International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048
Volume 10, Issue 9, pp 38-56, September-2026
https://doi.org/10.47001/IRJIET/2026.109005

Federated Learning for Multi-Disease Retinal Image Classification: A Systematic Review, Taxonomy, and Future Directions

[1]Divya Chunduri, [2]K. Suresh Babu

[1]Research Scholar, Department of Computer Science, Jawaharlal Nehru Technological University, Hyderabad, India
[2]Professor, Department of Computer Science, Jawaharlal Nehru Technological University, Hyderabad, India
E-mail: [1]divya.chunduri@gmail.com, [2]kare_suresh@jntuh.ac.in

Abstract **- Retinal diseases are becoming more frequent, leading to increased demand for large and diverse datasets in applying deep learning techniques in automated diagnoses. The problem with data from different health centers is that, due to various privacy laws, ethical concerns, and limitations on institutional data sharing, it is usually difficult to combine retina images into one centralized system. Federated learning offers a solution by allowing various institutions to train a model in collaboration without merging their patients' data into a single center. The objective of this paper is to review methods for federated learning in the context of classification of retinal diseases based on the analyses of patterns of retina images. A systematic literary search and selection process has been undertaken in order to get the relevant researches published between 2018 and 2025 from the major scientific databases. The selected works are analyzed and classified in accordance with federated learning strategies, distribution of data used, disease coverage, models used, and data protection methods. In addition, a comparative analysis was also conducted in order to figure out major trends, advantages of the discussed approaches, and weaknesses of the existing studies. The current study exposes the weakness of federated retinal learning frameworks, which mostly work towards single-disease classification and base their analyses on the overly simplified claims about data. The issues of communication efficiency, privacy-utility trade-offs, and lack of large-scale clinical validation seriously challenge the application of the existing frameworks. After discussing the results of the review, the authors identify the most important gaps in the available research and develop an approach for deal with the existing issues effectively.**

Keywords: Federated learning, retinal imaging, multi-disease classification, fundus photography, optical coherence tomography, privacy-preserving machine learning.

I. INTRODUCTION

Retinal diseases such as diabetic retinopathy, glaucoma, and age-related macular degeneration are important causes of vision impairment and require timely and accurate diagnosis for effective clinical management. Advances in retinal imaging, particularly fundus photography and optical coherence tomography, have enabled the development of automated diagnostic systems based on deep learning techniques [1]-[2].

In recent years, deep learning models have demonstrated considerable potential for automated retinal disease analysis. Gulshan *et al.* demonstrated the effectiveness of deep learning for detecting diabetic retinopathy from retinal fundus photographs, while Ting et al. developed a deep learning system capable of detecting diabetic retinopathy and related eye diseases using retinal images from multiethnic populations [1], [3]. De Fauw *et al.* further demonstrated the applicability of deep learning to three-dimensional optical coherence tomography scans for diagnosis and referral across multiple sight-threatening retinal diseases [2]. These studies demonstrate the potential of deep learning for supporting automated retinal disease detection across different imaging modalities and clinical conditions.

However, despite these developments, conventional deep learning systems rely on centralized training, where a large amount of patient data is gathered and processed in a single computing environment. Achieving such centralized data collection in the field of healthcare is complicated since retinal scans and information related to patients are sensitive and thus, prone to legal, data ownership, and institutional data sharing issues. Moreover, models built using data from a limited number of hospitals might prove ineffective when working with populations that exhibit distinctive demographic feature, types of equipment, disease incidences, or data collection practices [4].

ISSN (online): 2581-3048
Volume 10, Issue 9, pp 38-56, September-2026
https://doi.org/10.47001/IRJIET/2026.109005

Federated Learning (FL) has introduced a new paradigm of machine learning that allows numerous organizations to train a shared model without transferring data to a central server. The approach suggested by McMahan *et al.* requires that clients involved in learning only perform local training and communicate model parameters instead of sending the actual training data to the server [5]. The FL framework has garnered significant interest in the healthcare sector since it makes collaborative learning feasible without transferring sensitive patients' information [4].

However, there are many unsolved problems concerning the usage of federated learning in the classification of retinal diseases. Generally, datasets from various hospitals differ substantially in terms of the types of imaging devices, patient cohorts, prevalence of disease, annotations applied, and size of data, resulting in non-IID data across the clients, as well as influencing local updates of the model and global performance [5], [4]. Furthermore, the use of federated learning in retinal diseases is complicated by the communication burden, computational heterogenicity, privacy issues related to parameter updates received, and insufficient validation of the clinical result [5], [4].

Another major constraint is the prevalence of individual disease categorization in present-day federated retinal learning research work. Federated ophthalmic studies majorly focused on studying only single diseases or question-specific scenarios like diabetic retinopathy and glaucoma, with studies on multi-disease federated diagnostics being rare [6]- [9]. In practice, ophthalmic patients often present with several retinal problems, and hence ophthalmic diagnostic systems must distinguish multiple categories in order to provide adequate solutions. These considerations indicate the need for research work in the area of federated learning strategies that can achieve multi-disease classification in realistic conditions characterized by heterogeneous data distribution.

Even though a number of papers and reviews have been dedicated to the analysis of federated learning in healthcare, medical imaging, and ophthalmology, reports devoted specifically to its use in multi-disease retinal image classification are highly limited. The studies differ from one another in terms of their approaches to learning process, number of diseases investigated, distribution of data, architectures used, privacy measures applied, and evaluation methods employed. Therefore, a systematic review of the aforementioned aspects is necessary to get a clearer picture of the existing body of knowledge in the area and find the unresolved issues.

This paper addresses these concerns through a detailed review of the available federated learning methods applicable to retinal image-based disease classification, focussing on multi-disease diagnostic approaches. Specifically, this review makes three significant contributions: (i) it systematically analyzes existing federated learning studies in relation to their learning paradigms, data heterogeneity, and disease coverage; (ii) it develops a structured taxonomy to appropriately classify available methodologies based on relevant characteristics; and (iii) it provides an extensive discussion of the challenges that remain unsolved and the prospects for future developments in the area of sustainable, secure and applicable federated methodologies for retinal diagnostics.

II. BACKGROUND AND PRELIMINARIES

2.1 Retinal Imaging Modalities

Retinal imaging provides an important foundation for automated ophthalmic disease analysis. Fundus photography and optical coherence tomography (OCT) are among the widely used imaging modalities for retinal disease screening, diagnosis, and classification. These modalities provide complementary information about retinal appearance and structure and have been extensively used in deep learning-based retinal analysis [1]-[2].

Fundus images are widely used for screening and classification of retinal diseases such as diabetic retinopathy, glaucoma, and age-related macular degeneration. Research in deep learning has revealed that retinal fundus photographs can be used effectively for the diagnosis of diabetic retinopathy and other retinal disorders through automated image analysis [1], [3]. Because fundus photography is relatively easy to perform, it is the best technique for case finding in the mass screening of patients with retinal disease through computer-based analysis.

Optical coherence tomography (OCT) presents the images of the retina in two- and three-dimensional views and makes it possible to see each retinal layer and its pathological changes. It should be noted that OCT is more informative than two-dimensional fundus photography because it provides the additional information about the retina. De Fauw *et al.* showed the effectiveness of deep learning with the application of three-dimensional OCT images for clinical diagnosis and referral for several vision-threatening retinal diseases [2].

In spite of their advantages regarding diagnosis, there is a great discrepancy in retinal imaging technology between hospitals in practice. Variables such as the type of equipment used for imaging, the method of obtaining the scan, image clarity, lighting and the features of the patient cause differences in the resulting pictures, which is of great significance in the circumstances of federated learning, where retinal scans are distributed across various hospitals and

ISSN (online): 2581-3048
Volume 10, Issue 9, pp 38-56, September-2026
https://doi.org/10.47001/IRJIET/2026.109005

clinics. Therefore, it is important for the models that are used to classify diseases to be able to work with heterogeneous data and still give good results [4].

2.2 Multi-Disease Retinal Classification

The goal of retinal disease classification is to recognize settings by using different computer programs and deep learning and image technologies. The first research on the automated analysis of images has focused mainly on the specific diseases such as diabetic retinopathy, glaucoma, and age-related macular degeneration, and the systems based on deep learning technology showed good results in the successful classification of the mentioned diseases [1]-[2].

Despite that, it must be pointed out that retinal images may exhibit several different irritations at a time. Therefore, conducting automated analysis of one specific retinal disorder can lead to the loss of diversity of conditions that can be present in practice. Multi-disorder classification addresses this challenge by analyzing numerous retinal diseases in a single classification system, which allows producing more general assessments of retinal conditions of patients rather than focusing on one specific disease.

There is an important difference between multi-class and multi-label classifications of retina images. In traditional multi-class classification, it is possible to assign only one class to an image from a set of mutually-exclusive classes. However, in multi-label classification it is possible to associate an image with many labels at a time. This point of difference is especially important for retinal classification because retinal images may have several pathological processes at the same time.

The Retinal Fundus Multi-Disease Image Dataset (RFMiD) was developed specifically to support automatic detection of multiple retinal diseases and rare pathologies. The dataset contains 3,200 color fundus images with expert annotations covering a wide range of retinal conditions. The dataset includes 46 annotated disease or abnormality categories, and its associated challenge formulated disease or pathology classification as a 28-class multi-label classification problem [10]. The RFMiD challenge also demonstrated the importance of preprocessing, data augmentation, pretrained models, and ensemble approaches for multi-disease retinal classification [10].

Multi-disease classification introduces additional challenges compared with single-disease classification. Disease categories may have highly imbalanced sample distributions, rare pathologies may have very few training examples, and multiple diseases may occur together within the same image. These characteristics make model training and evaluation more difficult and require appropriate consideration of class imbalance, label co-occurrence, and generalization across different patient populations. Reviews of deep learning-based retinal analysis have also identified multi-disease detection as an important direction beyond conventional disease-specific classification [11].

In the context of federated learning, these challenges become more complex because multi-disease retinal data may be distributed unevenly across participating hospitals. Different institutions may observe different disease prevalences, patient populations, imaging devices, and combinations of pathological conditions. Consequently, a federated multi-disease retinal classification framework must address both the multi-label nature of the diagnostic task and the heterogeneous distribution of data across participating clients. These considerations form an important basis for the taxonomy and research-gap analysis presented in the subsequent sections.

III. FEDERATED LEARNING IN MEDICAL IMAGING

3.1 Federated Learning Architecture

Federated learning (FL) is a distributed machine learning paradigm that enables multiple clients to collaboratively train a shared model without directly transferring their local training data to a central server. The fundamental concept was introduced by McMahan *et al.*, who proposed Federated Averaging (FedAvg) for learning from decentralized and potentially non-IID datasets [5]. In this approach, the training data remain at the participating clients, while model parameters or model updates are communicated between the clients and a coordinating server.

A typical federated learning architecture consists of a central aggregation server and multiple participating clients. In a healthcare setting, the clients may represent hospitals, medical institutions, or diagnostic centres, each maintaining its own locally collected retinal image dataset. At the beginning of a federated training round, the server distributes the current global model to selected clients. Each client then trains the model using its local retinal images for a predefined number of local epochs. After local training, the clients transmit their updated model parameters or gradients to the central server rather than transmitting the underlying patient images [5], [4].

The server aggregates the received local model updates to produce an updated global model. In the conventional FedAvg approach, the aggregation is generally performed using a weighted average, where the contribution of each client is determined according to the amount of local training data. The resulting global model is subsequently distributed back to the participating clients for the next round of local training. This

ISSN (online): 2581-3048
Volume 10, Issue 9, pp 38-56, September-2026
https://doi.org/10.47001/IRJIET/2026.109005

process is repeated for multiple communication rounds until the model reaches a satisfactory convergence level [5].

The proposed architecture has a great advantage for healthcare usage as clinical centers can develop models together while not letting the patient information leave their local systems [4]. Sheller *et al.* proved that federated learning can be used in multi-institutional medical imaging collaboration employing the technology to train models without transferring patient information between institutions.

However, federated learning does not mean that all client institutions have identical data. Differences in the patient population, disease frequency, imaging instruments, procedures for obtaining images, and the number of imaging examinations lead to different and non-identical distributions of the data at various institutions. The original work on the subject notes the presence of unbalanced, and non-IID data as a specific problem of decentralized learning environments [5]. Therefore, the success of the federated learning system depends not only on the architecture of the model but also on the ability of the federated learning method to effectively operate under the conditions of variability among clients.

This system allows for the successful development of a common diagnostic model by different institutions while, at the same time, enabling institutions to keep their images in the local storage system. However, issues of communication efficiency, statistical variation of the data, and differences in availability of computing resources remain important factors that need to be discussed next.

3.2 Challenges in Healthcare Federated Learning

While federated learning creates an opportunity for effective model training without the need for exchanging patient data, its application in healthcare comes with a significant number of issues. Some of these issues include challenges posed by the differences in data distributions, computational and communications resources, privacy controls, and challenges in obtaining a trustworthy global model.

One important issue is the matter of statistical heterogeneity, also known as non-identically distributed data (Non-IID). In many healthcare applications, different hospitals may have different patient populations. This can lead to different patterns of occurrence of diseases, variations in demographic data, differences in imaging devices, acquisition protocols, and data quality. It is very possible that the local datasets kept by hospitals participating might differ significantly in terms of data distributions. This heterogeneity may put the different local model updates apart and also negatively affect convergence and generalization capabilities of any global model [12].

An additional major challenge is heterogeneous systems. Healthcare institutions that are participating in the federated learning process can differ in terms of their computational capacity, resource allocation and networking capabilities. Some clients can afford to finish local training before others, whereas some clients will have to go through local training longer due to lack of resources. That is why FedProx has been created to overcome statistical and system heterogeneities by performing different types of computations locally and introducing the proximal term to the local optimization function [12].

Another crucial drawback that Federated Learning is associated with is the communication overhead. Training requires a lot of communication between clients and the central aggregation server and is performed over several rounds. Furthermore, deep learning models have a lot of parameters, which means that model updates can require a lot of bandwidth. The communication overhead becomes even more significant due to the increasing number of participating institutions and the decreasing network connectivity [13].

Security and privacy are also key concerns. Federated learning keeps all raw patient information on local premises. However, even though the federated technique keeps the information locally, the updated model can still carry some traces of the original data used in training. Therefore, federated learning does not ensure total privacy. The secure aggregation technique makes it possible for a server to receive a summary of clients' data while keeping the information on the individual clients' contributions private [14]. Another method is the differential privacy approach, which includes noise in the original data so that it is impossible to learn about individual records or participants [15].

One more problem is that of convergence and generalization of the model. When the datasets belong to different institutions have a different structure, this might lead to a situation when the local optimization process will create different updates of the models. In this case, the worldwide model might obtain surprising results. Therefore, various federated optimization solutions (including FedProx are being studied) [12].

Such challenges are relevant to the classification of retinal diseases based on federated learning. The retinal datasets collected by different hospitals have different characteristics such as the distribution of disease types, quality of images, imaging technology, patient characteristics, and so on. Therefore, the design of a viable federated retina

ISSN (online): 2581-3048
Volume 10, Issue 9, pp 38-56, September-2026
https://doi.org/10.47001/IRJIET/2026.109005

classification system must take into consideration all items mentioned above.

IV. REVIEW METHODOLOGY

4.1 Data Sources and Search Strategy

A systematic literature review was performed to find papers on the subject of the application of federated learning to retinal image-based disease classification. The review was conducted on the most well-known databases such as IEEE Xplore, SpringerLink, ScienceDirect, PubMed, Google Scholar. The search combined all publications related to the topic between 2018 and 2025 in order to observe how robustly federated learning has developed in the area of health care.

The technique employed in this search was the use of an assortment of keywords associated with federated learning, retinal imaging, disease identification, data disparity, and privacy-focused ML. The primary keywords were "federated learning", "federated machine learning", "retinal classification", "retinal disease identification", "fundus imaging", "ophthalmic imaging", "classification of multiple diseases", "classification of multiple labels" and "privacy-based learning", etc. In this process, Boolean operators were appropriately utilized to come to the conclusions.

The initial screening was carried out on the records that were obtained through title and abstract screening to find articles relevant to the subject. The duplicate records were excluded from the screening process. Appropriate studies were then scrutinized through the full text based on the pre-determined inclusion and exclusion criteria.

4.2 Inclusion and Exclusion Criteria

Eligible studies were identified according to the following criteria: (i) the research had to be on federated learning or another decentralized learning method; (ii) the application process should concern retinal imaging, medical imaging, or some healthcare applications similar to the classification of retinal disease; (iii) the research must provide a methodological, experimental, comparative, and systematic analysis which is compliant with the demands of our review; (iv) the complete scholarly text had to be available for review, and (v) the research in question should be published in the time frame from 2018 to 2025.

Studies were eliminated if they had nothing to do with decentralized or federated learning, did not deal with retinal disease or similar applicable processes, represented previously published papers, contained insufficient methodological or experimental details, or presented non-scholarly sources. The studies for which full texts were not available were also excluded from our final analysis.

4.3 Study Selection and Analysis

The process of selecting studies included various complicated steps, among which was the identification of studies, removal of duplicate studies, screening for titles and abstracts, complete assessment of studies, evaluation for meeting inclusion criteria, and final selection of the studies. The search for studies through the databases resulted in identification of a number of different records, which were collected and checked for removing duplications. The titles and abstracts of the remaining records were evaluated to determine if any potentially relevant studies could be identified for screening. The studies meeting the eligibility criteria were included in the qualitative analysis conducting in the end.

The process of selecting studies followed the outlined steps, which were demonstrated in a PRISMA flow diagram used to illustrate all steps of the study-selection process. The PRISMA flow chart was used to summarize the procedure of searching for studies in different databases followed by removal of duplicates, title and abstract screening, complete assessment of studies, and selection of the studies into final.

For each study included, pertinent data was systematically extracted, such as the year of publication, field of application, imaging modality, disease coverage, and characteristics of datasets, as well as information on the federated learning method, client-server architecture, characteristics of data distribution, model design, privacy and security techniques, metrics, treatment of non-IID data, communication specifics, and computational limitations, along with evidence of the clinical validity.

The information taken was classified according to the basic parameters that have been defined in the review taxonomy. In particular, the parameters consist of federated learning method, data distribution characteristics, disease coverage, model design, and privacy and security mechanisms. Then the extracted data was utilized to qualitatively analyze the types of techniques available as well as detect the gaps and problems in the field of corresponding research.

V. TAXONOMY OF FEDERATED LEARNING FOR RETINAL DISEASE CLASSIFICATION

The literature reviewed shows significant differences in how federated learning systems for retinal disease classification are designed and evaluated. To allow structured learning from the available studies, this review has classified

ISSN (online): 2581-3048
Volume 10, Issue 9, pp 38-56, September-2026
https://doi.org/10.47001/IRJIET/2026.109005

previous research according to five study dimensions: federated learning strategy, data distribution characteristics, disease coverage and classification scope, model architecture and learning paradigm, and privacy and security mechanisms. This system of classification reflects the main methodological factors impacting the effectiveness, robustness, and clinical applicability of federated retinal learning systems.

5.1 Federated Learning Strategy

Federated learning methods used in retinal disease classification can be classified according to the strategy of coordinating local model training and global model aggregation. A method for determining the federated learning strategy used affects the convergence of the model, litigation requirements, sustainability of the use of heterogeneous client data, and the performance of the resulting model.

The most widely adopted approach is Federated Averaging (FedAvg), in which participating clients independently train a shared model using their local datasets and transmit the resulting model updates to a central server. The server aggregates the client updates, typically using a weighted average based on the number of local training samples, and distributes the updated global model to the clients for the next communication round [5]. FedAvg provides a fundamental baseline for collaborative federated learning and has been widely used in medical imaging applications [4].

However, the performance of conventional FedAvg can be affected when client datasets are highly heterogeneous. Healthcare institutions may differ substantially in patient populations, disease prevalence, imaging equipment, acquisition protocols, and data volume. Federated optimization approaches such as FedProx introduce additional mechanisms to improve learning under heterogeneous client conditions by constraining local optimization and reducing divergence among client updates [12]. The importance of addressing non-IID data in federated learning has also been demonstrated in earlier studies, where severe differences in local class distributions were shown to affect model performance and convergence [16].

Federated learning strategies can also be differentiated according to the degree of personalization provided to individual clients. Global federated learning attempts to obtain a common model that is shared across participating institutions, whereas personalized federated learning allows the learned model or selected model components to adapt to the characteristics of individual clients. Personalized strategies are potentially useful in retinal applications where hospitals may have substantially different disease distributions or imaging characteristics.

Another dimension is the degree of client participation during each communication round. In practical federated environments, it may not always be possible for every institution to participate in every round because of computational limitations, network availability, or institutional constraints. Consequently, federated strategies may employ partial client participation while maintaining iterative global model aggregation [5].

The reviewed retinal studies also demonstrate the use of different aggregation strategies beyond basic FedAvg. For example, federated retinal classification studies have investigated FedProx and other adaptive aggregation approaches to address heterogeneous client data and improve the robustness of collaborative training [12], [17], [9]. Therefore, the selection of a federated learning strategy should be considered in relation to the characteristics of the participating institutions and their local datasets.

Within the proposed taxonomy, federated retinal learning approaches are therefore classified into conventional global aggregation approaches, heterogeneity-aware approaches, personalized approaches, and other adaptive federated strategies. This categorization provides a basis for comparing the assumptions, advantages, and limitations of different federated learning strategies.

5.2 Data Distribution Characteristics

The distribution of training data across participating clients is an important dimension for categorizing federated learning approaches. In healthcare environments, data are rarely distributed identically across institutions because hospitals may serve different patient populations, use different imaging equipment, and follow different data acquisition and annotation practices. Consequently, federated retinal disease classification must consider the statistical characteristics of data available at each participating institution.

Federated learning studies can initially be classified according to whether the data distribution across clients is assumed to be independent and identically distributed (IID) or non-identically distributed (non-IID). Under an IID setting, participating clients are assumed to have approximately similar data distributions. This provides a relatively controlled environment for evaluating federated algorithms but may not accurately represent real-world healthcare conditions. In contrast, non-IID settings allow client datasets to differ in their underlying distributions and are therefore more representative of multi-institutional healthcare environments [16].

Non-IID data can occur in several forms. Label distribution skew occurs when the proportions of disease classes differ across clients. For example, one hospital may

ISSN (online): 2581-3048
Volume 10, Issue 9, pp 38-56, September-2026
https://doi.org/10.47001/IRJIET/2026.109005

contain a larger proportion of diabetic retinopathy cases, whereas another may contain more glaucoma or normal retinal images. Feature distribution skew occurs when images acquired at different institutions exhibit variations in visual characteristics because of differences in imaging devices, acquisition protocols, illumination, image quality, or patient populations. Quantity skew occurs when participating institutions possess substantially different numbers of training samples. These forms of heterogeneity may occur simultaneously in practical healthcare settings [12], [16].

For retinal disease classification, label distribution skew is particularly important because disease prevalence may vary substantially between hospitals. In a multi-disease setting, some institutions may contain sufficient samples for common diseases while having very few examples of rare retinal conditions. Consequently, local models may learn different disease-related patterns, resulting in divergent local updates during federated optimization.

Feature heterogeneity is also relevant to retinal imaging because images obtained using different fundus cameras or imaging protocols may exhibit variations in resolution, illumination, color characteristics, field of view, and image quality. These variations can create domain differences between clients even when the disease categories are identical. Therefore, a federated model may need to learn disease-related features that remain robust across different institutional imaging environments. Such heterogeneity is a recognized challenge in federated medical imaging systems [4].

The degree of data heterogeneity has important implications for federated model training. Under strongly non-IID conditions, local models may move toward client-specific solutions rather than a common global optimum. This can result in client drift, slower convergence, and reduced global model performance [12], [16]. Data heterogeneity across institutions is therefore an important factor affecting convergence and generalization in federated systems.

Within the taxonomy proposed in this review, federated retinal learning studies are therefore categorized according to their treatment of data distribution, including IID assumptions, naturally occurring non-IID distributions, artificially simulated non-IID distributions, and approaches specifically designed to mitigate data heterogeneity. This categorization helps distinguish studies evaluated under simplified experimental conditions from those attempting to model the heterogeneous data distributions encountered in real clinical environments.

5.3 Disease Coverage and Classification Scope

The disease coverage of a federated retinal classification system represents another important dimension for organizing existing studies. Current approaches can be broadly categorized according to whether they address a single retinal disease, multiple diseases using separate disease-specific tasks, or multiple diseases within a unified classification framework. This distinction is important because the complexity of the diagnostic task increases as the number and diversity of disease categories increase.

Single-disease classification focuses on the detection or grading of one particular retinal condition. Diabetic retinopathy is one of the frequently investigated examples, with deep learning systems demonstrating the feasibility of automated disease detection from retinal images [1]. Federated learning can extend such disease-specific approaches by enabling multiple institutions to collaboratively train a model without transferring their local retinal images [4].

Multi-disease classification considers several retinal diseases within a common learning framework. Such approaches can provide broader diagnostic coverage and may better reflect clinical environments in which patients can present with different retinal abnormalities. The RFMiD dataset, for example, was developed to support the detection of multiple retinal diseases and pathological conditions and provides a basis for studying multi-disease retinal image analysis [10].

Within multi-disease classification, it is important to distinguish multi-class and multi-label learning. In multi-class classification, each image is assigned to one mutually exclusive class. In multi-label classification, an individual image can simultaneously be associated with multiple disease or pathology labels. This distinction is particularly relevant to retinal image analysis because multiple abnormalities may occur in the same retinal image. The RFMiD challenge specifically demonstrated the relevance of multi-label formulation for multi-disease retinal pathology detection [10].

Federated retinal classification studies can therefore be categorized according to their disease coverage and classification formulation, including single-disease classification, multi-class multi-disease classification, and multi-label multi-disease classification. Studies can also differ in whether normal retinal images are included as an explicit class and whether rare disease categories are retained or grouped with other conditions.

Disease imbalance is an additional consideration within this taxonomy. Common retinal diseases may be represented by substantially more samples than rare pathologies. When such imbalanced distributions are combined with client-level heterogeneity, some participating institutions may contain only a small number of examples for particular disease

ISSN (online): 2581-3048
Volume 10, Issue 9, pp 38-56, September-2026
https://doi.org/10.47001/IRJIET/2026.109005

categories. This can affect local model training and may lead to unequal performance across disease classes.

Recent work has begun to address multi-disease retinal classification directly within federated settings. For example, Nabil *et al*. evaluated federated learning for seven retinal pathologies using OCT angiography and investigated both controlled and heterogeneous data distributions, including a Dirichlet-based non-IID setting [9].

The disease-coverage dimension therefore provides an important basis for evaluating the clinical scope of federated retinal learning systems. In particular, approaches that address multiple diseases and multiple simultaneous labels have greater relevance to comprehensive retinal screening, but they also introduce additional challenges related to class imbalance, label co-occurrence, and heterogeneous disease distributions across participating institutions.

5.4 Model Architecture and Learning Paradigm

Federated retinal disease classification approaches can also be categorized according to the model architecture and learning paradigm adopted by the participating clients. Since federated learning is an optimization and collaboration framework rather than a specific neural network architecture, different studies can employ different deep learning models while following the same client-server training process.

Convolutional neural networks (CNNs) have been widely used for retinal image analysis because of their ability to learn hierarchical visual representations directly from retinal images [1], [3]. CNN-based architectures can be trained locally at participating institutions, after which their model parameters or updates are aggregated by the federated server. The resulting global model can then be redistributed to the participating clients for subsequent training rounds.

The architectures used in retinal classification can include conventional CNNs as well as deeper pretrained networks and transfer-learning-based models. Transfer learning is particularly useful when retinal datasets are limited because a model pretrained on a large image dataset can provide a useful initialization for local training. In a federated setting, the same pretrained architecture can be distributed to participating clients, while subsequent optimization is performed locally using institution-specific retinal datasets.

Federated studies can also be distinguished according to the learning paradigm used during local training. Centralized global learning aims to obtain a single model shared across all clients, whereas personalized learning allows the learned model to adapt to individual client characteristics. Other approaches may combine global and client-specific components to achieve a balance between generalization across institutions and adaptation to local data distributions.

The choice of architecture and learning paradigm can influence the computational requirements, communication cost, convergence behaviour, and classification performance of a federated system. Large and computationally intensive models may provide stronger feature representations but can increase the amount of information exchanged during federated communication and impose greater computational demands on participating healthcare institutions [13].

For retinal disease classification, the architecture should therefore be considered together with the characteristics of the distributed datasets. A model that performs well under a centralized or homogeneous setting may not necessarily maintain the same performance when trained across institutions with different disease distributions and imaging characteristics. Accordingly, the proposed taxonomy considers both the underlying model architecture and the learning paradigm used to coordinate global and local learning.

5.5 Privacy and Security Mechanisms

Privacy and security mechanisms represent an important dimension for categorizing federated learning approaches in healthcare. Although federated learning avoids the direct transfer of raw patient data between participating institutions, the exchange of model parameters or updates may still create potential privacy risks. Therefore, additional protection mechanisms may be incorporated to strengthen the confidentiality of distributed medical data [4], [14], [15].

One important mechanism is secure aggregation, which enables a central server to obtain an aggregated model update from multiple participating clients without directly accessing the individual contribution of each client. Secure aggregation can therefore reduce the exposure of client-specific model updates during the federated training process [14].

Another commonly investigated mechanism is differential privacy (DP). Differential privacy introduces carefully controlled random noise into data, gradients, or model updates to reduce the possibility of inferring information about individual training records. In federated learning, differential privacy can be utilized at different levels and also at the user level, allowing for the extra level of privacy protection during joint training [15]. Privacy-preserving federated learning is especially critical for the realm of medical imaging since data of patients is protected by strict ethical, legal, and conceptual regulations.

The need for protection of privacy can create the need to compromise between security and model utility. Therefore,

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 10, Issue 9, pp 38-56, September-2026

https://doi.org/10.47001/IRJIET/2026.109005

any increase in the level of privacy protection can change the data in the updates to the model, which may affect its classification performance.

Thus, the federated learning approaches can be divided into approaches that mainly use the data-localization property of federated learning system and those that employ additional methods such as secure aggregation and differential privacy. Some approaches can use a number of different privacy-protecting methods.

In the case of federated classification of retinal diseases, privacy relates to various issues. A good federal system must take into account both classification quality and the level of privacy protection, computational costs, communication requirements, and privacy mechanisms influence on accuracy levels.

In this connection, we establish an important privacy and security aspect of the presented classification which would allow comparing the studies in federated retinal learning in terms of privacy mechanisms as well as finding out the trade-off among privacy, security, computational costs, and classification accuracy.

VI. COMPARATIVE ANALYSIS OF EXISTING STUDIES

The literature about federated learning and retinal imaging proves the viability of the model of collaborative building of the model across several healthcare institutions in a way of not exchanging raw images of the patients. But the studies review is still different from each other from the perspective of methods used for imaging, diseases covered, federated learning techniques applied, data distribution, model architecture, privacy mechanisms, and assessment methods.

Table 1 summarizes representative studies that illustrate the development of federated learning from general medical imaging applications to disease-specific ophthalmic applications and, more recently, multi-disease retinal classification.

6.1 Evolution of Federated Learning in Retinal and Ophthalmic Imaging

In the early days of medical-imaging analysis, researchers successfully performed federated learning, which involved working collaboratively without sharing patient data in a centralized way. Sheller *et al.*, for example, demonstrated multi-center federated learning to identify brain tumors and showed that federated learning can be carried out successfully while ensuring patient data remains at hospitals [4]. Although the study does not address retinal imaging, it is crucial for understanding the role of federated learning in multidisciplinary medical imaging.

Later studies further examined federated learning in retinal imaging. Lo *et al.* applied this technique to OCT data to classify diabetic retinopathy and microvasculature segmentation at several institutions [6]. The outcome was that the federated models were successfully able to work as accurately as the internal models, thus confirming the possibility of conducting collaborative retinal imaging without any direct data sharing.

The research conducted by Hanif *et al.* explores federated learning in multi-centre ophthalmic collaboration and its implications in clinical diagnosis and epidemiology of diseases [18]. The results of this research show that federated learning can enable collaboration between the different ophthalmic centres while keeping local data ownership intact. These studies imply the evolution of federated learning from being demonstrated in medical imaging to its application in ophthalmic systems. Still, most of those pioneering research works concentrated on one disease or one clinical task which limited the application of those studies for representing the greater multi-disease diagnosis of retina screening.

6.2 Comparison of Imaging Modalities and Disease Coverage

The reviewed studies employ several retinal and ophthalmic imaging modalities, including fundus photography, OCT, and OCT angiography (OCTA). The choice of imaging modality influences the type of disease information available to the model and the nature of the data heterogeneity encountered across institutions.

OCT-based federated learning has been investigated for conditions such as diabetic retinopathy and glaucoma. Lo *et al.* demonstrated federated classification and microvasculature segmentation using OCT data [6], while Ran *et al.* developed a multicentre federated framework for glaucoma detection using three-dimensional OCT data [7]. The latter study involved seven eye centres and evaluated the federated model on two unseen datasets, providing stronger evidence of cross-site and external generalization than studies evaluated only on local test data [7].

Fundus photography has also been used in federated ophthalmic learning. Tang et al. investigated privacy-preserving federated learning with domain adaptation for multi-disease ocular disease recognition using fundus images [17]. This work is particularly relevant to the present review because it considers both disease diversity and domain differences between participating sites.

ISSN (online): 2581-3048
Volume 10, Issue 9, pp 38-56, September-2026
https://doi.org/10.47001/IRJIET/2026.109005

In contrast, Nabil *et al.* investigated federated learning specifically for multi-disease retinal classification using OCTA [9]. The study evaluated five aggregation strategies—FedAvg, FedAdagrad, FedYogi, FedProx, and FedMRI—and considered seven retinal pathologies using combined public and private data. The study also evaluated controlled and heterogeneous data distributions, including a Dirichlet-based non-IID setting, making it particularly relevant to the multi-disease and non-IID dimensions of the present review.

Overall, the literature indicates that OCT and fundus imaging dominate earlier federated ophthalmic research, while recent work has begun extending federated learning toward OCTA-based multi-disease classification. However, the number of studies addressing several diseases simultaneously remains substantially smaller than the number addressing individual disease-specific tasks.

6.3 Comparison of Federated Learning Strategies

Dhillon, S., C2D, and C3D have employed different aggregation strategies for federated learning, among which FedAvg is a widely accepted approach thanks to its straightforward weighted averaging method [5]. In numerous studies involving eyes, FedAvg has been found to be the basic and the most popular aggregation method used in practice.

FedProx has been used by Ran, Xue, and Ma in their work on the detection of glaucoma with the help of three-dimensional OCT images [7]. The usage of FedProx was due to the requirement of taking into consideration variances in the characteristics of the hospitals involved and the study showed steady results on all involved hospitals and other previously unseen data [7].

Baptista and colleagues conducted research analysing various types of federated approaches like FedAvg, FedSGD, FedProx, FedBN, and FedYogi. They applied these methods to the diagnosis of glaucoma using distributed fundus datasets [8]. The outcome of this research established an important principle in that it showed how the selection of a certain aggregation method may cause variations in the process of federated learning under diverse conditions of clients' activity.

Tang and others incorporated federated learning into domain adaptation in the medical field in their study of the multi-disease ocular recognition problems [17]. The idea behind this study is to know how to unify knowledge from different institutions provided that their data may be entirely different.

More recently, Nabil and colleagues explored the comparison of the aggregation methods like FedAvg, FedAdagrad, FedYogi, FedProx and FedMRI in multi-disease retinal diagnosis [9]. They established various results in that some of the aggregation methods were better suitable for one evaluation criterion while the others were superior in terms of other criteria.

These findings suggest that aggregation strategy should be selected according to the characteristics of the distributed data rather than treated as an independent algorithmic choice. In particular, increasing client heterogeneity may require optimization or personalization mechanisms beyond conventional FedAvg.

6.4 Comparison of Data Heterogeneity

Data heterogeneity represents one of the most important differences between experimental federated learning studies. The reviewed literature includes studies using genuine multi-institutional datasets as well as studies that simulate client distributions by partitioning available datasets.

Multi-institutional studies provide a more realistic representation of clinical variability because differences in patient populations, imaging devices, acquisition protocols, and institutional practices can naturally occur across participating sites. Ran *et al.*, for example, used data from seven eye centres and additionally evaluated the model using unseen datasets, providing evidence of generalization beyond the participating training centres [7].

However, not all studies can obtain sufficiently large datasets from multiple institutions. Thus, non-IID simulated distributions are often utilized to test the reliability of various federated algorithms. Nabil *et al.* performed a proper assessment of heterogeneous data using the Dirichlet distribution method with $\alpha = 0.5$ for non-IID analysis concerning the problem of multi-disease classification in retina imaging [9].

While the experiments with simulated non-IID distributions play an important role in controlled comparisons, they don't always account for various types of clinical diversity. Real-life institutions may face different disease distributions, types of imaging systems, the quality of images and patients' characteristics, among other things. Thus, the results of artificial data partitioning experiments have to be treated differently from those obtained with the actual clinical databases.

The comparison thus highlights an opportunity for future investigations to present both the approach used to generate client distributions and the attributes of those distributions. Such disclosure would facilitate the differentiation between performance obtained in simplified experimental settings and performance achieved in clinical heterogeneity situations.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 10, Issue 9, pp 38-56, September-2026

https://doi.org/10.47001/IRJIET/2026.109005

6.5 Comparison of Privacy and Security Considerations

Privacy protection is of utmost importance in the context of federated learning and applications in medical imaging. In the studies reviewed in the paper, there is generally no transfer of raw patient data. Instead, model parameters or updates are shared between collaborating medical institutions and a main coordination entity [4], [14], [15].

Nevertheless, there are significant differences between the works in terms of implemented privacy-preserving mechanisms. While some methods mostly rely on performing local computation on the client's side, other techniques apply additional approaches, including secure aggregation, differential privacy, and others. For instance, Tang et al. used a randomized mechanism employed within a privacy-preserving federated learning framework for different diseases in the area of ocular recognition [17]. Nabil *et al.* used a more sophisticated approach in their federated learning based on ophthalmic imaging that included differential privacy and secure aggregation [9].

One major weakness is that confidentiality is not necessarily evaluated along with performance. The methods used to protect privacy may involve additional computation and can also affect modifications in the models, thus impairing classification performance. It means that stating merely the results of classification is insufficient for complete evaluation of a federated system applicability.

For this reason, future comparative studies must include information on confidentiality preservation, communication load, computation requirements, and performance characteristics of the models together.

6.6 Overall Comparative Findings

The comparison of the reviewed studies reveals several consistent trends.

First, federated learning has progressed from general medical-imaging proof-of-concept studies toward specialized ophthalmic applications. Early work established the feasibility of collaborative medical-imaging learning without centralized patient-data sharing [4], followed by federated studies targeting diabetic retinopathy, glaucoma, retinopathy of prematurity, and other ophthalmic tasks [6]-[8].

Second, single-disease classification remains more common than unified multi-disease classification. Diabetic retinopathy and glaucoma have received substantial attention, whereas relatively few studies address multiple retinal diseases within a common federated framework. Recent OCTA-based work by Nabil *et al.* represents an important step toward federated multi-disease retinal classification [9].

Third, data heterogeneity is increasingly recognized as an important factor affecting federated retinal learning. Studies have used multi-centre datasets, simulated non-IID distributions, domain adaptation, and heterogeneity-aware aggregation strategies. Nevertheless, the degree to which these experimental settings reproduce naturally occurring clinical heterogeneity varies considerably.

Fourth, the choice of federated strategy is becoming increasingly diverse. FedAvg remains an important baseline, while FedProx, FedBN, FedYogi, FedAdagrad, FedMRI, and domain-adaptive approaches have been investigated for addressing different aspects of client heterogeneity and model optimization [8], [9], [12], [17].

Finally, evaluation practices remain inconsistent across studies. Some studies evaluate only local or participating-site performance, whereas others include unseen external datasets or explicitly examine heterogeneous data distributions. Consequently, direct comparison of reported performance values across studies is difficult because datasets, disease categories, imaging modalities, client configurations, evaluation metrics, and experimental conditions are not standardized.

Overall, the comparative evidence indicates that existing federated retinal learning research has established the feasibility of privacy-preserving collaborative diagnosis but has not yet fully addressed the combination of multi-disease classification, realistic non-IID data, privacy protection, communication efficiency, fairness, and external clinical validation within a single standardized framework. These limitations provide the basis for the research gaps discussed in the next section.

Table 1: Federated Learning Studies in Retinal/Ophthalmic Disease Analysis

Study	**Imaging modality**	**Disease / task**	**Dataset / setting**	**FL approach**	**Data heterogeneity**	**Privacy / security**	**Main observation**
Sheller *et al.* (2020) [4]	MRI/medical imaging	Brain tumour segmentation	Multi-institutional medical imaging	Federated learning	Multi-institutional	Data retained locally	Demonstrated feasibility of collaborative medical-image learning without sharing patient data

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 10, Issue 9, pp 38-56, September-2026

https://doi.org/10.47001/IRJIET/2026.109005

Lo *et al.* (2021) [6]	OCT / OCTA	Diabetic retinopathy classification and microvasculature segmentation	Multi-centre OCT data	Federated learning	Cross-site variation	Federated training	Federated models achieved performance close to internally trained models
Hanif *et al.* (2022) [18]	Ophthalmic imaging	Multi-centre ophthalmic diagnosis / epidemiology	Multi-centre ophthalmic data	Federated learning	Multi-centre heterogeneity	Privacy-preserving FL	Demonstrated the potential of FL for multi-centre ophthalmic collaboration
Tang et al. (2024) [17]	Fundus photography	Multi-disease ocular disease recognition	Multi-site fundus dataset	FL + domain adaptation	Domain shift across sites	Gaussian randomized mechanism	Combined privacy protection with domain adaptation for multi-disease ocular recognition
Ran *et al.* (2024) [7]	3D OCT	Glaucoma detection	Seven-client multicentre setting + unseen datasets	FedProx	Multi-centre heterogeneity	Privacy-preserving FL	Reported strong generalization across participating and unseen datasets
Baptista *et al.* (2023) [8]	Fundus photography	Glaucoma classification	Distributed retinal fundus datasets	FedAvg, FedSGD, FedProx, FedBN, FedYogi	Client heterogeneity	Federated learning	Compared multiple FL strategies for glaucoma diagnosis
Nabil *et al.* (2025) [9]	OCTA	Seven-class multi-disease retinal classification	OCTA-500 + private UIC data	FedAvg, FedProx, FedAdagrad, FedYogi, FedMRI	Controlled + Dirichlet non-IID	Differential privacy + secure aggregation	Systematically evaluated aggregation, architectures, heterogeneity, privacy and utility

VII. OPEN CHALLENGES AND RESEARCH GAPS

The comparative analysis of existing federated learning studies reveals that, although collaborative retinal image analysis has demonstrated considerable potential, several methodological and clinical limitations remain unresolved. The major challenges concern realistic modelling of non-IID data, limited multi-disease learning, incomplete evaluation of privacy-utility trade-offs, communication and computational constraints, limited clinical validation, and the absence of standardized benchmarks. These gaps indicate that future federated retinal learning systems must move beyond proof-of-concept demonstrations toward robust, heterogeneous, privacy-aware, and clinically validated frameworks.

7.1 Limited Modelling of Realistic Non-IID Data

One significant gap within research is insufficient modelling of realistic non-identially distributed data. Despite federated learning technology being specifically aimed at decentralised data environments, many researchers studied federated retinal models using manipulated datasets or technology-controlled datasets. These approaches help with algorithm development, but they do not reflect the complexity of heterogeneity existing in real healthcare institutions [16].

In practical clinical environments, differences between institutions can arise from patient demographics, disease prevalence, imaging devices, acquisition protocols, image quality, annotation practices, referral patterns, and the number of available training samples. These factors can produce simultaneous variations in label, feature, and quantity distributions. Consequently, the local data available at one hospital may differ substantially from those available at another institution.

This limitation is particularly important for retinal disease classification. One institution may contain a high proportion of diabetic retinopathy cases, whereas another may observe more glaucoma, age-related macular degeneration, or normal retinal images. Rare disease categories may also be represented by very few samples at individual institutions. Such differences can influence local model updates and produce client drift during federated optimization [12], [16].

Recent retinal FL research has begun to address this issue by explicitly introducing heterogeneous client distributions.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 10, Issue 9, pp 38-56, September-2026

https://doi.org/10.47001/IRJIET/2026.109005

For example, Nabil et al. evaluated multi-disease retinal classification under a Dirichlet-based non-IID distribution with α = 0.5, providing a controlled representation of heterogeneous disease distributions [9]. However, simulated distributions remain an approximation of real clinical heterogeneity. These considerations further emphasize the importance of evaluating both feature-distribution skew and label-distribution imbalance in multicohort retinal data.

Therefore, an important research need is the development and evaluation of federated retinal classification frameworks using realistic multi-institutional data distributions. Future studies should report the characteristics of client-level distributions transparently and distinguish between naturally occurring and artificially simulated heterogeneity. Standardized protocols for evaluating different degrees and types of non-IID data would further improve reproducibility and enable meaningful comparison among federated retinal learning approaches.

7.2 Predominance of Single-Disease Federated Models

A second major research gap is the predominance of federated learning frameworks developed for individual retinal diseases. Many existing studies focus on specific diagnostic tasks such as diabetic retinopathy or glaucoma rather than addressing several diseases within a unified federated framework [6], [7], [8].

Disease-specific models provide valuable evidence regarding the feasibility of federated ophthalmic learning, but they do not fully represent the broader requirements of retinal screening. In real clinical practice, patients may have various problems with their retinas, thus a good diagnosis system should be able to differentiate multiple problems based on the symptoms the patient is presenting.

Multi-disease classification in the federated learning frameworks presents new challenges, such as class imbalance, various rare diseases, overlapping pathological properties, co-occurrence of labels as well as the variation in prevalence of diseases in different facilities. The situation is complicated by the unequal frequencies of diseases among the federated learners.

Recent research has shown that multi-disease classification is possible with the help of federated learning. For instance, Nabil and his colleagues have studied the presence of seven diseases of the retina in patients through the analysis of OCTA, comparing different models in federated learning [9]. Nevertheless, the existing evidence of the effectiveness of such methods is scarce in comparison to multi-disease frameworks.

That is why there is a clear need for a federated model able to handle several retinal conditions at the same time, especially when performing multi-label classification, which means that the same image may have multiple pathological problems.

7.3 Inadequate Evaluation of Privacy–Utility Trade-offs

Preserving privacy is one of the main reasons for using federated learning in the health care sector. However, simply keeping the original images locally does not automatically lead to full privacy as the information in the model parameters, gradients, and other transmitted data can still be used against the healthcare provider [14], [15].

Present-day work has studied mechanisms, including but not limited to secure aggregation and differential privacy techniques, yet these methods may increase computational load or diminish the predictive utility of the model. Therefore, there exists a conflict between privacy and the efficiency of the model.

An important issue with observational research in retinal FL is that privacy and diagnostics effectiveness are seldom evaluated together. While some papers assess model performance, other studies focus on privacy mechanisms and do not analyze how they influence the model's predictive performance.

Recent research in FL in terms of the retina has started to close this gap. Nabil et al. have studied the privacy techniques in terms of the model performance through multi-disease classification [9].

Starting from here, it is also needed to include to evaluations your results of the privacy protection and diagnostic efficacy permanently. You should also take into account the privacy budget if differential privacy is used, the costs related to communications and computations, the security of the system against different attacks, etc., as well as efficiency of the Classification under different privacy conditions. This would allow obtaining more valuable assessments connected to the balance between privacy and utility in the context of using federated retinal learning.

7.4 Communication and Computational Constraints

Federated learning requires repeated communication between participating institutions and the central aggregation server. For deep learning models with large numbers of parameters, transmitting model updates over many communication rounds can create substantial communication overhead [5], [13].

ISSN (online): 2581-3048
Volume 10, Issue 9, pp 38-56, September-2026
https://doi.org/10.47001/IRJIET/2026.109005

The problem becomes more significant in retinal image classification because modern CNN and other deep learning architectures may contain millions of parameters. Healthcare institutions may also differ substantially in available computational resources, network connectivity, storage capacity, and local processing capabilities. Consequently, some clients may complete local training rapidly while resource-constrained institutions may become communication or computation bottlenecks.

Several federated optimization and communication-efficient approaches have been proposed to reduce these limitations [12], [13]. However, retinal FL studies do not consistently report communication volume, number of communication rounds, local computational requirements, training time, or energy consumption. As a result, it is difficult to determine whether a reported improvement in classification performance is practically achievable in resource-constrained clinical environments.

Future federated retinal systems should therefore jointly evaluate diagnostic performance and system efficiency. Communication compression, adaptive client participation, efficient aggregation, lightweight architectures, asynchronous training, and parameter-efficient learning represent potential directions for reducing the computational and communication burden.

7.5 Limited Clinical Validation and Deployment Studies

On the other hand, another major research gap is a lack of validation of federated retinal learning systems. The point is that most of the studies are retrospective without the application of clinical datasets or controlled experiments and consequently do not provide convincing evidence of the model's applicability.

The inability to apply the developed models to the clinical practice decreases the significance of the research findings [7].

Nevertheless, potential clinical validation is still insufficient. A functional rollout also needs to take into account the integration of the workflow, understandability, interchangeability, monitoring, acceptance by doctors, and liability for breakdowns in the model.

Therefore, future investigations should get from retrospective proof-of-concept research to external and prospective validation. It is not enough just to check the quality of federated retina systems by standard classification metrics, they should also be assessed based on clinically significant criteria such as sensitivity for specific diseases, calibration, performance for subgroups, generalisability, and decision support.

7.6 Absence of Standardized Benchmarks

The last big research gap is the absence of standard mechanisms for evaluation of federated systems for retina disease classification. The discussed projects use different datasets, imaging modalities, disease categories, number of patients, data partitioning methods, federated algorithms, evaluation methods, privacy measures, and arrangements of experiments.

This difference in the presented projects makes direct comparisons of results complicated. For instance, a model that is tested with the help of a limited number of simulated patients in organized non-IID distribution cannot be compared to a model trained on a network of institutes that have naturally heterogeneous information. Similarly, performance values obtained for single-disease classification cannot be directly compared with multi-label multi-disease classification results.

The absence of common benchmarks also makes it difficult to determine whether improvements result from the federated learning strategy itself, differences in model architecture, dataset characteristics, data partitioning, or experimental configuration. Such methodological differences highlight the need for stronger evaluation standards.

A standardized benchmark for federated retinal classification should ideally specify common datasets or well-documented multi-institutional settings, client configurations, disease distributions, IID and non-IID scenarios, evaluation metrics, privacy settings, communication measures, and external validation procedures. Such a benchmark would improve reproducibility and allow researchers to compare federated approaches under consistent experimental conditions.

In summary, the identified gaps call for interventions that take into account various issues and challenges that the next generation of federated retinal image classification systems encounters, including heterogeneous data, multiple health conditions, balancing privacy and utility, and meeting clinical requirements.

VIII. FUTURE RESEARCH DIRECTIONS

In the previous sections we have analyzed the existing research gaps, which reveal that there is a need for further development of federated retinal imaging systems. Future research projects should focus on implementing advances in this technology and addressing issues related to non-

ISSN (online): 2581-3048
Volume 10, Issue 9, pp 38-56, September-2026
https://doi.org/10.47001/IRJIET/2026.109005

homogeneous clinical data, multiple diseases, privacy concerns, and communication obstacles. As a result, six primary research directions can be defined: personalized federated learning, multi-disease and multi-label learning, effective use of non-IID data, development of communication-efficient federated systems, privacy-centered learning, and approach to clinical implementation and validation.

8.1 Personalized Federated Learning for Retinal Imaging

Personalized federated learning is an essential approach to solving the problem of heterogeneous healthcare institutions. Traditional federated learning aims to develop global models that can be utilized by all clients of the system. Nevertheless, datasets on retinal images acquired from different healthcare institutions can have significant differences when it comes to the disease distribution, patient population, imaging devices, acquisition protocols, and image properties. Therefore, one global model cannot achieve the same level of performance at all sites [12].

Personalized federated learning solves this problem by enabling the application of the information obtained from different clients to the specific conditions of particular healthcare institutions. While traditional approaches usually assume that all clients use the same model, the personalized approach enables researchers to create a world-class model through a combination of universally applicable characteristics and local patterns, which allows for knowledge transfer while making it applicable to the unique data characteristics.

For retinal image classification, personalization can be particularly useful when participating hospitals have substantially different disease distributions. For example, one institution may have a higher proportion of diabetic retinopathy cases, whereas another may contain more glaucoma or age-related retinal diseases. Personalized strategies may therefore reduce the negative effects of statistical heterogeneity and improve client-specific performance while retaining the benefits of collaborative learning.

Future research should investigate personalized federated architectures specifically designed for multi-disease retinal classification under realistic non-IID conditions. Evaluation should include both global and client-specific performance, together with disease-wise sensitivity, specificity, F1-score, communication cost, convergence behaviour, and generalization to unseen institutions. Such evaluation would help determine whether personalization provides consistent benefits across different clinical environments.

Overall, personalized federated learning provides a promising direction for developing retinal diagnostic systems that can learn collaboratively while adapting to the characteristics of individual healthcare institutions. Its integration with multi-disease classification, non-IID modelling, privacy-preserving mechanisms, and robust deep learning architectures represents an important area for future investigation.

8.2 Federated Multi-Disease and Multi-Label Learning

Future federated retinal learning systems should move beyond disease-specific classification toward unified frameworks capable of handling multiple retinal diseases within the same learning environment. Classification of multiple diseases is crucial since a retinal image can have several diseases and symptoms from the different diseases can be present on clinical images. Hence, a federated system could enhance diagnostic capabilities while helping the institutions involved in the study work together and learn from each other's data without exchanging raw retinal images.

Classification of multiple diseases is crucial since a retinal image can have several diseases and symptoms from the different diseases can be present on clinical images. Hence, a federated system could enhance diagnostic capabilities while helping the institutions involved in the study work together and learn from each other's data without exchanging raw retinal images[10].

The classification of numerous illnesses is vital because a retinal photograph may exhibit multiple conditions. Moreover, various diseases' symptoms might be included in the clinical photographs. Therefore, a federated system can contribute to the improvement of diagnostic abilities and allow organizations participating in the research to cooperate and share their knowledge without revealing retinal images.

Despite the advantages of multi-illness federated systems in comparison to single-illness systems, many problems arise in the process of implementing multi-disease learning. Some of them include the limitations of systems based on class mismatch, disease coexistence, and different disease frequency among clients.

In order to address these issues, studies about federated structures should be conducted as they are meant for classifying several illnesses, and they need to take into account class problems and disease coexistence issues. It means that the data on the efficiency of the system should be based not only on the data on accuracy.

Overall, federated multi-disease and multi-label learning represents an important direction for developing comprehensive retinal diagnostic systems. Combining multi-disease learning with realistic non-IID distributions,

ISSN (online): 2581-3048
Volume 10, Issue 9, pp 38-56, September-2026
https://doi.org/10.47001/IRJIET/2026.109005

personalized learning, privacy-preserving mechanisms, and robust deep learning architectures could provide a stronger foundation for clinically applicable federated retinal AI.

8.3 Realistic Modelling of Non-IID Clinical Data

Future federated learning research should place greater emphasis on modelling realistic non-IID data distributions that reflect the heterogeneity encountered across healthcare institutions. In real clinical environments, participating hospitals rarely have identical patient populations, disease prevalence, imaging devices, acquisition protocols, image quality, or dataset sizes. These differences can produce substantial variations in both feature and label distributions across clients and may negatively affect the performance and generalization of a federated model [12], [16].

For retinal image classification, non-IID conditions should therefore represent clinically meaningful sources of heterogeneity rather than relying only on randomly divided datasets. Important forms of heterogeneity include label distribution skew, feature distribution skew, quantity skew, and domain shift. For example, different hospitals may have different proportions of diabetic retinopathy, glaucoma, age-related macular degeneration, and normal retinal images. Differences in fundus cameras, imaging protocols, patient demographics, and image quality may additionally introduce feature-level and domain-level variations.

Future studies should evaluate federated models under controlled but realistic client distributions. Strategies such as Dirichlet-based partitioning can be used to simulate varying degrees of label imbalance, while additional variations in image characteristics and client sample sizes can be incorporated to better represent clinical conditions. Nabil *et al.* provide an example by evaluating seven retinal pathologies under heterogeneous conditions using Dirichlet distribution partitioning with $\alpha = 0.5$ [9].

Importantly, evaluation should not be limited to overall global accuracy. Future experiments should report client-level and disease-wise performance together with sensitivity, specificity, precision, recall, F1-score, and AUC. Performance variation between clients should also be examined to determine whether a federated model benefits all participating institutions or primarily improves performance for clients with larger or more representative datasets.

Therefore, realistic non-IID modelling should become a standard component of federated retinal research. Establishing reproducible protocols for controlling the degree and type of heterogeneity would enable fairer comparison between federated algorithms and provide a more reliable assessment of their robustness under real clinical conditions.

8.4 Communication-Efficient Federated Frameworks

The improvement of communication capabilities is one of the key future research topics in the domain of federated learning applied to retinal images. In the federated setup, institutions are required to continuously transmit the model parameters or updates to the central server. When the deep learning models are large, this continuous information transfer can become a major obstacle, especially if the participating institutions lack advanced network infrastructure or computing power[5],[13].

Future federated retinal frameworks should therefore investigate methods that reduce the amount and frequency of information exchanged between clients and the server. Potential approaches include model compression, parameter-efficient updates, selective parameter transmission, client selection, adaptive communication intervals, knowledge distillation, and asynchronous federated optimization. These techniques may reduce communication costs while maintaining the diagnostic capability of the global model [13].

Communication efficiency becomes even more important under non-IID retinal data because heterogeneous clients may require different amounts of local computation to obtain effective model updates. Excessive synchronization can increase communication overhead, whereas reducing communication too aggressively may affect convergence or model performance. Future research should therefore jointly consider communication rounds, transmitted parameter volume, local computation, convergence rate, and classification performance.

For multi-disease retinal classification, communication-efficient strategies should additionally preserve performance across individual disease categories and heterogeneous clients. Evaluation should therefore report not only overall accuracy and AUC but also communication cost, number of communication rounds, client participation rate, training time, and disease-wise performance.

Overall, communication-efficient federated learning can help transform retinal FL from experimental multi-institutional studies into scalable and deployment-oriented diagnostic systems. Future frameworks should aim to achieve an appropriate balance between communication efficiency, computational requirements, model convergence, diagnostic performance, and robustness to non-IID clinical data.

8.5 Privacy-Aware and Regulation-Compliant Learning

Privacy protection should remain a central consideration in the future development of federated learning for retinal image classification. Although federated learning avoids direct

ISSN (online): 2581-3048
Volume 10, Issue 9, pp 38-56, September-2026
https://doi.org/10.47001/IRJIET/2026.109005

centralization of raw retinal images, model parameters, gradients, embeddings, and other intermediate information exchanged during training may still contain information that could potentially be exploited through privacy attacks [14], [15]. Federated learning should therefore be considered a privacy-enhancing approach rather than a complete privacy guarantee.

Future retinal federated learning systems should incorporate appropriate privacy-enhancing technologies according to the requirements of participating institutions. Secure aggregation can prevent the coordinating server from directly accessing individual client updates, while differential privacy can provide formal privacy guarantees by controlling the contribution of individual training records [14], [15]. Homomorphic encryption and other cryptographic approaches may provide additional protection, although their computational and communication overhead must be carefully considered.

An important future direction is the systematic evaluation of the privacy-utility trade-off. Stronger privacy protection may introduce additional noise, computational requirements, or communication overhead and can consequently affect model performance. Future studies should therefore report the privacy mechanism, privacy parameters, computational overhead, communication cost, and corresponding change in diagnostic performance.

Regulatory compliance should also be considered throughout the complete lifecycle of a federated retinal system. Healthcare federated learning involves questions concerning data ownership, institutional responsibilities, consent, cross-border data processing, model-update governance, security auditing, and accountability for deployed models. The use of federated learning alone does not automatically ensure compliance with applicable legal and institutional requirements.

For retinal disease classification, future studies should therefore evaluate privacy mechanisms and their effects on diagnostic performance, fairness, communication cost, computational requirements, and robustness against privacy attacks. Privacy evaluation should be integrated into the experimental design rather than treated as an additional component after model development.

Overall, future federated retinal frameworks should aim for a privacy-aware and regulation-conscious design from the beginning of model development. Combining secure aggregation, differential privacy, appropriate cryptographic protection, privacy auditing, and institutional governance can provide a stronger foundation for trustworthy multi-institutional retinal AI while maintaining acceptable diagnostic utility.

8.6 Clinical Validation and Deployment-Oriented Learning

Moving forward, future studies concentrating on federated classification of retinal images should transition from retrospective studies and simulated federated situations to real-life validation and implementation. While it is true that federated learning allows collaboration for model development without collecting data, the majority of the existing ophthalmic studies on federated learning are conducted using retrospective datasets or simulated multi-centre environments.

It is critical to assess federated models not only at one institution but also at many independent institutions with heterogeneous participants and diseases as well as variations in imaging systems and protocols, image quality, and practice. Using external validation – showing the performance of the model on unseen institutions – is even more important as the machine learning model could perform well on training sites, but this should not lead to the conclusion that it is transferable to other institutions. Therefore, any subsequent research should involve independent by means of external validation datasets and prospective studies as much as possible.

Research regarding the implementation of the solution should not rely solely on conventional performance metrics such as classification accuracy and other similar metrics. There are other important aspects including clinical workflow implementation, interoperability, explainability, calibration and fairness evaluation, work in the conditions of confidentiality, computational expenses, and communication efficiency. What is more important is the fact that federated systems should fit the technological infrastructure of healthcare institutions.

Additionally, for multi-class retinal diagnostics, clinical validation should report both disease-wise and institution-wise performance rather than a single performance parameter.

Consequently, future research ought to provide development-oriented federated learning frameworks in which model development, privacy protection, external validation, clinical integration, and post-deployment monitoring will be taken into account. The new research may gradually change from retrospective multi-centre study to prospective clinical validation, and further to real-world application.

In summary, clinical validation, as well as learning about deployment, is the key last stage of transforming the federated

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 10, Issue 9, pp 38-56, September-2026

https://doi.org/10.47001/IRJIET/2026.109005

retinal AI from an experimental form into a successful form of multipurpose diagnostic devices.

IX. CONCLUSION

Federated learning enables the practical approach to jointly analyze retinal images by using federated learning without the necessity to centralize and distribute sensitive patient information. In this systematic review, we explored the literature available in the area of federated learning for classification of retinal diseases and classified the papers analyzed according to the federated learning approaches used, the nature of data distribution, the scope of diseases studied, architectures of models used, and privacy/security techniques.

The comparison indicates that federated learning holds great promise for the joint analysis of ocular and retinal images in distributed health care institutions. Nonetheless, the existing literature is still splintered across imaging methods, types of disease, federated approaches, arrangements of the data, and evaluation procedures. Most studies have examined diseases affecting the retina in isolation while multi-disease and multi-class federated classification remain comparatively rare. Furthermore, the idea of non-IID data often comes from the usage of artificial partitions or controlled distributions obscuring the heterogeneity of the actual clinical environment.

The review further identifies important challenges related to privacy and security, communication overhead, computational constraints, clinical validation, and standardized evaluation. Although federated learning reduces the need for direct exchange of raw retinal images, additional privacy mechanisms may introduce privacy-utility trade-offs. Similarly, repeated communication and differences in computational resources across participating institutions can affect the scalability of federated systems. Limited external and prospective clinical validation further restricts the evidence available for real-world deployment.

Based on these findings, six major research directions were identified: personalized federated learning, federated multi-disease and multi-label learning, realistic modelling of non-IID clinical data, communication-efficient federated frameworks, privacy-aware and regulation-compliant learning, and clinical validation and deployment-oriented learning. These directions emphasize the need to evaluate federated retinal systems not only in terms of overall classification performance but also with respect to disease-wise performance, client-level fairness, privacy protection, communication efficiency, robustness to heterogeneous data, and generalization to unseen clinical institutions.

Overall, the review indicates that the future of federated retinal image classification lies in developing systems that can jointly address multi-disease learning, realistic non-IID data distributions, privacy preservation, communication and computational efficiency, fairness, and clinical reliability. Progress in these areas can help move federated retinal artificial intelligence from experimental proof-of-concept studies toward scalable, trustworthy, and clinically applicable multi-institutional diagnostic systems without requiring centralized sharing of sensitive retinal images.

REFERENCES

[1] 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.

[2] J. De Fauw *et al.*, "Clinically Applicable Deep Learning for Diagnosis and Referral in Retinal Disease," *Nature Medicine*, vol. 24, pp. 1342–1350, 2018.

[3] D. S. W. Ting *et al.*, "Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes," *JAMA*, vol. 318, no. 22, pp. 2211–2223, 2017.

[4] M. J. Sheller *et al.*, "Federated Learning in Medicine: Facilitating Multi-Institutional Collaborations Without Sharing Patient Data," *Scientific Reports*, vol. 10, Article 12598, 2020.

[5] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, vol. 54, pp. 1273–1282, 2017.

[6] J. Lo *et al.*, "Federated Learning for Diabetic Retinopathy Classification and Microvasculature Segmentation Using Optical Coherence Tomography Angiography," *Ophthalmology Scien*ce, 2021, Art. no. 100069. DOI: 10.1016/j.xops.2021.100069.

[7] H. Ran *et al.*, "Federated Learning for Glaucoma Detection With 3D Optical Coherence Tomography," *British Journal of Ophthalmology*, vol. 108, no. 8, 2024. DOI: 10.1136/bjo-2023-324188.

[8] T. Baptista, C. Soares, T. Oliveira, and F. Soares, "Federated Learning for Computer-Aided Diagnosis of Glaucoma Using Retinal Fundus Images," *Applied Sciences*, vol. 13, no. 21, Art. no. 11620, 2023. DOI: 10.3390/app132111620.

[9] A.Nabil *et al.*, "Federated Learning for Multi-Disease Ophthalmic Diagnostics using OCTA," *medRxiv*, 2025. DOI: 10.1101/2025.04.25.25326431.

ISSN (online): 2581-3048
Volume 10, Issue 9, pp 38-56, September-2026
https://doi.org/10.47001/IRJIET/2026.109005

[10] S. Pachade *et al.*, "Retinal Fundus Multi-Disease Image Dataset (RFMiD): A Dataset for Multi-Disease Detection Research," *Data*, vol. 6, no. 2, p. 14, 2021.

[11] T. Li, B. Wang, C. Hu, H. Kang, H. Liu, K. Wang, and H. Fu, "Applications of deep learning in fundus images: A review," *Medical Image Analysis*, vol. 69, Art. no. 101971, 2021.

[12] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated Optimization in Heterogeneous Networks," *Proceedings of Machine Learning and Systems*, vol. 2, 2020.

[13] J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, "Federated Learning: Strategies for Improving Communication Efficiency," *NIPS Workshop on Private Multi-Party Machine Learning*, 2016.

[14] K. Bonawitz *et al.*, "Practical Secure Aggregation for Privacy-Preserving Machine Learning," *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pp. 1175–1191, 2017.

[15] C. Dwork and A. Roth, "The Algorithmic Foundations of Differential Privacy," *Foundations and Trends in Theoretical Computer Science*, vol. 9, nos. 3–4, pp. 211–407, 2014.

[16] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, "Federated Learning with Non-IID Data," *arXiv preprint* arXiv:1806.00582, 2018.

[17] Z. Tang, H.-S. Wong, and Z. Yu, "Privacy-Preserving Federated Learning With Domain Adaptation for Multi-Disease Ocular Disease Recognition," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 6, pp. 3219–3227, 2024. DOI: 10.1109/JBHI.2023.3305685.

[18] A.Hanif *et al.*, "Federated Learning for Multicenter Collaboration in Ophthalmology: Implications for Clinical Diagnosis and Disease Epidemiology," *Ophthalmology Retina*, Vol. 6, no. 8, pp.650-656,2022.:10.1016/j.oret.2022.03.005.

Citation of this Article:

Divya Chunduri, & K. Suresh Babu. (2026). Federated Learning for Multi-Disease Retinal Image Classification: A Systematic Review, Taxonomy, and Future Directions. *International Research Journal of Innovations in Engineering and Technology - IRJIET*, 10(9), 38-56. Article DOI https://doi.org/10.47001/IRJIET/2026.109005

****