Applied AI Letters

WILEY

PERSPECTIVE OPEN ACCESS

A Hybrid Deep Learning Paradigm for Robust Feature Extraction and Classification for Cataracts

Akshay Bhuvaneswari Ramakrishnan[1] | Mukunth Madavan[2] | R. Manikandan[3] | Amir H. Gandomi[4,5,6]

[1]Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, Maryland, USA | [2]Mukunth Madhavan, AI Security Engineer, Banglore, India | [3]School of Computing, SASTRA Deemed University, Thanjavur, India | [4]Faculty of Engineering & Information Technology, University of Technology Sydney, Sydney, Australia | [5]University Research and Innovation Center (EKIK), Óbuda University, Budapest, Hungary | [6]Department of Computer Science, Khazar University, Baku, Azerbaijan

Correspondence: R. Manikandan (srmanimt75@gmail.com) | Amir H. Gandomi (gandomi@uts.edu.au)

Received: 30 September 2024 | **Revised:** 14 January 2025 | **Accepted:** 28 January 2025

Funding: The authors received no specific funding for this work.

Keywords: cataract | computer-assisted | fundus | hybrid | ocular | oneAPI | oneDNN | transfer learning

ABSTRACT

The study suggests using a hybrid convolutional neural networks-support vector machines architecture to extract reliable characteristics from medical images and classify them as an ensemble using four different models. Manual processing of fundus images for the automated identification of ocular disorders is laborious, error-prone, and time-consuming. This necessitates computer-assisted technologies that can automatically identify different ocular illnesses from fundus images. The interpretation of the photos also plays a massive role in the diagnosis. Automating the diagnosing procedure reduces human mistakes and helps with early cataract detection. The oneDNN library available in the oneAPI Environment provided by Intel has been used to optimize all transfer learning models for better performance. The suggested approach is verified through a range of metrics in experiments using the openly accessible Ocular Disease Intelligent Recognition dataset. The MobileNet Model outperformed other transfer learning techniques with an accuracy of 0.9836.

1 | Introduction

The primary cause of vision impairment at the global level is eye disease, including cataracts [1]. An early and accurate diagnosis of cataracts is important to make effective treatments and stop further vision loss. A cloudy patch on the lens of the eye, caused by aging and tissue deterioration of the eyes, is referred to as a cataract, which causes blurred vision. Early detection of eye conditions like cataracts is one economical and efficient way to prevent blindness. World Health Organization estimates that 2.2 billion people in our world are visually impaired, and at least a billion of them might have avoided it if not [2]. Identification of cataracts is essential as it is a widespread condition that affects many people and also, at the same time, has massive costs to health care.

Deep learning techniques have boosted the utilization of autonomous image recognition methods to diagnose ocular diseases, particularly with the use of transfer learning and convolutional neural networks (CNN) [3]. In the field of ophthalmic disease diagnosis, the technology is advantageous, for it can discretely learn about intricacies contained in large datasets and identify properties human perception cannot detect [4]. Deep learning has made advancements with promise in cataract or ocular disease detection, and the research of Menaka et al. demonstrated how well a deep learning model can detect cataracts using fundus pictures with an accuracy of 92.78% [5]. Just like this, Thanoon and Dawwd used a deep learning system with parallel architecture, which showed the model's ability to detect cataracts rapidly and accurately (96.7%) [6]. They also studied the use of more sophisticated architectures, like ResNet-50, and were able to diagnose cataracts in fundus images with over 97% accuracy, even in the presence of variable picture visibility. The models in this regard showed good accuracy but did not perform well with respect to cataract identification. It was studied by considering various standalone machine learning models and conventional image processing techniques. All the models were taken into consideration but discarded for use in the study. Often, created features are required when using traditional image processing techniques for medical photos, and they may be better captured by deep learning models [7]. In addition, although applicable, other machine learning algorithms may not be as good at detecting the subtle differences and spatial hierarchies that are crucial to a precise cataract diagnosis in the fundus image.

There are currently two diagnostic approaches to cataracts; these include manual or tonometry, visual acuity, and slit lamp examination: the time-consuming, subjective, and diagnostic discrepancies of these procedures. The human technique is improved over conventional machine learning models like random forest and support vector machines (SVM), which provide automatic categorization. Nevertheless, the trade-off is that it is computationally inefficient, low in accuracy, and requires manual feature extraction. However, deep learning models such as Efficient NetV2S, Resnet50, and other complex deep learning models, mainly CNNs, have higher loading and feature extraction efficiency and accuracy. However, such models are complicated to train efficiently as they require large labeled data sets.

This article introduces a novel approach to integrating SVMs and CNNs to extract features for HCI problems. Four distinct classification models are implemented: In the project, we implemented Inception, MobileNet, ResNet, and VGG19. To classify, an individual model is used with results rigorously compared to determine which model performs best. The oneDNN library is employed within a oneAPI environment to optimize each model individually for efficacies during the classification process [8]. The intricacies and the potential benefits of each model were compared. CNNs are reliable at extracting fine details from medical imagery, but noise and overlapping features might make them unreliable for fine-grained classification tasks. It is shown that using SVMs for feature selection increases the discriminative strength of the learnt features while, at the same time, concentrating on the most pertinent patterns in the data. It solves the problem of CNN-only techniques, improving the classification accuracy by combining SVMs for the fine-tuning of feature selection and CNNs for reliable feature extraction. This is particularly useful for diagnosing cataracts because it is important to spot minute changes in the fundus image. OneAPI integration offers a standard programming model used to create an integrated and optimized application on diverse hardware architectures. It thus enables the operation of our models and their ability to operate in numerous computing environments. Following this, several sections present a thorough investigation of aspects of our proposed methodology, including optimized classification strategies (using oneDNN and oneAPI) and hybrid CNN-SVM feature extraction. We seek to improve the precision and reliability of cataract diagnostics by combining insights from individual classification results with the knowledge that CNN-SVM has about features. It will also help advance automated medical image analysis.

2 | Related Studies

The contemporary improvement of deep learning, remarkably CNN, has totally changed medical imaging, offering notable gains in diagnostic efficiency and accuracy. This is because CNNs are extremely well suited to analyze medical pictures. After all, they are able to quickly spot characteristics such as edges and textures, which are imperative to tasks like tumor diagnosis and fracture detection. This characteristic makes CNNs very useful in medical imaging applications since they considerably decrease the number of parameters and improve the computing efficiency [9]. Specifically, CNNs have shown great promise in ocular diagnostics. Some models they used include VGG-16, VGG-19, and ResNet, and they stated that they have good accuracy rates in diagnosing various eye diseases (including cataracts) [10]. The complexity and unpredictability of retinal pictures are managed, but these systems provide accurate and dependable diagnoses. Complex CNNs with inception modules and residual networks demonstrate significant improvement in the diagnosis of diabetic retinopathy and cataracts and improve feature extraction and model performance. Raman et al. [11] further discussed the usage of these architectures. Moreover, as Pingat et al. [12] suggest, techniques that hybridize CNNs with traditional classifiers permit better interpretability as well as improved performance and, thus, represent an appropriate strategy for the diagnosis of eye disorders. Olaniyan et al. combined Siamese networks VGG16 and Grad-CAM for explainability and presented a hybrid deep learning model with high cataract detection accuracy and visual interpretations of the machine's predictions [13]. The effectiveness of the ResNet50 transfer learning model, handling the dataset imbalanced, and high accuracy in cataract detection, as reported in [14], was demonstrated by Mahmood et al., which included data augmentation and fine-tuning strategies. These studies collectively demonstrated how CNNs transformed medical imaging, particularly the diagnosis of ocular disease. They suggested CNNs as a superior architecture over other networks due to their better feature extraction abilities, computational efficiency, and flexibility with more convoluted medical image data. Multiclass multilabel ophthalmological illness diagnosis is done by researchers P. Khanna and N. Gour [15] using a CNN based on transfer learning. The recommended method uses four cutting-edge pre-trained CNN models, which are then fine-tuned with Ocular Disease Intelligent Recognition (ODIR) database. During validation, training, and testing phases, parametric evaluations showed that VGG-16 network performed better than CNN architectures of ResNet, Inception V3, and MobileNet. The F1 and AUC values obtained with VGG-16 using two input approaches with SGD optimizer were 85.57 and 84.93, respectively. Comparing other designs to the concatenated input approach with VGG-16 architecture, also known as Model2, we have found that AUC and F-1 scores are much higher, 68.88 and 85.57, respectively. This improvement is by using VGG-16 with SGD optimizer. Li et al. [16] used the ODIR dataset and developed a deep residual network (ResNet) based model for the automated categorization of diabetic retinopathy. The model compared favorably with other cutting-edge techniques with an accuracy of 91.3%. Cao et al. [17] used Haar wavelet to provide an automated cataract identification using retinal picture characteristics. Retinal pictures of mild, moderate, and severe cataracts and normal (non-cataract) conditions are automatically identified by the enhanced Haar wavelet. The four-class classification and the two-class classification, where the two-class classification classifies between cataract and non-cataract cases, achieve accuracy rates of 85.98% and 94.83%, respectively. Another investigation proposed by He et al. [18] was to identify eye disorders using the ODIR database. As a classifier to give classification scores, a feature extractor, and a module to establish spatial correlation, these pre-trained models work as a classifier. The proposed network reaches an AUC of 93% and an F1-Score of 91.3%. Nevertheless, its suggested feature correlation module leads to an excessively high cost of 74.2 million parameters. Wang et al. use CNN and self-attention, called MBSaNet, to identify fundus diseases. MBSaNet outperforms existing techniques and achieves state-of-the-art performance with fewer parameters [19]. In [20], authors thoroughly study cataract classification strategies, both conventional machine learning and deep learning. This study investigates several ocular imaging modalities, including slit lamp and AS-OCT images and grading methods such as LOCS III and Wisconsin grading. In article [21], the authors propose a novel CNN, GraNet, that combines region-based integration and recalibration attention blocks for the categorization of nuclear cataracts from AS-OCT images. This highlights the importance of using clinical expertise [22, 23] as well as attention processes to increase classification efficacy. Both researches add to the refinement of more precise and effective techniques for cataract diagnosis and classification.

Ref	Year	Model	Result (validation metrics accuracy)	Inference
[24]	2024	Utilizes VGG19, ResNet50, DenseNet201, MIRNet, Inception V3, Xception, EfficientNet B0 models to develop CSDnet Framework.	97.24%	A new deep learning methodology, named CSDNet, is presented in the study for the identification of cataract states. The framework is engineered specifically for situations with restricted memory or storage capacity to accommodate such situations and be lightweight and flexible. It is able to learn representations from data without the need to increase number of trainable parameters drastically. The model achieves reasonable computational expense and average running time compared with other pre-trained models by employing smaller kernels and a smaller number of training parameters and layers.
[25]	2024	CNN	98.93	We propose a new framework, named CSDNet, to improve cataract state identification using deep learning methodology, which is presented in the study. The framework is designed specifically to be lightweight and flexible to accommodate a situation with restricted memory (or storage) capacity. It is able to learn representations well from data with the smallest amount of trainable parameters. Employing this with smaller kernels, reduced training parameters, and layers, the model's compute study article is based on the application of convolutional neural networks (CNN) in differentiating cataract-subjected eyes from normal eyes. In this regard, the feature extraction approach and hyperparameter optimization are proposed. We methodically explore the hyperparameter space to identify the configuration with the best efficiency of the classification algorithm compared to other pre-trained models in terms of national expenses and average running time.
[26]	2024	VGG19 and ResNet-50.	The ResNet-50-93.41%. The VGG19-91.13%.	The study article relates the use of CNN to differentiate eyes affected with cataracts from normal eyes. We achieve this by proposing a novel feature extraction approach and hyperparameter optimization. We search the hyperparameter space to identify the optimal setting, maximizing the classification algorithm efficiency.
[26]	2023	EfficientNetB0	91.49%	In this experimental methodology, a deep learning model is trained to classify cataracts strictly for early detection and timely intervention.

Ref	Year	Model	Result (validation metrics accuracy)	Inference
[27]	2023	VGG-19	95%	Deep learning algorithms, in particular VGG 19 model categorization, were done to categorize fundus images of eye diseases. We preprocess the dataset and augment it to deal with data imbalance.

The designs and hybrid methodologies adopted were varied to enhance the accuracy and efficiency of diagnostics. In particular, they [13] have introduced a hybrid deep learning system that employs VGG16 architecture combined with Siamese networks and Grad-CAM for interpretable prediction. To increase interpretability, the shown technique increased the interpretability measures, such as AUC and the Pointing Game, for explanation quality with saliency map evaluation. A hybrid CNN model was introduced by Chun-Ling [28] to detect cataracts where a fundus image is first segmented into many regions, and predictions of five different models are fused through a majority voting mechanism. The methodology presented in this novel helped increase the accuracy of cataract diagnosis. A LeNet-CNN model was used by Ganokratanaa et al. [29] to perform first cataract identification with an accuracy of 96% and benchmarked its performance against an SVM classifier. For image diagnosis, their findings showed that CNN-based approaches perform significantly better than the conventional SVM model, specifically for the medical domain, and Gurjot and Neha [30] used VGG19 architecture with transfer learning and used data augmentation and regularization methods to address this issue of overfitting of data. Using a DCNN model for the Kaggle dataset, Menaka et al. [31] obtained a detection accuracy of 92.78%.

The study that they primarily focused on was their preprocessing method, primarily to achieve textural information, which included the contribution of contrast, correlation, and entropy for medical image analysis. Compared with previous research, which successfully utilized CNNs and hybrid models in cataract detection, our study is new. We achieved high-efficiency computational performance using oneAPI, a technology that allows parallel processing. The resultant reduced model runtime makes our technique suitable for use in real-time clinical applications. By using CNN-SVM hybrids, feature extraction, and categorization are reconciled, negating the effects of CNN-only-based systems. Medical image analysis techniques of today often struggle with noise sensitivity, model complexity, and processing requirements. Through high-quality preprocessing using uniform scaling, normalization, and contrast enhancement, our method successfully mitigates the influence of noise on model performance. Our model architecture is intricate, but it is designed with precision in mind while also being feasible to execute. To achieve performance optimization, we took advantage of Intel oneAPI libraries to accelerate computational efficiency and reduce processing time. The following sections present in-depth analyses of these criteria (Table 1).

3 | Materials and Methods

A thorough explanation of the essential data collection, preprocessing stages, model architectures, oneAPI optimization, training protocols, and assessment metrics employed in our investigation can be found in this section.

3.1 | Data Collection

In this article, we use colour fundus pictures from ODIR collection (ODIR-2019, 2022) [32]. The data in the collection pertain to eight different classifications of illnesses or disorders. We only used normal fundus and cataract imageries for our investigation. The research consists of 500 standards and 594 cataract images extracted from color fundus photographs of patient's left and right eyes. Surgeons' diagnostic keywords accompany the images and offer helpful information on the clinical assessment. Figure 1 shows a few cataracts and normal photos extracted from the dataset. The dataset is intended to provide an authentic collection of patient data in as accurate a form as possible as collected by Shanggong Medical Technology Co. Ltd. from many hospitals and healthcare institutions in China. Images of both the left and right eyes are included to expand the range of the dataset and to increase the level of knowledge on eye health.

3.2 | Data Preprocessing

An extensive preprocessing pipeline was implemented on fundus pictures collected from various camera sources, including Kowa, Zeiss, and Canon, before model training to standardize the input data and improve model performance. Discrepancies in dimension were eliminated, and the deep learning model was provided with equal input size by reducing all photos to a standard resolution of 224×224 pixels, which decreased computational complexity and memory demands [33]. Resizing is used to aid in balancing the preservation of critical picture features with reasonable resolution for efficient processing.

Pixel intensity distributions were normalized by scaling pixel values to [0, 1] interval. The input data size was standardized in this phase to enhance model convergence, thus reducing the chance of gradient disappearing or ballooning during backpropagation. In addition, histogram equalization was used to enhance contrast to increase the visibility of features in retinal pictures and increase the model's ability to detect subtle cataract-related anomalies by dispersing pixel intensity values.

The preprocessing method, which is done very carefully with scaling, normalization, and contrast enhancement, has been able to mitigate the inherent discrepancy in imaging conditions across different devices and prepare the data for robust feature extraction and efficient model training.

TABLE 1 | Cataract classification performance metrics.

Model	Dataset	Accuracy	Precision	Recall	F1 score
MobileNet	Train	0.9921	0.9862	0.9845	0.9853
	Test	0.9836	0.9754	0.9744	0.9769
VGG-19	Train	0.8031	0.6887	0.6192	0.6528
	Test	0.7697	0.6508	0.5478	0.6965
ResNet-50	Train	0.9044	0.8194	0.8346	0.8269
	Test	0.8777	0.7698	0.7907	0.7734
Inception-V3	Train	0.9264	0.7856	0.7926	0.7890
	Test	0.9246	0.6797	0.7853	0.6523

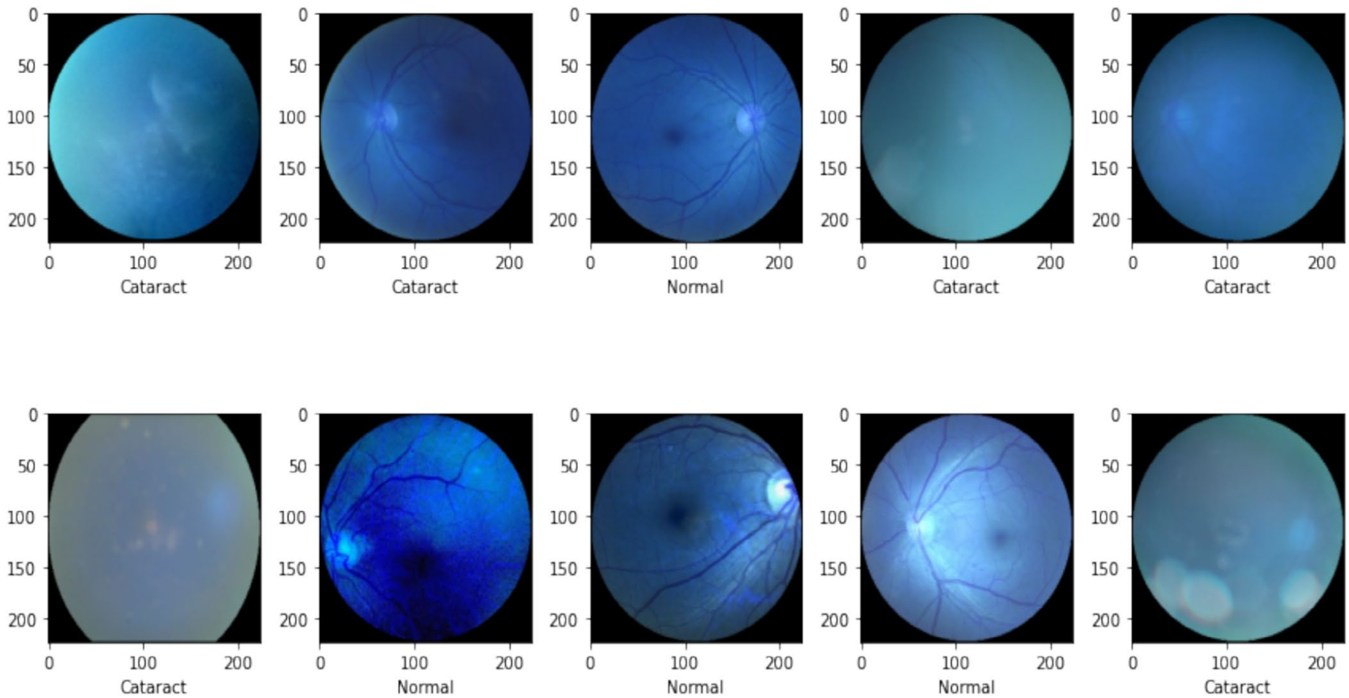

FIGURE 1 | Image samples from the dataset.

3.3 | Proposed Architecture

The proposed architecture recommends CNNs-SVMs with restricted capacity that are supported by architecturally architected hierarchies of feature extraction and refinement for cataract classification. A deep CNN is first used to extract highly patterned images from the fundus. After each of the five convolutional layers of the CNN architecture, we included rectified linear unit (ReLU) and batch normalization. Figure 2 shows the overall architectural design. The first layer has 64 initial filters of size 3×3, and a subsequent layer has 128 of those filters. The breakdown of exact down sampling for spatial dimensions can occur by putting max-pooling layers with a pool size of 2×2 in between each of two successive convolutional layers. To further increase discriminative strength, we combine these features with an SVM that serves as a feature selector. This method aims to combine the benefited features of CNN and SVM in cataract classification and categorization, and it is the most comprehensive and efficient one among them. In addition, architecture is optimized and implemented inside of the oneAPI environment using the oneDNN library so that the architecture is plug-and-play for various hardware architectures.

An analysis of four distinct classification models, MobileNet, ResNet, VGG19, and Inception, was conducted to classify cataract images based on features extracted by the hybrid CNN-SVM architecture. To achieve better performance and efficiency, the oneDNN library is used to optimize each model independently as part of the oneAPI environment. The features produced from the CNN-SVM hybrid are fed into separate classification models to make them suitable for autonomous classification. This comparative study is helpful for a detailed characterization of the strengths and capabilities of each model and for their utility in cataract detection. A comprehensive evaluation of results from

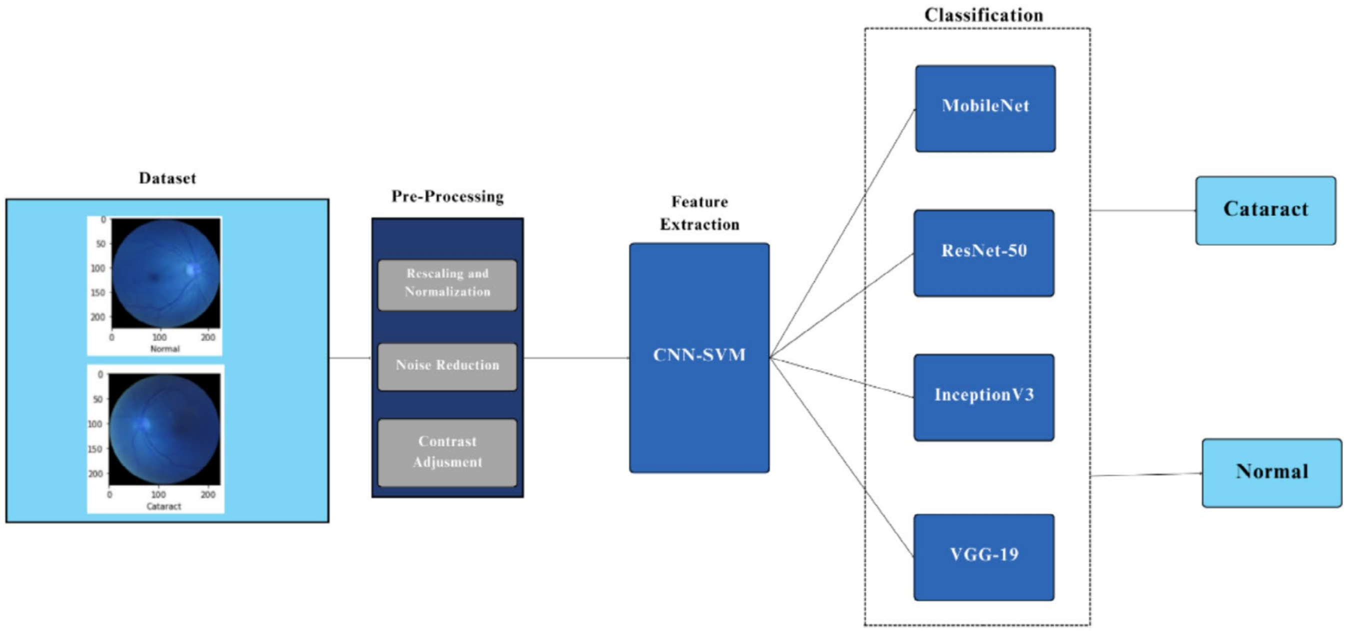

FIGURE 2 | Proposed hybrid architecture.

each model is performed with respect to terms such as precision, recall, accuracy, and f1 score to judge how effective each model is for categorizing cataracts. It is a comprehensive analysis that helps determine the model yielding the most dependable diagnoses in real-world settings.

3.4 | CNN-SVM Feature Extraction

First is the use of a carefully architected CNN architecture that constitutes the first part of the feature extraction process. After each of the five convolutional layers in the CNN, batch normalization and ReLU are applied [34]. After 128 3×3 kernel-sized filters are done, the first layer contains a filter size of 64 3×3 kernels. Each spatial down-sampling of 2×2 is done by strategically placing 2×2 max pooling layers. Two fully linked layers, both containing 512 neurons, ReLU activation, and batch normalization, follow before the final softmax layer. At the same time, it has the potential to improve the features through simultaneous integration of SVM. The use of SVM as a feature selector instead of a classifier [35] makes the SVM a powerful discriminator of CNN's features. To refine features, the SVM's regularization parameter and RBF kernel are adjusted. Through the hybrid architecture's ability to integrate the SVM's pattern recognition strength with CNN's ability of representation, a feature extraction method with the explicit goal of cataract classification is developed by using the architecture's capacity to combine them. Our feature extraction stage is built on an integrated CNN-SVM framework, showing how the common architectures in machine learning and deep learning can be presented in a unified form.

3.5 | Classification Models

For cataract diagnosis in fundus pictures, four unique CNN models (MobileNet, ResNet-50, VGG-19, and Inception-V3) are delineated with their topologies in this section. In terms of lightweight architecture, mobile, and embedded applications, MobileNet is a proposed architecture that can provide efficient performance for fewer computing resources. By the deep design and residual connections that have alleviated the vanishing gradient problem, ResNet-50 is well equipped to train the wide networks and learn high-level information from complex medical pictures. VGG is a deeper level of variation of VGG, but it has a very simple, uniform architecture with small convolutional filters to get subtle visual info. Inception-V3 can learn multi-scale features with many filter sizes concurrently, something beneficial for recognizing subtle patterns in cataract pictures. We select each model to address the challenges related to medical picture classification and offer varied approaches to cataract identification.

3.5.1 | MobileNet

MobileNet is CNN's mobile architecture. In Figure 2, we see that the core framework of the design is based on separable convolutions in depth. A specific family of complexities, namely factorized complexities, are complexities that transform a conventional complexity into a depth-wise complexity. Pointwise complexity is a 1×1 complexity. The depth-wise separable convolutions are followed by batch normal and ReLU operations after the depth-wise and pointwise layers, as shown in Figure 1. In the interim, the model includes two fundamental global hyperparameters to balance the trade-off between latency and accuracy [36]. In Equation (1), we can mathematically express the depth-wise separable convolution.

$$y = Pointwise(Depthwise(x)) \tag{1}$$

Furthermore, pointwise convolution is applied after the application of depth-wise convolution, which is expressed in Equation (2).

$$Y_{i,j,k} = \sum_{l=1}^{C'} + X_{i,j,l}.\, K_{l,k} \tag{2}$$

3.5.2 | VGG-19

VGG19 is an expansion of VGG16 architecture and a well-known CNN, which is known for its simplicity. In its architecture, VGG19 maintains the 2×2 pooling size and the 3×3 convolutional filter size of 2×2 throughout. A max-pooling layer inserted in front of a VGG19 can be written as a series of conv1, conv2, and conv3, as described [37]. Equation (3) represents the convolution layer mathematically.

$$Y_i = \sigma\left(W_{i*}X_{i-1} + b_i\right) \tag{3}$$

where Y_i is the output feature map, σ is the ReLU, X_{i-1} is the feature map, and W_i *and* b_i are the weights and biases. It consists of nineteen layers in total. Sixty-four convolutional layers×2, 128 strain neural network convolution×2, 256 filter convolution sheet×4, 64 convolutional locations×8, full engaged bed×2, and 1 output sheet [38]. We reduced ImageNet Gainsay's original 1000 product nodes to two to meet our requirements. Two entirely neighboring layers were used with a 0.5 dropout.

3.5.3 | ResNet-50

ResNet-18 and ResNet-50 are two versions of ResNet with sophisticated characteristics and broad applications in the field of picture categorization [39]. ResNet-50 model contains 48 coiled layers, 1 MaxPool layer, an intermediate pool layer, and a ReLU activation function. ResNet resolves the overfitting problem that occurs when the input is fed into a front-fed neural network, which has hidden layers with unreasonably many parameters. Hence, network training is linked with a fixed amount of input. Using linear techniques, a handful of specialized neurons in this hidden layer is sufficient [40] to fill the data. ResNet tries to learn from data and prevent the growth of layers by creating shortcut connections by skipping one or more levels of network topology to avoid mistakes during the training process. Equations (3) and (4) provide three equations defining the initial residual unit.

$$y_1 = h\left(x_1\right) + f\left(x_1, w_1\right) \tag{3}$$

$$x_{l+1} = f\left(y_1\right) \tag{4}$$

x_l is the input, W_l is the weights, and f is the residual function.

3.5.4 | Inception-V3

The Inception V3 model contains six convolution layers, two maxpool layers, a linear layer, a softmax layer, three inceptions of module A, five inceptions of module B, and two inceptions of module C. In an earlier version of Inception, the spatial layer was more significant, e.g., 5×5. This convolution will require much computing. To decrease it, two (3×3) convolution layers are used instead of the (5×5) convolution layer used in Inception V3 [41]. This will reduce computational costs because fewer parameters are used. Inception V3 also pays attention to the feature factorized convolutions to reduce the computational complexity. It uses batch normalization and ReLU activation functions for every convolutional layer to improve training stability and create non-linearity. Fully connected layers and classification-focused softmax output layer are at the end of the network. Inception V3 has been shown to be [42] especially well suited for challenging picture identification tasks since it can perform efficient feature extraction due to its unique design.

3.6 | One API Optimization

We optimize our four models of classification (ResNet, VGG19, InceptionV3, and MobileNet) using the oneAPI toolkit to utilize a range of hardware architectures better. Taking advantage of the smooth integration of the oneDNN library into the oneAPI environment, we guarantee that every model is optimized for peak performance on both CPUs, GPUs, and other accelerators. Our method enables the parallel execution of important operations, allowing it to speed up both training and inference for all models by taking advantage of data parallelism. This hardware-agnostic property of OneAPI enables the deployment effectively across a range of platforms with minimal changes to the model code [43]. In addition, basic linear algebra subprogram operations are optimized, and profiling tools from the oneAPI are applied to reduce bottlenecks and improve the overall performance of each classification model.

4 | Evaluation Metrics

In this study, standard classification measures (accuracy, precision, recall, and F1 score) were used to evaluate the model performance in the binary classification of cataract-affected and normal fundus pictures. We used a predetermined probability threshold of 0.5 by labeling predictions equal to or beyond 0.5 as cataract-affected and those under it as usual.

That is, accuracy measures the ratio of correctly classified events in both group's overall events. The precision measures the model's ability to pick out cataract cases from all cases projected as positive, hence reducing false positives. Sensitivity (recall) is the model's ability to select all true positives from all true positives correctly and is a measure of model's ability to reduce false negatives. The F1 score combines accuracy and recall in one statistic, which benefits from balanced performance measurement when there is a trade-off between these two criteria.

It was ensured that these assessment measures were applied uniformly to all models in this comparison study. It was simply for its simplicity as well as interpretability, using a simple set threshold of 0.5 and guaranteeing consistency in classification predictions. The future may look at dynamic or data-driven threshold adjustment for better classification performance.

5 | Results and Discussion

Based on findings from the assessment, it is observed that the four models considered in this study have particular performance traits that can be advantageous or disadvantageous. As we can see, MobileNet demonstrates the highest accuracy at

98.36%, precision at 97.54%, recall at 97.44%, and F1 score at 97.69%. These results demonstrate that while it accurately detects cataracts, it does so with an appropriate level of accuracy as well as memory. As with the higher accuracy of the model, its possibility of overfitting might pose that its generalizability is restricted on novel data. It is also notable that MobileNet is also very efficient computing: 1827 s of runtime is feasible for real-time applications.

After that comes ResNet-50 with an accuracy of 87.77%, accuracy of 76.98%, recall of 79.07, and F1 score of 77.34, indicating that this model is capable of extracting delicate and significant features from cataract photos. However, it provided highly commendable accuracy; its precision and recall suggest room for improvement, specifically in reducing false positives and certainly guaranteeing dependable detection. The longer 2423 s of training show us that ResNet-50 is more profound and, therefore, has more processing needs in computer resources, which could make it unsuitable for real-time.

The accuracy, precision, recall, and F1 score for VGG19 are 76.97%, 65.08%, 54.78%, and 69.65%, respectively. Even though VGG19 underperforms on the performance measures, it still learns and detects key traits reasonably well. However, challenges like class imbalance or overfitting on our model may lead to lower outcomes for cataract diagnosis. However, the accuracy and resilience are compromised, but the computational speed is exceptionally high, with training performed within 476 s using VGG19, thus making it an excellent choice when speed dominates.

Although the accuracy of 92.46% of inception-V3 in cataract classification ranks highly, its 78.53% recall and 67.97% precision, along with 65.23% F1 score, suggest the probable false positives and false negatives. The results show that Inception V3, although accurate, requires further refinement so that detection rates for positive and negative instances are achieved at equilibrium. However, despite these problems, Inception V3 represents an appealing option, given its limited training duration of 1345 s, which strikes a nice compromise between accuracy and computing efficiency.

All these models are equally accurate and performant, but we need to assess the trade-off between precision, recall, and F1 score when deciding what model to deploy. Although the accuracy and the balancing metrics of MobileNet are particularly impressive, more evaluation is needed to demonstrate its robustness and generality. In selecting the model, the given particular demands of the application, such as the speed, the accuracy, or the capacity to manage complicated aspects in cataract identification, must be guided.

6 | Conclusion and Future Study

We then present the results of our extensive evaluation and optimization of four cataract classification models, achieving 98.36%, with MobileNet showing the best results. On different types of hardware configurations, we showed how flexible our methodology was with the incorporation of oneAPI enhancements. We found that ResNet-50 achieved an accuracy of 87.77%, which also performed quite well, but VGG19 and InceptionV3 require improvement. Future work should focus on refining the identified regions of need for improvement, increasing the size of the datasets for better generalization, introducing Explainable AI (XAI) methods, and studying group methods for additional model resilience. Depending on the ongoing integration of the latest and the best design and optimization strategies, the sustained effectiveness of cataract classification models in practical clinical applications will critically depend on it. The efficacy of the models is robust in cataract identification, but implementation in real clinical environments may meet obstacles. Model dependability can be dependent on differences in imaging apparatus, patient characteristics, and ambient conditions. A simple example is that although MobileNet might be overfitting driven and hence less generalizable, ResNet-50 could be too long in its processing time to be real-time. VGG19 may perform diminished accuracy and precision, leading to misclassifications and false positives, and Inception V3 may have reduced precision, resulting in false positives. More training on various datasets and additional fine-tuning of the model is needed to increase resilience to real-world settings and to increase clinical relevance.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are openly available in ODIR dataset at https://www.kaggle.com/datasets/andrewmvd/ocular-disease-recognition-odir5k.

References

1. J. Chen, Y. Zhu, Z. Li, et al., “Global Impact of Population Ageing on Vision Loss Prevalence: A Population-Based Study,” *Global Transitions* 6 (2024): 28–36.

2. B. M. Sztonyk, “Adversarial Growth in Persons With Vision Impairment,” in *The Routledge International Handbook of Posttraumatic Growth* (Routledge, 2024), 380–392.

3. T. D. Nguyen, D. T. Le, J. Bum, S. Kim, S. J. Song, and H. Choo, “Retinal Disease Diagnosis Using Deep Learning on Ultra-Wide-Field Fundus Images,” *Diagnostics* 14, no. 1 (2024): 105.

4. S. Al-Fahdawi, A. S. Al-Waisy, D. Q. Zeebaree, et al., “Fundus-DeepNet: Multilabel Deep Learning Classification System for Enhanced Detection of Multiple Ocular Diseases Through Data Fusion of Fundus Images,” *Information Fusion* 102 (2024): 102059.

5. M. Menaka, V. P. Kolanchinathan, A. Yogaraj, et al., Cataracts Detection Using Convolution Neural Network 2024.

6. M. Thanoon and S. A. Dawwd, DCNN For Cataract Disease Detection Based on Model Parallelism 2024.

7. I. Khan, W. Akbar, A. Abdullah, et al., Enhancing Ocular Health Precision: Cataract Detection Using Fundus Images and ResNet-50 2024.

8. U. Utkarsh, V. Churavy, Y. Ma, et al., “Automated Translation and Accelerated Solving of Differential Equations on Multiple GPU Platforms,” *Computer Methods in Applied Mechanics and Engineering* 419 (2024): 116591.

9. S. I. Abir, S. Shoha, and S. A. Al Shiam, “Deep Neural Networks in Medical Imaging: Advances, Challenges, and Future Directions for Precision Healthcare,” *Journal of Computer Science and Technology Studies* 6, no. 5 (2024): 94–112, https://doi.org/10.32996/jcsts.2024.6.5.9.

10. S. Bano, B. Agrawal, and S. Bhardwaj, “Eye Disease Detection Using CNN, ResNet, and VGG16,” in *Universal Threats in Expert Applications and Solutions. UNI-TEAS 2024. Lecture Notes in Networks and Systems*, ed. V. S. Rathore, V. Piuri, R. Babo, and S. Karthik, vol. 1007 (Springer, 2024), https://doi.org/10.1007/978-981-97-5146-4_22.

11. R. Raman, V. Kumar, and D. Saini, Advancements in Diabetic Retinopathy and Cataract Identification Through Deep Learning 2024.

12. S. P. Pingat, D. H. Kulkarni, and P. N. Mahalle, “Convolutional Neural Networks (CNNs) for Detection of Eye Diseases in an Effective Manner,” *Panamerican Mathematical Journal* 34, no. 4 (2024): 361–371.

13. J. Olaniyan, D. Olaniyan, and I. C. Obagbuwa, “Transformative Transparent Hybrid Deep Learning Framework for Accurate Cataract Detection,” *Applied Sciences* 14, no. 21 (2024): 10041, https://doi.org/10.3390/app142110041.

14. S. S. Mahmood, S. Châabouni, and A. Fakhfakh, “Improving Automated Detection of Cataract Disease Through Transfer Learning Using ResNet50,” *Engineering, Technology & Applied Science Research* 14, no. 5 (2024): 17541–17547, https://doi.org/10.48084/etasr.8530.

15. N. Gour and P. Khanna, “Multiclass Multilabel Ophthalmological Disease Detection Using Transfer Learning-Based Convolutional Neural Network,” *Biomedical Signal Processing and Control* 66 (2021): 102329, https://doi.org/10.1016/j.bspc.2020.102329.

16. Y. Li, J. Xu, X. Zhang, and C. Liang, “A Novel Deep Learning-Based Approach for Diagnosing Diabetic Retinopathy Using Fundus Images,” *IEEE Journal of Biomedical and Health Informatics* 25, no. 5 (2021): 1515–1526.

17. L. Cao, H. Li, Y. Zhang, L. Zhang, and L. Xu, “Hierarchical Method for Cataract Grading Based on Retinal Images Using Improved Haar Wavelet,” *Information Fusion* 53 (2020): 196–208.

18. J. He, C. Li, J. Ye, Y. Qiao, and L. Gu, “Multilabel Ocular Disease Classification With a Dense Correlation Deep Neural Network,” *Biomedical Signal Processing and Control* 63 (2021): 102167.

19. K. Wang, C. Xu, G. Li, Y. Zhang, Y. Zheng, and C. Sun, “Combining Convolutional Neural Networks and Self-Attention for Fundus Disease Identification,” *Scientific Reports* 13, no. 1 (2023): 76.

20. X. Q. Zhang, Y. Hu, Z. J. Xiao, J. S. Fang, R. Higashita, and J. Liu, “Machine Learning for Cataract Classification/Grading on Ophthalmic Imaging Modalities: A Survey,” *Machine Intelligence Research* 19, no. 3 (2022): 184–208.

21. X. Zhang, Z. Xiao, H. Fu, et al., “Attention to Region: Region-Based Integration-And-Recalibration Networks for Nuclear Cataract Classification Using AS-OCT Images,” *Medical Image Analysis* 80 (2022): 102499.

22. X. Zhang, Z. Xiao, B. Yang, X. Wu, R. Higashita, and J. Liu, “Regional Context-Based Recalibration Network for Cataract Recognition in AS-OCT,” *Pattern Recognition* 147 (2024): 110069.

23. X. Zhang, Z. Xiao, X. Wu, et al., “Nuclear Cataract Classification in Anterior Segment OCT Based on Clinical Global–Local Features,” *Complex & Intelligent Systems* 9, no. 2 (2023): 1479–1493.

24. P. L. Lahari, R. Vaddi, M. O. Elish, V. Gonuguntla, and S. S. Yellampalli, “CSDNet: A Novel Deep Learning Framework for Improved Cataract State Detection,” *Diagnostics* 14, no. 10 (2024): 983.

25. M. Denisha, “Deep Learning-Based Classification of Cataract Infected Eyes and Normal Eyes Using Convolutional Neural Network,” in *2024 International Conference on Expert Clouds and Applications (ICOECA)* (IEEE, 2024), 792–798.

26. S. Kamepalli, N. Veeranjaneyulu, and J. D. Bodapati, “Exploring Deep Learning Techniques for Automated Cataract Detection in Digital Camera Imagery,” in *2024 International Conference on Advances in Modern Age Technologies for Health and Engineering Science (AMATHE)* (IEEE, 2024), 1–6.

27. S. S. Mahmood, S. Chaabouni, and A. Fakhfakh, “A New Technique for Cataract Eye Disease Diagnosis in Deep Learning,” *Periodicals of Engineering and Natural Sciences* 11, no. 6 (2023): 14–26.

28. L. Chun-Ling, Cataract Detection Using Hybrid CNN Model on Retinal Fundus Images 2023.

29. T. Ganokratanaa, M. Ketcham, and P. Pramkeaw, “Advancements in Cataract Detection: The Systematic Development of LeNet-Convolutional Neural Network Models,” *Journal of Imaging* 9, no. 10 (2023): 197, https://doi.org/10.3390/jimaging9100197.

30. K. Gurjot and S. Neha, Cataract Disease Detection Using VGG 19 Deep Learning Architecture 2024.

31. M. Menaka, V. P. Kolanchinathan, A. Yogaraj, N. Venkata Abhiram, V. Kirubesh, and S. Hariharan, “Cataracts Detection Using Convolution,” *Neural Networks* 2 (2024): 883–890, https://doi.org/10.1109/iccpct61902.2024.10672654.

32. Ocular Disease Recognition, “Kaggle,” 2020, https://www.kaggle.com/datasets/andrewmvd/ocular-disease-recognition-odir5k.

33. H. Ding, Q. Huang, and A. Alkhayyat, “A Computer-Aided System for Skin Cancer Detection Based on a Developed Version of the Archimedes Optimization Algorithm,” *Biomedical Signal Processing and Control* 90 (2024): 105870.

34. G. E. Rao, B. Rajitha, P. N. Srinivasu, M. F. Ijaz, and M. Woźniak, “Hybrid Framework for Respiratory Lung Diseases Detection Based on Classical CNN and Quantum Classifiers From Chest X-Rays,” *Biomedical Signal Processing and Control* 88 (2024): 105567.

35. Y. Pan, J. Qin, Y. Hou, and J. J. Chen, “Two-Stage Support Vector Machine-Enabled Deep Excavation Settlement Prediction Considering Class Imbalance and Multi-Source Uncertainties,” *Reliability Engineering & System Safety* 241 (2024): 109578.

36. F. J. Xavier, “ODMNet: Automated Glaucoma Detection and Classification Model Using Heuristically-Aided Optimized DenseNet and MobileNet Transfer Learning,” *Cybernetics and Systems* 55, no. 1 (2024): 245–277.

37. M. M. Khalid and O. Karan, “Deep Learning for Plant Disease Detection,” *International Journal of Mathematics, Statistics, and Computer Science* 2 (2024): 75–84.

38. O. C. Do, C. M. Luong, P. H. Dinh, and G. S. Tran, “An Efficient Approach to Medical Image Fusion Based on Optimization and Transfer Learning With VGG19,” *Biomedical Signal Processing and Control* 87 (2024): 105370.

39. D. G. Kumar and S. Chaudhari, Recxvgg: An Efficient Method for Remote Sensing Image Scene Classification Using Vgg19 and Xgboost.

40. J. F. Chen, Y. T. Liao, and P. C. Wang, “Development and Deployment of a Virtual Water Gauge System Utilizing the ResNet-50 Convolutional Neural Network for Real-Time River Water Level Monitoring: A Case Study of the Keelung River in Taiwan,” *Watermark* 16, no. 1 (2024): 158.

41. L. Kang, G. Huan, Z. Xuejin, et al., “Deep Learning and Thermographic Imaging Method for Thermal Comfort Prediction in Different Genders,” *International Journal of Thermal Sciences* 197 (2024): 108804.

42. N. Bansal and A. Jain, “Word Recognition From Indian Sign Language Using Transfer Learning Models and RNN Classifier,” *International Journal of Intelligent Systems and Applications in Engineering* 12, no. 9s (2024): 182–189.

43. A. B. Ramakrishnan, S. K. Vasudevan, T. S. Murugesh, and S. R. Pulari, “Enhancing Multiclass Classification of Knee Osteoarthritis Severity Grades Using oneDNN,” *International Journal of Bioinformatics Research and Applications* 19, no. 3 (2023): 200–212.