Journal of Information Systems Engineering and Management
2025, 10(53s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Evaluating Deep Learning and Traditional Approaches in the Automated Classification of Retinal Diseases

Andrae D. Bretaña[1], Daniela J. Comapon[2], Hermoso J. Tupas Jr.[3], Maureen M. Villamor[4]
[1] *Student, College of Information and Computing, University of Southeastern Philippines Davao City, Philippines*
[2] *Student, College of Information and Computing, University of Southeastern Philippines Davao City, Philippines*
[3] *Faculty, College of Information and Computing, University of Southeastern Philippines Davao City, Philippines*
[4] *Faculty, College of Information and Computing, University of Southeastern Philippines Davao City, Philippines*

ARTICLE INFO

Received: 30 Dec 2024

Revised: 19 Feb 2025

Accepted: 27 Feb 2025

ABSTRACT

Retinal diseases such as diabetic retinopathy, glaucoma, and age-related macular degeneration are leading causes of preventable blindness worldwide. Early and accurate diagnosis is essential, but manual interpretation of retinal images remains time-consuming and prone to variability. This study evaluates deep learning models, DenseNet121 and DenseNet201, against traditional machine learning models like Support Vector Machine (SVM) and Random Forest (RF) for the automated classification of retinal diseases using fundus images. The dataset consists of over 21,000 images across ten ocular diseases. DenseNet201 achieved the highest validation accuracy (85.77%), outpacing traditional models (SVM: 65.04%, RF: 64.87%). This research highlights the potential of AI in ophthalmology, demonstrating its ability to automate disease detection and improve diagnostic accuracy.

Keywords: Retinal Disease Classification, Deep Learning, DenseNet

INTRODUCTION

Retinal diseases, including diabetic retinopathy (DR), glaucoma, age-related macular degeneration (AMD), and myopia, are some of the common causes of visual impairment that are prone to blindness globally. As stated by the World Health Organization (2019) projects, there are over one billion cases of people with untreated or even preventable vision impairments. In which diabetic retinopathy is responsible for approximately 4.8 percent of global blindness cases (Flaxman et al., 2017).

Despite the advancements in ophthalmic imaging, diagnostic processes still rely heavily on eye specialists' manual diagnosis of retinal fundus images. It is time-consuming, prone to inter-observer variability, and predominantly unavailable in low-resource or resource-poor settings (Abdullah et al., 2024). These limitations demonstrate the need for scalable, accurate, and automatic diagnostic systems to support early disease detection and relieve clinical professionals of the burden.

Artificial Intelligence (AI) falls under the category of the emerging trends of the present day, which has been considered a breakthrough solution for analyzing, classifying, and diagnosing a range of medical images. Furthermore, AI has been shown to act as a substitute because it can at least match or exceed specialists' capacity to identify dermatological, radiological, and ophthalmologic diseases (Esteva et al., 2017; Gulshan et al., 2016; Ting et al., 2017) as demonstrated by the fact that these models are typically built on top of Convolutional Neural Networks (CNNs), which are known to be highly efficient at analyzing retinal images since they can quickly learn complex hierarchical patterns directly extracted from raw pixel images from the raw pixel image, without one of the most difficult and time-consuming tasks: feature engineering (Sengupta et al., 2019).

Today, CNNs are widely used in diabetic retinopathy, glaucoma, and AMD diagnosis (Rajalakshmi et al., 2018). Furthermore, traditional ML methods such as SVMs and RFs have been used widely in detecting retinal disease with inputs such as optic disc boundaries, blood vessel morphology, and texture descriptors (Al Marouf et al., 2023). Even

Journal of Information Systems Engineering and Management
2025, 10(53s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

though these models exhibit interpretability and low complexity in terms of computation, they lack flexibility and do not scale up when dealing with numerous diverse real-world datasets. Conversely, deep networks like DenseNet, ResNet, and VGG16 perform better in large image sets because they can learn deep visual representations (Iqbal et al., 2022; Abdullah et al., 2024).

Artificial intelligence advancements are revolutionizing medical diagnosis, particularly in ophthalmology. One key advancement is RETFound, a self-supervised model that can generalize across a range of tasks without further training following training on 1.6 million retinal images. This highlights the progress in AI when it comes to diagnosing systemic diseases, including Alzheimer's disease, cardiovascular risk, or eye disease (Li et al., 2023; Poplin et al., 2018). These advances also endorse the advantages of AI-assisted fundus imaging in improving diagnostic functions.

Although these encouraging advancements were made, deep learning models still present significant challenges to their clinical application. Some of the issues that have not been addressed yet (Shorten and Khoshgoftaar, 2019; Ting et al., 2019) are sensitivity to hyperparameter optimization, overfitting, and comprehending judgments of AI. Additionally, normal fundus images closely mimic specific examples, such as early-stage glaucoma, resulting in more difficult categorization and using models that can detect tiny variations required.

This paper, Evaluating Deep Learning and Traditional Approaches in the Automated Classification of Retinal Diseases, attempts to overcome these challenges by comparing the performance of two deep learning structures (DenseNet121, DenseNet201) against two conventional machine learning models (Support Vector Machine, Random Forest). Using a wide set of retinal fundus images for analysis, researchers will assess training behavior, how precise the training is, and how resistant the training becomes, with special emphasis on conditions with similar visual characteristics. While this research helps to build on the pervasiveness of AI-based tools in ophthalmology by making available informative insights into the model's performance and optimization, it can be concluded that the research has made a valuable contribution to the growing trends of AI-based tools in ophthalmology. It offers a framework for selecting and optimizing models to improve the detection of retinal diseases in real-world clinical environments, which is more important.

OBJECTIVES

The principal aim of this study is to compare the outcomes of deep learning models DenseNet121 and DenseNet201 with traditional machine learning methods SVM and RF when classifying retinal conditions using fundus images. The goal of the study is:

1. To evaluate and compare the classification accuracy and robustness of deep learning models (DenseNet121 and DenseNet201) versus traditional machine learning models (SVM and Random Forest) for retinal disease detection using fundus images.
2. To investigate the effectiveness of fine-tuning pre-trained DenseNet architectures on a large and diverse retinal disease dataset to improve disease classification performance.
3. To analyze model performance using key metrics such as accuracy, F1-score, and confusion matrices, focusing especially on challenging retinal diseases with subtle visual differences.

METHODS

Using color fundus images, the study contrasts deep learning's efficacy architectures, specifically the DenseNet-121 and DenseNet-201, with conventional machine learning models, such as Random Forest and Support Vector Machine (SVM), for the classification and identification of retinal disorders. To ascertain how well the models could correctly identify different ocular conditions, a carefully chosen dataset of photos of eye diseases was used to train and evaluate them.

The Eye Disease Fundus Image Dataset used in this study contains 21,577 images, including 16,242 augmented images. The dataset was provided by Anwara Hamida Eye Hospital and BNS Zahrul Haque Eye Hospital in Bangladesh. The images are divided into ten classes: nine retinal diseases (e.g., Diabetic Retinopathy, Glaucoma, Retinal Detachment) and one class representing healthy eyes.

Journal of Information Systems Engineering and Management
2025, 10(53s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Table 1. Ocular Diseases Dataset

Class Label	Description
Retinitis Pigmentosa	A genetic disorder causing progressive retinal degeneration and night blindness
Retinal Detachment	Separation of the retina from the underlying tissue, potentially leading to vision loss
Pterygium	Benign growth of fibrovascular tissue on the conjunctiva that can extend onto the cornea
Myopia	A refractive error in which distant objects appear blurry due to elongation of the eyeball
Macular Scar	Scarring or damage to the macula, affecting central vision, often due to trauma or disease
Glaucoma	A group of eye conditions that damage the optic nerve, commonly associated with elevated intraocular pressure
Disc Edema	Swelling of the optic disc, typically caused by increased intracranial pressure or inflammation
Diabetic Retinopathy	Retinal blood vessel damage caused by prolonged high blood sugar levels in diabetic patients
Central Serous Chorioretinopathy	A retinal condition involving fluid accumulation under the retina, leading to distorted vision
Healthy Eye	Fundus images exhibiting no visible signs of retinal or optic nerve disease

The dataset includes nine retinal disease classes and one class representing images of healthy eyes, all of which were both annotated by medical experts to ensure the clinical credibility and validity. The description of each ocular class is presented in **Table 1**.

Data preprocessing steps included resizing the images to 224 × 224 pixels, normalizing pixel intensity values to a range of 0-1, and applying various data augmentation techniques such as rotation, flipping, and zooming to address class imbalance and improve model generalization. These transformations helped to prevent overfitting and ensured diversity in the training data.

Deep Learning Models DenseNet121 and DenseNet201, two CNN-based models with dense connectivity, were selected for their ability to learn hierarchical features efficiently. Both models were initialized with pre-trained ImageNet weights and fine-tuned on the retinal fundus images.

Traditional Models Support Vector Machine (SVM) and Random Forest (RF) were selected as baseline models. Support Vector Machine was configured with a radial basis function (RBF) kernel, while RF was configured with 100 decision trees using Gini impurity for splitting.

Table 2. DenseNet's Parameters

Model	Learning Rate	Epochs	K-Fold
DenseNet201	0.01, 0.0001	10	5
DenseNet121	0.01	10	5

Journal of Information Systems Engineering and Management
2025, 10(53s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

On Table 2. Both DenseNet-121 and DenseNet-201 were both used and implemented with the Keras deep learning libraries, but with TensorFlow as the backend. The eye disease dataset was used to fine-tune the models after they were initialized using pre-trained weights on ImageNet. The table below lists all crucial parameters and training configurations

Table 3. Training Configurations

Parameter	Training Configurations
Batch Size	32
Optimizer	Adam
Loss Function	Categorical Cross-Entropy
Early Stopping	Implemented to prevent overfitting based on validation loss
Callbacks	ModelCheckpoint and ReduceLROnPlateau were used for learning rate scheduling and checkpointing the best performing models.

Based on the **Table 3**. The researchers assessed each network's sensitivity to changes in optimization parameters by using different learning rates for each DenseNet model. Fivefold cross-validation was used to train the two models to minimize variance and ensure high-performance metrics.

Correct hyperparameters were configured for each classification model by prior research and then fine-tuning. The choice of these parameters was to encourage model performances combined with maintaining the consistency of the model between the evaluations.

Table 4. Support Vector Machine (SVM) Parameters

Parameter	Value
Kernel	RBF
Cross-validation folds	5

In **Table 4,** the researchers outlined the parameters for the Support Vector Machine (SVM) model. The model uses a Radial Basis Function (RBF) kernel, which is particularly effective for classifying complex patterns in retinal images. In order to assess the stability and generalization of the model, 5-fold cross-validation was applied, this is to be able to ensure that the results are not overly dependent on any single subset of the data. This approach is vital in providing a more robust performance evaluation across various training and validation sets

Table 5. Random Forest Parameters

Parameter	Value
Number of trees (estimators)	100
Max depth	Auto (expanded until pure or all leaves contain <2 samples)
Cross validation folds	5

Table 5 shows the parameters of the Random Forest (RF) model. This model relies on 100 decision trees and by averaging what each tree predicts, the accuracy of predictions improves. The depth of the trees is left to “Auto,” so

every branch will expand until either the leaves are pure or include fewer than two pieces. The arrangement helps minimize data fitting to the model and keep accuracy high. A 5-fold cross-validation was performed on the RF model to make its performance assessment dependable and consistent with different datasets.

Fine-tuning refers to the adjustment or minor changes in a pre-trained model's parameters to render it suitable for a specific task or domain, as Bergmann (2024) defines. It is an inherent aspect of transfer learning, which retains the essential information learned from large data sets while allowing models to be adapted for more specific uses. The approach suits domain-specific or sparse data, e.g., medical imaging.

In this research, the researchers adapted the deep learning models DenseNet-121 and DenseNet-201, belonging to convolutional neural network architectures pre-trained initially on the ImageNet dataset. These models were employed to classify the retinal disease using the above-described curated eye disease image dataset.

The researchers can teach the unique visual features of retinal diseases using the taught models thanks to this fine-tuning technique. This greatly enhanced the models' capacity to discriminate between diseases such as diabetic retinopathy, central serous chorioretinopathy, and macular scar that have minimal differentiation. It may be challenging to differentiate between these conditions based on image processing algorithms alone because they often present similar clinical manifestations in fundus images, e.g., fluid leaks, patterns of retinal thickening, or vascular anomalies. Fine-tuning allowed the models to acquire localized texture, color gradient, lesion pattern, and microvascular alterations typical of each disorder while maintaining the global feature representations learned with pre-training from ImageNet.

The researchers used a collection of measurements frequently used across machine learning and medical image processing to evaluate the performance of classification models. F1 score and accuracy were the primary concerns of this work since they provide valuable information regarding model performance in case of class imbalance.

Formula 1 illustrates that the F-score, which combines recall and precision, is weighted by a parameter specifically labeled β. When β equals 1, the metric is the F1-score, the harmonic mean of precision and recall. When metrics are modeled to achieve a balanced proportion of false positives to false negatives, the F1 score is the most appropriate (Dalianis, 2018).

$$F\beta = (1 + B^2) * \frac{P \times R}{B^2 \times P + R} \qquad (1)$$

When β = 1, the standard F1-score is calculated; check **Formula 2**.

$$F1 = F = 2 \times \frac{P \times R}{P + R} \qquad (2)$$

This measure is helpful in multiple class classification problems with imbalanced data, like retinal disease classification, where some of the conditions have far fewer samples than others.

The researchers also used accuracy as a performance metric. In which accuracy refers to the proportion of correctly classified instances both positive and negative among the total number of predictions made. It is calculated as shown in **Formula 3**:

$$A = \frac{tp + tn}{tp + tn + fp + fn} \qquad (3)$$

Though accuracy is a general measure of model correctness, it can be deceiving when used in imbalanced datasets. Thus, the researchers decided to assess each model's performance based on confusion matrices, which give a visual breakdown of the classification results per class.

A k-fold cross-validation procedure with k = 5 was used to provide robust performance estimation and minimize variance. In this procedure, the dataset was split into five portions, successively training and validating the model to provide a stable average of performance measures.

This study evaluated how effective each model was using the combined results of the score, accuracy, and confusion matrix. The higher the score on these metrics, the better its classification capability and overall generalization among ocular disease categories.

Journal of Information Systems Engineering and Management
2025, 10(53s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Models were evaluated using 5-fold cross-validation to ensure consistent performance across different subsets of the data. The evaluation metrics included accuracy, F1-score, and loss measures. The goal was to determine which model provided the most reliable and accurate classification of retinal diseases.

RESULTS

This study evaluated and compared the performance of traditional machine learning models, Support Vector Machine (SVM), and Random Forest with intensive learning models such as DenseNet121 and DenseNet201 for classifying ocular diseases using color fundus retinal images. Evaluation metrics included mean accuracy, standard deviation, validation, training loss, training and validation accuracy, and classification reports of precision, recall, and F1-score.

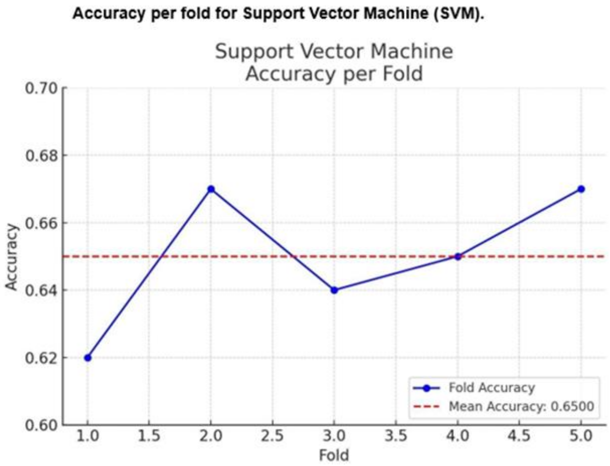

Figure 1. Accuracy per fold for Support Vector Machine (SVM)

Figure 1 gives the results of the SVM in a 5-fold cross-validation process. All five validation folds give results close to each other, except for fold one, which achieves a lower accuracy of ~0.62. The variation is reflected in the more minor standard deviation (1.66%) in the outcome. The plot reflects the model's reliability and accuracy changes when working on different parts of the data.

The uneven results in fold-wise accuracy imply that achieving consistent outcomes is challenging for the SVM model. Such fluctuations mean that SVM often fails to adapt well to unrelated data or complex imaging, probably because its feature measures are less advanced than those of existing deep learning models.

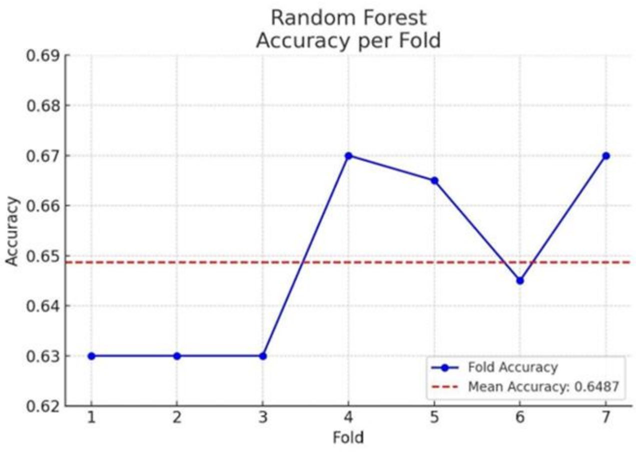

Figure 2. *Accuracy per fold for Random Forest (RF)*

The Random Forest model was evaluated with cross-validation; its results are shown in Figure 2. Just as with the SVM model, the Random Forest model's accuracy also changes from fold to fold, with folds four and seven doing best

Journal of Information Systems Engineering and Management
2025, 10(53s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

(67%) and folds 1 to 3 performing more consistently below (63%). The RF model has a slightly greater amount of variability in performance, as reflected by its higher standard deviation (1.75%). Therefore, Random Forest can work well sometimes, but it still has difficulty reliably classifying retinal diseases.

Figure 2 shows that the performance of the Random Forest model varied during the prediction process. Although some results are strong, the variation seen in other folds suggests that the model might be too rigid or struggle to handle complex, realistic data. The results indicate the model does not pick up on the necessary fine points to distinguish some retinal diseases.

Table 6. Classification Report Average of Densenet201 - 0.0001

Ocular Disease	Precision	Recall	F1-Score	Support
Central Serious Chorioretinopathy	0.96	0.90	0.93	98
Diabetic Retinopathy	0.93	0.92	0.87	121
Disc Edema	0.97	0.97	0.97	101
Glaucoma	0.62	0.55	0.58	123
Healthy Eye	69	0.81	0.74	130
Mascular Scar	82	0.71	0.76	140
Myopia	0.72	0.76	0.74	113
Pterygium	1.00	1.00	1.00	113
Retinal Detachment	0.98	0.98	0.98	113
Retinitis Pigmentosa	0.98	0.96	0.97	113
	Macro Average	0.86	0.85	1187
	Weighted Average	0.85	0.85	1187

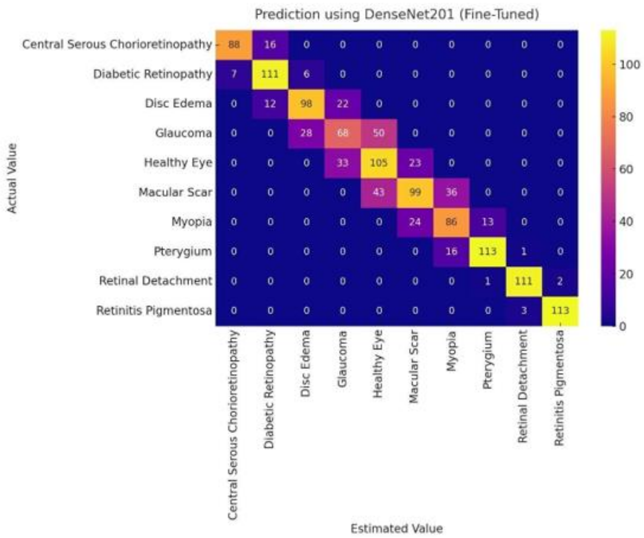

Figure 3. Classification Report of Densenet201 - 0.0001

The class-wise classification report for DenseNet201 (0.0001) is shown in **Figure 3** and **Table 6** highlighted with particularly strong performance on Pterygium and Retinal Detachment (F1-scores: 1.00 and 0.98). However, lower precision and recall were observed for Glaucoma (0.62 and 0.55), indicating difficulty in distinguishing this class.

Journal of Information Systems Engineering and Management
2025, 10(53s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Table 7. Performance Metrics of Deep Learning Models

Metrics	Densenet121	Densenet201 (0.0001%)	Densenet201 (0.01%)
Validation Loss	40.89	40.47	4.63
Validation Accuracy	85.73	85.77	77.89
Training Loss	37.37	34.72	12.30
Training Accuracy	87.22	88.03	70.81

Deep learning models were trained using different learning rates and evaluated based on training and validation accuracy, loss, and performance consistency. As presented in Table 7, DenseNet201 (0.0001) achieved the highest validation accuracy (85.77%) and superior generalization ability, as evidenced by the lowest validation loss (0.4047). This was closely followed by DenseNet121 (0.0001) with a validation accuracy of 85.73% and training accuracy of 87.22%.

On the other hand, DenseNet201 (0.01), which was trained using a higher learning rate, showed a comparatively lower validation accuracy (77.89%) and training accuracy (70.81%), although it recorded the lowest training and validation loss. These results suggest lower learning rates were more effective for stable and accurate model training in this context.

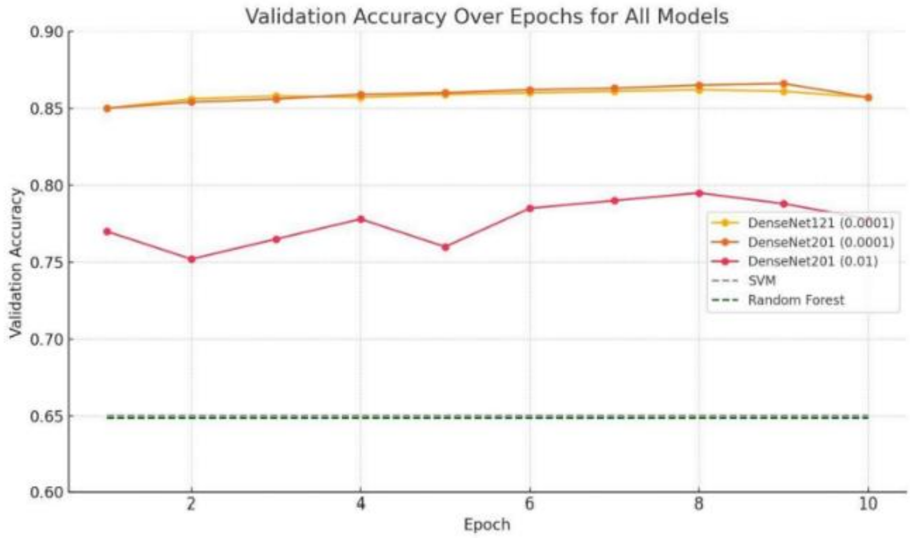

Figure 4. Overall Performance of all Models

Figure 4 presents the validation accuracy trends of all evaluated models across ten training epochs. The deep learning architectures, particularly DenseNet121 (0.0001) and DenseNet201 (0.0001), demonstrated the highest performance, maintaining stable accuracy within the 85% to 86.6% range. The line plots show minimal fluctuations, indicating consistent learning throughout the training. In contrast, DenseNet201 (0.01) exhibited a more volatile validation accuracy curve. The values fluctuated between 75% and 79.5%, suggesting that this learning rate introduced instability during training.

This trend diverges from the smoother curves observed in models trained with lower learning rates. For comparative reference, the Random Forest and Support Vector Machine (SVM) models are illustrated as horizontal lines due to their non-epoch-based nature. Their validation accuracies remained fixed at 65.04% and 64.87%, respectively, and were consistently lower than those achieved by the deep learning models.

Journal of Information Systems Engineering and Management
2025, 10(53s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Table 8. Model Performance Comparison Between Baseline and CNN-Based Models

Machine Learning	Model	Accuracy
Baseline Models	Support Vector Machine	65.04%
	Random Forest	64.85%
CNN-Based Models	DenseNet201(0.0001)	85.73%
	DenseNet201(0.0001)	85.77%
	DenseNet121 (0.01)	77.89%

Table 8 presents a comparative summary of validation accuracies between baseline traditional models and CNN-based architectures. Among the baseline models, Random Forest and Support Vector Machine (SVM) yielded similar performance, with accuracies of 65.04% and 64.87%, respectively. On the other hand, all CNN-based models significantly outperformed the baseline models. DenseNet201 (0.0001) achieved the highest validation accuracy of 85.77%, followed closely by DenseNet121 (0.0001) at 85.73%, while DenseNet201 (0.01) attained 77.89%. These results highlight the enhanced classification capabilities of deep learning models when applied to retinal fundus image data.

DISCUSSION

Performance of Conventional Machine Learning Models

Conventional machine learning models were used in this study. Specifically, Support Vector Machine (SVM) and Random Forest demonstrated limited effectiveness in classifying eye diseases. Both models reported moderate accuracy, with SVM achieving 65.04% and Random Forest at 64.87% (Table 5). These findings align with the work of Gulshan et al. (2016), who found that traditional machine-learning models struggle to handle the complexity of medical images, particularly in classifying intricate patterns in retinal scans. The fold-wise accuracy variability (1.66% for SVM and 1.75% for Random Forest) indicates poor generalizability, a challenge highlighted by Johnson and Khoshgoftaar (2019), who emphasized the limitations of conventional models across diverse datasets.

The confusion matrices (Figures 3 and 4) further reveal these models' difficulty distinguishing between conditions with subtle visual differences, such as early Glaucoma and normal fundus images. Rajalakshmi et al. (2018) similarly found that conventional models often misclassify Glaucoma, particularly in its early stages, when fundus images appear nearly identical to healthy eyes.

Deep Learning Model Performance

In contrast, the deep learning models, particularly DenseNet121 and DenseNet201, demonstrated substantial improvements, achieving validation accuracies of 85.77% and 85.73%, respectively (Table 5). DenseNet201 emerged as the most efficient model, with the lowest validation loss (0.4047) and highest accuracy. These results corroborate the findings of Gulshan et al. (2016) and Esteva et al. (2017), who found that convolutional neural networks (CNNs) excel at learning complex hierarchical features directly from raw data, outperforming traditional machine learning models in medical image classification tasks.

The F1 scores for the DenseNet models, particularly for classes like Pterygium, Retinitis Pigmentosa, and Retinal Detachment, were significantly above 0.95, indicating robust performance across these conditions. However, both models struggled with classifying Glaucoma, as evidenced by lower precision and recall scores. This issue is consistent with Ting et al. (2017), who highlighted that Glaucoma, especially in its early stages, presents minimal visual differences from normal fundus images, which poses challenges even for advanced deep learning models. Iqbal et al. (2022) also discussed how conditions with subtle visual features remain difficult for CNNs to classify accurately.

Model Hyperparameter Impact on Model Robustness

Model performance was heavily influenced by hyperparameter settings, particularly the learning rate. DenseNet201, after a high learning rate of 0.01 training, exhibited lower validation accuracy (77.89%) despite achieving low training and validation losses. This finding aligns with Bajwa et al. (2021), who observed that overly high learning rates often lead to unstable training, resulting in poor generalization. As illustrated in Figure 5, this instability highlights the importance of fine-tuning the learning rate to ensure stability and optimal performance, particularly in medical applications where consistency is essential.

Comparative Model Evaluation

For all cases, it was observed that CNN-based models performed better than SVM and Random Forest by about all of the metrics against deep learning models. All CNN models were proven to have been validated at biases over 77%, while traditional models were capped at around 65%. This is by Poplin et al. (2018), exhibiting that deep learning models are far superior to traditional models in medical imaging tasks, especially when the feature of interest is complex and nuanced. Sengupta et al. (2019) state that CNNs' success is due to the ability to learn features directly and easily scale with large datasets.

However, classical models such as SVM and Random Forest can still be used for interpretability and reasonably lower OCN (as shown in Gulshan et al. (2016)). Although deep learning models are better suited for use in clinical real-world applications, as applications require fast processing and classification over large and complex datasets of existing resources, support in machine learning research and practice cannot suffice.

Overall Model Performance Trends

For all cases, it was observed that CNN-based models performed better than SVM and Random Forest by about all of the metrics against deep learning models. All CNN models were proven to have been validated at biases over 77%, while traditional models were capped at around 65%. This is by Poplin et al. (2018), exhibiting that deep learning models are far superior to traditional models in medical imaging tasks, especially when the feature of interest is complex and nuanced. Sengupta et al. (2019) state that CNNs' success is due to the ability to learn features directly and easily scale with large datasets.

However, classical models such as SVM and Random Forest can still be used for interpretability and reasonably lower OCN (as shown in Gulshan et al. (2016)). Although deep learning models are better suited for use in clinical real-world applications, as applications require fast processing and classification over large and complex datasets of existing resources, support in machine learning research and practice cannot suffice.

Future Work and Model Improvement

Data Augmentation. The challenge in detecting conditions like glaucoma could be mitigated through advanced data augmentation techniques, such as synthetic data generation, to balance the dataset and improve model robustness. Studies like Iqbal et al. (2022) have shown that augmented data can improve model performance for underrepresented classes.

Model Interpretability. Future research should explore techniques to improve deep learning models' interpretability in medical applications. Techniques such as Grad-CAM could help make the decision-making process of CNNs more transparent, ensuring clinicians can trust the AI's predictions. Rajalakshmi et al. (2018) highlighted the value of accessibility in AI-driven systems for clinical implementation.

Ensemble Learning. Combining multiple models through ensemble learning could help address the challenges CNNs face in classifying difficult conditions. As Johnson & Khoshgoftaar (2019) suggested, techniques like bagging and boosting could be employed to improve accuracy and generalization.

This study demonstrates that deep learning models, particularly DenseNet201, significantly outperform traditional machine learning models in retinal disease classification. While challenges remain in classifying conditions like glaucoma, the superior performance of CNN-based models highlights their potential for real-world clinical applications. Finetuning hyperparameters, enhancing data augmentation strategies, and increasing model interpretability will improve the efficacy of AI-based systems in ophthalmology. These advancements promise to revolutionize retinal disease detection, improving diagnostic efficiency and accessibility in clinical settings.

Journal of Information Systems Engineering and Management
2025, 10(53s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

This work shows the revolutionizing ability of deep learning in retinal disease classification, indicating how state-of-the-art convolutional neural network (CNN) structures, specifically DenseNet architectures, can outperform conventional machine learning models concerning accuracy and consistency. With the help of deep feature extraction and hierarchical pattern learning, these newer models efficiently mitigate the complexities of intricate retinal image analysis.

Among the models under test, DenseNet201, using a 0.0001 learning rate, achieved the highest validation accuracy of 85.77%, as it generalized better across all classes of ocular diseases. The model was very robust in differentiating between conditions with minimal visual differences, such as early glaucoma and myopic maculopathy, which most traditional models normally find challenging to differentiate. On the other hand, traditional classifiers such as Support Vector Machine (SVM) and Random Forest struggled to properly distinguish between certain retinal diseases, with a significantly lower mean accuracy of about 65%. Though useful for interpretability, their reliance on hand-crafted features limited their ability to effectively capture the subtle differences happening in real retinal images.

These findings again validate the advantages of deep learning in medical imaging, particularly ophthalmic imaging. Compared to conventional machine learning models where feature extraction is performed manually and is labor-intensive, CNN-based architectures like DenseNet learn the high-level visual features automatically from input image data to report enhanced diagnostic accuracy and increased model flexibility. This ability is especially worthwhile in the case of diseases where visual overlap in attributes, subtle textures, vascular structure, or optic disc morphology are essential diagnostic features.

Apart from precision, the study also points to the potential of CNN-based models for clinical use in real-world practice, particularly in low-resource settings where trained ophthalmologists may not be available. AI-based automated diagnostic platforms can be utilized as cost-effective, scalable options for early disease detection, bridging healthcare gaps and enabling timely intervention for patients at risk of vision loss. In addition, deploying such models on telemedicine platforms and cell phone screening programs would enhance availability, providing uniform pre-screening estimates in underserved communities.

While deep learning has evident strengths, this paper also identifies hurdles to be addressed before clinical translation will succeed. Dataset bias, explainability, and model generalization are issues that require further development to deliver equally good performance with varied patient groups. Multi-modal approaches, attention mechanisms, and ensemble-based schemes are avenues for further research to continue to drive the resilience of AI-based retinal diagnosis systems.

Lastly, this study reiterates AI's crucial role in revolutionizing ophthalmology into more efficient, accurate, and accessible retinal disease diagnosis systems that have the potential to significantly improve patient care through early detection and accurate diagnosis.

REFRENCES

[1] Abdullah, A. A., Aldhahab, A., & Abboodi, H. (2024). Review of eye diseases detection and classification using deep learning techniques. BIO Web of Conferences. doi:doi.org/10.1051/bioconf/20249700012
[2] Abramoff, M. D. (2018). Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. NPJ Digital Medicine. doi:doi.org/10.1038/s41746-018-0040-6
[3] Al Marouf, A., Mottalib, M. M., Alhajj, R., Rokne, J., & Jafarullah, O. (2023). An efficient approach to predict eye diseases from symptoms using machine learning and ranker-based feature selection methods. Bioengineering. doi:doi.org/10.3390/bioengineering10010025
[4] Alowais, S. A. (2023). Revolutionizing healthcare: The role of artificial intelligence in clinical practice. BMC Medical Education, doi.org/10.1186/s12909-023-04698-z.
[5] Bajwa, J. M. (2021). Artificial intelligence in healthcare: Transforming the practice of medicine. Future Healthcare Journal. Retrieved from doi.org/10.7861/fhj.2021-0095
[6] Bergmann, D. (2024). What is fine-tuning in machine learning? IBM. Retrieved from https://www.ibm.com/topics/fine-tuning

[7] Dalianis, H. (2018). Clinical text mining: Secondary use of electronic patient records. Springer. doi:doi.org/10.1007/978-3-319-78503-5
[8] Esteva, A. K. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature. doi:doi.org/10.1038/nature21056
[9] Flaxman, S. R. (2017). Global causes of blindness and distance vision impairment 1990–2020: A systematic review and meta-analysis. The Lancet Global Health, 5(12), e1221–e1234, doi.org/10.1016/S2214-109X(17)30393-5.
[10] Gulshan, V. P. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA. doi:doi.org/10.1001/jama.2016.17216
[11] Iqbal, S. K. (2022). Recent trends and advances in fundus image analysis: A review. Computers in Biology and Medicine. doi:doi.org/10.1016/j.compbiomed.2022.106277
[12] Johnson, J. M. (2019). Survey on deep learning with class imbalance. Journal of Big Data. doi:doi.org/10.1186/s40537-019-0192-5
[13] Poplin, R. V. (2018). Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. Nature Biomedical Engineering. doi:doi.org/10.1038/s41551-018-0195-0
[14] Rajalakshmi, R. S. (2018). Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence. Eye. doi:doi.org/10.1038/s41433-018-0064-9
[15] Rashid, M. R. (2024). Eye disease fundus image dataset (Version 1) [Data set]. Mendeley Data. doi:doi.org/10.17632/s9bfhswzjb.1
[16] Sengupta, S. S. (2019). Ophthalmic diagnosis using deep learning with fundus images – A critical review. Artificial Intelligence in Medicine. doi:doi.org/10.1016/j.artmed.2019.101758
[17] Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. Journal of Big Data. doi:doi.org/10.1186/s40537-019-0197-0
[18] Ting, D. S. (2017). Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. JAMA. doi:doi.org/10.1001/jama.2017.18152
[19] World Health Organization. (2019). Eye care, vision impairment and blindness. Retrieved from World Health Organization.: https://www.who.int/health-topics/blindness-and-vision-loss#tab=tab_1