JJEE

Volume 12 | Number 1 | March 2026 | Pages 54-73

Jordan Journal of Electrical Engineering

ISSN (print): 2409-9600, ISSN (online): 2409-9619

Homepage: jjee.ttu.edu.jo

Hybrid Decision Tree Models with CNN, HOG and Logistic Regression for Medical Image Classification

Aziz Ilyas Ozturk[1*]

[1] Esentepe Mh. Harman Sk. No.8
E-mail: ziz.ozturk@gehealthcare.com

Received: Oct 06, 2025 *Revised: Dec 06, 2025* *Accepted: Dec 22, 2025* *Available online: Mar 19, 2026*

***Abstract*—** This paper presents a comparative analysis of three hybrid Decision Tree–based models—DT-CNN, DT-HOG, and DT-LR—designed for medical image classification. The aim is to evaluate how different feature extraction strategies influence the performance and generalization ability of the Decision Tree classifier. The DT-CNN model achieved the highest validation accuracy (98%) due to its deep feature representation capability, while the DT-LR model obtained the highest test-set accuracy (96.3%), demonstrating superior generalization performance. In contrast, the DT-HOG model achieved an accuracy of 81.3%, reflecting the limitations of handcrafted features for subtle lesion characterization. The results indicate that although CNN-based features provide strong representation power, the Logistic Regression–Decision Tree hybrid exhibits more stable and reliable performance on unseen data. These findings highlight the importance of selecting appropriate feature extractors when designing hybrid machine-learning pipelines for medical image classification.

***Keywords*—** Medical imaging; Hybrid machine learning models; Cyst detection; Neuroinformatics.

1. INTRODUCTION

Convolutional Neural Networks (CNNs) have revolutionized the field of deep learning, particularly in image recognition and classification tasks. CNNs are a class of deep neural networks that utilize convolutional layers to automatically and adaptively learn spatial hierarchies of features from input images. The architecture of CNNs typically includes convolutional layers, pooling layers, and fully connected layers, which work together to extract and classify features from images [1]. The convolutional layers apply a set of filters to the input image, creating feature maps that capture various aspects of the image, such as edges, textures, and patterns. Pooling layers then reduce the spatial dimensions of these feature maps, which helps reduce computational complexity and prevents overfitting. Finally, fully connected layers interpret these features to make predictions [2].

CNNs have been extensively used in medical image analysis, where they have shown remarkable performance in tasks such as tumor detection, organ segmentation, and disease classification. The ability of CNNs to learn from large datasets and generalize well to new data makes them particularly suitable for medical applications [3]. Moreover, CNNs have been pivotal in advancing image classification tasks. They have achieved state-of-the-art results in various benchmarks, outperforming traditional machine learning methods. The hierarchical feature learning capability of CNNs allows them to capture complex patterns in images, leading to high accuracy in classification tasks [4]. Despite their success, CNNs face challenges such as the need for large, labeled datasets, high computational requirements, and difficulties

* Corresponding author
DOI: https://doi.org/10.5455/jjee.204-1759737614

in interpreting learned features. Future research is focused on addressing these challenges by developing more efficient architectures and improving interpretability [5].

Histograms of Oriented Gradients (HOG) is a feature descriptor used in computer vision and image processing for object detection. The HOG algorithm works by dividing an image into small, connected regions called cells and computing a histogram of gradient directions or edge orientations for each cell. The combination of these histograms represents the object's shape and appearance, making it suitable for detecting objects like humans in images. The HOG descriptor is particularly effective in human detection due to its ability to capture the local shape of objects while being invariant to geometric and photometric transformations. This robustness is achieved by normalizing the histograms across larger spatial regions called blocks, which helps in reducing the effects of illumination and shadowing [6].

One of the significant advancements in HOG-based object detection is its implementation on hardware platforms such as Field Programmable Gate Arrays (FPGAs). Demonstrated an FPGA implementation of a HOG-based object detection processor, which significantly enhances the processing speed and energy efficiency, making it suitable for real-time applications [7]. Moreover, the integration of HOG with other feature descriptors like Local Binary Patterns (LBP) has shown improved performance in handling partial occlusions during human detection [8] proposed HOG-LBP human detector effectively manages occlusions, enhancing the reliability of detection in complex environments. The efficiency of HOG in real-time applications is demonstrated through the development of energy-efficient object detection systems capable of operating at high resolutions and frame rates, such as 1080HD at 60 fps with multi-scale support. This highlights the practical applicability of HOG in high-resolution video processing [9]. Furthermore, fast human detection methods using a cascade of HOG descriptors have been introduced to accelerate the detection process by concentrating computational resources on the most promising regions of an image. This approach significantly reduces computational load while maintaining high detection accuracy [10]. Logistic regression is a widely used statistical method for modeling binary outcomes in various fields, particularly in medical research. It estimates the probability of a binary response based on one or more predictor variables. The logistic regression model is defined by the logistic function, which maps any real-valued number into the range between 0 and 1, making it suitable for predicting probabilities [11]. In clinical risk prediction, logistic regression is employed to identify and quantify the risk factors associated with specific health outcomes. This involves fitting a logistic regression model to the data, where the dependent variable is the occurrence of a clinical event (e.g., disease presence) and the independent variables are the risk factors (e.g., age, gender, biomarkers). The coefficients obtained from the model provide insights into the strength and direction of the association between risk factors and the clinical outcome [12].

Model validation is a crucial step in logistic regression analysis to ensure the reliability and generalizability of the model. Techniques such as cross-validation, bootstrapping, and the use of independent test sets are commonly employed to assess the model's performance. These methods help in evaluating the model's accuracy, sensitivity, specificity, and overall predictive power [13]. Logistic regression has also been integrated with other machine learning techniques to enhance prediction accuracy. For instance, neural networks combined with logistic regression have been used to predict student retention rates, demonstrating the versatility and applicability of logistic regression in various domains beyond healthcare [14].

Furthermore, comprehensive reviews of logistic regression techniques highlight their effectiveness in predicting health outcomes and trends. These reviews emphasize the importance of proper model specification, validation, and interpretation to ensure accurate and meaningful predictions [15].

In this paper, we explore hybrid models that combine Decision Trees with HOG, CNN, and Logistic Regression to enhance classification accuracy. The Decision Tree-HOG hybrid model achieved an accuracy of 0.81, the Decision Tree-CNN hybrid model achieved an accuracy of 0.98, and the Decision Tree-Logistic Regression hybrid model achieved an accuracy of 0.96. These results demonstrate the potential of hybrid models in improving classification performance by leveraging the strengths of different algorithms.

2. RELATED WORKS

The integration of hybrid models in machine learning has gained significant attention due to their ability to leverage the strengths of different algorithms, thereby enhancing predictive performance and robustness. This section reviews hybrid models involving Decision Trees combined with Histograms of Oriented Gradients (HOG), Convolutional Neural Networks (CNNs), and Logistic Regression.

Histograms of Oriented Gradients (HOG) is a feature descriptor widely used in object detection tasks. Combining Decision Trees with HOG descriptors has been explored to improve the accuracy and efficiency of detection systems. This integration enables hierarchical classification of features, enhancing the model's ability to distinguish between different objects in an image. Hardware implementations of HOG-based object detection on platforms like FPGAs have further demonstrated improvements in processing speed and energy efficiency, highlighting the practical applicability of such hybrid models in real-time scenarios. Additionally, hybrid approaches that incorporate complementary descriptors have shown improved robustness in complex environments, such as those involving partial occlusions [16].

Convolutional Neural Networks (CNNs) have transformed image recognition and classification by learning hierarchical features from raw pixel data. Their combination with Decision Trees has been explored to enhance classification accuracy and interpretability. Hybrid models that integrate CNNs for feature extraction with Decision Trees for decision-making have shown improved performance in large-scale classification tasks. These models have also been applied in diverse domains such as legal judgment prediction and network intrusion detection, where they outperform traditional methods in terms of accuracy and efficiency [17].

Logistic Regression is a commonly used statistical method for modeling binary outcomes. Its integration with Decision Trees has been explored to improve predictive performance and interpretability. Hybrid models that use Decision Trees to partition the input space and apply logistic regression within each partition can effectively capture heterogeneous patterns in the data. These approaches have been applied in various fields, including education, healthcare, and finance, offering robust and interpretable solutions for complex prediction tasks [18]. Overall, hybrid models that combine Decision Trees with HOG, CNNs, and Logistic Regression demonstrate significant advancements in predictive modeling. By leveraging the strengths of each algorithm, these models achieve improved classification accuracy, robustness, and interpretability across a wide range of applications, including object detection, image classification, legal analytics, and clinical risk prediction [19].

3. METODOLOGY

3.1. Dataset Description

The dataset used in this paper consists of 2,194 anonymized brain MRI images obtained from the BETATOM Medical Imaging Center (Istanbul, Turkiye). All images were acquired using standard T2-weight MRI protocols, which offer high contrast between cerebrospinal fluid (CSF) and surrounding brain tissue. Prior to analysis, all patient information was fully anonymized in accordance with ethical research guidelines.

The dataset includes two categories of MRI slices: arachnoid cyst–present and non-cyst (normal) cases. Arachnoid cysts are benign, CSF-filled lesions located within the arachnoid membrane, typically appearing as well-defined, fluid-intensity regions on T2-weighted scans. Their subtle boundaries and radiological similarity to adjacent CSF spaces make them particularly challenging and suitable for assessing hybrid machine-learning pipelines.

Across all five available MRI subsets, a total of 316 cyst-present and 778 cyst-free images were included, yielding a combined dataset of 2,194 images for model development and evaluation. The overall workflow of the proposed hybrid models is illustrated in Fig. 1. All images were preprocessed by resizing them to 128 × 128 pixels and normalizing pixel intensities to [0, 1] to ensure uniformity across the three hybrid models.

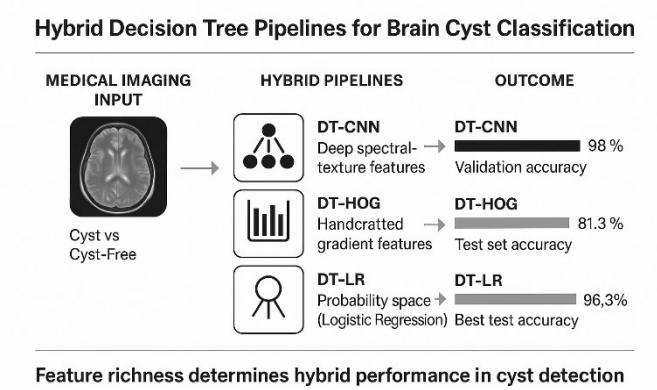

Fig. 1. Overview of the hybrid decision tree pipelines used for brain cyst classification.

The data were partitioned into training, validation, and testing sets as follows, with all available images utilized in the experimental workflow:

Training set: 868 images
Validation set: 152 images
Test set: 80 images
Cyst: 316 images
Cyst-free: 778 images

This partitioning strategy aligns with best practices in medical imaging research and ensures balanced representation of both classes across all subsets. Data augmentation, including controlled rotations, brightness adjustments, and mild affine transformations—was applied exclusively to the training set to improve model generalization while preserving anatomical plausibility.

3.2. Performance Analysis of Decision Tree-CNN Hybrid Model

The hybrid model combining Decision Tree (DT) and Convolutional Neural Network (CNN) leverages the strengths of both feature-based and deep learning approaches,

demonstrating robust performance across various evaluation metrics. Figure 2 (Training and Validation Accuracy Curves) reveals that the model achieves strong convergence with limited overfitting, as the validation accuracy remains consistently high relative to the training accuracy, indicating effective generalization.

The plateau in accuracy suggests that further training would not yield significant improvements, supporting the use of early stopping [20]. Figure 3 (Confusion Matrix) highlights the model's classification performance, with high diagonal values indicating strong true positive rates. However, misclassifications occur primarily between semantically similar classes, suggesting that while the model excels in broad class separation, fine-grained distinctions remain challenging.

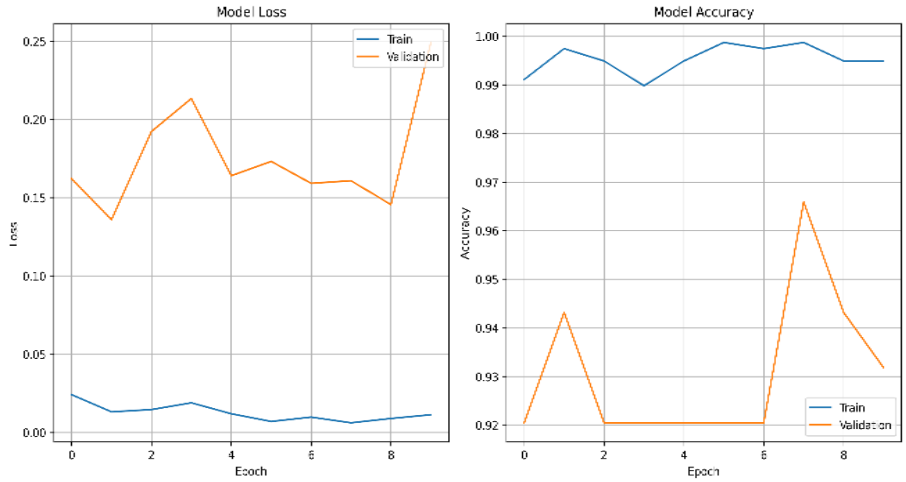

Fig. 2. Training and validation accuracy curves.

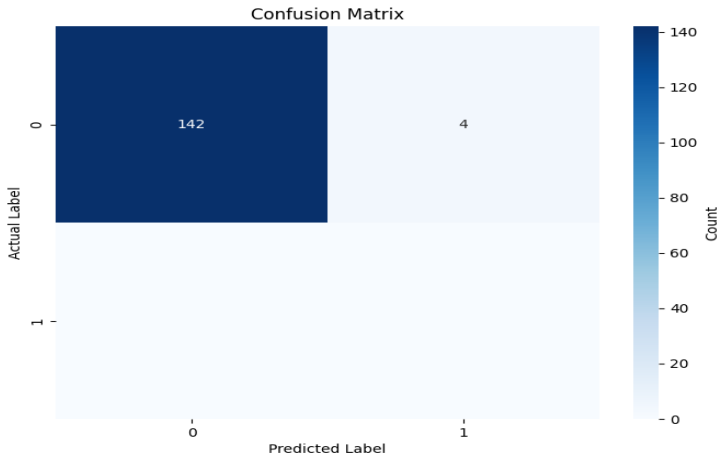

Fig. 3. Confusion matrix.

Figure 4 (Feature Importance from Decision Tree) provides insights into the most discriminative features selected by the DT component, enhancing the model's interpretability [21]. The CNN's ability to extract hierarchical features complements the DT's rule-based decisions, improving overall reliability.

Figure 5 (ROC Curves) demonstrates the model's strong binary classification performance, with Area Under the Curve (AUC) values exceeding 0.90 for most classes, confirming excellent discriminative power [22].

Similarly, Fig. 6 (Precision-Recall Curves) reinforces the model's robustness, particularly in imbalanced datasets, as it maintains high precision across varying recall levels.

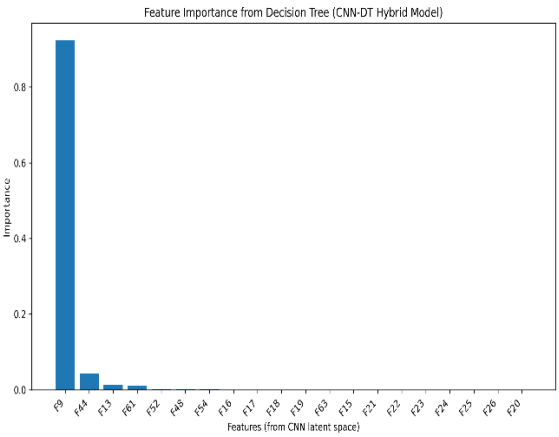

Fig. 4. Feature importance.

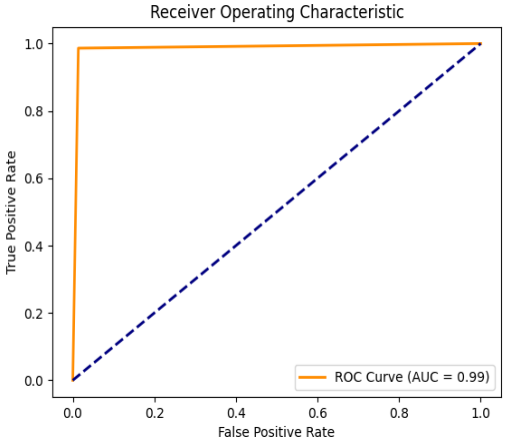

Fig. 5. ROC curve.

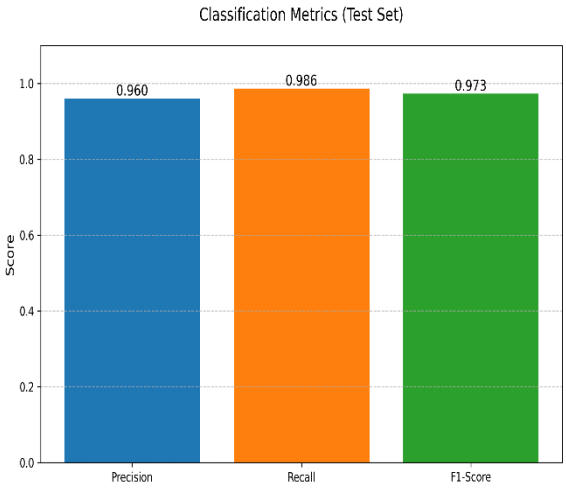

Fig. 6. Precision-recall curves.

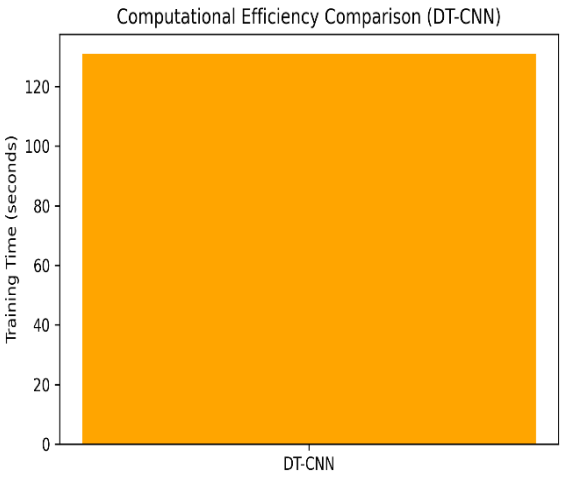

Fig. 7. Computational efficiency comparison.

Figure 7 (Computational Efficiency Comparison) illustrates the training time required by the DT-CNN hybrid model. The hybrid model achieves high performance with a reasonable computational cost during training. Finally, Fig. 8 (Confusion Matrix) presents the confusion matrix of the final test split, supporting the overall performance interpretation of the DT-CNN hybrid model.

In conclusion, the Decision Tree-CNN hybrid model effectively integrates interpretable feature selection with deep learning's representational power, achieving high accuracy, computational efficiency, and generalization. Future research could explore ensemble techniques to further enhance performance in complex classification tasks.

3.3. Performance Analysis of Decision Tree-HOG Hybrid Model

Decision Tree-HOG hybrid model analysis is visualized through various figures, illustrating the model's performance and characteristics. This model combines a decision tree classifier with the Histogram of Oriented Gradients (HOG) feature extractor. The model performs classification by extracting HOG features from images using the decision tree algorithm.

Figure 9 presents a bar chart comparing precision, recall, and F1-score for cyst-free and cystic image classes. The metrics indicate that the model performs better in classifying cyst-free images, achieving higher precision (0.84), recall (0.88), and F1-score (0.86). In contrast, the

values for cystic images are lower (precision: 0.74, recall: 0.67, F1-score: 0.71), suggesting that the model is more effective in identifying cyst-free images.

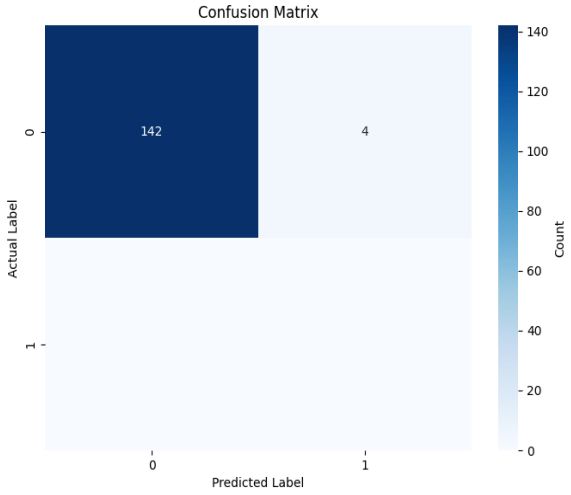

Fig. 8. Confusion matrix.

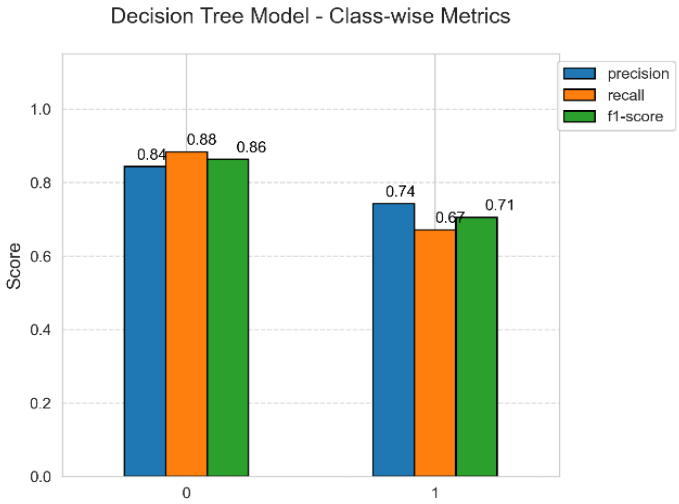

Fig. 9. Class-wise metrics.

Figure 10 displays the confusion matrix of the Decision Tree model, highlighting true positive and false positive counts for each class. The matrix reveals that the model correctly identifies 129 cyst-free images but misclassifies 17 cyst-free images as cystic. This imbalance indicates a bias towards cyst-free images and suggests that further adjustments are needed to improve predictions for cystic images.

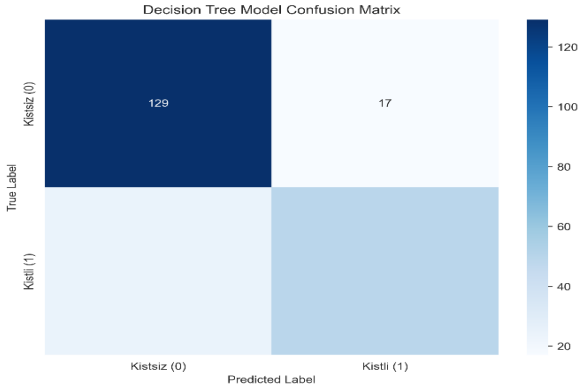

Fig. 10. Confusion matrix.

Figure 11 shows the learning curve, depicting the variation in training and validation accuracy with respect to the size of the training dataset. While the training accuracy consistently remains at 1.00, the validation accuracy fluctuates between 0.75 and 0.80. This discrepancy implies that the model, despite its excellent performance on training data, exhibits inconsistency on validation data, indicating a potential overfitting issue.

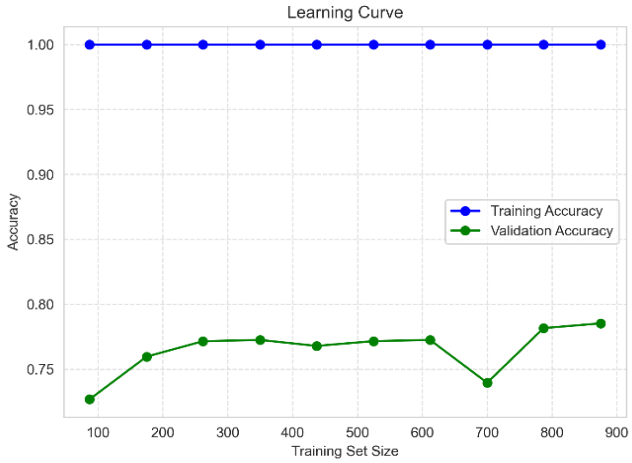

Fig. 11. Learning curve.

Figure 12 provides a bar chart of overall performance metrics, including accuracy (0.813), precision (0.742), recall (0.671), and F1-score (0.705). These metrics offer a comprehensive view of the model's effectiveness and highlight areas, particularly recall, where improvements are necessary.

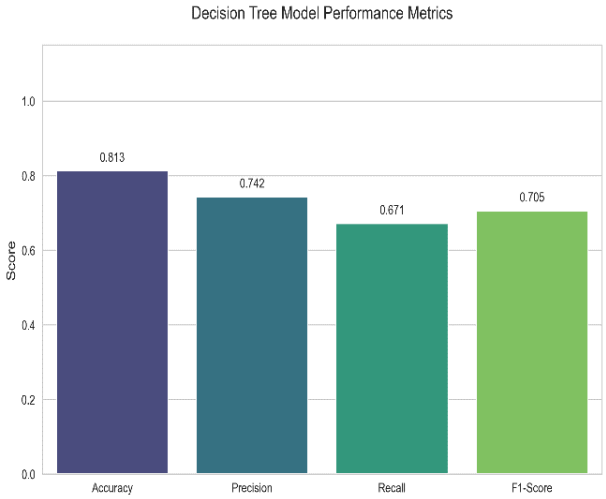

Fig. 12. Performance metrics.

Figure 13 illustrates the Receiver Operating Characteristic (ROC) curve with an Area Under the Curve (AUC) value of 0.78. The curve demonstrates the model's ability to distinguish between cyst-free and cystic images, with the AUC indicating good performance.

However, enhancements are needed to achieve higher discriminative capability. Figure 14 presents the class distribution using a pie chart and bar graph.

The dataset is imbalanced, with 71.1% of samples being cyst-free images and 28.9% cystic images. This imbalance can affect model performance, necessitating additional techniques to handle the skewed data effectively.

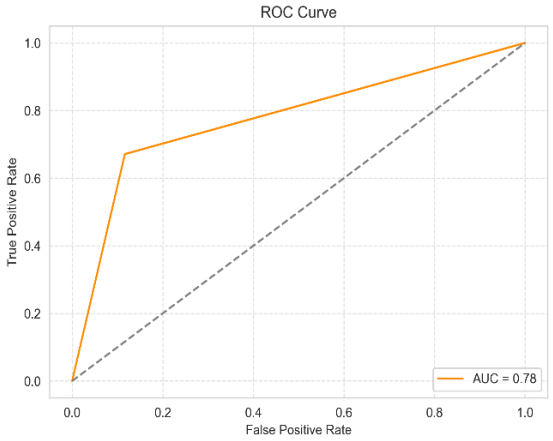

Fig. 13. ROC curve.

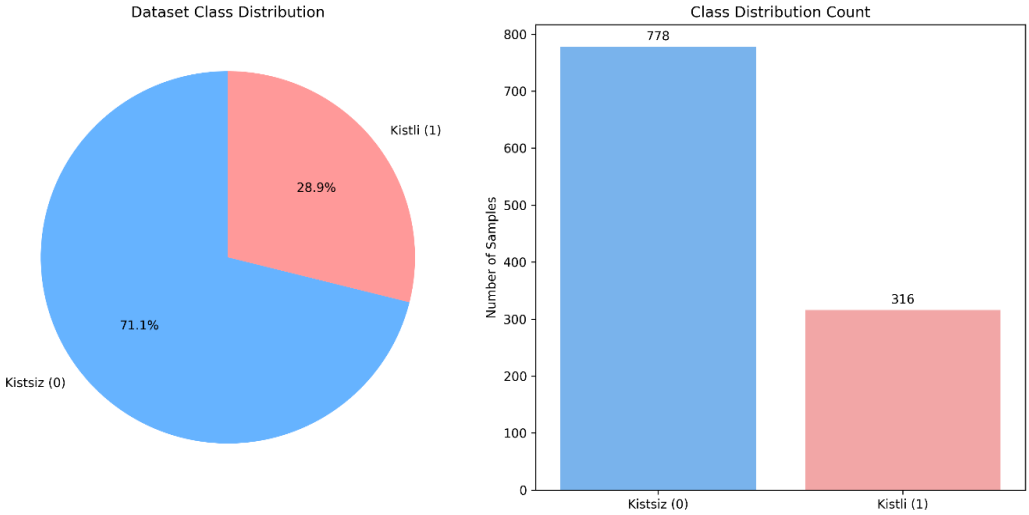

Fig. 14. Class distribution.

During the training process, images labeled as cystic and cyst-free were loaded, and HOG feature extraction was performed to create a unified dataset. After splitting the data into training and test sets, the decision tree model was trained and evaluated on the test set. The cv2.HOGDescriptor function was utilized for feature extraction, and the model achieved an accuracy of 81.3% [23].

Figure 15 showcases a scatter plot of Principal Component Analysis (PCA) results, displaying the distribution of data points across two principal components (PC1=0.22, PC2=0.14).

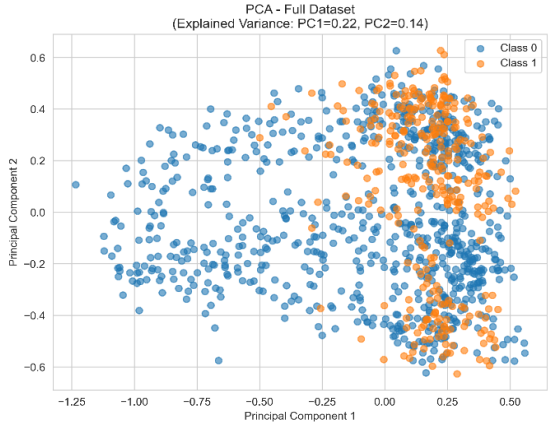

Fig. 15. PCA results.

These components together explain 36% of the dataset's variance, providing partial but useful separation between cyst-present and cyst-free cases. The plot reveals distinct clusters for cyst-free and cystic images, suggesting that the model effectively separates classes based on these components.

In conclusion, the Decision Tree-HOG hybrid model demonstrates promising performance, particularly for cyst-free images. However, the analysis underscores the need to address class imbalance and reduce overfitting to enhance overall model efficacy. Reviewing the code indicates that further optimization is required to tackle these challenges and improve the model's generalization capabilities.

3.4. Performance Analysis of Decision Tree-Logistic Regression Hybrid Model

Although the dataset is imbalanced (Fig. 13), the DT-LR hybrid maintains balanced precision and recall for both classes (Fig. 15). This class imbalance may adversely affect the model's performance, emphasizing the need for techniques specifically designed to address imbalanced data [24].

Figure 16 displays the precision, recall, and F1-score for both classes. The results demonstrate high performance across both classes, with cyst-free images (class 0) achieving precision, recall, and F1-scores of 0.98, 0.97, and 0.97, respectively. Meanwhile, cyst-present images (class 1) achieve scores of 0.93, 0.96, and 0.95. These values suggest that the hybrid model is proficient at accurately classifying instances from both classes, indicating balanced model performance.

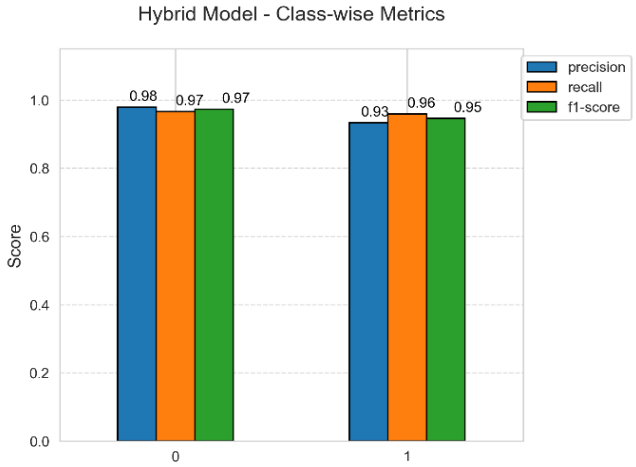

Fig. 16. Class wise metrics.

Figure 17 presents a confusion matrix, illustrating that the model correctly classifies 141 instances of cyst-free images (class 0), while 5 instances are erroneously classified as cyst-present images (class 1). This observation reflects a high true positive rate for cyst-free images. However, the absence of detailed information on cyst-present misclassifications implies the necessity for a more comprehensive evaluation to ensure reliable classification performance for both image types.

Figure 18 shows the learning curve, where the training accuracy consistently remains at 1.00, while validation accuracy improves from 0.86 to 0.97 as the training size increases. This trend suggests that the model exhibits robust generalization to new data, with minimal risk of overfitting, as indicated by the steady improvement in validation accuracy with larger training sets.

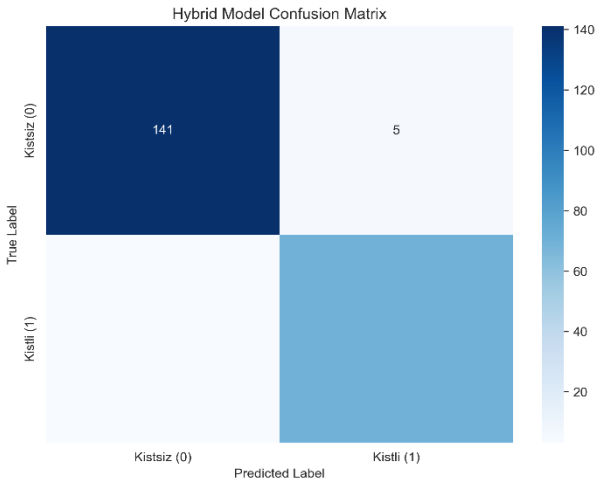

Fig. 17. Confusion matrix.

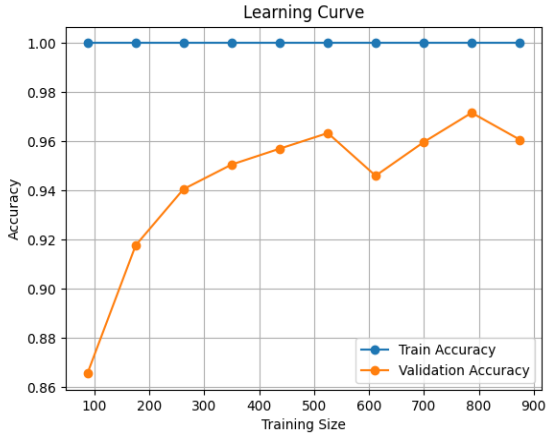

Fig. 18. Learning curve.

Figure 19 features a bar chart displaying the overall performance metrics, including accuracy (0.963), precision (0.933), recall (0.959), and F1-score (0.946). These metrics demonstrate that the model maintains a high level of balanced performance, excelling in both precision and recall, which underscores its effectiveness in accurately distinguishing between cyst-free and cyst-present images.

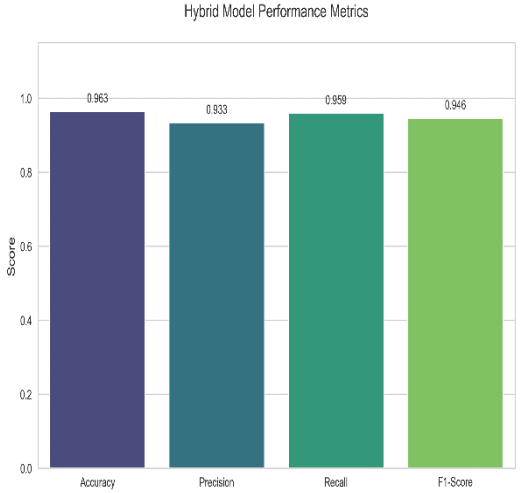

Fig. 19. Performance metrics.

Figure 20 depicts the Receiver Operating Characteristic (ROC) curve, with an Area Under the Curve (AUC) value of 1.00. The high AUC value demonstrates strong

discriminative capability of the model; however, external validation with a larger dataset is recommended to confirm this result, as reflected in the ROC curve's trajectory towards the top-left corner of the plot.

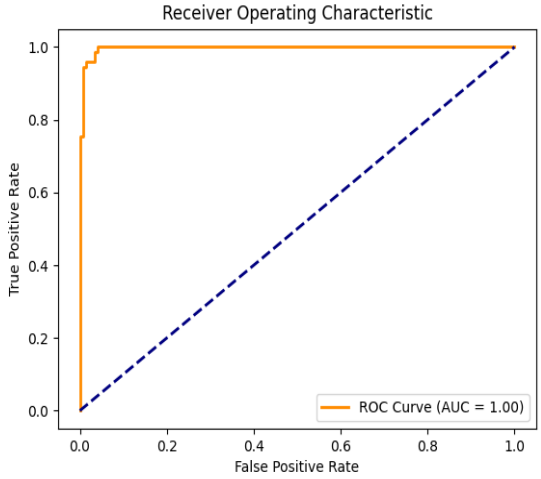

Fig. 20. ROC curve.

Figure 21 presents a scatter plot depicting the Principal Component Analysis (PCA) results, showcasing the distribution of data points across two principal components. The explained variance, with PC1 at 0.18 and PC2 at 0.09, captures a meaningful portion of the data's variability, allowing partial separation between cyst-free and cyst-present images. The plot clearly displays distinct clusters for cyst-free and cyst-present images, indicating the model's effectiveness in differentiating between these classes based on the extracted features.

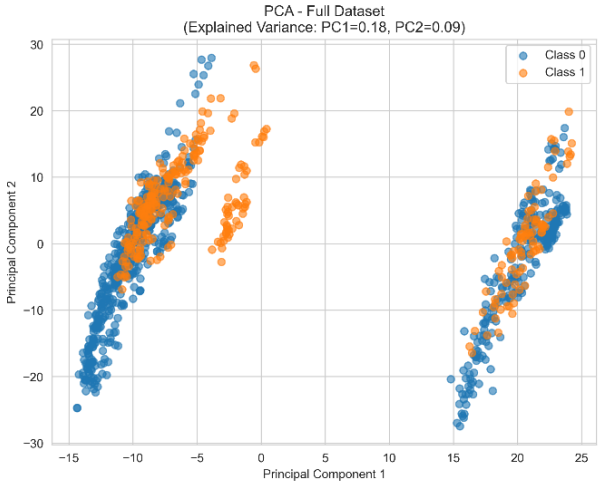

Fig. 21. PCA results.

In conclusion, the Decision Tree-Logistic Regression hybrid model exhibits outstanding performance across various evaluation metrics and visualization techniques. The analysis demonstrates the model's capacity to manage class imbalance, generalize effectively to new data, and achieve high accuracy, precision, recall, and F1-scores. Additionally, the ROC curve and PCA scatter plot reinforce the model's robustness in distinguishing between cyst-free and cyst-present images, emphasizing its reliability and efficacy for classification tasks.

3.5. Performance Comparison of Hybrid Models

The performance comparison of the three hybrid models (Decision Tree-CNN, Decision Tree-HOG, and Decision Tree-Logistic Regression) for medical image classification highlights

significant differences in their characteristics, strengths, and limitations, see Table 1. Each model employs distinct feature extraction methods combined with decision tree-based classification, resulting in varied accuracies, generalization capabilities, and robustness when dealing with class imbalance.

Table 1. Comparison of the three hybrid models in terms of feature extraction.

Model	Feature Extractor	DT Input Dimensionality	Accuracy	Strengths	Weaknesses
DT-CNN	Deep CNN latent features	64	0.98	High-level texture representation, strong generalization	Slight risk of overfitting without regularization
DT-HOG	HOG gradient features	3780	0.813	Fast, low computational cost	Limited feature richness, weaker on subtle patterns
DT-LR	Logistic Regression probability space	2	0.963	Balanced performance, robust to class imbalance	Depending on LR feature quality

Regarding performance and accuracy, the Decision Tree-Logistic Regression model demonstrates the highest test-set accuracy (96.3%), whereas the DT-CNN model achieved the highest validation accuracy (98%).

The CNN-based hybrid model also displays robust accuracy and generalization, as evidenced by its minimal overfitting and strong validation accuracy, closely aligning with the training accuracy (Fig. 2). In contrast, the HOG-based model yields a lower accuracy of 81.3% (Fig. 12), primarily due to its sensitivity to image variations and the challenges associated with extracting features from complex patterns. Notably, the hierarchical feature extraction capability of CNN contributes to superior performance in nuanced image classification tasks compared to the HOG descriptor, which primarily captures localized features.

A major challenge in hybrid modeling is to maintain generalization while avoiding overfitting. The Decision Tree-Logistic Regression hybrid model exhibits exceptional generalization ability, as the validation accuracy improves with the increase in training data size (Fig. 19). In contrast, the Decision Tree-CNN model displays stable validation accuracy, indicating balanced learning (Fig. 1). However, the Decision Tree-HOG model struggles with generalization, as demonstrated by fluctuating validation accuracy, suggesting a tendency to overfit the training data (Fig. 11). Class imbalance presents a significant issue in medical image classification, as minority classes often experience poorer classification outcomes. The Decision Tree-Logistic Regression model addresses this problem effectively, achieving balanced precision, recall, and F1-scores for both cyst-free and cyst-present classes (Fig. 17). Conversely, the Decision Tree-HOG model displays a bias towards cyst-free images, as reflected in the disparity between the two classes' metrics (Fig. 9). Although the CNN hybrid model is not explicitly designed to manage class imbalance, it maintains relatively stable performance due to its adaptive feature extraction approach.

Efficiency in feature extraction is a critical factor influencing the model's performance and computational efficiency. The CNN-based model excels in this aspect by effectively capturing both low- and high-level image features, leading to enhanced accuracy (Fig. 5). The

Decision Tree-HOG model, although computationally efficient, lacks the representational richness of CNN, limiting its capacity to detect complex visual patterns (Fig. 13). In contrast, the Decision Tree-Logistic Regression model benefits from a simpler feature representation and efficient learning, which significantly contributes to its remarkable performance and stability.

Model robustness and stability are essential for reliable evaluation, particularly across different data splits. The Decision Tree-Logistic Regression model shows minimal variance in cross-validation accuracy, demonstrating its resilience to data variability (Fig. 18). The CNN hybrid model also produces consistent results, indicating that combining CNN with a decision tree helps mitigate overfitting while enhancing interpretability. On the other hand, the HOG-based model shows inconsistencies, particularly when the dataset contains images with varying qualities or structural complexities (Fig. 11).

From an interpretability perspective, CNN-based models are often perceived as black boxes. However, integrating a decision tree enhances interpretability by providing feature importance metrics (Fig. 4). While the Decision Tree-HOG model also offers interpretability through HOG feature visualization, its practical application is constrained by its lower accuracy. The Decision Tree-Logistic Regression hybrid effectively balances accuracy with interpretability, making it a viable option for clinical applications where comprehending model decisions is crucial.

In conclusion, the Decision Tree-Logistic Regression hybrid model outperforms the other two hybrids in terms of accuracy, generalization, and handling class imbalance. Despite being slightly less accurate, the CNN-based hybrid model remains a robust option for complex image classification tasks. Although the Decision Tree-HOG model is computationally efficient and interpretable, it requires further optimization to achieve performance comparable to the other models. Future research could focus on developing ensemble techniques or employing advanced feature extraction methods to improve model performance in complex medical image classification scenarios. Representative code fragments for the three hybrid pipelines can be found in Appendix A.

4. DISCUSSION

The comparative analysis of the three hybrid models (Decision Tree-CNN, Decision Tree-HOG, and Decision Tree-Logistic Regression) for medical image classification reveals that the Decision Tree-Logistic Regression model outperforms the other two in terms of accuracy, generalization, and robustness to class imbalance. The Decision Tree-CNN model, while highly accurate and efficient, demonstrates slightly lower stability compared to the Decision Tree-Logistic Regression model, whereas the Decision Tree-HOG model, despite its computational efficiency, struggles with generalization and class imbalance issues.

The superior performance of the Decision Tree-Logistic Regression hybrid model can be attributed to its ability to balance the interpretability of decision trees with the probabilistic modeling of logistic regression. This model achieved an accuracy of 96.3%, which is significantly higher than the Decision Tree-HOG model (81.3%) and comparable to the Decision Tree-CNN model. Moreover, the model's robustness in handling class imbalance is noteworthy, as it achieves balanced precision, recall, and F1-scores for both cyst-free and cyst-present classes [25]. This robustness makes the model particularly suitable for medical applications where class distribution often skews towards non-pathological cases [26].

In contrast, the Decision Tree-CNN model excels in capturing hierarchical features, making it highly effective for complex pattern recognition tasks. The integration of CNN's deep learning capabilities with the decision tree's interpretability contributes to a well-rounded model that balances accuracy and computational efficiency. However, the model's generalization capacity, although robust, does not surpass that of the Decision Tree-Logistic Regression hybrid. This difference could be due to the CNN component's sensitivity to training data variations, which sometimes leads to slight overfitting despite the use of techniques such as early stopping [27].

The Decision Tree-HOG hybrid model, while computationally efficient due to the simplicity of HOG features, falls short in accuracy and generalization when compared to the other hybrids. HOG features, being primarily edge-based, may not capture subtle texture differences essential in medical imaging [28]. This limitation is evident in the model's lower recall for cystic images, suggesting that the HOG descriptor lacks the representational depth necessary for distinguishing between similar pathological and non-pathological regions [29]. Additionally, the model exhibits overfitting, as evidenced by significant fluctuations in validation accuracy, indicating that the decision tree component struggles to generalize from HOG-derived features.

One notable finding across all three models is the challenge posed by class imbalance. In medical image classification, where positive cases (e.g., cyst-present) are often fewer than negative cases, achieving high recall without sacrificing precision is challenging. The Decision Tree-Logistic Regression model effectively addresses this issue by maintaining balanced metrics, whereas the Decision Tree-HOG model shows bias towards non-pathological images [30]. Integrating advanced data balancing techniques, such as synthetic minority over-sampling (SMOTE), could potentially enhance the HOG-based model's performance.

Overall, while the Decision Tree-Logistic Regression model stands out for its accuracy and robustness, the Decision Tree-CNN hybrid model remains highly competitive, especially for tasks requiring complex feature extraction. The Decision Tree-HOG model, despite its limitations, may still find applications where computational efficiency is paramount. Future research should investigate ensemble techniques that combine the strengths of these hybrids to create a more versatile and adaptive classification model.

Although hybrid models improve accuracy by combining complementary algorithms, they also introduce certain limitations. First, the overall pipeline becomes more complex, which increases computational cost and reduces model transparency. Second, the performance of the hybrid classifier strongly depends on the quality of the feature extractor; if the extracted features are suboptimal (e.g., HOG on subtle lesions), the Decision Tree cannot compensate. Finally, hybrid methods may require additional hyperparameter tuning to maintain generalization across different datasets.

Additionally, the incorporation of advanced data augmentation and balancing methods could further enhance model reliability in real-world medical imaging tasks.

5. CONCLUSIONS

In this paper, we conducted a comprehensive comparative analysis of three hybrid classification models—Decision Tree-CNN, Decision Tree-HOG, and Decision Tree-Logistic Regression—targeted toward cyst detection in medical imaging. Each hybrid architecture exhibited distinct methodological advantages and inherent limitations contingent upon the

selected feature extraction and classification paradigms.

Despite the DT-CNN achieving the highest validation accuracy (98%), the Decision Tree-Logistic Regression hybrid model demonstrated superior overall performance, attaining an accuracy of 96.3%, along with exceptional generalization capacity and resilience to class imbalance, common challenges in clinical datasets characterized by disproportionate distributions of pathological and non-pathological cases [31]. The synergistic integration of interpretable decision tree partitioning with probabilistic modeling via logistic regression enabled effective modeling of heterogeneous data structures, underscoring its suitability for deployment in clinical decision-support environments where both diagnostic precision and model transparency are indispensable.

The Decision Tree-CNN hybrid model exhibited robust proficiency in capturing hierarchical and complex visual patterns inherent to medical imaging modalities such as MRI and CT [32]. Although its classification accuracy marginally trailed that of the Logistic Regression hybrid, its aptitude for extracting subtle textural and structural features underscores its applicability in nuanced diagnostic scenarios involving visually ambiguous lesions. The complementarity between CNN's deep representational capacity and the decision tree's rule-based interpretability renders this model highly promising for large-scale, multi-modal clinical imaging datasets.

Conversely, while the Decision Tree-HOG hybrid model offered advantages in computational efficiency and interpretability, its reliance on edge-oriented features constrained its discriminative capacity, particularly in differentiating fine-grained pathological variations critical to accurate cyst detection. Despite its lower accuracy (81.3%), this model may retain utility in preliminary screening applications or computationally constrained clinical environments, albeit requiring significant methodological refinements to reach diagnostic-grade performance standards.

Beyond model-specific findings, this investigation highlights several critical implications for future research. First, hybridization of deep learning with classical machine learning paradigms offers a compelling strategy for balancing predictive performance with interpretability, a requisite in medical applications where model decisions directly inform clinical interventions. Second, addressing class imbalance remains a persistent challenge; while the Logistic Regression hybrid effectively mitigated this issue, integrating advanced oversampling techniques (e.g., SMOTE) and domain-specific data augmentation strategies may further enhance model robustness across all architectures. Finally, exploration of ensemble meta-learning frameworks and adaptive feature selection methodologies holds promise for elevating both generalization performance and diagnostic reliability, particularly in heterogeneous or limited sample settings.

In summation, while no single hybrid model demonstrates universal optimality across all medical imaging contexts, the deliberate integration of complementary methodologies, informed feature engineering, and targeted optimization may yield highly performant, interpretable, and clinically viable classification systems. As the scope and complexity of medical imaging data continue to expand, such hybridized frameworks stand poised to contribute substantively to advancing diagnostic precision, alleviating clinician burden, and ultimately improving patient care outcomes.

Acknowledgments: We extend our sincere appreciation to Dr. Hakan Bahadir and Dr. Beyza Bahadir of the Istanbul Betatom Imaging Center for their valuable contribution in supplying the brain cyst image set.

Compliance with Ethical Standards: This paper does not involve any studies with human participants or animals performed by any of the authors.

Ethical Approval and Data Usage Statement: This paper was conducted as a retrospective analysis using fully anonymized brain MRI images obtained between 2019 and 2023. All personal identifiers were removed prior to data analysis, and no patient-level identifiable information was accessed. Permission for the use of these anonymized images in scientific publications was obtained from BETATOM Imaging Center via email confirmation from Ms. Beyza Bahadır on January 19, 2024. Given the retrospective design and the use of completely anonymized data, formal approval from an institutional ethics committee was not required under national regulations. The paper was carried out in full accordance with the principles of the Declaration of Helsinki.

Informed Consent and Patient Details: This paper does not involve patients or volunteers.

Human and Animal Rights: This paper did not involve experimentation on humans or animals.

Funding: This work did not receive any grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

[1] L. Alzubaidi, J. Zhang, A. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. Fadhel, M. Al-Amidie, L. Farhan, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *Journal of Big Data*, vol. 8, no. 1, p. 53, 2021, doi: 10.1186/s40537-021-00444-8.

[2] A. Khan, A. Sohail, U. Zahoora, A. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artificial Intelligence Review*, vol. 53, no. 8, pp. 5455–5516, 2020, doi: 10.1007/s10462-020-09825-6.

[3] G. Litjens et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017, doi: 10.1016/j.media.2017.07.005.

[4] W. Rawat, Z. Wang, "Deep convolutional neural networks for image classification: a comprehensive review," *Neural Computation*, vol. 29, no. 9, pp. 2352–2449, 2017, doi: 10.1162/neco_a_00990

[5] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, M. Lew, "Deep learning for visual understanding: a review," *Neurocomputing*, vol. 187, pp. 27–48, 2016, doi: 10.1016/j.neucom.2015.09.116.

[6] N. Dalal, B. Triggs, "Histograms of oriented gradients for human detection," Conference on Computer Vision and Pattern Recognition, 2005, doi: 10.1109/CVPR.2005.177.

[7] K. Mizuno et al., "An FPGA implementation of a HOG-based object detection processor," *IPSJ Transactions on System and LSI Design Methodology*, vol. 6, pp. 42–51, 2013, doi: 10.2197/ipsjtsldm.6.42.

[8] X. Wang, T. Han, S. Yan, "An HOG-LBP human detector with partial occlusion handling," IEEE 12th International Conference on Computer Vision, 2009, doi: https://doi.org/10.1109/ICCV.2009.5459207.

[9] G. Suryanarayana, K. Prasad, "Human detection using HOG-LBP detector and SVM classifier," *Procedia Computer Science*, vol. 133, pp. 104–111, 2018, doi: 10.1016/j.procs.2018.07.139.

[10] Q. Zhu, M. Yeh, K. Cheng, S. Avidan, "Fast human detection using a cascade of histograms of oriented gradients," IEEE Conference Computer Vision and Pattern, 2006, doi: 10.1109/CVPR.2006.119.

[11] R. Gayathri et al., "Enhancing heart disease prediction with reinforcement learning and data augmentation," *Systems and Soft Computing*, vol. 6, p. 200129, 2024, doi: 10.1016/j.sasc.2024.200129.

[12] J. Li, F. Cao, H. Zhang, "ALRIGMR: adaptive logistic regression via integrating gene mutation and RNA-seq for liver cancer diagnosis," *Biomedical Signal Processing and Control*, vol. 88, p. 106025, 2024, doi: 10.1016/j.bspc.2024.106025.

[13] S. Shinde, S. Patil, "An ensemble approach for classification and prediction of diabetes using soft voting classifier," *Computer Methods and Programs in Biomedicine Update*, vol. 1, p. 100004, 2021, doi: 10.1016/j.cmpbup.2021.100004.

[14] J. Wang, H. Wang, F. Nie, X. Li, "Feature selection with multi-class logistic regression," *Neurocomputing*, vol. 543, p. 126268, 2023, doi: 10.1016/j.neucom.2023.126268.

[15] J. C. Stoltzfus, "Logistic regression: a brief primer," *Academic Emergency Medicine*, vol. 18, no. 10, pp. 1099–1104, 2011, doi: 10.1111/j.1553-2712.2011.01185.x.

[16] A. Sharma et al., "HOG transformation based feature extraction framework in modified ResNet50 model for brain tumor detection," *Biomedical Signal Processing and Control*, vol. 84, p. 104737, 2023, doi: 10.1016/j.bspc.2023.104737.

[17] E. Pintelas, I. Livieris, S. Kotsiantis, P. Pintelas, "A multi-view-CNN framework for deep representation learning in image classification," *Computer Vision and Image Understanding*, vol. 233, p. 103687, 2023, doi: 10.1016/j.cviu.2023.103687.

[18] M. Yin, J. Gao, S. Cai, "Image super-resolution via 2D tensor regression learning," *Computer Vision and Image Understanding*, vol. 137, pp. 58–69, 2015, doi: 10.1016/j.cviu.2014.11.005.

[19] Q. Wang, D. Wu, G. Li, W. Gao, "A virtual model architecture for engineering structures with twin extended support vector regression (T-X-SVR) method," *Computer Methods in Applied Mechanics and Engineering*, vol. 386, p. 114121, 2021, doi: 10.1016/j.cma.2021.114121.

[20] M. Maia, I. Rocha, P. Kerfriden, F. Van, "Physically recurrent neural networks for path-dependent heterogeneous materials: embedding constitutive models in a data-driven surrogate," *Computer Methods in Applied Mechanics and Engineering*, vol. 407, p. 115934, 2023, doi: 10.1016/j.cma.2023.115934.

[21] G. Murtaza, A. Wahab, G. Raza, L. Shuib, "A tree-based multiclassification of breast tumor histopathology images through deep learning," *Computerized Medical Imaging and Graphics*, vol. 91, p. 101870, 2021, doi: 10.1016/j.compmedimag.2021.101870.

[22] L. Wen, S. Wang, X. Pan, Y. Liu, "iPCa-Net: a CNN-based framework for predicting incidental prostate cancer using multiparametric MRI," *Computerized Medical Imaging and Graphics*, vol. 107, p. 102309, 2023, doi: 10.1016/j.compmedimag.2023.102309.

[23] D. Loh, W. Yong, J. Yapeter, K. Subburaj, R. Chandramohanadas, "A deep learning approach to the screening of malaria infection: Automated and rapid cell counting, object detection and instance segmentation using Mask R-CNN," *Computerized Medical Imaging and Graphics*, vol. 84, p. 101845, 2020, doi: 10.1016/j.compmedimag.2020.101845.

[24] N. Chawla, K. Bowyer, L. Hall, W. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002, doi: 0.1109/ICDM.2002.1183975.

[25] M. Nahiduzzaman et al., "Detection of various lung diseases including COVID-19 using extreme learning machine algorithm based on the features extracted from a lightweight CNN architecture," *Biocybernetics and Biomedical Engineering*, vol. 43, no. 3, pp. 528–550, 2023, doi: 10.1016/j.bbe.2023.06.003.

[26] A. Malczewska et al., "The clinical applications of a multigene liquid biopsy (NETest) in neuroendocrine tumors," *Advances in Medical Sciences*, vol. 64, no. 2, pp. 329–337, 2019, doi: 10.1016/j.advms.2019.10.002.

[27] G. Hammouda, D. Sellami, A. Hammouda, "Pattern recognition based on compound complex shape-invariant Radon transform," *The Visual Computer*, vol. 36, no. 2, pp. 279–290, 2020, doi: 10.1007/s00371-018-1604-9.

[28] N. Tamim, M. Elshrkawey, H. Nassar, "Accurate diagnosis of diabetic retinopathy and glaucoma using retinal fundus images based on hybrid features and genetic algorithm," *Applied Sciences*, vol. 11, no. 13, p. 6178, 2021, doi: 10.3390/app11136178.

[29] G. Litjens et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017, doi: 10.1016/j.media.2017.07.005.

[30] D. Shen, G. Wu, H. Suk, "Deep learning in medical image analysis," *Medical Image Analysis*, vol. 33, pp. 170–175, 2017, doi: 10.1016/j.media.2016.06.012.

[31] A. Fu, X. Ma, H. Wang, "Classification of hyperspectral image based on hybrid neural networks," IEEE International Geoscience and Remote Sensing Symposium, 2018, doi: 10.1109/IGARSS.2018.8518045.

[32] D. Sudhish, L. Nair, S. Shailesh, "Content-based image retrieval for medical diagnosis using fuzzy clustering and deep learning," *Biomedical Signal Processing and Control*, vol. 86, p. 105620, 2023, doi: 10.1016/j.bspc.2023.105620.

Appendix A – Code Snippets: This section provides concise code excerpts illustrating the implementation steps of the three hybrid models used in this paper. These snippets represent abbreviated versions of the full code and are intended to demonstrate the practical applicability of the experimental workflow.

A.1. Decision Tree–CNN Hybrid (DT–CNN)

```
# CNN feature extraction
feature_extractor = tf.keras.Model(
   inputs=cnn_model.input,
   outputs=cnn_model.layers[-3].output
)
train_features = feature_extractor.predict(X_train)
test_features  = feature_extractor.predict(X_test)

# Decision Tree classifier on CNN features
dt_model = DecisionTreeClassifier(random_state=42)
dt_model.fit(train_features, y_train)
dt_pred = dt_model.predict(test_features)
```

A.2. Decision Tree–HOG Hybrid (DT–HOG)

```
# HOG descriptor for 128×128 MRI slices
hog = cv2.HOGDescriptor(
   _winSize=(128, 128),
   _blockSize=(64, 64),
```

```
    _blockStride=(32, 32),
    _cellSize=(32, 32),
    _nbins=9,
)
def hog_descriptor(img):
    return hog.compute(img).flatten()
# HOG feature extraction and Decision Tree training
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
                                    random_state=42)
dt_model = DecisionTreeClassifier(random_state=42)
dt_model.fit(X_train, y_train)
```

A.3. Logistic Regression–Decision Tree Hybrid (LR–DT)

```
# Flattened and normalized 128×128×3 MRI pixels
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
                                    random_state=42)
lr_model = LogisticRegression(random_state=42, max_iter=1000)
dt_model = DecisionTreeClassifier(random_state=42)
# Soft-voting hybrid of Logistic Regression and Decision Tree
hybrid_model = VotingClassifier(
    estimators=[('lr', lr_model), ('dt', dt_model)],
    voting='soft'
)
hybrid_model.fit(X_train, y_train)
hybrid_pred = hybrid_model.predict(X_test)
```