Multilingual Document Classification using XAI: A Review

Shalini Puri[1], Midhunchakkaravarthy Janarthanan[2], Ganesh Khekare[3]

[1]Lincoln University College Malaysia; [2]Lincoln University College Malaysia; [3]Vellore Institute of Technology, Vellore, India

eng.shalinipuri30@gmail.com

midhun@lincoln.edu.my

khekare.123@gmail.com

Abstract: As information systems have become more globalized, the importance of multilingual document classification has increased. Although effective, traditional AI models often function as opaque black boxes, they are being improved to improve the explainability of the models. This work employs machine learning and explainable artificial intelligence models to classify documents in multiple languages. It presents research questions regarding document classification utilizing XAI and ML models, while also optimizing outcomes with fuzzy logic and convolutional neural networks. The study reviews several current models and contrasts their approaches. Furthermore, it provides analytical findings derived from previous studies. Notably, few existing research efforts utilizing XAI have been proposed for multilingual document classification to date.

Keywords: XAI; Machine Learning; Sustainable Learning; Multilingualism; Document Categorization; Images.

Introduction

From the early conceptual theories, Artificial Intelligence (AI) and Machine Learning (ML) have advanced significantly and have become essential components of modern technological civilization [1]. While effective, the rapid development and widespread use of AI and ML in military and civilian applications have also introduced new challenges [1]. As some emerging AI and ML decision-making systems operate with minimal human intervention, it is crucial to understand their decision-making processes [1]. This need has led to the emergence of Explainable AI (XAI), a new area of AI research [1]. XAI focuses on creating AI systems to explain their decisions or predictions to human users [5]. This is particularly important in high-stakes fields such as healthcare, finance, and security, where transparency, reliability, and accountability are vital [5]. XAI and ML are applied in various domains, including legal, medical, and educational document classification, along with many other applications across diverse fields like policymaking [9].

Document image classification differs from plain-text document classification in that it involves understanding the structure and content of documents, such as emails, forms, and other documents [3].

A document is deemed “responsive” in review situations if any of its brief passages meet the criteria for responsiveness [4]. Consider an application that converts text into images by emulating ways people form mental images. Text-to-image technology enables computers to generate images based on textual input [2]. In this process, features are extracted from the input text, noise is removed, and then fed into a classifier [2]. This classifier relies on the competition between a discriminator and a generator to create images that closely resemble the originals. Although images have been produced from English text on a large scale, this method, which is based on multilingualism, is still in its early stages of development [2].

The second application is legal document classification. Corporations routinely spend millions of dollars to evaluate electronically stored documents in legal matters. To find relevant documents for these cases, lawyers use text categorization to sift through vast amounts of data effectively [4]. Although text classification is commonly utilized to reduce discovery costs in judicial proceedings, it faces a perception issue: lawyers often view this technology as a black box. In other words, they lack understanding of why certain documents are categorized as responsive. Recently, explainable ML (XML) has emerged as a significant area of research. In an XML system, the predictions or decisions of an ML model are presented in a way that humans can understand [4]. This transparency is significant for attorneys, allowing them to assess the model's document classification choices. If the reasons behind the categorization of documents are made clear, it becomes crucial in the context of XAI [4].

The third application is automatic news article classification, which is essential for helping readers quickly find relevant information [7]. Indic scripts, particularly Devanagari used in Hindi, Sanskrit, and Marathi, are complex compared to other languages, and many online datasets are imbalanced regarding class label distribution. The fourth application involves text categorization in sentiment analysis. The significant increase in user-generated content, primarily in text from social media posts, has been driven by the rapid growth and popularity of platforms like Twitter [9]. However, this data presents challenges due to its diversity and strong multilingual aspects, making it difficult to analyze and extract useful information [9]. This highlights the growing need for multilingual and domain-agnostic solutions.

This study discusses the classification of multilingual documents utilizing XAI and ML models. It outlines research questions concerning utilizing XAI and ML models for document classification and using fuzzy logic and convolutional neural networks (CNN) to optimize results. The paper discusses various existing models and compares their methodologies, providing analytical results based on existing research. Recent developments in fuzzy logic systems and artificial neural networks (ANN), combined with XAI approaches, have shown promise in improving classification accuracy while ensuring interpretability for human users.

This paper is organized as follows: The next section presents a set of research questions. Following that, we discuss related work in multilingual document classification using XAI and ML, comparing the existing research in these areas. The subsequent section provides analytical results highlighting the year-by-year contributions of XAI- and ML-based models for multilingual document classification. Finally, the paper concludes with a discussion on future directions.

Research Questions

- How can multilingual document classification systems be more accurate and interpretable using improved fuzzy logic and neural networks?
- Which approaches work best for incorporating XAI methods into multilingual document classification to maintain openness and human-centered decision-making?
- How much do optimization techniques like feature selection and hyperparameter tuning affect the suggested hybrid framework's computing efficiency, scalability, and performance?

Related work

XAI features, XAI themes, and several approaches to XAI system design and development were covered in the evaluation [1]. It also covered the objectives, methodologies, and assessment of XAI, significant turning points in XAI research, XAI for cybersecurity and security, difficulties, and metrics for assessing the resilience of XAI systems. Another study proposed a multilingual text-to-image model that can produce webtoon images when given multilingual input content [2]. It trained a generative adversarial network alongside the photos and used multilingual Bidirectional Encoder Representations from Transformers (BERT) to extract feature vectors for several languages. After training, when given multilingual input text, it produced graphics that resembled the originals.

Another work [3] proposed a multi-modal multilingual model to classify the document images of WIKI-DOC and MULTIEURLEX-DOC multilingual datasets. Another study [4] found justifications for explainable text classification methods in legal document review without using human-annotated training text samples. During a legal document examination, it was found that manually annotating training text excerpts was often impractical. However, the cost of reviewing huge documents might be significantly reduced with even modest improvements in precision.

The review [5] provided a systematic literature review of Recent developments and future trends for XAI approaches with different applications and observed many research works describing XAI development and applications in healthcare, manufacturing, transportation, and finance between January 2018 and October 2022. Those existing research works were retrieved from scholarly databases using pertinent keyword searches. Another study [6] detected and verified the multilingual text from a complex scene. It proposed a ML cum Deep Learning (DL) based solution irrespective of font size, as well as various background gradients applied on human machine interface.

The XAI-based multiclass news classification system [7] increased its performance and made it robust to handle imbalanced data. It trained the BERT with 9 classes of Bengali news, achieved 92% accuracy, and explained the model using Integrated Gradient. Another approach [8] looked at the Glance and Focus Network (GFNet) using the new local interpretable model-agnostic explanations and picture segmentation to simplify classification. It used the Integrated Gradient and the CARE framework to animate attentional CNN through transformations. It used CNN to visualize the regions and identify items within photos.

Another study [9] compared multilingual approaches for classifying the sentiment and the text of an examined multilingual corpus. It utilized four multilingual BERT-based classifiers and a zero-shot classification approach. A different study [10] investigated the DL application, which was built on the BERT and optimized for classifying goods and services provided in the Czech language descriptions of public procurement bids. By analyzing situations in which the model was either pretrained on English data and fine-tuned on automatically translated target descriptions or trained on data in multiple languages and fine-tuned for a target language, it compared various methods for pretraining and fine-tuning in a multilingual setup.

Table 1. Comparing existing XAI- and ML-based multilingual document classification research works

Ref. No.	Year	Problem-Focused
[1]	2022	Present situation, obstacles, and views on recent developments in reliable XAI
[2]	2022	Multilingual text-to-image models for webtoon generation
[3]	2023	Multilingual, multi-modal standard for classifying document images
[4]	2023	Methods for classifying explainable text in legal document reviews without using human-annotated training text excerpts
[5]	2023	Analysis of concepts and applications of XAI for current events and emerging patterns
[6]	2023	Analysis of multilingual text detection and verification from complex scenes
[7]	2023	XAI-based Bengali multiclass news classification
[8]	2023	XAI technique for classifying images and providing prediction interpretations
[9]	2023	Sentiment analysis and multilingual text classification for Twitter data
[10]	2025	DL techniques for classifying business documents in many languages

Analytical Results

Figure 1 compares various existing multilingual document classification models, specifically focusing on those utilizing XAI over different years. The analysis reveals a steady development of models and several review articles published in recent years. According to Figure 1, contributions are distributed as follows: 20% of the work was produced in 2022, 70% in 2023, and 10% in 2025. Additionally, research involving XAI is represented by 10% in 2022, 30% in 2023, and 10% in 2025.

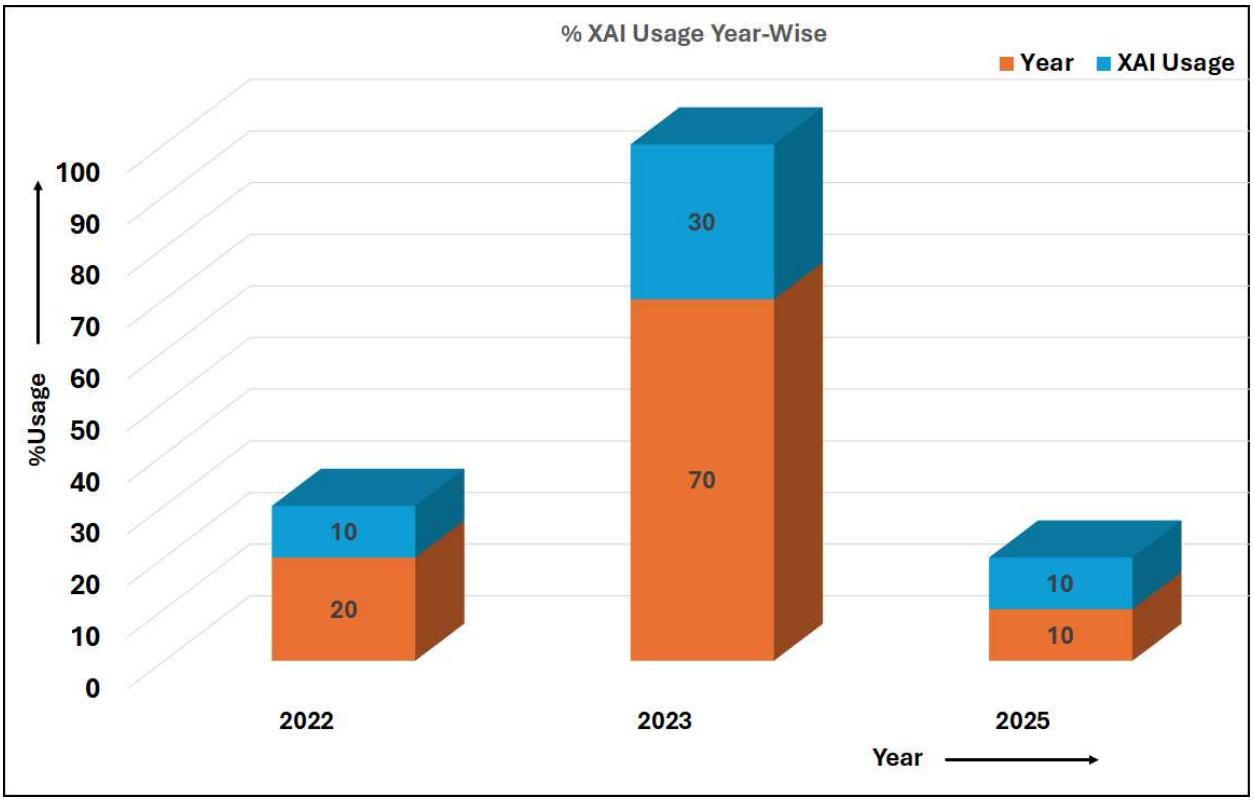

Figure 1. Comparing existing models based on the years and XAI technologies used.

Conclusions and Future Dimensions

This study discussed how XAI and ML techniques can simulate human intelligence to classify multilingual documents using various classifiers and methods. AI technologies, ML, and DL utilize algorithms to make more accurate predictions without human intervention. However, this accuracy can be achieved by applying cumulative model complexity and the opaque nature of black box models. Furthermore, the study presented analytical findings based on previous research while examining several currently used models and comparing their approaches. Recent advancements in fuzzy logic systems and ANN combined with XAI techniques have shown promise in improving classification accuracy while maintaining interpretability for human users. It is noted that most of the relevant work was proposed in 2023 using XAI techniques. This research can be further extended by conducting a more comprehensive review over the years and implementing work with XAI-based techniques.

References

1. A. Rawal, J. McCoy, D. B. Rawat, B. M. Sadler and R. S. Amant, "Recent Advances in Trustworthy Explainable Artificial Intelligence: Status, Challenges, and Perspectives", *IEEE Transactions on Artificial Intelligence*, vol. 3, issue 6, pp. 852-866, 2022. https://doi.org/10.1109/TAI.2021.3133846

2. K. Yu, H. Kim, J. Kim, C. Chun and P. Kim, “A study on generating webtoons using multilingual text-to-image models”, *Applied Sciences*, vol. 13, issue 12, pp. 1-13, 2022. https://doi.org/10.3390/app13127278
3. Y. Fujinuma, S. Varia, N. Sankaran, S. Appalaraju, B. Min and Y. Vyas, “A multi-modal multilingual benchmark for document image classification”, *Association for Computational Linguistics: EMNLP 2023,* pp. 14361–14376, 2023. https://aclanthology.org/2023.findings-emnlp.958/
4. C. Mahoney, P. Gronvall and J. Zhang, “Explainable text classification techniques in legal document review: Locating rationales without using human annotated training text snippets”, IEEE International Conference on Big Data (Big Data), pp. 2044-2051, 2023. https://doi.org/10.1109/BigData55660.2022.10020626
5. A. Saranya and R. Subhashini, “A systematic review of Explainable Artificial Intelligence models and applications: Recent developments and future trends”, Decision Analytics Journal, vol. 7, pp. 1-14, 2023. https://doi.org/10.1016/j.dajour.2023.100230
6. C. H. Patil, R. Zope and M. Jabde, "Comparative study of multilingual text detection and verification from complex scene”, 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, pp. 903-910, 2023. doi: https://doi.org/10.1109/ICAAIC56838.2023.10141373
7. M. F. Sikder, M. Ferdous, S. Afroz, U. Podder, K. Fatema, M. N. Hossain, M. T. Hasan and M. K. Baowaly, “Explainable Bengali multiclass news classification”, 26th International Conference on Computer and Information Technology (ICCIT), Cox's Bazar, Bangladesh, pp. 1-6, 2023. https://doi.org/10.1109/ICCIT60459.2023.10441218
8. D. Furman, M. Mach, D. Vranay and P. Sinčák, “Exploring interpretable XAI algorithms for image classification and prediction explanations”, World Symposium on Digital Intelligence for Systems and Machines (DISA), Košice, Slovakia, pp. 66-71, 2023. https://doi.org/10.1109/DISA59116.2023.10308941
9. G. Manias, A. Mavrogiorgou, A. Kiourtis, C. Symvoulidis and D. Kyriazis, “Multilingual text categorization and sentiment analysis: A comparative analysis of the utilization of multilingual approaches for classifying Twitter data”, Neural Comput & Applic, vol. 35, pp. 21415–21431, 2023. https://doi.org/10.1007/s00521-023-08629-3
10. P. Bednár, J. I. Vanko and E. Žiak, “Deep learning methods for multilingual classification of business documents”, IEEE 23rd World Symposium on Applied Machine Intelligence and Informatics (SAMI), Stará Lesná, Slovakia, pp. 000327-000330, 2025. https://doi.org/10.1109/SAMI63904.2025.10883330