

Fuzzy C-Means Based Soft Computing Model for Early Detection of Acute Respiratory Diseases from Confusable Symptom Sets

Chukwu, Emenike Gabriel¹, Ezema, Modester E.², Achi, Ikenna Kalu³, Mbagwu, Amarachi Austina³, Okereke, George Emeka²

¹Department of Software Engineering, Federal University of Technology Ikot-Abasi, Nigeria.

²Department of Computer Science, University of Nigeria Nsukka, Nigeria.

³Department of Computer Science, Michael Okpara University of Agriculture, Umudike.

DOI: <https://doi.org/10.5281/zenodo.23395754>

Received: 22-09-2026

Accepted: 05-10-2026

Published: 15-10-2026

ABSTRACT

Acute Respiratory Diseases (ARDs), including pneumonia, influenza, Severe Acute Respiratory Syndrome (SARS), Acute Respiratory Distress Syndrome (ARDS), and COVID-19, share overlapping symptoms that often lead to diagnostic confusion and delayed treatment. This study aimed to develop a Fuzzy C-Means (FCM)-based soft computing model for the early detection and classification of ARDs from confusable diseases. Specifically, it sought to create a symptom dataset, apply FCM clustering and classification techniques, develop an FCM-based diagnostic system, and compare its performance with Rule-Based Conventional Clinical Diagnosis (RCCD) and Symptom-Based Conventional Clinical Diagnosis (SCCD) systems. Using an analytical approach, symptom data comprising 20,373 instances and five attributes were collected from clinicians and medical archives. FCM was used for clustering, Random Forest for classification, and a Python-MySQL system was developed using Rapid Application Development. The FCM model achieved strong clustering metrics (PC=0.82, PE=0.19, ARI=0.78, NMI=0.81), while the classifier attained 90.3% accuracy and 0.95 AUC. The developed system outperformed RCCD and SCCD methods, demonstrating improved ARD detection and classification.

Keywords: Fuzzy C-Means, Acute Respiratory Diseases, Soft Computing, Medical Diagnosis, Clustering, Decision Support System

Subject Area

Artificial Intelligence, Soft Computing, Machine Learning, and Medical Informatics, with a focus on intelligent systems for healthcare diagnosis.

I. INTRODUCTION

Acute Respiratory Diseases (ARDs), including pneumonia, influenza, COVID-19, Severe Acute Respiratory Syndrome (SARS), and Acute Respiratory Distress Syndrome (ARDS), remain major causes of morbidity and mortality worldwide. These diseases often present with similar symptoms such as fever, cough, dyspnea, chest discomfort, and hypoxemia, making accurate diagnosis difficult, particularly during the early stages of illness. The overlap of clinical manifestations among respiratory diseases frequently leads to diagnostic uncertainty, delayed treatment, and poor patient outcomes.

Conventional diagnostic methods, including imaging and laboratory investigations, are valuable but are often resource-intensive and may not effectively address symptom-level ambiguity. Consequently, Artificial Intelligence (AI) and data-driven approaches have gained increasing attention for improving disease detection and clinical decision-making. According to Esteva et al. (2019), AI technologies can assist healthcare professionals

by identifying complex disease patterns that may not be readily apparent through traditional diagnostic procedures.

Recent studies have demonstrated the effectiveness of machine learning and deep learning in respiratory disease diagnosis. Kai-Chih et al. (2022) developed an ensemble CNN-based model for ARDS detection with high diagnostic accuracy, while Jaiswal et al. (2019) applied Mask R-CNN for pneumonia detection. Despite their promising performance, most existing approaches rely heavily on medical imaging, require sophisticated infrastructure, and often focus on single-disease detection rather than distinguishing among multiple confusable respiratory conditions.

To address uncertainty and overlapping symptom patterns, soft computing techniques have emerged as viable alternatives. The Fuzzy C-Means (FCM) clustering algorithm, developed by Bezdek (1981), enables data instances to belong to multiple classes with varying degrees of membership, thereby providing a more realistic representation of clinical uncertainty. The applicability of FCM in medical

diagnosis has been demonstrated by El-Ibrahimi et al. (2024) in chronic kidney disease prediction and by Haris et al. (2023) in patient classification based on symptom severity. Furthermore, Aniobi et al. (2020) highlighted the importance of intelligent approaches for diagnosing diseases with overlapping symptoms.

However, significant research gaps remain. Haruna et al. (2020) observed that most AI-driven studies during the COVID-19 era concentrated on imaging-based diagnosis, with limited attention given to symptom-based differential classification. In addition, recent clinical studies by Beitler et al. (2021) and Ellen A. et al. (2022) emphasized the heterogeneous nature of ARDS and the need for flexible, precision-oriented diagnostic frameworks capable of handling overlapping syndromes.

Motivated by these gaps, this study proposes a Fuzzy C-Means-based soft computing model for the early detection and classification of confusable Acute Respiratory Diseases using structured symptom datasets. By assigning membership degrees across disease categories, the proposed model effectively captures diagnostic uncertainty and supports more accurate early decision-making. The system is designed to provide an interpretable, scalable, and cost-effective solution suitable for both resource-constrained and advanced healthcare environments.

II. RELATED WORKS

A fuzzy cluster means approach was introduced by James et al. (2022) for the identification of confusable diseases. In their study, they contended that ambiguity of one kind or another has overtaken medical science in recent years, significantly impacting the decision-making process and resulting in instances of incorrect diagnoses and, in the worst circumstances fatalities.

Using the fuzzy cognitive map engine, Faith-Michael et al. (2021) provided a technique for early differential diagnosis of tropical confusable illnesses. Access to medical care is a major problem in developing countries, particularly those in Africa, Asia, and the Middle East.

An early infectious illness diagnosis method based on complex probabilistic hesitant fuzzy N-soft information was proposed by Shahzaib et al. (2023). The purpose of this article was to evaluate and address the difficulties faced by healthcare providers who treat infectious infections.

Hayat et al., (2021), in their work leverages the smartphone sensors for the preliminary diagnosis of COVID-19. According to them, deep learning, a significant advancement in artificial intelligence over the last ten years, has enormous potential for identifying relevant and suitable aspects in healthcare.

The study "Soft Clustering for Enhancing the Diagnosis of Chronic Diseases over Machine Learning Algorithms" was conducted by Theyazn et al. in 2020. The goal of their research is to enhance the chronic disease monitoring detection system, which is utilized to save people's lives.

Using key clinical traits and symptoms, Pius and Victor (2021) developed a fuzzy expert system diagnostic model to help with early infection risk assessment. Since some of the symptoms are indicators of other diseases, the fuzzy membership functions were constructed using pertinent research findings to address the imprecision present in this area.

III. METHODOLOGY

Following a thorough examination of the elements listed in the preceding introduction, our study would employ Analytical and Empirical Research Methodology (AERM). This strategy is essential since identifying the benefits and drawbacks of the current system requires a full, appropriate, and comprehensive evaluation. To avoid giving the incorrect medication for such a sensitive topic involving human health, we need such a system.

3.1 Data Collection and Preprocessing

These are important stages in our research work involving data analysis, machine learning, artificial intelligence, or software system development. These stages ensure that accurate, relevant, and high-quality data are gathered and prepared for effective analysis, classification, prediction, or decision-making. In our healthcare-related studies, we have adopted data collection and preprocessing to improve the reliability, accuracy, and efficiency of the developed intelligent system.

We adopted the following methods in the cause of collecting data and make it better: Information Gathering: a casewe gathered past patient information at the hospital, such as symptoms, age, duration, and diagnosis labels, using the Document Reviews Information Gathering Approach. Additionally, we gathered samples of characteristics including age, gender, length of sickness, and binary symptoms like fever or cough and Data Sources: in which case we visited many hospitals and spoke with medical lab technologists and virologists. We also looked through the hospital's records of cases of acute respiratory illness. For documentation purposes, the primary sources of the data we used for our research are given below: Records of Clinical Symptoms, Published Datasets and Medical Literature, Domain Experts, and Data Pre-processing. Sample outcome or result of the data collection and preprocessing is captured below.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Sex	Age	Duration	Cough	Night_Fever	Catarrh	Cold	Tenderness	Headache	Jerking	Movement	Throat_Fever	Fast_Breathing	Body_Rashes	Cough_Sills	Symptoms		
2	M	13	7	1	1	0	0	1	1	0	0	0	0	0	0	0	MALARIA	
3	M	13	7	1	1	0	0	1	1	0	0	0	0	0	0	0	URT	
4	F	4	14	0	0	0	0	0	0	0	0	0	0	0	0	0	HEAMORRHAGIC DISEASE	
5	F	4	14	0	0	0	0	0	0	0	0	0	0	0	0	0	PERINATAL ASPHYXIA	
6	F	5	28	0	0	0	0	0	0	1	1	0	0	0	0	0	FERNATAL ASPHYXIA	
7	F	5	28	0	0	0	0	0	0	1	1	0	0	0	0	0	NNS	
8	M	5	7	1	1	0	0	0	0	0	0	0	1	0	0	0	MACROSOMU BABY	
9	M	5	7	1	1	0	0	0	0	0	0	0	1	0	0	0	RVD-RETRO VIRUS DISEASE	
10	M	10	14	0	1	0	0	0	0	0	0	0	0	1	0	0	GET Sepsis	
11	F	12	21	0	1	0	0	0	0	0	0	0	0	0	0	1	SEVERE MALARIA	
12	F	12	21	0	1	0	0	0	0	0	0	0	0	0	0	1	HYPERPYREXIA	
13	M	17	14	0	0	0	0	0	0	0	0	0	0	0	0	0	MALARIA	
14	M	17	14	0	0	0	0	0	0	0	0	0	0	0	0	0	URT	
15	M	21	14	1	0	0	0	0	0	0	0	0	0	0	0	0	ACUTE URT	
16	M	21	14	1	0	0	0	0	0	0	0	0	0	0	0	0	PNEUMONIA	
17	F	12	49	1	1	0	0	0	0	0	0	0	0	0	0	0	MALARIA	
18	F	12	49	1	1	0	0	0	0	0	0	0	0	0	0	0	BPN	
19	F	12	21	0	1	0	0	0	0	0	0	0	0	0	0	0	DIARRHEA	
20	F	12	21	0	1	0	0	0	0	0	0	0	0	0	0	0	DEHYDRATION	
21	F	12	21	0	1	0	0	0	0	0	0	0	0	0	0	0	MALARIA	

Figure 1: Sample Data Collected

3.2 Fuzzy C-Means Clustering Engine

This is the basis of our soft computing methodology. Uses the Fuzzy C-Means technique to cluster symptom data according to different Acute Respiratory Diseases (ARDs). Because FCM allows overlapping memberships, it reflects real-world symptom ambiguity, where a symptom may partially belong to different disorders. This also

captures both diagnostic uncertainty and soft cluster membership. The reasoning for this was because FCM allows a patient to have different levels of membership (μ -values) in many sickness clusters. This is ideal for ARDs, since symptoms often overlap and the diagnosis is not binary. Supports differential diagnosis and "gray zones" in the real world.



Figure 2: TheArchitectureofthe Existing System

IV. ANALYSIS OF THE EXISTING SYSTEM

The rule-based or existing system for diagnosing Acute Respiratory Diseases (ARDs) largely depends on traditional, manual approaches carried

out in clinical settings. The diagnostic process is often sequential and involves the following components: Patient Input, Manual Physician Diagnostic, Laboratory Tests (if available), Challenges and Limitations, and Diagnostic Output

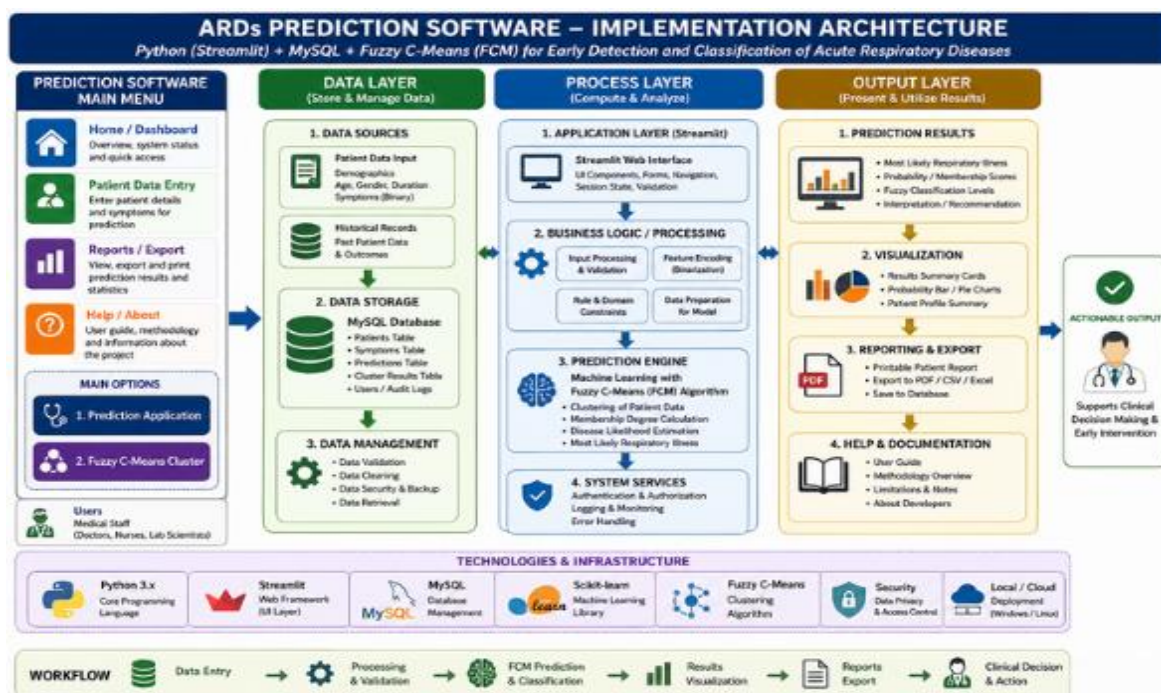


Figure 3: The Architecture of the Proposed System (ARDs Prediction Software System)

The goal of the suggested Fuzzy Cluster Means-Based Soft Computing System with Fuzzy Cognitive Map is to facilitate the early identification and differentiation of Acute Respiratory Diseases (ARDs) from their confusing counterparts. The design incorporates a number of linked modules, each of which is in charge of handling, processing, making decisions, and interpreting the results. Figure above provides comprehensive details regarding the system and it is of the following components: Prediction Software Main Menu where the user selects what operation to carry on, Data Layer used to store and manage data, Process Layer used to compute and analyses data, and Output Layer which is used to present and utilize result.

V. RESULTS & EVALUATION

Here, we examined the importance, significance, and relevance of our findings. It focuses on summarizing and evaluating what we found, showing how it relates to the subject of our paper

or dissertation and literature review, and making a case for our final conclusion.

5.1 Evaluation Metrics

Fuzzy C-Means was evaluated using the following standard classification metrics: Partition Coefficient (PC) for assessing the degree of fuzziness in a fuzzy clustering result, Partition Entropy (PE) for assessing fuzzy cluster assignment uncertainty, The Dunn Index (DI) for considering both cluster separation and compactness when assessing the quality of clustering, Davies-Bouldin Index (DBI) for evaluating the degree of cluster separation and compactness, The Adjusted Rand Index (ARI) as an external validation parameter that assesses the degree of agreement between the predicted cluster labels and the actual class labels and Normalized Mutual Information (NMI) for assessing the degree of agreement between the actual class labels and the anticipated cluster labels.

Table 1: Comparative Results Evaluation

Performance Metric	FCM-Based System	RCCD System	SCCD System
Accuracy	50%	5%	5%
Precision	39%	30%	30%
Recall	49%	28%	26%
F1-Score	39%	27%	27%
Support	61%	61%	61%

5.2 Overall Comparative Performance Ranking of the Diagnostic Systems

```

... Logistic Regression
Model performance for Training set
- Accuracy Score: 0.5000
- Classification Report:
precision    recall  f1-score   support

0           0.00      0.00      0.00         1
1           0.00      0.00      0.00         3
3           0.00      0.00      0.00         1
4           0.00      0.00      0.00         2
5           0.00      0.00      0.00         1
6           0.00      0.00      0.00         1
8           0.00      0.00      0.00         2
9           0.00      0.00      0.00         2
10          0.00      0.00      0.00         1
11          0.00      0.00      0.00         2
13          0.00      0.00      0.00         2
14          0.00      0.00      0.00         1
15          0.00      0.00      0.00         2
16          0.00      0.00      0.00         2
17          0.00      0.00      0.00         1
18          0.00      0.00      0.00         1
19          0.00      0.00      0.00         1
20          0.00      0.00      0.00         2
21          1.00      1.00      1.00         2
...

```

Figure 5: Model Performance for Training Set of the Existing System

```

Generate + Code + Markdown | Run All ...
... Accuracy Score: 49.18 %
Classification Report:
precision    recall  f1-score   support

0           1.00      1.00      1.00         1
1           1.00      1.00      1.00         1
2           0.00      0.00      0.00         2
3           0.00      0.00      0.00         1
6           1.00      1.00      1.00         1
7           0.00      0.00      0.00         1
10          0.00      0.00      0.00         1
11          0.00      0.00      0.00         2
12          0.00      0.00      0.00         1
14          0.00      0.00      0.00         0
18          0.00      0.00      0.00         1
19          0.00      0.00      0.00         0
20          0.00      0.00      0.00         0
24          0.50      1.00      0.67         1
25          0.33      0.11      0.17         9
28          0.00      0.00      0.00         1
29          1.00      1.00      1.00         1
32          0.00      0.00      0.00         1
37          1.00      1.00      1.00         1
38          0.00      0.00      0.00         1
40          0.00      0.00      0.00         1

...
accuracy                0.49         61
macro avg              0.30         61
weighted avg           0.39         61

```

Figure 6: Model Performance for Training Set of the New System

In summary, the FCM-based system greatly outperforms RCCD and SCCD systems across all main performance measures, as confirmed by the experimental comparison using Scikit-learn 1.1.1.

VI. CONCLUSION

We created a model that uses the Fuzzy Cluster Means Algorithm to diagnose Acute Respiratory Diseases (ARD). The work was divided into two main sections. In the first, a medical professional who is unsure about diagnosing a patient using

laboratory procedures but has overlapping symptoms that make the doctor unsure and indecisive in identifying the specific diseases based on presented symptoms was asked pertinent questions using the concept of SARS. This information was displayed in a membership function format that shows the doctor's assurance (truthfulness) that a disease exists as well as the uncertainty and inaccuracy of the diagnosis. In order to identify the precise illness, the second step was to send the knowledge base data set to the inference engine using the Multi-Criteria Decision Making (MCDM) technique. We closely examined the current systems and identified a few constraints to take into account. Finally, the suggested framework included an interface that allowed the system to record the patient's symptoms, calculate the confusability measure, and access the knowledge base. After diagnosing the patient, the inference process accurately pinpoints the specific illness within the disease set.

REFERENCES

- [1]. Aniobi, D. E., et al, (2020) Development of an Expert System Framework for the Diagnosis of Confusable Diseases Using Neutrosophic-Ahp Technique, FULafia Journal of Science & Technology, ISSN: 24490954, Vol. 6 No.1 March 2020.
 - [2]. Beitler, J. R., Malhotra, A., & Thompson, B. T. (2021). Precision medicine for acute respiratory distress syndrome. *The Lancet Respiratory Medicine*, 9(9), 995–1006.
 - [3]. Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*, Springer Science & Business Media Journal, ISBN: 978-1-4757-0452-5. <http://dx.doi.org/10.1007/978-1-4757-0450-1>
 - [4]. Ellen A. G., Cecilia M. O. and Daniel F. M. (2022), Acute Respiratory Distress Syndrome in Adults: Diagnosis, Outcomes, Long-Term Sequelae, and Management, the *Lancet journal of Respiratory Medicine*, Wellcome-Wolfson Institute for Experimental Medicine, Queen's University Belfast, Belfast BT9 7BL, UK. *Lancet* 2022; 400: 1157–70, [https://doi.org/10.1016/S0140-6736\(22\)01439-8](https://doi.org/10.1016/S0140-6736(22)01439-8).
 - [5]. El-Ibrahimi, A., Laghmati, S., Hicham, K., Hamida, S., & et al. (2024, May). Fuzzy logic based expert system for early predicting of chronic kidney disease. In *2024 4th International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*. IEEE. <https://doi.org/10.1109/IRASET60544.2024.10548858>
 - [6]. Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., ...& Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine*, 25(1), 24–29. <https://doi.org/10.1038/s41591-018-0316-z>
 - [7]. Faith-Michael, E. U. et al (2021), A framework for Early Differential Diagnostic of Tropical Confusable Diseases using the Fuzzy Cognitive Map Engine, Conference Paper · February 2021. <https://www.researchgate.net/publication/310141600>
 - [8]. Haris A., Muhammad A. A., Waleej H., et al (2023), A Fuzzy Clustering-based Approach for Classifying COVID-19 Patients by Age and Early Symptom Indicators, *LGU Research Journal of Computer Science*, SSN: 2521-0122, Vol. 7 Issue, doi:10.54692/lgurjcsit.2023.0702410
 - [9]. Haruna C., Absalom E. E., Fatsuma J. et al, Early Eurvey with Bibliometric Analysis on Machine Learning Approaches in Controlling Coronavirus, *PeerJ Computer Science Journal*, <https://doi.org/10.1101/2020.11.04.20225698>
 - [10]. Hayat K., Karim A., Abderrahim B. et al (2021), Deep Learning Based Early Detection Framework for Preliminary Diagnostic of COVID-19 via Onboard Smartphone Sensors, *MDPI Journal, Sensors* 2021, 21, 6853. <https://doi.org/10.3390/s21206853>
 - [11]. James, G.G., Okpako, A.E., and Ndunagu J.N., (2022) Fuzzy cluster means algorithm for the Diagnostic of confusable disease, *Journal of Computer Science and its application*, ISSN: 2006-5523, Vol. 23 No. 1 (2022). <https://www.ajol.info/index.php/jcsia/article/view/153911>
 - [12]. Jaiswal, A. K., Tiwari, P., Kumar, S., Gupta, D., Khanna, A., & Rodrigues, J. J. P. C. (2019). Identifying pneumonia in chest X-rays: A deep learning approach. *Measurement*, 145, 511–518.
 - [13]. Kai-Chih P., Wen-Cheng C., Yu-Len H., et al (2022), Artificial intelligence-aided diagnosis model for acute respiratory distress syndrome combining clinical data and chest radiographs, *Digital Health Journal*, Volume 8: 1–15, DOI:
-

- 10.1177/20552076221120317,
journals.sagepub.com/home/dhj.
- [14]. Pius E. U. and Ekong V. E., (2021), A Fuzzy Expert System Model for the Determination of Coronavirus Disease Risk, International Journal of Mechatronics, Electrical and Computer Technology (IJMEC, ISSN: 2305-0543 (Online), ISSN: 2411-6173 (Print), Vol. 11(39), Jan. 2021, PP. 4825-4831
- [15]. Shahzaib A., Muneeba K., Muhammad and Shazib H., Early infectious diseases identification based on complex probabilistic hesitant fuzzy N-soft information, Application of Soft Computing Journal, 27:18285–18310, <https://doi.org/10.1007/s00500-023-08083-2>
- [16]. Theyazn H. H. A., Ali S. A., and Mohammed Y. A., (2021), Soft Clustering for Enhancing the Diagnostic of Chronic Diseases over Machine Learning Algorithms, Journal of Healthcare Engineering, Volume 2020, Article ID 4984967, 16 pages, <https://doi.org/10.1155/2020/4984967>
-