ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

A Review on Explainable Artificial Intelligence in Healthcare BPOS-Application and Challenges for Sustainability of Business

Prof. Shreedhar Deshmukh

Assist Professor, NSB World Business school

DOI: https://doi.org/10.51583/IJLTEMAS.2026.150300126

Received: 27 March 2026; 02 April 2026; Published: 24 April 2026

ABSTRACT

The rapid advancement of Artificial Intelligence (AI) and Machine Learning (ML) across healthcare domains has significantly improved diagnostic accuracy, operational efficiency, and patient outcomes (Rajkomar et al., 2019; Jiang et al., 2017). Despite these benefits, concerns regarding the lack of transparency and interpretability in AI systems remain critical, particularly in high-stakes environments such as healthcare (Doshi-Velez & Kim, 2017; Tjoa & Guan, 2020). Many AI models operate as "black boxes," making it difficult for stakeholders to understand the rationale behind their decisions, thereby raising issues of trust, accountability, and ethical compliance (Arrieta et al., 2020; Floridi et al., 2018).

Healthcare Business Process Outsourcing (BPO) organizations increasingly utilize AI-driven tools to automate medical document processing for insurance claims. These documents, including prescriptions, lab reports, and radiology records, must be classified and chronologically organized to reflect patient history. Traditional ML approaches such as Count Vectorization and TF-IDF combined with classification algorithms achieve high accuracy; however, they rely heavily on word frequency, which can introduce bias and lead to misclassification (Wang et al., 2018; Obermeyer et al., 2019).

To address these limitations, Explainable Artificial Intelligence (XAI) has emerged as a promising solution that enhances transparency and interpretability in AI systems (Ribeiro et al., 2016; Lundberg & Lee, 2017). This study presents a systematic review of XAI techniques in healthcare, focusing on their application in medical document classification. It further identifies key challenges, including the lack of standardized evaluation metrics, and proposes future research directions to enhance transparency, fairness, and reliability in AI-driven healthcare systems (Arrieta et al., 2020; Tjoa & Guan, 2020).

Keywords: Artificial Intelligence, Explainable AI, Healthcare BPO, Medical Document Classification, Ethics, Fairness.

INTRODUCTION

Artificial Intelligence (AI) has transformed multiple industries, with healthcare witnessing particularly significant advancements in recent years (Jiang et al., 2017). The integration of AI technologies has enabled improved diagnostic precision, personalized treatment planning, and enhanced operational efficiency (Rajkomar et al., 2019). Machine Learning (ML) and Deep Learning (DL), as core components of AI, facilitate the analysis of large-scale medical datasets, supporting predictive analytics and informed clinical decision-making (Esteva et al., 2017; Rajkomar et al., 2019).

However, the growing complexity of AI models has led to reduced transparency, often resulting in "black-box" systems whose decision-making processes are not easily interpretable. This lack of explainability poses serious concerns in healthcare, where decisions directly impact patient safety and clinical outcomes (Doshi-Velez & Kim, 2017). Explainable Artificial Intelligence (XAI) addresses this challenge by providing interpretable insights into model behavior, thereby enhancing trust, accountability, and adoption (Tjoa & Guan, 2020; Arrieta et al., 2020).

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

Existing literature has extensively explored AI applications in healthcare, including predictive analytics, diagnostics, and drug discovery (Rajkomar et al., 2019; Obaido et al., 2024). Additionally, several studies have highlighted ethical and regulatory concerns associated with AI deployment (Floridi et al., 2018; Obermeyer et al., 2019). However, limited research focuses specifically on XAI within healthcare contexts, particularly in operational domains such as medical document processing. Therefore, this study aims to bridge this gap by providing a comprehensive review of XAI techniques, their applications, challenges, and future directions in healthcare.

LITERATURE REVIEW

The application of Artificial Intelligence (AI) in healthcare has expanded rapidly, with numerous studies examining its impact across both clinical and operational settings. Prior research has highlighted AI's potential to enhance diagnostic accuracy, improve patient care, and optimize healthcare delivery systems (Jiang et al., 2017; Rajkomar et al., 2019). For instance, Maleki and Forouzanfar emphasized the role of AI in clinical decision support systems, while Kalra et al. explored the integration of AI into electronic health systems, identifying key challenges related to workflow adaptation and system interoperability. Similarly, Lee et al. (2023) reviewed various AI models used in disease prediction and management, demonstrating their effectiveness in improving healthcare outcomes.

Explainable Artificial Intelligence (XAI) has emerged as a critical subfield focused on improving the transparency and interpretability of AI systems. Hulsen (2023) categorized XAI techniques into model-specific and model-agnostic approaches, emphasizing their importance in building trust and accountability in AI-driven decision-making. Lesley and Hernández (2024) further examined physician perspectives on XAI, highlighting the necessity of interpretable explanations for clinical adoption and decision support. Additionally, Longo et al. (2020) introduced the concept of XAI 2.0, advocating for interdisciplinary approaches and identifying future research challenges in achieving meaningful explainability.

Despite these contributions, much of the existing literature provides either a broad overview of AI applications in healthcare or focuses primarily on clinical use cases. There remains a lack of domain-specific research addressing XAI applications in healthcare Business Process Outsourcing (BPO), particularly in areas such as medical document classification and operational workflow optimization. This gap underscores the need for focused studies that integrate explainability into AI-driven document processing systems to enhance both efficiency and trustworthiness.

Problem Statement

Healthcare BPO organizations supporting insurance claims processing face significant challenges in managing large volumes of medical documents, including handwritten prescriptions, lab reports, and radiology images. These documents must be accurately classified and organized chronologically to reflect patient medical history.

Traditional **Healthcare BPO (Business Process Outsourcing)** relies heavily on **manual tasks**, including **data entry, claims processing, medical coding, patient record management, and billing**. These processes are time-consuming, error-prone, and require significant human effort.

Even with the adoption of AI techniques such as Optical Character Recognition (OCR) and Natural Language Processing (NLP), challenges persist in ensuring accurate classification and interpretability.

Conventional ML approaches using Count Vectorization and TF-IDF achieve high classification accuracy but lack explainability. These methods depend on word frequency, which can introduce bias and result in incorrect classifications without providing justification. In critical domains like healthcare, such limitations can lead to serious consequences, including incorrect claim processing and reduced trust in AI systems.

Therefore, there is a need for an explainable and reliable AI framework that not only classifies medical documents accurately but also provides interpretable insights into its decision-making process.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

Traditional Machine Learning Approach

The traditional approach to medical document classification involves the following steps:

1. **Text Extraction:** Conversion of PDF documents into machine-readable text.
2. **Preprocessing:** Removal of noise, including punctuation, symbols, and irrelevant data.
3. **Feature Extraction:** Application of Count Vectorization and TF-IDF to convert text into numerical representations.
4. **Model Training:** Use of ML algorithms such as Naïve Bayes, Support Vector Machines, or deep learning models.
5. **Classification:** Assignment of documents to predefined categories based on learned patterns.
6. **Evaluation:** Performance assessment using metrics such as precision, recall, and F1-score.

Although this approach yields high accuracy, it lacks transparency, making it difficult to validate classification decisions.

RESULTS AND DISCUSSION

The performance of the traditional machine learning model for medical document classification is presented in Table 1. The evaluation metrics include **precision, recall, and F1-score**, which are standard measures for assessing classification performance.

Class	Precision	Recall	F1-Score	Support
Affidavit	1	0.96	0.98	46
Anaesthesia Record	0.99	0.79	0.88	106
CT	0.94	0.99	0.96	423
Consent	0.93	0.96	0.95	269
Culture and Sensitivity Test	1	0.94	0.97	94

The results demonstrate that the model achieves **high classification performance across most categories**, with precision values ranging from 0.93 to 1.00 and F1-scores above 0.88. This indicates that the model is generally effective in correctly identifying and classifying medical documents.

Performance Interpretation

- The Affidavit and Culture_and_Sensitivity_Test classes achieve near-perfect precision (1.00), indicating that when the model predicts these classes, it is almost always correct.
- The CT (Computed Tomography) category shows a very high recall (0.99), meaning the model successfully identifies nearly all relevant CT documents. This is likely due to the presence of distinctive medical terminology associated with radiology reports.
- The Consent class also performs well, with balanced precision and recall, suggesting consistent classification performance.

However, the Anesthesia_Record category exhibits a comparatively lower recall (0.79), indicating that a significant number of relevant documents are not correctly identified. This may be due to:

- Overlapping terminology with other document types

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

- Variability in document structure and wording
- Insufficient distinguishing features in the dataset

Limitations of the Results

Despite achieving high accuracy, the model exhibits a critical limitation: lack of interpretability. The classification decisions are based on statistical patterns derived from techniques such as TF-IDF and word frequency, but:

- The model does not provide explicit reasoning for its predictions
- It is unclear which features (words or phrases) influenced the classification
- There is no mechanism to validate or justify predictions, especially in borderline cases
- Misclassifications cannot be easily diagnosed or corrected

This limitation is consistent with concerns raised in prior research regarding "black-box" machine learning models, which lack transparency and hinder trust in decision-making systems (Doshi-Velez & Kim, 2017; Arrieta et al., 2020).

Implications for Healthcare Applications

In healthcare BPO contexts, where document classification directly impacts insurance claims processing and patient records management, such lack of explainability poses serious risks:

- Incorrect classification may lead to claim rejection or delays
- Lack of justification reduces trust among domain experts (e.g., doctors, auditors)
- Regulatory compliance becomes difficult without traceable decision logic

Need for Explainable AI

These limitations highlight the necessity of integrating Explainable Artificial Intelligence (XAI) techniques. Methods such as LIME and SHAP can provide:

- Feature-level explanations
- Visualization of important words influencing classification
- Insights into model errors and biases

By incorporating XAI, the system can move beyond accuracy metrics and provide interpretable, trustworthy, and auditable predictions, which are essential in healthcare environments (Tjoa & Guan, 2020).

Explainable AI for Medical Document Classification

Explainable AI (XAI) enhances traditional ML models by providing interpretable explanations for their predictions. One widely used technique is **Local Interpretable Model-Agnostic Explanations (LIME)**.

LIME-Based Approach

- Generates perturbed samples of input data
- Observes changes in model predictions

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

- Assigns importance scores to features (words)
- Highlights key factors influencing classification

This enables users to understand why a document is classified into a particular category and helps identify potential biases or errors.

Gap Analysis

Despite advancements in XAI, several challenges remain:

- **Lack of Standard Metrics:** No universally accepted benchmarks for evaluating explanation quality
- **Interpretability vs Accuracy Trade-off:** Balancing model performance with explainability
- **User Trust Evaluation:** Limited studies on how explanations influence user trust
- **Domain-Specific Adaptation:** Need for healthcare-specific XAI frameworks

CONCLUSION

The integration of AI and XAI in healthcare BPO operations significantly improves efficiency, accuracy, and scalability in medical document processing. By automating classification and enabling explainability, organizations can reduce workload, enhance decision-making, and ensure compliance with regulatory standards.

Furthermore, XAI fosters trust and transparency, making AI systems more reliable and acceptable in critical healthcare applications. Future research should focus on developing standardized evaluation frameworks and domain-specific explainability techniques to further enhance AI adoption in healthcare.

REFERENCES

1. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., et al. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges. Information Fusion, 58, 82–115.
2. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. NAACL-HLT.
3. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.
4. Floridi, L., Cowls, J., Beltrametti, M., et al. (2018). AI4People—An ethical framework for a good AI society. Minds and Machines, 28(4), 689–707.
5. Jiang, F., Jiang, Y., Zhi, H., et al. (2017). Artificial intelligence in healthcare: Past, present and future. Stroke and Vascular Neurology, 2(4), 230–243.
6. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. NeurIPS.
7. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage populations. Science, 366(6464), 447–453.
8. Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347–1358.
9. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should I trust you? Explaining predictions of any classifier. KDD.
10. Selvaraju, R. R., Cogswell, M., Das, A., et al. (2017). Grad-CAM: Visual explanations from deep networks. ICCV.
11. Tjoa, E., & Guan, C. (2020). A survey on explainable artificial intelligence (XAI). IEEE Transactions on Neural Networks and Learning Systems, 32(11), 4793–4813.
12. Wang, Y., Wang, L., Rastegar-Mojarad, M., et al. (2018). Clinical information extraction applications: A literature review. Journal of Biomedical Informatics, 77, 34–49.