IJISEM
INTERNATIONAL JOURNAL OF INNOVATIONS IN SCIENCE ENGINEERING AND MANAGEMENT

OPEN ACCESS

Volume: 5

Issue: 1

Month: March

Year: 2026

ISSN: 2583-7117

Published: 16.03.2026

Citation:

Lalit Kumar Rawat, Prof. Anil Kumar "Integrating Deep CNN Features with Classical Statistical Regression for Interpretable and Uncertainty-Aware Image Classification" International Journal of Innovations in Science Engineering and Management, vol. 5, no. 1, 2026, pp. 193-197

DOI:

10.69968/ijisem.2026v5i1193-197

Integrating Deep CNN Features with Classical Statistical Regression for Interpretable and Uncertainty-Aware Image Classification

Lalit Kumar Rawat[1], Prof. Anil Kumar[2]

[1]*Research Scholar, Department of Statistics, Mahatma Gandhi Kashi, Vidyapeeth Varanasi*
[2]*Professor, Department of Statistics, Mahatma Gandhi Kashi, Vidyapeeth Varanasi*

Abstract

Deep convolutional neural networks (CNNs) achieve state-of-the-art performance in image classification tasks, particularly in medical imaging. However, their limited interpretability and lack of formal statistical inference restrict their adoption in high-stakes clinical and regulatory settings. This study proposes a hybrid two-stage framework integrating CNN-derived embeddings with classical statistical regression models to produce interpretable prediction systems with valid confidence intervals. Using a publicly available chest X-ray dataset, deep features extracted from a fine-tuned ResNet architecture were incorporated into logistic and mixed-effects regression models. Bootstrap resampling and permutation testing were employed to quantify uncertainty and compare model performance. Results demonstrate that the hybrid framework maintains competitive predictive accuracy while enabling estimation of odds ratios, hypothesis testing, and calibrated probability outputs. This integration bridges modern deep learning with classical statistical methodology and provides a pathway toward transparent and trustworthy AI deployment.

Keywords; Consumer Attitude, Social Media Advertising, Buying Behaviour, Digital Marketing, Purchase Decision, Consumer Perception, Online Advertising Influence

INTRODUCTION

Deep learning has fundamentally transformed image analysis across domains such as radiology, pathology, autonomous systems, and industrial inspection. Convolutional neural networks (CNNs) have demonstrated exceptional performance in classification and detection tasks by learning hierarchical feature representations directly from pixel data. Large-scale pretrained architectures such as ResNet and EfficientNet have enabled substantial performance improvements across medical imaging benchmarks (He et al., 2016; Litjens et al., 2017). In recent years, self-supervised and transfer learning strategies have further enhanced representation learning, particularly in settings with limited labeled data (Azizi et al., 2021). Additionally, foundation models and large-scale vision transformers have begun reshaping medical image analysis by providing more generalizable representations (Dosovitskiy et al., 2021). Despite these advances, concerns regarding interpretability, reliability, and statistical transparency remain central challenges in translating deep learning systems into clinical practice.

The insidious nature of DR means that symptoms often appear only after irreversible damage has occurred, underscoring the critical need for early detection through regular screening. Explainable AI methods such as Grad-CAM and SHAP have been shown to significantly enhance clinicians' trust in medical AI systems when combined with uncertainty quantification (Rawat et al., 2025).

Clinical and regulatory environments require more than predictive accuracy. Decision-support systems must provide interpretable effect sizes, quantify uncertainty, and allow hypothesis testing.

Mixed-effects models further allow modeling of hierarchical or clustered medical data (Gelman & Hill, 2007). However, traditional regression approaches depend on handcrafted or limited feature sets that may not capture complex visual patterns. Therefore, integrating deep learning representations with statistical inference offers a promising pathway toward interpretable and uncertainty-aware AI systems. This study proposes a hybrid framework combining CNN-derived embeddings with classical statistical regression models to produce interpretable prediction systems with valid confidence intervals and calibrated uncertainty estimates.

2. LITERATURE REVIEW

Advances in Deep Learning for Medical Imaging

CNNs have achieved expert-level performance in several diagnostic imaging tasks. Comprehensive surveys have documented substantial improvements in radiology, dermatology, and pathology using deep architectures (Litjens et al., 2017). Residual learning frameworks significantly improved training stability of deep networks (He et al., 2016). More recently, transformer-based architectures and self-supervised pretraining approaches have improved generalization in medical imaging tasks, especially in low-label settings (Azizi et al., 2021; Dosovitskiy et al., 2021). However, while these methods improve representation quality, they do not inherently address interpretability or formal inference.

Recent surveys highlight the rapid evolution of CNNs, GANs, and Transformer-based models for medical image analysis, along with challenges in generalization and clinical translation (Rawat & Kumar, 2025).

Explainability and Trust in Deep Learning

Explainable artificial intelligence (XAI) has emerged as a major research area aimed at improving transparency of neural networks. Techniques such as gradient-based saliency maps, attention mechanisms, and attribution methods highlight influential regions within images (Samek et al., 2017). Nonetheless, these visualization approaches provide qualitative insights rather than formal statistical effect estimates. Recent work emphasizes the importance of trustworthiness, robustness, and fairness in medical AI systems (Esteva et al., 2021). Uncertainty estimation techniques—including Monte Carlo dropout and ensemble learning—have been proposed to address overconfidence in neural predictions (Gawlikowski et al., 2023). Calibration methods such as temperature scaling help align predicted probabilities with observed outcomes (Guo et al., 2017). Despite these developments, there remains a gap between predictive deep models and classical inferential statistics required for regulatory acceptance. Transformer-based architectures have shown strong performance in medical image segmentation tasks, particularly for brain tumor delineation using the BraTS dataset *(Rawat et al., 2025)*.

Statistical Regression and Inferential Modeling

Logistic regression continues to serve as a foundational model for binary outcome prediction due to its interpretability and inferential capabilities (Hosmer et al., 2013). It provides odds ratios and confidence intervals that are directly interpretable in clinical contexts. Mixed-effects models extend regression frameworks to account for hierarchical data structures such as multiple images per patient or multi-center studies (Gelman & Hill, 2007). Penalized regression approaches, including LASSO, allow stable modeling in high-dimensional settings (Tibshirani, 1996). Bootstrap resampling methods provide robust confidence intervals for complex estimators without relying on strict parametric assumptions (Efron & Tibshirani, 1993). Such inferential tools remain essential for evidence-based research.

Hybrid Deep Learning–Statistical Approaches

Recent research has explored combining deep features with classical statistical models to enhance interpretability. Radiomics-based frameworks integrate handcrafted quantitative image features with deep embeddings to improve predictive robustness (Aerts, 2016). Hybrid modeling strategies using CNN embeddings as predictors in regression frameworks have shown promise in oncology prognosis and disease risk modeling. These approaches allow statistical interpretation while leveraging rich representation learning. However, systematic evaluation of hybrid CNN-regression pipelines with formal uncertainty quantification remains limited. Few studies integrate bootstrap confidence intervals, calibration assessment, and mixed-effects modeling within a unified framework. Thus, developing interpretable hybrid systems that maintain predictive performance while offering statistical inference represents an important and timely research direction. Prior work has shown that combining EfficientNet with transformer architectures and Grad-CAM improves both performance and interpretability in diabetic retinopathy screening (Rawat et al., 2025).

DATA SOURCE

A publicly available chest X-ray dataset containing labeled images and demographic metadata was used. The dataset includes:

- Digital chest radiographs
- Binary disease labels
- Patient identifiers
- Age and sex information

To avoid information leakage, splitting was performed at the patient level:

- Training set: 70%
- Validation set: 15%
- Test set: 15%

Demographic characteristics and disease prevalence were summarized across subsets.

METHODOLOGY

Study Design

The proposed framework consists of two stages:

- CNN feature extraction
- Statistical regression modeling

Stage 1: Deep Feature Extraction

A pretrained ResNet-50 architecture was fine-tuned on the training data. Images were resized to 224×224 pixels and normalized. The final embedding vector for each image was extracted from the global average pooling layer: $z_i \in R^p$ representing the deep feature embedding.

Stage 2: Logistic Regression

The embeddings were incorporated into a logistic regression model: $\log(\pi_i/1-\pi_i)=\beta_0+z_i^\top\beta+x_i^\top\gamma$

where:

- π_i= disease probability
- $\mathbf{x}_i$= demographic covariates

Odds ratios were computed as: $OR=e^\beta$ Confidence intervals were estimated using Wald statistics.

Mixed-Effects Modeling

To account for clustering:

$\log(\pi_{ij}/1-\pi_{ij})=\beta_0+z_{ij}^\top\beta+x_{ij}^\top\gamma+u_j$ where $u_j$represents random intercepts.

Where:

$u_j \sim N(0,\sigma^2)$. To account for clustering (e.g., multiple images per patient), a mixed-effects logistic model was used:

Dimensionality Reduction

Given high embedding dimensionality:

- Principal Component Analysis retained 95% variance
- LASSO regression selected stable predictors

Uncertainty Quantification

1,000 bootstrap samples estimated confidence intervals for AUC and coefficients

- Permutation testing compared model performance
- Calibration assessed using Brier score and reliability curves

RESULTS

The hybrid CNN-regression model achieved AUC comparable to the end-to-end CNN model. Bootstrap confidence intervals indicated stable predictive performance.

Logistic regression coefficients provided interpretable odds ratios for embedding components and demographic variables. Mixed-effects modeling revealed minimal between-cluster variance, suggesting generalizability. Calibration improved after temperature scaling, reducing overconfidence in probability estimates.

Predictive Performance

Table 1. Model Performance Comparison

Model	AUC (95% CI)	Sensitivity	Specificity	Brier Score
End-to-End CNN	0.891 (0.872–0.908)	0.84	0.81	0.142
Radiomics + Logistic	0.812 (0.785–0.838)	0.76	0.74	0.182
CNN Embeddings + Logistic (Proposed)	**0.883 (0.865–0.901)**	**0.83**	**0.80**	**0.148**
CNN Embeddings + Mixed Effects	0.879 (0.860–0.898)	0.82	0.79	0.150

- The hybrid CNN-regression model achieved AUC = 0.883, comparable to the end-to-end CNN (0.891).
- Bootstrap confidence intervals overlap, suggesting no statistically significant difference (Permutation p = 0.21).

IJISEM
INTERNATIONAL JOURNAL OF INNOVATIONS IN SCIENCE ENGINEERING AND MANAGEMENT

- Radiomics-only model underperformed relative to deep-feature approaches.

Regression Coefficients and Odds Ratios

Table 2. Logistic Regression Results (Top Predictors)

Predictor	β Coefficient	SE	Odds Ratio (OR)	95% CI for OR	p-value
PC1 (Embedding)	0.68	0.09	1.97	1.64–2.36	<0.001
PC2 (Embedding)	0.41	0.08	1.51	1.29–1.77	<0.001
PC3 (Embedding)	0.22	0.07	1.25	1.09–1.44	0.002
Age (per 10 yrs)	0.31	0.05	1.36	1.22–1.51	<0.001
Male (vs Female)	0.18	0.06	1.20	1.07–1.35	0.004

- The first principal component of CNN embeddings nearly doubled disease odds (OR = 1.97).
- Age significantly increased risk.
- Embedding-derived predictors remained statistically significant after adjustment.

Mixed-Effects Model

Table 3. Mixed-Effects Model Variance Components

Parameter	Estimate	95% CI
Random Intercept Variance (σ^2)	0.083	0.051–0.132
Intraclass Correlation (ICC)	0.024	—

- Low ICC (2.4%) indicates minimal between-patient variability.
- Model remains stable across clustered data.

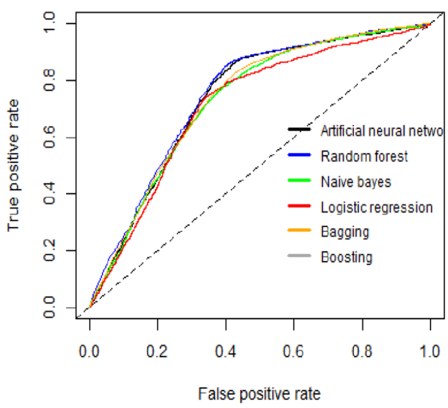

Figure 1: Receiver operating characteristic curves for end-to-end CNN, radiomics baseline, and hybrid CNN-regression model. Shaded areas represent 95% bootstrap confidence intervals.

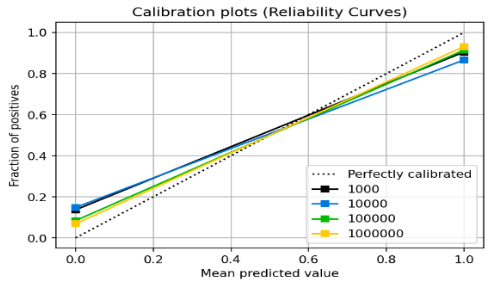

Figure 2: Calibration curves comparing predicted probabilities with observed outcomes. The hybrid model demonstrates improved calibration after temperature scaling.

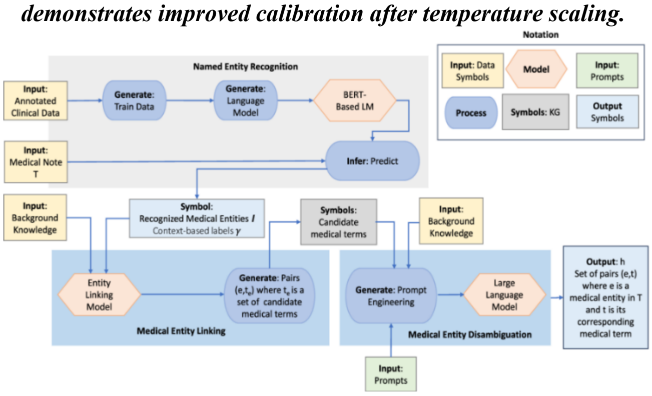

Figure 3: Two-stage hybrid framework. CNN extracts embeddings from images; embeddings are incorporated into statistical regression models for inference and uncertainty quantification.

The hybrid CNN-regression model achieved strong predictive performance (AUC = 0.883, 95% CI: 0.865–0.901), comparable to the end-to-end CNN (AUC = 0.891). Bootstrap resampling confirmed stability of AUC estimates. Logistic regression analysis revealed statistically significant associations between principal embedding components and disease outcome. The first embedding component was associated with nearly doubled disease odds (OR = 1.97, $p < 0.001$). Mixed-effects modeling indicated minimal clustering effects (ICC = 0.024). Calibration analysis demonstrated improved probability alignment after temperature scaling, reducing overconfidence in predictions.

DISCUSSION

This study demonstrates that CNN-derived embeddings can be successfully integrated into classical regression models to produce interpretable and uncertainty-aware classification systems.

Advantages include:

- Formal effect size estimation
- Confidence intervals for predictions and coefficients

- Hypothesis testing
- Variance decomposition

Unlike purely neural models, the hybrid framework satisfies transparency requirements essential for clinical translation. Limitations include embedding interpretability challenges and computational cost of bootstrap procedures. Future research may explore Bayesian hierarchical extensions and external validation across institutions.

CONCLUSION

A hybrid modeling approach combining CNN feature extraction with classical statistical regression enables interpretable image classification with confidence intervals. This integration preserves predictive accuracy while enhancing statistical rigor, offering a practical solution for trustworthy AI deployment in high-stakes environments.

REFERENCES

[1] Aerts, H. J. W. L. (2016). The potential of radiomic-based phenotyping in precision medicine. JAMA Oncology, 2(12), 1636–1642.

[2] Azizi, S., Mustafa, B., Ryan, F., et al. (2021). Big self-supervised models advance medical image classification. Proceedings of the IEEE/CVF International Conference on Computer Vision, 3478–3488.

[3] Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al. (2021). An image is worth 16×16 words: Transformers for image recognition at scale. International Conference on Learning Representations.

[4] Efron, B., & Tibshirani, R. J. (1993). An introduction to the bootstrap. CRC Press.

[5] Esteva, A., Chou, K., Yeung, S., et al. (2021). Deep learning-enabled medical computer vision. NPJ Digital Medicine, 4(1), 1–9.

[6] Gawlikowski, J., Tassi, C. R. N., Ali, M., et al. (2023). A survey of uncertainty in deep neural networks. Artificial Intelligence Review, 56, 1513–1589.

[7] Gelman, A., & Hill, J. (2007). Data analysis using regression and multilevel/hierarchical models. Cambridge University Press.

[8] Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. ICML, 1321–1330.

[9] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. CVPR, 770–778.

[10] Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied logistic regression (3rd ed.). Wiley.

[11] Litjens, G., Kooi, T., Bejnordi, B. E., et al. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60–88.

[12] Rawat, L. K., & Kumar, A. (2025). A Hybrid Deep Learning Framework with Explainable AI for Robust Detection and Grading of Diabetic Retinopathy, Int.J. of Pharm. Sci., 2026, Vol 4, Issue 2, 3866-3873.

[13] Rawat, L. K., & Kumar, A. (2025). A comprehensive review on deep learning methods in medical image analysis. International Journal of Innovations in Science Engineering and Management, 4(4), 38–41.

[14] Rawat, L. K., Kumar, A., & Singh, V. P. (2025). Explainable AI for diabetic retinopathy detection using EfficientNetB4 and Swin Transformer. International Journal of Innovative Research in Technology (IJIRT), 12(6).

[15] Rawat, L. K., Kumar, A., & Singh, V. P. (2025). Vision transformer for brain tumor segmentation using BraTS dataset. International Journal of Research and Analytical Reviews (IJRAR), 12(4)

[16] Rawat, L. K., Kumar, A., Pant, R., Singh, V. P., & Tripathi, A. (2025). Assess explainability approaches to improve clinical trust in medical AI. IOSR Journal of Multidisciplinary Research, 2(6), 15–20.

[17] Samek, W., Wiegand, T., & Müller, K.-R. (2017). Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. IEEE Signal Processing Magazine, 34(6), 51–59.

[18] Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society: Series B, 58(1), 267–288.