© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722241

Hierarchical Prototype Network for Interpretable Chest X-ray Disease Classification

AKSHAYA SREEKUMAR[1], SANKALP KUMAR SINGH[2], DR. SIBI AMARAN[3], R. SHOBANA[4]

[1,2]Student, Department of Computing Technologies, School of Computing, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India

[3]Assistant Professor, Department of Computing Technologies, School of Computing, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India

[4]Associate Professor, Department of CSE, Aarupadai Veedu Institute of Technology,Vinayaka, Mission's Research, Foundation (DU)

Abstract- In recent years, deep learning has made great strides in medical image categorization, especially for automated chest X-ray image analysis for illness screening and early detection. Even though they are quite good at producing predictions, most deep learning-based models behave as black-box decision-making systems that don't give much information about how an input is classified. This could make clinician trust and interpretability weaker. To address this challenge, this research presents a Hierarchical Prototype network (HierProtoPnet) for interpretable chest X-ray classification of various images which are provided.nThe suggested solution uses a ResNet50 backbone for strong feature learning and two levels of prototype learning. This lets it learn both fine-grained and coarse-grained visual patterns in medical images at the same time. Lower-level prototypes learn more detailed local properties, such textures and edge patterns. Higher-level prototypes, on the other hand, encode more abstract disease aspects. Because of its hierarchical structure, the model classifies based on learnt prototypes, which gives clear predictions.

The suggested method uses a ResNet50 backbone for strong feature extraction and has two levels of prototype learning. This allows it to learn both little and large visual patterns in medical images at the same time. Prototypes at a lower level learn more about the finer points of a place, like textures and edge patterns. Higher-level prototypes represent broader and more general disease characteristics. Because of this hierarchical structure, the model makes predictions by comparing input images with learned prototypes. This makes the reasoning behind each prediction clearer and easier to understand. The prototype-based approach also improves interpretability by showing how different visual patterns influence the final decision. The results indicate that hierarchical prototype learning supports diagnosis while improving both model accuracy and clarity, strengthening its suitability for clinical decision support systems. This study highlights the potential of combining deep convolutional feature extraction with hierarchical interpretability mechanisms to create accurate and informative models for medical image classification.

Keywords— Chest X-ray, Deep Learning, Hierarchical Prototype Network, ProtoPNet, ResNet50, Medical Image Classification, Explainable Artificial Intelligence (XAI)

I. INTRODUCTION

Medical imaging is an important feature of modern healthcare since it lets doctors detect and keep an eye on a lot of different illnesses. One of the most popular imaging tests is a chest X-ray. It is cheap, easy to get, and good at diagnosing lung diseases including pneumonia and respiratory infections [4], [5]. To get the necessary medical attention immediately, it's crucial to read chest X-ray images correctly. Radiologists, on the other hand, have to look at the pictures by hand, which takes a long time and might be affected by weariness and differences between observers, especially when there are a lot of medical photos to look at [6].Deep learning (DL) algorithms have done quite well at tasks like evaluating and categorizing medical pictures in the last several years. Convolutional Neural Networks (CNNs), particularly designs such as ResNet, have demonstrated the capability to autonomously discern distinctions in complicated medical images [1], [12]. These models can recognize tiny visual patterns that the human eye might not be able to see. Most deep learning models don't show you how they make judgments, even though they are incredibly accurate. This is because they are black-box systems. They are not employed in clinical practice because they are hard to understand, and trust, reliability, and explainability are very important [7].To overcome this constraint, interpretable deep learning models have been created to provide substantial justifications for classification decisions. One such approach is the Prototype Propagation Network . It puts images into groups by comparing them to a set of learnt prototypes that show how each class usually looks [2]. ProtoPNet helps individuals comprehend by indicating which aspects of an image are similar to the taught prototypes. ProtoPNet makes things easier to understand, but it only employs one level of prototype representation, therefore it might not fully demonstrate how visual patterns in medical images are organized into levels [10].Chest X-ray pictures show both little things, like local textures and small areas of low opacity, and big things, like the overall shape of the lungs and big areas that aren't typical. It may be insufficient to represent both categories of information using a singular tier of a prototype design. We need a hierarchical learning system that can take in both low-level and high-level visual representations at the same time [7] and [16].

This research proposes a Hierarchical Prototype Network (HierProtoPNet) for the straightforward classification of chest X-ray images. The proposed approach integrates a ResNet50 backbone for robust feature extraction [1] with a dual-level prototype learning framework that encompasses both fine-level and coarse-level prototypes. Fine-level prototypes learn about specific local visual cues, while coarse-level prototypes learn about more general and abstract disease patterns that derive from fine-level similarities. Because of its hierarchical structure, the model is able to reason at multiple levels of abstraction, which improves both its transparency and its classification performance.To assess the effectiveness of the proposed approach, HierProtoPNet is compared with two baseline models: a standard ResNet50 classifier and a single-level ProtoPNet architecture [2]. The experimental results show that HierProtoPNet achieves better classification performance while producing predictions that can be clearly understood through prototype-based reasoning.

The results show that hierarchical prototype learning makes diagnostic performance and explainability better, making the suggested model better for clinical decision support systems.

The main contributions of this work are summarized as follows:

- A hierarchical, prototype-based deep learning framework is introduced for structured and effective chest X-ray classification.
- Fine-level and coarse-level prototypes are used to capture both detailed local features and broader global patterns in medical images.
- ResNet50 and ProtoPNet are used as baseline models for comparative evaluation.
- The proposed HierProtoPNet model improves classification accuracy while maintaining clear interpretability.

II. RELATED WORK

A number of people have successfully utilized deep learning to examine and categorize medical pictures. Convolutional Neural Networks (CNNs) have proved to be very successful in extracting unique features from complicated medical pictures. He et al. in [1] designed a picture classification system known as ResNet, which allowed for the training of deep neural networks with the aid of residual connections.This made picture classification much more accurate. Numerous medical imaging apps have effectively utilized this design [12].

Several studies have focused on the categorization of chest X-ray images using deep CNN models. Wang et al. in their research [4], proposed a comprehensive chest X-ray data set and assessed the application of deep learning models for the diagnosis of thoracic diseases.CheXNet was made by Rajpurkar et al. [5] and can find pneumonia as well as a radiologist. Litjens et al. [6] examined deep learning techniques for the analysis of medical pictures and determined that CNN-based models were the most effective.

Most CNN-based methods are like black boxes that can't be comprehended, which makes them impossible to utilize in clinical settings, even when they work. Researchers have developed explainable

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722241

artificial intelligence (XAI) solutions to address this issue. Selvaraju et al. [9] developed Grad-CAM to identify significant regions influencing classification outcomes, whereas Bach et al. [8] introduced Layer-Wise Relevance Propagation (LRP) for pixel-level elucidations. These solutions, on the other hand, simply explain things after they happen and don't change the reality that CNN models are black boxes.
The purpose of prototype-based neural networks is to help people understand how they work. Chen et al. [2] proposed the use of ProtoPNet, which differentiates images on the basis of their comparison with learned prototypes that contain unique visual characteristics for the classes. Zhou et al. [10] demonstrated that prototype-based models can achieve accuracy comparable to other models while providing valuable explanations in medical picture diagnosis. Other studies coupled CNNs with easier-to-understand approaches to make chest X-ray classification clearer [12], [16].

But most prototype-based methods only use one level of prototype representation, which might not be able to completely show how medical images are put together. Chest X-ray images have both little, detailed details and large, general patterns. You could need more than one level of a prototype structure to show both types of information. Although hierarchical feature learning has been examined in numerous vision tasks, there exists a paucity of research on hierarchical prototype learning for interpretable medical picture categorization.

This restriction leads to the development of the Hierarchical Prototype Network (HierProtoPNet), which integrates fine-level and coarse-level prototype layers to depict both local and global visual patterns. By combining deep feature extraction with hierarchical prototype reasoning, the proposed method seeks to increase the accuracy of classification while making medical picture diagnosis simple and easy to grasp.

III. METHODOLOGY

There are a number of steps proposed in the suggested framework based on HierProtoPNet, which occur sequentially as depicted in Fig. 1. All of these steps are extremely important for generating correct and understandable results.

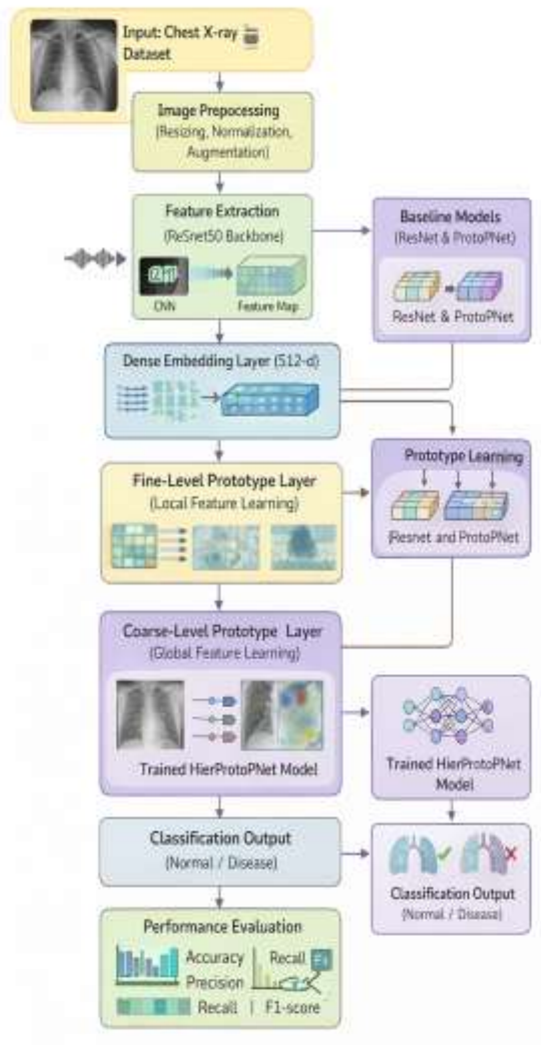


Fig. 1. Workflow of proposed HierProtoPNet model

A. Input Dataset

The technique works using a collection of chest X-ray pictures that have been marked as either sick or normal. You use these photographs to train and test the model that classifies things.

Let the chest X-ray image be represented as:

$$I \in \mathbb{R}^{H \times W \times C}$$

where H, W, and Cdenote the height, width, and number of channels of the image.

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722241

B. Image Preprocessing
To make the input images bigger, more even, and better, data augmentation techniques are used, and the images are rotated and flipped. This step also refines the image and the model as a whole.

All input images are resized and normalized:

$$I' = \text{Normalize}(\text{Resize}(I))$$

Data augmentation is applied to improve generalization:

$$\hat{I} = T(I')$$

where $T(\cdot)$ represents transformations such as rotation, flipping, zooming, and contrast adjustment.

Feature Extraction using ResNet50
The convolutional neural network, ResNet50, processes the images and transforms them into deep features. The features allow you to view the most important patterns in the chest X-ray images.

ResNet50 transforms the preprocessed image into deep features:

$$F = f_{ResNet}(\hat{I})$$

where Fdenotes the feature representation learned by the CNN.

D. Dense Embedding Layer
A tiny, dense embedding vector is made from the features that were stripped out. This representation is then utilized to compare learnt prototypes in the future steps.

The extracted feature map is converted into a dense embedding vector:

$$Z = \phi(W_1 F + b_1)$$

where:

- W_1 and b_1 are trainable parameters,
- $\phi(\cdot)$ is the ReLU activation function.

Fine-Level Prototype Learning
Fine-level prototypes learn detailed and local visual patterns such as textures and small abnormal regions. Similarity between the embedding and fine-level prototypes is computed to capture local evidence for classification.

Let the set of fine-level prototypes be:

$$P^f = \{p_1^f, p_2^f, \ldots, p_m^f\}$$

The distance between embedding Zand each fine prototype is computed as:

$$d_i^f = \| Z - p_i^f \|_2, i = 1,2,\ldots,m$$

These distances capture similarity to local and detailed visual patterns.

The fine-level feature vector is obtained as:

$$F_f = \phi(W_2 d^f + b_2)$$

F. Coarse-Level Prototype Learning
Coarse-level prototypes are those that show higher-level and global visual patterns, such as the shape of the lungs and large areas that are not typical.These prototypes are based on answers to fine-level prototypes.

The set of coarse-level prototypes is defined as:

$$P^c = \{p_1^c, p_2^c, \ldots, p_n^c\}$$

Distances to coarse-level prototypes are computed:

$$d_j^c = \| F_f - p_j^c \|_2, j = 1,2,\ldots,n$$

These prototypes represent global and abstract disease-related patterns.

G. Trained HierProtoPNet Model
The fine-level and coarse-level prototype layers together form the trained HierProtoPNet model. This model combines deep feature extraction with hierarchical prototype reasoning.

Training is performed in two phases:

Phase 1 – Prototype Training (Backbone Frozen):

$$\theta_{ResNet} = constant$$
$$\theta_{proto} \leftarrow \theta_{proto} - \eta \nabla L$$

Phase 2 – Fine-Tuning (Backbone Unfrozen):

$$\theta^{*}_{ResNet} \leftarrow \theta_{ResNet} - \eta \nabla L$$

H. Classification Output

The final classification result is produced by the trained model, and it shows whether the image of the chest X-ray is normal or sick. Predictions are based on how well something fits with what we've learnt before.

The final classification output is obtained using a sigmoid activation function:

$$\hat{y} = \sigma(W_3 d^c + b_3)$$

where:

- $\hat{y}$ represents the predicted class (Normal / Disease),
- $\sigma(\cdot)$ is the sigmoid function.

Loss Function

The model is trained using Binary Cross-Entropy loss; prototype diversity is analyzed/defined but not included in the final optimization objective.

Binary Cross-Entropy Loss

$$L_{cls} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)]$$

Prototype Diversity Loss

$$L_{proto} = \frac{1}{K}\sum_{i \neq j} cosine\,(p_i, p_j)$$

Total Loss

$$L = L_{cls}$$

J. Performance Evaluation

Model The performance of the model can be evaluated by calculating the accuracy, precision, recall, and F1 score of the proposed classification framework.

The performance of the trained model can be evaluated by calculating the accuracy, precision, recall, and F1 score of the proposed classification framework:

- Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- Precision

$$Precision = \frac{TP}{TP + FP}$$

- Recall

$$Recall = \frac{TP}{TP + FN}$$

- F1-score

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

IV. RESULTS AND DISCUSSION

This section demonstrates the experimental evaluation of the proposed HierProtoPNet model and how it performs compared to other models like ResNet50 and ProtoPNet. The performance of the models is evaluated based on how well it works with the help of performance measures like accuracy, precision, recall, and F1-score.

A. Performance Comparison

Table I shows how well the three models did in classifying things.

Table I: Performance Comparison of Models

Model	Accuracy (%)
ResNet50	62.5
ProtoPNet	76.44
HierProtoPNet (Proposed)	81.25

It can be shown from Table I that the standard ResNet50 model has an accuracy of 62.50%. ResNet50 is good at extracting deep features, however it works like a black box and is hard to understand.

With the help of the prototype-based reasoning, the accuracy of the ProtoPNet model improves to 76.44%. This proves the point that classification improves by learning the representative visual prototypes and makes the process easier to understand.

The accuracy of the proposed HierProtoPNet model reaches the highest accuracy of 81.25%, which is higher than the ResNet50 and ProtoPNet models. This is because of the proposed learning mechanism of the hierarchical prototypes, which detects fine-level and coarse-level visual patterns from the images of the chest X-rays. Fine-level prototypes include small features like texture and small areas that do not seem right, while coarse-level prototypes include the overall lung structure and disease patterns.

B. Training and Validation Analysis
The training and validation accuracy curves for the proposed HierProtoPNet model are closer and closer with time. This indicates that the model is learning in a constant manner. In addition, the validation accuracy is relatively close to the training accuracy. This implies that the model performs well with new data and does not overfit much.

The loss curves for the model are showing a constant decrease, indicating that the classification layers and the hierarchical prototype layers are performing well.

Figure 2: Training vs Validation Accuracy (HierProtoPNet)

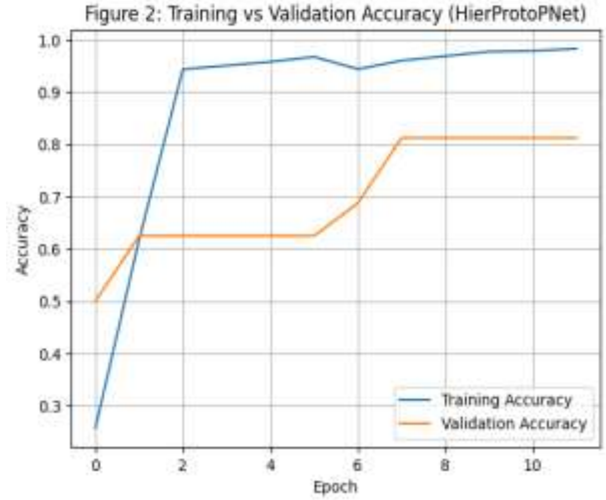
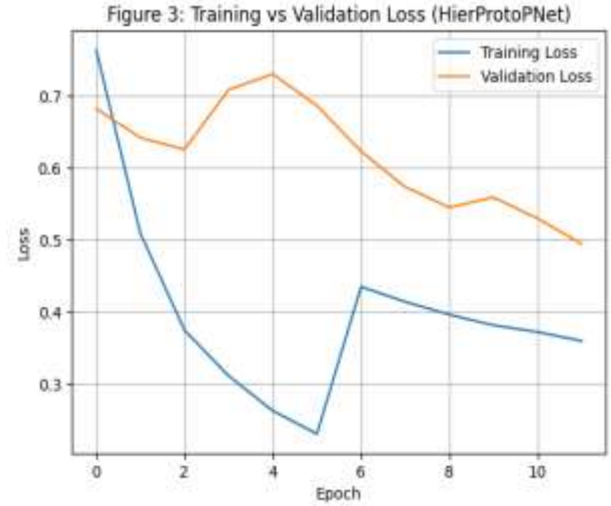
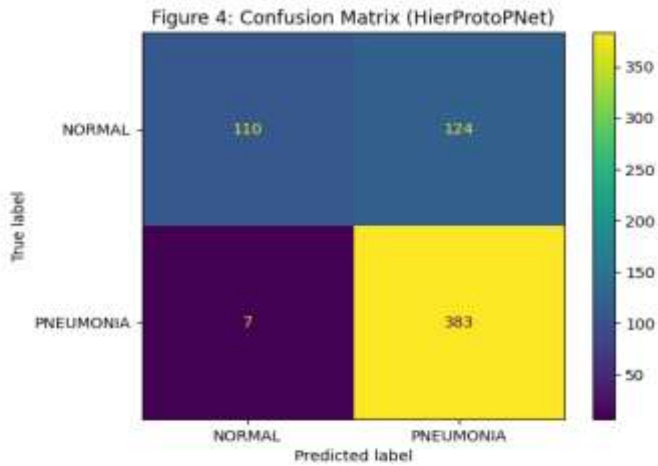


Figure 3: Training vs Validation Loss (HierProtoPNet)



C. Confusion Matrix and Metric Evaluation
We designed a confusion matrix for our proposed model, HierProtoPNet, to be able to look closer into its performance in classifying items. The confusion matrix shows how many normal and diseased cases are correctly or incorrectly classified.

The confusion matrix shows that our model has a high percentage of true positives and true negatives. This shows that our model can distinguish normal and pathological chest X-ray images well.

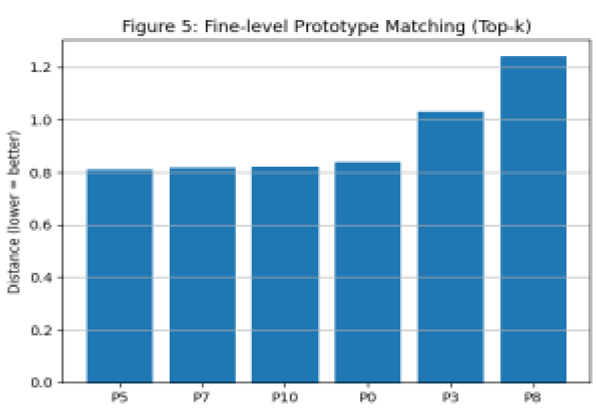


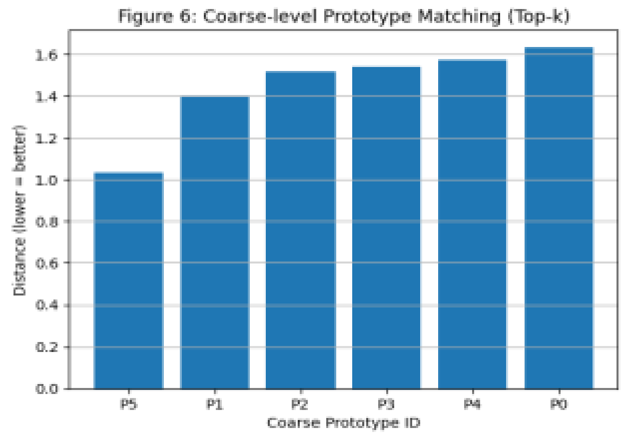
D. Interpretability and Prototype Visualization

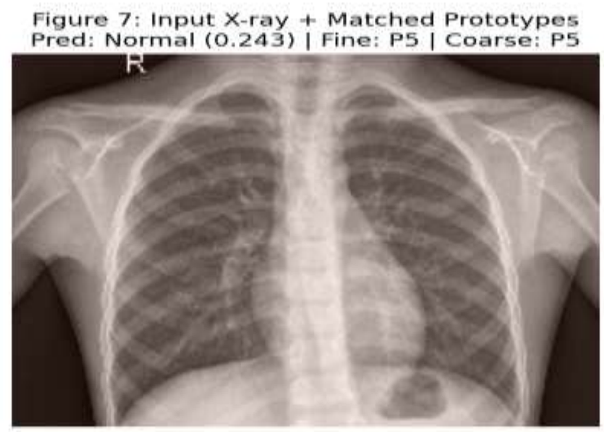
The proposed HierProtoPNet model is straightforward to understand, which is one of its best features. ResNet50 just tells you what class something is, while HierProtoPNet uses learned prototypes to explain its decisions.

The fine-level prototypes focus on small areas of interest, such as variations in texture and opacity patterns in the lungs. The coarse level prototypes show the large aberrant patches and the structure of the lungs. These are examples of global disease patterns. The explanations based on the prototypes help doctors understand the learned visual patterns that influenced the final judgment of the categorization.

The hierarchical explanation method makes things simpler and easier to accept the predictions made by the models. This is very important for the decision-support systems.







E. Comparative Discussion

The proposed HierProtoPNet is able to provide predictions that are more precise and interpretable compared to the predictions made by ResNet50. ResNet50 only has the ability to carry out feature extraction and classification. The proposed HierProtoPNet, on the other hand, is able to make decisions based on how similar something is to the trained prototypes.

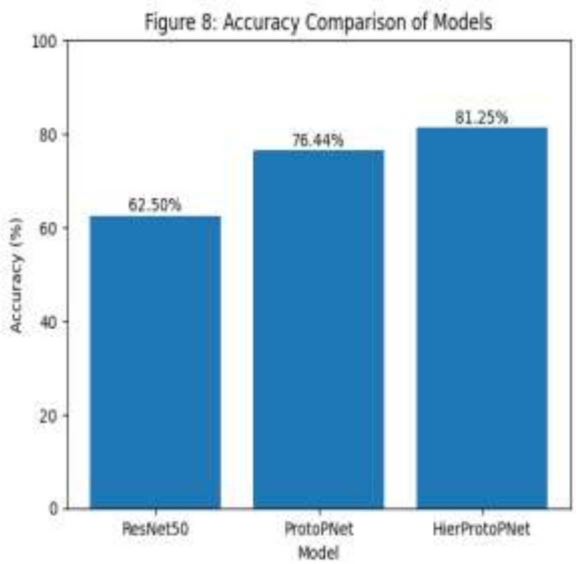
The proposed model is an improvement compared to the existing ProtoPNet model because it has the ability to incorporate hierarchical prototype layers. The existing ProtoPNet model only has the ability to incorporate a single level of representation for the prototypes. This is not able to capture all the information that is needed from the images, both local and global. The proposed HierProtoPNet solves this problem by incorporating both fine level and coarse level prototype learning.

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722241

F. Overall Discussion

The experimental results demonstrate that hierarchical prototype learning facilitates the classification of chest X-rays while maintaining the model's interpretability. The model has the ability to learn discriminative and explicable representations through the use of ResNet50 in deep feature learning and the application of two-level prototype reasoning.

The proposed HierProtoPNet model has tremendous potential in the creation of a clinical decision support system. The model makes accurate predictions and has the ability to provide explanations to the predictions made.



V. CONCLUSION

The study proposed a Hierarchical Prototype Network, or HierProtoPNet, for interpretable chest X-ray image categorization, incorporating a ResNet50 backbone and fine-level and coarse-level prototype learning. The hierarchical structure allows the model to perceive local and global visual patterns and provides interpretable explanations for its categorization decisions. Based on the experimental results, the proposed HierProtoPNet outperforms regular ResNet50 and ProtoPNet models for categorizing objects. The proposed method is suitable for image analysis and decision support in medical images, as it performs well and is interpretable. In future, there are prospects for expanding the model to support multi-class disease categorization, working with larger datasets, and improving prototype visualization tools.

REFERENCES

[1] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778, 2016.

[2] C. Chen, O. Li, D. Tao, A. Barnett, C. Su, and C. Rudin, “This Looks Like That: Deep Learning for Interpretable Image Recognition,” Advances in Neural Information Processing Systems (NeurIPS), vol. 32, pp. 8930–8941, 2019.

[3] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 234–241, 2015.

[4] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. Summers, “ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases,” IEEE CVPR, pp. 2097–2106, 2017.

[5] A. Rajpurkar et al., “CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning,” arXiv preprint arXiv:1711.05225, 2017.

[6] G. Litjens et al., “A Survey on Deep Learning in Medical Image Analysis,” Medical Image Analysis, vol. 42, pp. 60–88, 2017.

[7] R. Tjoa and C. Guan, “A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI,” IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 11, pp. 4793–4813, 2021.

[8] S. Bach et al., “On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation,” PLOS ONE, vol. 10, no. 7, 2015.

[9] R. R. Selvaraju et al., “Grad-CAM: Visual Explanations from Deep Networks via

Gradient-Based Localization," IEEE International Conference on Computer Vision (ICCV), pp. 618–626, 2017.

[10] Y. Zhou et al., "Prototype-Based Interpretable Deep Learning for Medical Image Diagnosis," IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 9, pp. 3423–3433, 2021.

[11] S. Lundberg and S. Lee, "A Unified Approach to Interpreting Model Predictions," Advances in Neural Information Processing Systems (NeurIPS), pp. 4765–4774, 2017.

[12] J. Shin et al., "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning," IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1285–1298, 2016.

[13] P. Mooney, Chest X-ray Pneumonia Dataset, Kaggle, 2018. [Online]. Available: https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia

[14] A. Esteva et al., "Dermatologist-level Classification of Skin Cancer with Deep Neural Networks," Nature, vol. 542, pp. 115–118, 2017.

[15] T. A. Al-Antari et al., "Fast Deep Learning Computer-Aided Diagnosis of COVID-19 Based on Digital Chest X-ray Images," Applied Intelligence, vol. 51, pp. 2890–2907, 2021.

[16] J. G. Silva et al., "Explainable Deep Learning for Chest X-ray Image Classification," IEEE Access, vol. 8, pp. 180034–180045, 2020.

[17] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, no. 7639, pp. 115–118, 2017.

[18] G. Litjens et al., "A survey on deep learning in medical image analysis," Medical Image Analysis, vol. 42, pp. 60–88, 2017.

[19] H. R. Roth et al., "Deep learning and its application to medical image segmentation," Medical Image Analysis, vol. 45, pp. 1–19, 2018.

[20] J. Irvin et al., "CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison," AAAI Conference on Artificial Intelligence, 2019.

[21] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," KDD, 2016.

[22] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," NeurIPS, 2017.

[23] B. Kim, M. Wattenberg, J. Gilmer, et al., "Interpretability beyond feature attribution: Quantitative testing with concept activation vectors (TCAV)," ICML, 2018.

[24] Y. Zhou, Y. Li, and W. Li, "Interpretable deep learning models for medical image diagnosis," IEEE Journal of Biomedical and Health Informatics, 2021.

[25] [D. G. T. Denil et al., "Learning where to attend with deep architectures for image tracking," Neural Computation, 2012.