SHM PUBLISHER
SURYA HIJAU MANFAAT

Journal of Information System Exploration and Research

https://shmpublisher.com/index.php/joiser

p-ISSN 2964-1160 | e-ISSN 2963-6361

Ensemble Deep Learning: A State-Of-The-Art Comprehensive Review

Mahmoud Mohamed Abdelsamie[1*]

[1] *Departement of Computer Science, Cairo University, Cairo, Egypt*

DOI: https://doi.org/10.52465/joiser.v4i1.614
Received 26 August 2025; Accepted 26 January 2026; Available online 28 January 2026

Article Info

Keywords:
Ensemble learning;
LLMs ensembles;
Transformers ensembles;
Taxonomy

Abstract

Ensemble learning has been a cornerstone of machine learning, providing improved predictive performance and robustness by combining multiple models. However, in the era of deep learning, the landscape of ensemble techniques has rapidly evolved, influenced by advances in neural architectures, training models, and practical application requirements. This review provides a state-of-the-art survey of ensemble deep learning approaches, focusing on recent developments of ensemble methods. We introduce a classification of ensemble strategies based on model diversity, fusion mechanisms, and task alignment, and highlight emerging techniques such as attention-based ensemble fusion, neural architecture search-based ensembles, and large ensembles of language or vision models. The review also examines theoretical foundations, practical tradeoffs, and domain-specific adaptations in some fields. Compiling state-of-the-art benchmarks, we evaluate ensemble performance in terms of accuracy, efficiency, robustness, and interpretability. We also identify key challenges such as scalability, overfitting, and deployment limitations and present open research directions, including ensemble learning for continuous learning, federated learning, and learning from scratch. By connecting key insights with current trends, this review aims to guide researchers and practitioners in designing and implementing ensemble deep learning systems to address the next generation of AI challenges.

1. Introduction

Machine learning is a field of artificial intelligence (AI) that encompasses many traditional approaches, enabling systems to learn and enhance the performance of various tasks without being explicitly programmed [1][2]. Deep learning, a part of machine learning, focuses on training neural networks using large datasets to accomplish various tasks [3]. It has transformed the artificial intelligence landscape by enabling breakthroughs in diverse fields, including computer vision, natural language processing, speech recognition, and scientific discovery. Despite the impressive capabilities

[*] ***Corresponding Author:***

Mahmoud Mohamed Abdelsamie
Department of Computer Science,
Cairo University, Cairo, Egypt
Email: 12422021452995@pg.cu.edu.eg

of deep neural networks, individual models often suffer from limitations such as overfitting, sensitivity to initialization, poor robustness to irrelevant inputs, and suboptimal generalization to unseen data [4], which are evaluated on the basis of various factors like performance and scalability [5].

Traditional learning methods always rely on baseline models or learners to handle small-scale tasks. Ensemble learning is one of the most widely used approaches across different domains and it outperforms the traditional machine learning methods [5][6].

Ensemble deep learning methods emerged when ensemble neural networks were combined, and they have been shown to reduce generalization errors [7][8]. Ensemble learning can be defined as a technique Which combines multiple base models to create one more powerful and robust model [5][7][9][10]. This combination of learners achieved superior results and success in various applications in terms of better generalization of intelligent learning system [5][9][11]. Mixing predictions from many models to enhance predictive performance is also another definition for ensemble learning [12]. In ensemble learning, using a variety of base models can reduce the risk of overfitting [5][13][14] due to the diverse perspectives each model provides [15], leading to more generalized predictions. There are several ensemble methods that differ in how the underlying models are trained and combined. The most commonly used ensemble methods include averaging, bagging, random forests, stacking, and boosting [5][16][17][18]. Each of these techniques has its own strengths and weaknesses, depending on the specific problem and characteristics of the data [19][20].

Accordingly, ensemble methods have traditionally been associated with classical machine learning algorithms [21], such as decision trees or support vector machines, and have gained renewed importance in the era of deep learning [5]. Contemporary ensemble strategies have evolved beyond simple clustering or boosting [22][23][24]. Modern approaches include a combination of diverse neural architectures, integration through attention mechanisms or gating modules [25][26][27], automated model selection using neural architecture search (NAS) [28][29], and the ensemble of large-scale models such as transformers [30] and diffusion models [31]. These developments have led to significant gains in accuracy, calibration, and robustness, particularly in high-stakes domains such as medical diagnosis [32], autonomous systems [33], and language understanding [34].

Despite the growing interest, existing surveys and reviews on ensemble deep learning suffer from several limitations. Many are outdated and focus on primarily methods [5][9][10][13][19], neglecting recent advances such as induction-based clustering for large language models (LLMs) [35], self-assembly techniques, and federated ensemble frameworks [36][37]. Other studies provide overly general overviews and lack a systematic classification or detailed assessment of trade-offs, scalability, and application-specific limitations. Furthermore, few reviews have systematically analyzed ensemble methods in the context of emerging trends such as continuous learning [38], resource-constrained scenarios [39][40], and ethical considerations such as fairness and interpretability.

This review addresses these gaps by providing a comprehensive and up-to-date study of ensemble deep learning methods, focusing on the recent trends and developments. Our main contributions are:

- A new taxonomy of ensemble deep learning methods, organized by model diversity, fusion strategy, and task alignment, offering a fresh perspective beyond traditional classifications.
- A detailed analysis of recent developments, including attention-based ensembles, ensemble techniques for large language and vision models, NAS-driven ensembles, and domain-specific adaptations.
- An in-depth evaluation of ensembles performance across dimensions such as predictive accuracy, robustness, interpretability, and computational efficiency.
- Specific case studies and application insights from emerging fields, including low-resource natural language processing tasks, medical diagnostics, and real-time systems.
- Identifying challenges and future directions, including the need for scalable clustering techniques, integration with self-supervised learning, and the design of ethical and explainable clustering systems.

Through this review, we aim to provide researchers and practitioners with the theoretical foundations, practical considerations, and strategic insights needed to leverage deep learning clustering in current and future AI systems.

2. Taxonomy of Ensemble Deep Learning

Designing ensemble deep learning models involves critical decisions about how to select, train, and integrate models [13]. While traditional ensemble deep learning methods such as clustering, boosting, and stacking provide fundamental strategies [5], they are often insufficient to describe the diversity and complexity of modern deep learning systems. In this section, we propose a new taxonomy that categorizes ensemble deep learning methods according to four orthogonal dimensions:

1. Model Diversity
2. Task Alignment
3. Fusion Stage
4. Ensemble Strategy

This taxonomy not only highlights structural differences but also provides insight into when and why different group configurations may be effective.

2.1. Based on Model Diversity

Model diversity is essential for the success of ensemble learning [5][41][42]. In deep learning, diversity can be introduced through different architectures, training models, hyperparameters, or subsets of data [43][44]. In this matter the types are presented as follows:

2.1.1. Homogeneous Ensembles:

Homogeneous ensembles consist of models with identical structures, typically trained using different initializations, random seeds, or data partitions [45][46]. A common example is training multiple ResNet-50 models [47][48][49]on re-run subsets of the same dataset. These ensembles are relatively easy to implement and are effective at reducing variance [50], leading to more stable predictions. However, their drawback is that they may not provide enough diversity [51] to significantly improve generalization, especially when the models converge to similar representations or errors.

2.1.2. Heterogeneous Ensembles:

Heterogeneous ensembles incorporate a variety of model architectures, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformers [13][52]. This structural diversity enables the ensemble to capture different feature representations and learning biases [53], often leading to improved performance [54] and robustness across all tasks. However, the problem is the increased complexity of model integration [55], calibration, and tuning, as different architectures may produce features at different scales or formats.

2.1.3. Data-Induced Diversity:

Ensembles can also achieve diversity through variations in the input data [5][56][57]. This can be achieved by training models on different perspectives, increments, or subsets of the data. A common example is self-ensemble training techniques [58], where models are exposed to strong and weak increments of the same input during training or inference. This type of diversity enhances generalization, particularly in domains with limited labeled data. However, its effectiveness may be limited if the data is inherently homogeneous or lacks meaningful variation [59].

2.1.4. Task-Induced Diversity:

Task-induced ensembles involve training models on different but related tasks such as hate speech detection, sentiment analysis, and fake news detection and then combining their outputs to achieve a common prediction goal [60]. These ensembles are particularly useful in multi-task learning scenarios, where auxiliary tasks help the model learn richer and more abstract representations. While this approach may lead to improved generalization across domains [61], it poses additional challenges in balancing the objectives of multiple tasks and ensuring that the auxiliary tasks contribute positively to the main goal [62].

Moreover, Figure.1 provides a visual diagram showing model diversity branching into architecture, data, task, and training variance.

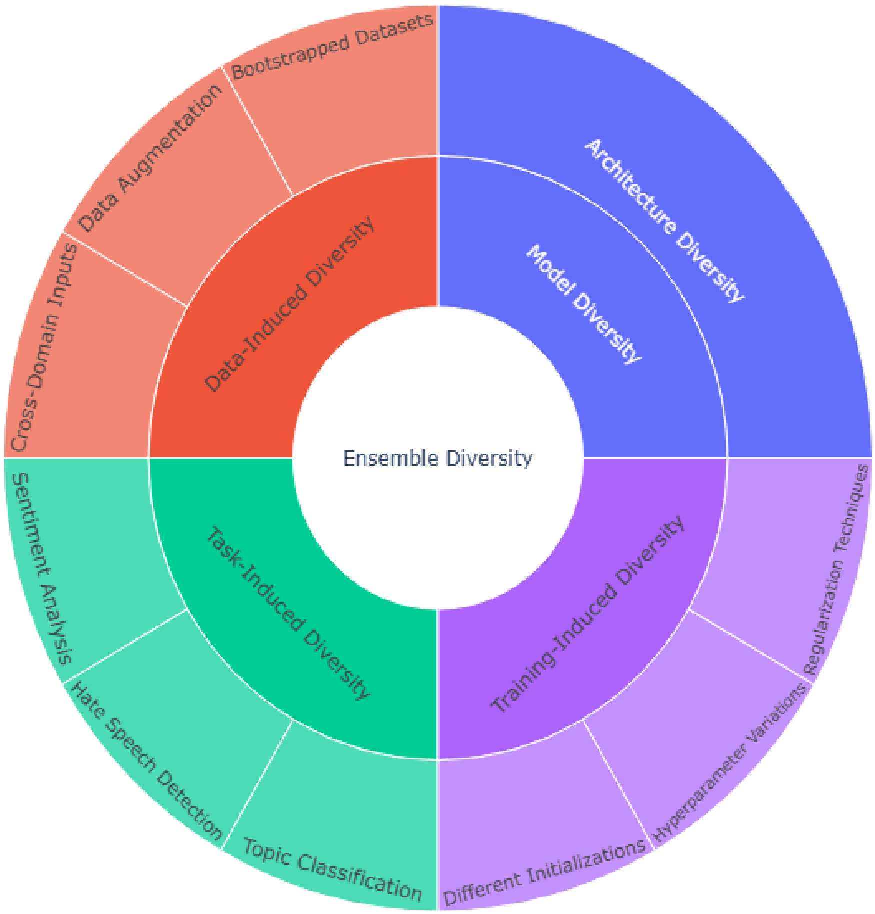

Figure. 1 Ensembles Based on Model Diversity

2.2. Based on Task Alignment

Deep ensemble methods can be optimized differently depending on the nature of the task [63]. The complexity of clustering increases with the difficulty and scope of the task.

For classification tasks, ensemble adaptation often involves techniques such as majority voting, flexible voting, or probabilistic averaging [5][64]. These methods combine the predictions of multiple models to make a final decision [5]. For examples, ensemble classifiers for image recognition [65][66] and hate speech detection systems [67][68].

For regression problems, ensemble methods typically combine model outputs using arithmetic mean or median calculations, or through weighted averaging [5][69], where some models contribute more based on their performance. This approach is frequently used in applications such as house price prediction [70][71] and weather forecasting [72].

Multi-task learning ensembles are designed to handle multiple related tasks simultaneously [73]. This is often achieved by sharing underlying models with task masters, allowing the ensemble to perform joint aggregation across tasks or maintaining separate ensembles for each task. Examples include joint modeling for fake reviews detection and helpfulness prediction [74].

For sequence-to-sequence tasks, such as neural machine translation or text summarization, ensembles often include multiple encoders [75]. These ensembles use strategies such as majority voting where each model generates a prediction, and the final output is chosen based on the most common sequence or token across models [5].

In generative modeling, ensembles can be formed by sampling multiple generative models or by averaging their latent representations [76]. This technique is particularly useful for complex generative tasks, such as combining multiple diffusion models [77] or generative adversarial networks (GANs) [78].

To conclude, Figure 2. present the mapping of task types and ensemble strategies.

Figure. 2 Mapping Ensemble strategies to task types

2.3. Based on Fusion Stage

Fusion refers to the stage at which outputs of base models are combined [5][79]. This affects both the interpretability and complexity of the ensemble system.

Early Fusion involves combining features before the model training phase [80]. This typically means concatenating the output from different feature extractors into a single, unified representation that the model then learns from. Early fusion is commonly used in scenarios like multi-modal classification [81]and sensor fusion, where data from various sources needs to be integrated from the start to improve model performance.

In contrast, Late Fusion combines outputs or decisions after individual models have completed their inference processes [80]. Instead of merging features, late fusion aggregates predictions, often through methods such as majority voting in ensemble classifiers. This approach allows each model to specialize independently before their outputs are combined to make a final decision [82].

Hybrid Fusion represents a combination of early and late fusion techniques [83]. In this approach, both features and model outputs contribute to the final decision, leveraging the strengths of both strategies. Hybrid fusion is particularly effective in complex systems like those combining natural language processing with computer vision or in task-aware ensemble, where different fusion stages provide complementary information.

Finally, Late fusion is most common due to simplicity and model modularity [84]. Early fusion can lead to richer representations but require architectural alignment. Hybrid fusion enables greater flexibility but is more resource-intensive [85]. Figure 3. Shows a flowchart of ensemble learning pipeline with fusion strategies.

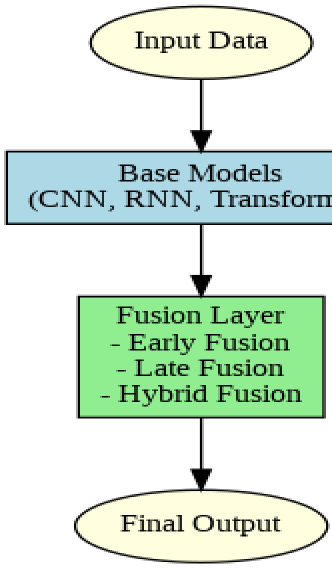

Figure 3. An example of ensemble learning pipeline with fusion strategies.

2.4. Based on Ensemble Strategy

How models are combined statistically, deterministically, or adaptively has a significant impact on ensemble behavior. We categorize the most prominent strategies as follows:

Voting (hard/soft) is a simple ensemble strategy, where predictions are combined either by majority vote (hard voting) or by the average of the class probabilities (soft voting) [5][86][87]. This method is easy to implement and often effective, especially when the reliability of individual models is close to the set. However, it assumes that all models are equally reliable, which may not always be true.

Averaging involves calculating the arithmetic mean or weighted average of the regression outputs or likelihood scores from multiple models [88][89]. This approach helps reduce variability in predictions and can improve overall stability. On the other hand, averaging can be sensitive to outliers [90][91], which can disproportionately affect the combined output.

Stacking refers to training a metamodel that learns how to optimally combine the outputs of the underlying models [19]. This method is characterized by learning an optimal ensemble strategy tailored to the specific data and task, often leading to improved performance [5]. However, stacking requires fine-tuning of the metamodel and can be more complex to implement.

Gating strategies use a learned function to dynamically assign weights to different models based on the input context [92][93]. This context-sensitive ensemble allows the ensemble to adaptively emphasize the most appropriate models for a given input. The main limitation is the need to train an additional gate network, which increases model complexity.

Attention-based ensemble methods apply attention mechanism to dynamically weight model outputs or their internal representations [94]. Attention is widely used in LLMs, where it allows the ensemble to recognize the varying importance of different components. While these methods are powerful, they come at a high computational cost due to their complexity [95].

Finally, neural architecture search methods automate the process of selecting and combining ensemble components [96]. This automation provides scalability and simplifies model selection without manual intervention. However, NAS requires significant computational resources [97] making it less feasible in resource-constrained environments.

2.5. Multi-dimensional Perspective

Most modern ensemble systems combine multiple aspects of the above classification. For example:

- A heterogeneous, multi-task ensemble, using late fusion with an attention-based strategy, may outperform single-task homogeneous ensemble on complex tasks [98].

- A homogeneous NAS-based ensemble, using smooth voting, may offer the best balance between performance and deploy ability in low-latency environments.

To guide researchers in selecting an appropriate configuration, we recommend evaluating task complexity, resource constraints, and the need for interpretability, in line with the proposed classification.

3. Theoretical Foundations

Ensemble methods are fundamentally motivated by the classic bias-variance decomposition [99], where the total prediction error of a model can be attributed to:

- Bias: the error due to erroneous or overly simplistic assumptions in the model [100][101].
- Variance: the error due to sensitivity to fluctuations in the training set [101].
- Irreducible error: noise inherent in the data [102].

In traditional machine learning, ensembles reduce variance by averaging predictions across diverse models [5]. This principle extends to deep learning: while deep neural networks are typically low-bias, high-variance models due to their overparameterization [103], ensembles can reduce variance without significantly increasing bias [104]. For instance, training multiple deep models with different initializations, data partitions, or architectures allows averaging their predictions to stabilize outputs, improving robustness and predictive accuracy.

From a theoretical standpoint, ensemble methods often improve generalization of the model's performance on unseen data by aggregating the hypotheses of multiple learners [105][106]. Generalization bounds for ensemble learning show that combining diverse hypotheses can reduce Rademacher complexity and tighten PAC (Probably Approximately Correct) bounds, especially when individual models are weakly correlated.

In deep learning, however, the theoretical landscape is more complex. Due to high-capacity models and non-convex optimization, classical generalization bounds often fail to tightly characterize performance [107]. Nonetheless, ensembles benefit from error decorrelation: if base models make uncorrelated errors, their aggregated prediction is more likely to correct them [108]. This is particularly effective when model diversity is induced through different architectures, data subsets, or training hyperparameters.

Ensembles tend to work especially well with deep learning for several reasons:

- Deep networks are highly sensitive to initialization and training noise ensembles naturally stabilize this [109].
- Diversity among deep models is easier to induce such as through dropout, data augmentation, or batch stochasticity [13].
- Deep ensembles provide better calibration, producing probabilistic outputs that reflect true confidence levels [110][111].

However, ensembles also pose challenges in deep learning [6]:

- Computational cost: training and storing multiple deep networks is resource-intensive.
- Diminishing returns: if models are too similar such as trained with same seeds and data, ensembles offer little gain.
- Deployment complexity: ensembles can be harder to optimize, compress, or deploy at scale compared to a single model.

In summary, ensemble learning enhances generalization and uncertainty modeling in deep neural networks by reducing variance and stabilizing predictions.

4. State-of-the-Art Techniques

Recent years have seen a surge in ensemble learning innovations, particularly within deep architectures like transformers, diffusion models, and AutoML pipelines. These ensemble methods go beyond traditional averaging or voting by leveraging advanced mechanisms such as attention-based fusion, neural gating, and adaptive weighting, all while addressing scalability, uncertainty, and deployment constraints in modern AI systems [6].

4.1. Ensemble of Transformers:

Transformers have become the backbone of state-of-the-art models in NLP, vision, and multi-modal learning [112]. Researchers now explore layer-wise, checkpoint ensembles, and parameter sharing across transformers to reduce memory [113] while retaining diversity. We introduce this through some domains as follows:

- In NLP, models such as BERT, mT5, and GPT-based ensembles have been used in sentiment analysis, hate speech detection [68], and zero-shot learning. For example, ensemble approaches combining multilingual transformers such as Twitter-roberta-base + Bertweet-base + GPT-3 have shown gains in cross-lingual tasks [114].
- In vision, ViT (Vision Transformer) ensembles improve robustness in image classification [115], while SAM (Segment Anything Model) ensembles help improve spatial coherence in segmentation tasks [116].
- Attention-based ensemble is gaining traction, where attention weights are used to selectively integrate representations from different transformer heads or entire models, improving interpretability and dynamic adaptation to inputs [25].

4.2. Ensemble Diffusion Models:

Diffusion models, which have risen to prominence for generative tasks (e.g., image synthesis, text-to-image, audio generation), are also being ensembled to enhance sample quality and diversity [77].

- Recent work shows that ensembling diffusion trajectories, averaging noise schedules, or sampling from multiple learned priors can reduce artifacts and mode collapse [117].
- Ensembles of text-to-image diffusion models, such as eDiff-I, has been applied for multi-prompt conditioning, improving alignment with user intents [118].

4.3. AutoML and NAS-Driven Ensembles:

The integration of AutoML and NAS with ensemble learning has automated the discovery of optimal ensemble configurations [119].

- From instance, systems like AutoGluon, Google's AutoEnsemble, and NASBench-Ensemble optimize not just base model architecture but also the ensemble topology (e.g., parallel vs. cascaded).
- Techniques such as evolutionary ensembling or multi-objective NAS [120] (balancing accuracy, latency, and diversity) are increasingly used in production pipelines, especially for edge deployment.

4.4. Federated Ensemble Learning:

With the rise of federated learning, ensemble learning has found new relevance in distributed, privacy-preserving settings [121].

- Each local client trains a distinct model, and a central aggregator forms an ensemble instead of averaging parameters[122] (as in FedAvg).
- Federated ensembles preserve heterogeneous model behavior [123], useful for non-IID data distributions common in medical imaging[124].
- One of the recent trends and techniques FedStack [125] (which trains a global meta-model on model outputs across clients).

4.5. Fusion Mechanisms: Attention, Gating, and Learned Weights:

Modern ensembles increasingly rely on learnable fusion layers to combine base model outputs more effectively than static strategies [126].

- Gating modules use an auxiliary neural network to weight base model contributions dynamically based on the input.
- Attention mechanisms allow the model to learn which ensemble members to trust for different parts of the input (e.g., spatial regions in images or tokens in text).
- In multi-modal ensembles (e.g., CLIP, LLaVA, Flamingo), fusion layers align vision and language representations, with adaptive weights depending on task context or modality relevance [127].

In summary, the recent landscape of ensemble learning reflects a significant shift toward task-specific, adaptive, and learned ensemble structures. Whether through federated systems, AutoML pipelines, or multi-modal architectures, ensembles are evolving from simple aggregators to intelligent integrators making them central to the next generation of robust, generalizable AI systems.

5. Applications Across Domains

Ensemble learning continues to make a substantial impact across a wide range of real-world domains. Its ability to combine multiple models to improve robustness, generalization, and uncertainty estimation has proven especially valuable in safety-critical, data-scarce, and high-variability environments. Below, we explore ensemble use cases across several domains, highlighting recent innovations and comparative insights.

5.1. Medical Imaging

In medical imaging, ensemble learning has become a de facto standard for improving diagnostic performance [128]. Tasks such as tumor segmentation, disease classification (e.g., diabetic retinopathy, COVID-19 detection), and organ localization benefit from ensemble models, which mitigate the risk of false negatives in critical scenarios [129].

- Case Study: In brain tumor segmentation (e.g., BraTS dataset), ensembles of 3D U-Net and Transformer-based models consistently outperform single models by improving dice scores and reducing prediction variance across patient populations [130][131].
- Recent Innovations: Attention-guided fusion in ensemble layers has improved lesion detection under ambiguous conditions [132]. Techniques such as Test-Time Ensembling (TTA) and Bayesian Deep Ensembles have also been used for uncertainty-aware diagnoses [133][134].

5.2. Hate Speech and Toxic Language Detection (Arabic and Low-Resource Languages)

Hate speech detection, especially in Arabic dialects and low-resource languages, poses unique challenges due to data scarcity, dialectal variation, and semantic ambiguity [135]. Ensemble learning offers a robust solution by combining different language models, dialect embeddings, and tokenization strategies [68].

- Case Study: For Arabic hate speech, ensembles of AraBERT and CAMeLBERT have shown significant improvements in F1 scores compared to individual models [136].
- Underexplored Potential: In dialect-rich regions like North Africa or the Levant, bootstrapped and cross-dialectal ensembles can leverage labeled data from one dialect to enhance detection in another.
- Low-Resource Extension: Ensemble strategies could be competitive in improving the accuracy of few-shot relation extraction and mitigating high variance risks [137].

5.3. Finance and Fraud Detection

The finance sector has adopted ensemble learning extensively for fraud detection, credit scoring, and stock price forecasting [138][139]. These applications benefit from ensembles' ability to capture rare and adversarial patterns that single models may overlook.

- Case Study: In credit card fraud detection, ensembles of gradient-boosted trees, deep autoencoders, and recurrent neural networks outperform any individual technique, especially when trained on highly imbalanced datasets [140].
- Emerging Trends: Federated ensembles are being adopted to preserve data privacy while combining fraud detection models trained across banks [141].

5.4. Climate and Remote Sensing

Ensemble models are useful for any future development of climate reanalysis, land surface models, and remote sensing retrievals [142]. These models process satellite data, temperature records, and emissions predictions such as [143][144][145].

5.5. Autonomous Systems

Autonomous vehicles, drones, and robotics rely heavily on real-time perception and decision-making [146]. Ensemble learning enhances safety by increasing robustness to edge cases and unseen environments [147].

- Case Study: In autonomous driving, ensembles of object detectors (YOLO, Faster R-CNN) and semantic segmentation models (e.g., DeepLab + SegFormer) improve situational awareness under adverse conditions [148][149] (e.g., night, fog, or urban clutter).
- Innovation: Attention-based ensemble fusion modules are being tested to dynamically weigh sensors (e.g., camera, LiDAR) based on environmental noise or occlusion. Multi-agent ensembles, where each agent contributes a policy component, are also being explored in drone swarms.

5.6. Cross-Domain Observations

- Uncertainty Matters: Domains like medicine, finance, and autonomy benefit most when ensembles are used not just for accuracy, but for uncertainty estimation and calibrated decision-making [150].
- Low-Resource Potential: Hate speech and remote sensing are examples of domains where data scarcity or noise make ensembles particularly advantageous.
- Compute vs. Performance Trade-off: While ensembles are compute-intensive [151], recent methods (e.g., knowledge distillation, snapshot ensembles) are bridging this gap, making ensembles more viable for real-time or edge applications.

6. Performance Analysis & Benchmarking

Ensemble learning, while widely appreciated for boosting predictive performance, must also be scrutinized through the lens of empirical benchmarking and practical trade-offs. This section presents a meta-analysis of ensemble versus non-ensemble approaches across benchmark datasets, discusses key trade-offs (e.g., accuracy vs. complexity), and evaluates ensemble models in terms of robustness, fairness, interpretability, and computational efficiency.

6.1. Accuracy vs. Complexity Trade-offs

While ensemble methods typically enhance accuracy, they introduce significant computational and deployment complexity. Table 1 provides a comparison between single and ensemble models.

Table 1. A comparison between single and ensemble models

Factor	Single Model	Ensemble Models
Training Time	Faster	Slower (due to multiple models or training phases)
Inference Latency	Low	Higher (especially for soft voting, stacking)
Model Size	Compact	Larger memory footprint
Maintenance	Simple	Requires versioning & orchestration

To mitigate these drawbacks, techniques like snapshot ensembling, knowledge distillation, or NAS-guided ensemble compression are increasingly adopted.

6.2. Robustness, Fairness, and Interpretability

6.2.1. Robustness:

Ensemble models, especially those involving diverse architectures or input perturbations, are more resilient to adversarial inputs and noisy data [152].

For example, in Arabic hate speech detection, ensembling models trained on different dialects improved classification under code-switching and sarcasm.

6.2.2. Fairness:

When the underlying models carry different biases, groups can mitigate the effects of individual bias [153]. However, if all base models are biased in the same way (e.g., trained on unbalanced datasets), ensembles may still propagate unfairness. Recent work focuses on bias-aware ensembling, using fairness constraints in meta-learners or attention fusion weights.

6.2.3. Interpretability:

Interpretability often decreases with ensemble complexity [154]. While individual models may be explainable, an ensemble of deep models becomes opaque. Mitigations include:

- Using SHAP or LIME on ensemble outputs.
- Visualizing attention fusion weights in Transformer ensembles.
- Applying feature importance aggregation across members.

6.3. Efficiency Optimizations

To address ensemble inefficiency, the following strategies are gaining traction:

- Snapshot Ensembles: Saving multiple checkpoints from a single training run with cyclic learning rates.
- Knowledge Distillation: Compressing the ensemble into a single "student" model without sacrificing performance.
- Federated Ensembles: Combining model outputs from distributed nodes without centralizing data.

To conclude, ensemble learning remains a powerful paradigm for improving performance across a wide array of tasks and datasets. However, the trade-offs in complexity, resource demand, and interpretability must be carefully balanced, especially in low-resource and real-time scenarios. As the field moves forward, future work must strive to make ensembles leaner, fairer, and more interpretable without compromising on the robustness that makes them valuable in the first place. Figure 4 shows robustness, fairness and accuracy,…etc.

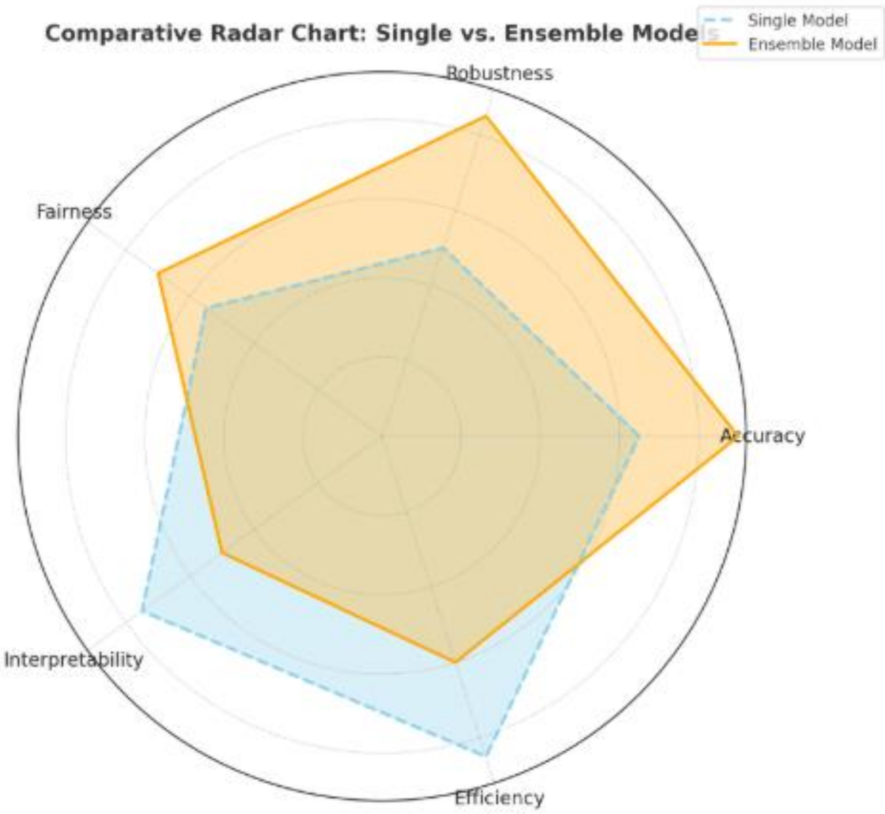

Figure 4. Capabilities of ensemble models vs single models.

7. Challenges and Limitations

Despite the widespread success of ensemble learning across domains, it faces several critical challenges that hinder its broader deployment especially in real-time, resource-constrained, or regulated environments. This section outlines the key limitations that need to be addressed to make ensemble methods more practical and accessible.

7.1. Computational Cost and Latency

One of the most significant drawbacks of ensemble models is their increased computational demand. Training multiple base learners (especially large deep neural networks) requires substantial resources in terms of GPU memory, compute time, and storage. This challenge is amplified during inference, where the ensemble must aggregate outputs from several models leading to higher latency, which is unsuitable for time-sensitive applications like autonomous driving or fraud detection.

7.2. Interpretability and Explainability

As ensemble architectures grow more complex especially those combining heterogeneous models they become increasingly opaque to users and auditors. While individual models may be interpretable, the composite behavior of the ensemble often defies straightforward analysis.

This is a pressing issue in regulated sectors like healthcare and finance, where explainability is essential for decision accountability. Traditional tools like SHAP or LIME must be adapted to aggregate or average explanations across ensemble members, but this may still obscure the individual decision pathways.

7.3. Data and Model Diversity Requirements

The success of ensemble learning heavily depends on diversity among base models, both in architecture and in the data they are trained on. However:

- Access to diverse labeled datasets is limited in many domains (e.g., Arabic dialect NLP, rare diseases).
- Training models with different hyperparameters or architectures significantly increases resource demands.
- Techniques like data augmentation and bootstrapping introduce artificial diversity, which may not always reflect real-world variability.

Insufficient diversity can cause the ensemble to behave like a redundant average, reducing its potential performance gain.

7.4. Overfitting in Ensemble Learners

Although ensembles generally reduce variance, they are not immune to overfitting especially when:

- The base learners are all high-capacity models (e.g., deep CNNs or Transformers).
- The ensemble over-learns patterns specific to the training data (e.g., stacked ensembles with powerful meta-learners).
- There is data leakage during cross-validation used for meta-model training.

Careful cross-validation, early stopping, and regularization are necessary to mitigate this risk.

7.5. Real-Time Deployment Issues

Deploying ensemble models in real-time applications poses several engineering and logistical challenges:

- Synchronizing predictions across distributed systems or devices.
- Efficiently loading and maintaining multiple large models in memory.
- Performing ensemble-level decision fusion without compromising response time.

These issues are particularly acute in edge computing, mobile apps, and IoT systems, where computational budgets are tight.

7.6. Model Compression for Ensembles

To bridge the gap between performance and deployability, researchers are increasingly exploring:

- Knowledge distillation: Compressing the ensemble's collective knowledge into a single "student" model.
- Snapshot ensembling: Using checkpoints from a single training run to simulate an ensemble.

- Weight sharing and pruning: Reducing parameter duplication across ensemble members.
- Transformer-based routing/gating: Activating only relevant ensemble paths during inference.

While promising, these techniques often involve complex training procedures and may still trade off some robustness or diversity. To summarize, we provide table 2 to summarize the above challenges.

Table 2. Ensemble models challenges and mitigation strategies.

Challenge	Impact	Mitigation Strategies
Computational Cost	Slower inference, resource bottlenecks	Model distillation, pruning, quantization
Interpretability	Opaque decisions, poor explainability	SHAP/LIME aggregation, attention visualization
Data/Model Diversity	Reduced generalization	Bootstrapping, cross-domain training, hybrid architectures
Overfitting	Poor generalization	Regularization, bagging, early stopping
Real-Time Deployment	Latency, synchronization issues	Lightweight ensembling, snapshot ensembles
Model Compression	Loss of performance or robustness	NAS, knowledge distillation, routing modules

8. Future Research Directions

As the field of deep learning continues to evolve, ensemble methods must also adapt to meet new challenges and opportunities. In this section, we outline several promising directions for future research in ensemble deep learning, particularly emphasizing underexplored paradigms that intersect with recent advances in large models, low-resource learning, and generalization across tasks and domains.

8.1. Prompt Ensembling in Large Language Models (LLMs)

With the rise of prompt-based learning in large language models such as GPT, LLaMA, and T5, a novel paradigm emerges prompt ensembling. Instead of ensembling models, this approach ensembles multiple prompt formulations to reduce variance in outputs, improve robustness, and boost generalization across tasks. Future research can explore optimal strategies for selecting and weighting prompts and combining them dynamically based on input characteristics or confidence metrics.

8.2. Self-Ensembling and Test-Time Augmentation

Self-ensembling leverages the predictions of a model across multiple augmented versions of an input at test time, treating them as pseudo-ensemble members. Though already popular in semi-supervised learning, this technique remains underutilized in broader ensemble contexts. Research could investigate more sophisticated augmentation pipelines, dynamic weighting mechanisms, and applications in noisy or domain-shifted data settings. Combining self-ensembling with transformer-based architectures or diffusion models is another promising avenue.

8.3. Cross-Lingual and Cross-Domain Ensembling

The application of ensemble learning to cross-lingual or cross-domain problems is still in its infancy. For example, combining models trained on different languages or domains can mitigate the scarcity of labeled data and improve robustness to domain shifts. In NLP, ensembles of multilingual models like mBERT, XLM-R, or mT5 can enhance performance on low-resource or zero-shot tasks. For computer vision, domain-specific ViTs or diffusion models can be ensembled to tackle variation across datasets (e.g., satellite vs. aerial imagery). Key research questions include how to align representations across models and how to ensure fair and balanced outputs across diverse inputs.

8.4. Semi-Supervised and Self-Supervised Ensemble Frameworks

There is growing interest in applying ensemble methods to semi-supervised and self-supervised learning paradigms. Ensembles can benefit from pseudo-labeling consistency, representation diversity, and improved convergence. Future directions include:

- Ensembles of contrastive learners or masked autoencoders.

- Bootstrapped ensembling with confidence-aware pseudo-label filtering.
- Joint optimization of representation diversity and task performance.

These approaches can particularly benefit applications in healthcare, speech recognition, and other domains with limited labeled data.

8.5. Ensemble Learning in Continual and Online Settings

Traditional ensemble methods assume static training data, which limits their applicability in continual and online learning settings. Emerging techniques should focus on:

- Incremental ensembling strategies that adapt as data evolves.
- Memory-efficient approaches to managing model pools in streaming environments.
- Catastrophic forgetting mitigation via ensemble diversity. Such methods can empower real-time applications like adaptive recommendation systems or autonomous driving under changing environmental conditions.

8.6. Low-Resource and Zero-Shot Ensemble Frameworks

As AI systems expand to underserved languages and domains, low-resource and zero-shot ensemble methods gain relevance. Possible directions include:

- Leveraging pre-trained models with lightweight adapters in ensemble setups.
- Ensembling distilled models to reduce resource demands.
- Multilingual or multi-domain ensembling for zero-shot generalization.

These approaches are crucial for democratizing AI, especially in humanitarian applications such as disaster response or multilingual content moderation.

9. Conclusion

Ensemble learning continues to play a crucial role in improving the performance and robustness of deep learning models across various domains. This review presented a new taxonomy, theoretical insights, recent advancements, and diverse applications, highlighting the strengths and challenges of ensemble approaches. Key future directions include prompt ensembling in large language models, cross-domain and cross-lingual ensembles, and semi/self-supervised ensemble frameworks. Researchers should focus on adapting ensembles to emerging architectures and dynamic settings, while practitioners can benefit from lightweight and domain-specific ensemble strategies. Overall, ensemble deep learning is an essential tool for building reliable and adaptable AI systems, and ongoing innovation will further unlock its potential in practical applications.

References

[1] Y. Hu *et al.*, “Artificial Intelligence Approaches,” Aug. 2019, doi: 10.22224/gistbok/2019.3.4.

[2] I. H. Sarker, “Machine Learning: Algorithms, Real-World Applications and Research Directions,” *SN Comput. Sci.*, vol. 2, no. 3, p. 160, May 2021, doi: 10.1007/s42979-021-00592-x.

[3] I. H. Sarker, “Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions,” *SN Comput. Sci.*, vol. 2, no. 6, p. 420, Nov. 2021, doi: 10.1007/s42979-021-00815-1.

[4] S. F. Ahmed *et al.*, “Deep learning modelling techniques: current progress, applications, advantages, and challenges,” *Artif. Intell. Rev.*, vol. 56, no. 11, pp. 13521–13617, Nov. 2023, doi: 10.1007/s10462-023-10466-8.

[5] A. Mohammed and R. Kora, “A comprehensive review on ensemble deep learning: Opportunities and challenges,” *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 2, pp. 757–774, Feb. 2023, doi: 10.1016/j.jksuci.2023.01.014.

[6] N. Rane, S. P. Choudhary, and J. Rane, “Ensemble deep learning and machine learning: applications, opportunities, challenges, and future directions,” *Studies in Medical and Health Sciences*, vol. 1, no. 2, pp. 18–41, Jul. 2024, doi: 10.48185/smhs.v1i2.1225.

[7] Y. Cao, T. A. Geddes, J. Y. H. Yang, and P. Yang, “Ensemble deep learning in bioinformatics,” *Nat. Mach. Intell.*, vol. 2, no. 9, pp. 500–508, Aug. 2020, doi: 10.1038/s42256-020-0217-y.

[8] L. K. Hansen and P. Salamon, “Neural network ensembles,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 10, pp. 993–1001, 1990, doi: 10.1109/34.58871.

[9] Y. Yang, H. Lv, and N. Chen, “A Survey on ensemble learning under the era of deep learning,” *Artif. Intell. Rev.*, vol. 56, no. 6, pp. 5545–5589, Jun. 2023, doi: 10.1007/s10462-022-10283-5.

[10] O. Sagi and L. Rokach, “Ensemble learning: A survey,” *WIREs Data Mining and Knowledge Discovery*, vol. 8, no. 4, Jul. 2018, doi: 10.1002/widm.1249.

[11] J. Jia, W. Liang, and Y. Liang, “A Review of Hybrid and Ensemble in Deep Learning for Natural Language Processing,” Aug. 2024, [Online]. Available: http://arxiv.org/abs/2312.05589

[12] P. Mahajan, S. Uddin, F. Hajati, and M. A. Moni, “Ensemble Learning for Disease Prediction: A Review,” *Healthcare*, vol. 11, no. 12, p. 1808, Jun. 2023, doi: 10.3390/healthcare11121808.

[13] M. A. Ganaie, M. Hu, A. K. Malik, M. Tanveer, and P. N. Suganthan, “Ensemble deep learning: A review,” *Eng. Appl. Artif. Intell.*, vol. 115, p. 105151, Oct. 2022, doi: 10.1016/j.engappai.2022.105151.

[14] Venugopal. Y. R and Dr. S. V, “Ensemble Learning Approaches for Improved Predictive Analytics in Healthcare,” *International Journal of Research Publication and Reviews*, vol. 5, no. 3, pp. 757–760, Mar. 2024, doi: 10.55248/gengpi.5.0324.0629.

[15] I. Marçal and R. E. Garcia, “A comprehensible analysis of the efficacy of Ensemble Models for Bug Prediction,” Oct. 2023, [Online]. Available: http://arxiv.org/abs/2310.12133

[16] A. A. Khan, O. Chaudhari, and R. Chandra, “A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation,” *Expert Syst. Appl.*, vol. 244, p. 122778, Jun. 2024, doi: 10.1016/j.eswa.2023.122778.

[17] A. Goyal and N. Sardana, “Empirical Analysis of Ensemble Machine Learning Techniques for Bug Triaging,” in *2019 Twelfth International Conference on Contemporary Computing (IC3)*, IEEE, Aug. 2019, pp. 1–6. doi: 10.1109/IC3.2019.8844876.

[18] D. Opitz and R. Maclin, “Popular Ensemble Methods: An Empirical Study,” *Journal of Artificial Intelligence Research*, vol. 11, pp. 169–198, Aug. 1999, doi: 10.1613/jair.614.

[19] I. D. Mienye and Y. Sun, “A Survey of Ensemble Learning : Concepts , Algorithms , Applications , and Prospects,” *IEEE Access*, vol. 10, no. September, pp. 99129–99149, 2022, doi: 10.1109/ACCESS.2022.3207287.

[20] S. Ali, T. Sreenivas Sremath, and A. Sarrafzadeh, “Ensemble learning methods for decision making: Status and future prospects,” *International Conference on Machine Learning and Cybernetics (ICMLC), Guangzhou, China*, pp. 211–216, 2015, doi: 10.1109/ICMLC.2015.7340924.

[21] T. G. Dietterich, “Ensemble Methods in Machine Learning,” *Multiple Classifier Systems*, vol. 1857, 2000, doi: https://doi.org/10.1007/3-540-45014-9_1.

[22] T. Dinh *et al.*, “Data clustering: an essential technique in data science,” 2024, doi: https://doi.org/10.48550/arXiv.2412.18760.

[23] A. Ghorbanian, “A new method based on ensemble time series for fast and accurate clustering,” *Data Technologies and Applications*, vol. 57, no. 5, pp. 756–779, 2023, doi: 10.1108/DTA-08-2022-0300.

[24] A. A. Wani, “Comprehensive analysis of clustering algorithms : exploring limitations and innovative solutions,” *PeerJ Comput. Sci.*, vol. 10, p. e2286, 2024, doi: 10.7717/peerj-cs.2286.

[25] I. Al, B. H. Hammo, and O. Al-kadi, “An ensemble model with attention based mechanism for image captioning,” *Computers and Electrical Engineering*, vol. 123, no. PA, p. 110077, 2025, doi: 10.1016/j.compeleceng.2025.110077.

[26] S. Wang, H. Jiao, X. Su, and Q. Yuan, “An Ensemble Learning Approach With Attention Mechanism for Detecting Pavement Distress and Disaster-Induced Road Damage,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 10, pp. 13667–13681, Oct. 2024, doi: 10.1109/TITS.2024.3391751.

[27] X. Geng, X. He, L. Xu, and J. Yu, “Attention-based gating optimization network for multivariate time series prediction,” *Appl. Soft Comput.*, vol. 126, p. 109275, Sep. 2022, doi: 10.1016/j.asoc.2022.109275.

[28] K. Chen, L. Yang, Y. Chen, K. Chen, Y. Xu, and L. Li, “GP-NAS-ensemble: a model for NAS Performance Prediction,” Jan. 2023.

[29] J. Artin, A. Valizadeh, M. Ahmadi, S. A. P. Kumar, and A. Sharifi, “Presentation of a Novel Method for Prediction of Traffic with Climate Condition Based on Ensemble Learning of Neural Architecture Search (NAS) and Linear Regression,” *Complexity*, vol. 2021, no. 1, Jan. 2021, doi: 10.1155/2021/8500572.

[30] M. R. Al-Sinan, A. F. Haneef, and H. Luqman, "Ensemble Learning using Transformers and Convolutional Networks for Masked Face Recognition," Oct. 2022.
[31] C. Wang *et al.*, "Ensembling Diffusion Models via Adaptive Feature Aggregation," Feb. 2025.
[32] E. Lella, A. Pazienza, D. Lofù, R. Anglani, and F. Vitulano, "An Ensemble Learning Approach Based on Diffusion Tensor Imaging Measures for Alzheimer's Disease Classification," *Electronics (Basel).*, vol. 10, no. 3, p. 249, Jan. 2021, doi: 10.3390/electronics10030249.
[33] H. Manjunatha and P. Tsiotras, "Beyond One Model Fits All: Ensemble Deep Learning for Autonomous Vehicles," Dec. 2023.
[34] A.-A. Preda, D.-C. Cercel, T. Rebedea, and C.-G. Chiru, "UPB at IberLEF-2023 AuTexTification: Detection of Machine-Generated Text using Transformer Ensembles," Aug. 2023.
[35] H. Yang, M. Li, H. Zhou, Y. Xiao, Q. Fang, and R. Zhang, "One LLM is not Enough: Harnessing the Power of Ensemble Learning for Medical Question Answering," Dec. 24, 2023. doi: 10.1101/2023.12.21.23300380.
[36] A. Mabrouk, R. P. Díaz Redondo, M. Abd Elaziz, and M. Kayed, "Ensemble Federated Learning: An approach for collaborative pneumonia diagnosis," *Appl. Soft Comput.*, vol. 144, p. 110500, Sep. 2023, doi: 10.1016/j.asoc.2023.110500.
[37] V. Hegiste, T. Legler, and M. Ruskowski, "Federated Ensemble YOLOv5 – A Better Generalized Object Detection Algorithm," in *2023 Eighth International Conference on Fog and Mobile Edge Computing (FMEC)*, IEEE, Sep. 2023, pp. 7–14. doi: 10.1109/FMEC59375.2023.10305958.
[38] F. E. Casado, D. Lema, R. Iglesias, C. V. Regueiro, and S. Barro, "Ensemble and continual federated learning for classification tasks," *Mach. Learn.*, vol. 112, no. 9, pp. 3413–3453, Sep. 2023, doi: 10.1007/s10994-023-06330-z.
[39] M. Đumić and D. Jakobović, "Ensembles of priority rules for resource constrained project scheduling problem," *Appl. Soft Comput.*, vol. 110, p. 107606, Oct. 2021, doi: 10.1016/j.asoc.2021.107606.
[40] F. S. Lay, A. Dömel, N. Y. Lii, and F. Stulp, "Orchestrating Method Ensembles to Adapt to Resource Requirements and Constraints During Robotic Task Execution," *IEEE Robot. Autom. Lett.*, vol. 10, no. 2, pp. 1186–1193, Feb. 2025, doi: 10.1109/LRA.2024.3518077.
[41] S. Osei, A. R. Masegosa, and A. D. Masegosa, "Understanding the Role of Diversity in Ensemble-Based AutoML Methods for Classification Tasks," *IEEE Access*, vol. 13, pp. 63566–63586, 2025, doi: 10.1109/ACCESS.2025.3554093.
[42] S. Mao, J.-W. Chen, L. Jiao, S. Gou, and R. Wang, "Maximizing diversity by transformed ensemble learning," *Appl. Soft Comput.*, vol. 82, p. 105580, Sep. 2019, doi: 10.1016/j.asoc.2019.105580.
[43] A. M. Durán-Rosal, T. Ashley, J. Pérez-Rodríguez, and F. Fernández-Navarro, "Global and Diverse Ensemble model for regression," *Neurocomputing*, vol. 647, p. 130520, Sep. 2025, doi: 10.1016/j.neucom.2025.130520.
[44] T. Anande, S. Alsaadi, and M. Leeson, "Enhanced Modelling Performance with Boosting Ensemble Meta-Learning and Optuna Optimization," *SN Comput. Sci.*, vol. 6, no. 1, p. 12, Dec. 2024, doi: 10.1007/s42979-024-03544-3.
[45] M. Sabzevari, G. Martínez-Muñoz, and A. Suárez, "Building heterogeneous ensembles by pooling homogeneous ensembles," *International Journal of Machine Learning and Cybernetics*, vol. 13, no. 2, pp. 551–558, Feb. 2022, doi: 10.1007/s13042-021-01442-1.
[46] J. Wilson, S. Chaudhury, and B. Lall, "Homogeneous–Heterogeneous Hybrid Ensemble for concept-drift adaptation," *Neurocomputing*, vol. 557, p. 126741, Nov. 2023, doi: 10.1016/j.neucom.2023.126741.
[47] W. Li *et al.*, "Machine Learning Model of ResNet50-Ensemble Voting for Malignant–Benign Small Pulmonary Nodule Classification on Computed Tomography Images," *Cancers (Basel).*, vol. 15, no. 22, p. 5417, Nov. 2023, doi: 10.3390/cancers15225417.
[48] Q. Zhang, Q. Yuan, J. Li, Z. Yang, and X. Ma, "Learning a Dilated Residual Network for SAR Image Despeckling," Jan. 2018.
[49] Z. Lu, B. Xu, L. Sun, T. Zhan, and S. Tang, "3-D Channel and Spatial Attention Based Multiscale Spatial–Spectral Residual Network for Hyperspectral Image Classification," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 13, pp. 4311–4324, 2020, doi: 10.1109/JSTARS.2020.3011992.
[50] M. Sharkas, "Ear recognition with ensemble classifiers; A deep learning approach," *Multimed. Tools Appl.*, vol. 81, no. 30, pp. 43919–43945, Dec. 2022, doi: 10.1007/s11042-022-13252-w.

[51] S. Bian and W. Wang, “On diversity and accuracy of homogeneous and heterogeneous ensembles,” *Int. J. Hybrid Intell. Syst.*, vol. 4, no. 2, pp. 103–128, Jun. 2007, doi: 10.3233/HIS-2007-4204.
[52] Z. Fang, Y. Wang, L. Peng, and H. Hong, “A comparative study of heterogeneous ensemble-learning techniques for landslide susceptibility mapping,” *International Journal of Geographical Information Science*, vol. 35, no. 2, pp. 321–347, Feb. 2021, doi: 10.1080/13658816.2020.1808897.
[53] K. Jiang, Z. Xiong, Q. Yang, J. Chen, and G. Chen, “An interpretable ensemble method for deep representation learning,” *Engineering Reports*, vol. 6, no. 3, Mar. 2024, doi: 10.1002/eng2.12725.
[54] H. Saleh, S. Mostafa, A. Alharbi, S. El-Sappagh, and T. Alkhalifah, “Heterogeneous Ensemble Deep Learning Model for Enhanced Arabic Sentiment Analysis,” *Sensors*, vol. 22, no. 10, p. 3707, May 2022, doi: 10.3390/s22103707.
[55] Y. Xia, J. Zhao, L. He, Y. Li, and M. Niu, “A novel tree-based dynamic heterogeneous ensemble method for credit scoring,” *Expert Syst. Appl.*, vol. 159, p. 113615, Nov. 2020, doi: 10.1016/j.eswa.2020.113615.
[56] M. S. Abassi, “Diversity of Ensembles for Data Stream Classification,” Feb. 2019.
[57] B. Pes, “Ensemble feature selection for high-dimensional data: a stability analysis across multiple domains,” *Neural Comput. Appl.*, vol. 32, no. 10, pp. 5951–5973, May 2020, doi: 10.1007/s00521-019-04082-3.
[58] M.-R. Amini, V. Feofanov, L. Pauletto, L. Hadjadj, É. Devijver, and Y. Maximov, “Self-training: A survey,” *Neurocomputing*, vol. 616, p. 128904, Feb. 2025, doi: 10.1016/j.neucom.2024.128904.
[59] M. E.Almandouh, M. F. Alrahmawy, M. Eisa, M. Elhoseny, and A. S. Tolba, “Ensemble based high performance deep learning models for fake news detection,” *Sci. Rep.*, vol. 14, no. 1, p. 26591, Nov. 2024, doi: 10.1038/s41598-024-76286-0.
[60] A. Melleng, A.-J. Loughrey, and D. P, “Multi-task Ensemble Learning for Fake Reviews Detection and Helpfulness Prediction: A Novel Approach,” in *Proceedings of the Conference Recent Advances in Natural Language Processing - Large Language Models for Natural Language Processings*, INCOMA Ltd., Shoumen, BULGARIA, 2023, pp. 721–729. doi: 10.26615/978-954-452-092-2_078.
[61] R. Caruana, “Multitask Learning,” *Mach. Learn.*, vol. 28, no. 1, pp. 41–75, Jul. 1997, doi: 10.1023/A:1007379606734.
[62] K.-H. Thung and C.-Y. Wee, “A brief review on multi-task learning,” *Multimed. Tools Appl.*, vol. 77, no. 22, pp. 29705–29725, Nov. 2018, doi: 10.1007/s11042-018-6463-x.
[63] M. Oner and A. Ustundag, “Combining predictive base models using deep ensemble learning,” *Journal of Intelligent & Fuzzy Systems*, vol. 39, no. 5, pp. 6657–6668, Nov. 2020, doi: 10.3233/JIFS-189126.
[64] A. A. Khan, O. Chaudhari, and R. Chandra, “A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation,” *Expert Syst. Appl.*, vol. 244, p. 122778, Jun. 2023, doi: 10.1016/j.eswa.2023.122778.
[65] M. Harandi, J. Taheri, and B. C. Lovell, “Ensemble Learning for Object Recognition and Tracking,” in *Pattern Recognition, Machine Intelligence and Biometrics*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 261–278. doi: 10.1007/978-3-642-22407-2_11.
[66] X. Zhou and Y. He, “Deep Ensemble Learning-Based Sensor for Flotation Froth Image Recognition,” *Sensors*, vol. 24, no. 15, p. 5048, Aug. 2024, doi: 10.3390/s24155048.
[67] P. Garg, M. K. Sharma, and P. Kumar, “Improving Hate Speech Classification Through Ensemble Learning and Explainable AI Techniques,” *Arab. J. Sci. Eng.*, vol. 50, no. 15, pp. 11631–11644, Aug. 2024, doi: 10.1007/s13369-024-09540-2.
[68] M. M. Abdelsamie, S. S. Azab, and H. A. Hefny, “A comprehensive review on Arabic offensive language and hate speech detection on social media: methods, challenges and solutions,” *Soc. Netw. Anal. Min.*, vol. 14, no. 1, p. 111, May 2024, doi: 10.1007/s13278-024-01258-1.
[69] D. M. Jose, A. M. Vincent, and G. S. Dwarakish, “Improving multiple model ensemble predictions of daily precipitation and temperature through machine learning techniques,” *Sci. Rep.*, vol. 12, no. 1, p. 4678, Mar. 2022, doi: 10.1038/s41598-022-08786-w.

[70] S. V. Boyapati, M. S. Karthik, K. Subrahmanyam, and B. R. Reddy, “An Analysis of House Price Prediction Using Ensemble Learning Algorithms,” *Research Reports on Computer Science*, pp. 87–96, May 2023, doi: 10.37256/rrcs.2320232639.

[71] Z. Yang, X. Zhu, Y. Zhang, P. Nie, and X. Liu, “A Housing Price Prediction Method Based on Stacking Ensemble Learning Optimization Method,” in *2023 IEEE 10th International Conference on Cyber Security and Cloud Computing (CSCloud)/2023 IEEE 9th International Conference on Edge Computing and Scalable Cloud (EdgeCom)*, IEEE, Jul. 2023, pp. 96–101. doi: 10.1109/CSCloud-EdgeCom58631.2023.00025.

[72] H. Shaiba *et al.*, “Weather Forecasting Prediction Using Ensemble Machine Learning for Big Data Applications,” *Computers, Materials & Continua*, vol. 73, no. 2, pp. 3367–3382, 2022, doi: 10.32604/cmc.2022.030067.

[73] E. Pei, Z. Hu, L. He, H. Ning, and A. D. Berenguer, “An ensemble learning-enhanced multitask learning method for continuous affect recognition from facial images,” *Expert Syst. Appl.*, vol. 236, p. 121290, Feb. 2024, doi: 10.1016/j.eswa.2023.121290.

[74] A. Melleng, A.-J. Loughrey, and D. P, “Multi-task Ensemble Learning for Fake Reviews Detection and Helpfulness Prediction: A Novel Approach,” in *Proceedings of the Conference Recent Advances in Natural Language Processing - Large Language Models for Natural Language Processings*, INCOMA Ltd., Shoumen, BULGARIA, 2023, pp. 721–729. doi: 10.26615/978-954-452-092-2_078.

[75] X. Sun, D. Wu, A. Zinflou, and B. Boulet, “Anomaly Detection with Ensemble of Encoder and Decoder,” Mar. 2023.

[76] H. Shen, C. Lin, Y. Ma, and E. Xie, “Ensemble Deep Learning Classification Method Based on Generative Adversarial Networks,” in *2024 16th International Conference on Computer and Automation Engineering (ICCAE)*, IEEE, Mar. 2024, pp. 46–53. doi: 10.1109/ICCAE59995.2024.10569429.

[77] Y. Shin, G. Son, D. Hwang, and T. Eo, “Ensemble and low-frequency mixing with diffusion models for accelerated MRI reconstruction,” *Med. Image Anal.*, vol. 101, p. 103477, Apr. 2025, doi: 10.1016/j.media.2025.103477.

[78] Y. Wang, L. Zhang, and J. van de Weijer, “Ensembles of Generative Adversarial Networks,” Dec. 2016.

[79] K.-L. Du and M. N. S. Swamy, “Combining Multiple Learners: Data Fusion and Ensemble Learning,” in *Neural Networks and Statistical Learning*, London: Springer London, 2019, pp. 737–767. doi: 10.1007/978-1-4471-7452-3_25.

[80] F. Abdullakutty and U. Naseem, “Decoding Memes: A Comprehensive Analysis of Late and Early Fusion Models for Explainable Meme Analysis,” in *Companion Proceedings of the ACM Web Conference 2024*, New York, NY, USA: ACM, May 2024, pp. 1681–1689. doi: 10.1145/3589335.3652504.

[81] A. Dore, M. Pinasco, and C. S. Regazzoni, “Multi-Modal Data Fusion Techniques and Applications,” in *Multi-Camera Networks*, Elsevier, 2009, pp. 213–237. doi: 10.1016/B978-0-12-374633-7.00011-2.

[82] L. M. Pereira, A. Salazar, and L. Vergara, “A Comparative Analysis of Early and Late Fusion for the Multimodal Two-Class Problem,” *IEEE Access*, vol. 11, pp. 84283–84300, 2023, doi: 10.1109/ACCESS.2023.3296098.

[83] X. Huang, T. Ma, L. Jia, Y. Zhang, H. Rong, and N. Alnabhan, “An effective multimodal representation and fusion method for multimodal intent recognition,” *Neurocomputing*, vol. 548, p. 126373, Sep. 2023, doi: 10.1016/j.neucom.2023.126373.

[84] H. R. V. Joze, A. Shaban, M. L. Iuzzolino, and K. Koishida, “MMTM: Multimodal Transfer Module for CNN Fusion,” Mar. 2020.

[85] S. Kalamkar and G. M. A., “Multimodal image fusion: A systematic review,” *Decision Analytics Journal*, vol. 9, p. 100327, Dec. 2023, doi: 10.1016/j.dajour.2023.100327.

[86] O. O. Awe, G. O. Opateye, C. A. G. Johnson, O. T. Tayo, and R. Dias, “Weighted Hard and Soft Voting Ensemble Machine Learning Classifiers: Application to Anaemia Diagnosis,” 2023, pp. 351–374. doi: 10.1007/978-3-031-41352-0_18.

[87] K. Cao-Van, T. C. Minh, L. G. Minh, T. T. B. Quyen, and H. M. Tan, “Soft-Voting Ensemble Model: An Efficient Learning Approach for Predictive Prostate Cancer Risk,” *Vietnam Journal of Computer Science*, vol. 11, no. 04, pp. 531–552, Nov. 2024, doi: 10.1142/S2196888824500155.

[88] D. Arpit, H. Wang, Y. Zhou, and C. Xiong, “Ensemble of Averages: Improving Model Selection and Boosting Performance in Domain Generalization,” Oct. 2022.
[89] D. A. Drew and S. L. Passman, “Ensemble Averaging,” 1999, pp. 92–104. doi: 10.1007/0-387-22637-0_10.
[90] B. Avci and A. Boduroglu, “Contributions of ensemble perception to outlier representation precision,” *Atten. Percept. Psychophys.*, vol. 83, no. 3, pp. 1141–1151, Apr. 2021, doi: 10.3758/s13414-021-02270-9.
[91] Y. Zhao and M. K. Hryniewicki, “DCSO: Dynamic Combination of Detector Scores for Outlier Ensembles,” Nov. 2019.
[92] E. M. Burger and S. J. Moura, “Gated ensemble learning method for demand-side electricity load forecasting,” *Energy Build.*, vol. 109, pp. 23–34, Dec. 2015, doi: 10.1016/j.enbuild.2015.10.019.
[93] N. Farhan, I. T. Awishi, M. H. K. Mehedi, MD. M. Alam, and A. A. Rasel, “Ensemble of Gated Recurrent Unit and Convolutional Neural Network for Sarcasm Detection in Bangla,” in *2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC)*, IEEE, Mar. 2023, pp. 0624–0629. doi: 10.1109/CCWC57344.2023.10099157.
[94] Z. Song *et al.*, “Attention-based multi-label neural networks for integrated prediction and interpretation of twelve widely occurring RNA modifications,” *Nat. Commun.*, vol. 12, no. 1, p. 4011, Jun. 2021, doi: 10.1038/s41467-021-24313-3.
[95] P. Ratadiya and D. Mishra, “An Attention Ensemble Based Approach for Multilabel Profanity Detection,” in *2019 International Conference on Data Mining Workshops (ICDMW)*, IEEE, Nov. 2019, pp. 544–550. doi: 10.1109/ICDMW.2019.00083.
[96] E. J. Herron, S. R. Young, and T. E. Potok, “Ensembles of Networks Produced from Neural Architecture Search,” 2020, pp. 223–234. doi: 10.1007/978-3-030-59851-8_14.
[97] T. Yamasaki, Z. Wang, T. Luo, N. Chen, and B. Wang, “RBFleX-NAS: Training-Free Neural Architecture Search Using Radial Basis Function Kernel and Hyperparameter Detection,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 36, no. 6, pp. 10057–10071, Jun. 2025, doi: 10.1109/TNNLS.2025.3552693.
[98] K. Sriprateep *et al.*, “Heterogeneous ensemble machine learning to predict the asiaticoside concentration in centella asiatica urban,” *Intelligent Systems with Applications*, vol. 21, p. 200319, Mar. 2024, doi: 10.1016/j.iswa.2023.200319.
[99] G. Valentini and T. G. Dietterich, “Bias—Variance Analysis and Ensembles of SVM,” 2002, pp. 222–231. doi: 10.1007/3-540-45428-4_22.
[100] A. Mikołajczyk-Bareła and M. Grochowski, “A survey on bias in machine learning research,” Aug. 2023.
[101] M. Deprez and E. C. Robinson, “Machine learning basics,” in *Machine Learning for Biomedical Applications*, Elsevier, 2024, pp. 41–65. doi: 10.1016/B978-0-12-822904-0.00007-8.
[102] W. Blanzeisky and P. Cunningham, “Algorithmic Factors Influencing Bias in Machine Learning,” Apr. 2021.
[103] Z.-H. Zhou, “Why over-parameterization of deep neural networks does not overfit?,” *Science China Information Sciences*, vol. 64, no. 1, p. 116101, Jan. 2021, doi: 10.1007/s11432-020-2885-6.
[104] N. Gupta, J. Smith, B. Adlam, and Z. Mariet, “Ensembling over Classifiers: a Bias-Variance Perspective,” Jun. 2022.
[105] R. Baradaran and H. Amirkhani, “Ensemble learning-based approach for improving generalization capability of machine reading comprehension systems,” *Neurocomputing*, vol. 466, pp. 229–242, Nov. 2021, doi: 10.1016/j.neucom.2021.08.095.
[106] P. Gupta, A. Pratap Singh, and V. Kumar, “A Review of Ensemble Methods Used in AI Applications,” 2023, pp. 145–157. doi: 10.1007/978-981-99-5080-5_13.
[107] J. Wang, Y. Qian, F. Li, J. Liang, and Q. Zhang, “Generalization Performance of Pure Accuracy and its Application in Selective Ensemble Learning,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 2, pp. 1798–1816, Feb. 2023, doi: 10.1109/TPAMI.2022.3171436.
[108] J. Kazmaier and J. H. van Vuuren, “The power of ensemble learning in sentiment analysis,” *Expert Syst. Appl.*, vol. 187, p. 115819, Jan. 2022, doi: 10.1016/j.eswa.2021.115819.
[109] K. Kawaguchi, “Deep Learning without Poor Local Minima,” Dec. 2016.
[110] R. Rahaman and A. H. Thiery, “Uncertainty Quantification and Deep Ensembles,” Nov. 2021.

[111] T. Buddenkotte *et al.*, “Calibrating ensembles for scalable uncertainty quantification in deep learning-based medical image segmentation,” *Comput. Biol. Med.*, vol. 163, p. 107096, Sep. 2023, doi: 10.1016/j.compbiomed.2023.107096.

[112] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, “Transformers in Vision: A Survey,” *ACM Comput. Surv.*, vol. 54, no. 10s, pp. 1–41, Jan. 2022, doi: 10.1145/3505244.

[113] Q. Fournier, G. M. Caron, and D. Aloise, “A Practical Survey on Faster and Lighter Transformers,” *ACM Comput. Surv.*, vol. 55, no. 14s, pp. 1–40, Dec. 2023, doi: 10.1145/3586074.

[114] M. S. U. Miah, M. M. Kabir, T. Bin Sarwar, M. Safran, S. Alfarhood, and M. F. Mridha, “A multimodal approach to cross-lingual sentiment analysis with ensemble of transformer and LLM,” *Sci. Rep.*, vol. 14, no. 1, p. 9603, Apr. 2024, doi: 10.1038/s41598-024-60210-7.

[115] S. Bhojanapalli, A. Chakrabarti, D. Glasner, D. Li, T. Unterthiner, and A. Veit, “Understanding Robustness of Transformers for Image Classification,” in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE, Oct. 2021, pp. 10211–10221. doi: 10.1109/ICCV48922.2021.01007.

[116] L. Nanni, “Augmentation and Ensembles: Improving Medical Image Segmentation with SAM and Deep Networks,” 2024.

[117] C. Li, G. Feng, Y. Li, R. Liu, Q. Miao, and L. Chang, “DiffTAD: Denoising diffusion probabilistic models for vehicle trajectory anomaly detection,” *Knowl. Based. Syst.*, vol. 286, p. 111387, Feb. 2024, doi: 10.1016/j.knosys.2024.111387.

[118] Y. Balaji *et al.*, “eDiff-I: Text-to-Image Diffusion Models with an Ensemble of Expert Denoisers,” Mar. 2023.

[119] I. Salehin *et al.*, “AutoML: A systematic review on automated machine learning with neural architecture search,” *Journal of Information and Intelligence*, vol. 2, no. 1, pp. 52–81, Jan. 2024, doi: 10.1016/j.jiixd.2023.10.002.

[120] Y. Ji *et al.*, “An Active Learning based Latency Prediction Approach for Neural Network Architecture,” in *2024 4th International Conference on Neural Networks, Information and Communication (NNICE)*, IEEE, Jan. 2024, pp. 967–971. doi: 10.1109/NNICE61279.2024.10498710.

[121] N. Saber Rashid and H. Maseeh Yasin, “Privacy-preserving machine learning: a review of federated learning techniques and applications,” *International Journal of Scientific World*, vol. 11, no. 1, pp. 30–39, Feb. 2025, doi: 10.14419/af03y111.

[122] B. Fan, S. Jiang, X. Su, S. Tarkoma, and P. Hui, “A Survey on Model-heterogeneous Federated Learning: Problems, Methods, and Prospects,” Nov. 2024.

[123] D. P. Nguyen, S. Yu, J. P. Muñoz, and A. Jannesari, “Enhancing Heterogeneous Federated Learning with Knowledge Extraction and Multi-Model Fusion,” in *Proceedings of the SC ’23 Workshops of the International Conference on High Performance Computing, Network, Storage, and Analysis*, New York, NY, USA: ACM, Nov. 2023, pp. 36–43. doi: 10.1145/3624062.3626325.

[124] F. S. Alhafiz and A. A. Basuhail, “Non-IID Medical Imaging Data on COVID-19 in the Federated Learning Framework: Impact and Directions,” *COVID*, vol. 4, no. 12, pp. 1985–2016, Dec. 2024, doi: 10.3390/covid4120140.

[125] T. Shaik *et al.*, “Clustered FedStack: Intermediate Global Models with Bayesian Information Criterion,” *Pattern Recognit. Lett.*, vol. 177, pp. 121–127, Jan. 2024, doi: 10.1016/j.patrec.2023.12.004.

[126] K.-L. Du, R. Zhang, B. Jiang, J. Zeng, and J. Lu, “Foundations and Innovations in Data Fusion and Ensemble Learning for Effective Consensus,” *Mathematics*, vol. 13, no. 4, p. 587, Feb. 2025, doi: 10.3390/math13040587.

[127] T.-H. Huang, C. Shin, S. J. Tay, D. Adila, and F. Sala, “Multimodal Data Curation via Object Detection and Filter Ensembles,” Jan. 2024.

[128] M. R. Supriyadi *et al.*, “A systematic literature review: exploring the challenges of ensemble model for medical imaging,” *BMC Med. Imaging*, vol. 25, no. 1, p. 128, Apr. 2025, doi: 10.1186/s12880-025-01667-4.

[129] V. K. Prasad *et al.*, “Revolutionizing healthcare: a comparative insight into deep learning’s role in medical imaging,” *Sci. Rep.*, vol. 14, no. 1, p. 30273, Dec. 2024, doi: 10.1038/s41598-024-71358-7.

[130] F. Ullah *et al.*, “Brain MR Image Enhancement for Tumor Segmentation Using 3D U-Net,” *Sensors*, vol. 21, no. 22, p. 7528, Nov. 2021, doi: 10.3390/s21227528.

[131] M. M. Saleh, M. E. Salih, M. A. A. Ahmed, and A. M. Hussein, “From Traditional Methods to 3D U-Net: A Comprehensive Review of Brain Tumour Segmentation Techniques,” *J. Biomed. Sci. Eng.*, vol. 18, no. 01, pp. 1–32, 2025, doi: 10.4236/jbise.2025.181001.

[132] A. A. Munia *et al.*, “Attention-guided hierarchical fusion U-Net for uncertainty-driven medical image segmentation,” *Information Fusion*, vol. 115, p. 102719, Mar. 2025, doi: 10.1016/j.inffus.2024.102719.

[133] Z. Sherkatghanad, M. Abdar, M. Bakhtyari, P. Plawiak, and V. Makarenkov, “BayTTA: Uncertainty-aware medical image classification with optimized test-time augmentation using Bayesian model averaging,” Aug. 2024.

[134] M. M. A. Valiuddin, R. J. G. van Sloun, C. G. A. Viviers, P. H. N. de With, and F. van der Sommen, “A Review of Bayesian Uncertainty Quantification in Deep Probabilistic Image Segmentation,” Dec. 2025.

[135] M. M. Abdelsamie, S. S. Azab, and H. A. Hefny, “The dialects gap: A multi-task learning approach for enhancing hate speech detection in Arabic dialects,” *Expert Syst. Appl.*, vol. 295, p. 128584, Jan. 2026, doi: 10.1016/j.eswa.2025.128584.

[136] I. El Karfi and S. El Fkihi, “An Ensemble of Arabic Transformer-based Models for Arabic Sentiment Analysis,” *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 8, 2022, doi: 10.14569/IJACSA.2022.0130865.

[137] Q. Lin, Y. Liu, W. Wen, Z. Tao, C. Ouyang, and Y. Wan, “Ensemble Making Few-Shot Learning Stronger,” *Data Intell.*, vol. 4, no. 3, pp. 529–551, Jul. 2022, doi: 10.1162/dint_a_00144.

[138] P. Tomar, S. Shrivastava, and U. Thakar, “Ensemble Learning based Credit Card Fraud Detection System,” in *2021 5th Conference on Information and Communication Technology (CICT)*, IEEE, Dec. 2021, pp. 1–5. doi: 10.1109/CICT53865.2020.9672426.

[139] C. L. Perera and S. C. Premaratne, “An Ensemble Machine Learning Approach for Forecasting Credit risk of Loan Applications,” *WSEAS TRANSACTIONS ON SYSTEMS*, vol. 23, pp. 31–46, Dec. 2023, doi: 10.37394/23202.2024.23.4.

[140] N. M. Ibrahim and A. C. Aguguo, “AN ENSEMBLE OF UNSUPERVISED DEEP LEARNING MODELS FOR CREDIT CARD FRAUD DETECTION,” *FUDMA JOURNAL OF SCIENCES*, vol. 9, no. 5, pp. 263–270, May 2025, doi: 10.33003/fjs-2025-0905-3379.

[141] M. Abdul Salam, K. M. Fouad, D. L. Elbably, and S. M. Elsayed, “Federated learning model for credit card fraud detection with data balancing techniques,” *Neural Comput. Appl.*, vol. 36, no. 11, pp. 6231–6256, Apr. 2024, doi: 10.1007/s00521-023-09410-2.

[142] M. Valipour and J. Dietrich, “Developing ensemble mean models of satellite remote sensing, climate reanalysis, and land surface models,” *Theor. Appl. Climatol.*, vol. 150, no. 3–4, pp. 909–926, Nov. 2022, doi: 10.1007/s00704-022-04185-3.

[143] T. Mollick, G. Hashmi, and S. R. Sabuj, “A perceptible stacking ensemble model for air temperature prediction in a tropical climate zone,” *Discover Environment*, vol. 1, no. 1, p. 15, Sep. 2023, doi: 10.1007/s44274-023-00014-0.

[144] Z. Jin, Y. Ma, L. Chu, Y. Liu, R. Dubrow, and K. Chen, “Predicting spatiotemporally-resolved mean air temperature over Sweden from satellite data using an ensemble model,” *Environ. Res.*, vol. 204, p. 111960, Mar. 2022, doi: 10.1016/j.envres.2021.111960.

[145] S. Hu, S. Li, L. Gong, D. Liu, Z. Wang, and G. Xu, “Carbon emissions prediction based on ensemble models: An empirical analysis from China,” *Environmental Modelling & Software*, vol. 188, p. 106437, Apr. 2025, doi: 10.1016/j.envsoft.2025.106437.

[146] T. Miller, I. Durlik, E. Kostecka, P. Borkowski, and A. Łobodzińska, “A Critical AI View on Autonomous Vehicle Navigation: The Growing Danger,” *Electronics (Basel).*, vol. 13, no. 18, p. 3660, Sep. 2024, doi: 10.3390/electronics13183660.

[147] M. Wyatt, B. Radford, N. Callow, M. Bennamoun, and S. Hickey, “Using ensemble methods to improve the robustness of deep learning for image classification in marine environments,” *Methods Ecol. Evol.*, vol. 13, no. 6, pp. 1317–1328, Jun. 2022, doi: 10.1111/2041-210X.13841.

[148] M. Carranza-García, P. Lara-Benítez, J. García-Gutiérrez, and J. C. Riquelme, “Enhancing object detection for autonomous driving by optimizing anchor generation and addressing class imbalance,” *Neurocomputing*, vol. 449, pp. 229–244, Aug. 2021, doi: 10.1016/j.neucom.2021.04.001.

[149] G. Wang, Z. Li, G. Weng, and Y. Chen, "An overview of industrial image segmentation using deep learning models," *Intelligence & Robotics*, vol. 5, no. 1, pp. 143–80, Feb. 2025, doi: 10.20517/ir.2025.09.

[150] T. Abe, E. K. Buchanan, G. Pleiss, R. Zemel, and J. P. Cunningham, "Deep Ensembles Work, But Are They Necessary?," Oct. 2022.

[151] T. N. Palmer, "The economic value of ensemble forecasts as a tool for risk assessment: From days to decades," *Quarterly Journal of the Royal Meteorological Society*, vol. 128, no. 581, pp. 747–774, Apr. 2002, doi: 10.1256/0035900021643593.

[152] L. Alzubaidi *et al.*, "MEFF – A model ensemble feature fusion approach for tackling adversarial attacks in medical imaging," *Intelligent Systems with Applications*, vol. 22, p. 200355, Jun. 2024, doi: 10.1016/j.iswa.2024.200355.

[153] K. Domijan, "What is to be gained by ensemble models in analysis of spectroscopic data?," *Chemometrics and Intelligent Laboratory Systems*, vol. 241, p. 104936, Oct. 2023, doi: 10.1016/j.chemolab.2023.104936.

[154] N. Rane, S. P. Choudhary, and J. Rane, "Ensemble deep learning and machine learning: applications, opportunities, challenges, and future directions," *Studies in Medical and Health Sciences*, vol. 1, no. 2, pp. 18–41, Jul. 2024, doi: 10.48185/smhs.v1i2.1225.