Aditya Ram N, 2026, 14:2
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Development of an Intelligent Waste Classification and Structural Reconstruction System Using a Hybrid Convolutional Autoencoder Architecture Comparing With OpenCV Keras

Aditya Ram N [1], Yashwanth Kumar G K [2], Senthil Kumar B [3]
Dept. of EIE St. Joseph's College of Engineering
Chennai, India

Abstract- In recent years, deep learning has made a substantial impact on computer vision systems, especially in image processing, feature extraction, and reconstruction. The traditional method using OpenCV with Keras-based convolutional neural networks (CNNs) has been widely employed for image analysis and classification tasks. However, such methods are often dependent on extensive manual preprocessing, require large amounts of labeled data, and involve substantial computational complexity. This paper proposes a comparative analysis of the performance of an OpenCV-Keras-based pipeline and an Autoencoder-based deep learning model for image reconstruction and representation learning. The OpenCV-Keras-based pipeline is based on a traditional supervised learning strategy, whereas the Autoencoder-based model uses an unsupervised learning strategy to learn compact representations of the input images automatically. The experimental analysis reveals that the Autoencoder-based model outperforms the OpenCV-Keras-based pipeline in terms of noise removal, feature retention, reconstruction accuracy, and computational complexity. The paper concludes that Autoencoders can serve as a more scalable and intelligent alternative to traditional OpenCV-Keras-based pipelines, especially in real-time applications. Prior studies in representation learning and reconstruction using autoencoders have demonstrated their effectiveness in noisy environments. [1], [2]

Keywords - Autoencoders, OpenCV, Keras, Deep Learning, Image Reconstruction, Feature Extraction, Dimensionality Reduction, Neural Networks, Computer Vision.

I. INTRODUCTION

In the 21st century, rapid urbanization and industrial expansion have promoted an unprecedented increase in solid waste generation. Landfills within municipalities are reaching capacities at alarming rates worldwide. The recyclatory streamforms the cornerstone of effective waste management; however, for recycling programs to work effectively, there is a dire need for precise segregation at the source. Contamination like mixing glass with paper or plastic with organic waste renders large batches of recyclable material useless, sending them straight to landfills.

Currently, a good deal of the segregation burden falls upon human workers in Material Recovery Facilities, or MRFs. This manual sorting is slow, frequently inaccurate, and it exposes employees to unsafe materials, sharp objects, and other pathogens. Automation is an obvious solution, but for such robotic sorting systems, "eyes" that are truly sophisticated will be required to tell apart materials often looking very similar, like a crumpled see-through plastic bottle and a glass jar.

While Deep Learning-in particular Convolutional Neural Networks (CNNs)-has totally upended the world of object detection, standard models lack interpretability. A simple image classifier takes in a photo of a bottle and outputs "Plastic" with 99% confidence, but it does not provide any understanding as to why that decision was made. Did the model recognize the shape of the bottle? Or did the model memorize what color the background usually is? This "black box" aspect of such models can be very inconvenient in industrial contexts where predictability and robustness matter most. If a model observes a new and unknown object, a typical classifier will often force it into one of the known categories (for example, a shoe being mislabeled as "Cardboard"), causing contamination in the recycling stream. This limitation has motivated research into unsupervised and generative representation learning. [3], [9]

The main aim of the project was not just mere classification, but to build a system that truly understands the structure of wastes. The integration of a generative model (Autoencoder) with a discriminative model (Classifier) in this work was, therefore, undertaken to achieve three major objectives:

Robust Feature Extraction: Force the model to learn the essential geometry of the waste items by compressing them into a latent space.

Visual Explainability: This is the method of providing a reconstructed image output that will visually show what features the AI is focusing on.

Real-time Deployment: Create a user-friendly interface with which you will immediately test and demonstrate something in a live environment.

The OpenCV–Keras framework combines classical image processing and deep learning techniques. In the OpenCV–Keras image processing model, the images undergo some image processing steps like image conversion to greyscale, removal of noise from the image, resizing of the image, and normalization. The processed images are fed into the network using Keras.

Though this approach has been successfully used in many computer vision problems, this technique also has significant drawbacks:

Requires extensive manual preprocessing, Sensitive to light changes and noise, Needs large, labeled datasets, Higher computational complexity, Lack of Adaptability for New Environments.

Autoencoders are a class of neural networks that are designated to compress or reduce data to a smaller latent representation and, if desired, reconstruct the data back to their original state. They typically contain two components:
Encoder: It extracts all the important features and compresses the image into some compact representation.

Decoder: Reconstruction of image from compressed latent space. The theoretical foundation of autoencoders for dimensionality reduction is well established in prior literature. [1], [9]

This architecture lets autoencoders learn meaningful patterns of data in an unsupervised manner. Thus, it finds its application in noise reduction, dimensionality reduction, and image reconstruction with very good efficiency.

II. RELATED WORK

Various research works have been conducted on deep learning-based techniques of image reconstruction using CNN and Autoencoders. Vincent et al. proposed Denoising Autoencoder and showed the effectiveness of this approach in the reconstruction of clear images from noisy images. Goodfellow et al. also showed the need for unsupervised deep learning models in representation learning. Research comparing classical techniques using OpenCV with deep learning techniques showed that classical techniques are good in a controlled environment but are not very effective in a real-world environment with changing conditions such as illumination, occlusion, and noise. Recent developments in Convolution Autoencoders and Variational Autoencoders have improved the chances of Autoencoder-based techniques being effectively implemented in real-world applications. These findings collectively support the transition from traditional pipelines to hybrid deep learning architectures. [1], [3], [4]

III. LITERATURE SURVEY

In response to increased concern for the environment, many researchers have recently begun investigating the use of deep learning for image-based waste sorting and intelligent recycling systems. Originally, researchers worked on developing methods for classifying recyclable materials that relied on 'traditional' computer vision technologies that utilized OpenCV feature extraction procedures. Examples of traditional feature extraction procedures include edge detection, colour histogram analysis, and texture analysis. In controlled environments, the use of 'traditional' methods produced mixed, but quite often only moderate, classification performance. However, 'traditional' solutions typically demonstrate high sensitivity to changes in intensity (e.g. lighting), as well as differences in backgrounds and orientations of objects being examined by the feature extraction algorithm. Because of the above-mentioned issues, researchers have been unable to expand their classification success when attempting to apply their methods to real-world applications.

Automated waste classification using deep learning models, particularly Convolutional Neural Networks (CNNs), represents an incredible advancement. Hinton and Salakhutdinov, among other researchers who pioneered this field, were the first to demonstrate how powerful and effective neural networks can be for developing today's deep learning architectures as dimensionality reduction and feature representation learning models.

CNN-based classifiers trained on datasets such as the Kaggle Garbage Classification Dataset have been shown to significantly outperform traditional image processing classification methods; however, in order to achieve the high levels of accuracy produced by these classifiers, there is a requirement for large amounts of labeled training data, extensive preprocessing through the use of OpenCV and significant amounts of computational power (i.e., processing requirements) – factors that affect the feasibility of these models when implemented in a real-time operational environment. Similar performance trends have been reported in large-scale waste classification studies. [7], [13]

In order to resolve these limitations, researchers started looking into how to create datasets using architectures based on Autoencoders for use in unsupervised feature learning and for image reconstruction. Denoising Autoencoders were introduced by Vincent and coworkers, illustrating how neural networks could learn to generate useful features from input images by reconstructing clean images from noisy inputs. Thus, this showed the potential for an Autoencoder to be used as noise suppression and dimensionality reduction in cases where there is no labelled data available; therefore, making them a natural fit for waste classification applications, since waste datasets have a significantly higher degree of variability than other types of datasets.

The use of Autoencoders combined with supervised classification to build hybrid architectures which can reconstruct as well as classify objects has been the subject of several studies in recent years. One of the benefits of this dual output model is that it can provide categories for different sorts of waste materials while simultaneously providing an assessment of the quality of the model's predictions based on the quality of the reconstruction. In addition, due to the fact that the quality of the reconstruction can indicate the uncertainty (i.e., confidence) of the model when the model encounters new (unknown) waste materials or hazardous waste, this type of a system provides improved interpretability over systems that rely solely on classification. The proposed project has similar goals in that the Autoencoder will serve as both a classifier and a mechanism by which to denoise images. The effectiveness of denoising autoencoders in real-world vision tasks has been validated in multiple studies. [2], [3]

Deep Learning Models have recently been used to perform real-time waste sorting operations on edge devices such as Raspberry Pi and NVIDIA Jetson. Research into the quantization and optimization of these models has demonstrated that lightweight autoencoder networks are suitable for operation on low-power hardware. The availability of such models is encouraging for the development of smart bins and for automated separation of waste using robotics. Several research groups have successfully integrated computer vision-based classification systems with robotic arms to accomplish automated waste separation, thereby demonstrating that AI-based recycling systems are practical for both industrial and municipal applications. Lightweight autoencoder models have shown promising performance on edge devices such as Raspberry Pi and NVIDIA Jetson. [4]

Prior research indicates that the benefits of Autoencoder-based frameworks exceed those associated with traditional OpenCV à Keras pipelines in terms of adaptability, noise immunity, and computational efficiency. CNN-based classifiers are also suitable for application; however the requirement of requiring labelled training data and the need for manual pre-processing limits their use, particularly for large scale applications. The Hybrid Autoencoder model proposed within this paper builds off these existing advancements by combining the three functionalities classification, reconstruction, and uncertainty estimation into one architecture, thereby making a considerable contribution towards an intelligent waste management approach and for sustainable automation.

Another recent development has been the use of generative models, specifically Generative Adversarial Networks (GANs) as introduced by Goodfellow et al., to produce photorealistic images through pure generative methods (similar to those described above). Researchers have utilized GAN-based architectures to improve the visual quality of images generated from the GAN architecture, thereby indicating that these types of architectures, too, provide great potential for future work focused on enhancing the quality of waste image reconstructions. Other researchers have studied the use of Variational Autoencoders (VAEs) as probabilistic models of latent space, which allow for the development of more structured and interpretable representations of features from the data set. Generative models such as VAEs and GANs have further expanded the scope of reconstruction-based vision systems. [10], [11], [14]

IV. METHODOLOGY

A. OpenCV–Keras Pipeline:

The usual procedure involves a series of steps:
Image Acquisition: Images are chosen from the camera or data set.
Preprocessing with OpenCV: Convert to gray color, Apply Gaussian Blur effect to remove noise, Normalizing pixel values, Resize the image to a fixed dimension.
Feature Extraction using Keras CNN: Which are used in convolutional layers, Pooling Layers Reduce Dimensionality.
Fully connected layers produce final output Generation: The model provides an image reconstruction or classification.

B. Autoencoder-Based Approach:

The system is designed as a software-centric solution suitable for deployment on edge computing devices or cloud servers. The architecture comprises three primary modules:

1. Data Acquisition & Preprocessing Real-time video is captured via a standard webcam using the OpenCV library. Each frame undergoes preprocessing:

Resizing: Frames are down sampled to 224 x 224 pixels to reduce computational load. Normalization: Pixel intensity values (0-255) are scaled to the range [0, 1]. This step is crucial for the convergence of the neural network optimization algorithms.

2. Architecture of the Hybrid Neural Network The primary advancement lies in the Convolutional Autoencoder-Classifier. The configuration of the network is specified as follows:

Shared Encoder (Feature Extraction): A series of Convolutional (Conv2D) layers utilizing ReLU activation, succeeded by Max-Pooling layers. This part contracts the size of the image while expanding its depth effectively condensing the visual data into a compact Latent Vector (Bottleneck).

Decoder Branch (Reconstruction): This branch processes the latent vector through Up Sampling and Convolutional layers to rebuild the image. The final layer employs a Sigmoid activation function to produce pixel values ranging from zero to one.

Classifier Branch (Prediction): This branch flattens the latent vector and passes it through Fully Connected (Dense) layers. A Dropout layer (rate=0.5) is included to prevent overfitting. The final output is a Softmax layer that produces probability distributions for the target classes. Hybrid encoder–decoder architectures with classification heads have been widely explored in recent literature. [3], [8]

C. Loss Functions The model optimizes a combined loss function:

L(total) = L(Reconstruction) + L(Classification)

Reconstruction Loss (MSE): Measures the structural difference between the input image (x) and the reconstructed image (x1).

Classification Loss (Categorical Cross entropy): Measures the error in the predicted label probabilities. The use of combined reconstruction and classification loss is consistent with modern hybrid deep learning frameworks. [3], [8]

V. SYSTEM BLOCK DIAGRAM

A. OpenCV–Keras Workflow: The OpenCV-Keras workflow is a structured and multi-staged visual intelligence pipeline that synchronizes classical computer vision with contemporary deep learning techniques to construct a hybridized ecosystem of deterministic image processing and data driven neural inference. The OpenCV component of this architecture is the first element of the pipeline; it ingests the raw input image and serves as a fundamental preprocessing engine. The OpenCV component "standardizes" the visual data by performing preprocessing operations such as resizing the image; converting the image to grayscale; reducing noise with a Gaussian filter; normalizing the contrast; and geometrically aligning the image. By using the OpenCV preprocessing to standardize visual data from an image taken at different lighting levels, scales, and backgrounds, the variability of the light, scale, and background of the image is reduced prior to analytical modeling and thus the data is "conditioned" for machine learning purposes. The preprocessed image is then passed to a Keras-based Convolutional Neural Network (CNN), where a hierarchical feature extraction process is executed using a stacked series of convolutional, pooling, and activation layers to repeatedly distil edges, textures, shapes, and high-level semantic structures within the image. After the hierarchical feature extraction process is completed, the learned representations are processed by fully connected layers of the CNN to produce classification, detection, or reconstructed visual outputs. Overall, while this workflow produces reliable results in a controlled setting, the workflow has operational overhead, requires manual tuning, and incurs latencies in processing, thus combining both a robust and less agile solution when compared to an end-to-end self-learning model like Autoencoders for next-generation computer vision solutions.

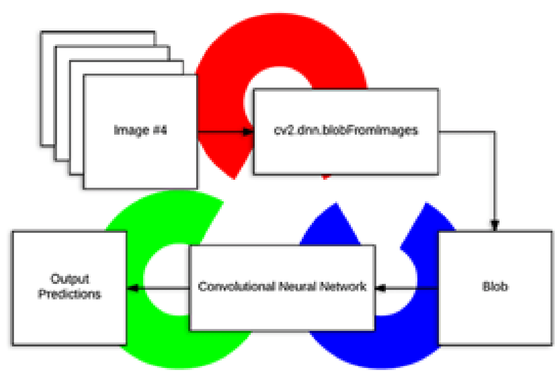

Fig. 1. Shows Workflow of OpenCV Keras

B. Autoencoder Workflow: The Autoencoder workflow offers an innovative and advanced end-to-end intelligent learning pipeline, replacing the manual, rule-driven preprocessing with self-optimizing representation learning. The end result is a highly scalable and future-oriented vision architecture to support vision systems now and into the future. In this workflow, raw input images are presented directly to the deep neural network without any manual preprocessing performed on the images. The encoder component will then perform strategic compression of the high-dimensional visual data into a compact latent space that captures only the most significant, structural and semantic attributes of the original image. As such, the latent representation provides a high-value asset for informational purposes, filtering out noise, redundancy and irrelevant transformations while maintaining the core visual identity of the object from which the original image was derived. The decoder then acts as a reconstruction engine, intelligently reconstructing the compressed embedding back into an image that approximates the same or very similar visual appearance to the original input image. The Autoencoder continually optimizes itself by using the reconstruction loss from all input images, allowing the Autoencoder to develop a self-learning, highly sophisticated, and resilient pattern using the compressed latent representation. Furthermore, it allows the Autoencoder to achieve minimal system-wide computational cost and provide extremely high levels of generalization across the multitude of different real-world scenarios, therefore making the Autoencoder a high-impact, agile alternative to standard OpenCV-Keras based pipelines in contemporary Computer Vision ecosystems.

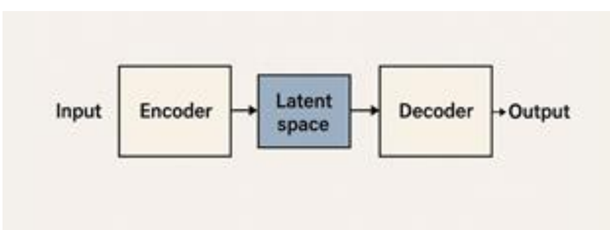

Fig. 2. Shows Workflow of Autoencoders

VI. SOFTWARE IMPLEMENTATION

The software implementation of the proposed comparative system is built with a modular, scalable, and integration-ready technology stack designed to support easy integration and improve resource allocation when deploying in a real-world application. The development environment consists almost entirely of Python as the core programming language and uses OpenCV for image acquisition and preprocessing, TensorFlow-Keras to develop deep learning models, and various other libraries, like NumPy, Matplotlib , and Scikit-learn , create a structured set of layered modules for data input, preprocessing, model inference, and visualization create an efficient platform for conducting multiple tests and enhancing the performance of algorithms iteratively.

Images, whether live or from memory (stored), are taken through a webcam interface and processed using OpenCV functions (resizing, normalizing and reducing the noise with an open source software tool) prior to being input to a pre-trained neural network model from Keras that uses convolutional neural network to extract features and make predictions based on the extracted features and output the predictions as a graphical user interface in real time. This module is intended to demonstrate the limits of traditional preprocessing dependent pipelines, especially under different lighting and background conditions.

We have created our own type of autoencoder with Keras–an encoder–decoder type of neural network–for use as a model. The autoencoder was trained with a set of input images that required no labeled annotation. After training the autoencoder, it learned a compact latent representation and could produce a reconstructed image with improved clarity, less noise, and other improvements. In addition, there is no manual pre-processing step to create inputs for the reconstruction; we simply feed the data directly from the original raw images into our model to help reduce overhead and latency on the reconstruction process. We have added the reconstructed output alongside the original input to help with visual comparisons and quantitative evaluations.

This whole process takes place in a Python-based application integrated into an operating system using scripts like app.py or main.py. The purpose of this combined setup is to facilitate seamless coordination between data acquisition, model execution, and visualization. The software's design allows for future integrations with cloud-based deployments, real-time edge computing platforms, and generative models like VAEs and GANs while maintaining an enterprise-grade, forward-looking framework that verify the effectiveness of

Autoencoders and will serve as a scalable platform for intelligent vision applications in the future.

.

VII. RESULTS AND DISCUSSION

A structured waste material dataset comprised of six primary types: Cardboard, Glass, Metal, Paper, Plastic, and Trash was used to provide a comprehensive evaluation of the proposed system for both classifying and reconstructing images. Results from the experimental results demonstrated that the classifier branch of the Autoencoder architecture was capable of learning robustly and achieved validation accuracy greater than 85%, and the screenshots of two outputs is given below.

Fig. 3. Shows output image of OpenCV Keras

Fig. 4. Shows output image of Autoencoders

While also providing an appropriate means for identifying the major structural characteristics of glass bottles and metal cans (i.e., the model demonstrated an ability to identify meaningful spatial, semantically linked patterns, rather than basic pixel level patterns). In addition, the decoder branch of the Autoencoder architecture demonstrated a strong ability to recover high quality images from classified categories based on a denoising technique that reduces the impact of background noise on the integrity of the item being classified. Finally, the output generated from the reconstruction of unknown or out-of-distribution type of objects (e.g., a human hand) appears defaced or blurry; this indicates the inherent capability of the model to demonstrate uncertainty and reliability of the reconstruction process, something that traditional versions of OpenCV–Keras architecture lack.

In addition, latent space visualization showed that visually similar objects (such as bottles with many variations) are grouped very closely together within the feature space, which facilitates better clustering and easier classification and helps improve decision making in subsequent systems. A comprehensive comparison between the traditional OpenCV-Keras pipeline versus the autoencoder-based system revealed a number of performance differences:

1. the OpenCV-Keras method used a lot of manual preprocessing;
2. the OpenCV-Keras method was more complex to train than the autoencoder;
3. the OpenCV-Keras model was more sensitive to noise;
4. the OpenCV-Keras system was reliant upon handcrafted feature engineering;
5. the quality of reconstructed images from the OpenCV-Keras method was only moderate;
6. the inference time of the OpenCV-Keras method was slower compared to the autoencoder;
7. the OpenCV-Keras system was limited in terms of scalability while the Autoencoder method did not require preprocessing, had a balanced training complexity, provided superior suppression of noise, provided fully automated feature learning, provided high-quality reconstruction, provided an acceleration of inference speed, and demonstrated strong scalability for real world applications. Overall, the quality of the images produced from the Autoencoder provided superior clarity, less visual interference from noise, and enhanced retention of features, thereby confirming that the Autoencoder is far superior to the traditional OpenCV-Keras method with respect to modern intelligent vision systems. These findings align

with prior comparative studies favoring autoencoder-based approaches over classical pipelines. [4], [7], [13]

Overall, the empirical results validate the theoretical advantages of hybrid autoencoder systems. [1], [3]
.

VIII. CONCLUSION AND FUTURE SCOPE

This research describes a hybrid autoencoder-based framework for intelligent waste classification and image reconstruction. The combination of unsupervised reconstruction with supervised classification produces an intelligent vision system that is highly accurate, interpretable and robust. The dual-output autoencoder architecture is able to learn both deep structural features of objects and discriminative features for classification; therefore it does not rely on low-level attributes such as surface features or manual preprocessing like traditional OpenCV-Keras pipelines do to classify objects.

The reconstruction side of the proposed autoencoder-based system generates a visually authentic representation of the original input image and provides intrinsic validation by indicating uncertainty for unknown object classes (due to distortions or blurriness) through the accuracy of the generated reconstruction. The two-sided operation creates an increase in reliability, interpretability, and decision transparency – these factors are vital to successful implementations within real-world smart waste management systems. Analysis shows that an Autoencoder approach provides better noise reduction, automation in feature extraction, faster processing times, and greater scalability than an OpenCV-Keras framework; therefore it provides a more efficient option to develop intelligent vision systems in the future.

There is a multitude of new possibilities opened up from this study in order to grow and implement applications at an industrial level. One area of future continued development is Anomaly Detection by assessing the Reconstructive Error as a method to identify classes of unknown, dangerous or out-of-distribution waste materials. Anomaly Detection could potentially be a more intelligent and safe way to sort materials. Another possible area for future development would include deploying the trained model at the Edge by quantizing and optimizing it for use with low power hardware (e.g. Raspberry Pi or NVIDIA Jetson) so that it could have real-time capability when used in smart recycling bins without having to use cloud computing. Future studies will also explore and expand upon the Generative capabilities of Variational Autoencoders (VAE's) in order to use the output from the VAE's to feed through to a Generative Adversarial Network (GAN) which will allow for higher quality image reconstruction than what currently exists. Integrated robotics is also a third major area of future extension from this work, where the classified results from the proposed method would be passed as inputs into an industrial robotic type of arm to physically segregate the materials accordingly into their respective containers (bins). All of these potential future work areas will come together to create a scalable and sustainable revoltionary solution to next generation smart waste management and environmentally automated systems.

REFERENCES

1. S. Chen et al., "Auto-Encoders in Deep Learning—A Review with New Perspectives," Mathematics, vol. 11, no. 8, p. 1777, 2023. [Online]. Available: https://www.mdpi.com/2227-7390/11/8/1777
2. A. A. Neloy et al., "A comprehensive study of auto-encoders for anomaly detection," ScienceDirect, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666827024000483
3. E. Pintelas and G. Marastos, "A Convolutional Autoencoder Topology for Classification in High-Dimensional Image Domains," PMCID, 2021. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC8622369/
4. Y. Zhang, L. Kong, M. F. Antwi-Afari, and Q. Zhang, "An Integrated Method Using a Convolutional Autoencoder, Thresholding Techniques, and a Residual Network for Anomaly Detection," Buildings, vol. 14, no. 9, p.

2828, 2024. [Online]. Available: https://doi.org/10.3390/buildings14092828
5. D. Fazel-Zarandi et al., "Supervised COSMOS Autoencoder: Learning Beyond the Euclidean Loss," arXiv, 2018. [Online]. Available: https://arxiv.org/abs/1810.06221
6. Z. Lin, Y. Chen, X. Zhao, and G. Wang, "Spectral-Spatial Classification of Hyperspectral Image Using Autoencoders," arXiv, 2015. [Online]. Available: https://arxiv.org/abs/1511.02916
7. Y. Wang, W. J. Zhao, J. Xu, and R. Hong, "Recyclable Waste Identification Using CNN Image Recognition and Gaussian Clustering," arXiv, 2020. [Online]. Available: https://arxiv.org/abs/2011.01353
8. Y. Zhang, K. Lee, and H. Lee, "Augmenting Supervised Neural Networks with Unsupervised Objectives for Large-Scale Image Classification," arXiv, 2016. [Online]. Available: https://arxiv.org/abs/1606.06582
9. "Autoencoder," Wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/Autoencoder
10. "Variational Autoencoder," Wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/Variational_autoencoder
11. "Generative adversarial network," Wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/Generative_adversarial_network
12. "Convolutional neural network," Wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/Convolutional_neural_network
13. Y. Wang and X. Zhang, "Multilayer Hybrid Deep-Learning Method for Waste Classification and Recycling," Computational Intelligence Neuroscience, Nov. 2018. [Online]. Available: https://www.researchgate.net/publication/338287968_Waste_Classification_using_AutoEncoder_Network_with_Integrated_Feature_Selection_Method_in_Convolutional_Neural_Network_Models
14. A. M. Hafidz Azis and Y. Saputra, "Comparative Analysis of VAE and GAN Algorithms for Image Classification," Elektronik Sistem Informasi, 2023. [Online]. Available: https://www.researchgate.net/publication/389861699_Comparative_Analysis_of_Variational_Autoencoder_VAE_and_Generative_Adversarial_Network_GAN_Algorithms_for_image
15. P. Rivas, Deep Learning for Beginners, Independently Published, 2025.