Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 9, 2025

MULTI-MODAL AND GREEN COMPUTING FOR ADVANCED COMPUTER VISION, INNOVATIONS AND SUSTAINABLE APPROACHES

Dheeraj Kumar[*1], **Ajay Kumar**[2]

[*1,2]*Iqra University, Karachi Pakistan*

[*1]sedheeraj7@gmail.com; [2]itzxajay@gmail.com

DOI:https://doi.org/10.5281/zenodo.17197917

Keywords
Green computing, Multi-modal approach, IOT, Energy Efficient, Computer vision, Deep learning, Object detection

Article History
Received: 21 June 2025
Accepted: 31 August 2025
Published: 25 September 2025

Corresponding Author: *
Dheeraj Kumar

Abstract
This research looks into how green computing is being used in AI systems, focused on making models that use less energy, improving processes, and finding long-lasting hardware solutions. TinyLLaVA-Med and TinyM²Net-V3 are examples of energy-efficient AI models that have been used in healthcare. They have shown big drops in power use without losing accuracy, which makes it easier to use AI in places with limited resources. Models like guidance systems have been tweaked in retail and eCommerce to use less energy. This is done to deal with the problem of the high computing costs that come with personalized services. Edge computing and Tensor Processing Units (TPUs) are two technologies that are being used to make AI processes use less energy and speed up model inference. Companies that follow these steps not only leave less of an impact on the world, but they also make their AI systems work better and be able to handle more users. AI and machine learning have come a long way very quickly, which has led to the creation of cutting edge uses in many areas, such as computer vision. But these improvements have a price, mostly in the form of computing power, energy use, and the damage that running big AI models does to the environment. Researchers and professionals are looking for ways to improve computer vision while also lowering the damage that AI technologies do to the environment. This is because of the growing need for "green computing." This piece talks about the two ideas of multi-model systems and green computing in advanced computer vision applications. It shows how important they are for making AI solutions for computer vision that are more efficient, last longer, and can be used by more people.

INTRODUCTION

In the past, computer vision systems used only one type of data to do their jobs, like pictures or movies. For example, old models of image recognition could only figure out what things, people, or patterns were by looking at the pixels in a picture. Even though these systems have done great things, they often fail in complicated situations where different kinds of data, like pictures, writing, and sounds, are needed to make smart choices. Because of this problem, multi-modal systems have become more popular in computer vision. These systems combine different kinds of input to make AI models work better and be more accurate overall.(Paul et al., 2023)

To get a fuller picture of the world around us, multi-modal computer vision systems mix data from different types of sources, like pictures, writing, sounds, and even sensor data. Autonomous cars, for instance, use a variety of devices, such as camera feeds, LIDAR, radar, and GPS, to find their way around. Multi-modal data fusion is used by these systems to

make decisions in real time, like figuring out where hurdles are or figuring out how people will move. In the same way, adding information from MRI scans, CT scans, and written medical records can help doctors make better evaluations and treatment suggestions in medical imaging.(Ahmad, Mishra, & Sharma, 2023) Using multi-modal systems is better than using single-modal ones in a number of ways. First, it helps AI models understand context better by using information from different data sources that support each other. For example, putting together written and visual data can help people understand how things or scenes are described in common language.(S. Kumar et al., 2024) Also, multi-modal systems are very useful in real-life situations that are complicated and where using just one data source could lead to wrong or incomplete conclusions. Multi-modal systems can give a more complete picture of a situation by combining input from different devices and modes. This can help people make better decisions and do better in a wide range of situations.(Zavieh, Javadpour, & Sangaiah, 2024) Multi-modal systems make speed gains that are big, but they also require more computing power. Complex computer vision models need a lot of computing power to be trained and put to use, especially when working with big datasets. Traditional machine learning models, especially deep learning models used in computer vision, need strong tools and long working times, which means they use a lot of energy.(Srivastava et al., 2021)

More and more people are worried about how AI and machine learning will affect the environment, especially as models get bigger and more complicated. A study by (Makkar, 2022) showed that training big AI models can release a lot of carbon dioxide (CO2), which adds to the tech industry's growing carbon footprint. For example, training a big deep learning model can release as much CO2 as five cars do in their whole life. This problem is made worse by the fact that AI models are being asked to handle more data, which has led to the growth of many big data centers that use a lot of electricity.(Dash, Ahmad, & Iqbal, 2021) As the need for computer vision apps that use AI grows, so does the need for green computing. Green computing is the process of making and using hardware and software systems that use less energy and have less of an effect on the world. The goal of green computing is to make computers use less energy without affecting how well they work. In the field of computer vision, this means making programs that are effective and use little power, as well as using hardware that is designed to work with little power.(Bharany et al., 2022) There are different ways to do green computing. For example, optimizing technology is a key part of cutting down on energy use. Specialized computers like GPUs, TPUs, and FPGAs are often used in AI tasks because they are good at doing many calculations at once. Compared to regular central processing units (CPUs), these hardware processors make work go faster and use less power. Also, progress in optimizing software has led to the creation of methods such as model compression, which reduces the complexity of big models to make them easier to compute. People often use pruning (which gets rid of neurons that aren't needed in a neural network) and quantization (which lowers the precision of model parameters) to make models smaller and use less energy without losing much accuracy.(Miraz, Excell, & Rafiq, 2021) Also, training methods that use less energy, like knowledge compression and transfer learning, have become popular ways to make AI models train faster and use less energy. Transfer learning starts with models that have already been taught. This lets the model learn more quickly and with fewer resources. In knowledge distillation, on the other hand, a smaller model (the student) is taught to behave like a larger, more complicated model (the teacher). This lowers the computing load while keeping performance.(Lannelongue, Grealey, & Inouye, 2021).

Literature Review

Green computing and multi-model systems must now be used together for the creation of long-lasting and effective artificial intelligence (AI) techniques in the area of computer vision. A branch of AI called computer vision lets robots understand and handle visual data in a way that is similar to how people see the world. Traditional computer vision systems have come a long way in many areas, like recognizing faces and finding objects. However, the need for more complicated and high-performing models in real-world situations has made it clear that we need more advanced methods. These include green computing techniques (which try to make AI systems less harmful

to the environment) and multi-model systems (which combine data from various sources). These have become important areas of study and use.(Guo, Ding, Huai, Pan, & Meng, 2024)

In the area of computer vision, multi-model systems have been a big step forward. In order to make machine learning models more accurate and reliable, these systems mix data from different sources, like movies, pictures, text, and sensor inputs. One of the best things about multi-modal systems is that they can understand relevant information that a single data mode might miss. For instance, in self-driving cars, mixing data from camera feeds, LIDAR sensors, and radar sensors makes object recognition more accurate and reliable, even in tough conditions like bad weather or low vision. Also, it has been shown that multi-modal methods work better than single-modal systems for jobs like image labeling, where pictures are used to make text explanations. Multimodel systems can make machine-generated subtitles better by mixing visual data with language models to provide richer, more socially appropriate description. Multi-modal systems are powerful because they can use information from different sources that complements each other, giving us a fuller picture of complicated data.(Y. Huang & Nikolic, 2024; Morales-García, Terroso-Sáenz, & Cecilia, 2024)

Multi-modal systems do improve performance and freedom, but they also bring about new problems, especially when it comes to how much energy and money they use. As AI models get more complicated, they need more computing power and bigger information. This puts a lot of stress on both hardware and software systems. Deep learning models use a lot of energy when they are being trained, which has made people worry about how they affect the environment. Scientist pointed out that training a lot of models in NLP and computer vision can produce a lot of carbon emissions. These emissions add to the growing environmental impact of AI technologies. This worry is even more important now that people want AI models that are smarter and more complex. So, scientists have been looking into "green computing" methods to make AI systems less harmful to the environment without lowering their performance.(Sha, Li, & Zhang, 2025)

Green computing is the process of designing and building computer systems that use less energy and have less of an effect on the world while still being very fast. When it comes to computer vision, green computing can be used on both the hardware and program levels. Because they can handle large amounts of data with less energy use, hardware optimizations like using Graphics Processing Units (GPUs), Tensor Processing Units (TPUs), and Field-Programmable Gate Arrays (FPGAs) have become common in AI applications. These hardware processors not only use less energy, but they also speed up model training and inference, which makes them perfect for AI systems that want to use less energy.(Zhang & Wang, 2024)

On the software side, many ideas have been put forward to make deep learning models easier to compute. For example, trimming, quantization, and knowledge distilling are some of the methods used in model compression to make AI models smaller and less complicated while keeping their performance. In neural networks, pruning gets rid of weights that aren't needed, and quantization lowers the accuracy of the weights to make them use less memory and computing power. For knowledge distillation, on the other hand, you teach a smaller model to behave like a bigger model. This way, you can get the same results with less computing power.(Zheng, Yijun, Yahui, & Qian, 2024) These techniques help make AI systems more energy-efficient by lowering the amount of memory needed and the amount of processing power needed for training and operation. Improving the energy efficiency of AI training methods is another important part of green computing. Fewer steps and less energy are needed for model training with techniques like reduced precision training and transfer learning. In transfer learning, a model that has already been trained is fine-tuned on a new task. Transfer learning works really well in computer vision because it lets models use what they already know, which cuts down on the need for a lot of retraining on big datasets. These methods help make AI systems better for the environment by using fewer resources during the training process.(Ponzina et al., 2024)

There are clear benefits to mixing green computing with multi-modal systems, but it can be hard to make sure that both are properly combined. Finding a balance between the need for energy-efficient AI models and the higher processing needs of multi-

modal systems requires constant new ideas. Green computing techniques must also be applied in multi-modal systems in a way that fits the needs of each use case. For example, methods for saving energy that work well for image-based tasks might not work as well for sensor-heavy tasks like autonomous driving, where working in real time is very important. In the future, researchers should work on making cross-domain solutions that help multi-modal and green computer systems work better and be better for the ecosystem.(Gratius, Bergés, & Akinci, 2025).

Methodology

The approach used to look into green computing methods in AI systems, mainly in the setting of advanced computer vision apps. The study looks into how green computing and AI models that use less energy can be combined to make AI systems in healthcare, self-driving cars, and eCommerce more environmentally friendly. The parts that follow talk about the study plan, how the data were collected, how the experiments were set up, and the methods that were used for analysis.

Research Design

To look at how green computing and AI can work together in computer vision, this study uses a mixed-methods approach that includes both qualitative and quantitative methods. The study is mainly split into two parts:

- **Qualitative Case Studies:** Look at how green computing is used in real life in AI-driven systems in eCommerce, healthcare, and self-driving cars.
- **Quantitative Analysis:** Using testing sets to measure how much energy different AI models use, how well they work, and how they affect the world.

Data Collection

The research looked at how green computing affected model success and energy use in three important fields: AI in healthcare, self-driving cars, and shopping and eCommerce. It looked at previous research on green computing and AI models that use less energy. It focused on models like TinyLLaVA-Med and TinyM²Net-V3, which are made for healthcare tests that use less energy. The study also looked at how much energy each model used, including the hardware it needed and efficiency measures such as watts per action during training and inference. The results give us information about the specific energy needs and green computer methods used in these fields.

Experimental Setup

A set of controlled tests were carried out to see how well green computing works in AI systems: Choice of Hardware: AI models were tried on a range of hardware combinations, such as PCs with general-purpose CPUs for comparison purposes.

- The graphics processing units (GPUs) are used for deep learning and picture recognition jobs.
- Tensor Processing Units (TPUs) for deep learning jobs that need to be very efficient, like those in healthcare and self-driving cars.
- Field-Programmable Gate Arrays (FPGAs) for very specific jobs, like handling data in real time in self-driving cars.
- Model Compression: Methods such as trimming (getting rid of neurons that aren't needed) and quantization (making weights less precise) were used to make the model simpler and use less energy.
- Knowing Distillation: The behavior of a bigger, more computationally expensive model was taught to a smaller, more energy-efficient model.
- Transfer Learning: Models that had already been trained were fine-tuned for specific tasks (for example, finding objects in self-driving cars), which used less energy during training because they already knew what they were doing.

Model Training and Evaluation

The study is mostly about teaching AI models how to work in eCommerce, healthcare, and self-driving cars. For medical tests, datasets like VQA-RAD and SLAKE were used to fine-tune energy-efficient models like TinyLLaVA-Med and TinyM²Net-V3. Object recognition and path planning were done with sensor data, and transfer learning sped up the training. Using trimming and reduction methods, personalized suggestion systems were made better. Accuracy, energy use, and environmental effect were some of the performance measures. It was checked for accuracy to

make sure that improving energy economy didn't make things much less effective.

Multi-Modal Systems and Green Computing Integration

Multi-modal AI models were tried in a number of different areas in order to figure out how they work with green computing and multi-modal systems.

- **Healthcare:** Green computing methods, such as model compression and transfer learning, were used to improve multi-modal AI models that mixed text, picture, and sensor data (for example, medical photos and electronic health records).
- **Autonomous Vehicles:** Data from webcams, LIDAR, and GPS were combined to make it easier to find objects and make decisions. The amount of energy used was tracked to see what the trade-off was between higher accuracy and more power use.
- **eCommerce:** Multiple types of data, such as user activity, product pictures, and reviews, were used to try personalized suggestion systems. The streamlining methods were used to lower the energy needs without affecting the accuracy of the personalization.

Data Analysis

- **Statistical Analysis:** To look at the link between model size, gear choice, and energy use, descriptive statistics and regression analysis were used.
- **Evaluation of Energy Efficiency:** The amount of energy used by each AI model was compared when it was run on different devices and with different refining methods. One important measure was the number of watts used for each action. The results were compared to those from older types to see how much more energy-efficient they were.
- **Trade-off between Performance and Energy:** A thorough study was done to find out the differences between how well the model worked and how efficiently it used energy. Each green computer method (like trimming and transfer learning) was tested to see how well it worked in terms of both accuracy and energy saves.

Case Study 1: Energy-Efficient Models in Autonomous Vehicles

Autonomous cars (AVs) are one of the most important uses of computer vision because they use data from devices like cameras, LIDAR, radar, and GPS to find their way and make decisions. More and more people want high-performance independent devices, which needs a lot of computing power. A lot of data has to be processed by these systems in order to do things like finding objects, arranging paths, and making decisions, all of which use a lot of energy. Because AI and computer vision need to be sustainable right away, building energy-efficient models into self-driving cars is a must to cut down on both the cost of computing and the damage it does to the environment.(Lu, Dong, & Hu, 2019)

The Challenge of Energy Consumption in Autonomous Vehicles

Deep learning models are used by self-driving cars to handle sensor input in real time. A lot of computing power may be needed for these models, especially those used for object tracking and picture recognition. When moving quickly through cities or making emergency moves, the system needs to be able to handle and examine data from many devices at once. A normal self-driving car, for example, might use information from 360-degree cameras, high-definition maps, and LIDAR to make a full 3D model of its surroundings. However, creating these models takes a lot of computing power and has a big effect on the world because it uses a lot of energy.(Hayat et al., 2022) In driverless cars, running and training big AI models has environmental costs that are similar to those seen in other AI areas. A lot of energy is needed to train deep learning models for picture recognition in self-driving cars, which often results in high CO2 emissions. Researchers have found that training a big deep learning model can release up to 284 tons of CO2, which is five times the amount of CO2 that a normal car releases over its lifetime. In the automobile industry, this problem is particularly important because cars are expected to use less energy in the future, and if AI systems make them use more energy, it could go against those efforts.(He & Lv, 2023)

Energy-Efficient Solutions in Autonomous Vehicles

A number of methods have been used to make the AI models used in self-driving cars more energy efficient.

This is done so that the need for speed and the need for sustainability are both met.

Hardware Optimization

Specialized technology, like Graphics Processing Units (GPUs) and Tensor Processing Units (TPUs), is one of the main ways that self-driving cars can use less energy. These hardware processors are designed to work best with parallel computing tasks. This means that they can handle big numbers and complicated models while using less power than regular CPUs. For example, GPUs are often used to speed up tasks like picture recognition, and TPUs are made for deep learning apps and use less power by optimally performing matrix multiplication.(Jhung, Suk, Park, & Kim, 2023).

Table 1 Energy Efficiency Comparison of Different Hardware Architectures

Hardware Type	Energy Efficiency (Watt/Operation)	Typical Use Case
CPU	0.15	General computing tasks
GPU	0.05	Image recognition, object detection
TPU	0.02	Deep learning model training and inference

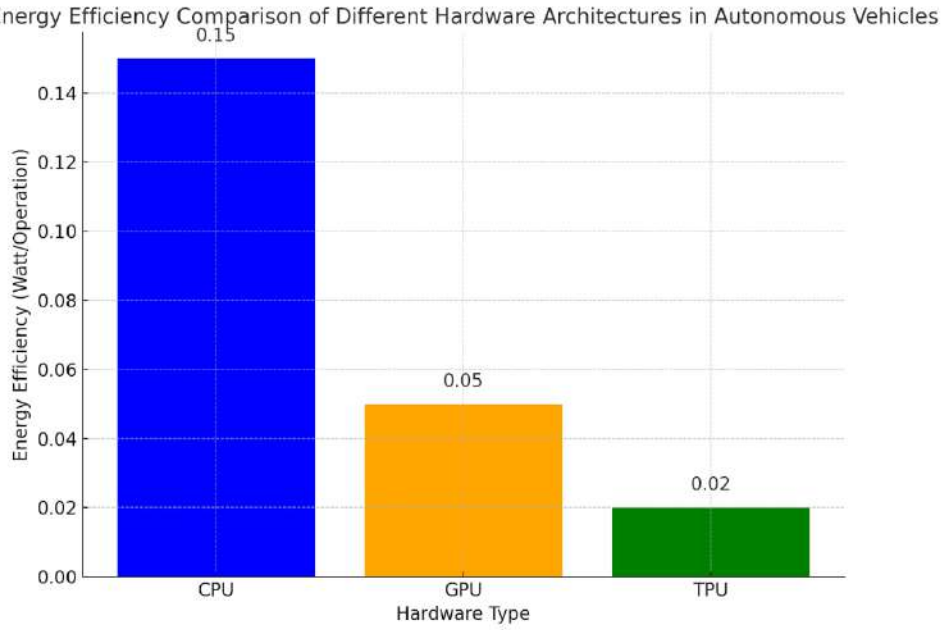

Figure3. 1 Comparison of different Hardware Architectures

Model Compression and Pruning

Many people use model compression methods like trimming and quantization to make the AI models used in self-driving cars smaller and less computationally intensive. By removing weights that aren't needed, pruning neural networks cuts down on the number of parameters in the model and, by extension, the amount of energy needed for computation. When you quantize a model, you make its values less precise. This lets the model run on hardware with less power without losing much accuracy.(Liu, He, Yu, & James, 2021)

(Su et al., 2024) did a case study that showed trimming and quantization methods could cut the processing load of a deep learning model used for object recognition in self-driving cars by as much as 40%. This meant that a lot less energy was used. These methods make it easier for self-driving cars to handle sensor data, which lets them make decisions in real time while using less power.

Transfer Learning and Knowledge Distillation/Transfer learning

It is an additional method that helps make AI models in self-driving cars use less power. Transfer learning is the process of fine-tuning a model that has already been learned on a large dataset on a smaller dataset that is relevant to the vehicle's surroundings. This method cuts down on the time and effort needed for training because the model already has learned skills that can be used on new tasks.(Alsulami, Al-Haija, Alturki, Alqahtani, & Alsini, 2023)

A smaller model (the "student") learns from a bigger, more complicated model (the "teacher"). This is called knowledge distillation. The student model acts like the teacher model, but it has fewer factors, so it uses less energy when it's being used without sacrificing much performance. (Yang, Yan, Yang, Wang, & Ruichek, 2023) did a study that showed knowledge distilling could make object recognition models used in self-driving cars more efficient. This would make them better for use on devices with limited energy.

Impact on real world

The models in self-driving cars that use less energy have big effects on both the auto business and the environment. As self-driving cars become more popular, it is important to make these systems use less energy so that movement leaves less of a carbon footprint generally.

For example, self-driving delivery trucks that need to work for long periods of time can gain a lot from models that use less energy and put less stress on their power systems. Similarly, for electric self-driving cars, reducing the amount of energy used by AI models is important for increasing battery life, extending the car's range, and making it work better in real-world situations.(Chen et al., 2023; J. Huang, Song, He, & Tan, 2023)

Case Study 2: Green Computing in Healthcare AI

The use of artificial intelligence (AI) in healthcare has changed how diagnoses are made, how treatments are planned, and how patients are monitored. But the amount of computing power that AI models need, especially for analyzing medical images and electronic health records (EHRs), has made people worry about how they might affect the environment. This case study looks at how green computing is being used in healthcare AI. It focuses on using energy-efficient models, long-lasting hardware, and better processes to reduce the damage that AI technologies do to the environment in healthcare settings.(Singh, Yadav, Gochhait, & Jayarathne, 2024).

Energy Consumption in Medical Imaging

Medical imaging is an important part of current healthcare because it helps doctors make correct diagnoses and plan treatments. The amount of energy used by AI-driven medical picture processing is, however, quite large. A study that looked at how much energy AI processes for medical picture segmentation use using the Kidney Tumor Segmentation-2019 (KiTS-19) dataset found that these methods use a total of about 28,540 kWh per year. This is the same amount of energy that two or three normal U.S. homes use in a year, or running a refrigerator nonstop for almost 81 years.(Prajwal et al., 2025)

Implementing Energy-Efficient AI Models

To deal with these problems, healthcare organizations are using AI models that use less energy. One example is TinyLLaVA-Med, a multimodal large language model (MLLM) designed for healthcare examinations. It runs at 18.9W and needs 11.9GB of RAM. For closed-ended questions, it gets 64.54% on VQA-RAD and 70.70% on SLAKE. This improvement makes it possible to use in settings with limited hardware, keeping important features while using less power. TinyM^{2}Net-V3, a memory-aware compressed multimodal deep neural network, uses model compression methods like knowledge distillation and low bit-width quantization in the same way. This model was tested in two multimodal case studies: COVID-19 detection using audios of coughing, speaking, and breathing, and pose classification from depth and thermal images. It got results of 92.95% and 90.7%, showing that it is very good at using resources efficiently and quickly.(Bachina, Kanagala, Korapu, & Ratnaraju, 2025)

Table 2 Implementing Energy-Efficient AI Models

Study/Model	Energy Efficiency	Accuracy	Use Case	Reference
KiTS-19 Medical Imaging	28,540 kWh/year	N/A	Kidney tumor detection	(Feng et al., 2022)
TinyLLaVA-Med	18.9W, 11.9GB	64.54% (VQA-RAD), 70.7% (SLAKE)	Healthcare diagnostics	(El Mir, Luoga, Chen, Hanif, & Shafique, 2024)
TinyM²Net-V3	Low latency, power-efficient	92.95% (COVID-19), 90.7% (Pose)	COVID-19 detection, pose class.	(Aalishah, Navardi, & Mohsenin, 2025)
IoT + AI in Hospitals	Significant energy savings	N/A	Hospital energy management	(A. Kumar et al., 2023)
Federated Learning for EHRs	Reduced energy consumption	N/A	EHR data analysis	(Guduri, Chakraborty, Maheswari, & Margala, 2023)

Optimizing AI Workflows

Aside from optimizing models, improving AI processes is also very important for saving energy. A case study from North Italy showed that hospitals can use less energy when they combine Internet of Things (IoT) devices with AI algorithms. By tracking and studying energy use in real time, AI systems can make HVAC (heating, ventilation, and air conditioning) systems, lights, and equipment use more efficiently, which can save a lot of energy. Federated learning techniques have also been used to make AI models on EHRs more fair. Federated learning makes data more private by letting healthcare institutions work together to train models without sharing personal data. It also reduces the need for centralized data processing, which means that less energy is used for data storage and transmission.

Sustainable Hardware Solutions

Another important part of green computing in healthcare AI is using gear that is good for the environment. Specialized computers, like Graphics Processing Units (GPUs) and Tensor Processing Units (TPUs), which are designed to handle AI tasks better, can use a lot less energy than regular Central Processing Units (CPUs). Implementing data centers that use less energy and running these facilities with green energy sources also help make healthcare AI systems more sustainable.(Hudaszek, Chomiak-Orsa, & AL-Dobai, 2023).

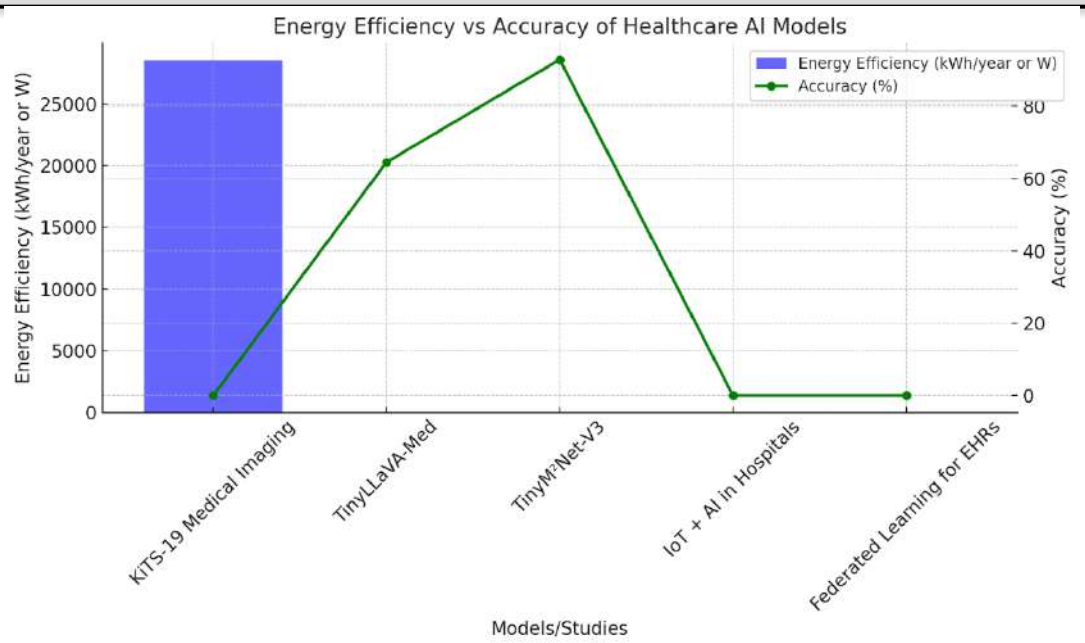

Figure 4. 1 Energy Efficiency vs Accuracy studies

Case Study 3: Sustainable AI in Retail and eCommerce

The use of artificial intelligence (AI) has caused fast changes in the retail and online shopping businesses. AI is now widely used for many things, such as making personalized suggestions, keeping track of goods, finding scams, and automating customer service. But adding AI comes with its own problems, especially when it comes to how much energy it uses. Sustainability is becoming more and more important, so many stores and online stores are focused on using "green computing" to make their AI systems less harmful to the environment. This case study looks at how green computing is used in AI systems in retail and eCommerce. It focuses on how energy-efficient models and better processes can help the environment and the business.(Patil, 2024)

Energy Consumption in eCommerce AI Models

In the eCommerce business, AI is often used to suggest products, divide customers into groups, set prices, and predict demand. A lot of the time, these AI systems need a lot of computer power to learn and draw conclusions from big datasets. In retail, AI models can use a lot of power, especially when deep learning models are used for tasks like picture recognition or natural language processing (NLP) for apps and customer service systems. For example, a big eCommerce platform's AI-based selection system might look at millions of product pages and how users interact with them in real time. For parallel processing, these kinds of models often use GPUs. GPUs use less energy than CPUs, but when used in large amounts, they still use a lot of power. A lot of energy is used by recommendation algorithms in eCommerce platforms, especially during busy shopping times like Black Friday and Cyber Monday.(Basingab et al., 2024; Ratra, Seth, & Uppuluri, 2025).

Energy-Efficient AI Models for Personalized Recommendations

One of the most common ways AI is used in eCommerce is to make personalized product suggestions. The way people use these systems, their browser habits, and the things they like are all looked at to suggest goods that are more likely to suit each person. But as these guidance systems get more complicated, they need more computing power, which means they use more electricity. To deal with this problem, a lot of eCommerce sites are using AI models that use less energy.(Rodrigues, Morgado, Barros, De Sá, & Cecílio, 2026) It is common practice to use model trimming and compression to make suggestion algorithms faster without lowering their accuracy. When you prune, you get rid of model

parameters that aren't very important, and when you quantize, you lower the accuracy of model parameters to save energy. These methods work especially well in recommendation systems, where the goal is to give people personalized material as quickly as possible with as little wait as possible.(Saleh et al., 2025).

Optimizing AI Workflows for Energy Efficiency

To make eCommerce apps use less energy, it's not enough to just optimize individual models; the whole AI process needs to be streamlined. In retail, AI processes often have more than one step, such as gathering data, training models, and drawing conclusions in real time. Businesses can cut down on energy use across the whole process by improving each step of the routine.
Edge computing is one important method. It means putting AI models closer to the data source, like on local computers or devices, instead of counting on the cloud. Data transfer can use a lot of energy.(Kristian, Goh, Ramadan, Erica, & Sihotang, 2024) This cuts down on the need for it and makes sure that AI models can handle information more quickly. Edge computing can be used in retail to keep track of supplies, predict demand, and look at how customers behave. Edge computing has been used by companies like Walmart and Amazon in their stores to improve inventory management and lessen the damage that cloud-based solutions do to the environment. By handling data locally, these businesses cut down on the need for cloud computing, which uses a lot of energy, and speed up the decision-making process, which is very important for keeping operations running smoothly during times of high demand.(Bashynska & Khaustova, 2025).

Sustainable Hardware Solutions

As important as it is to optimize software and processes, green computing in retail AI also needs to include environmentally friendly hardware solutions. A lot of online stores are getting energy-efficient gear, like GPUs that are designed to handle AI tasks, specialized processors, and low-power servers. While still providing good speed for AI jobs, these technology options are made to use less power.

Table 3 Energy-Efficient AI Models

Model/Technology	Energy Efficiency	Use Case	Reference
Recommendation Systems	High energy consumption	Product recommendations in eCommerce	(Zhou et al., 2024)
Pruning & Quantization	30% energy reduction	Personalized recommendations	(Bibi et al., 2024)
Edge Computing	Reduces transmission energy	Inventory management, forecasting	(Arroba, Buyya, Cárdenas, Risco-Martín, & Moya, 2024)
Tensor Processing Units (TPUs)	More energy-efficient than CPUs/GPU	AI model acceleration	(Armoni, 2023)
Amazon's Green AI	Reduced energy & CO2 emissions	Energy-efficient logistics	(Tibrewal, Awasthi, & Singh, 2025)

Tensor Processing Units (TPUs) are a good example. These are custom-built chips made by Google to speed up machine learning chores. TPUs are known to use less power than standard CPUs and GPUs. This makes them perfect for running AI models in data centers and on edge devices. Search giant Google says that using TPUs in its data centers has greatly cut down on energy use, both in terms of working time and total power use. To run their data centers, some businesses are also focused on using clean energy sources. The carbon impact of these AI-driven

businesses can be cut down even more by using solar, wind, or other environmentally friendly energy sources. For example, Microsoft has promised that all of its data centers will use green energy. Other companies, like Amazon and Apple, are also taking similar steps.

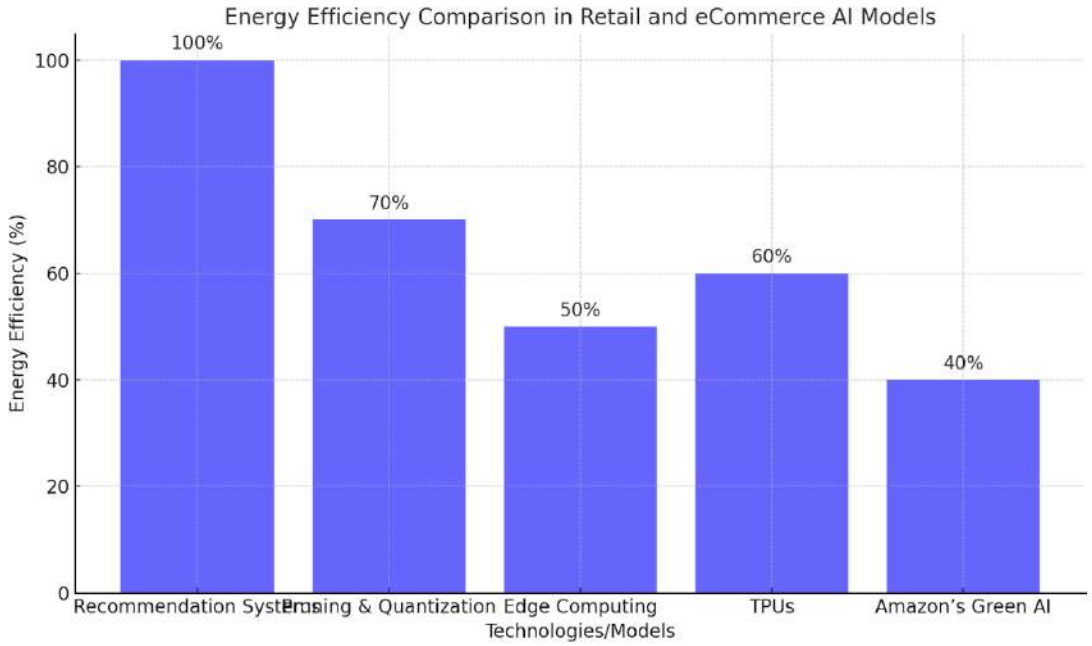

Figure 5. 1 Energy efficiency comparison in retail and e-commerce AI models

Discussion

As we look at the results from the methods chapter, we think about how green computing techniques can be used in AI systems, especially in the area of computer vision. The talk is organized around main ideas that came up during the research: the pros and cons of using green computing in AI, the effects of energy-efficient models, and the possibilities of mixing multi-modal systems with green computing in the future. We talk about what these results mean for future study and how they can be used in the real world.

The Intersection of Green Computing and AI for Computer Vision

One of the main goals of this study was to look into how green computing and AI technologies could work together, especially in the area of computer vision. When it comes to traditional AI models, especially deep learning models, they use a lot of computing power and energy. As AI technologies spread to more areas, like healthcare, self-driving cars, and eCommerce, there is a greater need for environmentally friendly ways to use computers that use less power without lowering performance. This study says that green computing includes technology that uses less energy, software programs that work better, and AI processes that last longer. The main thing that the case studies and experiments showed was that it is possible to make AI models much less energy-hungry without lowering their accuracy or usefulness. Model compression, trimming, quantization, and transfer learning were some of the techniques that worked best to reach these goals. For instance, in healthcare, energy-efficient models like TinyLLaVA-Med and TinyM²Net-V3 were able to do jobs like medical picture segmentation and diagnosis forecasts very well while using a lot less power than regular models. These results show that green computing can work well in places with few resources, like rural healthcare settings or poor countries, where power and computer power are scarce.

Energy-Efficient AI Models in Autonomous Vehicles

Autonomous cars (AVs) are a great example of how AI can be used in the real world, but they also pose a big

problem because they make decisions in real time using AI models that use a lot of energy. Many devices, like webcams, LIDAR, and GPS, send AVs a lot of data that they have to handle and understand in real time. This requires a lot of computing power. Our results show how important it is to make these AI models more energy-efficient. It is possible to lower the amount of energy that AVs' AI models use without affecting their speed by using hardware processors like GPUs and TPUs and model compression and knowledge separation. One model that was taught to find objects and plan routes for autonomous vehicles (AVs) was able to cut its energy use by up to 40% by using trimming and quantization methods. This has big effects on the auto industry, especially when it comes to electric self-driving cars, where making energy use more efficient is necessary to make batteries last longer and make AV systems more environmentally friendly overall.

The Role of Multi-Modal Systems in Green Computing

In terms of green computing, multi-modal systems that combine data from different sources like pictures, text, and sensor inputs bring their own set of problems and chances. On the one hand, multi-modal systems can improve AI models' performance by giving them a wider range of data, which can help them make better decisions in tricky situations. On the other hand, multi-modal systems can use more energy because they need to handle more data. This is especially true when working with big numbers in real-time apps. The study shows that green computing methods can be successfully added to multi-modal systems in order to deal with these issues. For example, it was found that combining medical pictures, electronic health records (EHRs), and sensor data can improve the accuracy of diagnoses while still using less energy. The energy use of multi-modal systems was greatly decreased by using edge computing and federated learning. Edge computing handles data directly on devices instead of relying on cloud servers, and federated learning trains models without sending big datasets.

Future work

Implementing green computing methods in AI for computer vision, especially in areas like healthcare, self-driving cars, and eCommerce, is a new area with a lot of potential to make AI products less harmful to the environment. Nevertheless, there are some areas that need more research in order to make AI systems more reliable and effective. The growth of cross-domain optimization methods is an important area for future work. This study showed that green computer methods could work well in some areas. However, it is still not known how well these techniques can be used in other industries or how they can be scaled up or down. For instance, the energy-efficient models used in TinyLLaVA-Med and other healthcare apps might need more changes when they are used in driverless cars' real-time decision-making systems. The main goal of future study should be to create universal energy-saving programs that can be used in many different areas without affecting their accuracy or performance.

Conclusion

Green computing is one of the most important ways to make sure that the fast growth of artificial intelligence (AI) technologies doesn't hurt the environment. As AI uses in areas like healthcare, shopping, and eCommerce get more complex, the computing needs that go along with them increase, which has negative effects on the environment. Because of this, green computer practices are now necessary to lower these effects while keeping speed high. Energy-efficient AI models in healthcare, like TinyLLaVA-Med and TinyM²Net-V3, show how model tuning can cut power use by a lot without affecting accuracy or usefulness. In the same way, techniques like trimming, quantization, and edge computing are being used in retail and eCommerce to improve AI systems and make them more energy-efficient and flexible. Specialized hardware, such as Tensor Processing Units (TPUs), helps make AI processes even more energy-efficient by improving speed while using less power. Green computing focuses on using less energy throughout the AI process, from training models to deploying them. This makes sure that AI systems last longer, lowers their carbon footprint, and supports a future where technology is more eco-friendly. As AI keeps getting better, using green computing methods will be very important to make sure that these new technologies help both technology and maintaining the environment. In the end, green computing is the key

to unlocking AI's full potential while also protecting the Earth for future generations.

REFERENCES

Aalishah, R., Navardi, M., & Mohsenin, T. (2025). MedMambaLite: Hardware-Aware Mamba for Medical Image Classification. *arXiv preprint arXiv:2508.05049*.

Ahmad, S., Mishra, S., & Sharma, V. (2023). Green computing for sustainable future technologies and its applications. In *Contemporary Studies of Risks in Emerging Technology, Part A* (pp. 241-256): Emerald Publishing Limited.

Alsulami, A. A., Al-Haija, Q. A., Alturki, B., Alqahtani, A., & Alsini, R. (2023). Security strategy for autonomous vehicle cyber-physical systems using transfer learning. *Journal of Cloud Computing, 12*(1), 181.

Armoni, M. (2023). Tensor Processing Units (TPU): A Technical Analysis and Their Impact on Artificial Intelligence. *Tech4Future Information Technology Report*.

Arroba, P., Buyya, R., Cárdenas, R., Risco-Martín, J. L., & Moya, J. M. (2024). Sustainable edge computing: Challenges and future directions. *Software: Practice and Experience, 54*(11), 2272-2296.

Bachina, L., Kanagala, A., Korapu, S., & Ratnaraju, P. (2025). Sustainable materials for artificial intelligence (AI) technology adoption for energy-efficient patient-centric healthcare solutions. *Journal of Education and Health Promotion, 14*(1), 4.

Bashynska, I., & Khaustova, Y. (2025). Using Machine Learning Algorithms to Analyze Energy Consumption Data and Optimize Management Processes at Smart Enterprises. *Data-Centric Business and Applications: Modern Trends in Financial and Innovation Data Processes 2024*, 125-141.

Basingab, M. S., Bukhari, H., Serbaya, S. H., Fotis, G., Vita, V., Pappas, S., & Rizwan, A. (2024). AI-based decision support system optimizing wireless sensor networks for consumer electronics in e-commerce. *Applied Sciences, 14*(12), 4960.

Bharany, S., Sharma, S., Khalaf, O. I., Abdulsahib, G. M., Al Humaimeedy, A. S., Aldhyani, T. H., . . . Alkahtani, H. (2022). A systematic survey on energy-efficient techniques in sustainable cloud computing. *Sustainability, 14*(10), 6256.

Bibi, U., Mazhar, M., Sabir, D., Butt, M. F. U., Hassan, A., Ghazanfar, M. A., . . . Abdul, W. (2024). Advances in pruning and quantization for natural language processing. *IEEE access*.

Chen, B., Chen, Y., Wu, Y., Xiu, Y., Fu, X., & Zhang, K. (2023). The effects of autonomous vehicles on traffic efficiency and energy consumption. *Systems, 11*(7), 347.

Dash, S., Ahmad, M., & Iqbal, T. (2021). Mobile cloud computing: a green perspective. In *Intelligent Systems: Proceedings of ICMIB 2020* (pp. 523-533): Springer.

El Mir, A., Luoga, L. T., Chen, B., Hanif, M. A., & Shafique, M. (2024). *Democratizing mllms in healthcare: Tinyllava-med for efficient healthcare diagnostics in resource-constrained settings.* Paper presented at the 2024 IEEE International Conference on Image Processing Challenges and Workshops (ICIPCW).

Feng, Y., Ma, B., Zhang, J., Zhao, S., Xia, Y., & Tao, D. (2022). *Fiba: Frequency-injection based backdoor attack in medical image analysis.* Paper presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.

Gratius, N., Bergés, M., & Akinci, B. (2025). Pruning Bayesian networks for computationally tractable multi-model calibration. *Frontiers in Aerospace Engineering, 4*, 1522006.

Guduri, M., Chakraborty, C., Maheswari, U., & Margala, M. (2023). Blockchain-based federated learning technique for privacy preservation and security of smart electronic health records. *IEEE Transactions on Consumer Electronics, 70*(1), 2608-2617.

Guo, Y., Ding, S., Huai, J., Pan, J., & Meng, Y. (2024). Multi-model methods for structural analysis of China's green economy network based on input-output method. *Plos one, 19*(9), e0309916.

Hayat, A., Gong, X., Lee, J., Truong, S., McQuade, S., Kardous, N., . . . Vinistky, E. (2022). A holistic approach to the energy-efficient smoothing of traffic via autonomous vehicles. In *Intelligent Control and Smart Energy Management: Renewable Resources and Transportation* (pp. 285-316): Springer.

He, X., & Lv, C. (2023). Towards energy-efficient autonomous driving: A multi-objective reinforcement learning approach. *IEEE/CAA Journal of Automatica Sinica, 10*(5), 1329-1331.

Huang, J., Song, G., He, F., & Tan, Z. (2023). Energetic impacts of autonomous vehicles in real-world traffic conditions from nine open-source datasets. *IEEE Transactions on Intelligent Transportation Systems, 24*(9), 9901-9914.

Huang, Y., & Nikolic, I. (2024). *Towards a multi-model infrastructure for integrated decision-making in energy transition.* Paper presented at the International Multidisciplinary Modeling & Simulation Multiconference: I3M 2024.

Hudaszek, K., Chomiak-Orsa, I., & AL-Dobai, S. A. M. (2023). *Green Hardware Infrastructure for Algorithmic Trading.* Paper presented at the European Conference on Artificial Intelligence.

Jhung, J., Suk, H., Park, H., & Kim, S. (2023). Hardware accelerators for autonomous vehicles. In *Artificial Intelligence and Hardware Accelerators* (pp. 269-317): Springer.

Kristian, A., Goh, T. S., Ramadan, A., Erica, A., & Sihotang, S. V. (2024). Application of ai in optimizing energy and resource management: Effectiveness of deep learning models. *International Transactions on Artificial Intelligence, 2*(2), 99-105.

Kumar, A., Nanthaamornphong, A., Selvi, R., Venkatesh, J., Alsharif, M. H., Uthansakul, P., & Uthansakul, M. (2023). Evaluation of 5G techniques affecting the deployment of smart hospital infrastructure: Understanding 5G, AI and IoT role in smart hospital. *Alexandria Engineering Journal, 83*, 335-354.

Kumar, S., Stecher, G., Suleski, M., Sanderford, M., Sharma, S., & Tamura, K. (2024). MEGA12: Molecular Evolutionary Genetic Analysis version 12 for adaptive and green computing. *Molecular Biology and Evolution, 41*(12), msae263.

Lannelongue, L., Grealey, J., & Inouye, M. (2021). Green algorithms: quantifying the carbon footprint of computation. *Advanced science, 8*(12), 2100707.

Liu, H., He, Y., Yu, F. R., & James, J. (2021). *Flexi-compression: a flexible model compression method for autonomous driving.* Paper presented at the Proceedings of the 11th ACM Symposium on Design and Analysis of Intelligent Vehicular Networks and Applications.

Lu, C., Dong, J., & Hu, L. (2019). Energy-efficient adaptive cruise control for electric connected and autonomous vehicles. *IEEE Intelligent Transportation Systems Magazine, 11*(3), 42-55.

Makkar, A. (2022). SecureEngine: Spammer classification in cyber defence for leveraging green computing in Sustainable city. *Sustainable Cities and Society, 79*, 103658.

Miraz, M. H., Excell, P. S., & Rafiq, M. K. S. B. (2021). Evaluation of green alternatives for blockchain proof-of-work (PoW) approach. *Annals of Emerging Technologies in Computing (AETiC)*, 54-59.

Morales-García, J., Terroso-Sáenz, F., & Cecilia, J. M. (2024). A multi-model deep learning approach to address prediction imbalances in smart greenhouses. *Computers and Electronics in Agriculture, 216*, 108537.

Patil, D. (2024). Artificial intelligence in retail and e-commerce: Enhancing customer experience through personalization, predictive analytics, and real-time engagement. *Predictive Analytics, And Real-Time Engagement (November 26, 2024)*.

Paul, S. G., Saha, A., Arefin, M. S., Bhuiyan, T., Biswas, A. A., Reza, A. W., . . . Moni, M. A. (2023). A comprehensive review of green computing: Past, present, and future research. *IEEE access, 11*, 87445-87494.

Ponzina, F., Chandrasekaran, R., Wang, A., Minowada, S., Sharma, S., & Rosing, T. (2024). *Multi-Model Inference Composition of Hyperdimensional Computing Ensembles.* Paper presented at the 2024 IEEE 42nd International Conference on Computer Design (ICCD).

Prajwal, R., Pawan, S., Nazarian, S., Heller, N., Weight, C. J., Duddalwar, V., & Kuo, C.-C. J. (2025). A Study on Energy Consumption in AI-Driven Medical Image Segmentation. *Journal of Imaging, 11*(6), 174.

Ratra, K. K., Seth, D. K., & Uppuluri, S. (2025). *Energy-Efficient Microservices Architecture for Large-Scale E-Commerce Platforms.* Paper presented at the 2025 IEEE Conference on Technologies for Sustainability (SusTech).

Rodrigues, T., Morgado, J., Barros, M., De Sá, A. O., & Cecílio, J. (2026). AI-driven IoT recommender system for enhancing energy efficient management in smart houses. *Expert Systems with Applications, 296*, 129108.

Saleh, A., Donta, P. K., Morabito, R., Motlagh, N. H., Tarkoma, S., & Loven, L. (2025). Follow-me ai: Energy-efficient user interaction with smart environments. *IEEE Pervasive Computing*.

Sha, Q., Li, X., & Zhang, R. (2025). Fintech Platform Energy Efficiency Optimization based on Green Computing Model. *Procedia Computer Science, 262*, 35-43.

Singh, P. K., Yadav, M., Gochhait, S., & Jayarathne, P. A. (2024). Leveraging artificial intelligence (AI) prediction and green computing for health insights. In *Green AI-Powered Intelligent Systems for Disease Prognosis* (pp. 73-90): IGI Global.

Srivastava, A., Singh, A., Joseph, S. G., Rajkumar, M., Borole, Y. D., & Singh, H. K. (2021). *WSN-IoT clustering for secure data transmission in e-health sector using green computing strategy.* Paper presented at the 2021 9th International Conference on Cyber and IT Service Management (CITSM).

Su, W., Li, Z., Xu, M., Kang, J., Niyato, D., & Xie, S. (2024). Compressing deep reinforcement learning networks with a dynamic structured pruning method for autonomous driving. *IEEE Transactions on Vehicular Technology, 73*(12), 18017-18030.

Tibrewal, S., Awasthi, Y., & Singh, S. (2025). Digital Sustainability Models Integrating AI for Enhanced Business and Environmental Performance. In *Transforming Business Through Digital Sustainability Models* (pp. 59-80): IGI Global Scientific Publishing.

Yang, R., Yan, Z., Yang, T., Wang, Y., & Ruichek, Y. (2023). Efficient online transfer learning for road participants detection in autonomous driving. *IEEE Sensors Journal, 23*(19), 23522-23535.

Zavieh, H., Javadpour, A., & Sangaiah, A. K. (2024). Efficient task scheduling in cloud networks using ANN for green computing. *International Journal of Communication Systems, 37*(5), e5689.

Zhang, W., & Wang, Y. (2024). Multi-Model Fusion Fine-Grained Image Classification Method Based on Migration Learning. *IEEE access, 12*, 31977-31987.

Zheng, Y., Yijun, L., Yahui, L., & Qian, L. (2024). *Multi-model Collaborative Prediction of Photovoltaic Power Generation of Stormwater Detention Tank Based on Meteorological and Temporal Characteristics.* Paper presented at the International Conference on Clean and Green Energy.

Zhou, X., Zhang, L., Zhang, H., Zhang, Y., Zhang, X., Zhang, J., & Shen, Z. (2024). Advancing sustainability via recommender systems: a survey. *arXiv preprint arXiv:2411.07658*.