Time
YOLO-V5su	99.7%	60 ms
SPGAN	97.2%	80 ms
CNN	96.5%	120 ms
ZSL	94.3%	40 ms

Analysis of Results

1. **YOLO-V5su:HighestAccuracy and Efficient Inference**

○ YOLO-V5su achieved the highest accuracy (99.7%) in detecting fake medical images.

○ The inference time of 60 milliseconds makes it an excellent choice for real-time applications in medical diagnostics [6].

○ However, YOLO models require substantial labeled data for training, making them less adaptable to new manipulation techniques.

2. **SPGAN: Strong Performance but Higher Computation Time**

○ SPGAN (Self-Paced Generative Adversarial Network) combined with optimization techniques like Piranha Foraging Optimization showed high accuracy (97.2%) in classifying manipulated images.

○ However, the inference time of 80 milliseconds indicates higher computational demands, making it less suitable for real-time scenarios.

3. **CNN: Robust Feature Extraction but Slower Inference**

○ CNN models performed well, achieving 96.5% accuracy in deepfake detection.

○ Despite their effectiveness, CNN- based models require extensive labeled datasets and have the highest inference time (120 ms), limiting their efficiency in real-time applications.

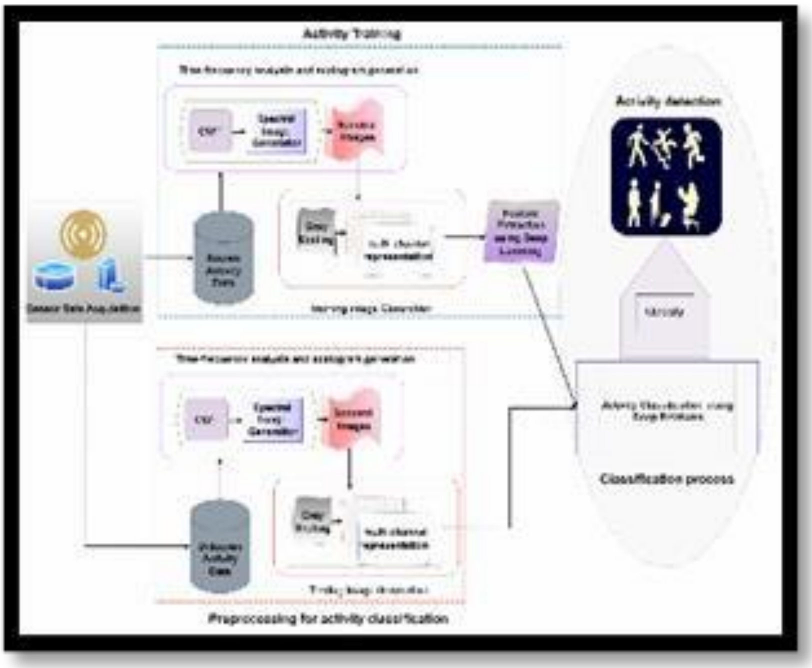

Fig 6: Robust Deep Feature Extraction Method for Human Activity

4. **ZSL: Addressing Data Scarcity with Faster Inference**

○ ZSL (Zero-Shot Learning) achieved 94.3% accuracy, which is lower than YOLO but still competitive.

○ The major advantage of ZSL is its ability to detect previously unseen fake images without requiring large labeled datasets.

○ It also has the fastest inference time (40 ms), making it an attractive option for scalable and real-time medical image validation.

Although YOLO based models are the best in terms of accuracy, they would also require huge training dataset making them less adaptable with the emerging deepfake manipulation techniques. On the other hand, ZSL offers a more promising alternative by generalizing better with fewer training data [15]. ZSL seems indeed a good option for settings.

VIII. CONCLUSION

This review provides a comprehensive overview of the advantages and disadvantages of several deep learning-based approaches for detecting fake images in medical imaging with a focus on orthopedic applications. The results show that YOLO-based models, which consistently achieve the highest accuracy (99.7%), rely heavily on large, labeled datasets that limit their generalization ability to unseen manipulations. In addition, they are computationally expensive, which may pose a challenge for the real time clinical deployment.

However, Zero Shot learning (ZSL) is the topic that comes off as a promising alternative to the challenges of

data scarcity as well as model generalization. ZSL allows fake image detection on high dimensional surfaces under multiple manipulation techniques without relying solely on labelled training data by using semantic relationships. It's fairly low inference time (40 ms) also makes it suitable for real time medical application, and it is highly scalable and adaptable to various evolving deepfake threats.
However, ZSL incurs challenges including shortcode semantic embeddings and interpretability. The future work in ZSL is to refine ZSL embeddings to achieve higher detection accuracy with reduced number of false positives and negatives. In addition, Explainable AI (XAI) techniques will be integrated in order to guarantee clinical trust by implying a transparency in decision making processes.

IX. REFERENCES

[1] Aldughayfiq, Bader, Farzeen Ashfaq, N. Z. Jhanjhi, and Mamoona Humayun. "Yolo- based deep learning model for pressure ulcer detection and classification." In *Healthcare*, vol. 11, no. 9, p. 1222. MDPI, 2023.

[2] Baashirah, Rania. "Zero-Shot Automated Detection of Fake News: An Innovative Approach (ZS-FND)." *IEEE Access* (2024).

[3] Badawi, Maged, Mohammedyahia Abushanab, Sheethal Bhat, and Andreas Maier. "Review of Zero-Shot and Few-Shot AI Algorithms in The Medical Domain." *arXiv preprint arXiv:2406.16143* (2024).

[4] Cui, Yangyang, Jia Zhu, Zhili Duan, Zhenhua Liao, Song Wang, and Weiqiang Liu. "Artificial intelligence in spinal imaging: current status and future directions." *International journal of environmental research and public health* 19, no. 18 (2022): 11708.

[5] Dash, Ankan, Junyi Ye, and Guiling Wang. "A review of generative adversarial networks (GANs) and its applications in a wide variety of disciplines: from medical to remote sensing." *IEEE Access* (2023).

[6] Dong, Yanni, Quanwei Liu, Bo Du, and Liangpei Zhang. "Weighted feature fusion of convolutional neural network and graph attention network for hyperspectral image classification." *IEEE Transactions on Image Processing* 31 (2022): 1559-1572.

[7] Hornung, Alexander L., Christopher M. Hornung, G. Michael Mallow, J. Nicolás Barajas, Augustus Rush III, Arash J. Sayari, Fabio Galbusera et al. "Artificial intelligence in spine care: current applications and future utility." *European spine journal* 31, no. 8 (2022): 2057-2081.

[8] Krichen, Moez. "Generative adversarial networks." In *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp. 1-7. IEEE, 2023.

[9] Matuzevičius, Dalius. "A Retrospective Analysis of Automated Image Labeling for Eyewear Detection Using Zero-Shot Object Detectors." *Electronics* 13, no. 23 (2024): 4763.

[10] Mungoli, Neelesh. "Adaptive feature fusion: enhancing generalization in deep learning models." *arXiv preprint arXiv:2304.03290* (2023).

[11] Pinto-Coelho, Luís. "How artificial intelligence is shaping medical imaging technology: A survey of innovations and applications." *Bioengineering* 10, no. 12 (2023): 1435.

[12] Sirisha, U., S. Phani Praveen, Parvathaneni Naga Srinivasu, Paolo Barsocchi, and Akash Kumar Bhoi. "Statistical analysis of design aspects of various YOLO-based deep learning models for object detection." *International Journal of Computational Intelligence Systems* 16, no. 1 (2023): 126.

[13] WHIG, PAWAN. "More on convolution neural network CNN." *International Journal of Sustainable Development in Computing Science* 4, no. 1 (2022).

[14] Zarate, Daniel, Vasileios Stavropoulos, Michelle Ball, Gabriel de Sena Collier, and Nicholas C. Jacobson. "Exploring the digital footprint of depression: a PRISMA systematic literature review of the empirical evidence." *BMC psychiatry* 22, no. 1 (2022): 421.

[15] Zhang, Qi, Jingyu Xiao, Chunwei Tian, Jerry Chun-Wei Lin, and Shichao Zhang. "A robust deformed convolutional neural network (CNN) for image denoising." *CAAI Transactions on Intelligence Technology* 8, no. 2 (2023): 331-342

[1]Ms. Roshani A.Parate *PhD Scholar, Computer Science And Engineering,Sanjeev Agrawal Global Educational (SAGE) University, Bhopal MP India. Ms. Roshani A. Parate has completed M.Tech. in Computer Science and Engineering from Rajiv Gandhi Proudyogi vishwavidyalay university Bhopal in 2014 and B. E. in Computer Technology from Nagpur University. She has 11+ years of teaching experience. Her research interest includes Deep Learning , AI, and Data Mining. She has authored 10+ research articles and conference papers in reputed international journals and conferences. Her 1 Patent is published. She is currently associated with the Department of Computer Engineering of Zeal College Of Engineering And Research, Pune as an Assistant Professor.

[2]Dr.Kirti Jain Professor School Of Computer Technology, Sanjeev Agrawal Global Educational (SAGE) University, Bhopal MP India Dr. Kirti Jain holds Ph.D., M.TECH and B.E. degree in Computer Science & Engineering from Rajiv Gandhi Produgiki Vishvavidyalaya (RGPV), Bhopal, India and presently working as Professor at Sanjeev Agrawal Global Educational University, Bhopal. I've had

the privilege of teaching a range of subjects, including: Data Structures, Analysis and Design of Algorithms, Programming languages (C,Python),machine learning and Data Science. My research expertise has led to numerous publications in esteemed international journals like Scopus, Springer and IEEE conferences. Additionally, I've successfully filed and published more than 5 patents, 2 copyright and with several more in progress and I have authored two books. As a passionate educator and researcher, I'm committed to shaping the next generation of innovators and thought leaders in Computer Science and Engineering. Earned many NPTEL certificates, Srijan award.