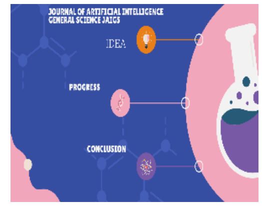
Vol.4, Issue 01, April 2024
Journal of Artificial Intelligence General Science JAIGS

Home page http://jaigs.org




# Advancements in Self-Supervised Learning for Remote Sensing Scene Classification: Present Innovations and Future Outlooks

**José Gabriel Carrasco Ramírez**

**Lawer graduated at Universidad Católica Andrés Bello. Caracas. Venezuela. / CEO, Quarks Advantage. Jersey City, United States. / Director at Goya Foods Corp., S.A. Caracas. Venezuela**

*Corresponding Author: José Gabriel Carrasco Ramírez

Doi: https://doi.org/10.60087/jaigs.vol4.issue1.p56

***ARTICLEINFO***
***Article History:***
*Received: 01.04.2024*
*Accepted: 15.04.2024*
*Online: 23.04.2024*
Keyword: self-supervised learning; representation learning; scene classification; remote sensing

*ABSTRACT*

Deep learning methodologies have significantly advanced the fields of computer vision and machine learning, enhancing performance across various tasks like classification, regression, and detection. In remote sensing for Earth observation, deep neural networks have propelled state-of-the-art results. However, a major drawback is their dependence on large annotated datasets, necessitating extensive human effort, especially in specialized domains like medical imaging or remote sensing. To mitigate this reliance on annotations, several self-supervised representation learning techniques have emerged, aiming to learn unsupervised image representations applicable to downstream tasks such as image classification, object detection, or semantic segmentation. Consequently, self-supervised learning approaches have gained traction in remote sensing. This article surveys the foundational principles of various self-supervised methods, focusing on scene classification tasks. We elucidate key contributions, analyze experimental setups, and synthesize findings from each study. Furthermore, we conduct comprehensive experiments on two public scene classification datasets to evaluate and benchmark different self-supervised models.

## Introduction:

Modern supervised deep learning methods in computer vision heavily depend on large annotated datasets to learn pertinent image features. However, annotating such datasets is arduous and time-consuming. Notably, ImageNet stands as one of the largest annotated image recognition datasets, comprising over 14 million training images, which took considerable human effort to annotate. In many practical applications of vision-based fields, leveraging supervised models pre-trained on ImageNet has become customary to enhance the performance of deep neural networks through transfer learning or fine-tuning on smaller, domain-specific image data. Utilizing pre-trained ImageNet weights in transfer learning improves performance compared to initializing network weights randomly (i.e., training from scratch). These pre-trained weights offer superior representation capabilities, particularly in initial network layers. However, fine-tuning of deeper layers on domain-specific data is necessary for the network to extract task-relevant features effectively. In Earth observation through aerial and satellite remote sensing, vast amounts of data are generated daily, making meticulous annotation impractical. If annotated, this data could train supervised models for scene classification and serve as backbone models for other tasks by leveraging neural activations from coarse to deeper layers as image-level or patch-level representations. Self-supervised learning (SSL) offers a method to train generalized image representations without heavy reliance on annotated data. SSL learns deep feature representations invariant to sensible transformations, or augmentations, of input data. These models rely solely on unlabeled data to define their own training objective (i.e., pretext task), circumventing the need for time-consuming annotations. Features generated by SSL methods should possess discriminative properties for future downstream tasks while being generalized enough for application to new tasks without requiring retraining. Given recent SSL advancements in image representation, this paper explores how these developments can benefit remote sensing, specifically in scene classification tasks. Thus, the paper aims to review SSL methods developed for scene classification in recent years, providing guidance to researchers interested in this potential research area within remote sensing. This paper is structured as follows: we briefly outline remote sensing scene classification approaches, from classical feature engineering to modern deep learning, followed by an overview of significant self-supervised methods in computer vision inspiring the remote sensing community. Section 3 provides a detailed survey of SSL approaches developed for scene classification tasks. Section 4 presents our experimental study, benchmarking and comparing current state-of-the-art SSL frameworks on two public scene classification datasets. Section 5 discusses the role of image augmentation strategies and the transfer learning ability of SSL pre-trained models based on ablation analysis and additional experiments.

## Background

### Scene Classification

Scene classification in remote sensing involves predicting a label for an image from a dataset containing various semantic categories of land cover. Due to the visual similarities and shared objects across scenes, methods solely focusing on pixel or object-level modeling have been insufficient for accurate scene classification. Instead, a deeper understanding and characterization of the relationships among objects and regions within each scene type are required. For instance, both residential and industrial scenes may feature manmade structures, roads, and trees. Therefore, effective scene classification methods need to capture coarse-to-fine features from images while considering the spatial appearance and relationships among semantic elements. Typically, scene classification involves two steps: first, encoding the image into a feature representation, and then training a classifier on these representations to differentiate between semantic classes. Depending on the representations, classification can be achieved using simple linear classifiers or more complex ones like random forests or support vector machines (SVMs). Early methods relied on feature engineering to craft representations tailored to the task. Techniques such as Histogram of Oriented Gradients (HOG) and Scale-Invariant Feature Transform (SIFT) were commonly used to extract local features, which were then aggregated using methods like Bag-of-Visual-Words (BOVW) or Fisher Vector (FV) representation.

In recent years, deep learning models have demonstrated remarkable representation capabilities across various domains, including remote sensing, leading to state-of-the-art performance in scene classification. Convolutional Neural Networks (CNNs) have particularly dominated remote sensing scene classification, extracting features from local image neighborhoods using shared weights across convolutional kernels. CNNs' early layers capture low-level features, while deeper layers extract object-level features during image classification. Spatial features are aggregated and processed by fully connected layers to generate scores for each semantic class. Furthermore, the adoption of transformer models, treating images as sequences of visual tokens, has shown promising advancements in image classification.

A pragmatic approach in scene classification involves utilizing pre-trained weights from models trained on large datasets for initialization, rather than starting training from scratch. Research has demonstrated that pre-trained weights from networks trained on ImageNet already enhance classification performance, even when the target dataset differs visually from ImageNet. This approach, known as transfer learning, could yield even greater benefits if pre-training occurs on a large remote sensing dataset instead of ImageNet, as it can offer more relevant features. This notion is supported by studies indicating significant differences between ImageNet samples and remote sensing images. Objects in remote sensing datasets are often dispersed throughout the entire image, contrasting with ImageNet samples where objects are typically centered.

**Figure 1. Comparison between Object-Centric Natural Images (from the ImageNet [1] dataset) and Remote Sensing Scene Images (from the Resisc-45 [3] dataset).**

To assess and compare scene classification methods, the remote sensing community has compiled diverse datasets for benchmarking, ranging from simple three-channel RGB images to more complex multi-spectral, hyperspectral, or time series datasets. Here, we briefly introduce some commonly used datasets for benchmarking scene classification methods, focusing on optical images. One of the earliest and most renowned datasets is the UC-Merced dataset, comprising 21 classes, each with 100 images sized 256x256 pixels and a resolution of 0.3 meters. However, due to the demand for larger datasets with more classes, several bigger datasets like NWPU-RESISC45, the Aerial Image Dataset (AID), and the RSI-CB were created by collecting and extracting data from Google Earth. Among these, NWPU-RESISC45 (Resisc-45 hereafter) is widely utilized, containing over 31,500 high-resolution images covering 45 different scene categories, with resolutions ranging from 30 meters to 0.2 meters. For lower-resolution images sourced from open-access data, EuroSAT and BigEarthNet are prominent choices for benchmarking. EuroSAT consists of 27,000 small images sized 64x64 pixels, with spatial resolutions varying from 10 meters to 30 meters per pixel, covering 10 scene categories. BigEarthNet, on the other hand, is one of the largest remote sensing scene classification datasets, comprising more than 590,000 samples sized 120x120 pixels extracted from Sentinel-2 data, encompassing 44 classes. For further details on these datasets and others, readers are directed to in-depth descriptions and analyses of scene classification benchmarks in recent review papers.

While deep learning methods have become the predominant approach for solving remote sensing scene classification problems, the demand for domain-specific labeled data poses a scalability challenge for improving classification performance. Studies have shown that the performance of deep learning models can indeed improve with increased labeled data and network size. However, labeling data is costly, time-consuming, and can be biased depending on the annotator. Consequently, the computer vision community has been driven to develop unsupervised representation learning methods to address this issue. Recent methods, which formulate their own training objectives using data, are referred to as self-supervised learning (SSL) methods, a topic we will review in the following section.

## Self-Supervised Methods

Acquiring a large annotated dataset for a specific task can be labor-intensive, prompting the development of algorithms to learn effective image representations without supervision, known as unsupervised learning techniques. When the training objective is derived from the data itself, these methods are termed self-supervised learning methods. Essentially, a feature representation is encoded from an image using a deep neural network trained on a pretext task for which labels are automatically generated without human annotation. These learned representations, designed to solve pretext tasks, can later serve as a foundation for supervised downstream tasks. In this section, we provide a brief overview of the most significant state-of-the-art self-supervised methods, primarily proposed within the machine learning and computer vision communities. Without sacrificing generality, we categorize these methods into four groups: generative, predictive, contrastive, and non-contrastive SSL. It's worth noting that in the literature, contrastive and non-contrastive approaches can be amalgamated into a single joint-embedding approach. However, for clarity, we opt to differentiate between these two without any loss of generality. The aim is to trace their chronological evolution and offer sufficient background for our primary survey in Section 3. For more comprehensive surveys of self-supervised approaches, readers are encouraged to explore dedicated review papers.

**Generative**

A prevalent pretext task involves reconstructing the input image after compression using an autoencoder. By minimizing the reconstruction loss, the model learns to compress all relevant information from the image into a lower-dimensional latent space using the encoder component. Subsequently, the decoder component attempts to reconstruct the image from this latent space. Denoising autoencoders have also demonstrated efficacy in generating robust image representations by learning to eliminate artificial noise from images. Variational autoencoders (VAE) enhance the autoencoder framework by encoding the parameters of the latent space distribution. They are trained to minimize both the reconstruction error and an additional term, reducing the Kullback-Leibler divergence between a known latent distribution, often a unit-centered Gaussian distribution, and the one produced by the encoder. This regularization over the latent space facilitates sampling from the generated distribution. More recently, the advent of vision transformers has facilitated the development of large masked autoencoders operating at a patch level instead of pixel-wise, reconstructing entire patches with only a subset of visible patches. This reconstruction task yields robust image representations by appending a class token to the sequence of patches or employing global average pooling on all the patch tokens.

**Predictive**

The second category of SSL methods encompasses models trained to predict the outcome of an artificial transformation applied to the input image. This approach is motivated by the idea that predicting the transformation necessitates learning relevant characteristics of semantic objects and regions within the image. For instance, by pre-training a model to predict the relative position of two image patches, reference [31] achieved performance enhancements compared to random initialization, approaching the performance level of ImageNet pre-trained weights in well-established computer vision datasets. Various other predictive pretext tasks have been proposed to learn representations. One such task is image colorization, introduced in reference [32]. In this method, the input image is initially converted to grayscale, and then an autoencoder is trained to recolorize the grayscale version back to the original color image by minimizing the mean squared error between the reconstruction and the original. The feature representations extracted by the encoder are then utilized for downstream tasks. Another notable predictive SSL method is RotNet [33], which trains a model to predict the randomly applied rotation to the input image. This rotation prediction task compels the model to extract meaningful features that aid in comprehending the semantic content of the image. Similarly, another SSL model tackles a jigsaw puzzle task [34], predicting the relative positions of image partitions that were previously shuffled. Furthermore, the Exemplar CNN [35] is trained to predict the augmentations applied to images by considering various types of augmentations, including cropping, rotation, color jittering, and contrast modification.

**Contrastive**

Another approach to obtaining effective image representations is by encouraging the features of multiple views of an image to be similar. This ensures that the final representations are invariant to the augmentations used to create the different image views. However, if not properly managed, the network may converge to a constant representation that is independent of the input image, satisfying the invariance constraint (known as the collapsing problem).

To address this challenge and promote diverse representations while preventing the collapsing issue, contrastive loss is commonly employed. This loss function aims to compel the model to differentiate between representations of views from the same image (i.e., positives) and those from different images (i.e., negatives). Essentially, it strives to generate similar feature representations for positive pairs while pushing apart representations for negative pairs. Among methods in this category, the simplest objective is the triplet loss. In triplet loss, a model is trained to minimize the distance between representations of an anchor and its positive instance more than the distance between that anchor and a randomly selected negative instance. This concept is illustrated in Figure 4. The triplet loss function can be expressed as follows:

3. Self-Supervised Remote Sensing Scene Classification

Remote sensing scene classification data possess distinct characteristics compared to natural images in computer vision. Remote sensing images typically exhibit heterogeneous backgrounds with abundant textures and structural information. Unlike vision images, where primary objects are typically the focal point, images captured by aerial and satellite platforms may contain various object classes with different sizes, shapes, and orientations, influenced by the sensor's spectral and spatial resolution. Consequently, despite the significant advancements in deep learning models from the machine learning and computer vision communities, most learning methods have been adapted for scene classification to generate more relevant feature representations suitable for downstream remote sensing tasks. In this section, we delve into existing self-supervised remote sensing scene classification methods and their specifics compared to general methods developed in computer vision. We categorize these methods based on the aforementioned approaches and discuss their application in the remote sensing domain.

**Generative**

One of the pioneering generative SSL methods applied to scene classification is MARTA GANs (Multiple-layer Feature-matching Generative Adversarial Networks), as proposed in reference [52]. Similar to the concept of GAN-based generative SSL, MARTA GANs involve training a GAN to generate artificial scene images as a pretext task to create image representations. The core concept of MARTA GANs involves extracting multi-level features from different network layers and aggregating them through concatenation. Additionally, the generator is trained to maximize the similarity of activations between fake and real images at every layer of the discriminator, defining the multilevel feature matching loss. MARTA GANs demonstrate promising results on the UC-Merced and Brazilian coffee scene datasets, providing high-quality fake samples while delivering competitive classification performance.

Another early utilization of generative models in SSL for remote sensing is presented in reference [54], where a split-brain autoencoder is evaluated for self-supervised image representation. Split-brain autoencoders address the challenge of learning relevant information from data distribution by splitting the input data into two different nonoverlapping subsets of data channels (or spectral bands in the remote sensing context) and learning to reconstruct one subset from the other. The overall training loss is defined, and the final image representation is obtained by concatenating the output of both encoders into a single discriminative embedding vector. Experimental results on the Resisc-45 and AID datasets demonstrate competitive performance, particularly with few unlabeled training images.

**Predictive**

In [56], a comparative study of different SSL methods applied to remote sensing scene classification is conducted, including image inpainting, relative position prediction, and instance discrimination. By employing linear classification as a downstream task, the instance discrimination (IDSSL) model outperforms predictive approaches while being less sensitive to the amount of labeled data. Furthermore, utilizing IDSSL pre-trained weights significantly boosts classification performance, particularly with limited labeled samples.

Another approach proposed in [57] combines self-supervised and supervised training strategies using multitask learning with a mixup loss function. By jointly training a model with self-supervised loss for image rotation prediction and supervised cross-entropy loss for label prediction, the model learns features dependent on both classification and rotation, leading to competitive classification results.

An alternative predictive method is presented in [59], where parts of a sample are masked, and a single encoder-decoder is trained to reconstruct the entire original sample. This method, called SITS-BERT, is based on the masking technique from the BERT model adapted for satellite image time series. SITS-BERT learns spectral-temporal representations related to land cover contents from satellite image time series data.

These predictive methods showcase diverse approaches to self-supervised learning in remote sensing, each offering unique advantages and demonstrating competitive performance in scene classification tasks.

**Contrastive Learning Approaches in Remote Sensing Scene Classification**

In recent years, the remote sensing community has embraced various contrastive joint-embedding methods, tailoring them to develop novel algorithms for scene classification. Among these methods, Tile2Vec stands out as one of the pioneers. Tile2Vec utilizes the triplet loss function to glean compressed representations from unlabeled remote sensing data. It leverages geographical proximity to establish positive and negative sample sets for each image patch, ensuring similar representations for close tiles and distinct ones for distant ones. Although Tile2Vec's reliance on geographic data may limit its applicability in some datasets, its efficacy is evident in specific contexts where geo-information is available. Evaluations on datasets like the National Agriculture Imagery Program (NAIP) and the Cropland Data Layer (CDL) demonstrate Tile2Vec's superiority over other unsupervised feature extraction methods.

Another noteworthy approach modifies the triplet loss to suit remote sensing images. This adaptation reformulates the loss as a binary classification problem, employing fully-connected layers with sigmoid activation to output scores. By optimizing these scores, the model learns to distinguish positive and negative pairs effectively. Moreover, the introduction of randomized predictor network weights enhances representation quality compared to traditional methods. Evaluation on datasets like NAIP and CDL showcases the method's potential, outperforming Tile2Vec in certain scenarios.

Contrastive methods often necessitate the creation of multiple views for each instance to build discriminative representations. Leveraging the specifics of remote sensing data, contrastive multiview coding exploits multi-spectral images to produce consistent representations across different spectral bands. This approach splits an original image into two views based on its channels and employs a contrastive loss to encourage proximity between positive pairs and separation between negative pairs. Extensive experiments confirm the superiority of SSL pre-training on remote sensing images over natural images. Furthermore, proper pre-training on multi-spectral data is essential for downstream tasks reliant on such imagery.

Geography-aware SSL methods capitalize on temporal and spatial metadata available in certain datasets. By using this information, models can learn consistent representations across different timestamps and locations, improving performance in downstream tasks. The momentum contrast (MoCo) method, alongside geolocation classification, enhances feature representations compared to direct MoCo application on remote sensing datasets.

Overall, contrastive learning approaches show immense promise in remote sensing scene classification, offering robust representations and outperforming traditional methods in various contexts. With ongoing research focusing on multimodal fusion, temporal invariance, and novel pre-training strategies, the field continues to evolve, paving the way for more accurate and efficient remote sensing applications.

References List:

[1]. Shivakumar, S. K., & Sethii, S. (2019). *Building Digital Experience Platforms: A Guide to Developing Next-Generation Enterprise Applications*. Apress.

[2]. Sethi, S., & Panda, S. (2024). Transforming Digital Experiences: The Evolution of Digital Experience Platforms (DXPs) from Monoliths to Microservices: A Practical Guide. *Journal of Computer and Communications*, *12*(2), 142-155. https://doi.org/10.4236/jcc.2024.122009

[3]. Sethi, P. Karmuru, & Tayal.(2023). Analyzing and Designing a Full-Text Enterprise Search Engine for Data-Intensive Applications. *International Journal of Science, Engineering and Technology*, *11*. https://www.ijset.in/wp-content/uploads/IJSET_V11_issue6_628.pdf

[4]. Sethi, S., Panda, S., & Kamuru, R. (2023). Comparative study of middle tier caching solution. *International Journal of Development Research*, *13*(11), 64225-64229.

[5]. Gitte, M., Bawaskar, H., Sethi, S., & Shinde, A. (2014). Content based video retrieval system. *International Journal of Research in Engineering and Technology*, *3*(06), 123-129.

[6]. Gitte, M., Bawaskar, H., Sethi, S., & Shinde, A. (2014). Content based video retrieval system. *International Journal of Research in Engineering and Technology*, *3*(06), 123-129.

[7]. Sethi, S., & Shivakumar, S. K. (2023). DXPs Digital Experience Platforms Transforming Fintech Applications: Revolutionizing Customer Engagement and Financial Services. *International Journal of Advance Research, Ideas and Innovations in Technology*, *9*, 419-423.

[8]. Sethi, S., Panda, S., & Hooda, S. (2024). Design and Optimization of a Zookeeper and Kafka-Based Messaging Broker for Asynchronous Processing in High Availability and Low Latency Applications. *J Curr Trends Comp Sci Res*, *3*(2), 01-07.

[9]. Li, Z., Zhu, H., Liu, H., Song, J., & Cheng, Q. (2024). Comprehensive evaluation of Mal-API-2019 dataset by machine learning in malware detection. *arXiv preprint arXiv:2403.02232*.

https://doi.org/10.48550/arXiv.2403.02232

[10]. Jhurani, J. REVOLUTIONIZING ENTERPRISE RESOURCE PLANNING: THE IMPACT OF ARTIFICIAL INTELLIGENCE ON EFFICIENCY AND DECISION-MAKING FOR CORPORATE STRATEGIES.

[11]. Jhurani, J. Enhancing Customer Relationship Management in ERP Systems through AI: Personalized Interactions, Predictive Modeling, and Service Automation.

[12]. Jhurani, J. DRIVING ECONOMIC EFFICIENCY AND INNOVATION: THE IMPACT OF WORKDAY FINANCIALS IN CLOUD-BASED ERP ADOPTION.

[13]. Smith, J. D. (2024). The Impact of Technology on Sales Performance in B2B Companies. *Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023*, *3*(1), 86-102.

https://doi.org/10.60087/jaigs.vol03.issue01.p102

[14]. Deep Smith, J. (2024). 2019 to 2024, The State of Sales Management Research in the Latin America. *International Journal of Engineering, Management and Humanities (IJEMH)*, *5*(1), 223-226.

[15]. Smith, J. D. Influence of Self-Efficacy, Stress, and Culture on the Productivity of Industrial Sales Executives in Latin American Sales Networks.

[16]. Miah, S., Rahaman, M. H., Saha, S., Khan, M. A. T., Islam, M. A., Islam, M. N., ... & Ahsan, M. H. (2013). Study of the internal structure of electronic components RAM DDR-2 and motherboard of nokia-3120 by using neutron radiography technique. *International Journal of Modern Engineering Research (IJMER)*, *3*(60), 3429-3432

[17]. Rahaman, M. H., Faruque, S. B., Khan, M. A. T., Miah, S., & Islam, M. A. (2013). Comparison of General Relativity and Brans-Dicke Theory using Gravitomagnetic clock effect. *International Journal of Modern Engineering Research*, *3*, 3517-3520.

[18]. Miah, M. H., & Miah, S. (2015). The Investigation of the Effects of Blackberry Dye as a Sensitizer in TiO2 Nano Particle Based Dye Sensitized Solar Cell. *Asian Journal of Applied Sciences*, *3*(4).

[19]. Miah, S., Miah, M. H., Hossain, M. S., & Ahsan, M. H. (2018). Study of the Homogeneity of Glass Fiber Reinforced Polymer Composite by using Neutron Radiography. *Am. J. Constr. Build. Mater*, *2*, 22-28.

[20]. Miah, S., Islam, G. J., Das, S. K., Islam, S., Islam, M., & Islam, K. K. (2019). Internet of Things (IoT) based automatic electrical energy meter billing system. *IOSR Journal of Electronics and Communication Engineering*, *14*(4 (I)), 39-50.

[21]. Nadia, A., Hossain, M. S., Hasan, M. M., Islam, K. Z., & Miah, S. (2021). Quantifying TRM by modified DCQ load flow method. *European Journal of Electrical Engineering*, *23*(2), 157-163.

[22]. Miah, S., Raihan, S. R., Sagor, M. M. H., Hasan, M. M., Talukdar, D., Sajib, S., ... & Suaiba, U. (2022). Rooftop Garden and Lighting Automation by the Internet of Things (IoT). *European Journal of Engineering and Technology Research*, *7*(1), 37-43.

DOI: https://doi.org/10.24018/ejeng.2022.7.1.2700

[23]. Prasad, A. B., Singh, S., Miah, S., Singh, A., & Gonzales-Yanac, T. A Comparative Study on Effects of Work Culture on employee satisfaction in Public & Private Sector Bank with special reference to SBI and ICICI Bank.

[24]. Ravichandra, T. (2022). A Study On Women Empowerment Of Self-Help Group With Reference To Indian Context.

https://www.webology.org/data-cms/articles/20220203075142pmwebology%2019%20(1)%20-%2053.pdf

[25]. Kumar, H., Aoudni, Y., Ortiz, G. G. R., Jindal, L., Miah, S., & Tripathi, R. (2022). Light weighted CNN model to detect DDoS attack over distributed scenario. *Security and Communication Networks*, *2022*.

https://doi.org/10.1155/2022/7585457

[26]. Ma, R., Kareem, S. W., Kalra, A., Doewes, R. I., Kumar, P., & Miah, S. (2022). Optimization of electric automation control model based on artificial intelligence algorithm. *Wireless Communications and Mobile Computing*, *2022*.

https://doi.org/10.1155/2022/7762493

[27]. Devi, O. R., Webber, J., Mehbodniya, A., Chaitanya, M., Jawarkar, P. S., Soni, M., & Miah, S. (2022). The Future Development Direction of Cloud-Associated Edge-Computing Security in the Era of 5G as Edge Intelligence. *Scientific Programming*, *2022*.

https://doi.org/10.1155/2022/1473901

[28]. Al Noman, M. A., Zhai, L., Almukhtar, F. H., Rahaman, M. F., Omarov, B., Ray, S., ... & Wang, C. (2023). A computer vision-based lane detection technique using gradient threshold and hue-lightness-saturation value for an autonomous vehicle. *International Journal of Electrical and Computer Engineering*, *13*(1), 347.

[29]. Patidar, M., Shrivastava, A., Miah, S., Kumar, Y., & Sivaraman, A. K. (2022). An energy efficient high-speed quantum-dot based full adder design and parity gate for nano application. *Materials Today: Proceedings*, *62*, 4880-4890.https://doi.org/10.1016/j.matpr.2022.03.532

[30]. Pillai, A. S. (2023). Advancements in Natural Language Processing for Automotive Virtual Assistants Enhancing User Experience and Safety. *Journal of Computational Intelligence and Robotics*, *3*(1), 27-36.

[31]. Rehan, H. (2024). Revolutionizing America's Cloud Computing the Pivotal Role of AI in Driving Innovation and Security. *Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023*, *2*(1), 189-208. DOI: https://doi.org/10.60087/jaigs.v2i1.p208

[32]. Rehan, H. (2024). AI-Driven Cloud Security: The Future of Safeguarding Sensitive Data in the Digital Age. *Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023*, *1*(1), 47-66.
DOI: https://doi.org/10.60087/jaigs.v1i1.p66

[33]. Rahman, S., Mursal, S. N. F., Latif, M. A., Mushtaq, Z., Irfan, M., & Waqar, A. (2023, November). Enhancing Network Intrusion Detection Using Effective Stacking of Ensemble Classifiers With Multi-Pronged Feature Selection Technique. In *2023 2nd International Conference on Emerging Trends in Electrical, Control, and Telecommunication Engineering (ETECTE)* (pp. 1-6). IEEE. https://doi.org/10.1109/ETECTE59617.2023.10396717

[34]. Latif, M. A., Afshan, N., Mushtaq, Z., Khan, N. A., Irfan, M., Nowakowski, G., ... & Telenyk, S. (2023). Enhanced classification of coffee leaf biotic stress by synergizing feature concatenation and dimensionality reduction. *IEEE Access*.
https://doi.org/10.1109/ACCESS.2023.3314590

[35]. Irfan, M., Mushtaq, Z., Khan, N. A., Mursal, S. N. F., Rahman, S., Magzoub, M. A., ... & Abbas, G. (2023). A Scalo gram-based CNN ensemble method with density-aware smote oversampling for improving bearing fault diagnosis. *IEEE Access*, *11*, 127783-127799.

https://doi.org/10.1109/ACCESS.2023.3332243

[36]. Irfan, M., Mushtaq, Z., Khan, N. A., Althobiani, F., Mursal, S. N. F., Rahman, S., ... & Khan, I. (2023). Improving Bearing Fault Identification by Using Novel Hybrid Involution-Convolution Feature Extraction with Adversarial Noise Injection in Conditional GANs. *IEEE Access*. https://doi.org/10.1109/ACCESS.2023.3326367

[37]. Latif, M. A., Mushtaq, Z., Arif, S., Rehman, S., Qureshi, M. F., Samee, N. A., ... & Al-masni, M. A. Improving Thyroid Disorder Diagnosis via Ensemble Stacking and Bidirectional Feature Selection. https://www.techscience.com/cmc/v78n3/55928/html

[38]. Jimmy, F. N. U. (2024). Cyber security Vulnerabilities and Remediation Through Cloud Security Tools. *Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023*, *3*(1), 196-233. https://doi.org/10.60087/jaigs.vol03.issue01.p233

[39]. Gunasekaran, K. P., Babrich, B. C., Shirodkar, S., & Hwang, H. (2023, August). Text2Time: Transformer-based Article Time Period Prediction. In *2023 IEEE 6th International Conference on Pattern Recognition and Artificial Intelligence (PRAI)* (pp. 449-455). IEEE. https://doi.org/10.1109/PRAI59366.2023.10331985

[40]. Gunasekaran, K., & Jaiman, N. (2023, August). Now you see me: Robust approach to partial occlusions. In *2023 IEEE 4th International Conference on Pattern Recognition and Machine Learning (PRML)* (pp. 168-175). IEEE. https://doi.org/10.1109/PRML59573.2023.10348337

[40]. Kommaraju, V., Gunasekaran, K., Li, K., Bansal, T., McCallum, A., Williams, I., & Istrate, A. M. (2020). Unsupervised pre-training for biomedical question answering. *arXiv preprint arXiv:2009.12952*. https://doi.org/10.48550/arXiv.2009.12952

[41]. Bansal, T., Gunasekaran, K., Wang, T., Munkhdalai, T., & McCallum, A. (2021). Diverse distributions of self-supervised tasks for meta-learning in NLP. *arXiv preprint arXiv:2111.01322*. https://doi.org/10.48550/arXiv.2111.01322

[42]. Padmapriya, V. M., Thenmozhi, K., Hemalatha, M., Thanikaiselvan, V., Lakshmi, C., Chidambaram, N., & Rengarajan, A. (2024). Secured IIoT against trust deficit-A flexi cryptic approach. *Multimedia Tools and Applications*, 1-28. https://doi.org/10.1007/s11042-024-18962-x

[43]. Bansal, T., Gunasekaran, K., Wang, T., Munkhdalai, T., & McCallum, A. (2021). Diverse distributions of self-supervised tasks for meta-learning in NLP. *arXiv preprint arXiv:2111.01322*.

https://doi.org/10.48550/arXiv.2111.01322

[44]. Gunasekaran, K., Tiwari, K., & Acharya, R. (2023, June). Utilizing deep learning for automated tuning of database management systems. In *2023 International Conference on Communications, Computing and Artificial Intelligence (CCCAI)* (pp. 75-81). IEEE.

https://doi.org/10.1109/CCCAI59026.2023.00022

[45]. Gunasekaran, K. P. (2023, May). Ultra sharp: Study of single image super resolution using residual dense network. In *2023 IEEE 3rd International Conference on Computer Communication and Artificial Intelligence (CCAI)* (pp. 261-266). IEEE.

https://doi.org/10.1109/CCAI57533.2023.10201303