DOI:10.14421/ijid.2025.5200 IJID (International Journal on Informatics for Development), *e-ISSN: 2549-7448*
Vol. 14, No. 2, December 2025, Pp. 616-628

Early Detection of Diabetic Retinopathy Through Explainable AI Models: A Systematic Review

Tinashe Ngwazi
Department of Informatics and Analytics
National University of Science and Technology
Bulawayo, Zimbabwe
tinashengwazi@gmail.com

Belinda Ndlovu
Department of Informatics and Analytics
National University of Science and Technology
Bulawayo, Zimbabwe
belinda.ndlovu@nust.ac.zw

Kudakwashe Maguraushe
School of Computing
University of South Africa
Johannesburg, South Africa
magark@unisa.ac.za

Article History
Received May 12th, 2025
Revised August 1st, 2025
Accepted August 4th, 2025
Published December 2025

***Abstract*— Diabetes, if not detected early, can lead to serious complications such as vision loss, known as diabetic retinopathy. Explainable Artificial Intelligence (XAI) can enhance traditional Machine Learning methods, which are not understandable and transparent in diagnostic tasks. This Systematic Literature Review explores data inputs that influence the performance of XAI models in detecting diabetic retinopathy, how XAI techniques can enhance early detection outcomes in diabetic retinopathy, the challenges in implementing these techniques and the ethical implications of using these models in clinical practice. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses approach guided the search in 4 databases, Springer, Science Direct, PubMed and IEEE Xplore. The findings reveal that XAI techniques like Local Interpretable Model-agnostic Explanations (LIME), SHapley Additive exPlanations (SHAP) and Gradient-weighted Class Activation Mapping (GRAD-CAM) offer opportunities like early detection outcomes, integration with existing clinical processes, enhancing trust in AI systems, improving accuracy and personalised treatment. XAI can also facilitate collaboration among clinicians, maintaining fairness in AI systems and supporting adherence to ethical standards. However, research on clinical validation of these models, as well as standardised performance evaluation metrics, is lacking.**

Keywords—diabetes; diagnostic task; ethical implication; hybrid models; interpretability

1 INTRODUCTION

Diabetic retinopathy (DR) is one of the most common complications of diabetes, which causes visual impairment and blindness, especially in the working-age population [1]. It is estimated that once a patient is classified as having diabetic retinopathy, it will progress to the vision-threatening stage in approximately 11% of patients each year, which makes diabetic retinopathy a public health concern [2].

Detecting diabetic retinopathy in its early stages is vital in helping healthcare providers make informed decisions about patient management and effective treatment strategies [3]. The current management of diabetic retinopathy is based on recognising fundus images in earlier stages, using methods like fundus photography [4]. In low-income countries where there is a lack of healthcare resources [5] and a lack of trained caregivers for retinopathy screening and early detection, there is growing evidence to support the use of Artificial Intelligence (AI) in screening the population at risk of sight loss due to diabetic retinopathy complications [6].

Advances have been made in predicting and managing the disease, but most of the ML algorithms fail to give insight beyond the provided data, and they require extensive debugging and deciphering to understand them [7]. There is a substantial positive impact of promoting the reliance of clinicians on AI-based models for early detection of diabetic retinopathy, leading to innovative autonomous systems that can reduce the costs of screening and address the shortage of caregivers in that field [8]. ML is regarded as a concept with great success in engineering and sciences, but data-driven insights have their limits, especially when it comes to data availability [9].

Explainable Artificial Intelligence (XAI) is a collection of ML techniques that are designed to make AI-based systems more interpretable and transparent, allowing the end-users to understand and trust these systems [10]. XAI is capable of boosting interactivity by facilitating collaborative explanations, allowing end-users to engage and manipulate inputs to discover varying patterns [11]. XAI is also capable of bringing about fairness in AI systems by analysing and mitigating biases from training data, and also supports adherence to ethical standards to ensure the responsible use of AI systems [12].

Ref. [13] and [14] identified opportunities in XAI, highlighting that it has the potential to utilise low computational resources, meaning that XAI can be more accessible even to smaller healthcare facilities with low budgets. Ref. [15], [16], [17] highlight the opportunity that XAI creates to make AI systems more understandable, increasing trust and adoption by clinicians. Ref. [18], [19], [20] all use fundus image datasets in their studies, showing that the most commonly used type of data in the early detection of diabetic retinopathy is retinal image data. Ref. [21] identified the challenges of implementing XAI in clinical settings, stating the scarcity of high-quality data needed for training models as well as the difficulty in complying with regulatory and ethical guidelines, which mostly complicate the implementation of these AI-based systems.

Despite all the research on XAI in predictive models, existing studies tend to focus more on general ophthalmologic applications. This manuscript provides a focused review of how XAI has been applied in the early diagnosis of diabetic retinopathy, adding novelty by including the types of data inputs and analysing the ethical implications of XAI, to highlight the effects of applying this technology, whether positive or negative. Our research questions are:

1. What data inputs significantly influence the performance of XAI models in detecting diabetic retinopathy?
2. What innovative XAI techniques are emerging as pivotal tools in the early diagnosis of diabetic retinopathy?
3. How might XAI techniques create new pathways for improving early detection outcomes in diabetic Retinopathy?
4. What challenges need to be addressed when incorporating XAI techniques into the early detection processes for diabetic retinopathy?
5. What are the ethical implications of using XAI models for detecting diabetic retinopathy in clinical practice?

2 METHOD

A Systematic Literature Review was used to answer the posed questions, with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) approach applied in the identification, screening, and determining the eligibility of studies. This review uses quantitative analysis to understand the findings and collected data.

Science Direct, Springer, PubMed and IEEE Xplore were used as databases for searching relevant literature. Various specific keywords and synonyms were used to search the databases. The search string used Boolean connectors, which were layered as follows: ("early detection" OR "early diagnosis") AND ("diabetes retinopathy" OR "diabetic retinopathy") AND ("explainable AI" OR "interpretable AI" OR "explainable artificial intelligence" OR "interpretable machine learning". To enhance the reproducibility of our search results, the number of results associated with key terms in the search string was recorded as follows: ("early detection" OR "early diagnosis") = 178, ("diabetes retinopathy" OR "diabetic retinopathy") = 202, ("explainable AI" OR "interpretable AI" OR "explainable artificial intelligence" OR "interpretable machine learning") = 299. Some of the combinations were overlapping, which means individual block counts did not sum up to 202 studies. In this case, the Boolean logic ensured that only the studies that appeared in all three blocks were included in the final query. The search was executed on 3 February 2025.

Due to the large number of articles found in online databases, the search string returned thousands of papers, a large number of which were irrelevant. In this case, the focus was on studies directly related to the use of XAI in diabetic retinopathy detection; studies strictly written in the English language and published from 2020 to 2025. The choice to focus on the studies in the specified period was based on the notion that recent studies reflect current challenges and

technologies and provide more context that may be useful for future research. Studies were excluded if they were non-peer-reviewed, non-original research, like commentaries and reviews, less detailed (lack of description for the ML, AI tool applied or performance metrics), and studies applying XAI outside the context of diabetes retinopathy or early detection. To assess the methodological quality and risk of bias of the included studies, we applied the QUADAS-2 tool, which is a validated instrument for evaluating diagnostic accuracy studies. **BN** evaluated *Patient Selection*, focusing on whether the studies used a representative sample of patients, specifically, individuals whose retinal images were included for diabetic retinopathy detection and whether any inappropriate exclusions were made (e.g., excluding mild cases or poor-quality images without justification). **KM** assessed the *Reference Standard*, examining whether the expert annotation of retinal images was valid and consistently applied; studies involving multiple expert graders or clinically verified labels were rated as low risk. **TN** reviewed the index test and the flow and timing domains, determining whether the ML/AI technique was applied consistently across all subjects. Studies that followed a consistent protocol scored low bias risk. The limitations in using this method are that extracting data from the selected studies can be prone to human error, and also variations in the reporting style of the papers may make it difficult to maintain consistency in comparing results. To overcome these limitations, the authors applied double data extraction, whereby two reviewers, **TN** and **BN,** extracted data separately and resolved any discrepancies through discussion with **KM**. All authors also set clear inclusion and exclusion criteria to focus on studies that meet a specific reporting standard (PRISMA in this case), to address the issue of variations in reporting style.

A total of 202 were identified through four electronic databases: Springer =146, IEEE Xplore = 8, PubMed = 6 and Science Direct = 42.

After removing 4 duplicate records, 198 records remained for title and abstract screening. Of these, 125 records were excluded because their titles did not specifically align with the research topic (unrelated to diabetic retinopathy or ML/XAI).

A total of 43 full-text articles were assessed, and 13 studies were excluded for lacking sufficient detail about the machine learning techniques or XAI methods applied. A further 13 studies were removed as they focused on treatment strategies and general ophthalmology, with no application of algorithms and XAI techniques. 3 Studies were excluded for not providing algorithm accuracy. 1 study was not written in the English language, leading to exclusion. Following the overall eligibility assessment, 13 studies met all inclusion criteria and were included in the final quantitative synthesis and methodological quality assessment.

3 RESULT AND DISCUSSION

The following PRISMA diagram in Fig. 1 shows the identification and screening of records from the searched databases. Figure 1 shows a Prisma flow diagram to illustrate the process of identifying and screening records from the selected databases. The process led to the selection of 13 studies that satisfied the inclusion criteria. These studies are given in Table 1.

3.1 Number of Publications per Continent

Figure 2 shows a bar graph that represents the studies published for each continent. From the bar chart in Fig. 2, Asia (6 publications) has a high prevalence of studies and research when it comes to diabetic retinopathy, which may be attributed to the large population of people with diabetes mellitus in India [22].

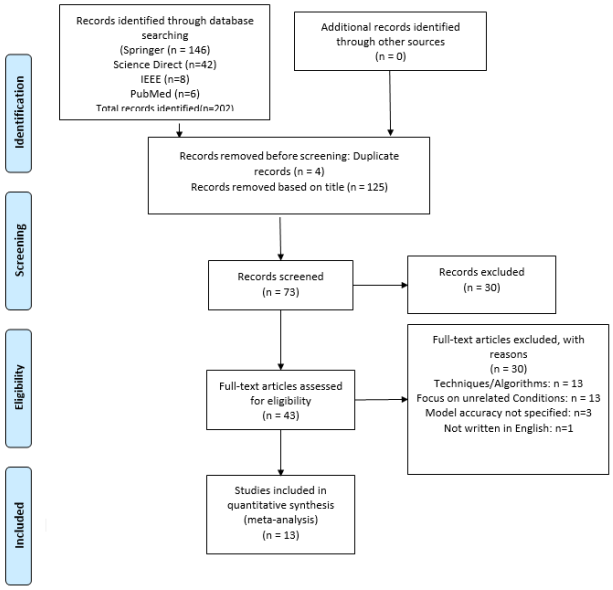

Figure 1. Prisma flow diagram

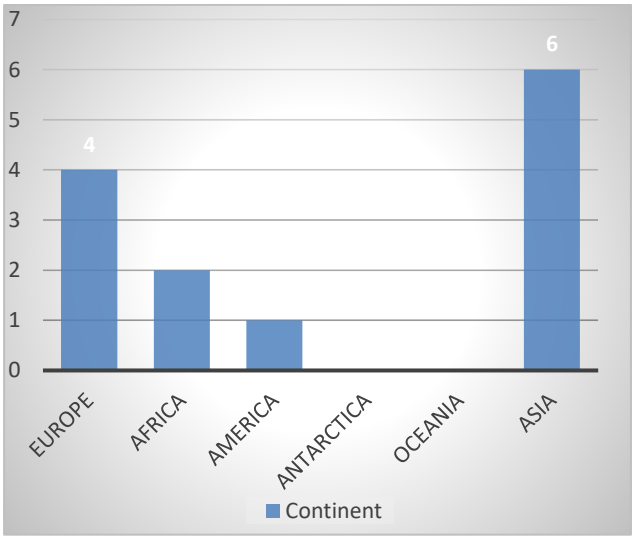

Figure 2. Number of publications per continent

Table 1. Papers that Met the Inclusion Criteria

Author	Dataset Used	Origin	Research aim	Opportunities	Challenges	ML and XAI Algorithms Used	Accuracy	Data Inputs	Ethical Implications
[23] (Khan et al., 2021)	Kaggle Diabetic Retinopathy Detection	Pakistan	Diabetic Retinopathy Detection Using VGG-NIN, a Deep Learning Architecture	• Increased trust • Enhanced performance • Low computing resources.	• Small dataset	• Visual Geometry Group of the University of Oxford (Vgg16) VGG-NiN model	• 85%	• Labelling information • Colored image data	• Bias and fairness • Transparency
[24] (Taifa et al., 2024)	EyePACS	Bangladesh	A hybrid approach for enhancing diabetic retinopathy	• Personalised treatment • Improved accuracy	• Limited dataset size • Difficult to interpret • Image quality variations	• Decision trees • Random forests • Support Vector Machines	• 95.50%	• Labelling information • Colored image data	• Clinical responsibility • Data privacy • Bias and Fairness • Transparency and interpretability
[25] (Shahzad et al., 2024)	Kaggle Diabetic Retinopathy Detection dataset	South Africa, Pakistan	Developing a transparent diagnosis model for Diabetic Retinopathy Using Explainable AI	• Integration with clinical systems. • Efficient patient management. • Enhanced accuracy. • Interpretable systems • Leverage less powerful computational resources.	• Dataset limitation • Requires powerful computational resources • Lack of robust preprocessing layers • Limited use of Explainable AI	• CNN • LIME (Local Interpretable Model-agnostic Explanations	• 95%	• Retinal images • Demographic data (age, gender, BMI)	• Bias and fairness • Transparency
[26] (Li et al 2022)	Kaggle Diabetic Retinopathy Detection dataset	Malaysia	The adoption of deep learning interpretability techniques on diabetic retinopathy analysis	• Integration with clinical practices. • Increased trust • Enhanced accuracy	• Need for validation	• CAM (Class Activation Mapping) • CNN	• 98.34%	• Retinal images	• Clinical responsibility
[19] (Ikran et al 2025)	Kaggle Diabetic Retinopathy Detection dataset	Pakistan	ResViT FusionNet Model: An explainable AI-driven approach for automated grading of diabetic retinopathy in retinal images	• Integration with clinical processes • Improved accuracy • Enhanced trust • Enhanced interpretability.	• Interpretability limitations • Complexity • Validation of explanations • Variability in image quality • • Imbalanced datasets, • Need for interpretable models •	• Grad-CAM (Gradient-weighted Class Activation Mapping) • LIME (Local Interpretable Model-agnostic Explanations) •	• 93.01%	• Retinal images	• Clinical responsibility and patient safety • Bias in data training • Transparency
[21] Bidwai et al., 2021)	Kaggle Diabetic Retinopathy Detection	India	Multimodal image fusion for the detection of diabetic retinopathy using	• Regulatory compliance • Enhanced trust • Improved decision making	• Bias and fairness • Computational resource requirements	• Multimodal image fusion • SHapley Additive Explanations (SHAP) •	• 96.47%	• Image data	• Clinical responsibility and patient safety

			optimised explainable AI-based Light GBM classifier		• Data quality and availability • Complexity				• Bias in data training • Transparency Trust
[27] (Parmar et al., 2024)	Kaggle Diabetic Retinopathy Detection	India, USA, Italy, Nigeria.	Artificial Intelligence (AI) for Early Diagnosis of Retinal Diseases	• Improved accuracy • Personalised treatment • Integration into clinical practice	• Lack of interpretability.	• CNN	• 99.37%	• Clinical data (patient demographics) • Retinal images	• Bias in data representation Clinical • Responsibility and accountability
[28] . (Romero-Oraá et al., 2024)	Kaggle Diabetic Retinopathy dataset	Spain	Attention-based deep learning framework for automatic fundus image processing to aid in diabetic retinopathy grading	• High accuracy • High performance	• Dataset limitations • Lack of transparency.	• CNN	• 77.2 %	• Retinal images	• Transparency • Data quality • Bias and fairness
[29] (Yagin et al., 2024)	Patient Cohort	Norway	Hybrid Explainable Artificial Intelligence Models for Targeted Metabolomics Analysis of Diabetic Retinopathy	• Improved decision making • Integration into clinical practice • High accuracy • Personalised treatment	• Computational resource requirements • Data availability • Complexity	• SHapley Additive exPlanations (SHAP)	• 89.58%	• Clinical parameters (glucose levels) • Metabolomic data (glucose, amino acids) • Demographic data (age) •	• Clinical responsibility and patient safety • Bias in data • Transparency
[30] (Yagin et al 2023)	Open Access Data on T2D Patients	Norway	Explainable Artificial Intelligence Paves the Way in Precision Diagnostics and Biomarker Discovery	• Enhanced trust • Increased accuracy	• Need for validation • Integration with clinical processes • Complexity.	• Explainable Boosting Machine (EBM)	• 89.33	• Metabolomic data (metabolites associated with DR), • Demographic data (age, gender, BMI)	• Clinical responsibility • Bias in data interpretability • Transparency
[16] (Ahnaf et al., 2024)	Aptos 2019 Blindness Detection Dataset	Bangladesh	Enhancing Early Detection of Diabetic Retinopathy Through the Integration of Deep Learning Models and Explainable AI.	• Improved trust • Enhanced accuracy • Integration with clinical settings.	• Not transparent • Image quality variation	• CNN •	• 95.27%	• Retinal images	• Transparency • Clinical accountability • Bias in data representation
[31] (Wang et al., 2024)	Diabetes Complication Early Warning Dataset	China	Prediction and analysis of risk factors for diabetic retinopathy based on machine learning and interpretable models.	• Basis for future research • Integration with clinical operations. • Enhanced trust	• Data availability	• SHAP Framework (SHapley Additive exPlanations)	• 82.5 %	• Retinal images	• Clinical accountability • Bias and fairness
[32] (Quellec et al., 2021)	OPHDIAT	France	Explanatory artificial intelligence for diabetic retinopathy diagnosis	• Improved decision making • Enhanced trust	• Balancing explainability and performance • Complexity	• ExplAIn, • Generalised Occlusion Method •	• 99.78% for severe DR • 99.39% (moderate DR)	• Labelled data • Image data	• Bias • Clinical accountability

Europe has the second most publications (4 publications) on diabetic retinopathy. Africa has 2 publications, followed by 1 study from America. The least number of publications is in Antarctica and Oceania, with both continents recording 0 publications related to diabetic retinopathy. Despite the shortage of eye-care professionals and healthcare resources in Africa, Africa has 2 publications, revealing a research gap in the region. Region-specific research is required to ensure fair representation across diverse populations and trigger initiatives to support low-resource settings.

3.2 Algorithms Adopted

Figure 3 shows the count of algorithms that were leveraged in each publication. Five (5) studies used CNN as the traditional/baseline algorithm of choice for their study. SHAP was leveraged by three (3) studies, Gradient-weighted Class Activation Map (Grad-CAM) was utilised by 3 studies, LIME (2 studies), ExplAIn and EBM (1 study). Leveraging these XAI methods with powerful algorithms like CNN could boost the results even further and lead to effective hybrid models. SHAP and Grad-CAM are the most utilised XAI techniques to improve the explainability of the traditional algorithms, with both of these techniques amounting to a total of 6 studies out of the 13 studies. The consistent appearance of CNN architectures enhanced with XAI across the high-performing studies suggests that hybrid approaches are becoming the new standard in clinical decision support systems.

3.3 Data Inputs per Study

Table 2 shows the data inputs that were used to predict diabetic retinopathy in each study. Retinal images (11 studies) are the most common data input in the diagnosis of DR, showing their dominance in the early detection of diabetic retinopathy. The wide use of retinal images as data inputs is due to the visual nature of the condition and the reliance on image-based deep learning techniques like CNN. Clinical data is also useful, as it is utilised in 4 studies. Labelled data (3 studies) is also important as it shows the classes present in the datasets used for training. Metabolomic data were leveraged in 2 of the reviewed studies. The integration of metabolomic, clinical and labelled data with retinal images indicates that multimodal fusion of data could enhance diagnosis and uncover additional information that images alone may otherwise miss.

3.4 Opportunities per Publication

Table 3 illustrates the number of opportunities that were recorded as they appeared in each reviewed study. From the table, integration into clinical practice is an opportunity that was mentioned in 7 of the reviewed studies.

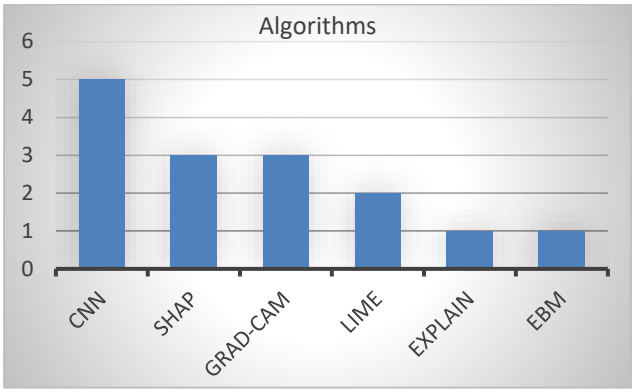

Figure 3. Algorithms used per study

Table 2. Data Inputs per Study

Data input	Occurrence/studies
Retinal images	11
Clinical data	4
Labeled data	3
Metabolomic data	2

One study mentioned the ability of XAI methods to leverage low-resource computation tools as having good potential to enhance diabetic retinopathy diagnosis. Eight of the reviewed studies highlighted that XAI can enhance trust in AI systems. Out of the reviewed papers, 3 studies agree that XAI personalised treatment is another opportunity emerging from applying XAI methods. 9 studies state that XAI can enhance the accuracy of traditional ML systems. These findings indicate the practical implications and goals of leveraging XAI in healthcare, which builds confidence among clinicians and ensures real-world applicability. This evidence suggests that XAI not only benefits model interpretability but also improves model performance. This is a counter-fact to the misconception that interpretability compromises accuracy.

3.5 Challenges per Study

Figure 4 illustrates the challenges that were faced in the reviewed studies. This research identified four factors that affect the adoption of ML in the early diagnosis of diabetic retinopathy.

Table 3. Opportunities per Publication

Opportunity	Occurrence (studies)
Integration into clinical practice	7
Leveraging resources with lower computational power.	1
Enhanced trust	8
Personalized treatment	3
Enhanced Accuracy	9

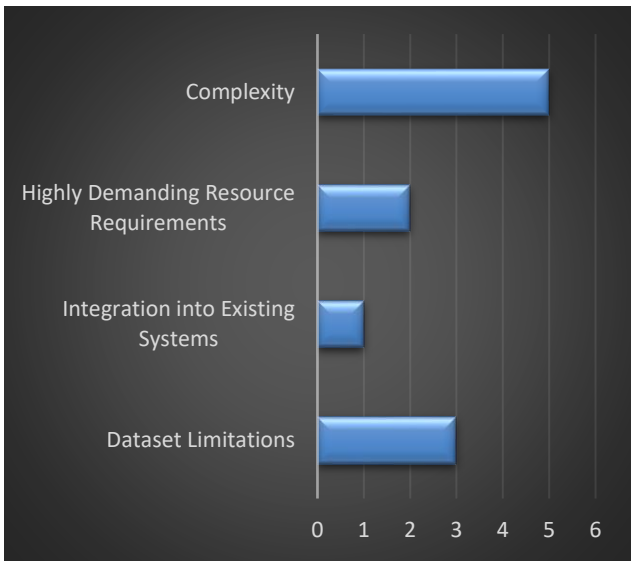

Figure 4. Challenges per study

Five (5) studies highlight that the complexity of XAI methods poses a challenge for leveraging these tools in diabetic retinopathy diagnosis. 2 studies state that XAI techniques have demanding resource requirements. 1 out of the reviewed studies mentions that XAI techniques may be difficult to integrate with existing systems. The challenge of dataset limitations in data training is mentioned in 3 studies in the review.

3.6 Ethical Implications of XAI

Table 4 shows the ethical implications that were flagged in each reviewed study. Nine (9) studies highlight that transparency is an ethical implication when it comes to using XAI systems in diagnosing diabetic retinopathy. Twelve (12) studies suggest that the use of XAI techniques raises the issue of bias and fairness, especially in training datasets. Clinical responsibility is another ethical issue to consider when dealing with XAI tools, as 10 studies mention that fact. The reviewed studies did not discuss legal accountability or regulatory approval issues, which are crucial for real-world implementation. This suggests a need for further research involving experts in the legal sector, clinicians, and ethicists to anticipate involved risks.

This section gives a discussion that answers the research questions posed in this study.

3.7 Emerging Trend

From the 13 reviewed studies, several emerging trends have been observed. CNNs (Convolutional Neural Networks) were the widely leveraged baseline models, mainly combined with explanatory tools like GRAD-CAM or LIME.

Table 4. Ethical Implications per Publication

Implication	Occurrence (studies)
Transparency	9
Bias and Fairness	12
Clinical Responsibility	10

Across several datasets, these hybrid combinations resulted in high classification performance. Ref. [14] reported 95% accuracy by combining CNN and LIME, and [19] reported 93.01% leveraging CNN, Grad-CAM and LIME. Despite this, performance varied depending on the objectives, for example, the study by [26] achieved 98.34% accuracy, Class Activation Mapping (CAM) with CNN using retinal images as data input, without providing sensitivity and or specificity. Ref. [27] used achieved 99.37% accuracy using both clinical and retinal image data.

In contrast to this, the Explainable Boosting Machine (EBM) was used in the study by and SHAP-based models were used by [33], [29], were effective with structured data such as metabolomic and demographic features, providing a more transparent view into feature importance but with modest accuracy values (89%). These observations prove that hybrid models (combining CNN-based architectures with XAI tools (LIME, SHAP) offer a balance between performance and explainability, but it is important to note that no single XAI tool fits all clinical solutions.

3.8 What Data Inputs Significantly Influence the Performance of Explainable Artificial Intelligence (XAI) Models in Detecting Diabetic Retinopathy?

3.8.1 *Retinal images:* Retinal images refer to photographic visualisations of the eye's surface, usually captured using special tools like fundus cameras [34]. The study by [19] used fundus retinal images from the APTOS 2019 dataset for data training. The research by [21] and [27] also uses retinal images, which are fundus images (for visualising retinal structures), and Optical Coherence Tomography Angiography (OCTA) images, which provide insights into blood flow in the retina. The study by [32] used two datasets from OPHDIAT and EyePACS to obtain images for learning (model training). These images are useful in classification training that leads to the prediction of diabetic retinopathy. XAI techniques like Grad-CAM use these retinal images to create heatmaps that showcase areas that the model focuses on for making decisions [35].

3.8.2 *Clinical data:* The study by [36] uses clinical data as data inputs for training in the early diagnosis of diabetic retinopathy among type 2 diabetic patients, which comprises clinical indicators like duration of diabetes, glycated haemoglobin (HbA1c) and systolic blood pressure (SBP). This is further supported by [29] where they utilised biochemical data, including glycine and creatinine, to better understand their role in diabetic retinopathy progression. Clinical data is useful in providing more context in data training, other than just using retinal images. Involving more features may lead to improved performance of XAI tools when it comes to decision-making, because these XAI tools rely on large datasets to make accurate decisions [37]. The research by [30] utilised demographic data that

included age and BMI, as age is a factor in the likelihood of being affected by diabetic retinopathy.

3.8.3 *Labelled data:* The research carried out by [32] used a labelled image dataset which is categorised into No_DR, Mild, Moderate, Severe and Proliferative, based on the International Clinical Diabetic Retinopathy (ICDR) severity scale. Other authors ike [25], [29] also used labelled images to help train their models. Alternatively, the study by [23] used a set of labelled images (35,126) and unlabeled images (53,576), which could be useful to determine how the model generalises unlabeled data in the real world, making XAI systems more accurate.

3.8.4 *Metabolomic data:* Metabolomic data is the quantitative measurement of metabolites found in biological samples that can provide useful insights into the biochemical state of cells within an organism [38]. The research by [29] used metabolomics data, which included serum samples from patients and 122 metabolites were selected for the model across various diabetic retinopathy patient groups. The research by [36] supports the use of metabolomic data, as serum samples were analysed to get 532 metabolites, which were further preprocessed to help identify diabetic retinopathy in type 2 diabetes patients. This may be useful in tracking the progression of diabetic retinopathy, as changes in metabolites may be correlated with the progression of the disease, which could aid in efficiently monitoring it.

3.9 What Innovative Explainable Artificial Intelligence Techniques are Emerging as Pivotal Tools in the Early Diagnosis of Diabetic Retinopathy?

3.9.1 *Local Interpretable Model-agnostic Explanations (*LIME is a technique used in ML to explain complex models transparently [39]. LIME is used in providing explanations for varying instances and also identifying critical features of the model [40]. This leads to increased transparency. The study by [19] used the LIME technique to develop a Res-ViT model that is capable of automated DR grading in retinal images, which saw the accuracy rocket to 93.01%. While LIME's simplicity facilitates clinician understanding, it does not provide insights into the model's behaviour across the entire dataset due to a lack of global explanation [41], which limits its utility for a comprehensive understanding.*LIME):*

3.9.2 *SHapley Additive exPlanations (SHAP):* SHAP is a technique used when applying interpretability to ML models by calculating how much each feature contributes to the final predictions of the model using SHapley values [42]. SHAP is useful when it comes to explaining model predictions, analysing feature importance, and providing global and local predictions [39]. Ref. [21] leveraged SHAP in a multimodal image fusion approach for diabetic retinopathy detection, showing a good accuracy of 94.32% even when it was trained with 90% of the data. The SHAP technique is limited in the sense that it has a high computational cost due to the Kernel SHAP approach, which has a quadratic time complexity in relation to both the dimensionality and size of the dataset, which can be prohibitive for larger datasets [43].

3.9.3 *Explainable Boosting Machine (EBM):* EBM is a modelling method that is leveraged to enhance the interpretability of ML models while maintaining a good predictive performance measure [44] achieving an accuracy of 87%. Ref. [30] applied this technique to improve diagnostic precision and incorporate biomarker discovery in diabetic retinopathy diagnosis. These papers show that the ability of EBM to represent the model as a sum of learned functions for each predictor makes the contributions of each variable transparently visible. Despite that, EBM is strong on tabular data instead of unstructured data, making it less suitable for image classification [45].

3.9.4 *Gradient-weighted Class Activation Map (Grad-CAM):* Grad-CAM is an interpretability technique that is used to outline and visualise the decision-making process of Convolutional Neural Networks (CNNs) in image classification processes [46]. Grad-CAMs are useful in generating heat maps, highlighting important features and improving model interpretability [35] While being so advantageous, Grad-CAM struggles to localise multiple occurrences of the same class within a single image, which can limit its effectiveness in complex situations like less clear images [47].

3.10 How might Explainable Artificial Intelligence Techniques Create New Pathways for Improving Early Detection Outcomes in Diabetic Retinopathy?

3.10.1 *Personalised treatment:* Personalised treatment refers to treatment that is tailored to individuals based on their characteristics, including genes, lifestyle, and environment, to develop an effective treatment strategy/plan for them [48]. Healthcare providers can leverage XAI and ML to streamline treatment regimens [49]. The research by [29] highlights that XAI supports personalised treatment strategies, whereby the collaboration between clinicians and AI systems enables tailored treatment plans that will ultimately improve patient outcomes. This was further affirmed by [24], where personalised treatment is stated as a strength in managing diabetic retinopathy. The study [34] utilised biomarkers for early detection of diabetic retinopathy, allowing continuous monitoring of the condition and tailored treatment. This is crucial in the efficient management of diabetic retinopathy,

taking into account varying patient profiles, leading to efficient treatment.

3.10.2 *Integration into clinical practice:* Integrating XAI with existing clinical processes leads to more positive health outcomes for patients, making it suitable for real-world application [50]. The studies [16], [19], [51], [52] agree that XAI is crucial in improving clinical decision-making and providing reliable systems with flexible usability. Integration can include systems like IoT systems [53], Smart Technologies [54], Quantified Self Technologies [55], [56], [48] which could help in continuous patient monitoring and automated predictions.

3.10.3 *Regulatory Compliance:* Both [57] and [58] discuss that XAI is important when it comes to compliance with regulatory standards such as GDPR, which requires clear explanations when it comes to decisions that affect patient outcomes. By adhering to these standards, the role of health practitioners and XAI is made clear by defining specific roles and responsibilities, leading to accountability and patient safety during diagnostic processes [59].

3.10.4 *Increased Trust:* Making AI-based systems much more interpretable allows clinicians to gain trust in these systems, which will drive the acceptance of AI into the medical field [26]. The studies [13], [19], [21], [26], [30] affirm that as trust increases, it will drive the growth of AI-based systems in the healthcare industry by allowing clinicians and patients to understand how AI systems work, which improves collaboration, ultimately leading to improved patient outcomes.

3.10.5 *Increased Accuracy:* The study by [18] utilized the ExplAIn framework and achieved high accuracy through end-to-end training and using a generalised occlusion method to reduce fake positives, leading to a focus on relevant lesions, thereby boosting accuracy. Similarly, the study by [36] highlights increased accuracy and effective robustness against overfitting and the combination of clinical and metabolomics data.

3.11 What Challenges need to be Addressed when Incorporating Explainable Artificial Intelligence Techniques into the Early Detection Processes for Diabetic Retinopathy?

3.11.1 *Integration into existing systems:* The research by [58] raises the issue of human interpretability, stating that these XAI systems should be able to be understood even by non-expert users, which may be a problem given their complexity. The study [30] highlights that it is difficult to integrate XAI into existing systems because of the complexity of medical data and the fact that clinicians may need further training to be able to effectively use these systems.

3.11.2 *Complexity:* The complexity of XAI techniques is worth noting as a concerning challenge. A study by [32] demonstrates that XAI techniques deal with high-dimensional data, which usually complicates feature selection. Also, as much as these techniques seek to improve model interpretability, it may be difficult to understand how they make decisions as well. The study by [29] further affirms this by acknowledging the complexity of XAI techniques, particularly when it comes to implementing them into clinical practice due to issues like resource requirements.

3.11.3 *Dataset Limitations and Quality:* When it comes to datasets, [51] states that their study utilised a limited dataset size, which affected the efficiency of the system when it came to findings and the interpretations extracted from the XAI model. The studies [14], [19], [23], [28], [60] further support that the limitation of datasets when it comes to size is a hindrance. Other critical dataset challenges are evident. The study by [61] highlights dataset labelling inconsistencies in expert annotations, which may introduce noise in model training. Differences in imaging equipment, patient demographics, and clinical settings bring the issue of domain shift across datasets, which may reduce a model's performance when subjected to external data [62].

3.11.4 *High computational power resource requirement:* The study done by [57] highlights that there is a huge trade-off between interpretability and performance, meaning that achieving a higher accuracy score while maintaining a high level of transparency may need system resources and hardware that can handle the high resource demand of this complex task. Resource constraints pose a notable challenge when it comes to XAI, as using it may require more resources than the healthcare facility may be able to meet. [26] highlights that low computational power resources may result in misleading interpretations due to inefficiency.

While related, the relationship between XAI complexity and demanding resource requirements is distinct in the sense that complexity refers to the intricacy of the XAI techniques in detecting diabetic retinopathy, while demanding resource requirements relates to the computational power required to implement and execute the models [63]. Despite the distinction, these two features are interconnected as the complexity of the models increases, so do the resource requirements. Addressing these challenges leads to a balance between model performance and efficient use of resources [64].

3.12 What are the Ethical Implications of Using Explainable Artificial Intelligence Models for Detecting Diabetic Retinopathy in Clinical Practice?

3.12.1 Transparency: There is potential for bias when it comes to training data, usually due to unrepresentative datasets, which may affect the predictive accuracy of the XAI methods across different populations [30]. The research conducted by [29] outlines the use of SHAP as a method that provides useful insight into the model's predictions in diabetic retinopathy, improving transparency. This transparency is important for clinicians to understand the recommendations made by XAI tools, enhancing trust and widespread adoption by practitioners. The study by [16] further supports this point by stating that the transparency enabled by XAI methods means that the systems are more likely to be accepted by clinicians. This is so because they can explain and justify the AI's actions to the patients, meaning that engaging and educating populations about diabetic retinopathy becomes easier and effective, leading to better management of the disease.

In clinical settings, there is a trade-off between explainability and accuracy, as algorithms may be highly transparent but sacrifice accuracy, or it could be the other way around [65]. This can be frustrating for doctors and patients who want to know the reasoning behind a diagnosis. Explainable AI (XAI) tries to make these models more understandable, but making them easier to explain can sometimes mean giving up a bit of accuracy. Finding the right balance between performance and clarity is essential to using AI responsibly and effectively in healthcare [66].

3.12.2 Bias and Fairness: The research carried out by [25] suggests that XAI is capable of addressing biases that may exist within AI systems by clarifying the decision-making processes of those systems. This gives room for criticising and adjusting the models to ensure fairness among all diabetic retinopathy patients. The study by [15] also points out that XAI is essential in providing fair treatment and reducing discrimination in decisions related to healthcare, in this case, diagnosing diabetic retinopathy. Reducing biases and improving fairness leads to the timely diagnosis of diabetic retinopathy because factors such as race and gender are eliminated, focusing all the resources on diagnosis and treatment.

3.12.3 Regulatory Compliance: There are substantial regulatory challenges to incorporating XAI models into clinical applications for diagnosing diabetic retinopathy. In the context of healthcare AI, almost all AI products are categorised as medical devices and are subject to rigorous approval processes by agencies like the FDA in the United States, the EMA in Europe, or national health authorities [67]. These involve validation processes, documentation of clinical safety, risk management and post-market surveillance, which many AI models (in particular those developed quickly or trained using non-standardised data) are likely to find challenging. Furthermore, despite offering some level of interpretability, current laws may not supply explicit thresholds on how much transparency is required for clinical approval [68]. There is uncertainty as to who would be legally at fault: a developer, a health care provider, or an institution, if something goes wrong and an AI system makes an incorrect prediction. Beyond that, models need to adhere to data protection laws, such as GDPR in the EU or HIPAA in the U.S., that require strict control over how patient data is collected, processed, and integrated with other data. But laws were not drafted with AI in mind, and this legal terrain is complicated to navigate.

3.12.4 Clinical responsibility: The study by [36] highlights the need for accountability when it comes to AI-based diagnostic processes, posing questions about who is responsible for the mistakes made by the AI model during the diagnosis of diabetic retinopathy. This may be useful in clarifying the role of XAI, the medical institution or the practitioners in the case of misdiagnosis, including lawsuits that may arise between the patient and the mentioned parties. The research by [29] further affirms this point by stating that clinicians need to be in control of these AI tools, making sure that they do not replace human judgment, but rather support it. While AI can enhance the accuracy of diagnostic models, the responsibility for patient care remains in the hands of healthcare professionals [16]. A successful data breach could lead to theft or loss of important healthcare information, which could be used for wrong purposes, like identity theft [69]. There is also a need for informed consent from the healthcare sector when dealing with sensitive patient data, making sure that patients understand that AI technology is involved in their diagnostic process, and this helps them make informed decisions and improve their trust in healthcare systems [70].

3.13 Implications

The findings in this study indicate that the introduction of XAI, in combination with ML models, can significantly improve the early diagnosis of diabetic retinopathy. This ensures flexibility and interaction between patients and caregivers, which could lead to improved healthcare. Predicting health outcomes, as in the case of early detection of diabetic retinopathy, leads to more informed decisions both for the caregiver and the patient on how to handle the disease progression and efficient treatment. This is especially crucial in involving patients in their treatment process, leading to a patient-centred approach. This study is also important as it

contributes to the rapidly growing literature and outlines the work of other authors, as well as the impact of their studies. This presents enormous opportunities, such as the use of AI with wearable devices like smart wristbands, which could enable diagnosis to be done on the go, regardless of location.

Practical implementation

The study focuses on the potential that XAI presents, especially in the healthcare system, to assist practitioners in enhancing diagnostic accuracy and patient outcomes when it comes to diabetic retinopathy. This introduces technologically advanced initiatives for healthcare professionals while allowing policymakers and regulators to implement effective policies while keeping up with rapid technological advancements. The findings in this study highlight the importance of efficient training and not neglecting data privacy in the process. Furthermore, most studies were conducted in countries with significantly stronger economies, indicating that more research and alternative perspectives from developing countries are needed to gather diverse facts that could lead to even better innovations. Based on this study, the author highlights the strong potential for expanding the use of AI-based diabetic retinopathy tools beyond clinical settings. As interesting as the idea of self-quantification and early diabetic retinopathy monitoring is, it is essential to note the clinical, regulatory, and ethical considerations involved (e.g., Potential misuse of the technology by patients). The diagnosis of diabetic retinopathy requires high-quality fundus imagery and expert intervention, which can be challenging to access, may require substantial financial investments, and may be difficult to fully integrate into consumer-grade applications for use by the general population. Therefore, future works must be guided by clinical validation, expert oversight, and clear standards to ensure safety and ethical compliance.

3.14 Research Gaps

From the bar graph shown in Figure 2, only 2 studies were carried out in Africa (South Africa and Nigeria). This indicates that the topic of diabetic retinopathy has been under-researched, which may be due to limitations in healthcare resources and finances [71]. There is a shortage of studies specifically applying XAI techniques in diabetic retinopathy diagnosis, even though XAI has the potential to improve understanding of the disease and improve patient outcomes [32].

4 CONCLUSION

This study focuses on using XAI to diagnose early stages of DR and has several novel contributions to the current body of literature. We observed several interesting gaps, especially in the area of integrating XAI with classical machine learning (ML) techniques. Though CNNs have achieved promising performance on conventional diagnostics, they remain opaque and uninterpretable methods. We stress that the combination of XAI methods with CNNs may help to circumvent these limitations by efficiently explaining model decision processes. This hybrid not only increases clinician trust but also enables more realistic deployment in clinical practice. In addition, the study categorised different XAI techniques and particular challenges regarding their application in practice. To enhance application into patient care, we suggest that further studies need to investigate the synergy of these predictive models with new technology, such as wearable sensors, with consideration of ethical practice. This SLR adds to the knowledge about the potential of XAI to innovate the detection of diabetic retinopathy and provides directions for future work. This paper specifically explores how XAI methods enhance model trust, clinical interpretability, and ethical operationalisation. In this respect, it provides a structured view of how certain XAI techniques serve specific clinical applications. This addresses a key gap in the literature, as the majority of non-traditional studies focus on technical performance and often fail to report on real-world usability and clinician-facing functionality. Furthermore, our study brings to light the underrepresentation of low-resource regions, particularly Africa, in XAI-related diabetic retinopathy research. Despite a significant diabetes burden in these areas, there is limited deployment and testing of interpretable models in such contexts. Our analysis encourages the development of lightweight, explainable models optimised for low-infrastructure settings and emphasises the need for broader geographic inclusion in validation efforts. Importantly, while previous reviews typically report model accuracy (frequently exceeding 90%), they neglect to discuss other critical clinical diagnostic performance metrics such as **s**ensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV**)**. These measures are essential for evaluating how closely a model mimics clinical decision-making. This review highlights that omission and strongly recommends these metrics be standardised across future studies to better reflect real-world diagnostic utility.

CREDIT AUTHOR STATEMENT

This review, "Early Detection of Diabetic Retinopathy Through Explainable AI Models: A Systematic Review", was made possible thanks to the contributions of three authors: Tinashe Ngwazi (TN), Belinda Ndlovu (BN), and Kudakwashe Maguraushe (KM). Conceptualisation: TN, BN, KM; Initial draft writing: TN; Methodology: TN, BN, KM; Software: TN; Validation: BN, KM; Writing, Review, Editing: BN, KM; Supervision: BN, KM.

COMPETING INTERESTS

The authors have no competing interests.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

AI tools, including Grammarly, were used to support language and grammar improvement during the preparation of this manuscript. The authors confirm that all content, including ideas, interpretations, analyses, and conclusions, is their own, and that AI tools were not used to generate any text

or content, data manipulation, or perform any part of the literature review. The final manuscript was reviewed and approved entirely by the authors.

ACKNOWLEDGMENT

The authors acknowledge the valuable insights and feedback from the reviewers, which significantly improved the quality of this manuscript.

REFERENCES

[1] R. Simó and C. Hernández, "What else can we do to prevent diabetic retinopathy?," *Diabetologia*, vol. 66, no. 9, pp. 1614–1621, Sep. 2023, doi: 10.1007/s00125-023-05940-5.

[2] Q. H. Yang, Y. Zhang, X. M. Zhang, and X. R. Li, "Prevalence of diabetic retinopathy, proliferative diabetic retinopathy and non-proliferative diabetic retinopathy in asian t2dm patients: A systematic review and metaanalysis," *Int J Ophthalmol*, vol. 12, no. 2, pp. 302–311, Feb. 2019, doi: 10.18240/ijo.2019.02.19.

[3] H. Y. Tsao, P. Y. Chan, and E. C. Y. Su, "Predicting diabetic retinopathy and identifying interpretable biomedical features using machine learning algorithms," *BMC Bioinformatics*, vol. 19, Aug. 2018, doi: 10.1186/s12859-018-2277-0.

[4] E. A. Alfonso-Muñoz, R. Burggraaf-Sánchez de las Matas, J. Mataix Boronat, J. C. Molina Martín, and C. Desco, "Role of oral antioxidant supplementation in the current management of diabetic retinopathy," *Int J Mol Sci*, vol. 22, no. 8, Apr. 2021, doi: 10.3390/ijms22084020.

[5] B. Mutunhu, B. Chipangura, and H. Twinomurinzi, "Internet of Things in the Monitoring of Diabetes," *International Journal of Health Systems and Translational Medicine*, vol. 2, no. 1, pp. 1–20, 2022, doi: 10.4018/ijhstm.300336.

[6] D. M. Mukona, P. Dzingira, M. Mhlanga, and M. Zvinavashe, "Uptake of Screening for Diabetic Retinopathy and Associated Factors among Adults with Diabetes Mellitus Aged 18-65 Years: A Descriptive Cross Sectional Study," *European Journal of Medical and Health Sciences*, vol. 2, no. 4, Jul. 2020, doi: 10.24018/ejmed.2020.2.4.247.

[7] I. Murere, B. Ndlovu, S. Dube, M. Muduva, and F. Jacqueline Kiwa, "Comparative Analysis of Machine Learning Techniques for Predicting Diabetes," in *Proceedings of the International Conference on Industrial Engineering and Operations Management*, Michigan, USA: IEOM Society International, Jul. 2024. doi: 10.46254/EU07.20240073.

[8] P. Uppamma and S. Bhattacharya, "A multidomain bio-inspired feature extraction and selection model for diabetic retinopathy severity classification: an ensemble learning approach," *Sci Rep*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-45886-7.

[9] L. Von Rueden *et al.*, "Informed Machine Learning - A Taxonomy and Survey of Integrating Prior Knowledge into Learning Systems," *IEEE Trans Knowl Data Eng*, vol. 35, no. 1, pp. 614–633, Jan. 2023, doi: 10.1109/TKDE.2021.3079836.

[10] R. Dwivedi *et al.*, "Explainable AI (XAI): Core Ideas, Techniques, and Solutions," *ACM Comput Surv*, vol. 55, no. 9, Sep. 2023, doi: 10.1145/3561048.

[11] A. B. Arrieta *et al.*, "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI," Oct. 2019, [Online]. Available: http://arxiv.org/abs/1910.10045

[12] L. Longo, R. Goebel, F. Lecue, P. Kieseberg, A. Holzinger, and A. H. Explainable, "Explainable Artificial Intelligence: Concepts, Applications, Research Challenges and Visions," pp. 1–16, 2020, doi: 10.1007/978-3-030-57321-8_1ï.

[13] Z. Khan *et al.*, "Diabetic Retinopathy Detection Using VGG-NIN a Deep Learning Architecture," *IEEE Access*, vol. 9, pp. 61408–61416, 2021, doi: 10.1109/ACCESS.2021.3074422.

[14] T. Shahzad, M. Saleem, M. S. Farooq, S. Abbas, M. A. Khan, and K. Ouahada, "Developing a Transparent Diagnosis Model for Diabetic Retinopathy Using Explainable AI," *IEEE Access*, 2024, doi: 10.1109/ACCESS.2024.3475550.

[15] D. Bhulakshmi and D. S. Rajput, "A systematic review on diabetic retinopathy detection and classification based on deep learning techniques using fundus images," *PeerJ Comput Sci*, vol. 10, 2024, doi: 10.7717/PEERJ-CS.1947.

[16] K. Ahnaf Alavee *et al.*, "Enhancing Early Detection of Diabetic Retinopathy Through the Integration of Deep Learning Models and Explainable Artificial Intelligence," *IEEE Access*, vol. 12, pp. 73950–73969, 2024, doi: 10.1109/ACCESS.2024.3405570.

[17] B. Ndlovu, K. Maguraushe, and O. Mabikwa, "Machine Learning and Explainable AI for Parkinson's Disease Prediction: A Systematic Review," *The Indonesian Journal of Computer Science*, vol. 14, no. 2, Apr. 2025, doi: 10.33022/ijcs.v14i2.4837.

[18] G. Quellec, H. Al Hajj, M. Lamard, P. H. Conze, P. Massin, and B. Cochener, "ExplAIn: Explanatory artificial intelligence for diabetic retinopathy diagnosis," *Med Image Anal*, vol. 72, Aug. 2021, doi: 10.1016/j.media.2021.102118.

[19] A. Ikram and A. Imran, "ResViT FusionNet Model: An explainable AI-driven approach for automated grading of diabetic retinopathy in retinal images," *Comput Biol Med*, vol. 186, Mar. 2025, doi: 10.1016/j.compbiomed.2025.109656.

[20] Y. Zou, Y. Wang, X. Kong, T. Chen, J. Chen, and Y. Li, "Deep Learner System Based on Focal Color Retinal Fundus Images to Assist in Diagnosis," *Diagnostics*, vol. 13, no. 18, Sep. 2023, doi: 10.3390/diagnostics13182985.

[21] P. Bidwai *et al.*, "Multimodal image fusion for the detection of diabetic retinopathy using optimized explainable AI-based Light GBM classifier," *Information Fusion*, vol. 111, Nov. 2024, doi: 10.1016/j.inffus.2024.102526.

[22] R. Raman *et al.*, "Prevalence of Diabetic Retinopathy in India. Sankara Nethralaya Diabetic Retinopathy Epidemiology and Molecular Genetics Study Report 2," *Ophthalmology*, vol. 116, no. 2, pp. 311–318, Feb. 2009, doi: 10.1016/j.ophtha.2008.09.010.

[23] Z. Khan, F. Khan, A. Khan, Z. Rehman, … S. S.-I., and undefined 2021, "Diabetic retinopathy detection using VGG-NIN a deep learning architecture," *ieeexplore.ieee.org*, Accessed: Nov. 12, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9409084/

[24] I. A. Taifa, D. M. Setu, T. Islam, S. K. Dey, and T. Rahman, "A hybrid approach with customized machine learning classifiers and multiple feature extractors for enhancing diabetic retinopathy detection," *Healthcare Analytics*, vol. 5, p. 100346, Jun. 2024, doi: 10.1016/J.HEALTH.2024.100346.

[25] T. Shahzad, M. Saleem, M. Farooq, … S. A.-I., and undefined 2024, "Developing a Transparent Diagnosis Model for Diabetic Retinopathy Using Explainable AI," *ieeexplore.ieee.org*, Accessed: Nov. 12, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10706847/

[26] W. X. Lim, Z. Y. Chen, and A. Ahmed, "The adoption of deep learning interpretability techniques on diabetic retinopathy analysis: a review," *Med Biol Eng Comput*, vol. 60, no. 3, pp. 633–642, Mar. 2022, doi: 10.1007/S11517-021-02487-8.

[27] U. Parmar, P. Surico, R. Singh, F. Romano, C. S.- Medicina, and undefined 2024, "Artificial intelligence (AI) for early diagnosis of retinal diseases," *mdpi.com*, Accessed: Nov. 11, 2024. [Online]. Available: https://www.mdpi.com/1648-9144/60/4/527

[28] R. Romero-Oraá, M. Herrero-Tudela, M. I. López, R. Hornero, and M. García, "Attention-based deep learning framework for automatic fundus image processing to aid in diabetic retinopathy grading," *Comput Methods Programs Biomed*, vol. 249, Jun. 2024, doi: 10.1016/j.cmpb.2024.108160.

[29] F. H. Yagin, C. Colak, A. Algarni, Y. Gormez, E. Guldogan, and L. P. Ardigò, "Hybrid Explainable Artificial Intelligence Models for Targeted Metabolomics Analysis of Diabetic Retinopathy," *Diagnostics*, vol. 14, no. 13, Jul. 2024, doi: 10.3390/diagnostics14131364.

[30] F. H. Yagin *et al.*, "Explainable Artificial Intelligence Paves the Way in Precision Diagnostics and Biomarker Discovery for the Subclass of Diabetic Retinopathy in Type 2 Diabetics," *Metabolites*, vol. 13, no. 12, Dec. 2023, doi: 10.3390/metabo13121204.

[31] X. Wang, W. Wang, H. Ren, X. Li, and Y. Wen, "Prediction and analysis of risk factors for diabetic retinopathy based on machine learning and interpretable models," *Heliyon*, vol. 10, no. 9, p. e29497, May 2024, doi: 10.1016/J.HELIYON.2024.E29497.

[32] G. Quellec, H. Al Hajj, M. Lamard, P. H. Conze, P. Massin, and B. Cochener, "ExplAIn: Explanatory artificial intelligence for diabetic retinopathy diagnosis," *Med Image Anal*, vol. 72, p. 102118, Aug. 2021, doi: 10.1016/J.MEDIA.2021.102118.

[33] P. Bidwai, S. Gite, K. Pahuja, and K. Kotecha, "A Systematic Literature Review on Diabetic Retinopathy Using an Artificial Intelligence Approach," Dec. 01, 2022, *MDPI*. doi: 10.3390/bdcc6040152.

[34] G. Chondrozoumakis *et al.*, "Retinal Biomarkers in Diabetic Retinopathy: From Early Detection to Personalized Treatment," *J Clin Med*, vol. 14, no. 4, p. 1343, Feb. 2025, doi: 10.3390/jcm14041343.

[35] K. Vinogradova, A. Dibrov, and G. Myers, "Towards Interpretable Semantic Segmentation via Gradient-Weighted Class Activation Mapping (Student Abstract)." [Online]. Available: www.aaai.org

[36] J. Li *et al.*, "Interpretable machine learning-derived nomogram model for early detection of diabetic retinopathy in type 2 diabetes mellitus: a widely targeted metabolomics study," *Nutr Diabetes*, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41387-022-00216-0.

[37] S. O'Sullivan *et al.*, "Explainable artificial intelligence (XAI): closing the gap between image analysis and navigation in complex invasive diagnostic procedures," May 01, 2022, *Springer Science and Business Media Deutschland GmbH*. doi: 10.1007/s00345-022-03930-7.

[38] A. K. Smilde *et al.*, "Dynamic metabolomic data analysis: A tutorial review," *Metabolomics*, vol. 6, no. 1, pp. 3–17, Mar. 2010, doi: 10.1007/s11306-009-0191-1.

[39] R. Younisse, A. Ahmad, and Q. Abu Al-Haija, "Explaining Intrusion Detection-Based Convolutional Neural Networks Using Shapley Additive Explanations (SHAP)," *Big Data and Cognitive Computing*, vol. 6, no. 4, Dec. 2022, doi: 10.3390/bdcc6040126.

[40] M. Rehman Zafar and N. Khan, "machine learning & knowledge extraction Deterministic Local Interpretable Model-Agnostic Explanations for Stable Explainability," 2021, doi: 10.3390/make.

[41] R. O. Alabi, M. Elmusrati, I. Leivo, A. Almangush, and A. A. Mäkitie, "Machine learning explainability in nasopharyngeal cancer survival using LIME and SHAP," *Sci Rep*, pp. 1–14, 2023, doi: 10.1038/s41598-023-35795-0.

[42] Y. Nohara, K. Matsumoto, H. Soejima, and N. Nakashima, "Explanation of Machine Learning Models Using Shapley Additive Explanation and Application for Real Data in Hospital," Dec. 2021, doi: 10.1016/j.cmpb.2021.106584.

[43] W. E. Marcilio-jr, D. M. Eler, and P. Brazil, "From explanations to feature selection : assessing SHAP values as feature selection mechanism".

[44] A. E. Maxwell, M. Sharma, and K. A. Donaldson, "Explainable boosting machines for slope failure spatial predictive modeling," *Remote Sens (Basel)*, vol. 13, no. 24, Dec. 2021, doi: 10.3390/rs13244991.

[45] E. Guldogan, F. H. Yagin, A. Pinar, C. Colak, S. Kadry, and J. Kim, "OPEN A proposed tree - based explainable artificial intelligence approach for the prediction of angina pectoris," *Sci Rep*, pp. 1–12, 2023, doi: 10.1038/s41598-023-49673-2.

[46] J. C. Chien, J. Der Lee, C. S. Hu, and C. T. Wu, "The Usefulness of Gradient-Weighted CAM in Assisting Medical Diagnoses," *Applied Sciences (Switzerland)*, vol. 12, no. 15, Aug. 2022, doi: 10.3390/app12157748.

[47] A. Chattopadhyay, A. Sarkar, and P. Howlader, "Grad-CAM ++ : Improved Visual Explanations for Deep Convolutional Networks".

[48] B. Mutunhu, B. Chipangura, and S. Singh, "An Exploration of Opportunities for Quantified-Self Technology in Diabetes Self-Care: A Systematic Literature Review," *J Health Inform Afr*, vol. 11, no. 2, pp. 17–30, 2024, doi: 10.12856/JHIA-2024-v11-i2-481.

[49] C. Fuyana, B. Ndlovu, S. Dube, K. Maguraushe, and L. Malungana, "Optimizing HIV Care Through Machine Learning-Assisted Prediction and Personalized Treatment," 2025, pp. 149–161. doi: 10.1007/978-981-96-2124-8_11.

[50] T. Shahzad, M. Saleem, M. S. Farooq, S. Abbas, M. A. Khan, and K. Ouahada, "Developing a Transparent Diagnosis Model for Diabetic Retinopathy Using Explainable AI," *IEEE Access*, 2024, doi: 10.1109/ACCESS.2024.3475550.

[51] X. Wang, W. Wang, H. Ren, X. Li, and Y. Wen, "Prediction and analysis of risk factors for diabetic retinopathy based on machine learning and interpretable models," *Heliyon*, vol. 10, no. 9, p. e29497, May 2024, doi: 10.1016/J.HELIYON.2024.E29497.

[52] U. Parmar, P. Surico, R. Singh, F. Romano, C. S.- Medicina, and undefined 2024, "Artificial intelligence (AI) for early diagnosis of retinal diseases," *mdpi.com*, Accessed: Nov. 11, 2024. [Online]. Available: https://www.mdpi.com/1648-9144/60/4/527

[53] B. Mutunhu, B. Chipangura, and H. Twinomurinzi, "A Systematized Literature Review: Internet of Things (IoT) in the Remote Monitoring of Diabetes," in *Lecture Notes in Networks and Systems*, Springer Science and Business Media Deutschland GmbH, 2023, pp. 649–660. doi: 10.1007/978-981-19-1610-6_57.

[54] K. Maguraushe and B. M. Ndlovu, "The use of smart technologies for enhancing palliative care: A systematic review," *Digit Health*, vol. 10, Jan. 2024, doi: 10.1177/20552076241271835.

[55] B. Mutunhu Ndlovu, B. Chipangura, and S. Singh, "Towards a quantified-self technology conceptual framework for monitoring diabetes," 2024, doi: 10.36303/SATNT.2024.43.1.970E.

[56] B. M. Ndlovu, B. Chipangura, and S. Singh, "Factors Influencing Quantified SelfTechnology Adoption in Monitoring Diabetes," in *Proceedings of Ninth International Congress on Information and Communication Technology*, X.-S. Yang, S. Sherratt, N. Dey, and A. Joshi, Eds., Singapore: Springer Nature Singapore, 2024, pp. 469–479.

[57] A. M. Antoniadi *et al.*, "Current challenges and future opportunities for xai in machine learning-based clinical decision support systems: A systematic review," *Applied Sciences (Switzerland)*, vol. 11, no. 11, Jun. 2021, doi: 10.3390/app11115088.

[58] A. Das and P. Rad, "Opportunities and Challenges in Explainable Artificial Intelligence (XAI): A Survey," Jun. 2020, [Online]. Available: http://arxiv.org/abs/2006.11371

[59] K. Reel, "The benefits of practice standards and other practice-defining texts: And why healthcare ethicists ought to explore them," *HEC Forum*, vol. 24, no. 3, pp. 203–217, Sep. 2012, doi: 10.1007/s10730-012-9186-9.

[60] I. A. Taifa, D. M. Setu, T. Islam, S. K. Dey, and T. Rahman, "A hybrid approach with customized machine learning classifiers and multiple feature extractors for enhancing diabetic retinopathy detection," *Healthcare Analytics*, vol. 5, Jun. 2024, doi: 10.1016/j.health.2024.100346.

[61] M. T. Hagos and S. Kant, "Transfer Learning based Detection of Diabetic Retinopathy from Small Dataset," 2019.

[62] S. Chokuwa and M. H. Khan, "Divergent Domains, Convergent Grading: Enhancing Generalization in Diabetic Retinopathy Grading," *Proceedings - 2025 IEEE Winter Conference on Applications of Computer Vision, WACV 2025*, pp. 3667–3677, 2025, doi: 10.1109/WACV61041.2025.00361.

[63] C. The- and T. H. Tower, "Chapter 34 (I) Computational Complexity Play with Hanoi (3) and beyond ...," vol. 34, no. I, pp. 1–13.

[64] C. Jean-quartier, K. Bein, L. Hejny, E. Hofer, A. Holzinger, and F. Jeanquartier, "The Cost of Understanding — XAI Algorithms towards Sustainable ML in the View of Computational Cost," pp. 1–14, 2023.

[65] P. Sengupta and F. Eliassen, "Balancing Explainability-Accuracy of Complex Models," no. Ml, pp. 1–18.

[66] Y. Luo, H.-H. Tseng, S. Cui, L. Wei, and R. K. Ten Haken, "Review article Balancing accuracy and interpretability of machine learning approaches for radiation treatment outcomes modeling," no. April 2019, pp. 1–12, 2022.

[67] A. I. Ml *et al.*, "FDA-Approved Artificial Intelligence and Machine Learning," 2024.

[68] A. Don, "Regulatory Perspectives of AI as Medical Device – a scoping review," no. November, 2024.

[69] H. Bleher and M. Braun, "Diffused responsibility: attributions of responsibility in the use of AI-driven clinical decision support systems," *AI and Ethics*, vol. 2, no. 4, pp. 747–761, Nov. 2022, doi: 10.1007/s43681-022-00135-x.

[70] B. Pickering, "Trust , but Verify : Informed Consent , AI Technologies , and Public Health Emergencies," 2021.

[71] C. Bascaran, M. Zondervan, C. Walker, N. J. Astbury, and A. Foster, "Diabetic retinopathy in Africa," May 01, 2022, *Springer Nature*. doi: 10.1038/s41433-022-01999-3.