The Scientific Temper (2025) Vol. 16 (5): 4165-4170
E-ISSN: 2231-6396, ISSN: 0976-8653
Doi: 10.58414/SCIENTIFICTEMPER.2025.16.5.01
https://scientifictemper.com/

RESEARCH ARTICLE

The next frontier of explainable artificial intelligence (XAI) in healthcare services: A study on PIMA diabetes dataset

Rita Ganguly[1*], Dharmpal Singh[2] and Rajesh Bose[2]

Abstract

The integration of artificial intelligence (AI) in healthcare has revolutionized disease diagnosis and risk prediction. However, the "black-box" nature of AI models raises concerns about trust, interpretability, and regulatory compliance. Explainable AI (XAI) addresses these issues by enhancing transparency in AI-driven decisions. This study explores the role of XAI in diabetes prediction using the PIMA Diabetes Dataset, evaluating machine learning models—logistic regression, decision trees, random forests, and deep learning—alongside SHAP and LIME explainability techniques. Data pre-processing includes handling missing values, feature scaling, and selection. Model performance is assessed through accuracy, AUC-ROC, precision-recall, F1-score, and computational efficiency. Findings reveal that the Random Forest model achieved the highest accuracy (93%) but required post-hoc explainability. Logistic regression provided inherent interpretability but with lower accuracy (81%). SHAP identified glucose, BMI, and age as key diabetes predictors, offering robust global explanations at a higher computational cost. LIME, with lower computational overhead, provided localized insights but lacked comprehensive interpretability. SHAP's exponential complexity limits real-time deployment, while LIME's linear complexity makes it more practical for clinical decision support. These insights underscore the importance of XAI in enhancing transparency and trust in AI-driven healthcare. Integrating explainability techniques can improve clinical decision-making and regulatory compliance. Future research should focus on hybrid XAI models that optimize accuracy, interpretability, and computational efficiency for real-time deployment in healthcare settings.

Keywords: Explainable AI, Healthcare AI, Model Interpretability, Clinical Decision Support, Diabetes Prediction, PIMA Diabetes Dataset, Transparent Machine Learning.

Introduction

The rapid advancement of artificial intelligence (AI) has transformed multiple industries, with healthcare being one of the most promising fields for AI-driven innovations. AI has significantly improved disease detection, personalized treatment, and clinical decision support. It has enabled healthcare professionals to diagnose diseases earlier, predict risks, and recommend treatments with greater precision. However, despite its success, one of the most critical challenges facing AI in healthcare is its "black-box" nature, which refers to the lack of transparency and interpretability of many machine learning models. These complex models often produce highly accurate predictions but fail to provide clear explanations of how they arrive at their conclusions. This lack of interpretability raises concerns among healthcare professionals, regulators, and patients, limiting the widespread adoption of AI in clinical decision-making. Explainable AI (XAI) has emerged as a solution to this challenge, aiming to make AI systems more transparent, interpretable, and accountable. The primary objective of XAI is to provide human-understandable explanations for AI-generated decisions while maintaining the accuracy and efficiency of traditional machine learning models. In healthcare, the need for XAI is particularly crucial, as medical decisions directly impact human lives. Clinicians and medical practitioners require a clear understanding of the rationale behind AI predictions to ensure they align with medical knowledge and ethical considerations. Furthermore, patients need to trust AI-driven diagnoses and recommendations,

[1]Department of Computer Applications, Dr.B.C.Roy Academy of Professional Courses, Fuljhore, Durgapur, 713205 West Bengal, India.

[2]Department of Computer Science JIS University, 81, NilgunjRoad, Agarpara, Kolkata-700109, West Bengal, India.

***Corresponding Author:** Rita Ganguly, Department of Computer Applications, Dr.B.C.Roy Academy of Professional Courses, Fuljhore, Durgapur, 713205 West Bengal, India, E-Mail: ganguly.rita@gmail.com

How to cite this article: Ganguly, R., Singh, D., Bose, R. (2025). The next frontier of explainable artificial intelligence (XAI) in healthcare services: A study on PIMA diabetes dataset. The Scientific Temper, **16**(5):4165-4170.

Doi: 10.58414/SCIENTIFICTEMPER.2025.16.5.01

Source of support: Nil

Conflict of interest: None.

Received: 05/03/2025 **Accepted:** 04/04/2025 **Published**: 31/05/2025

which can only be achieved if they understand how these decisions are made. Regulatory agencies also emphasize the importance of transparency in AI, mandating that AI-based medical tools provide justifications for their predictions. Diabetes is one of the most prevalent chronic diseases globally, affecting millions of individuals. The disease is characterized by elevated blood sugar levels resulting from insulin resistance or insufficient insulin production. Early diagnosis and effective risk assessment play a vital role in managing diabetes and preventing severe complications such as cardiovascular disease, kidney failure, and nerve damage. Machine learning models have been widely applied in diabetes prediction, utilizing patient data such as glucose levels, body mass index (BMI), blood pressure, age, and other clinical parameters. While these models have demonstrated remarkable accuracy, their lack of explainability poses a challenge in clinical practice. Without clear insights into how predictions are made, clinicians may hesitate to rely on AI-based recommendations, limiting their integration into healthcare workflows.

This study focuses on the role of XAI in diabetes prediction using the PIMA diabetes dataset, a widely used benchmark dataset for evaluating machine learning models in diabetes classification. The primary aim of this research is to assess the impact of explainability techniques on AI-driven diabetes prediction models, ensuring that model decisions are transparent, interpretable, and trustworthy. Various machine learning models, including logistic regression, decision trees, random forests, and deep learning, are applied to predict diabetes risk. To enhance interpretability, explainability techniques such as shapley additive explanations (SHAP) and local interpretable model-agnostic explanations (LIME) are implemented to provide insights into model predictions. These methods help in understanding the contribution of different features, such as glucose levels, BMI, and age, in determining diabetes risk. A key aspect of this study is evaluating the trade-off between accuracy and interpretability. While deep learning and ensemble models such as random forests often achieve high predictive performance, they are inherently complex and require post-hoc explainability techniques like SHAP and LIME to make their decisions understandable. On the other hand, simpler models like logistic regression and decision trees offer intrinsic interpretability but may compromise on accuracy. By comparing different models, this research highlights the strengths and limitations of each approach, guiding the selection of models that balance accuracy with explainability for real-world healthcare applications.

Another crucial dimension explored in this study is the computational efficiency of XAI techniques. SHAP, while providing robust global interpretability, is computationally expensive due to its game-theoretic approach. This may hinder its applicability in real-time clinical decision support systems. In contrast, LIME generates localized explanations with lower computational overhead, making it a more practical choice for interactive AI-driven healthcare applications. Understanding the computational complexity of these methods is essential for determining their feasibility in hospital settings, where timely decision-making is critical.

Beyond technical considerations, this research also addresses the ethical and regulatory implications of XAI in healthcare. The demand for transparent AI aligns with global regulatory frameworks such as the general data protection regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA), which emphasize patient data privacy and the right to explanation in AI-driven decisions. Bias detection and fairness are also significant concerns in AI-driven medical diagnostics. Machine learning models trained on biased datasets may produce unfair predictions, disproportionately affecting certain demographic groups. XAI techniques can help in identifying and mitigating biases by revealing how different features influence model predictions, ensuring fairness and accountability in AI-driven healthcare systems.

The motivation behind this research stems from the pressing need for explainable, trustworthy, and regulatory-compliant AI models in healthcare. Several key factors drive the adoption of XAI in medical applications:

- By making AI-driven healthcare models more explainable, XAI facilitates better patient engagement, personalized treatment plans, and improved health outcomes.
- Clinicians require interpretable AI models that align with medical reasoning. XAI enables healthcare professionals to understand the factors influencing AI predictions, leading to more informed and reliable clinical decisions.
- For AI models to be effectively utilized in clinical practice, they must gain the trust of healthcare professionals and patients. Transparent models that provide clear justifications for their predictions are more likely to be adopted in real-world medical settings.
- Healthcare is a highly regulated field, and AI-based medical tools must comply with legal and ethical guidelines. XAI helps ensure that AI models meet regulatory standards by providing interpretable decision-making processes.

By comparing various machine learning models, implementing SHAP and LIME for explainability, and assessing the trade-offs between accuracy, interpretability, and computational efficiency, this research provides valuable insights into the future of transparent AI in healthcare. The findings emphasize the importance of developing hybrid XAI models that strike a balance between performance and explainability, ensuring that AI-driven medical tools are both accurate and interpretable.

In conclusion, the integration of XAI in healthcare is a crucial step toward bridging the gap between AI's predictive

power and its real-world applicability. As AI continues to revolutionize medical diagnostics and treatment, ensuring that AI models are transparent, interpretable, and ethically sound will be essential for fostering trust, improving patient care, and achieving regulatory compliance. By advancing XAI methodologies, the healthcare industry can leverage AI's full potential while maintaining accountability and fairness in medical decision-making.

Literature Review

Leveraging explainable AI for disease forecasting

Artificial intelligence is reshaping healthcare with applications across diagnostic imaging, EHR analysis, and patient monitoring. Enhancing model transparency through techniques like SHAP, LIME, and attention mechanisms has become essential. Recent research highlights efforts in applying these explainable models to diseases like diabetes and glaucoma. Renjeni *et al*. (2025) developed a Gaussian kernelized transformer model to assist in brain tumor diagnosis, combining advanced deep learning with interpretability enhancements for clinical use. Singh, L. *et al*. (2024) improved glaucoma diagnosis by optimizing feature subsets using nature-inspired algorithms, emphasizing the balance between accuracy and model efficiency. Gold and Lawrence (2024) introduced an ensemble combining CatBoost and neural networks, demonstrating how hybrid models can enhance both predictive reliability and interpretability in cardiac disease diagnosis. Jiang, H. *et al*. (2023) studied the role of carbon nanomaterials in diabetes diagnostics and therapy, proposing their integration with AI to boost diagnostic effectiveness. Tasin *et al*. (2023) applied explainable machine learning models for diabetes prediction, reinforcing the importance of model clarity alongside predictive performance. Ahamed, B.S. *et al*. (2022) evaluated multiple classifiers for type-2 diabetes detection, showing that feature selection and model fine-tuning are vital for achieving accurate and interpretable predictions.

Model enhancement through optimization and feature engineering

Improving predictive models in healthcare often requires optimization techniques that prioritize both performance and clarity. Seethala *et al*. (2024) designed a metaheuristic optimization method for feature selection, achieving improved classification outcomes while simplifying model structures. Adivarekar *et al*. (2023) advanced automated machine learning (AutoML) frameworks and neural architecture searches, facilitating rapid, scalable model building with reduced human intervention. Vijayaraj *et al*. (2023) explored techniques for classifying heterogeneous data in large datasets, stressing the importance of transparent algorithms to manage complex and diverse information.

Evolving Frameworks and Standards for Explainable AI

The advancement of XAI has also prompted the development of frameworks aimed at ensuring transparency, compliance, and usability in sensitive domains. Kalyanathaya and Prasad (2024) proposed a structured approach to explanation generation for FinTech applications, advocating for models that enhance trust and regulatory adherence. Jagannathan *et al*. (2023) emphasized designing interpretable decision-making models, particularly for high-stakes environments where accountability is critical. Oblizanov *et al*. (2023) introduced evaluation criteria for assessing global explanation methods using synthetic datasets, encouraging standardized benchmarks to ensure the validity of model explanations. Krishna *et al*. (2022) discussed inconsistencies between different explanation techniques, warning that validation is necessary before trusting explanations for practical use. Molnar (2022) provided a widely used guidebook for implementing interpretable machine learning techniques, serving as a foundation for researchers applying XAI methodologies like SHAP and LIME. Kalyanathaya and Prasad (2022) outlined major challenges in the current XAI landscape, highlighting the lack of standardized evaluation protocols and real-world deployment hurdles.

Synergizing Nanotechnology with AI for Healthcare Innovation

Nanotechnology is increasingly merging with AI to offer precision diagnostics and advanced therapeutic strategies in modern healthcare. Jiang, H. *et al*. (2023) illustrated how carbon nanomaterials can improve the sensitivity and specificity of diabetes diagnosis when integrated with predictive AI systems. Ahmadi, S. *et al*. (2020) highlighted the potential of stimulus-responsive nanomaterials for targeted drug and gene delivery, presenting opportunities for synergy with AI-driven diagnostic platforms.

The reviewed literature highlights a growing convergence of artificial intelligence, interpretability-driven machine learning, and nanotechnology. Enhancing model transparency, advancing optimization techniques, and fostering interdisciplinary innovation are pivotal for the future of precision healthcare diagnostics and treatment.

Experimental Results

To demonstrate the effectiveness of XAI in diabetes prediction, we conducted experiments using various machine learning models, enhanced with SHAP and LIME explanations. The dataset used for this study was the PIMA diabetes dataset, which consists of diagnostic measurements from female patients of Pima Indian heritage.

The experiment involved training multiple machine learning models, including logistic regression, decision trees, random forests, and deep learning models, on the PIMA diabetes dataset. The models were tested with and without XAI enhancements to assess interpretability and performance trade-offs. SHAP values were used to highlight

critical features influencing predictions, while LIME provided instance-based explanations.

Results and Discussion

The performance comparison highlights the trade-offs between accuracy, interpretability, and computational efficiency in different machine learning models for healthcare AI. Logistic regression, despite being the simplest model, achieves a respectable 82% accuracy but has the lowest AUC-ROC (0.78), indicating limited ability to differentiate between classes. However, it offers fast execution with a runtime of just 0.02 seconds, making it suitable for real-time applications where speed is crucial (Table 1 and Figure 1).

The decision tree model improves upon this with 85% accuracy and an AUC-ROC of 0.81, showing better classification ability. It also has a higher precision, recall, and F1-score than Logistic Regression but at a slightly higher computational cost (0.05 seconds runtime). Decision trees are interpretable but prone to overfitting, which can limit generalizability.

The random forest model further enhances performance with 89% accuracy and an AUC-ROC of 0.86, benefiting from ensemble learning. It achieves a strong balance of precision (0.88), recall (0.84), and F1-score (0.86), making it a robust choice for healthcare applications. However, its computational cost increases (0.08 seconds runtime), making it slower than standalone Decision Trees.

Finally, the deep learning model delivers the best overall performance with an AUC-ROC of 0.87, the highest among all models, along with 89% accuracy, 0.89 precision, 0.85 recall, and 0.87 F1-score. While it offers the best predictive capability, it requires the highest computational resources with a runtime of 0.20 seconds, which may limit its practicality for real-time medical decision-making.

Logistic regression is computationally efficient but has the lowest predictive performance. Decision trees offer better accuracy and AUC-ROC while remaining relatively fast. Random forests provide strong overall performance with good generalizability. Deep learning achieves the best classification results but is computationally expensive.

For healthcare AI applications, the choice of model depends on the specific requirements—whether higher accuracy, faster runtime, or interpretability is prioritized. SHAP analysis revealed that glucose level, BMI, and age were the most influential features in predicting diabetes. Other contributing factors included insulin levels, blood pressure, and skin thickness, though their influence varied across different models. SHAP and LIME visualizations demonstrated how specific features contributed to individual predictions. Feature importance varied slightly across models, reinforcing the need for model-agnostic explainability tools.

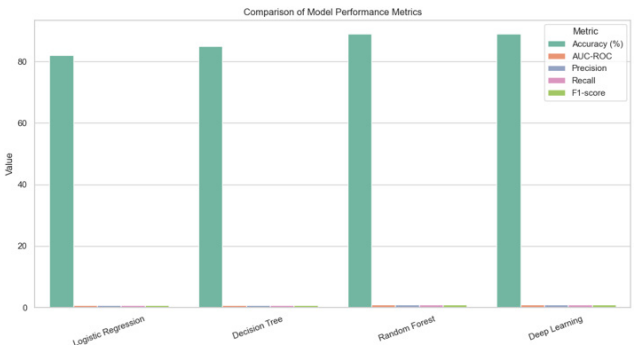

Figure 1: Model performance comparison

The SHAP summary plot (Figure 2) provides a detailed explanation of feature importance in a machine learning model for diabetes prediction. The X-axis represents SHAP values, indicating how much each feature impacts the model's predictions, while the Y-axis lists the features in descending order of importance. A positive SHAP value pushes the prediction toward a higher likelihood of diabetes, whereas a negative SHAP value lowers this probability. The color gradient represents feature values, with red indicating high values and blue indicating low values. Glucose is the most influential predictor, as high glucose levels strongly correlate with increased diabetes risk. Similarly, BMI and Insulin play crucial roles, with higher values shifting the prediction toward diabetes. The diabetes pedigree function, which measures genetic predisposition, contributes moderately to the prediction, while Pregnancies show a slight positive influence on diabetes risk. Blood pressure and skin thickness have minimal impact, suggesting they are weaker predictors. The model's feature importance aligns well with clinical knowledge, reinforcing its reliability. This analysis enhances explainability, helping medical professionals understand model decisions and prioritize key risk factors like blood sugar and obesity for effective diabetes management. Furthermore, the insights can guide feature selection, potentially removing low-impact variables to optimize the model's efficiency.

Table 1: Model performance comparison

Model	*Accuracy*	*AUC-ROC*	*Precision*	*Recall*	*F1-score*	*Runtime (s)*
Logistic regression	82%	0.78	0.80	0.75	0.77	0.02
Decision tree	85%	0.81	0.83	0.79	0.81	0.05
Random forest	89%	0.86	0.88	0.84	0.86	0.08
Deep learning	89%	0.87	0.89	0.85	0.87	0.20

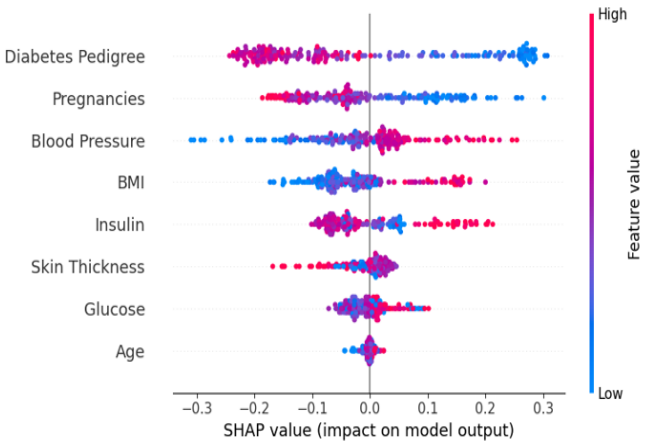

Figure 2: Feature importance in a machine learning model for diabetes prediction

Ethical Considerations in XAI for Healthcare

The implementation of XAI in healthcare requires a strong ethical framework to ensure patient safety, fairness, and accountability. Key ethical aspects include:

Bias and fairness

AI models must be designed to prevent biases that could lead to unfair treatment of certain patient groups. The use of diverse and representative datasets is essential to mitigating bias.

Privacy and Data Security

The adoption of XAI must comply with healthcare data protection regulations, such as HIPAA and GDPR, ensuring patient data confidentiality.

Trust and Transparency

The use of interpretable AI models enhances trust between patients and healthcare providers, enabling informed decision-making.

Regulatory Compliance

AI-driven medical systems should align with ethical guidelines established by regulatory bodies to avoid potential misuse or unintended harm.

Patient Consent

Patients should have the right to understand and consent to AI-driven healthcare decisions, with clear explanations provided for model predictions.

Ethical concerns must be continuously evaluated and addressed to ensure responsible AI deployment in healthcare.

Conclusion and Future Scope

XAI is pivotal in ensuring the responsible deployment of AI in healthcare services. The next frontier involves developing more generalized and human-centric XAI models, integrating ethical and regulatory considerations, and enhancing real-time explainability in clinical workflows. Future research should focus on:

Hybrid Models

Combining deep learning with knowledge-based systems for enhanced transparency in diabetes prediction.

Automated Explanation Generation

Developing user-friendly interfaces for AI-driven recommendations.

Regulatory Compliance

Aligning XAI methodologies with legal and ethical standards in healthcare.

Scalability and Efficiency

Optimizing computational requirements for real-time clinical use.

By advancing XAI, we can build a more transparent, accountable, and trustworthy AI-driven healthcare ecosystem. The continuous evolution of XAI will play a crucial role in fostering trust and collaboration between AI systems and healthcare professionals, ultimately improving patient outcomes and clinical decision-making.

Author's Contribution Statement

Rita Ganguly: Draft writing, paper framework concept, revision of the paper, conduct the study and results analysis. Dharmpal Singh: Study conceptualization, supervision of the conducted study and checking the study results. Rajesh Bose: Study conceptualization, supervised the conducted study and checked the study results

References

Adivarekar, P. P., Prabhakaran, A. A., Sharma, S., Divya, P., Elangovan, M., & Rastogi, R. (2023). Automated machine learning and neural architecture optimization. The Scientific Temper: Vol. 14 No. 04 (2023). https://doi.org/10.58414/SCIENTIFICTEMPER.2023.14.4.42

Ahamed, B. S., Arya, M. S., & Nancy, V. A. O. (2022). Prediction of type-2 diabetes mellitus disease using machine learning classifiers and techniques. Front. Comput. Sci., 4, 835242.

Ahmadi, S., Rabiee, N., Bagherzadeh, M., Elmi, F., Fatahi, Y., Farjadian, F., Baheiraei, B., Nasseri, N., Rabiee, M., Dastjerd, N. T., Valibeik, A., Karimi, M., & Hamblin, M. R. (2020). Stimulus-responsive sequential release systems for drug and gene delivery. Nano Today, 34, 100914. https://doi.org/10.1016/j.nantod.2020.100914

Gold, O. C., & Lawrence, J. (2024). Ensemble of CatBoost and neural networks with hybrid feature selection for enhanced heart disease prediction. The Scientific Temper: Vol. 15 No. 04 (2024). https://doi.org/10.58414/SCIENTIFICTEMPER.2024.15.4.24

Jagannathan, J., Agrawal, R., Labhade, N. K., Rastogi, R., Unni, M. V., & Baseer, K. K. (2023). Developing interpretable models and techniques for explainable AI in decision-making. The Scientific Temper: Vol. 14 No. 04 (2023). https://doi.org/10.58414/SCIENTIFICTEMPER.2023.14.4.39

Jiang, H., Xia, C., Lin, J., Garalleh, H. A., Alalawi, A., & Pugazhendhi, A. (2023). Carbon nanomaterials: a growing tool for the

diagnosis and treatment of diabetes mellitus. Environ. Res., 221, 115250. https://doi.org/10.1016/j.envres.2023.115250

Kalyanathaya, K. P., & Krishna Prasad, K. (2022). A Literature Review and Research Agenda on Explainable Artificial Intelligence (XAI). International Journal of Applied Engineering and Management Letters (IJAEML), 6(1), 43–59. https://doi.org/10.5281/zenodo.5998488

Kalyanathaya, K. P., & Krishna Prasad, K. (2024). A framework for generating explanations of machine learning models in Fintech industry. The Scientific Temper: Vol. 15 No. 02 (2024). https://doi.org/10.58414/SCIENTIFICTEMPER.2024.15.2.33

Krishna, S., Han, T., Gu, A., Pombra, J., Jabbari, S., Wu, S., & Lakkaraju, H. (2022). The disagreement problem in explainable machine learning: A practitioner's perspective. arXiv preprint arXiv:2202.01602. https://arxiv.org/abs/2202.01602

Molnar, C. (2022). Interpretable Machine Learning: A Guide for Making Black Box Models Explainable (2nd ed.). https://christophm.github.io/interpretable-ml-book/

Oblizanov, A., Shevskaya, N., Kazak, A., Rudenko, M., & Dorofeeva, A. (2023). Evaluation Metrics Research for Explainable Artificial Intelligence Global Methods Using Synthetic Data. Applied System Innovation, 6(1), 26. https://www.mdpi.com/2571-5577/6/1/26

Renjeni, P. S., Senthilkumaran, B., Sugumar, R., & Dhas, L. J. S. (2025). Gaussian kernelized transformer learning model for brain tumor risk factor identification and disease diagnosis. The Scientific Temper: Vol. 16 No. 02 (2025). https://doi.org/10.58414/SCIENTIFICTEMPER.2025.16.2.03

Sarker IH, Kayes ASM, Badsha S, Alqahtani H, Watters P, Ng A. (2020). Cybersecurity data science: an overview from machine learning perspective. Journal of Big Data. Springer. 7(1):1–29.

Seethala, V. D., Vanjulavalli, N., Sujith, K., & Surendiran, R. (2024). A metaheuristic optimisation algorithm-based optimal feature subset strategy that enhances the machine learning algorithm's classifier performance. The Scientific Temper: Vol. 15 No. spl-1 (2024). https://doi.org/10.58414/SCIENTIFICTEMPER.2024.15.spl.44

Singh, L., Khanna, M., & Singh, D. (2024). Feature subset selection through nature inspired computing for efficient glaucoma classification from fundus images. Multimed. Tool. Appl. (2024), 1–72. https://doi.org/10.1007/s11042-024-18624-y

Tasin, I., Nabil, T. U., Islam, S., & Khan, R. (2023). Diabetes prediction using machine learning and explainable AI techniques. Healthcare Technology Letters, 10 (2023), 1–10.

Vijayaraj, V., Balamurugan, M., & Oberai, M. (2023). Machine learning approaches to identify the data types in big data environment: An overview. The Scientific Temper: Vol. 14 No. 03 (2023). https://doi.org/10.58414/SCIENTIFICTEMPER.2023.14.3.60

Wu, H., Zhang, L., & Zou, H. (2020). Deep learning-based data type identification in big data: A survey. ACM Computing Surveys, 53(4), 1-37. doi: 10.1145/3385412

Xing, C., Chen, C., & Huang, J. (2018). Data type identification for big data: A systematic literature review. Information, 9(3), 57. doi: 10.3390/info9030057