INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

Assistant Tools for Medical Diagnostics through Rough Set-Based Data Analysis

Kamesh Kumar[1]*

1 Assistant Professor,, K.D. College Simbhaoli, Hapur, India

OPEN ACCESS

Received: 16-06-2024

Accepted: 21-07-2024

Published: 02-08-2024

Citation: Kumar K (2024) Assistant Tools for Medical Diagnostics through Rough Set-Based Data Analysis. Indian Journal of Science and Technology 17(31): 3174-3182. https://doi.org/10.17485/IJST/v17i31.1987

* **Corresponding author**.

fuzzmath@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Objective: This study aims to identify and prioritize critical symptoms of pneumonia, determining their relative importance. Based on these findings, a decision rule base is developed to enhance efficiency of pneumonia diagnosis. **Methods:** A disease may concern with a set of symptoms, also same set of symptoms may appear in different diseases. To make the diagnostic decision apparent, it is advantageous to identify and assigning extra importance to some critical symptoms. We applied the reduction of attributes and importance of attributes indices of rough set theory to characterize the critical symptoms or core symptoms. Thereafter, an algorithm has been proposed to set up assistant tools for medical diagnosis. **Findings:** The reduction of attributes and importance of attribute, indices are calculated for conditional attributes namely respiratory rate (RR), cough level (CO), chest in drawing (CI) and temperature (T) concerned to the decision attribute (disease) pneumonia. The findings reveal insights into identifying the critical symptoms and also its degree of likelihood of importance. Chest in drawing and temperature are found in the list of core symptoms. Also, degree of Importance for these symptoms are computed as 9/14 and 1/21, respectively. A rule base is optimized to assist the diagnostic process of disease pneumonia effectively. **Novelty:** This research contributes in presenting a novel mathematical algorithm that identifies the critical symptoms and irrelevant symptoms for pneumonia disease. This aids the diagnostic process by redefining the decision rule base. The numerical computation provides a practical and visual tool to assess potential outcomes of the proposed technique. **Applications:** The proposed idea can assist medical experts in better and robust diagnostics when quantity of symptoms is increased and linguistically expressed that creates non-specificity type incompleteness.

Keywords: Rough Set; Lower and Upper Approximations; Indiscernibility Relation; Reduction of Attribute; Importance of Attribute; Medical Diagnosis

1 Introduction

Rough set theory (RST) proposed by Pawlak[1], is motivated by its characteristics to handle with raw data comprising uncertainty, imprecision and incompleteness. It provides a systematic scientific procedure for feature selection, reduction of the dispensable data, makes decision-making processes easy, produces comprehensible results and classification of the techniques on behalf of their performance. It promotes knowledge search across a variety of areas. It has been utilized in different domains that include knowledge representation and expert decision systems. Also, it is considered as one of the first non-statistical approaches in data analysis [2,3].

The rapid evolution of rough-set theory has prompted the need for enhanced methodologies in medical diagnostics and healthcare also. Tsumoto examined the characteristics of experts and introduces a new approach to generate diagnostic rule [4]. Pawlak and Skowron presented some extensions of rough set based technique; also outline a challenge for the rough set based research [5] Tripathy et al.studied the rules generation algorithm of rough set theory. These suitable rules are explored to identify the chief characteristics affecting the relationship between heart disease and its attributes by using formal concept analysis. This helps the decision maker a priori detection of the heart disease [6] Vamsidhar et al. used the rough set theory technique to select the most relevant features, which helps to provide the efficient classification of medical data and disease detection. The proposed method, RST-RNN shows accuracy of 98.57% for heart disease dataset [7]. Kumar and Pandey improved a switching function and it was applied in medical decisions [8]. El-Bably et al. [9] enhanced soft rough sets from topological perspective and proposed topological soft rough sets. They suggested an algorithm of diagnosing heart failure disease. Kumar and Pandey [10] developed a soft linear programming model to set up waiting time targets in an outpatient department. Mishra and Acharjya [11] hybridized the rough set degree of dependency with the red dear optimization algorithm for feature selection and knowledge discovery. They employed the proposed technique for diagnosing hepatitis B. Their rough set red deer procedure helps in finding the prime features and outperforms with an accuracy of 91.7%. Singh and Mantri [12,13] developed another feature selection technique using rough set theory, rough fuzzy and machine learning. They examined and analyzed important features of hepatitis, dermatology conditions, hepatic disease, and autism with accuracy values of 88.66 %, 97.29 %, 91.58 %, and 100 %. Xiaoli et al. [14] addressed a long-term clinical efficacy evaluation decision making problem with temporal correlation hybrid attribute characteristics. They proposed a novel approach that combines a temporal correlation featured rough set model with machine learning techniques and non-additive measures. Ramanna et al. [15] identified rough-set based machine learning model with high accuracy that could be utilized for future studies regarding cannabinoids and precision medicine. They got high degree of accuracy by the developed approach with the patients affected by COVID-19 [15]. Abu-Gdairi and EL-Babli [16] have introduced refined mathematical techniques based on topological structures (called nearly initial-rough sets) derived directly from initial-rough sets. A rule-based classification system for COVID-19 variants was established. A. Hosny et al. [17] introduced the concepts of initial neighborhoods and ideals and mixing these with rough set theory, an exhaustive analysis of the dengue fever information system is conducted to validate the efficacy of the proposed approaches in maximizing accuracy and minimizing boundary regions. They achieved 63% accuracy. Nawar et al. [18] advanced rough set theory by defining tritopological approximation space and employed the defined notion of data reduction in the context of rheumatic fever disease.

Existing contributions utilizing rough set theory for medical diagnosis has explored various advanced techniques to address the uncertainty and developed diagnostic assisting models with high accuracy. They do not concentrate on searching the critical symptoms for the disease Pneumonia. Furthermore, none of these studies have proposed decision rule base generation concerned to the core set of attributes. Additionally, most research relies on deterministic data analysis while in the diagnostic scenario medical symptoms are often described using linguistic terms to indicate severity. So, the present approach is devoted to initiate reduction of dispensable symptoms and to investigate the core symptoms using information table having linguistic terms. A decision rules' basis is established to assist the diagnostics of pneumonia, aids to the degree of precision of diagnosis. In brief, the method identifies the core symptoms for concerned disease, finds its degrees of importance and establishes decision support system to aid diagnostic process. The article is organized as follows; Section-2 is concerned to the study of material and methodology. In Section-3 a numerical computation is illustrated for the pneumonia disease. It further provides an outcome analysis of the proposed research. Then, comparative analyses namely result and discussion is provided. Before deriving a conclusion, limitations of the research is presented. Finally, the paper puts an end to it by bringing it to a conclusion.

2 Material and Methodology

2.1 Information Systems and Decision Table [2,19]

An information system is distinguished by two disjoint classes of attributes, condition attributes and decision attributes respectively, denoted by S =(U,C,D) where U - universe of discourse, C and D are disjoints sets of conditional and decision attributes, respectively.

Example 1: An example of an information system (decision table) is shown in Table 1. [2]

$$C = \{Solar\ Energy,\ Volcanic\ activity,\ Residual\ CO_2\}\ and\ \ D = \{Tempreture\}$$

Table 1. Information system with conditional and decision attributes in linguistic form

Fact	Solar Energy	Volcanic Activity	Residual CO_2	Temperature	Days
p	Medium	High	Low	High	25
q	High	High	High	High	35
r	Medium	Low	High	High	80
s	Low	Low	Low	Low	150
t	High	High	Medium	High	75
u	Medium	Low	High	Low	30

The data set is inconsistent because facts r and u are contradictory, therefore the problem cannot be solved exactly but only approximately. Let us observe the data:

- Facts *p, q, r* and *t* can be certainly classified as causing high temperature.
- Fact *s* and *u* can be certainly classified as causing low temperature.

Example 2: Another example of a decision table is shown in Table 2.[19]

Table 2. Quantitative information system with conditional and decision attributes

Entity	BOD_5	NH_3-N	KN	Pb	WQE
A	2	2	3	2	2.2
B	1	2	3	1	1.7
C	1	2	3	2	1.7
D	1	1	1	1	1.0
E	1	3	1	1	1.7
F	1	1	1	1	1.0

Where grade-1 for very good; grade-2 for good; grade-3 for slightly polluted.

2.2 Indiscernibility Relation[19]

Let $B \subset A$, an indiscernibly relation on *B* (*Ind B*) is defined as $(x, y) \in Ind(B)$ if $b(x) = b(y)$ for every $b \in B$. Here *b(x)* denotes the value of attribute *b* for element *x*, it is the intersection of all equivalence relations (equivalence classes), *Re* (*b*), $b \in B$.

$$Ind\,(B) = \bigcap_{b \in B} Re(b) \tag{2.1}$$

Re (b) is the equivalence relation for the attribute b. From Table 2, $Re\,(BOD_5) = \{\{1\},\ \{2,3,4,5,6,7,8\}\}$,
$Re\,(NH_3\text{-}N) = (\{1,2,\ 3,\ 8\},\ (4,\ 6\}\{5,\ 7\}\}$, $Re\,(KN) = \{\{1,2,\ 3,\ 8\},\ (4,\ 5,6\}\{7\}\}$,
$Re\,(Pb) = \{\{1,\ 3\},\ (2,\ 4,\ 5,\ 6,\ 7,\ 8\}\}$, $B = \{BOD_5,\ NH_3\text{-}N,\ KN,\ P\}$. Using Equation (2.1) we get
$Ind\,(B) = (\{1\},\ \{3\},\ (2,\ 8\},\ (4,6\},(5\},(7\}\}$.

2.3 Lower approximation and Upper approximation[19]

Let $X \subseteq U$, R be an equivalence relation on U; U/R denotes the family of all disjoint equivalence relations (equivalence classes) induced on U by R. A rough set, R(X) is denoted as $\langle \underline{R}\,(X), \overline{R}(X)\rangle$ where, $\underline{R}(X)$ and $\overline{R}(X)$ are the lower approximation and upper approximation of X, respectively with respect to equivalence relations in U/R. The lower approximation is defined by
$\underline{R}\,(X) = \bigcup_x (Re\,(x)|\ Re\,(x) \subseteq X,\ x \in U)$.

And the upper approximation is defined as
$\overline{R}\,(X) = \bigcup_x (Re\,(x)|\ Re\,(x) \cap X \neq \phi,\ x \in U)$.

The boundary region of a set X is defined as $BN\,(X) = \overline{R}\,(X) - \underline{R}\,(X)$ and R - positive region with respect to X is $pos_R\,(X) = \underline{R}\,(X)$. If the boundary region of X is empty set then X is crisp set otherwise it is a rough set.

2.4 Reduction of Attributes and Importance of Attributes[19]

Let P and Q be two equivalence relations. Then P-positive region relative to Q defined as

$$pos_P(Q) = \bigcup_{X \in U/Q} \underline{P}(X) \tag{2.2}$$

If C- set of conditional attributes and D - set of decision attributes, the attribute c is said D-superfluous in C or c is said to be D-independency if

$$pos_{Ind\ C}(Ind\ D) = pos_{Ind\ C-(c)}(Ind\ D)\ or\ pos_C(D) = pos_{C-(c)}(D) \tag{2.3}$$

Importance Degree of Attributes

Let $C' \subseteq C$, importance degree of C' in C relative to D is measured by

$$\sigma_{CD}(C') = \gamma_C(D) - \gamma_{C-C'}(D),\ particularly\ if\ C' = \{c\},\ \sigma_{CD}(c\} = \gamma_C(D) - \gamma_{C-(c)}(D). \tag{2.4}$$

Where $\gamma_C(D) = \frac{\sum_{X_i \in U/D} |\underline{C}X_i|}{|U|}$, i.e. the percentage of objects in the lower approximation of X in C that surely belongs to D. Where

$$\sum_{X_i \in U/D} |\underline{C}X_i| = |pos_C D| \tag{2.5}$$

Then

$$\gamma_C(D) = \frac{|pos_C(D)|}{|U|}. \tag{2.6}$$

σ_{CD}, is a measure of the importance of the attributes in C relative to D. A large σ means that a large change in the classification of the attributes will result from taking away the attributes in C' .

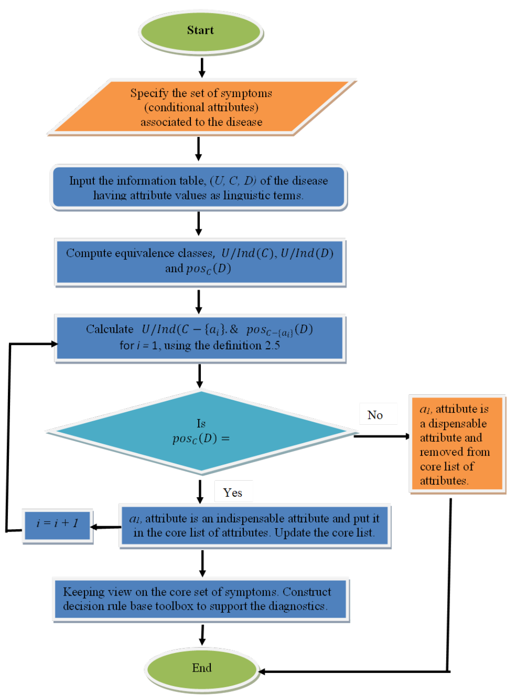

Fig 1. Simple flowchart showing decision system of medical diagnostics after core set formation

2.5 Proposed Algorithm

Methodology of the current article adheres the following algorithm:

Step 1: Specify associated symptoms of the disease and make them as conditional attributes.

Step 2: Construct possible knowledge rule base (decision table) showing behavior of the disease under associated conditional symptoms states linguistically.

Step 3: The knowledge rule base is improved by using Genetic algorithm based tuning.

Step 4: Input grades for different linguistic values of symptom characters and disease.

Step 5: Compute equivalence classes, $U/Ind(C), U/Ind(D)$ and $pos_C(D)$.

Step 6: Determine $U/Ind(C-(a_i\}, \ pos_{C-\{a_i\}}(D) \ for \ i=1$. If $pos_C(D) = \ pos_{C-\{a_1\}}(D)$, symptom a_1 is removed from core set. Set i = i + 1 until the list of the condition attributes is not absorbed. Repeat step 5 and step 6 for each i.

Step 7: Compute the degree of importance using $\sigma_{CD}\{c\} = \gamma_C(D) - \gamma_{C-\{c\}}(D)$ given in the Equation (2.4).

Step 8: Finally a simplified and precise decision support rule base is obtained.

3 Numerical Computation

The data in Table 3 is evaluated to assess the performance of the proposed method. It is a refined and tailored rule base using genetic algorithm in Roychowdhary et al. [20]. In this data set decision attribute is pneumonia (*D*) and conditional attributes are respiratory rate (*RR*), Cough level (*CO*), Chest in drawing *(CI*) and Temperature (*T*). Each of these attributes is coded into three qualitative categories: *L-Low, M-Medium, H-High* and quantitative categories: High-1, Medium-2, and Low-3. Here, initially, 42 entities (opinion) are available i.e. *U* = {1, 2, 3,…, 42}, decision table in qualitative form is given below:

Table 3. Information system of the pneumonia disease with symptoms in linguistic form

S.No.	RR	CO	CI	T	Pneumonia (D)	S.No.	RR	CO	CI	T	Pneumonia (D)
1	L	L	L	L	L	22	M	H	L	M	L
2	L	M	L	L	L	23	M	H	H	M	H
3	L	M	M	L	M	24	H	L	M	M	M
4	L	H	M	L	M	25	H	L	H	M	H
5	M	L	M	L	M	26	H	M	L	M	L
6	M	L	H	L	H	27	H	M	H	M	H
7	M	M	L	L	L	28	H	H	L	M	L
8	M	M	M	L	M	29	H	H	M	M	M
9	M	M	H	L	H	30	H	H	H	M	H
10	M	H	M	L	M	31	L	L	L	H	L
11	H	L	L	L	L	32	L	L	H	H	H
12	H	L	H	L	H	33	L	M	H	H	H
13	H	M	L	L	L	34	L	H	M	H	M
14	H	M	M	L	M	35	M	L	L	H	L
15	H	H	M	L	M	36	M	M	L	H	L
16	L	L	L	M	L	37	M	M	H	H	H
17	L	M	M	M	M	38	M	H	H	H	H
18	L	H	M	M	M	39	H	L	H	H	H
19	M	M	L	M	L	40	H	M	H	H	H
20	M	M	M	M	M	41	H	H	L	H	L
21	M	M	H	M	H	42	H	H	H	H	H

Quantitatively the above table is described as:

Table 4. Information system of the pneumonia disease and associated symptoms in quantitative format

Continued on next page

Table 4 continued

S.No.	RR (a_1)	CO (a_2)	CI (a_3)	T (a_4)	Pneumonia (D)	S.No.	RR (a_1)	CO (a_2)	CI (a_3)	T (a_4)	Pneumonia (D)
1	3	3	3	3	3	22	2	1	3	2	3
2	3	2	3	3	3	23	2	1	1	2	1
3	3	2	2	3	2	24	1	3	2	2	2
4	3	1	2	3	2	25	1	3	1	2	1
5	2	3	2	3	2	26	1	2	3	2	3
6	2	3	1	3	1	27	1	2	1	2	1
7	2	2	3	3	3	28	1	1	3	2	3
8	2	2	2	3	2	29	1	1	2	2	2
9	2	2	1	3	1	30	1	1	1	2	1
10	2	1	2	3	2	31	3	3	3	1	3
11	1	3	3	3	3	32	3	3	1	1	1
12	1	3	1	3	1	33	3	2	1	1	1
13	1	2	3	3	3	34	3	1	2	1	2
14	1	2	2	3	2	35	2	3	3	1	3
15	1	1	2	3	2	36	2	2	3	1	3
16	3	3	3	2	3	37	2	2	1	1	1
17	3	2	2	2	2	38	2	1	1	1	1
18	3	1	2	2	2	39	1	3	1	1	1
19	2	2	3	2	3	40	1	2	1	1	1
20	2	2	2	2	2	41	1	1	3	1	1
21	2	2	1	2	1	42	1	1	1	1	1

For above information table, $U/a_1, U/a_2, U/a_3$ and U/a_4 are given as follows:
$U/a_1 = \{\{1,2,3,4,16,17,18,31,32,33,34\}, \{5,6,7,8,9,10,19,20,21,22,23,35,36,37,38\},$
$\{11,12,13,14,15,24,25,26,27,28,29,40,41,42\}\}$
$U/a_2 = \{\{1,5,6,11,12,16,24,25,31,32,35,39\}, \{2,3,7,8,9,13,14,17,19,20,21,26,27,33,36,37,40\},$
$\{4,10,15,18,22,23,28,29,30,34,38,41,42\}\}$
$U/a_3 = \{\{1,2,7,11,13,16,22,26,28,31,35,36,41\}, \{3,4,5,8,10,14,15,17,18,20,24,29,34\},$
$\{6,9,12,21,23,25,30,32,33,37,38,39,40,42\}\}$
$U/a_4 = \{\{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15\}, \{16,17,18,19,20,21,22,23,24,25,26,2,28,29,30\},$
$\{31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42\}\}$
Taking intersection of four equivalence relations, we get
$U/Ind(C) = \{\{1\},\{2\},\{3\},\{4\},\{5\},\{6\},\{7\},\{8\},\{9\},\{10\},\{11\},\{12\},\{13\},\{14\},\{15\},\{16\},\{17\},$
$\{18\}, \{19\}, \{20\},\{21\},\{22\},\{23\},\{24\},\{25\},\{26\},\{27\},\{28\},\{29\},\{30\},\{31\},\{32\}$
$\{\{33\},\{34\},\{35\},\{36\},\{37\},\{38\},\{39\},\{40\},\{41\},\{42\}\}$
$U/Ind(D) = \{\{1,2,7,11,13,16,19,22,26,28,31,35,36\}, \{3,4,5,8,10,14,15,17,18,20,24,29,34\},$
$\{6,9,12,21,23,25,27,30,32,33,37,38,40,41,42\}\}$.
Here the different equivalence classes in $U/Ind(D)$ are
X_1 = {1, 2,7,11,13,16,19, 22, 26, 28, 31, 35, 36},
X_2 = {3,4,5,8,10,14,15,17,18,20,24,29,34}
X_3 ={6,9,12,21,23,25,27,30,32,33,37,38,39,40,41,42.
Since, $\underline{C}(X_1) = \{1, 2, 7, 11, 13, 16, 19, 22, 26, 28, 31, 35, 36\}$,

$\underline{C}(X_2) = \{3, 4, 5, 8, 10, 14, 15, 17, 18, 20, 24, 29, 34\}$

$\underline{C}(X_3) = \{6, 9, 12, 21, 23, 25, 27, 30, 32, 33, 37, 38, 39, 40, 41, 42\}$

Therefore, using Equation (2.5) C-positive region of D is $pos_C(D) = \bigcup_{X_i \in \frac{U}{D}} \underline{C}(X_i)$ = X.

Now to calculate $pos_{C-\{a_1\}}(D)$, we have

$U/Ind(C-\{a_1\} = \{\{1,11\},\{2,13\},\{3,8,14\},\{4,10,15\},\{5\},\{6,12\},\{7,13\},\{9\},\{16\},\{17,20\},$
$\{18,29\},\{19,26\},\{21,27\},\{22,28\},\{23,30\},\{24\},\{25\},\{31,35\},\{32,39\},\{33,37,40\},$
$\{\{34\},\{36\},\{38,42\},\{41\}\}$
$X_1 = \{1,2,7,11,13,16,19,22,26,28,31,35,36\}$,

$X_2 = \{3,4,5,8,10,14,15,17,18,20,24,29,34\}$
$X_3 = \{6,9,12,21,23,25,27,30,32,33,37,38,39,40,41,42\}$ are three sets are same as previous one.
Suppose $C-(a_1\} = C^{'}$. Lower approximations are $\bar{C'}X_1 = \{1,2,7,13,16,19,22,26,28,31,35,36\}$
$\bar{C'}X_2 = \{3,4,5,8,10,14,15,17,18,20,24,29,34\}$, $\bar{C'}X_3 = \{6,9,12,21,23,25,27,30,32,33,37,38,40,41,42\}$
$pos_{C-\{a_1\}}(D) = \{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,$
$25,26,27,28,29,30,31,32,33,34,35,36,38,39,40,41,42\}$

Since $pos_C(D) = pos_{C-\{a_1\}}(D)$, therefore attribute, a_1 is declared as dispensable. Similar analysis is employed for attributes a_2, a_3 and a_4, following results are obtained:

$pos_C(D) = pos_{C-\{a_2\}}(D), pos_C(D) \neq pos_{C-\{a_3\}}(D)$, and $pos_C(D) \neq pos_{C-\{a_4\}}(D)$.

Thus, the attribute a_2 is reduced and attributes a_3 and a_4 are indispensable attributes. To get the importance degree of the attributes Equation (2.6) is used. For this

$pos_{C-\{a_3\}}(D) = \{1, 4, 10, 15, 16, 17, 18, 33, 34, 35, 38, 39, 40, 41, 42\}$ &
$pos_{C-\{a_4\}}(D) = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27,$
$28, 29, 30, 31, 32, 33, 34, 35, 36, 38, 39, 40, 42\}$

Importance degree of attribute for symptom chest in drawing:
$\sigma_{CD}(a_3) = 1 - \frac{15}{42} = \frac{9}{14}$.
For symptom temperature:
$\sigma_{CD}(a_4) = 1 - \frac{40}{42} = \frac{1}{21}$.

4 Result and discussion

Finally, attributes, a_1 and a_2 are found as dispensable for pneumonia disease. These symptoms typically do not require immediate medical attention. Chest in drawing, a_3 and temperature, a_4 are found in the core attribute list. Importance degrees for both attributes a_3 and a_4 are evaluated as 9/14 and 1/21 respectively. It suggests that during the diagnostic process behavior of core symptoms must be studied more carefully. Also, to finalize rule base decision support system to assist medical diagnosis core list of symptoms is prioritized.

Following optimized decision rule base is obtained for medical diagnostics of the disease pneumonia using current approach. Support, Non- support, strength[2] and accuracy[11] for each rule are also rendered. Initially 42 entities are reduced to 9 certain rules with 100% accuracy.

Table 5. Decision rules basis of the disease pneumonia confining proposed approach

S.No.	Narration of decision rule	Support	Non- Sup-port	Strength	Accuracy
1	If chest in drawing is low and temperature is low then pneumonia is low	5	0	100	100
2	If chest in drawing is medium and temperature is low then pneumonia is medium	7	0	100	100
3	If chest in drawing is high and temperature is low then pneumonia is high	3	0	100	100
4	If chest in drawing is low and temperature is medium then pneumonia is low	5	0	100	100
5	If chest in drawing is medium and temperature is medium then pneumonia is medium	5	0	100	100
6	If chest in drawing is high and temperature is medium then pneumonia is high	5	0	100	100
7	If chest in drawing is low and temperature is high then pneumonia is low	4	0	100	100
8	If chest in drawing is high and temperature is high then pneumonia is high	7	0	100	100
9	If chest in drawing is medium and temperature is high then pneumonia is medium	1	0	100	100

Table 6. Comparative study with relevant existing approaches

Ip et al.[19] approach	Result of our approach of generating decision rules of the disease pneumonia is same to Ip et al. approach but our approach requires different idea and comparatively less computation in generating rules.				
Validation of proposed technique on the data El-Bably et al.[9]					
	Generated decision rules	**Support**	**Non-support**	**Strength**	**Accuracy**
Decision rule base prepared from the data set of El-Bably et al.[9] and Also comparative study of the results of both techniques	If the patient has symptoms the orthopnia and the ankle swelling then patient has disease heart failure	4	1	80	80
	If the patient has the symptom orthopnia then heart failure persist	2	0	100	100
	If the patient has symptoms reduce exercise tolerance and the ankle swelling then heart failure persist	2	0	100	100
	If the patient has symptoms the orthopnia and the reduce exercise tolerance then heart failure persist	5	0	100	100
	If the patient has symptom reduce exercise tolerance then heart failure persist	1	0	100	100
	If the patient has symptoms orthopnia, reduce exercise tolerance and ankle swelling then heart failure persist	3	0	100	100
	Bably et al. approach accuracy	**Current approach accuracy**			
$\{p_4\}$	25%	100%			
$\{p_{10}\}$	0%	100%			
$\{p_2, p_4, p_{10}\}$	50%	100%			
$\{p_2, p_4, p_5, p_7, p_{10}, p_{20}\}$	57%	100%			
Abu-Gdairi et al.[16] A. Hosny et al.[17] Nawar et al.[18]	Although these approaches are very strong for mathematical perspective. These have been employed on the deterministic data set similar to EL- Bably et al. technique. These methods have been developed to find the accuracy of the decision rules. While current technique is contrast from these techniques as present technique modifies the rule base and reduces the inconsistent rule, produces rules having better accuracy.				

Limitation of the study: A clinical information retrieval system is proposed based on the reduction of attribute in the study. The proposed approach is technically sound. It has been implemented on the secondary data which is adopted from Roychowdhury et al.[20]. The training data utilized for this study was processed via genetic algorithm. Because, the genetic algorithm may be problem specific and might be unable to apply in all problems. Also, data set can be scaled empirically. These can be limitations of the present study.

5 Conclusion

Available data often exhibits uncertainties which are frequent characteristic. It takes a lot of work to analyze such data and produce any useful, accurate information. In order to achieve this, current work introduces the core set based decision support retrieval system for disease diagnosis. Over the pneumonia diagnosis system, retrieval system is being examined. The proposed technique outperforms mentioned procedures. Obtained accuracy of the current approach is contrasted to Bably et al.[9], Abu-Gdairi et al.[16], A. Hosny et al.[17], Nawar et al.[18] illustrated in Table 6. Also, the present approach reduces the number of decision rules for the decision support system. The projected information retrieval system uses core set of attributes for the diagnosis of the disease pneumonia. It helps in reducing the cost of treatment for patients while diagnosis of pneumonia. These rules assist the physician for the pneumonia diagnosis. It is anticipated, will help doctors to choose the best course of action.

In the information system, it is advantageous that for all features, the range of the occurrence is partitioned into three linguistic terms. It can be increased appropriately. So, it is our future intention to introduce PSO, fuzzy set and rough set theory to establish a clinical decision model.

Acknowledgments

The author expresses his deepest gratitude to the Editor-in-Chief, Area Editor, and anonymous reviewers for their rigorous review and insightful comments. Their invaluable feedback has significantly strengthened the clarity and scientific rigor of this manuscript.

References

1) Pawlak Z. Rough sets. *International Journal of Computer and Information Sciences*. 1982;11(5):341–356. Available from: https://doi.org/10.4236/am.2016.712116.
2) Pawlak Z. Rough sets and intelligent data analysis. *Information Sciences*. 2002;147:1–12. Available from: https://doi.org/10.1016/S0020-0255(02)00197-4.
3) Wu WZ, Mi JS, Zhang WX. Generalized fuzzy rough sets. *Information Sciences*. 2003;151:379–380. Available from: https://doi.org/10.1016/S0020-0255(02)00379-1.
4) Tsumoto S. Mining diagnostic rules from clinical databases using rough sets and medical diagnostic model. *Information Sciences*. 2004;162:65–80. Available from: https://doi.org/10.1016/j.ins.2004.03.002.
5) Pawlak Z, Skowron A. Rough sets: Some extensions. *Information Sciences*. 2007;177:28–40. Available from: https://doi.org/10.1016/j.ins.2006.06.006.
6) Tripathy B, Acharjya D, Cynthya V. A framework for intelligent medical diagnosis using rough set with formal concept analysis. *International Journal of Artificial Intelligence & Applications*. 2013;2:45–66. Available from: https://doi.org/10.5121/IJAIA.2011.2204.
7) Vamsidhar T, Kotakonda M, Meghana CM, Naga JA, Lakshmi SK. The prediction of diseases using rough set theory with recurrent neural network in big data analytics. *International Journal of Intelligent Engineering and Systems*. 2020;13(5):10–18. Available from: http://www.inass.org/2020/2020103102.pdf.
8) Kumar K, Pandey D. Discussion on the switching between type-2 fuzzy sets and intuitionistic fuzzy sets: an application in medical diagnosis. *Journal of Information and Optimization Sciences*. 2018;39(2):427–444. Available from: http://dx.doi.org/10.1080/02522667.2017.1411014.
9) El-Bably MK, Ali MI, Abo-Tabl ESA. New topological approaches to generalized soft rough approximations with medical applications. *Journal of Mathematics*. 2021;2559495. Available from: https://doi.org/10.1155/2021/2559495.
10) Kumar K, Pandey D. Setting up waiting time targets for out-patients using fuzzy linear programming. *Indian Journal of Science and Technology*. 2023;28(16):2133–2143. Available from: http://dx.doi.org/10.17485/IJST/v16i28.1124.
11) Mishra M, Acharjya DP. A hybridized red deer and rough set clinical information retrieval system for hepatitis B diagnosis. *Scientific Reports*. 2024;14:3815–3815. Available from: https://doi.org/10.1038/s41598-024-53170-5.
12) Singh KN, Mantri JK. A clinical decision support system using rough set theory and machine learning for disease prediction. *Intelligent Medicine*. 2024;p. 1–10. Available from: https://doi.org/10.1016/j.imed.2023.08.002.
13) Singh KN, Mantri J. Clinical decision support system based on RST with machine learning for medical data classification. *Multimedia Tools Applications*. 2024;83:39707–39730. Available from: https://doi.org/10.1007/s11042-023-16802-y.
14) Xiaoli C, Juan X, Xiaodong C, Bingzhen S, Yan Z, Kun B. A non-additive rough set model for long-term clinical efficacy evaluation of chronic diseases in real-world settings. *Artificial Intelligence*. 2024;57. Available from: https://doi.org/10.1007/s10462-023-10672-4.
15) Ramanna S, Ashrafi N, Loster E, Debroni K, Turner S. Rough-set based learning: Assessing patterns and predictability of anxiety, depression, and sleep scores associated with the use of cannabinoids-based medicine during COVID-19. *Frontiers in Artificial Intelligence*. 2023;6:981953–981953. Available from: https://doi.org/10.3389/frai.2023.981953.
16) Abu-Gdairi R, El-Bably MK. The accurate diagnosis for COVID-19 variants using nearly initial-rough sets. *Heliyon*. 2024;17(10). Available from: https://doi.org/10.1016/j.heliyon.2024.e31288.
17) Hosny AR, Abu-Gdairi R, El-Bably MK. Enhancing dengue fever diagnosis with generalized rough sets: utilizing initial-neighborhoods and ideals. *Alexandria Engineering Journal*. 2024;94:68–79. Available from: https://doi.org/10.1016/j.aej.2024.03.028.
18) Nawar AS, Abu-Gdairi R, El-Bably MK, Atallah HM. Enhancing Rheumatic Fever Analysis via Tritopological Approximation Spaces for Data Reduction. *Malaysian Journal of Mathematical Sciences*. 2024;18(2):321–341. Available from: https://dx.doi.org/10.47836/mjms.18.2.07.
19) Ip WC, Hu BQ, Wong H, Xia J. Applications of rough set theory to river environment quality evaluation in China. . *Water Resource*. 2007;34:459–470. Available from: https://doi.org/10.1134/S0097807807040112.
20) Roychowdhury A, Pratihar DK, Bose N, Sankaranarayanan KP, Sudhahar N. Diagnosis of the diseases- using a GA- fuzzy approach. *Information Sciences*. 2004;162(2):105–120. Available from: https://doi.org/10.1016/j.ins.2004.03.004.