A Survey of Deep Learning-Based Pattern Classification Frameworks for Medical and Agricultural Image Analysis

Deepak Mane[1,3], Bharati Ainapure[2,3]
[1]Post Doctoral Research Scholar, Lincoln University College, Malaysia, India;
[2] Adjunct Research Faculty under the Lincoln Global Postdoctoral Researcher (LGPR) Programme at Lincoln University College, Malaysia;
[3]Professor, Department of Computer Engineering, Vishwakarma University, Pune, India

Email:
pdf.deepakmane@lincoln.edu.my
ainapuressa@gmail.com

Abstract: Disease classification from images plays a central role in decision support in today's world of medicine and agriculture, and scientists working on the problem in these two disciplines developed their solutions almost independently. The paper provides a survey of ten deep learning approaches to disease classification (2016-2025), five in medicine and five in plant sciences, focusing on the four key aspects: architecture, dataset, performance reported, and limitations. While benchmark classification accuracy is now close to 97-99% level, the actual generalization is not well tested, computational requirements far exceed those of edge devices, and interpretability is an option. None of the ten surveyed papers evaluated their model on data from the other discipline even though they use exactly the same set of visual features for classification - texture abnormalities, irregular boundaries, and gradient color changes. Based on this gap analysis, this paper presents the idea of a general and unified approach in computer vision which is architecture agnostic framework it could be implemented using any deep learning or any hybrid deep learning method and operates seamlessly in both domains without the need for complete retraining.

Keywords: Pattern Classification; medical images; deep learning; vision transformer; disease detection, convolutional neural network; Fuzzy hypersphere neural network

---

1. Introduction

Image classification into categories of diseases is one of those problems that, computationally, is already solved using many benchmarks. What remains unsolved and constitutes the essence of this review is the reliability of that classification when applied in real-world scenarios involving various image formats and machine learning models that the end-user would have enough confidence and control over. This will be the underlying theme of all the following sections. In both cases, the basic computational challenge is the same – disease imagery classification by means of fast, accurate and understandable model. This paper makes three contributions: Firstly, it offers a systematic comparative review of ten papers (five in the field of medicine and five in the field of agriculture) analyzed from the same perspective. Secondly, it outlines the main gaps in the current state of art based on the literature review analysis. And thirdly, it develops a framework of deep learning-based CNN-Fuzzy Hypersphere Neural Network for classification of both medical and agricultural disease images using the same trained backbone.

Organization of paper goes as follows. Section 2 describes review of the ten articles. Section 3 describes the observations made during analysis. Section 4 describes the gap analysis and proposed framework. Section 5 concludes.

2. Related work

This survey examines ten deep learning-based image classification studies published between 2016 and 2025, five drawn from medical imaging and five from agricultural disease detection. Each study is reviewed against the same four-dimensional lens: the architecture employed, the dataset used, the reported performance, and the limitation the authors themselves acknowledge or that becomes apparent on closer reading. Studies were selected to represent the full architectural arc of both fields, from early CNN baselines to recent hybrid and explainable designs, rather than to maximise topical coverage. The following two subsections present the medical and agricultural studies in turn, followed by a comparative summary table.

2.1 Medical Imaging Studies

Hemdan et al. [1] (2020) employed seven pre-trained CNNs on a small public COVID-19 chest radiography dataset; among them, VGG19 and DenseNet201 had achieved over 90% accuracy, yet the dataset consisted of less than 200 images and the results were not externally verified. Khan et al. [2] (2021) suggested an end-to-end CNN for four-class brain tumour classification (glioma, meningioma, pituitary, healthy) using MRI images, obtaining 96% accuracy on the Cheng et al. dataset without using any attention mechanism or saliency-based verification of predictions. Liu et al. [3] (2021) developed Swin Transformer using self-attention across shifted local windows, reducing the complexity of attention operations from quadratic to linear, allowing transformers to process high-resolution medical images for the first time. Swin attained 97-98% accuracy but was still 3-5 times more expensive in computations than ResNet-50 in terms of accuracy improvement. Hadhoud et al. [4] (2024) developed a two-stage hybrid approach that uses the ResNet-50 and ViT-B/16 for binary and three-class classifications of diseases of the chest. Binary classification achieved 98.97% accuracy for tuberculosis using X-ray images from Qatar and Dhaka University datasets, and 96.18% for three classes (tuberculosis, viral pneumonia, bacterial pneumonia) using pneumonia X-rays from Guangzhou Women's and Children's Medical Center. But Hadhoud et al.'s approach was tested on chest radiography only without separating the contributions of the CNN and ViT stages in their pipeline. Patel et al. [5] (2024) proposed a multi-class classification of COVID-19, pneumonia, tuberculosis, and pulmonary nodules in chest X-ray images by using an EfficientNet-B4 transfer-learning model with the Grad-CAM based explainability which achieved 96% accuracy. But Patel et al.'s approach was verified only for one organ system.

2.2 Agricultural diseases

Mohanty, Hughes & Salathe [6] (2016) introduced the PlantVillage benchmark, fine-tuning AlexNet and GoogLeNet on 54,306 studio-quality images with 26 classes and 14 crop types at 99.35% accuracy. It was noted by the authors themselves, that the network could not identify the same crops in real-world scenarios, and the accuracy was found to be only around 70-80% in field tests. Too et al. [7] (2019) finetuned four backbones (DenseNet121, ResNet50, InceptionV3, VGG16) under similar conditions on PlantVillage dataset, and got the highest accuracy with DenseNet121, scoring 98% but using only the one dataset, composed of studio-quality images. The work by Chen et al. [8] (2022) seems to be the most eco-friendly from the ones considered: EfficientNet-B4 trained on the combination of PlantVillage and the small field-collected image set achieved 97% accuracy, losing some performance on field images — an important but incomplete step towards ecological validity. Jia et al. [9] (2025) suggested ConvTransNet-S, which is a novel combination of CNN and Transformer, incorporating the Local Perception Unit with lightweight self-attention achieving 98.85% accuracy on PlantVillage and 88.53% on a self-constructed complex background field dataset of 10,441 images – an unprecedented lab-to-field comparison. Like all other papers related to agriculture mentioned in this essay, it never explores the possibility that medical imaging, which is facing similar problems, can serve as a complementary area. Pal et al. [10] (2025) presented Mob-Res, which is a lightweight residual MobileNetV2-based CNN consisting of just 3.51 million parameters, along with explainable-AI based visualization of the prediction results. It achieved an accuracy of 99.47% on PlantVillage and 97.73% on a more challenging 58-class Plant Disease Expert dataset. This architecture-specific lightweight model remains untested against transformer-based models.

Table 1. Comparison between the ten papers analyzed, divided into five medical papers and five agricultural papers published between 2016-2025.

Ref	Domain	Year	Approach / Architecture	Accuracy (%)	Challenges
[1]	Medical	2020	COVIDX—Net, VGG19, DenseNet201, MobileNetV2	>90	[N<200 images; not validated externally
[2]	Medical	2021	Custom CNN, brain MRI classification (4 classes)	96	No saliency/attention verification
[3]	Medical	2021	Swin Transformer (shifted window attention mechanism)	97-98	High computational cost compared to ResNet-50 for slight improvement
[4]	Medical	2024	Hybrid ResNet 50 + ViT B/16	96 to 98	Chest domain only
[5]	Medical	2024	EfficentNet-B4 + Grad-CAM	96	Classifies single organ system (chest only)
[6]	Agricultural	2016	AlexNet, GoogLeNet (PlantVillage dataset)	99.35	Failure rate increases to 70-80% in the field
[7]	Agricultural	2019	DenseNet121 vs. 3 other architectures (transfer learning)	98	Single studio dataset only
[8]	Agricultural	2020	EfficientNet-B4 + field image subset	97	Field subset small, drop unquantified
[9]	Agricultural	2025	Hybrid CNN-Transformer, 2 datasets	98	No cross-domain consideration
[10]	Agricultural	2025	Lightweight MobileNet V2 based residual CNN	97 to 99	Architecture specific-untested on transformers

3. Cross-cutting Findings

Three trends stand out in all ten papers. First, the accuracy range has shrunk to 96-99.35%, independent of the publication date or model architecture, suggesting that current benchmarks (Plant Village, ChestX-Ray14, the MRI data of Cheng et al.) have reached saturation point and not that the algorithms have found an optimal architecture. Second, the combination of CNN and Transformers has emerged as a standard high-accuracy architecture in both cases post-2021, without any citation across literature between the two hybrid architectures for medicine and agriculture; two communities came up with the same architecture independently of each other. Third, explainability and computational costs are not routinely reported; only two out of ten studies have provided explainability ([5],[10]), and computational cost metrics such as parameter count and FLOPs are not available in over half of the studies

4. Gap Analysis and Proposed Unified Framework

4.1 Critical Gaps

G1 – Domain lock-in: all reviewed models are constructed and evaluated in the context of one domain alone, and porting them to another domain requires complete reconstruction.

G2 – Separate research streams: medical and agricultural classification research share neither datasets nor codebase nor conference venue, even though they are solving the same type of problems from a structural perspective.

G3 – Lab-to-field failure: clean benchmarks have little resemblance to real-world motion blur, changing light conditions, and heterogeneous devices, with accuracy dropping by 15-20 points when applied to field or clinical imagery.

G4 – Computing capacity wall: latest hybrid or transformer architectures consume more parameters and memory than can be afforded by mobile/embedded deployment target.

G5 – Explainability gap: eight out of ten reviewed approaches give no output other than class, which becomes increasingly unacceptable in light of regulatory frameworks such as EU AI Act and its high-risk classification of medical diagnostic AI.

G6 – No unified framework: after nine years of parallel research efforts, none of the studies has developed or tested the same architecture in both domains, even though the same low-level visual cues – texture, edge, colour – power both disease-detection pipelines.

4.2 Generalized Unified Framework Proposal

In order to resolve the issues of G1, G2, and G6, we propose here an architecture-agnostic unified framework rather than a specific model. This framework comprises one shared component – feature-extraction backbone, which is pre-trained simultaneously on a combined dataset of medical and agricultural disease images, along with a lightweight domain-specific adapter which customizes the shared features for the domain of chest radiographs, dermoscopic images, or leaves without re-training the backbone. This backbone is intentionally left general and can be any kind of deep or hybrid deep learning architecture which demonstrates the best results for a particular moment of time – from convolutional network to vision transformer, from CNN-Transformer hybrid to CNN-Fuzzy Hypersphere Neural Network (proposed as one of the instantiations of such a backbone). In the last case, the standard soft-max classification head is substituted with Fuzzy Hypersphere Neural Network layer, where each class of diseases is modeled as a hypersphere in feature space – an approach more appropriate for modeling overlapping classes and uncertain predictions, which is quite typical for the ambiguous early stages of the disease, and also interpretable geometrically. In this manner, the contribution remains focused on the unified cross-domain architecture and the evaluation framework rather than on a particular backbone. In the future, one can use any other deep learning or hybrid deep learning architecture in place of CNN-FHNN without violating the idea of shared-backbone with adapter. Whether any given instantiation will perform better than single-domain baselines on both benchmarks is an empirical question which is raised but not resolved in this paper.

4.3 Research Priorities for Study

(1) Construct a public heterogeneous benchmark utilizing real-world medical and agricultural images with compulsory out-of-distribution splits. (2) Test and evaluate the proposed approach using several different types of backbones, including CNN, ViT, CNN-Transformer, and CNN-FHNN, while also carrying out cross-domain transfer learning experiments as well as domain adapter ablation. (3) Perform cross-domain pre-training research on shared and single-domain backbones. (4) Design the sub-10M parameter, sub-100ms hybrid models by using techniques such as pruning, knowledge distillation, and mobile-aware architecture search.

5. Conclusion

A review of ten papers related to image classification utilizing deep learning methods was conducted; the review covers both medical image classification and agricultural diseases detection applications, published during 2016 and 2025. In this review, a four-element approach: problem, data, network architecture and an open limitation was used to assess all of the selected papers. Current benchmark accuracy levels have hit the ceiling to the extent where performance of a convolutional neural network from 2016 and hybrid transformer from 2025 differs in less than a few percent points on the same data. Therefore, future development should focus on more difficult out-of-distribution testing and not on marginal benchmark improvements. The difference between the medical image classification research field and the agricultural disease detection research field arises due to organization of respective research communities, but not because of different problems solved there, as the visual task is the same in both cases: use of texture, boundaries and color to classify images. Creation of a common data set, backbone and evaluation protocol would be technically easy but not tried so far. The proposed generalized and architecture agnostic unified framework which can be implemented using either CNN-FHNN or other deep learning and hybrid deep learning backbones is such a step towards it.

References

1. Hemdan, E. E. D., Shouman, M. A., & Karar, M. E., " COVIDX-Net: A framework of deep learning classifiers to diagnose COVID-19 in X-ray images", arXiv:2003.11055, 2020 https://arxiv.org/abs/2003.11055
2. Khan, H. A., Jue, W., Mushtaq, M., & Mushtaq, M. U. , "Brain tumor classification in MRI image using convolutional neural network", Mathematical Biosciences and Engineering, 17(5), 6203–6216, 2020. https://doi.org/10.3934/mbe.2020328
3. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B., "Swin Transformer: Hierarchical vision transformer using shifted windows", In Proceedings of the International Conference on Computer Vision (ICCV 2021), 10012–100222021. https://doi.org/10.1109/ICCV48922.2021.00986
4. Hadhoud, Y., Mekhaznia, T., Bennour, A., Amroune, M., Kurdi, N. A., Aborujilah, A. H., & Al-Sarem, M. , "From binary to multi-class classification: A two-step hybrid CNN-ViT model for chest disease classification based on X-ray images", *Diagnostics*, 14(23), 2754, 2024. https://doi.org/10.3390/diagnostics14232754
5. Patel, A. N., Murugan, R., Srivastava, G., Maddikunta, P. K. R., Yenduri, G., Gadekallu, T. R., & Chengoden, R. ,"An explainable transfer learning framework for multi-classification of lung

diseases in chest X-rays", *Alexandria Engineering Journal*, 98, 328–343, 2024. https://doaj.org/article/e23edc9bb26a46bbb6ab3242a4532b15
6. Mohanty, S. P., Hughes, D. P., & Salathé, M. , "Using deep learning for image-based plant disease detection", Frontiers in Plant Science, 7, 1419, 2016. https://doi.org/10.3389/fpls.2016.01419
7. Too, E. C., Yujian, L., Njuki, S., & Yingchun, L., "A comparative study of fine-tuning deep learning models for plant disease identification", Computers and Electronics in Agriculture, 161, 272–279, 2019. https://doi.org/10.1016/j.compag.2018.03.032
8. Chen, J., Chen, J., Zhang, D., Sun, Y., & Nanehkaran, Y. A. , "Using deep transfer learning for image-based plant disease identification", Computers and Electronics in Agriculture, 173, 105393, 2020. https://doi.org/10.1016/j.compag.2020.105393
9. Jia, S., Wang, G., Li, H., Liu, Y., Shi, L., & Yang, S., "ConvTransNet-S: A CNN-Transformer hybrid disease recognition model for complex field environments", *Plants*, 14(15), 2252, 2025. https://doi.org/10.3390/plants14152252
10. Pal, C., Karmakar, S., Mukherjee, I., & Chakrabarti, P. P. , "A lightweight and explainable CNN model for empowering plant disease diagnosis", *Scientific Reports*, 15, 30720, 2025. https://doi.org/10.1038/s41598-025-94083-1