# A Novel Approach to Multilingual Text Document Classification using Fuzzy Logic and XAI

*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:** The fast expansion of worldwide digital content and language-diverse information systems has made multilingual document classification more crucial. Deep learning techniques achieve extremely high classification accuracy, but their opaque nature restricts transparency and trust, especially in critical regulated applications. To classify multilingual documents, this study suggests an explainable fuzzy-based framework that combines explainable artificial intelligence, fuzzy logic, and semantic embeddings. Contextual embeddings are extracted to capture cross-lingual semantics after multilingual texts have been collected and prepared. Interpretable feature representation and reasoning are made possible by modeling linguistic uncertainty and ambiguity using fuzzy logic and membership functions. It incorporates explainability by using both post-hoc analysis and intrinsic fuzzy rules, providing understandable explanations for categorization outcomes. Model performance is improved while maintaining interpretability using an iterative optimization and evaluation loop. The proposed framework aims to balance accuracy, robustness, and transparency, making it suitable for real-world multilingual high-stakes applications.

**Keywords**: Multilingual Document Classification; Fuzzy Logic; XAI; Semantic Embeddings; Text Classification; Interpretability; Sustainable Learning.

---

## Introduction

Due to numerous social, economic, and cultural factors, multilingual picture document categorization has emerged as a research area in the current era. By helping international corporations streamline processes, including marketing and documentation, across linguistically disparate regions, it promotes economic globalization. These days, explainable artificial intelligence (XAI) aims to address issues in a way that people can comprehend. Nevertheless, just a few strategies have been put forth thus far, and several of them prioritize interpretability over explainability. Due to their rapid expansion, machine learning (ML) and artificial intelligence (AI) have been widely embraced in a range of application industries in recent years. But most contemporary AI systems rely on complex black-box models with opaque decision-making, which significantly restricts their use in regulated and safety-critical areas like banking, healthcare, and automotive systems [4].

Explainable Artificial Intelligence (XAI), which attempts to enable AI systems to provide human-understandable rationale for their predictions, has emerged as a response to this constraint [1]. Even interpretable models may become opaque when handling high-dimensional or multilingual inputs [1] [4]. Despite the fact that several interpretable and explainable systems have been presented, many prioritize interpretability over actual explainability. This problem is further highlighted by recent developments in developing neuro-fuzzy systems, since fuzzy rule structures' lack of transparency continues to erode confidence and dependability in practical applications [12] [13].

This study discusses a novel technique to classify multilingual text documents using fuzzy logic and XAI. This paper is organized as follows: The next section presents the background for multilingual document classification using fuzzy logic and XAI, comparing the existing research in these areas. It distinguishes the year-by-year contributions, their classification techniques, performance results, and use of XAI. The subsequent section proposes a novel approach to classify the multilingual documents using fuzzy logic and XAI. The final section concludes the paper with future directions.

**The State-Of-The-Art**

Table 1 compares various parameters of existing multilingual text classification methods, including classification techniques, accuracy, and XAI usage. The work [1] learned the fuzzy relations and fuzzy properties, extracted frequent relations, classified the toy dataset, and generated explanations for the decisions. Another work [2] applied Support Vector Machine (SVM) and fuzzy matching for the classification of Hindi printed characters and documents, respectively, into the pre-defined categories. It segmented images into lines and words by profiling and obtained Shirorekha Less (SL) isolated characters. It calculated the character confidence values, mapped them into Roman words, and classified the documents using fuzzy-based matching. It achieved SL character execution times of 0.22675 sec. and 0.20375 sec for training and testing, respectively. The work in [2] was extended in [3] to classify English printed and handwritten text images by following the steps of preprocessing, segmentation, feature extraction, SVM-based character classification, word association, and fuzzy-matching document classification. The subsequent work [4] combined DL with a type-2 fuzzy logic system for XAI to achieve a highly interpretable model.

*Table 1. Comparing multilingual text classification methods using fuzzy logic*

| Ref. No. | Year | Classification Techniques | Accuracy | XAI Use |
|---|---|---|---|---|
| [1] | 2018 | Fuzzy | - | ✓ |
| [2] | 2019 | SVM and Fuzzy | 74.61% (Train) and 80.73% (Test) | ✗ |
| [3] | 2020 | SVM and Fuzzy | 88.65% (Printed) and 74.97% (Handwritten) | ✗ |
| [4] | 2020 | DL + Type-2 Fuzzy | Low mean absolute errors | ✓ |

| | | | | |
|---|---|---|---|---|
| [5] | 2022 | TSK Fuzzy rules | High | XAI with regression |
| [6] | 2022 | DT, KNN, & Fuzzy Set | >94% with fuzzy sets | ✗ |
| [7] | 2023 | FuzzyWuzzy, Cosine Similarity & CNN | 98.6% with Dataset-III | ✗ |
| [8] | 2023 | Fuzzy + ANN + ACPSO | 97.38% | Interpretable rules |
| [9] | 2023 | DNN | High | ✗ |
| [10] | 2023 | BERT | 68% - 75% | ✗ |
| [11] | 2024 | DL + Fuzzy | Improved accuracy: 19% more with residual networks and 3% with ConvNeXt | ✓ |
| [12] | 2024 | ANN + Fuzzy | 98.04% | ✓ |
| [14] | 2025 | CNN & Fuzzy Logic | 97.1% (RVL-CDIP and MASK-RCNN) with 2.3 sec processing time. | ✗ |
| [15] | 2025 | Fuzzy rules, DeBERTa & LSTM | 93% | ✓ |

The federated learning for Takagi-Sugeno-Kang Fuzzy Rules (TSKFR) in [5] applied explainable fuzzy regression. Here, the independent data owner nodes generated their own local TSKFR, which were forwarded to a server to generate a global TSK-FRBS. The contribution [6] evaluated the performance of 43 similarity and 19 distance measures of text document classification using fuzzy sets. It worked on the BBC News and BBC Sports datasets and optimized the system using hyperparameter tuning and leave-one-out cross-validation. The fuzzy-based classification outperformed Decision Tree (DT) and K-Nearest Neighbor (KNN). Another study [7] detected multilingual text in complex scenes using FuzzyWuzzy, cosine similarity, and Convolutional Neural Networks (CNNs) in three datasets. The text had varying font sizes and background gradients. Furthermore, the approach [8] performed missing value imputation using k-means clustering, selected features, classified using fuzzy logic and Alexnet neural networks, and optimized the network parameters using hybrid Ant Colony Particle Swarm Optimization (ACPSO).

Furthermore, the research work [9] classified multilingual image documents using a Deep Neural Network (DNN) by integrating five new activation functions. Similarly, the work [10] categorized the multilingual text and analyzed semantics using four multilingual BERT-based classifiers and a zero-shot classification approach for Twitter data, and the method [11] classified artistic images and used XAI to map known visual traits of an image to DL and fuzzy rule features. Another work [12] applied an interpretable neuro-fuzzy learner for data streams by integrating Local Interpretable Model-agnostic Explanations (LIME) to provide local explanations and evaluated them using faithfulness and monotonicity metrics. The dynamic stream dataset included diverse concept drifts, such as abrupt, gradual, recurring, contextual, and cyclical drifts.

Another study [13] discussed the multilingual document classification using XAI. It reviewed several recent models to contrast their approaches and analyzed the findings. The contribution [14] classified real-time on-device documents by combining visual and textual features using OCR and FuzzyWuzzy, and classified them using CNN and a voting method. It standardized the inputs by converting documents to images, extracted text using EasyOCR, used TF-IDF vectors with CNN, and applied FuzzyWuzzy to match text with class keywords. Another method [15] applied fuzzy rule mining, Decoding-enhanced BERT with disentangled attention (DeBERTa), metadata features, and Long Short-Term Memory (LSTM) technique for complex prediction of English Language proficiency levels, including unstructured and structured data streams. Additionally, it used XAI methods like SHAP and DeepSHAP and statistical significance tests to provide explanations.

**The Proposed Model**

The proposed XAI-based multilingual text document classification system performs a sequence of steps using fuzzy logic and membership. It integrates fuzzy reasoning, explanation-driven optimization, and semantic embeddings, as shown in Figure 1. It acquires and preprocesses multilingual text data from many sources. It pre-processes the data including linguistic variances, tokenization, noise reduction, and language normalization. It then applies multilingual embedding techniques to extract semantic features, allowing text to be transformed into dense vector representations that capture contextual and cross-lingual semantics. It feeds these embeddings into a fuzzy logic layer, which models the language ambiguity and uncertainty present in multilingual text by applying fuzzification using clearly defined membership functions.

It then analyses features to determine the most discriminative and comprehensible traits, creating a clear basis for categorization and subsequent reasoning. Furthermore, it integrates XAI by combining inherent fuzzy-rule-based explanations with post-hoc interpretability techniques to increase trust and transparency. This makes it easier to comprehend fuzzy rule activations and feature contributions by enabling the algorithm to generate comprehensible explanations for categorization outcomes.

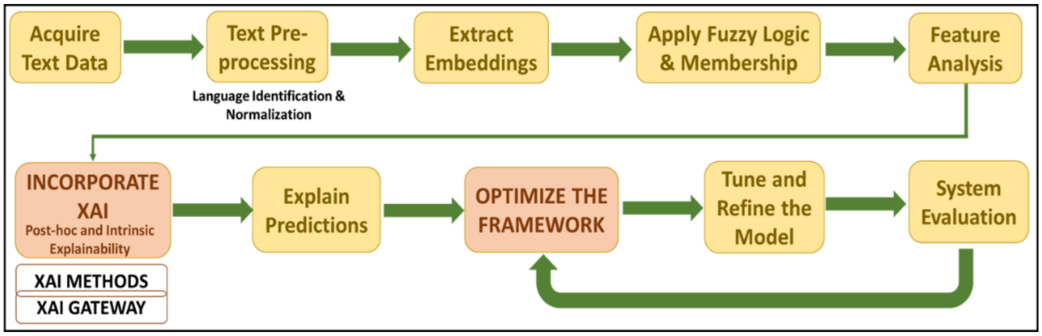


*Figure 1. Proposed XAI and fuzzy-based architecture for multilingual document classification.*

Using the given explanations as a guide for framework optimization, fuzzy memberships, rule structures, and the model are iteratively improved. After fine-tuning the model to balance explainability and classification performance, the proposed system is rigorously assessed using common performance indicators. The methodology is appropriate for real-world multilingual and high-stakes applications since assessment feedback is incorporated into the optimization loop, guaranteeing ongoing improvements in both predicted accuracy and interpretability.

**Conclusions and Future Directions**

The scientific body of knowledge created in the pursuit of techniques that elucidate the internal logic of a learning algorithm, a model generated from data, or a knowledge-based strategy for inference is collectively referred to as XAI. It is a fundamental aspect of AI. As such several contributions have been proposed in recent years for multilingual text recognition were proposed. From data collection to optimization and tuning, a series of procedures was followed in the suggested multilingual document text categorization utilizing fuzzy and XAI. Future research may extend this paradigm to multimodal data by integrating text and visual elements, exploring adaptive fuzzy rule learning for streaming data, and evaluating explanation quality using human-centered metrics. Additionally, deploying the system in real-time industrial settings and federated learning environments offers promising opportunities for scalable privacy-aware multilingual intelligence.

**References**

1. R. Pierrard, J. -P. Poli and C. Hudelot, "Learning Fuzzy relations and properties for explainable artificial intelligence," IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Rio de Janeiro, Brazil, pp. 1-8, 2018. https://doi.org/10.1109/FUZZ-IEEE.2018.8491538
2. S. Puri and S. P. Singh, "A Hybrid Hindi Printed Document Classification System Using SVM and Fuzzy: An Advancement", *Journal of Information Technology Research*, vol. 12, issue 4, pp. 107-131, 2019. https://doi.org/10.4018/JITR.2019100106
3. S. Puri and S. P. Singh, "A Fuzzy Matching based Image Classification System for Printed and Handwritten Text Documents", *Journal of Information Technology Research*, vol. 13, issue 2, pp. 155-194, 2020. https://doi.org/10.4018/JITR.2020040110

4. R. Chimatapu, H. Hagras, M. Kern and G. Owusu, “Hybrid deep learning type-2 fuzzy logic systems for explainable AI,” IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Glasgow, UK, pp. 1-6, 2020. https://doi.org/10.1109/FUZZ48607.2020.9177817
5. J. L. Corcuera Bárcena, P. Ducange, A. Ercolani, F. Marcelloni and A. Renda, “An approach to federated learning of explainable fuzzy regression models,” IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Padua, Italy, pp. 1-8, 2022. https://doi.org/10.1109/FUZZ-IEEE55066.2022.9882881
6. G. K. Sidiropoulos, N. Diamianos, K. D. Apostolidis and G. A. Papakostas, "Text Classification Using Intuitionistic Fuzzy Set Measures—An Evaluation Study," *Information*, vol. 13, issue 5, pp. 1-21, 2022. https://doi.org/10.3390/info13050235
7. 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
8. P. Patro, K. Kumar, G. S. Kumar and A. K. Sahu, “Intelligent Data Classification Using Optimized Fuzzy Neural Network and Improved Cuckoo Search Optimization,” *Iranian Journal of Fuzzy Systems*, vol. 20, issue 6, pp. 155-169, 2023. https://doi.org/10.22111/ijfs.2023.44767.7887
9. S. Banerjee and D. Shende, “MLing-Net: A computationally inexpensive deep neural framework designed to perform multilingual image document classification”, 7th International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech), Kolkata, India, pp. 1-6, 2023. https://doi.org/10.1109/IEMENTech60402.2023.10423510
10. 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
11. J. Fumanal-Idocin, J. Andreu-Perez, O. Cordón, H. Hagras and H. Bustince, "ARTxAI: Explainable Artificial Intelligence Curates Deep Representation Learning for Artistic Images Using Fuzzy Techniques," *IEEE Transactions on Fuzzy Systems*, vol. 32, no. 4, pp. 1915-1926, 2024. https://doi.org/10.1109/TFUZZ.2023.3337878
12. M. M. Ferdaus, T. Dam, S. Alam and D. -T. Pham, "X-Fuzz: An Evolving and Interpretable Neuro-Fuzzy Learner for Data Streams," *IEEE Transactions on Artificial Intelligence*, vol. 5, no. 8, pp. 4001-4012, 2024. https://doi.org/10.1109/TAI.2024.3363116
13. S. Puri, M. Janarthanan and G. Khekare, “Multilingual document classification using XAI: a review,” International Conference LGPR, SGS - Engineering & Sciences, vol. 1, issue 2, pp. 1-4, 2025. https://spast.org/techrep/article/view/5397
14. S. M. Satapathy, A. Victor, K. K. Gounder, V. Agrawal, R. S. Panchal and A. K. Mahale, “A lightweight hybrid CNN-fuzzy logic approach for real time on-device document classification,” Proceedings of the 18th Innovations in Software Engineering Conference, Association for Computing Machinery, New York, USA, pp. 1-11, 2025. https://doi.org/10.1145/3717383.3717387
15. X. Zhao X, “A Hybrid Deep Learning and Fuzzy Logic Framework for Feature-Based Evaluation of English Language Learners,” *Scientific Reports*, vol. 15, issue 1, pp. 1-40, 2025. https://doi.org/10.1038/s41598-025-17738-z