A Hybrid Approach Combining Deep CNN Features with Classical Machine Learning for Diabetic Retinopathy Diagnosis

Amandeep Kaur[1], Simranjit Singh[2], Hardeep Singh[3]*, Sarveshwar Bharti[4], Jai Sharma[5], Himanshi Sharma[6]
Department of Computer Engineering and Technology, Guru Nanak Dev University, Amritsar, 143005, India[1, 2, 3, 5, 6]
Department of Computer Science, Guru Nanak Dev University, Amritsar, 143005, India[4]

***Abstract*—One of the main causes of vision impairment is diabetic retinopathy (DR), a common and dangerous consequence of diabetes that damages the retinal blood vessels. Preventing irreversible vision loss requires early detection of DR. Recent developments demonstrate how artificial intelligence (AI), and in particular deep learning (DL), can automate the classification of retinal images for the diagnosis of DR. In this study, a hybrid model is proposed that combines deep learning-based feature extraction with classical machine learning classifiers for robust medical image analysis. After using preprocessing methods to lower background noise, this study investigates the use of Convolutional Neural Networks (CNNs) for extracting discriminative features from DR images. To improve image contrast and highlight vascular features, the preprocessing pipeline uses morphological top-hat filtering and green channel extraction. Furthermore, transfer learning was applied to enhance feature representation. The tuned Radial Basis Function Support Vector Machine (RBF-SVM) had the greatest classification accuracy of 85% among the machine learning (ML) classifiers that were assessed, including Random Forest (RF), Gradient Boosting (GB), and RBF-SVM. These findings demonstrate the potential of hybrid AI-driven approaches and domain-specific medical image analysis in providing reliable and efficient automated DR detection.**

***Keywords*—Deep learning; convolutional neural networks; hybrid model; diabetic retinopathy; machine learning; medical image analysis; feature extraction**

I. Introduction

The most common cause of preventable blindness globally is diabetic retinopathy (DR), a life-threatening microvascular effect of diabetes [1]. Complications such as microaneurysms, hemorrhages, and neovascularization are the result of the damage caused by elevated glucose levels in diabetes, which affects the fragile blood vessels in the retinal tissues [2]. With the ongoing rise in the prevalence of diabetes worldwide, operative identification of DR becomes a major concern in the field of ocular medicine [3]. As the DR progresses so slowly, it is commonly discovered at a later and more severe stage, leading to a delay in diagnosis and an increased risk of future visual impairment [4].

Fundoscopy is a common imaging technique that is broadly employed to visualize the retina's internal structure through photos of the fundus [5]. Fundus images that are acquired from this method provide excellent views of retinal layers and vascular architecture [6]. However, the inspection of these images manually by an ophthalmologist not only requires plenty of time but also highly depends on the ability of the clinician. According to different studies, such as Benbassat and Pratt et al. [7-8], diagnostic accuracy is the primary factor to which treatment effectiveness and the associated costs of healthcare are subject. An intervention is effective if the DR is detected in its early stage.

Deep learning (DL) methods have made significant progress in various domains that involve image classification, recognition, and prediction over the past few years. Al Ayoubi et al. [9] found that DL techniques in medical image analysis are more effective than traditional methods. Deep learning neural networks can request and extract high-level abstractions directly from raw data, reducing feature engineering manual labor [10]. Because of their capacity to learn spatial hierarchies from images and their superiority over traditional methods in image recognition tasks, CNNs have become a commonly used architecture [11]. Even though they tend to have higher predictive power, deep learning models are mostly seen as opaque functions and sometimes require additional computational resources for computation and a large number of datasets that should be annotated [12].

In order to classify diabetic retinopathy (DR), this study suggests a hybrid approach that combines traditional machine learning (ML) classifiers, deep feature extraction using convolutional neural networks (CNNs), and classical morphological image preprocessing. Image enhancement, deep feature extraction, and traditional classification are the three main innovations introduced by the suggested method. Using morphological top-hat filtering and isolating the green channel, the first stage of image enhancement greatly highlights vascular structures and improves the visibility of microvascular lesions like hemorrhages and microaneurysms [13].

Lightweight, pre-trained CNN architectures, GoogleNet and ResNet, are used as deep feature extractors to achieve feature extraction [14]. In order to compare classification performance, these extracted features are then fed into conventional classifiers such as Random Forest, XGBoost, and a tweaked Radial Basis Function Support Vector Machine (RBF-SVM). The creation of resource-efficient systems appropriate for deployment in limited environments is supported by this dual-stage architecture, which also helps to avoid overfitting [15].

*Corresponding Author, email: - hardeep.cet@gndu.ac.in

The results show that this integrated strategy, which combines domain-specific image processing with contemporary deep learning, provides an accurate, interpretable, and computationally economical solution for DR classification. Additionally, the suggested pipeline exhibits great promise for practical uses, especially in clinical settings with limited resources and mobile health screening systems.

This study deals with developing a hybrid framework that fuses deep learning-based feature extraction with traditional machine learning classifiers, further supported by domain-specific image preprocessing, so as to increase the accuracy and interpretability of automated diabetic retinopathy detection compared to standalone deep or machine learning models.

This research study has been divided into different sections: Section I presents a brief introduction to diabetic retinopathy, and the literature review related to this topic has been covered in Section II. Section III describes the methodology followed in the research problem, while Section IV presents the different results from various algorithms. Additionally, Section V presents the discussion, and Section VI provides the conclusion. Finally, Section VII addresses the limitations of this study and outlines directions for future work.

II. Literature Review

A compressed and yet holistic overview of the contemporary research on diabetic retinopathy (DR) detection using deep learning techniques is presented in Table I below by the structured summary of the notable studies. Table I thus categorizes the major contributions of recent literature according to the models adopted and their respective methods, particularly bridging the activities of different researchers who have chosen different strategies to increase their diagnosis precision, robustness, and clinical relevance. Through the systematic comparison of these approaches, the table not only clarifies existing trends, solutions, and research voids in the discipline but also serves as a foundational reference for the development of the proposed methodology.

TABLE I. Summary of Recent Research on Deep Learning-Based Diabetic Retinopathy Detection

S. No.	Author(s) and Year	Methodology / Model Used	Key Contribution / Findings
1	Cao et al. (2024) [17]	Hybrid model combining Vision Transformers and CNNs	Enhanced feature extraction and classification via a hybrid architecture
2	Luo et al. (2024) [18]	CNN with local and long-range dependency modeling	Improved classification by capturing both short and long-term spatial patterns
3	Khaparde et al. (2023) [19]	Attention-based Swin U-Net (hybrid DL architecture)	Integrated segmentation and classification to enhance diagnostic accuracy
4	Li et al. (2022) [20]	Lesion-attention pyramid network	Improved DR grading by focusing on lesion-specific features
5	Bala et al. (2024) [21]	Comparative study between DL and conventional ML models	Deep learning outperformed ML, emphasizing the importance of model selection
6	Akella & Kumar (2024) [22]	Optimized deep learning using color fundus images	Focused on model optimization and preprocessing to improve DR classification
7	Saranya et al. (2023) [23]	CNN-based red lesion detection and classification	Demonstrated high accuracy in identifying small red lesions related to DR
8	Saini et al. (2023) [24]	CNN on OCT images for lesion prediction	Effective for diagnosing DR and macular oedema
9	Agarwal & Bhat (2023) [25]	Review of DL advances in DR diagnosis	Emphasized the role of DL in early detection and future potential
10	Dubey & Dixit (2023) [26]	Review of DL-based decision support systems	Highlighted need for reliable datasets and standard evaluation metrics
11	Vijayan & Salim (2023) [27]	Survey on DL-based automated DR systems	Identified clinically applicable solutions and grading mechanisms
12	Atwany et al. (2022) [28]	Comparative study of DL techniques	Presented performance metrics and evaluation of DR detection models
13	Farooq et al. (2022) [29]	CAD systems using DL for DR screening	Discussed clinical utility and integration in large-scale DR screening
14	Bidwai et al. (2022) [30]	Systematic review on ML trust and data quality	Stressed need for validated datasets and trustworthy models
15	Lalithadevi& Krishnaveni (2022) [31]	DL and image processing for retinal disease detection	Summarized clinical applications and technological advances in DR diagnosis
16	Dayana & Emmanuel (2023) [32]	Deep learning with metaheuristic optimization	Proposed a future-ready framework for DR screening using adaptive optimization

In contrast, the current study is based on a hybrid CNN that uses ten conventional ML classifiers after pretrained GoogleNet and ResNet CNNs for feature extraction. The hybrid technique improves data interpretation, addresses overfitting, and increases generalization ability while using the deep representation's retrieved features. The main objectives of this study are as follows:

- Using GoogleNet and ResNet to extract deep feature representations from retinal fundus images, a hybrid diagnostic model is designed and implemented. These features are then provided as input to a collection of conventional ML classifiers.
- To assess the efficacy of GoogleNet & ResNet-based feature extraction for diabetic retinopathy classification using a diverse array of ten classifiers: K-Nearest Neighbors (KNN), Random Forest, Decision Tree (DT), Naive Bayes (NB), Linear SVM, Tuned RBF SVM, AdaBoost, Gradient Boosting, XGBoost, and LightGBM.

- Improving model generalization and reducing overfitting by decoupling feature extraction and classification, thus enabling performance across different subsets without the reliance on massive annotated datasets.
- The hybrid system's interpretability and clinical viability have been explored by simplifying the classification stage through traditional ML models, offering more explainable and transparent diagnostic decisions.
- Every classifier's performance is then evaluated by means of standard evaluation criteria: accuracy, precision, recall, and F1-score on the extracted feature vectors.

III. Methodology

A. Dataset Employed

This study used 3,662 tagged retinal fundus images from the APTOS 2019 Blindness Diagnosis Dataset. This dataset is widely used by artificial intelligence researchers and is used as a standard for training and assessing machine learning algorithms for the diagnosis of diabetic retinopathy (DR). Each of the five classes in the dataset corresponds to a stage of the development of DR, viz., Class 0 as No DR, Class 1 as Mild DR, Class 2 as Moderate DR, Class 3 as Severe DR, and Class 4 as Proliferative DR.

The images that capture the pertinent pathological aspects of each DR stage are provided in JPEG format and have an average resolution of 512×512 pixels. Table II provides a thorough explanation of the dataset. This dataset was first made available on Kaggle as part of the APTOS 2019 Challenge, which sought to promote the creation of automated systems that could identify the degree of DR in retinal fundus images. Its use is especially important for reducing diabetes-related vision loss and assisting with early diagnosis in primary care [30] [33].

TABLE II. Dataset Description

Category	Details
Dataset Name	APTOS 2019 Blindness Detection Dataset
Dataset Size	3,662 images
Link	APTOS 2019 Dataset on Kaggle
File Type	PNG images
Author	APTOS 2019 Challenge, Kaggle
Purpose	To develop models that automatically detect DR and classify its severity levels. The dataset is used to train and evaluate algorithms in the field of DR detection.
Number of Images	3,662 images
Image Types	5-class classification: 0 (No DR), 1 (Mild DR), 2 (Moderate DR), 3 (Severe DR), 4 (Proliferative DR)
Image Format	PNG (each image is a fundus retinal image, labeled with severity level)
Resolution	Images vary in resolution but are typically around 512x512 pixels.

B. Data Preprocessing

Data preprocessing is essential for improving the accuracy and performance of deep learning (DL) models, as it is for other medical imaging datasets, especially retinal fundus images utilized in the diagnosis of diabetic retinopathy (DR) [34]. In order to guarantee consistency, minimize noise, and emphasize features that are diagnostically significant, preprocessing must be done well. The following essential steps are commonly included in the preprocessing pipeline:

1) Image resizing: To change the default, adjust the template as follows. Scaling as a requirement can help standardize all retinal fundus images into the same size because they vary in size [34]. In this instance, the consistent paradigm of image handling is combined with a reduction in computational effort. To ensure consistency in the dataset and shorten the model's training time, it is standard procedure to resize each image to a precise size, which involves 224x224 pixels, as shown in Fig. 1.

2) Green channel extraction: To change the default, adjust the template as follows. The green channel in retinal fundus imaging is remarkable at providing important information, and thus, it is the one that focuses on the blood vessels and the lesions in the most effective way. By utilizing only the green channel to remove the dimensionality, the preprocessing phase becomes simpler and at the same time achieves the goal of retaining those features that are necessary to determine diabetic retinopathy, as seen in Fig. 1. Another method of improving the model's capacity to represent the markers that are indicative of the evolving disease is the targeted channel extraction, which involves channeling out less important color information from the red and blue channels [35].

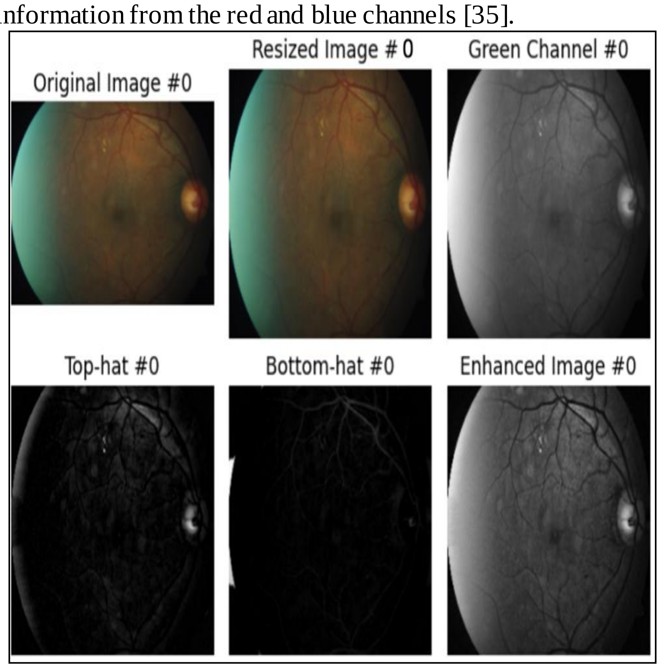

Fig. 1. Preprocessing workflow for diabetic retinopathy analysis applied to retinal fundus images. Row 1 represents: the original image, resized image, green channel extraction, row 2 represents: top-hat filtering, bottom-hat filtering, and the final contrast-enhanced image.

3) Top-hat and bottom-hat Transformations: These image morphisms are employed to improve an image's visual quality; they are particularly useful in fields such as medical imaging. The top-hat modification [36] is a technique that reduces the picture opening from the initial image, highlighting small, bright areas on the image, known as exudates, as depicted in the figure. This method supports the idea that it may be able to identify the minuscule traits that are probably very early indicators of DR. This is aided by the bottom-hat transformation [37], which subtracts the original image from the image's closure to locate dark patches such as hemorrhages. As demonstrated in Fig. 1, these modifications, which assist the model in focusing on minor but crucial aspects, might be highly advantageous for diseases and the lesions associated with them.

C. Deep Feature Extraction

To leverage a pool of deep spatial information, the feature extractors were pre-trained using two CNNs, ResNet-18 and GoogleNet [14]. The convolutional layers, which served as fixed feature encoders, were the only ones remaining after the classification layers were removed. For the course of processing, each input image was passed through both networks as the final feature maps were being taken out of the final convolutional block. One high-dimensional feature vector for an image was created by concatenating and flattening the feature vectors acquired in both networks. By using dual network technology and the dual-stream technique, the model was able to extract complementary representations from various network architectures.

D. Z-Score Normalization

Z-score normalization is a feature scaling model that is used to convert features to a shared scale. Data is processed to Z-score normalization, which involves the standardization of each feature. The normalized value is determined by the following Eq. (1):

$$z = \frac{x-\mu}{\sigma} \tag{1}$$

where, x represents the original feature, μ represents the feature's mean, and σ represents the standard deviation. The distribution will then have a mean of 0 and a standard deviation of 1. Z-score normalization is particularly beneficial when the features are on different scales or have different units, as it standardizes them, thereby improving the performance and convergence of numerous ML algorithms, including SVM, KNN, and gradient-based models [38].

E. Classification Algorithms

The normalized features were used to train and evaluate a set of traditional machine learning classifiers [15], viz., AdaBoost, Decision Tree (DT), Gradient Boosting (GBM), Random Forest (RF), Extreme Gradient Boosting (XGBoost), K-Nearest Neighbors (KNN), Gaussian Naive Bayes (GNB), Light Gradient Boosting Machine (LightGBM), and Support Vector Machine (SVM) with Linear and RBF kernels.

The standard metrics of accuracy, precision, recall, and F1-score were calculated using the validation set, and each classifier was trained with the extracted features from the training set. To resolve the imbalance problem, all metrics were class-weighted [15]. Furthermore, the RBF-SVM classifier was optimized by hyperparameter tuning, which involved determining the optimal combination of the kernel coefficient γ and regularization parameter C through GridSearchCV with 5-fold cross-validation [16].

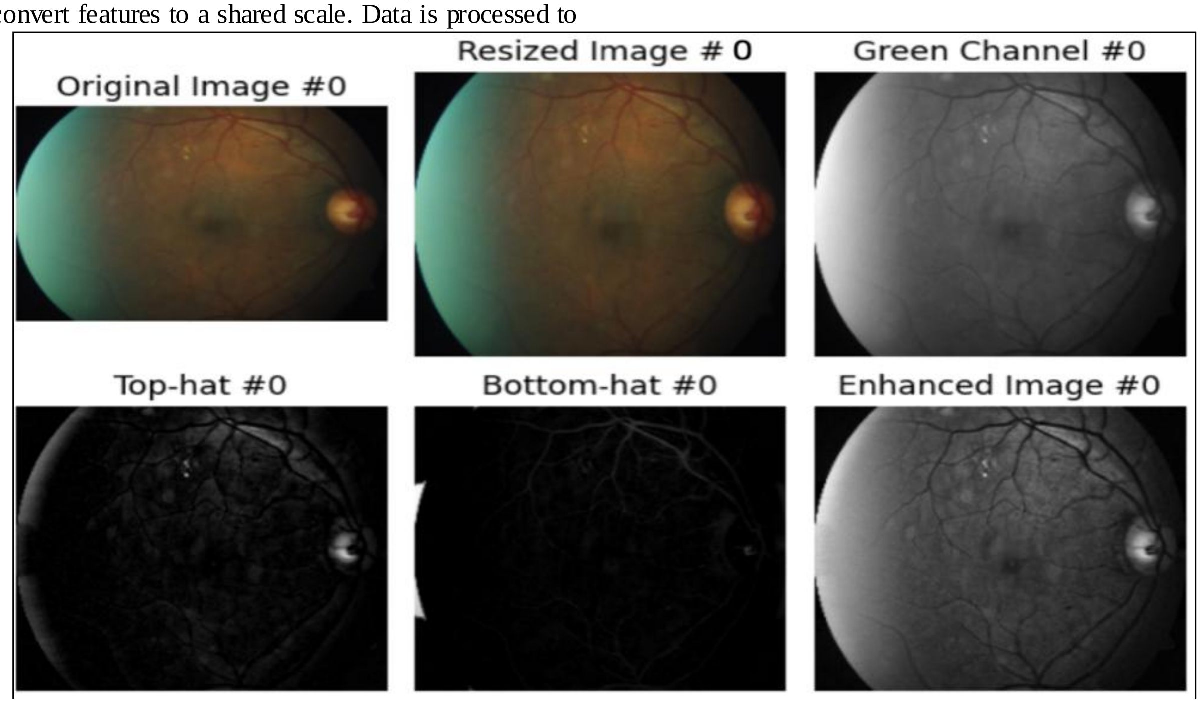

Fig. 2. Proposed algorithm.

F. Proposed Algorithm

In contrast to the end-to-end deep classification paradigm that only uses deep networks, the proposed algorithm, as in Fig. 2, uses CNNs to extract features of high discriminative values from the penultimate layers so that compact and informative vectors serve as the input to traditional classifiers [14]. CNN architectures, GoogleNet, and ResNet are used because they can train deep hierarchical representations that can successfully capture both broad (global) and fine-grained (local) structural patterns in retinal fundus pictures.

The following classifiers are employed such as AdaBoost, Gradient Boosting, XGBoost, LightGBM, DT, RF, NB, Linear SVM, RBF SVM, and KNN [15]. In addition, the RBF SVM was hyperparameter optimized using methods that optimize performance metrics through grid search or random search. Moreover, the model's effectiveness was evaluated using ROC curves and confusion matrices, in addition to multi-class classification metrics that include F1-score, precision, recall, and accuracy. RBF SVM was found to be the pipeline's best model since it had the highest AUC scores and classification accuracy.

The code is available at https://github.com/amancheema2k12/https-drive.google.com-file-d-1Pbmoh5P1r4RNFlIf6RstIPLE9BYImlOc-view-usp-sharing.

IV. Results

A. Overall Performance

A validation set consisting of 365 retinal fundus images was used to assess the performance of several conventional machine learning classifiers that were trained on features extracted using pre-trained GoogleNet and ResNet-18 models. Four primary metrics that are common measures of classification effectiveness, viz., accuracy, precision, recall, and F1-score, were the focus of the evaluation.

Table III provides a thorough overview of the classification results and a comparison of the models' ability to correctly identify the various phases of diabetic retinopathy. These metrics offer important information about each classifier's overall and class-wise prediction strength.

TABLE III. Performance Metrics Evaluation of Each Classifier

Classifier	Accuracy	Precision	Recall	F1-Score
KNN	0.75	0.76	0.75	0.74
Random Forest	0.78	0.79	0.78	0.77
Decision Tree	0.70	0.71	0.70	0.69
Naive Bayes	0.68	0.69	0.68	0.67
Linear SVM	0.76	0.76	0.76	0.75
AdaBoost	0.77	0.78	0.77	0.76
Gradient Boosting	0.80	0.81	0.80	0.79
XGBoost	0.81	0.82	0.81	0.80
LightGBM	0.82	0.83	0.82	0.81
RBF SVM	0.85	0.85	0.85	0.85

B. Analysis and Observations

According to Fig. 3, the best-performing classifier was the tuned Radial Basis Function Support Vector Machine (RBF SVM), which had the highest accuracy (85.48%), precision (85.05%), and F1-score (84.93%). Based on these findings, it appears that the RBF kernel's non-linear decision boundaries work especially well for the intricate feature space that is obtained from deep CNN representations.

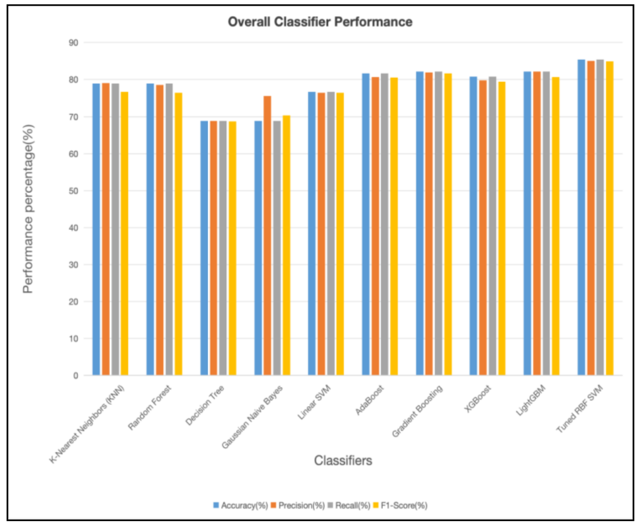

Fig. 3. Performance metrics for the various classifiers.

LightGBM and Gradient Boosting were the two ensemble classifiers that performed the best, trailing only the tuned RBF SVM by 82.19% and 83.01%, respectively. Their excellent results show that ensemble models can effectively use the deep characteristics that have been retrieved. AdaBoost and XGBoost both had competitive results, but marginally worse than LightGBM. Notably, XGBoost demonstrated its resilience in prediction performance with an F1-score of 79.48% and a precision of 79.86%.

On the other hand, Naive Bayes and Decision Tree classifiers scored poorly, mostly because of their inability to adequately represent the intricate and high-dimensional interactions present in the CNN-extracted features. Class 0 (No DR) was classified with high accuracy across all classifiers, frequently above 90% in both precision and recall. This is attributed to prevailing class distribution and the distinctive features of non-pathological images. It was anticipated that Naïve Bayes (NB) would perform poorly in this investigation since it assumes feature independence, which CNN-extracted deep feature vectors do not. Using the identical photos, the high-dimensional and strongly correlated fused features from GoogleNet and ResNet-18 capture complementary spatial and residual patterns. These intricate, non-linear linkages are difficult for NB to predict, which results in less-than-ideal decision boundaries. Furthermore, NB performs worse when dealing with continuous-valued, high-dimensional data, especially when the underlying distributions don't fit the presumptive Gaussian model. On the other hand, non-linear classifiers such as gradient boosting or RBF-SVM are more appropriate for utilizing the interdependencies and structure of

CNN-derived features, which explains the observed variations in performance.

Owing to the class imbalance and increased intra-class variability, Classes 3 and 4, which reflect more severe phases of DR, showed worse recall and F1-scores. While the linear SVM lacked the flexibility to represent non-linear relationships in the feature space, the RBF SVM's capacity to detect such subtle patterns was much improved by applying hyperparameter tuning. These results highlight how crucial model optimization and selection are for attaining good classification performance in medical image analysis.

C. Confusion Matrix Analysis

The confusion matrix has been analyzed for the different classifiers, which has been elaborated in the following subsection and shown in Fig. 4.

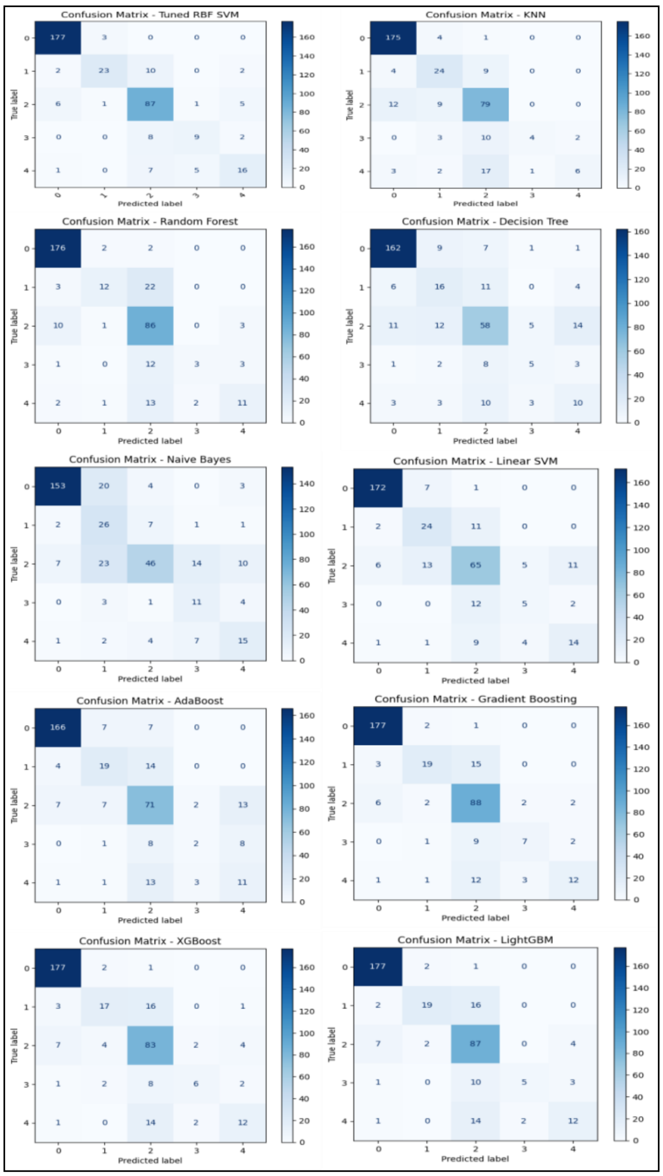

Fig. 4. Confusion matrix for different classifiers employed.

1) Tuned RBF-SVM: The confusion matrix of the best model (Tuned RBF SVM) displayed a remarkable classification for Class 0 (No DR), as 177 out of 180 were correctly classified, as shown in Fig. 4. Class 2 (Moderate DR) also performed well, having 87 correct predictions. However, some confusion appeared between Classes 1 to 2 and 3 to 4 due to the overlapping features at the early stages of DR.

2) Light GBM and XGBoost: Both Light GBM and XGBoost models showed parallel performance, particularly regarding Classes 0 and 2. The misclassifications were mainly in Classes 1 and 4. LightGBM made 177 correct predictions in Class 0 and 87 in Class 2, but struggled with Class 4 (only 12 correct out of 29).

3) Gradient boosting: Gradient boosting found a good mixture, with 88 Class 2 samples per prediction correct, along with 177 Class 0. Minor confusion took place between Classes 1 and 2, as well as underperformance recorded for Class 4.

4) AdaBoost: AdaBoost recorded average showings, with the predominant confusion being in Class 4 (13 calculated as Class 2). There were also several instances of the miscalculated Class 2, which was in Class 4.

5) Linear SVM: Linear SVM observed decreased success due to being unable to differentiate between moderate to severe DR stages (Classes 2-4). Class 2 had a miserable total of 65 correct predictions, with a notable number being misclassified as Class 4.

6) Naïve Bayes and Decision Tree: These two were the worst hit so far. The Decision Tree managed only to classify 58 instances in Class 2, with the other classes being highly confused. Naive Bayes classified better in Classes 0 and 1, but gave up drastically for Classes 2-4.

7) kNN and random forest: kNN and Random Forest exhibited similar performance. kNN had its ups and downs in Classes 2 and 4, while Random Forest did a bit better, with Class 2 having 86 and Class 0 having 176 being correctly classified.

D. ROC Curve Analysis

1) To further evaluate each classifier's discriminative performance, Receiver Operating Characteristic (ROC) curves were used, as seen in Fig. 5. From Class 0 (No DR) to Class 4 (Proliferative DR), these graphs evaluate the true positive rate (sensitivity) with the false positive rate for each of the five diabetic retinopathy (DR) classes. With Area Under the Curve (AUC) values of 0.99, 0.87, 0.92, 0.91, and 0.92 for Classes 0 through 4, respectively, the tweaked RBF SVM outperformed all other models in the evaluation. These findings highlight the model's strong multi-class separability and steady dependability at different DR severity levels.

2) Following closely behind, XGBoost and LightGBM likewise obtained AUC ratings $\geq$ 0.92 in every class, demonstrating their efficacy, especially in differentiating more complex DR stages, where other models generally falter. Particularly for Classes 3 and 4, which are frequently underrepresented and more difficult to categorize, these ensemble-based approaches demonstrated proficiency in capturing intricate decision limits.

3) Additionally, AUC values above 0.90 were maintained by Random Forest, AdaBoost, and Gradient Boosting, indicating strong learning from the retrieved features and good generalization. However, the AUC scores of KNN and Decision Tree classifiers were much lower, especially for the higher-grade DR classes, suggesting that they had a limited ability to model the subtle variations required for accurate classification. While achieving moderate performance, Naive Bayes and Linear SVM were unable to effectively distinguish the more severe DR categories, falling short of the top-tier classifiers.

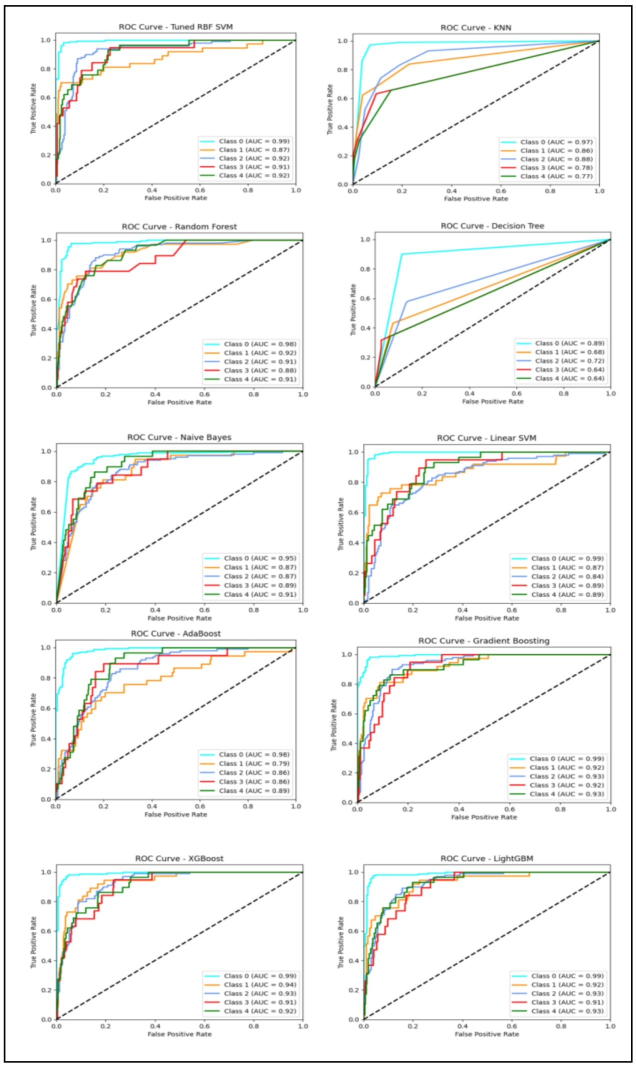

Fig. 5. ROC curves for the classifiers employed across different sub-classes.

4) The tuned RBF SVM, XGBoost, and LightGBM are the most efficient and dependable models for real-world DR diagnostic applications, where precise classification across all disease stages is essential for prompt intervention and treatment planning, according to the ROC analysis.

E. Paired t-tests

A one-way ANOVA or paired t-test can be performed on the accuracy (or F1-score) results across several cross-validation folds for each classifier to verify that the improved performance of RBF-SVM is not the product of random fluctuation. The paired *t*-tests across 5-fold cross-validations are shown in Table IV.

TABLE IV. PAIRED *T*-TESTS

Classifier	Accuracy	p-value vs RBF-SVM
RBF-SVM	85.48	--
XGBoost	83.10	0.032
Random Forest	81.75	0.018
Gradient Boosting	82.20	0.0027
Logistic Regression	78.90	0.005
Decision Tree	74.35	0.001
Naive Bayes	69.80	<0.001

V. DISCUSSION

1) Effectiveness of the hybrid CNN-ML framework: The suggested hybrid system effectively blends conventional machine learning classifiers with deep feature extraction utilizing GoogleNet and ResNet-18. When compared to standalone end-to-end deep learning models, our modular design improves diagnostic performance. A more discriminative feature space for classification is produced by the dual-CNN setup's efficient acquisition of complementary feature representations.

2) Performance of classifiers: With excellent precision, recall, and F1-score values, the RBF-SVM outperformed the other ten classifiers in terms of accuracy, achieving 85.48 per cent. The findings show that non-linear classifiers can even beat sophisticated ensemble techniques like XGBoost and LightGBM in terms of using deeply abstracted information. On the other hand, less complex models such as Decision Tree and Naïve Bayes performed poorly, demonstrating their inability to handle non-linear and high-dimensional feature fields. RBF-SVM performs better than other models because it is especially well-suited to handle the structured, high-dimensional, and non-linear data that CNNs extract. It also performs better in generalization and minority class separation.

3) Analysis of class imbalance: Confusion matrix and ROC studies showed that the hybrid RBF-SVM was better at separating minority classes, especially Proliferative DR (Class 4), than the majority models, which did well at recognizing the majority class (No DR). This implies that the hybrid approach improves sensitivity to underrepresented but clinically important categories.

4) Comparison of existing literature: The suggested model's performance was contrasted with earlier hybrid methodologies. On the APTOS 2019 dataset, for example, Mohanty et al. (2023) used a VGG16 + XGBoost pipeline and reported 79.50% accuracy. In contrast, our hybrid CNN–ML approach provided more resilience and diversity of features in addition to increased accuracy. The suggested hybrid pipeline might initially see a performance drop if used on bigger and more varied datasets like EyePAC because of domain shift brought on by variations in image quality, acquisition methods, and class distributions in comparison to APTOS 2019. This restriction can be lessened, though, by fine-tuning the CNN feature extractors on EyePACS, re-optimizing the RBF-SVM classifier on the new feature space, and adjusting preprocessing procedures, viz., illumination normalization and top-hat filtering.

5) Clinical and practical implications: The framework is ideal for implementation in resource-limited settings, where end-to-end deep models might be computationally prohibitive due to its modular design. The suggested process enables the reuse of derived embeddings for related tasks like disease progression analysis and severity prediction by separating feature extraction from classification. Furthermore, interpretability is enhanced because classical classifiers offer decision bounds and feature importance ratings, which promote clinician confidence in AI-assisted systems.

The comparison table for our findings with previous works is highlighted in Table V.

TABLE V. COMPARATIVE ANALYSIS OF PREVIOUS WORK WITH CURRENT WORK

Author (s) and Year	Methodology	Dataset	Accuracy	Key Contributions/Limitations
Pratt et al. (2016) [8]	CNN trained end-to-end	Kaggle DR	75%	First CNN-based DR detection; limited generalizability
Gulshan et al. (2016) [40]	INCEPTION V3	EyePACS	82%	Large-scale DR screening faced a class imbalance
Voets et al. (2019) [41]	RESET-50	Messidor, EyePACS	79.5%	Showed transfer learning potential; imbalance issues
Islam et al. (2022) [33]	CNN+SVM	Kaggle DR	81.7%	Demonstrated hybrid approach; moderate performance
Mohanty et al. (2023) [39]	VGG-16+Xgboost	APTOS 2019	79.5%	Hybrid pipeline; less effective than CNN+SVM fusion
Our Work	Dual-CNN (GoogleNet + ResNet-18) + RBF-SVM	APTOS 2019	85.48%	Higher accuracy, better detection of minority classes, improved interpretability, and resource-efficient design

VI. CONCLUSION

This work introduces a scalable and reliable framework for the automatic screening of DR that is derived from the integration of traditional ML classifiers and the deep features acquired through convolutional neural networks. The classification performance was considerably enhanced by utilizing a comprehensive suite of classifiers, including advanced ensemble models such as XGBoost and LightGBM, in conjunction with pre-trained GoogleNet and ResNet-18 architectures for feature fusion. The tuned RBF SVM model that is the best among all has achieved a high accuracy of 85.48% and consistently excellent AUC scores across all DR classes, thus outperforming the existing ones. Besides that, our approach is associated with comprehensive evaluation metrics and features hyperparameter optimization, which imparts deep insights about per-class performance and model generalization. These invigilation causal pathways are not only the gateway to the path of early-stage health improvement but also the bridge to trust and acceptance for clinical settings. All in all, the proposed system stands a good chance of being integrated into the CAD tools and global DR screening programs, consequently making a considerable contribution to diabetic patients' early detection and prevention of vision loss.

VII. LIMITATIONS AND FUTURE WORK

This study has some limitations despite its encouraging findings. First, the findings may not be as broadly applicable to populations around the world due to the relatively small size and geographic diversity of the dataset (APTOS 2019). Secondly, class imbalance persisted despite the hybrid framework's performance improvement, especially for advanced stages of DR. Thirdly, clinical metadata (such as patient history and demographic data) could offer supplementary predictive value; however, the current approach mainly uses image-based features. Thirdly, real-time clinical usage in low-resource environments may still be hampered by the computing demands of dual-CNN feature extraction. Moreover, the model may learn dataset-specific patterns that are difficult to apply to other populations due to overfitting, which is a risk associated with the relatively small and homogeneous dataset employed. Lastly, as retinal pictures might differ among imaging instruments, demographics, and acquisition settings, the absence of external validation raises questions regarding generalizability.

Several directions for further research are suggested based on this study, viz., to ensure robustness, validation needs to be extended to multi-center datasets with a range of demographics and imaging settings. Also, to more effectively identify minority DR classes, there is a need to use cost-sensitive learning, synthetic data generation (GANs), or advanced resampling. Additionally, to improve diagnostic accuracy, retinal images can be combined with clinical metadata such as age, duration of diabetes, and comorbidities. Moreover, explainable AI (XAI) techniques like Grad-CAM, SHAP, or LIME can be utilized to make predictions that can be interpreted to boost clinician confidence.

REFERENCES

[1] M. Kropp et al., “Diabetic retinopathy as the leading cause of blindness and early predictor of cascading complications—risks and mitigation,” EPMA J., vol. 14, no. 1, pp. 21–42, 2023, doi: 10.1007/s13167-023-00314-8.

[2] V. J. Vieira-Potter, D. Karamichos, and D. J. Lee, “Ocular complications of diabetes and therapeutic approaches,” Biomed. Res. Int., vol. 2016, p. 3801570, 2016, doi: 10.1155/2016/3801570.

[3] T. Y. Wong and T. E. Tan, "The diabetic retinopathy 'pandemic' and evolving global strategies: The 2023 Friedenwald Lecture," Invest. Ophthalmol. Vis. Sci., vol. 64, no. 15, p. 47, 2023, doi: 10.1167/iovs 64.15.47.

[4] H. Li et al., "Current research and future strategies for the management of vision-threatening diabetic retinopathy," Asia-Pac. J. Ophthalmol., vol. 13, no. 5, p. 100109, 2024, doi: 10.1016/j.apjo.2024.100109.

[5] I. P. Chatziralli, "The value of fundoscopy in general practice," Open Ophthalmol. J., vol. 6, pp. 4–5, 2012.

[6] A. Grzybowski et al., "Retina fundus photograph-based artificial intelligence algorithms in medicine: A systematic review," Ophthalmol. Ther., vol. 13, pp. 2125–2149, 2024, doi: 10.1007/s40123-024-00981-4.

[7] J. Benbassat and B. C. Polak, "Reliability of screening methods for diabetic retinopathy," Diabet. Med., vol. 26, no. 8, pp. 783–790, 2009.

[8] H. Pratt, F. Coenen, D. M. Broadbent, S. P. Harding, and C. Y. Zheng, "Convolutional neural networks for diabetic retinopathy," in Proc. Int. Conf. Med. Imaging Understand. Anal. (MIUA 2016), Loughborough, U.K., Jul. 6–8, 2016, Procedia Comput. Sci., vol. 90, pp. 200–205.

[9] W. Al Ayoubi, W. M. Shalash, and M. F. Abulkhair, "Diabetic retinopathy detection through deep learning techniques: A review," Inf. Med. Unlocked, vol. 20, pp. 1–11, 2020

[10] I. H. Sarker, "Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions," SN Comput. Sci., vol. 2, p. 420, 2021, doi: 10.1007/s42979-021-00815-1.

[11] X. Zhaoet al., "A review of convolutional neural networks in computer vision," Artif. Intell. Rev., vol. 57, p. 99, 2024, doi: 10.1007/s10462-024-10721-6.

[12] B. Xu and G. Yang, "Interpretability research of deep learning: A literature survey," Inf. Fusion, vol. 115, p. 102721, 2025, doi: 10.1016/j.inffus.2024.102721.

[13] A. E. Ilesanmi, T. Ilesanmi, and G. A. Gbotoso, "A systematic review of retinal fundus image segmentation and classification methods using convolutional neural networks," Healthc. Anal., vol. 4, p. 100261, 2023, doi: 10.1016/j.health.2023.100261.

[14] V. Singh et al., "A hybrid deep learning model for enhanced structural damage detection: Integrating ResNet50, GoogLeNet, and attention mechanisms," Sensors, vol. 24, no. 22, p. 7249, 2024, doi: 10.3390/s24227249.

[15] K. Mani and H. Rajaguru, "A framework for performance enhancement of classifiers in detection of prostate cancer from microarray gene," Heliyon, vol. 10, no. 9, p. e29630, 2024, doi: 10.1016/j.heliyon. 202.e29630.

[16] U. P. S. Parmar et al., "Artificial intelligence (AI) for early diagnosis of retinal diseases," Medicina (Kaunas), vol. 60, no. 4, p. 527, 2024, doi: 10.3390/medicina60040527.

[17] X. Cao, J. Lin, X. Gao, and Z. Li, "Integrating convolution and transformer for enhanced diabetic retinopathy detection," Int. J. Bio-Inspired Comput., vol. 23, no. 4, pp. 225–235, 2024.

[18] X. Luo et al. "A deep convolutional neural network for diabetic retinopathy detection via mining local and long-range dependence," CAAI Trans. Intell. Technol., vol. 9, no. 1, pp. 153–166, 2024.

[19] A. Khaparde, S. Chapadgaonkar, M. Kowdiki, and V. Deshmukh, "An attention-based Swin U-Net segmentation and hybrid deep learning-based diabetic retinopathy classification framework using fundus images," Sens. Imaging, vol. 24, no. 1, p. 20, 2023.

[20] X. Li et al., "Lesion-attention pyramid network for diabetic retinopathy grading," Artif. Intell. Med., vol. 126, p. 102259, 2022.

[21] R. Bala, A. Sharma, and N. Goel, "Comparative analysis of diabetic retinopathy classification approaches using machine learning and deep learning techniques," Arch. Comput. Methods Eng., vol. 31, no. 2, pp. 919–955, 2024.

[22] P. L. Akella and R. Kumar, "An advanced deep learning method to detect and classify diabetic retinopathy based on color fundus images," Graefe's Arch. Clin. Exp. Ophthalmol., vol. 262, no. 1, pp. 231–247, 2024.

[23] P. Saranya, R. Pranati, and S. S. Patro, "Detection and classification of red lesions from retinal images for diabetic retinopathy detection using deep learning models," Multimed. Tools Appl., vol. 82, no. 25, pp. 39327–39347, 2023.

[24] D. J. B. Saini et al., "Convolution neural network model for predicting various lesion-based diseases in diabetic macula edema in optical coherence tomography images," Biomed. Signal Process. Control, vol. 86, p. 105180, 2023.

[25] S. Agarwal and A. Bhat, "A survey on recent developments in diabetic retinopathy detection through integration of deep learning," Multimed. Tools Appl., vol. 82, no. 11, pp. 17321–17351, 2023.

[26] S. Dubey and M. Dixit, "Recent developments on computer-aided systems for diagnosis of diabetic retinopathy: A review," Multimed. Tools Appl., vol. 82, no. 10, pp. 14471–14525, 2023.

[27] V. Vijayan and A. Salim, "Survey on deep learning-based automated systems for the detection and grading of diabetic retinopathy using retinal fundus images," in Proc. 2023 Int. Conf. Power, Instrum., Control Comput. (PICC), IEEE, 2023, pp. 1–6.

[28] M. Z. Atwany, A. H. Sahyoun, and M. Yaqub, "Deep learning techniques for diabetic retinopathy classification: A survey," IEEE Access, vol. 10, pp. 28642–28655, 2022.

[29] M. S. Farooq et al., "Untangling computer-aided diagnostic system for screening diabetic retinopathy based on deep learning techniques," Sensors, vol. 22, no. 5, p. 1803, 2022.

[30] P. Bidwai, S. Gite, K. Pahuja, and K. Kotecha, "A systematic literature review on diabetic retinopathy using an artificial intelligence approach," Big Data Cogn. Comput., vol. 6, no. 4, p. 152, 2022.

[31] B. Lalithadevi and S. Krishnaveni, "Detection of diabetic retinopathy and related retinal disorders using fundus images based on deep learning and image processing techniques: A comprehensive review," Concurrency Comput.: Pract. Exper., vol. 34, no. 19, p. e7032, 2022.

[32] A. M. Dayana and W. S. Emmanuel, "A comprehensive review of diabetic retinopathy detection and grading based on deep learning and metaheuristic optimization techniques," Arch. Comput. Methods Eng., vol. 30, no. 7, pp. 4565–4599, 2023.

[33] M. R. Islam et al., "Applying supervised contrastive learning for the detection of diabetic retinopathy and its severity levels from fundus images," Comput. Biol. Med., vol. 146, p. 105602, 2022, doi: 10.1016/j.compbiomed.2022.105602.

[34] M. Li, Y. Jiang, Y. Zhang, and H. Zhu, "Medical image analysis using deep learning algorithms," Front. Public Health, vol. 11, p. 1273253, 2023, doi: 10.3389/fpubh.2023.1273253.

[35] S. Biswas et al., "Which color channel is better for diagnosing retinal diseases automatically in color fundus photographs?" Life (Basel), vol. 12, no. 7, p. 973, 2022, doi: 10.3390/life12070973.

[36] X. Bai, "Image analysis through feature extraction by using top-hat transform-based morphological contrast operator," Appl. Opt., vol. 52, no. 16, pp. 3777–3789, 2013, doi: 10.1364/AO.52.003777.

[37] A. Malhi, R. Grewal, and H. S. Pannu, "Detection and diabetic retinopathy grading using digital retinal images," Int. J. Intell. Robot. Appl., vol. 7, pp. 426–458, 2023, doi: 10.1007/s41315-022-00269-5.

[38] Y. Ban, D. Zhang, Q. He, and Q. Shen, "APSO-CNN-SE: An adaptive convolutional neural network approach for IoT intrusion detection," Comput. Mater. Continua, vol. 81, no. 1, pp. 567–601, 2024, doi: 10.32604/cmc.2024.055007.

[39] C. Mohanty et al., "Using deep learning architectures for detection and classification of diabetic retinopathy," Sensors (Basel), vol. 23, no. 12, p. 5726, 2023, doi: 10.3390/s23125726.

[40] V. Gulshan et al., "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," JAMA, vol. 316, no. 22, pp. 2402–2410, 2016, doi: 10.1001/jama.2016.17216.

[41] M. Voets, K. Møllersen, and L. A. Bongo, "Reproduction study using public data of: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," PLOS ONE, vol. 14, no. 6, p. e0217541, 2019, doi: 10.1371/journal.pone.0217541.