ESP-JETA
ESP Journal of Engineering & Technology Advancements
ISSN: 2583-2646/ Volume 6 Issue 4 October-December 2026
Paper Id: JETA-V6I4P106/ Page No: 57-67

Original Article

Convolutional Neural Networks for Real-World Classification: Lessons from Applied Image Recognition

Rahamath Mohamed Razikh Ulla
Capitol Technology University, Maryland

Received Date: 14 October 2026 Revised Date: 18 October 2026 Accepted Date: 22 October 2026

Abstract: *Image classification using convolutional neural networks (CNNs) has revolutionized the manner in which images are recognized because the manually constructed descriptors have been substituted by the hierarchical learning of features which have been directly applied to the image pixel data. This family of architectures has enabled high performance in medical diagnostics, agriculture, remote sensing, manufacturing, food inspection and waste sorting but much of that performance has been demonstrated so far in conditions that are not as representative of the environment of operational practice as actual deployment environments. This survey explains the development of CNNs since the initial document-recognition systems to domain-specific applications in the real world, based on what can be learnt about implementation, and not on architecture novelty. Themes such as transfer learning, dataset curation, class imbalance, domain shift, compression-aware deployment, lightweight inference, interpretability, and performance reporting by operation constraints are some of the important ones. The literature examined suggests that the gains in accuracy can be substantial, but the issue of reliability in the area is a question of the quality of the information, the conditions under which it was obtained, the validity of the labelling, and the infrastructure of usage just as much as the issue of the complexity of the model underlying it. Restrictive variables such as the poor extrapolation of external variables, benchmark inflation using controlled imaging, insufficient analysis of calibration, and failure mode and fairness treatment are common. The sphere itself is highly important as the mediation of visual classification of the screening, monitoring, industrial inspection and management of the environment becomes more and more quick. Complex research agenda, therefore, requires higher-levels of assessment protocols, broader implementation using multimodal and more resource sensitive methods, and more integration of consistency between laboratory and actual operational judgment.*

Keywords: *Agricultural Image Analysis, Convolutional Neural Networks, Image Classification, Medical Imaging, Real-World Deployment, Transfer Learning*

I. INTRODUCTION

Image classification has always been one of the key areas of computer vision as the categorical processing of the visual data is the basis of screening, diagnosis, monitoring, retrieval and automated decision support. The earlier paradigms relied on features that were handcrafted, featured pipelines and feature selection and were prone to work well under limited imaging conditions but failed to work well as the variation of the scene was expanded. Convolutional architectures re-architected this pipeline, adding local receptive fields, shared weights and hierarchical representation learning to one trainable architecture. There was a breakthrough in document recognition, with convolutional networks providing examples that indicated that strong image recognition could be provided even when there was variability in shape, noise on the segmentation and heterogeneity of operation in traditional pattern-recognition pipelines that were complex to execute [1].

The next rediscovery of deep learning expanded the range of practice and theory of CNNs. Literature has demonstrated that the modern CNN innovations were not due only to a single innovation but as a combination of more profound architectures, non-saturating activations, regularization methodologies, larger labelled datasets, and hardware acceleration [2]. In parallel, additional general-purpose research of visual representation places CNN-based image classification in a broader trend of end-to-end representations research to classification, detection, and segmentation tasks, thus image recognition is a gateway task to a diverse range of downstream tasks [3]. CNNs applicability to the present research environment is thus not limited to benchmark performance, the architecture has proven to be a workable informative in converting visual information to a workable computational estimation in the realm of science and industry.

This interdisciplinary usefulness of such development is particularly apparent in areas where human know-how is costly, irregular, space bound, or time bound. In dermatology, CNN-based classification achieved performance comparable to that of human dermatologists on several skin-lesion tasks when trained on an extremely large image collection that showed that large-scale visual learning would be useful in generating clinically significant contrast across broad diagnostic groups [4]. More broadly in medical image analysis, survey evidence showed that convolutional models rapidly became a dominant

This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/2.0/)

analytical paradigm across classification and related tasks because their feature hierarchies could adapt to texture, contrast, morphology, and other patterns that were difficult to encode manually [5]. The advances made it so that applied image recognition is a trans-disciplinary area of computer science and medicine, agriculture, manufacturing, environmental management, and geospatial analytics.

Although there have been substantial advances, the research-related environment is characterized with unaddressed problems. Reported high accuracy in curated settings does not necessarily mean that the model will be reliable in real-world applications with class imbalance, changing lighting conditions, acquisition noise, compression artifacts, device variability, changing label definitions and population changes. In disease recognition in agriculture, e.g. tightly controlled image recognition can fail in field conditions, in medical screening institutional variation can transpire through model validity, in manufacturing throughput and latency limits impose feasible architectures as hard as accuracy does. The other continuing challenge is the lack of alignment between evaluations based on a leaderboard and usefulness. Uncertainty calibration, cost of false-negative, maintenance load and complexity of integration are usually hidden behind accuracy, area under the curve or F1 score. These limitations indicate that real-world classification cannot be understood through architectural ranking alone; dataset realism, calibration, failure modes, and deployment constraints are equally important.

Another major gap concerns the limited transfer of lessons across application areas into a shared interpretive framework. The focus of medical imaging research is on validation and expert comparison, field deployability and data scarcity in agricultural research, transfer learning across scene distributions in remote sensing, efficiency and defect-pattern recognition have been central to manufacturing research, and demographic and material variation to waste sorting research. Each area has thus decided to look at a different facet of the same fundamental issue of how CNNs can act when image classification is exposed to operational uncertainty. An analysis of applied image recognition can thereby show cross-domain similarities that domain surveys are inclined to ignore.

The current review aims at analysing CNN-based real-world classification based on lessons learned in applied image recognition literature as opposed to an architecture-by-architecture chronology. The sections which follow are approximately an assessment of the literature thematically, followed by a tabulation of the main methodological approaches adopted in the domains, a subsequent analytical treatment of results reported, limitations and translational implications. The review ultimately aims to identify directions for moving CNN classification from high-performing experimental systems toward more reliable operational intelligence.

II. LITERATURE REVIEW

The applied CNN literature reveals a consistent pattern: the success can be observed most evidently when domain structure is equal to visual regularities that can be captured by convolutional hierarchies in an efficient manner. CNN-based systems were shown to be effective in assisting with high sensitivity and specificity in referral-oriented care in diabetic retinopathy screening (high volume retinal classification) [6]. These medical articles were not merely additions to proof of concept work, but they offered a model in which image classification has become an element of clinical workflow design with specific implication of triage capacity, expert access and quality assurances.

The contribution of realism to the data was disclosed in the field of agricultural image recognition that took another route to broaden the field. CNNs were shown to be highly effective in controlled plant-disease datasets, which achieved very high classification accuracy [8]. These findings were further confirmed by larger plant disease studies that estimated the diagnostic performance of various deep models using larger combinations of diseases classes [9]. However, field-based studies revealed a more sobering result: model capacity alone is not the sole determinant of performance in operational situations: symptom visibility, complexity of the background, and the conditions of device-level deployment are equally important factors [10]. A significant difference between controlled-image success and field-image robustness was thus brought about by the agricultural literature.

Another set of methodological and conceptual input came with remote sensing. Classification of scenes based on aerial or satellite imagery is different than medical and agricultural classification, as land-use patterns are a mixture of texture, scale, spatial distribution and sensor-specific attributes. The transfer-learnings in remote sensing demonstrated that models that are trained with natural images could still be very useful, but the choice of layers, fine-tuning strategy and the size of the datasets have a strong impact [11]. Connected work on scene-classification stressed that more training completeness and data heterogeneity can have a significant impact on performance which is particularly significant in geospatial work where category boundaries are confounded by inter-scene variation and intra-class heterogeneity [12]. The remote-sensing literature therefore showed that representational reuse can be powerful, although it is not inherently domain-agnostic.

The conflict between the requirements of precision and functionality was emphasized in industrial and infrastructural applications. Compression-conscious wireless sensor network classification showed that the performance of CNN can be

maintainable even in the case of lossy upstream image transmission, as long as compression is regulated in respect of classification tolerance [13]. Surface-defect and wafer-map analyses led to systems that were resource-conscious and with low latency where inference speed, memory footprint and class imbalance were the primary design variables and not secondary engineering considerations [14]–[15]. Such studies are particularly significant since classification systems in industrial settings are typically designed to work in continuous production settings where false alarms involve costs in the process and where latency in inference can be very detrimental to utility even in cases where the nominal accuracy is great.

Table 1: Summary of Key Findings

Ref	Focus	Key Findings
[6]	Diabetic retinopathy screening from retinal fundus photographs	High sensitivity and specificity supported automated referral-oriented screening, indicating strong suitability for high-volume ophthalmic triage.
[7]	Multi-pathology chest radiograph classification	Concurrent prediction of 14 thoracic findings showed expert-comparable discrimination on retrospective evaluation, while raising questions about institutional generalization.
[8]	Controlled plant disease classification across crop species	CNN training on 54,306 leaf images achieved very high held-out accuracy, demonstrating strong separability under controlled acquisition conditions.
[9]	Large-scale plant disease detection and diagnosis	Deep models trained on 87,848 images and 58 classes reached high diagnostic performance, confirming the value of deeper architectures for fine-grained crop pathology.
[10]	Field-image cassava disease recognition	Transfer learning enabled deployable classification under agricultural field conditions, but overall accuracy declined relative to laboratory-style datasets.
[11]	Remote-sensing scene classification via transfer learning	Fine-tuned CNNs pretrained on natural images remained effective, yet performance varied with network depth, dataset size, and layer specialization.
[12]	Remote-sensing scene classification with enhanced training methodology	Expanded training completeness and representation transfer improved scene-class discrimination in complex geospatial imagery.
[13]	Compression-aware classification in wireless sensor networks	Lossy transmission remained compatible with backend CNN classification up to quality thresholds, showing that pipeline design can offset bandwidth constraints.
[14]	Workpiece surface defect classification and detection	CNN-based feature extraction improved defect discrimination and detection performance for industrial inspection tasks with visually subtle anomalies.
[15]	Wafer map failure recognition in semiconductor manufacturing	Multi-class CNN recognition on WM-811K supported automatic pattern identification and reduced dependence on handcrafted defect descriptors.
[16]	Food image recognition and safety-oriented classification	CNN-based recognition improved categorization of food imagery while illustrating the importance of texture and shape representation under heterogeneous appearance.
[17]	Municipal solid waste image classification	EfficientNet-based transfer learning improved litter categorization, highlighting demographic-region specificity in waste appearance.
[18]	Inorganic packaging waste sorting	CNN classification achieved test accuracy of 95.6% with precision and recall of 0.96, indicating practical promise for automated sorting pipelines.

More CNN applications were introduced to consumer-oriented and sustainability-oriented environments through food inspection and waste sorting. Recognition of food image experiments revealed that CNNs could be used to assist in safety detection and category recognition in domains where the appearance variation is contingent on the packaging, occlusion, surface texture and illumination [16]. Research on waste classification showed that transfer learning and EfficientNet-scale architectures can substantially improve sorting performance of municipal solid waste categories, with the subsequent studies

on inorganic packaging wastes showing high levels of precision and recall and test-scaling above 95 per cent, which suggested that automated sorting processes have a practical potential [17-18]. The above application areas highlighted the social and environmental applicability of visual classifiers but also revealed the complexity of preserving accuracy when there is a change in the appearance of material across sites, between practices and camera set-ups.

Another literature train was on field realism and deployability that were discussed more specifically. Multi-species training in perennial ryegrass to detect weeds enhanced the difficulty category recall, which demonstrates that the class design and ecological diversity influence the quality of detectability [19]. This argument was further extended to edge-oriented plant disease classification by demonstrating that lightweight architectures like MobileNetV3-small can maintain the extremely high test accuracy of heavier backbones at resource limits that are far more realistic [20]. The same study of semiconductor wafer-bin devices revealed that light models can be made to be approximately as accurate as larger networks and the number of parameters and inference overhead can be significantly lowered [21]. This shifted the literature from asking whether CNNs could classify images accurately to asking which CNN designs could be relied upon under the constraints of the target environment.

The thematic record also indicates repetitive restrictions. Numerous applied research remains based on internally curated datasets of limited acquisition variety. External validation remains limited, error analysis is uneven, and calibration is reported less frequently than discrimination performance. A number of domains exhibit optimism in benchmarks due to manipulated backgrounds or focused objects or a small heterogeneity of devices. The literature is now mature enough to display a strong domain value as well as evidence of a strong need of more rigorous standards of evidentiary rigor. CNN-based classification has become plausible in a wide range of industry, but plausibility at this point is not based on the metrics of single headline but rather is based on the clarity in treating the realism of the dataset, failure mechanisms and application scenarios.

Table 1 demonstrates that the best reported results are most likely to be obtained in cases where variability of images is partly restricted or in cases where the transfer learning is matched with the target distribution. The table also gives a wider lesson, that domain success is contingent to varying bottlenecks in various sectors, with label reliability predominating in some of the medical tasks, environmental variation predominating in agriculture and waste sorting, and computational efficiency predominating in industrial inspection.

III. METHODOLOGY

This review was conducted as a narrative synthesis of applied CNN literature published between 1998 and 2024, with primary emphasis on work from 2016 onward, reflecting the period of large-scale deep learning adoption. Literature was identified through searches of IEEE Xplore, Scopus, PubMed, and Google Scholar using the following term combinations: “convolutional neural network image classification”, “CNN transfer learning real-world”, “deep learning medical imaging”, “CNN agricultural disease detection”, “remote sensing scene classification deep learning”, “industrial defect detection CNN”, and “waste sorting image classification”. The inclusion criteria included that the papers (a) used CNN-based models to a real-world image classification task, (b) gave quantitative performance measures on a held-out or validation set, and (c) were published in a peer-reviewed journal or conference proceedings. Articles were not included where they were solely theoretical without empirical analysis, they dealt with other tasks (e.g., pure object detection or segmentation without a classification part), or they did not provide full-text access. Foundational architectural works were included selectively to provide historical context. The final corpus encompasses 21 primary sources synthesized thematically across six application domains: medical imaging, agricultural disease recognition, remote sensing, industrial inspection, food recognition, and waste sorting. The review follows narrative synthesis conventions and does not employ formal meta-analytic aggregation, given the heterogeneity of evaluation protocols across domains.

The methodological core of applied CNN classification is better understood as a multilayered decision pipeline than as the selection of a single model. In the majority of real world investigations, the domain-specific image acquisition phase starts with a domain-definite image acquisition phase that is influenced by device type, operator actions, compression, complex background, and prevalence of targets. The second step is the construction of the dataset by annotation, definition of classes and by partitioning it into train, validation and test sets. The third step is to select a representation strategy, typically either training either on a large enough domain data or fine-tuning an existing backbone. The next phases deal with augmentation concern, optimization concern, evaluation concern, and deployment constraint. This sequence has been found over and over again in the medical, agricultural, industrial, and environmental uses, although network families differ considerably.

One of the key trends in the literature is a high prevalence of transfer learning. Remote sensing, cassava diagnosis, waste classification, and numerous industrial tasks are done using CNNs that have been pretrained on large natural-image corpora since because target datasets are often too small, too imbalanced, or too expensive to annotate effectively for training

from scratch [10]–[11], [17]. Transfer learning acts as a computational shortcut as well as a regularization scheme that performs stabilization of the training process and speed-up convergence. Yet the literature has shown that pretrained features are not necessarily the best. In cases where the statistics of texture or scale or sensor physics change rapidly with those of natural images, domain mismatch arises, and the issue of selective fine-tuning and representation adaptation is of special interest.

The other common methodological axis is one that concerns architectural scaling and deployment effectiveness. The implemented, previous systems were more likely to encourage parameter laden networks, as the leaderboard advancement encouraged a deeper, and denser, network structure. More recently, in applications that are less network-focused, notably edge agriculture and semiconductor inspection, researchers have abandoned heavyweight backbones in favour of lightweight families like MobileNet and ShuffleNet and inference requirements are more often addressed by operating on a constrained hardware platform or in a time-constrained environment [20-21]. This change is a radical conceptual change. In practice classification, the maximum that is the highest isolated classification is not the optimum performance, but rather a performance envelope, a balance between discriminatory power, latency, memory, energy requirement and maintainability.

A methodological component is comprised equally of data engineering. Repeated studies performed in applied CNN have demonstrated that augmentation, class balancing, background control and annotation quality may have as significant an impact as architecture choice. Training Multispecies weed enhanced the recollection of challenging categories as it diversified the representation of classes [19]. It was demonstrated that compression-conscious classification in wireless sensor networks that compression-induced data degradation could be eliminated only to task-dependent limits, thus, directly relating data fidelity to the reliability of the backend model [13]. Clinical label construction based on report or expert rating is another contributor to methodological variation which can cause strong effects on the apparent performance [6-7]. This is why the literature increasingly treats dataset design as an ongoing methodological determinant of model behaviour rather than as a one-time preliminary step.

Interpretability and evaluation methodology occupy a disproportionately important place in applied classification. There are domains that compare model outputs and expert judgments and those domains that mainly use held-out benchmark measures. Class activation visualisation, saliency mapping, confusion analysis and threshold tuning are only mentioned here and there but a lot of the literature is still focused on discrimination scores and not operational performance calibration. Such an imbalance is methodologically contingent. In the case of a classifier to be used during screening, sorting or defect triage, it is not only necessary to divide the classes but also to communicate where confidence is weak, prevalence is low, or image quality compromises trustworthiness. The demonstration of lack of systematic calibration and uncertainty treatment can be listed as one of the most obvious cross-domain methodological gaps.

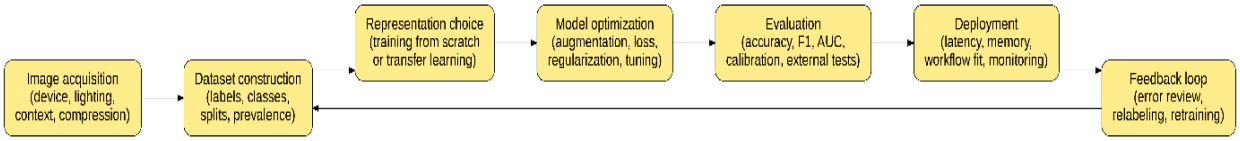


Figure 1: Conceptual Framework

Figure 1 synthesizes the prevailing methodological stages reported in the literature, from image acquisition to deployment feedback. The diagram highlights that applied performance is not dependent on architecture alone as it is determined by the joint conditions of data, model design, evaluation strategy and operational constraints.

The last pattern of methodology entails hybridization of classification and task specific system design. Mobile and edge deployment, in the agriculture sector, modifies the model budget acceptable and could demand detector-classifier combinations. In manufacturing, image classification is often embedded within larger production pipelines. In sorting waste, definition of categories relies on downstream infrastructure of handling, and not just on pure visual semantics. The literature therefore suggests that CNN classification is most useful when treated as one component of a broader socio-technical system. The latter observation can be used to understand why benchmark success is not always followed by field success although laboratory indicators are high.

IV. DISCUSSION

The literature consistently reports strong discriminative performance of CNNs across many applied settings; however, the strength and significance of that performance significantly depend on the realism of data. High sensitivity and specificity in diabetic retinopathy screening favoured the opinion of CNNs contributing large-scale referral processes, particularly during circumstances where the specialists lack capacity to screen patients [6]. Similar results were obtained in the classification of chest radiograph which showed high retrospective discrimination with a variety of pathology labels [7]. However, the two disciplines point out an important fact, that an effective screening classifier is not a full-fledged general diagnostic system.

Reported outcomes are determined by the determinants of label provenance, case mix, institutional imaging conventions and prevalence structure. The literature as such encourages the use of CNN as opposed to free autonomous interpretation on assistive triage and prioritization.

In the studies of agriculture, the contrast between the controlled and the operational performance is even more evident. The remarkably high degree of disease-recognition on the images of the leaf with low-background variation was made possible by PlantVillage paradigm [8-9]. This finding was corroborated by the findings of this study that confirmed that symptom carrying textures and patterns and colour change of lesions were highly learnable under normal imaging conditions. But field-based tests of cassava fields were showing poorer aggregate performance, yet, at the individual category level, individual disease categories are still showing high class-level performance [10]. This disparity implies that the clutter of the background, the angle of the leaves, delicacy of the symptoms and mixed conditions in the environment bring about domain shift which is more extreme than in such an environment as the benchmarks have. A significant conclusion regarding applied image recognition at the time is drawn; the stability of the visual world under representation is one of the limitations on performance of the real world classification not due to the richness of the representations under application.

The outcome of remote sensing explains another but a similar problem. Scene classification-to-transfer experiments showed the usefulness of pretrained CNNs and how hierarchical representations learned with natural images can be useful in terms of containing useful mid-level abstractions of geospatial patterns [11]. Meanwhile, the size of datasets was used to define transfer performance, and the extent of specialization in the fine-tuning process [11-12]. The evidence described suggests that the reuse of representations may be the most effective in such cases when target scenes contain edge, texture, and compositional features that most closely match those found in pretraining data to accept a greater range of global deviations in spatial scale, sensor modality or land-use heterogeneity. The biggest thing to keep in mind is that transfer learning is not solely an accuracy-enhancing tool but an empirical hypothesis about the continuity of the features across domains and therefore has to be tested and not assumed.

There is particularly clear evidence of the fact that model value is not independent of operational constraints in industrial inspection studies. The compression-sensitive wireless sensor network classification revealed that the upstream bandwidth constraint could be accommodated without instant failure of the backend classification service, yet compression severity should remain within recognition tolerable limits [13]. Studies of workpiece defects demonstrated that CNN-based feature extraction was more effective with regard to finding and classifying defects on objects, which are very fine on the surface, compared to more inflexible hand-engineered pipelines [14]. Investigations of wafer-map semiconductor production then applied the same concept to semiconductor production, where CNN classifiers were used to recognise multi-class patterns on operationally significant datasets [15,21]. These results are important as pipelines in industries are usually configured in terms of constant throughput, rapid response, and low hardware specifications. A slightly less accurate but much faster model may be more valuable than a slower one with only marginally better benchmark performance.

Another point of lenses of applied outcomes is provided through food inspection and waste sorting. Food-image studies indicate that CNNs can handle classification tasks involving fine-grained variation in shape and texture, although performance is also affected by presentation, occlusion, and surface condition [16]. Strong and practically good results in waste-sorting were reported on the basis of transfer learning and model scaling, such as region-specific EfficientNet tuning and classification of packaging-waste with a precision and recall of 0.96 [17-18]. But here too lie the contradictions of a structural problem: the target classes in waste management and the food system are not entirely characterized by the downstream handling or quality criterion, but by appearance. A classifier which has been trained using one set of disposal taxonomies or packaging profiles in a region can thus not generalize well to a different working situation. The broader implication is that class ontology itself becomes a deployment variable in field classification.

One pattern of results that is repeated is the performance-efficiency trade-off. The edge-oriented agricultural study achieved approximately 99.50% test accuracy on the selected plant dataset using MobileNetV3-small, while offering substantially greater deployment realism under hardware constraints [20]. Wafer-bin studies of semiconductor also achieved about 98 percent accuracy and an F1 of 89.5 percent when using MobileNetV3 in classification with much fewer parameters and much faster training and inference compared to heavier comparators [21]. These findings indicate that a larger model does not necessarily produce better practical classification performance, particularly when computational constraints are considered. The efficient CNNs are able to explain the ideal practical operating point in the event of visually consistent target classes, and binding computational funds.

It is also clear in the literature that it can be composed of data which can affect more than the architectural substitution. Perennial Ryegrass Weed identification demonstrated that it was more effective to train on the multi-species than on the single-species and suggest that a greater ecological coverage could enhance representational strength [19]. The importance of

the agricultural disease recognition in a controlled environment with many labelled corpora was also necessitated and these adaptations and mobile-oriented design were necessitated in field-deployable cassava systems [8-10]. It is only significant that a comparison of performance to the clinicians in medical imaging is well under control in terms of the reference standards, the selection of cases and the partitioning of data [6-7]. The analogous moral here is that CNNs are very sensitive to the quality of epistemic of the input: the quality of labels, delimitation, prevalence realness, and diversity of acquisition are in an interactive sense as much contributing to quality of results as is depth or breadth.

The other theme is remarkable to high interpretation of headlines accuracy. The near ceiling accuracies of controlled set of plant diseases can give the impression that classification problem is not even of interest. It has been demonstrated that field studies that are so misinformed have led to substandard performance with less standardized backgrounds, symptom severity and capture conditions [10]. Similar forms of benchmark optimism appear in other domains, including waste sorting where over selectively labelled datasets can exaggerate the actual performance of sorting-lines and in clinical imaging where obscure not-inference occurs when labels are extracted retrospectively. Out of distribution behaviour, error stratification, calibration and maintenance in the presence of temporal drift requires more emphasis in real world classification. In the absence of such analysis, the outcomes reported can be characterized by the mastery of the dataset rather than the mastery of the environment.

The implication to translation is extensive. CNN-based classifier has become a viable part of screening systems, crop surveillance devices, inspection, and sustainability infrastructure. Meanwhile, the evidence recorded indicates that the surest path towards value is by scoped deployment: massive triage, first-pass classification, anomaly pre-screening, and human-oversight decision support or process monitoring. The literature is far less convincing when deployment claims imply broad autonomy across devices, populations, or operating environments. CNNs have thus worked out not due to the fact that classification was available everywhere, but due to the fact that numerous real tasks are partially automatable when task boundaries are established well and analysis corresponds to operational reality.

Table 2: Method Comparison

Ref	Method	Strengths	Limitations
[8]	CNN training on controlled leaf-image dataset	Very high class separability; scalable disease categorization across many crop-disease combinations	Performance shaped by controlled backgrounds and limited field realism
[10]	Transfer learning for field-image cassava disease classification	Better alignment with deployable field capture; reduced data burden	Lower aggregate robustness than controlled-image settings; background complexity remains challenging
[11]	Transfer learning from natural-image pretrained CNNs for remote-sensing scenes	Effective feature reuse; strong results with modest target data	Sensitivity to layer selection, dataset size, and domain mismatch
[13]	Compression-aware backend CNN classification for sensor-network images	Maintains recognition under bandwidth constraints; system-level efficiency	Performance degrades once compression exceeds task-tolerant thresholds
[17]	EfficientNet-based municipal waste classification	Strong transfer-learning performance; adaptable to domain-specific waste categories	Region-specific appearance and category definitions limit portability
[20]	MobileNetV3-small for edge plant disease classification	High test accuracy with low resource burden; suitable for constrained hardware	Validation remains dataset-specific; broader field heterogeneity requires further testing
[21]	Lightweight CNNs for semiconductor wafer-bin classification	Strong accuracy-speed-parameter balance; realistic for industrial hardware budgets	Performance still depends on defect-distribution stability and dataset representativeness

Table 2 indicates that no single methodological pattern dominates across all practical classification tasks. Transfer learning is the most effective, but its utility can only be determined by compatibility of the domain and lightweight design becomes decisive in circumstances where the resources available to deploy are constrained.

Table 3: Results Comparison

Ref	System	Metric	Outcome
[6]	Retinal fundus CNN screening system	Sensitivity / specificity	High values sufficient for referral-oriented diabetic retinopathy screening
[7]	CheXNeXt chest radiograph classifier	Comparative radiologist-level discrimination	Strong retrospective performance across 14 pathologies
[8]	Plant disease CNN on controlled leaf images	Accuracy	99.35% held-out accuracy
[9]	Deep plant disease diagnosis models	Accuracy	Best models reported accuracy above 99% on curated dataset
[10]	Field cassava disease transfer-learning classifier	Overall accuracy	93% on data excluded from training
[18]	Inorganic packaging waste sorting CNN	Test accuracy / precision / recall / F1	95.6% accuracy; 0.96 precision; 0.96 recall; 0.96 F1
[20]	Edge plant disease classifier	Test accuracy	Around 99.50% with lightweight MobileNetV3-small
[21]	MobileNetV3 wafer-bin classifier	Accuracy / F1 / efficiency	About 98% accuracy and 89.5% F1 with major parameter and speed advantages

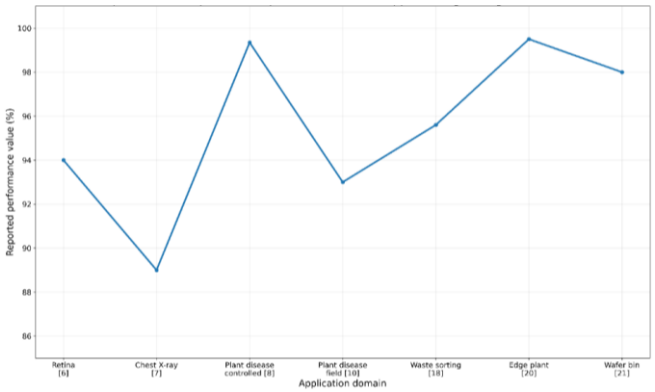


Figure 2: Illustrative Trend in Reported Applied CNN Performance Across Domains

Figure 2 provides an illustrative domain-level comparison based on outcome values reported in representative studies. The graph should not be perceived as a rigid standard as domains of evaluation protocols vary among themselves, rather, it graphically reflects how curated agricultural activities tend to deliver almost ceiling scores, and field and industrial environments demonstrate a more balanced but operationally significant performance.

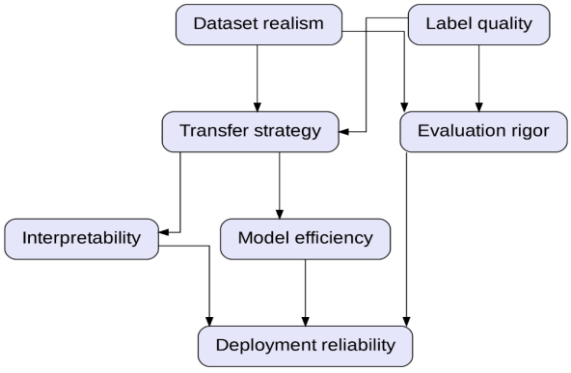


Figure 3: Relationship Diagram Linking Data Realism, Model Design, and Deployment Outcome

The relations between the dataset realism, transfer strategy, computational budget, interpretability, and deployment reliability are highlighted in Figure 3. The relationship diagram reinforces the major argument of this review since it demonstrates that operational success is a product of interdependent design decisions and not independent model enhancements.



Figure 4: Integrated Model for Translating CNN Classification from Benchmark Performance to Operational Value

Figure 4 combines the lessons of the review regarding the analytical model into a translational model between benchmark success, validation strategy, domain adaptation, efficient deployment, and post-deployment monitoring. The figure emphasizes the fact that real world value is determined by the feedback mechanisms, since classification environments change and fixed model may become unreliable with changing data distributions.

V. FUTURE DIRECTIONS

Future advances in CNN-based real-world classification will depend first on more realistic data. Several fields continue to be based on datasets with controlled frame, limited device variety or even on class ontologies that are not fully reflective of operational practice. It should then be given a higher priority to multi-site, multi-device, temporally updated datasets having explicit metadata that records the acquisition circumstances, label provenance, and uncertainty. With these datasets, a more significant analysis of robustness, drift, and failure concentration across sub conditions would be possible as opposed to just aggregate classes.

The second priority is the methodological improvement in evaluation. Accuracy and area-under-the-curve metrics remain useful, but the next-generation efforts should add regular reports of calibration, abstention behaviour, threshold sensitivity, subgroup analysis and external validation under distribution shift. The classification systems in the real world are often in the asymmetric risk structure where there are very different costs of false negatives and false positives. Assessment plans must then match measures to workflow implications as opposed to abstract benchmark norm. In medicine, agriculture and industrial inspection, intervention costs and missed-event costs vary significantly, and this change is particularly important.

Direction is the third direction with respect to resource-aware intelligence. The high performance and low number of parameters of lightweight CNNs have already been demonstrated, as well as their faster inference in semiconductor and agricultural tasks [20-21]. Future directions will be constructed along the line to compression-awareness training, quantization-friendly architecture, adaptation inference, and cooperation with the cloud boundaries. Not only an engineering breakthrough, but it will also directly spell out whether image recognition can be made a practical tool in farms, low-resource clinics, factories and municipal sorting systems where neither can it be assumed of a high-end hardware or continuous connectivity.

More emphasis should also be placed on interpretability and human interaction with the system. Domain experts often require models that not only provide predicted labels but also explain why a prediction was made and when that prediction should be treated with caution, and also the mistakes that the model makes are systematic. This is to be combined with explanation quality and auditing of errors, but this should be placed in the context of validation and retraining practice in future studies. This will be more crucial in areas where CNN outputs are used to pursue action priority and not just batch statistics. The research on interpretability should thus be working on whether explanations are able to enhance the trust calibration, expert review as well as the corrective feedback loops.

Another trend which is promising is interdisciplinary expansion. The cultures of validation have grown somewhat distinct in remote sensing, agriculture, medicine and manufacturing, but each field can enlighten the others. Clinical research provides more traditions of comparative validation, agriculture provides realism regarding field variability and device constraints, remote sensing provides knowledge of transfer-learning in case of a domain mismatch, and manufacturing provides system design that is efficiency-centred. The controlled cross-domain exchange would be beneficial in future studies, especially in the context of powerful labelling protocol, shift identification, lifelong learning and maintenance control.

VI. CONCLUSION

The use of CNNs has transformed visual recognition from a handcrafted-feature problem into a broadly applicable representation-learning problem. Medicine, agriculture, remote sensing, manufacturing, food inspection, and waste sorting show evidence of CNNs being capable of providing high performance in a large number of task formulations in the real world. Meanwhile, the literature suggests that architecture is not the only factor that determines applied success. Dataset realism,

label quality, transfer strategy, calibration, computational budget, and workflow fit jointly determine whether a classifier becomes operationally reliable.

A number of important insights are repeated in the reported studies. To begin with, high benchmark performance is most readily acquired when the acquisition situations are controlled and must not be equated with general environmental robustness. Second, transfer learning is still among the most useful methodological instruments in applied image recognition, the usefulness of which is based on the continuity of source and target visual statistics. Third, lightweight architectures have become a strategic issue since most deployment environments have efficiency as an equally strong reward compared to marginal accuracy improvements. Fourth, the most plausible near-term role for CNN classification is in scoped decision support, triage, and pre-screening rather than fully autonomous judgment.

The current gaps in research are still significant. External validity remains uneven, uncertainty treatment has not been well developed, and data shift has not been well addressed in most studies. The further development will be based on more detailed datasets, more robust evaluation design, feedback conscious deployment, and enhanced cooperation between algorithm developers and experts in the field. Practical classification is already past the demonstration of concept, but fully developed operational trust will require solutions and demonstrations as complicated as the environment CNNs can now operate in.

- **Conflicts of Interest:** The author declares no conflict of interest.

VII. REFERENCES

[1] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278–2324.

[2] Rawat, W., & Wang, Z. (2017). Deep convolutional neural networks for image classification: A comprehensive review. Neural Computation, 29(9), 2352–2449.

[3] Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S. (2016). Deep learning for visual understanding: A review. Neurocomputing, 187, 27–48.

[4] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115–118.

[5] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60–88.

[6] Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., Kim, R., Raman, R., Nelson, P. C., Mega, J. L., & Webster, D. R. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA, 316(22), 2402–2410.

[7] Rajpurkar, P., Irvin, J., Ball, R. L., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C., Shpanskaya, K., Lungren, M. P., & Ng, A. Y. (2018). Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLOS Medicine, 15(11), e1002686.

[8] Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. Frontiers in Plant Science, 7, 1419.

[9] Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. Computers and Electronics in Agriculture, 145, 311–318.

[10] Ramcharan, A., Baranowski, K., McCloskey, P., Ahmed, B., Legg, J., & Hughes, D. P. (2017). Deep learning for image-based cassava disease detection. Frontiers in Plant Science, 8, 1852.

[11] Pires de Lima, R., & Marfurt, K. (2020). Convolutional neural network for remote-sensing scene classification: Transfer learning analysis. Remote Sensing, 12(1), 86.

[12] Yu, X., Liu, X., Zhou, X., & Huang, C. (2017). Deep learning in remote sensing scene classification: A data augmentation enhanced convolutional neural network framework. Geo-spatial Information Science, 20(2), 142–149.

[13] Ahn, J., Park, J., Park, D., Paek, J., & Ko, J. (2018). Convolutional neural network-based classification system design with compressed wireless sensor network images. PLOS ONE, 13(5), e0196251.

[14] Xing, J., Ng, T.-T., Chen, Y., Wang, H., Xie, M., & Yang, X. (2021). A convolutional neural network-based method for workpiece surface defect detection. Measurement, 176, 109022.

[15] Zheng, H., Li, X., Li, Y., Kim, M., & Kim, J. (2021). A deep convolutional neural network-based multi-class image classification for automatic wafer map failure recognition in semiconductor manufacturing. Applied Sciences, 11(20), 9769.

[16] Wang, Y., Chen, J., & Wang, J. (2021). Food image recognition and food safety detection method based on neural network. Computational Intelligence and Neuroscience, 2021, 5562784.

[17] Malik, M., Sharma, S., Uddin, M., Chen, C.-L., Wu, C.-M., Soni, P., & Chaudhary, S. (2022). Waste classification for sustainable development using image recognition with deep learning neural network models. Sustainability, 14(12), 7222.

[18] Laksono, P. W., Anisa, A., & Priyandari, Y. (2024). Deep learning implementation using convolutional neural network in inorganic packaging waste sorting. Franklin Open, 8, 100146.

[19] Yu, J., Schumann, A. W., Cao, Z., Sharpe, S. M., & Boyd, N. S. (2019). Weed detection in perennial ryegrass with deep learning convolutional neural network. Frontiers in Plant Science, 10, 1422.

[20] Khan, A. T., Jensen, S. M., Khan, A. R., & Li, S. (2023). Plant disease detection model for edge computing devices. Frontiers in Plant Science, 14, 1308528.
[21] Shin, E., & Yoo, C. D. (2023). Efficient convolutional neural networks for semiconductor wafer bin map classification. Sensors, 23(4), 1926.