International Journal of Advanced Trends in Computer Applications
www.ijatca.com

Adaptive Properties of Tolerance Rough set in Healthcare feature selection and X-ray analysis

Indrani Kumari Sahu[1],Dr. Susant Kumar Das[2]
[1]Indrani Kumari Sahu
Dept of Master of Computer Application, KK (Auto) College
Berhampur, Odisha, India
Khallikote Autonomous College
[2]Susant Kumar Das
PG Dept of Computer Science
Berhampur University, Berhampur, Odisha, India
Berhampur University
[1]indranisahu@gmail.com, [2]dr.s.k.das.1965@gmail.com

Abstract: *Size reduction mechanism in real life data sets are very important and an essential factor in healthcare based machine learning (ML) analysis due to high dimension in nature. ML based feature selection aims in determining a minimal feature subset from a problem domain while retaining a suitably high accuracy in representing the original features. Rough Set (RS) theory provides the mechanism of discovery of data dependencies in the data set and the novel reducts facilitates, the reduction of the number of conditional attributes and the set of associated objects contained in a dataset in preserving the information of the original dataset. The process use the data alone and does not need any additional information. This paper presents the fundamental concepts of RS and Tolerance RS approaches and adapts the related feature selection for two relevant healthcare applications. Firstly, the TRS based feature selection method is used in latest developments of three medical dataset classification analysis, secondly the method is used in Chest X-Ray image analysis for nCOVID19 diagnose or test classifications and non-invasive thermal imaging process to detect inflammation and vascular dysfunction for sensitive screening of nCOVID19 cases.*

Keywords: Rough set, Tolerance Rough Set, Feature Selection, Reducts, Chest X-Ray image analysis, Thermal Imaging classification**.**

I. INTRODUCTION

Intelligent computational methods, prediction techniques, and deep learning based clinical analytics from raw medical data and are an eye-catching research aspect in recent theory and applications of IT and Medicine. Most of the approaches conquer to approximately solving real-life problems in decision making, predictions, learning and classification based intelligent systems [8],[11],[24]. Researchers and medical practitioners use some of the evolutionary methods like, neural networks, bayesian classifiers, genetic algorithms, decision trees, fuzzy sets, roughs sets and variances [6],[17],[18],[20],[21],[25].

Neural networks provide exhaustive methods to approximating functional values comprising real-valued, discrete-valued and vector-valued datasets. Back propagation algorithm uses gradient descent to tune network parameters to best fit the training set with input-output pair. This methodology of neural network has been applied to numerous problems including in medical field [14],[16]. In addition to solving broad range of problems, it also gained recognition in applying medical datasets though image processing, classification and machine learning, training of neural networks and system control. Decision trees and case based reasoning [1],[2] are also widely used to solve data analysis problems. Fuzzy sets deal with uncertainty ranging the precision of classical methods and the inherent imprecision of the real-life problems, especially in analysis of the segmented medical image.

Machine learning (ML) is significantly finds its applicability in the area of medical genomics, the study of the complete set of genes within an organism. Collaborative research across the domain pays attention how to model genetic sequencing and gain perspectives on the particular genetic blueprint that orchestrates all activities of that organism. In particular, sequence and analyze DNA, something that AI make faster, cheaper and more accurate. This would make decisions about care, what an organism might be susceptible to in the future, what mutations might cause different diseases and how to prepare for the future. Some of the latest developments in this space include, Microsoft's Elevation Project, Google's Deep Variant, Synpromics'sPromPT, Canada's Startup Deep Genomics, SOPHiAGenetics's Data-driven medicine [2]. ATUM's Leap-In Transposase Tech., Indian Startups Artivatic Data Labs and Orbuculum.

Rough Set (RS) introduced by Prof. Zdzislaw Pawlak[26] and recognized as a first non-statistical methods used in data analysis and classifications. Approximation of sets conquers in finding useful forms of granular computing which is known to be a part of computational intelligence. Information granulation used to discover to what extent a given set of objects approximates another set of objects of interest. RSmodels outperforms the results and analysis even if the healthcare data set is incomplete, imprecise or uncertain. Some applications includes, knowledge analysis, identification of imprecise or uncertain knowledge/information, quality evaluation and information analysis with respect to consistency, reasoning based uncertain or reduct of information data, attribute reduction, rule acquisition and intelligent algorithms. There are a number of practical outcomes of the RS approach, e.g., Cognitive sciences, engineering, machine learning, inductive reasoning, expert systems, pattern recognition, medicine, pharmacology, banking, financial and market analysis, decision analysis, process control, social networks and in various forms of machine learning [10], [12]. A study with RS in medical / clinical datasets is discussed in detailed in[7].RS found to be a powerful tool for data reduction based on dependency between attributes. The existing feature selection methods are not effective in selecting features which comprise real valued data. In this study, the methods use the strengths of Tolerance Rough Sets (TRS) to overcome from these issues.

The recent outbreak of the novel coronavirus (nCOVID-19) as been infecting millions of people across globe and killed several individuals. The World Health Organization has already declared this pandemic a global health emergency. The virus mainly infects the human respiratory tract leading to severe bronchopneumonia with symptoms of fever, dyspnea, dry cough, fatigue and respiratory failure. The mucormycosis (black fungus) variant of infection is observed to be more devastating effect and appears to be virulent. The reverse transcription polymerase chain reaction (RT-PCR) test is known to be the standard process in detecting nCOVID-19 but complex, time consuming and low access in high density of infected patients and community. Chest X-Ray is found to be an important and alternative screening process for the detection of nCOVID19. In this study we aim to use TRS based feature selection method for ML in identifying the abnormal cases during the diagnosis or testing for nCOVID19.

At the same time, rapid and sensitive screening tools are essential to limit the spread of nCOVID19 and to properly allocate at places with high and essential movement zones including clinics, hospitals and various healthcare units. Infrared thermography scanners used manually are risky and biased in nature. In this study, we aim in discussing a ML analysis with non-invasive thermal imaging process to detect inflammation and vascular dysfunction. The processing of thermal images by TRS based feature extraction or selection in ML identifies a solution for regions with limited diagnostic resources.

This article has the following organization. In section II we discuss foundations of classical RST and tolerance rough set,to be used in the paper. Section III provides an exhaustive discussion on feature selection on medical knowledge predictions with RS and TRS. Healthcare classification using reducts are analyzed in Section IV. Section V provides a brief review of image classification concerned to nCOVID-19. We discuss the usage of TRS based feature selection for X-Ray chest image classification and thermal imaging classification which would be useful in this pandemic. Finally, Section VI is the conclusion followed by bibliography.

II. FOUNDATIONS OF ROUGH SET

In keeping the view of MI for medical predictions and diagnosing process through observed decisions and available clinical datasets a brief explanation of the basic framework of RS along with some of the key definitions are presented in this section. A more comprehensive review can be found in [26],[27], [28].

The mathematical theory is based on an approximation space-based approach to classifying objects. Objects that possess similar characteristics known to be common set. An object in real-life finds a place in dataset (information table) and corresponds in a row and treated as a distinct entity. Necessary features concern to objects corresponds in columns of the dataset. In reality some information or relevance of the data (column value) may still associate with other objects in the dataset. In medical/clinical decision data sets, conditional attributes change according to the observations of experts and hence are not fixed. So, based on such dynamic characteristics from available information, it is required to identify the knowledge description and to approximate the likely sets. In other words, the set of indiscernible objects need to be classified automatically using any computational tools. Such identified objects are termed as "*knowledge granule*". These granules identify the knowledge of the relational data sets. RS models facilitates in eliminating such redundant information from high-dimensional datasets. The elimination, filters to a minimum number of attributes that would still allow each data record to be distinguished from the others. The optimized attribute set is termed as a “reduct”. We discuss such concepts, definitions, notations, as follows.

2.1 Classical Rough Set

In the universe of discourse U and an equivalence relation R over U, the family of all equivalence class of R is denoted as *U/R*, represent the categories or concepts of R. $[X]_R$ is termed as the equivalence class of an element$x \in U$. A relational system in a knowledge base is represented by$K = (U, R)$, with R family of equivalence relations on the universe.

The Indiscernibility relation (I) is the central component of RS. Indiscernibility relations are supposed to be equivalences such that two objects are equivalent if we cannot distinguish them by information. In other words, (I) is the intersection of all equivalence relations in S. S is any non-empty subset of R.

Definition-1. There are three distinct regions based on the equivalent classes induced by the feature values of objects. For any set$x \subseteq U, R \in IND(S)$,

$\underline{R}X = \bigcup\{Y \epsilon I(R): Y \subseteq X\}$ (1)

$\overline{R}X = \bigcup\{Y \epsilon I(R): Y \cap X \neq \emptyset\}$ (2)

$B_R(X) = \overline{R}X - \underline{R}X$ (3)

$\underline{R}X$and$\overline{R}X$are the lower and upper approximation respectively and $B_R(X)$ is the rough-boundary. Members of $\underline{R}X$ can be definitely classified as members of X but members of $\overline{R}X$can possibly be classified as members of X. In both-cases it retains the knowledge base of R. Moreover, the set X is said to be rough with respect to R if and only if both approximations are not equal or the rough-boundary must not be a null set. Similarly X is said to be rough-definable, if and only if both approximations are equal and the rough-boundary of X is null set.

2.2 Tolerance Rough Set

The major limitation of classical RS is the need of discrete data to process its methods, but many data models are continuous in real life applications. In time, the classical RS is blended with numerous hybrid approaches and extensions such as, Fuzzy Rough sets (1992), Variable precision RS(1993), Tolerance RS (1996), Neighborhood RS(1988), Dominance-based RS (2001), Covering based RS(2004). A more comprehensive review on these can be found in [7].

The indiscernibility relations which are equivalences arise naturally when one considers a given set of attributes: two objects are equivalent when their values of all attributes in the set are the same. However, some of the natural indiscernibility relations encountered in nondeterministic information systems are not transitive.

Definition-2. Let a decision table $T=(U, C \cup\{d\})$, where U is a set of real life objects, C is a set of condition attributes, d is a decision attribute. The procedure of reduction decision table is finding a new optimized table $T_r=(U_r, C_r \cup\{d\})$ that satisfies the conditions $U_R \subset U$, $C_r \subset C$ and decision rules constructed from C_r have the same quality of classification as decision rules constructed from C.If a binary relation R on a set U is reflexive and symmetric, it is called a tolerance relation on U. The set of all tolerance relations on U is denoted by *Tol(U)*. For all $R \in Tol(U)$ and $a \in U$, the set $a/R = \{b \in U \mid aRb\}$ is the R-neighborhood of a.

Definition-3. A tolerance decision table is defined by (C, τ(Tol(U), F_C)), where C is the condition attributes, τ(Tol(U), F_C) is a tolerance relation on the set of objects, F_C is a family of subsets C. The tolerance set of object $x$is defined as:

$TS(x) = \{x_r \in U | xTol(U), F(C)x_r\}$ (4)

Definition-4. Let U be a set of objects and let R be a tolerance on U. The lower-*R-approximation*, the upper-R-approximation, and the R-boundary of a set $x \subseteq U$ are:

$X_R = \{x \in U | x/R \subseteq X\}$ (5)

$X^R = \{x \in U | x/R \cap X \neq \emptyset\}$ (6)

$B_R X = X^R - X_R$ (7)

The set X_R consists of elements which surely belong to X in view of the knowledge provided by R. The R-boundary is the actual area of uncertainty. It consists of elements whose membership in X cannot be decided when R-related objects cannot be distinguished from each other.

Definition-5. Let $R \in Tol(U)$. A set $x \subseteq U$ is R-definable if $X_R = X^R$, where R is a tolerance.A set X is R-definable if and only if its R-boundary $B_R(X)$ is empty.

Lemma-1. Let $R \in Tol(U)$ and $X, Y \subseteq U$
(a) $X \in Def(R)\ iff\ X^R = X\ iff\ X_R = X$
(b) if $X \in Def(R), then\ (X \cup Y)_R = X_R \cup Y_R$ and

$$(X \cap Y)^R = X^R \cap Y^R$$

The set of all equivalences on U is denoted by *Eq(U)*. $x \subseteq U$ is saturated by an equivalence $E \in Eq(U)$, if X is the union of some equivalence classes of $E\ or\ x = \emptyset$. The set of all sets saturated by E is denoted by *Sat(E)*. A family $F \subseteq \varphi(U)$ is called a complete field of sets if $\emptyset, U \in F, X^C \in F$ for all $X \in F$and $\bigcup H, \bigcap H \in F$for all $H \subseteq F$. For $R \in Tol(U)$, $R^e = \bigcap\{E \in Eq(U) | R \subseteq E\}$.

Proposition-1. If $R \in Tol(U)$, then
(a) Def(R) = Sat(R^e)
(b) Def(R) is a complete field of sets

Note that $X^{R^e}$is the least R-definable set including X and that X_{R^e} is the greatest R-definable set included in X.

Let $E \in Eq(U)$, by Proposition 1(a), the E-definable sets are the unions of some (or none) E-classes and this actually is Pawlak's original definition of E-definable sets[15]. The sets X^E and X_E are E-definable for all $x \subseteq U$, but the approximations induced by tolerances are not necessarily definable.

Definition-6. Let $R \in Tol(U)$. The equivalence classes of the R-equality $\equiv_R$are called R-rough sets.

Further,R-rough sets can be equivalently viewed as pairs (X_R, X^R), where $x \subseteq U$, since each $\equiv_R - class\ C$ is uniquely determined by the pair (X_R, X^R), where X is any member of C.

For $R \in Tol(U)$and $x \subseteq U$, the pair R<X> =(X_R, X^R) is known as the R-approximation of X. The set of all R-approximations of the subsets of U is $A(R) = \{R < x > | X \subseteq U\}$.

III.FEATURE SELECTION

Real world healthcare data analytics accumulate several types of associated data (*data-driven*, *knowledge-driven*, *device-driven*) for future diagnosis, treatment, prognosis, and predictive analysis. During the analysis, both ML developers and clinical practitioners come across with situations to limit or even reduce the collected data. They often employ the largest number of attributes available or even conceivable from medical automation systems, in a well circumscribed experimental settings to train a set of ML models on specific relevant targets. These may relevant on classifying cases in terms of either the associated diagnosis or a stratum of expected improvements or outcomes and then perform a quantitative feature selection to determine the most useful N-features for each predictive task at hand. These features could then be indicated as being the data that it is recommended to report as carefully and accurately as possible for the secondary analysis.

Classification analysis is widely adopted for healthcare applications for supporting medical diagnostic decisions, improving quality of patient care, etc. Some data may be irrelevant to a particular case but not for all. If a training dataset contains irrelevant features (attributes) for the specific process, the classification analysis may produce less accurate and less understandable results.

Feature selection (FS) determines a minimal feature subset from a problem domain while retaining a suitably high accuracy in representing the original features. For such specific healthcare analysis tasks, FS is a must to tackle the abundance of noisy, irrelevant or misleading features and on removing these factors, learning from data techniques can benefit greatly. Aspects of FS is open (filter, wrapper) approach and several FS methods have been evolved in recent research. However, we limit our discussion for FS using RS and its variances.

3.1Rough Set based Feature Selection

As discussed earlier, RSis used as a tool to discover data dependencies and to reduce the number of attributes contained in a dataset using the data alone, requiring no additional information. Over the past years, it has become a topic of great interest to researchers [28]and has been applied to many domains. Given a dataset with discretized attribute values, it is possible to find a subset (termed a reduct) of the original attributes using RS that are the most informative; all other attributes can be removed from the dataset with minimal information loss. From the dimensionality reduction perspective, informative features are those that are most predictive of the class attribute.There are two main approaches to finding reducts, one, considers the degree of dependency and the other concerned with the discernibility matrix.

Let *I=(U, A)* be an information system, where U is a non-empty set of finite objects, the universe and A is a non-empty finite set of attributes such that a:U→V_a for every $a \in A$. V_a is the set of values that attribute a may take. With any $P \subseteq X$ there is an associated equivalence relation *IND(P)*.

$$IND(P) = \{(x, y) \in \mathrm{U}^2 | \forall \mathrm{a} \in \mathrm{P}, \mathrm{a(x)} = \mathrm{a(y)}\} \qquad (8)$$

The partition of U generated by *IND(P)* is denoted *U/IND(P)* or *U/P*. If $(x, y) \in IND(P)$, then x and y are indiscernible by attributes from P. The equivalence classes of the P-indiscernibility relation are denoted $[x]_P$.

Let $X \subseteq U$, X can be approximated using only the information contained within P by constructing the P-lower and P-upper approximations of X.

$$\underline{P}(X) = \{x | [x]_P \subseteq X\} \qquad (9)$$

$$\overline{P}(X) = \{x | [x]_P \cap X \neq \emptyset\} \qquad (10)$$

Let P and Q be equivalence relations over U, then the positive region can be defined as:

$$POS_P(Q) = \cup_{X \in U/Q} \underline{P}X \qquad (11)$$

The positive region contains all objects of *U* that can be classified to classes of *U/Q* using the information in attributes *P*. Using this definition of the positive region, the rough set degree of dependency of a set of attributes *Q* on a set of attributes P is defined in the following way:

For $P, Q \subset A$, it is said that *Q* depends on *P* in a degree $k(0 \le k \le 1)$, denoted $P \to_k Q, if\ k = \gamma_P(Q) = |POS_P(Q)|/|U|$ (12)

Definition-7. The threshold matrix $TM(c)\ for\ all\ c \in C$ is $card(U) \times card(U)$matrix, where $TM_{ij}(c) = S_c\left(c(x_i), c(x_j)\right)$, x_i, x_j are indiscernible with highest value of the threshold, and the element number of different values in threshold matrix *TM(c)* is less or equal to $\{\left(card(V_c)\right)^2 - card(V_c)\}/2$ +1.

Definition-8. In a tolerance set *Tol(c)* of attribute *c* with given threshold value *t(c)*, the tolerance matrix can be defined as:

$$TSM_{ij}(A) = \begin{cases} 1\ if \left(\sum_{c \in C} TSM_{ij}(c)\right) \ge t.card(A) \\ 0, otherwise \end{cases}$$

3.1.1 Quick Reduct Algorithm

The reduction of attributes is achieved by comparing equivalence relations generated by sets of attributes. Attributes are removed so that the reduced set provides the same predictive capability of the decision feature as the original. A reduct*R* is a subset of minimal cardinality of conditional attribute set *C* such that $\gamma_R(D) = \gamma_C(D)$.

QUICK-REDUCT ALGORITHM

```
C, the set of all conditional features
D, the set of decision features
R ← { }
do
  T ← R
  ∀x ∈ (C-R)
    if $\gamma_{R\cup\{x\}}$ (D) >$\gamma_T$ (D)
      T ← R∪{x}
    R ←T
until$\gamma_R$ (D) = $\gamma_C$ (D)
return R
```

The above QUICK-REDUCT algorithm attempts to calculate a reduct without exhaustively generating all possible subsets. It starts off with an empty set and adds in turn, one at a time, those attributes that result in the greatest increase in the rough set dependency metric, until this produces its maximum possible value for the dataset. The dependency of each attribute is calculated and the best candidate chosen. The generated reduct shows the way of reducing the dimensionality of the original dataset by eliminating those conditional attributes that do not appear in the set.

This, however, is not guaranteed to find a minimal subset. Using the dependency function to discriminate between candidates may lead the search down a non-minimal path. It is impossible to predict which combinations of attributes will lead to an optimal reduct based on changes in dependency with the addition or deletion of single attributes. It does result in a close-to-minimal subset, though, which is still useful in greatly reducing dataset dimensionality.

3.2 Tolerance Rough Set based Feature Selection

As discussed above, the limitation of RS in dealing with discrete value observes bottleneck with real-valued dataset, that can be taken care using the concept of measure of similarity of feature values. By relaxing the transitivity constraint of equivalence classes, a further degree of flexibility (with regard to indiscernibility) is carried out by TRS. TRS can be conceptualized as an approximation space (lower and upper) based on these similarity measures. In classical RS, objects are grouped into equivalence classes if their attribute values are equal. This requirement might be too strict for real-world healthcare data, where values might differ only as a result of noise.

3.2.1 Similarity Measures

Features signing through attributes need to be related by suitable similarity relation using similarity measures for the TRS. Although the same definition can be used for all features if applicable, a standard measure for this purposeis defined as follows:

$$SIM_a(x, y) = 1 - \frac{|a(x) - a(y)|}{|a_{max} - a_{min}|} \quad (13)$$

Where a is the attribute under consideration, a_{max} and a_{min} denote the maximum and minimum values respectively for the features (attribute) taken. When considering more than one attribute, the defined similarities must be combined to provide a measure of the overall similarity of objects. For a subset of features, *P*, this can be achieved in many ways, in this context, there are two commonly adopted approaches are:

$$(x, y) \in \mathrm{SIM}_{P,\tau} \text{ iff } \prod_{a \epsilon p} SIM_a(x, y) \ge \tau \quad (14)$$

$$(x, y) \in \mathrm{SIM}_{P,\tau}\ iff\ \frac{\sum_{a \in P} SIM_a(x, y)}{|P|} \ge \tau \quad (15)$$

Where$\tau \in [0,1]$ is a global similarity threshold (τ)that determines the required level ofsimilarity for inclusion within tolerance classes. It can be seen that thisframeworkallows for the specific case of classical

RS by defining a suitable similaritymeasure (e.g. equality of feature values and Equation-14) and threshold ($\tau = 1$).Tolerance classes generated by the similarity relation for an object x are defined as follows:

$$\mathrm{SIM}_{P,\tau}(x) = \{y \in U | (x,y) \in \mathrm{SIM}_{P,\tau}\} \quad (16)$$

Lower and upper approximations can be defined as:

$$\underline{\mathrm{P}_{\tau}}(x) = \{x | \mathrm{SIM}_{P,\tau}(x) \subseteq X\} \quad (17)$$

$$\overline{\mathrm{P}_{\tau}}X = \{x \,| \mathrm{SIM}_{P,\tau}(x) \cap X \neq \emptyset\} \quad (18)$$

The positive region and dependency functions are:

$$\mathrm{POS}_{P,\tau}(Q) = \cup_{X \in U/Q} \underline{\mathrm{P}_{\tau}}(x) \quad (19)$$

$$\gamma_{P,\tau}(Q) = \frac{|\mathrm{POS}_{\,P,\tau}(Q)|}{|U|} \quad (20)$$

In using the above definitions, attribute reduction methods can be constructed that use the tolerance based degree of dependency, $\gamma_{P,\tau}(Q)$ to gauge the significance of feature subsets.

3.2.2 Quick Reduct Algorithm with TRS

In order to overcome from the real-valued quantification limitation, either the real-valued measurements need to be discretized to near value or quantified subjectively. In both ways it compromises the degree of accuracy. In such cases, TRS and Fuzzy rough set offerpromising solutions to handle such limitations. The TRS replaces the traditional equivalence classes of crisp RS with alternatives that are better suited in dealing with this type of data. Indiscernibility relations are replaced with similarity relations that permit a limited degree of variability in attribute values. Approximations are derived based on these tolerance classes in a manner similar to that of classical RS.

Tolerance-Based-Quick-Reduct Algorithm

```
C, the set of all conditional features
D, the set of decision features
τ, similarity threshold
R ←{ }; $\gamma^T$ best =0
do
   T ← R
   $\gamma^T$prev = $\gamma^T$ best
   ∀x ∈ (C-R)
      if $\gamma_{R\cup\{x\},\tau}$ (D) >$\gamma_{T,\tau}$ (D)
          T ← R∪{x}
          $\gamma^T$ best = $\gamma_{T,\tau}$ (D)
      R ←T
until $\gamma^T$ best == $\gamma^T$prev
return R
```

In using the data (Table1) and definitions discussed above we derive, *P*={b,c}, *U/IND(P)* ={{3}, {1,5}, {4}, {2,7,8}, {6}}. For *P*={b,c} and *Q*={e}, then $POS_P(Q)$={3,4,6}, $\gamma_{\{b,c\}}$=3/8. Attribute *d* generates the highest dependency degree, so that attribute is chosen and the sets {a,d}, {b,d} and {c,d}are evaluated.

Table 1: An Example Dataset (Crisp Data)

X	*a*	*b*	*c*	*d*	*e*
1	1	0	2	2	0
2	0	1	1	1	2
3	2	0	0	1	1
4	1	1	0	2	2
5	1	0	2	0	1
6	2	2	0	1	1
7	2	1	1	1	2
8	0	1	1	0	1

In using the data (Table-2) and Equation (13-15) the similarity measure for all conditional attributes, the threshold of (0.7 for rest) and τ=1 for decision feature. This allows attribute values brings close values to be considered as identical anddiffers to a limited extent. The derived seven tolerances classes for the Table-2 are:

$U/SIM_{\{a\},\tau}$ = {{1,2,3},{4,5,6}}
$U/SIM_{\{b\},\tau}$ = {{1,3,4,5},{2},{6}}
$U/SIM_{\{c\},\tau}$ = {{1},{2,4,5,6},{3}}
$U/SIM_{\{q\},\tau}$ = {{1,3,6}, {2,4,5}}
$U/SIM_{\{a,b\},\tau}$ = {{1,3}, {2}, {4,5}, {4,5,6}, {5,6}}
$U/SIM_{\{b,c\},\tau}$ = {{1,3}, {2,6}, {4,5,6}, {2,4,5,6}}
$U/SIM_{\{a,b,c\},\tau}$ = {{1,3}, {2},{4,5,6}}

Table 2: An Example Dataset (Real Value Data)

X	*a*	*b*	*c*	*D*
1	-0.4	-0.3	-0.5	No
2	-0.4	0.2	-0.1	Yes
3	-0.3	-0.4	-0.3	No
4	0.3	-0.3	0	Yes
5	0.2	-0.3	0	Yes
6	0.2	0	0	No

In the above derived tolerance classes, it can be seen that, there are some objects found to be available repeatedly and belong to more than one tolerance class. This is due to the relaxed and flexibility in identifying classes in TRS in comparison to classical RS. The degree of dependency for conditional attribute subsets with significance depends on the above derived partitions. The degree of dependency$\gamma_{a,\tau}$(D) =0/6,$\gamma_{b,\tau}$(D) =2/6, $\gamma_{c,\tau}$(D) =2/6, $\gamma_{\{a,b\},\tau}$(D) =5/6, $\gamma_{\{b,c\},\tau}$(D) =2/6, $\gamma_{\{a,b,c\},\tau}$(D) =3/6. The Tolerance-based-quick-reductalgorithm identifies the output-reduct as{a,b}.

IV. HEALTHCARE DATASET CLASSIFICATION

In this section, we describe the conducted experiments to evaluate the performance of Tolerance RS based quick reduct (TRQ) model for medical classification models and compare its performance to other well-regarded and recent models using GA, PSO, ABC and BAT optimizations. The medical classification problems, evaluation measures, the experiments setup, and results are as follows.

The datasets concern with four different diagnostic problems in medicine where each focuses on identification of a specific disease.We demonstrate the performance analysis of classification models using following four popular healthcare datasets.

The Breast cancer dataset describes about 569instances that had undergone surgery for breast cancer. There are 357 instances for *benign* label and 212 instances for *malignant* label. More details of the dataset is available in [29].

The Parkinson dataset describes about 23 persons with Parkinsons disease and 8 healthy persons. There are 195 voice recordings and 23 biomedical voice measurements. The decision attribute *status*: 0 for healthy and 1 for PD. More details of the dataset is available in [30].

The Diabetes dataset describes on diagnostic measures of type 2 diabetes in women of the Pima Indian heritage[31]. It contains 768 observations with eight conditional attributes and the decision attribute identifies, whether the person is diagnosed with type 2 diabetes. There are diabetes positive cases of 268 patients and marked with binary value of 1 and 500 cases of diabetes negative with binary value of 0.*Pregnancies*, *Glucose*: The blood plasma glucose concentration after a 2 hour oral glucose tolerance test, *BloodPressure*: Diastolic blood pressure (mm/HG), *SkinThickness*: Skinfold thickness of the triceps (mm), *Insulin*: 2 hour serum insulin (mu U/ml), *BMI*: Body mass index (kg/m squared), *DiabetesPedigreeFunction*: A function that determines the risk of type 2 diabetes based on family history, the larger the function, the higher the risk of type 2 diabetes and *Age*.The decision attribute, *Outcome*:whether the person is diagnosed with type 2 diabetes (1=yes, 0=no).

The SAheart dataset [32] represents sample of males in a heart-disease high-risk region of the Western Cape, South Africa. It describes about 462 samples of males in heart disease with high risk. The dataset has nine conditional attributes, *sbp*: systolic blood pressure, *tobacco*: cumulative tobacco (kg), *ldl*: low density lipoprotein cholesterol, adiposity, *famhist*: family history of heart disease (Present=1, Absent=0), *typea*: type-A behavior, *obesity*, *alcohol*: current alcohol consumption, *age*: age at onset. The decision attribute *chd*: coronary heart disease (yes=1 or no=0).

In Table3, we put forward the evaluated parameters of classification diagnostic testing results of breast cancer, parkinson disease, diabetes and SAheart dataset of four well known algorithms (GA, PSO, ABC, BAT) and TRQ. The evaluation measures ofclassificationin terms of specificity, sensitivity, geometric mean, area under the ROC curve and accuracy are presented with the following detailed descriptions.

$$Specificity = \frac{TN}{TN+FP} \tag{21}$$

Specificity is very useful in the healthcare analysis which measures the percentage of the correctly predicted negative samples to the actual total number of negative samples. This type of measure is used to detect correctly the patients who do not carry disease.

$$Sensitivity = \frac{TP}{TP+FN} \tag{22}$$

Sensitivity measures the percentage of the correctly predicted positive samples to the actual total number of positive samples. In healthcare analysis, the high sensitivity values represent that the classification model can correctly predict the patients who carry disease.

$$G-Mean = \sqrt{Specificity \times Sensitivity} \tag{23}$$

The geometric mean (G-Mean) combines the measures between the sensitivity and specificity values.

$$AUC = \int_0^1 \frac{TP}{(TP+FP)} d\frac{FP}{(TN+FN)} = \frac{1}{\{(TP+FP).(TN+FN)\}}\int_0^1 TP\, d\, FP \tag{24}$$

Area under the receiver operating characteristic curve (ROC) curve (*AUC*) measures the area under the curve of receiver operating characteristics (ROC) graph. The *ROC* curve plots between true positive rate and false positive rate. In healthcare analysis, *AUC* plays a vital role, when the total number of positive and negative samples have unequal distribution in the dataset.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{25}$$

Accuracy (classification rate) presents the percentage of the correctly classified samples (positive or negative) to the actual total number of samples.

Table3: Classification: Diagnostic testing evaluation results

Dataset	*Algorithm*	*Specificity*	*Sensitivity*	*G-Mean*	*AUC*	*Accuracy*
BREAST CANCER	GA	0.9436	0.9798	0.9615	0.9955	0.9675
	PSO	0.9477	0.9822	0.9647	0.9956	0.9704
	ABC	0.9621	0.9724	0.9671	0.9854	0.9689
	BAT	0.9346	0.9764	0.9549	0.9708	0.9622
	TRQ	0.9531	0.9805	0.9666	0.9954	0.9711
DIABETES	GA	0.8572	0.5691	0.6980	0.8389	0.7516

	PSO	0.8626	0.5035	0.6565	0.8189	0.7310
	ABC	0.8480	0.5757	0.6956	0.8225	0.7482
	BAT	0.8729	0.5750	0.7078	0.8435	0.7637
	TRQ	0.8837	0.5594	0.7030	0.8553	0.7649
PARKINSONS	GA	0.6000	0.9693	0.7579	0.9282	0.8811
	PSO	0.4604	0.9654	0.6636	0.8512	0.8448
	ABC	0.4375	0.9640	0.6464	0.8000	0.8383
	BAT	0.5562	0.9588	0.7211	0.8464	0.8627
	TRQ	0.7250	0.9647	0.8308	0.9189	0.9075
SAHEART	GA	0.4259	0.7981	0.5865	0.7667	0.7181
	PSO	0.8510	0.4870	0.6413	0.7602	0.7266
	ABC	0.8228	0.4975	0.6364	0.7445	0.7116
	BAT	0.8365	0.5074	0.6509	0.7664	0.7240
	TRQ	0.4938	0.8545	0.6491	0.7555	0.7312

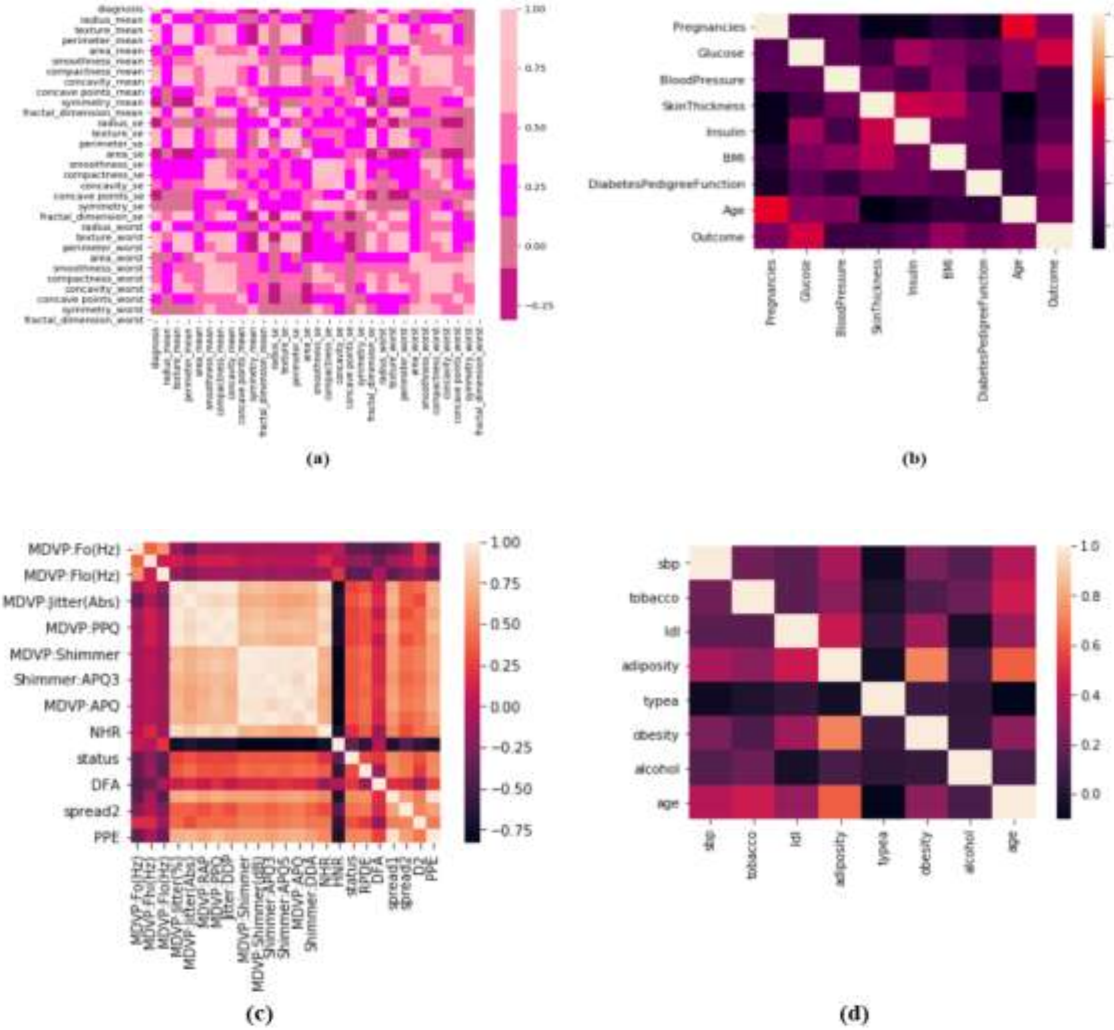

Figure1: Heatmap of features of datasets: (a) Breast cancer (b) Diabetes (c) Parkinsons (d)SAheart

V. HEALTHCARE IMAGE CLASSIFICATION

The most hilarious effect in present healthcare system is the deadly affect of the disease nCOVID-19.The pathogenic zoonotic novel coronavirus is termed 2019_nCov and is the third time coronavirus effect to infect human population. The previous two are, the severe acute respiratory syndrome coronavirus (SARS-CoV) outbreak in 2002 and the middle east respiratory syndrome coronavirus (MERS_CoV) outbreak in 2012.

To date, the world witness with 155,541,473 infections and 3,248,553 recorded deaths, where around 1,121,499,817 vaccine doses have been administrated. The disease is caused by severe acute respiratory syndrome coronavirus-2 (SARS-COV-2), leading to high morbidity and mortality worldwide. Several studies reveal that, chest X-Ray is one of the important role in the detection of such visual responses associated with SARS-COV-2 infection.

5.1 X-Ray Image Segmentation using TRS FS

In diagnosing and testing phase, the most reliable and the gold standard test for nCOVID-19 diagnosis is a real-time reverse transcription-polymerase chain reaction (RT-PCR). However, in clusters with high population, the test dependency factor per person becomes very high. At the same time, chest radiological imaging through X-ray or computed tomography (CT) are found to be effective in the early diagnosis and treatment of nCOVID-19. Application of machine learning (ML) for image classifications have recently gained popularity and used in many healthcare based automatic diagnosis systems and deep learning is one of the promising approaches. To overcome from the problems like, heavy load of RT-PCR machines, insufficient or availability of test kits, assays, test costs and waiting time of test results, MI based Chest X-Ray (CXR)image classification can be used to overcome such problems and can provide timely assistance to medical healthcare.

Recently many radiology images have been studied for nCOVID-19 detection [3], [4], [13]. For instant normal chest X-Ray images of healthy samples are shown in Figure 2(a-c) and Figure 2(d-f) represents the infected chest X-Ray image samples of some patients [33]. Deep-learning based classification models like deep convolutional network, recursive network, transfer learning models have been implemented to analyze the radiological disease characteristics [23]. Convolutional neural network (CNN) assigns a class label to different super pixels extracted from the lungs parenchyma and localize tuberculosis-infected regions in CXR images with average dice index of 0.67. In [5] authors used backpropagation neural network model that takes input of weekly labeled CXR images and generates visual attention feedback for accurate localization of pulmonary lesions; further they used reinforcement learning-based recurrent attention model which learns the sequence of images to find the nodules. In [19] authors came up with CheXNet-deep learning based model integrated with LibreHealth radiology information system, which analyzes uploaded CXR images and assigns one of 14 diagnostic labels. To avoid the overfitting of the models, some studies used a data augmentation technique, which generates different variants of the source image by applying random photometric transformations like blurring, sharpening, contrast adjustments etc.

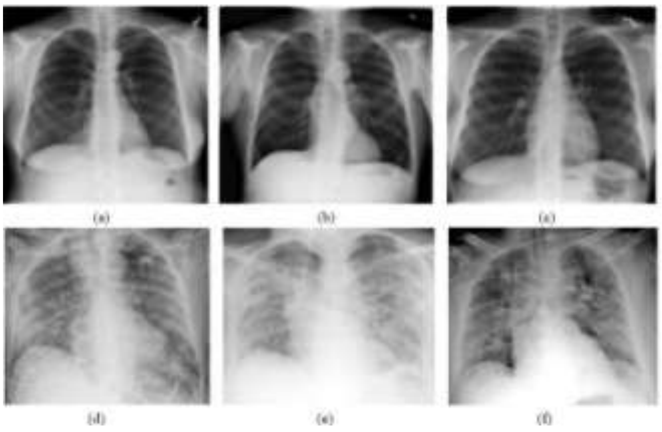

Figure 2: Chest X-Ray Images
(a-c) Normal chest X-Ray images, (d-f) nCOVID-19 infected chest X-Ray images

In analyzing the above, we observed that the existing studies had been performed using a limited number of input CXR or CT images which may lead to under-fitting of the data-hungry DL models. Conventional ML techniques can be better integrated with Tolerance RS based feature selection using quick reduct to overcome these short comings. The architectural outline of the CXR classification process with Tolerance Rough Set and Reduct approach, is presented in Figure3.

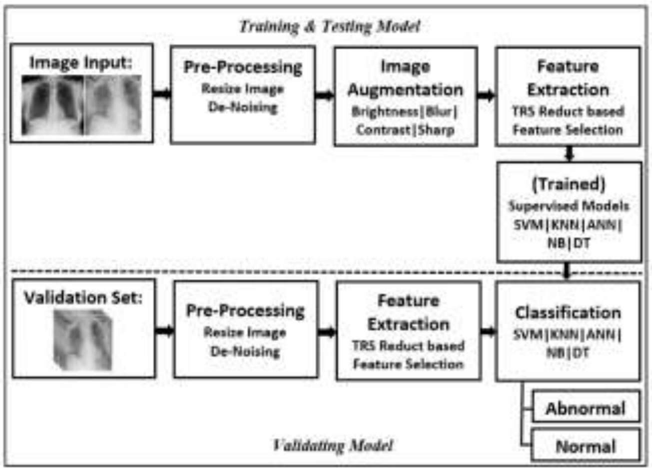

Figure3: The Architecture of Classification Model with Tolerance RS Reduct FS

In the training and testing model, all input images are preprocessed, with an image resizing of (512×512)pixels, format conversion (portable network graphics) and color space conversion (Gray Scale). Subsequently, the texture preserving guided filter is applied to reduce the inherent quanum noise [15]. The choice of the denoising filter is based on [22].The preprocessed images are divided into two sub sets: training-testing set (80%) and validation set (20%).Further the image augmentation technique is applied to the images of the training-testing set to build a generalized model.

In the feature extraction and feature selection phase, the TRS based reduct feature selection method is adapted in the model. The nCOVID-19 infected patients exhibit different radiographic texture patterns (see Figure-4) such as patchy ground-glass opacities, pulmonary consolidations, reticulonodular opacities. These characteristics can be efficiently represented with the help of radiomic texture descriptors.

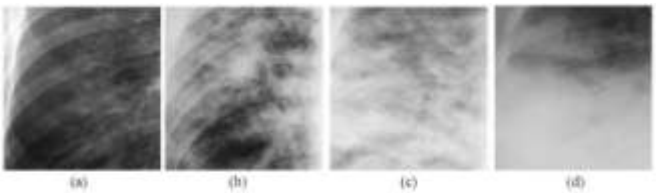

Figure 4: nCOVID-19 infected Chest X-Ray images (a) Mild Opacities (b) Ground-glass opacities, (c) Reticular Opacities, (d) Pulmonary Consolidation

A normal study needs 8 first order statistical features (FOSF), 88 grey level co-occurrence matrix (GLCM) features (in four different orientations) and 8100 histogram of oriented gradients (HOG), a total of 8196 features from each CXR image. It is to note that, all the extracted features are not relevant for accurate characterization of visual indicators associated with nCOVID-19. Thus, to select the most informative features, TRS based Quick Reduct algorithm is used which outperforms in classifying the X-ray images of normal or abnormal type (see Figure 2) cases.

5.2 Thermal Imaging and TRS Classification

Fever or higher body temperature is observed as the possible first-hand symptom of nCOVID-19 infection. Thermal imaging system (TIS) aims to detect a high body temperature accurately without human intervenes. TIS are extremely useful in the context of high throughput public places in healthcare based medical environments. In order to protect the medical practitioners or other patient visitors in the clinic, TIS would be more appropriate method to initially assess and triage people, especially if there is a risk that infected people would not be identified away. Moreover, in a clinic or nursing home, inaccurate biased temperature measurement or a missed contagious person without a fever could spread infection among nursing home residents. So in this case, other assessment options and following infection control practices may be more effective. Similarly, in mass temperature screening areas like workplaces, grocery stores, railway platforms, airports or other areas where there is a need to screen large groups of people in very short time span, diagnostic testing and manual infrared thermal screening may be too difficult. These systems will likely miss most individuals with nCOVID-19, who are contagious. TIS could be considered as one of the promising method for initial temperature assessment to tackle the risk management. Non-contact infrared thermometers (NCIT) can measure surface temperatures without contact, in a single location and generally operated by a person at the entrance. TIS can measure temperature differences across multiple locations, creating a relative temperature map of a region of the body.We present the schematic representation of capturing thermal image in Figure 5.The model captures thermal images of the entrants' exposed upper backs (over the lungs) using a portable thermal camera that connects to any computing machine. The system requires the use of a calibrated blackbody (infrared temperature sensor) at once.

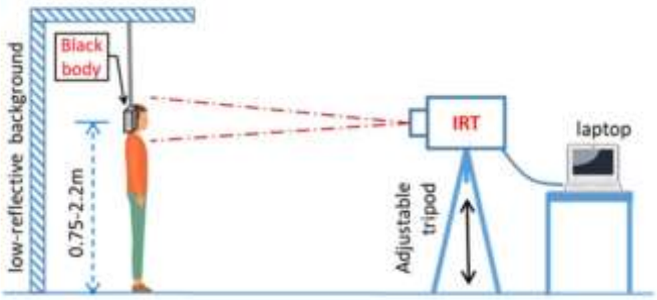

Figure5: Schematic representation of thermal imaging

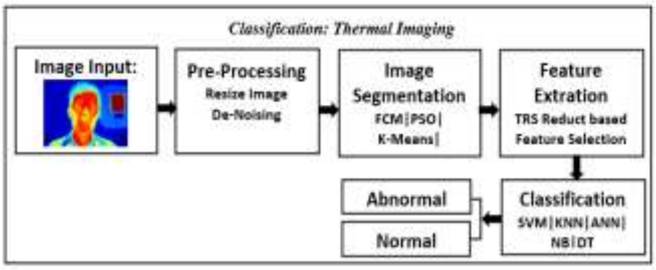

Figure 6: Architecture of Thermal Imaging Model

The input thermal image is passed through the pre-processing phase to get reduce the image into blocks and filters the noise associated with the image. The architecture of the model is shown in Figure 6. The sub images are extracted over 100 texture and shape features of temperature distribution across the skin. The fractal dimension of the gradient and Hofferian parameters compare with the rest of the cohort (n=101). Sensitivity, specificity and AUC were evaluated on both parameters to find the thermal scores. It is to note that, a combined cutoff of either a low FD score (<=1.82) and/or low HF score (<=13.5) identifies a 92% sensitivity and 62% specificity in detecting the color code for abnormal cases. The feature extraction phase adapts TRS based quick reduct method and allows classifying into the case of normal or abnormal type in accordance to the colour pixels associated with the image.

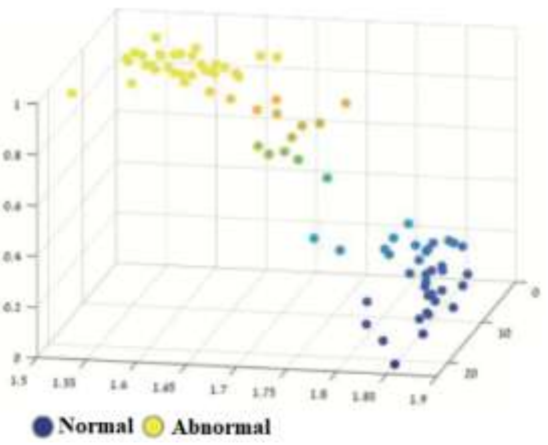

Figure 7: Classification Output of Thermal Imaging

VI. CONCLUSION

This work employed rough set theory as a mathematical tool for feature selection methods. Tolerance rough set found to be promising in handling real-valued data. As healthcare based datasets have high dimension attributes, the TRS based reducts offer promising results for feature selection process in machine learning based classification systems. The discussed method is also used in nCOVID19 chest X-Ray image classification to identify whether the input image is of normal or abnormal type.The thermal imaging process exhibits an alternative solution to identify and make a model for auto identification at public. The limitation of study is based on the specified local and surrounding parametric controls.

REFERENCES

[1] A. Hassanien, "Classification and feature selection of breast cancer data based on decision tree algorithm," Int. J. of Studies in Informatics and Control, vol. 12, no. 1, pp. 33–39, 2003.

[2] A. Hassanien, J. Ali, "Image classification and retrieval algorithm based on rough set theory," South African Computer Journal, vol. 30, pp. 9–16, 2003.

[3] A. Narin, C Kaya, Z. Pamuk, "Automatic Detection of Coronavirus Disease (COVID-19) Using X-Ray Images and Deep Convolutional Neural Networks", 2020.

[4] E.E.D. Hemdan, M.A. Shouman, M.E. Karar, "COVIDX-Net: A Framework of Deep Learning Classifiers to Diagnose COVID-19 in X-Ray Images, 2020.

[5] E. Pesce. S. W. Joseph, P.P. Ypsilantis, R. Bakewell, V. Goh, G. Montana, "Learning to detect chest radiographs containing pulmonary lesions using visual attention networks", Medical Image Analysis, 2019.

[6] G.K. Panda, I.K. Sahu, S.K. Das, "A Rough Set Approach for Knowledge Acquisition on Vector-Borne Diseases", IEEE Conf., 2nd Int. Conf. On Advanced Computational and Communication Paradigms, ICACCP-2019, Sikkim, India.

[7] I.K. Sahu, G.K. Panda, S.K. Das, "Rough Set Classifications and Performance Analysis in Medical Healthcare", Proceedings in Progress in Advanced Computing and Intelligent Engineering, Springer, vol. 1199, pp. 411-422, 2020.

[8] J. Bezdek, "What is computational intelligence?" in Computational Intelligence: Imitating Life", J. Zurada, R. Marks, and C. Robinson, Eds. Piscataway, NJ: IEEE Press, pp. 1–12, 1994.

[9] J.P. Cohen, P. Morrison, L. Dao, "COVID-Chestxray Database", https://github.com/ieee8023/covid-chestxray-dataset.

[10] J. Peters, S. Shahfar, S. Ramanna, T. Szturm, "Biologically-inspired adaptive learning: A nearset approach," in Frontiers in the Convergence of Bioscience and Information Technologies, Jeju City, Korea, pp. 403–408, 2007.

[11] J Peters, W. Pedrycz, "Computational intelligence," Electrical & Electronics Engineering Encyclopedia. NY: John Wiley & Sons, Ltd.

[12] J. Peters, "Perceptual granulation in ethology-based reinforcement learning," in Handbook on Granular Computing, W. Pedrycz, A. Skowron, and V. Krinovich, Eds. West Sussex, England: John Wiley & Sons, Ltd, 2008, pp. 671–689.

[13] L. Wang, A. Wong, "A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest Radiography Images, 2020.

[14] N. Karssemeijer, M. Thijssen, J. Hendriks, E. L. VanErning, "Digital Mammography: Computational Imaging and Vision", The Netherlands: Kluwer, 1998.

[15] P. Sprawls, "The physical principles of medical imaging", 2nd ed., 2018.

[16] R. Setiono, "Generating concise and accurate classification rules for breast cancer diagnosis," Artificial Intelligence in Medicine, vol. 18, no. 3, pp. 205–219, 2000.

[17] R. Swiniarski, "Rough sets and bayesian methods applied to cancer detection," in First Int. Conf. on Rough Sets and Current Trends in Computing, pp. 609–616, 1998.

[18] S. Das, A. Abraham, S. Sarkar, "A hybrid rough set–particle swarm algorithm for image pixel classification," in Proc. of the Sixth Int. Conf. on Hybrid Intelligent Systems, 2006, pp. 26–32.

[19] S. Purkayastha, S.B. Budi, S. Nuthakki, B Yadav, J.W. Gichoya, "Evaluating the implementation of deep learning in LibreHealth Radiology on Chest X-Rays", Advances in Intelligent Systems and Computing, 2020.

[20] S. Tsumoto, "Mining diagnostic rules from clinical databases using rough sets and medical diagnostic model," Information Sciences: An International Journal, vol. 162, no. 2, pp. 65–80, 2004.

[21] S. Widz, D. Slezak, K. Revett, "Application of rough set based dynamic parameter optimization to mri segmentation," in IEEE Annual Meeting of the Info. Processing NAFIPS, vol.1, pp.440– 445, 2004.

[22] T.B. Chandra, K Verma, "Pneumonia Detection on Chest X-Ray Using Machine Learning Paradigm", Proceedings of third Intl. Conf. on Computer vision & image processing, Springer, pp.21-33, 2020.

[23] V. Chouhan, S.K. Singh, A. Khamparia, D. Gupta, P. Tiwari, C. Moreira, "A novel transfer learning basedapproach for pneumonia detection in chest X-ray images". Applied Sciences (Switzerland), 2020.

[24] W. Pedrycz, "Computational Intelligence: An Introduction", BocaRaton, FL: CRC Press, 1998.

[25] Y. Yao, "A comparative study of fuzzy sets and rough sets," Journal of Information Sciences, vol. 109, pp. 227–242, 1998.

[26] Z. Pawlak, "Rough Sets", Int. J. Inf. Comp. Sc., II, pp. 341-356, 1982.
[27] Z. Pawlak: Rough Sets: Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, 1991.
[28] Z. Pawlak, A. Skowron, "Rudiments of rough sets," Information Sciences, vol. 177, pp. 3–27, 2007.
[29] https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)
[30] https://archive.ics.uci.edu/ml/datasets/parkinsons
[31] https://www.kaggle.com/uciml/pima-indians-diabetes-database
[32] https://github.com/harpreetSinghGuller/Data-Science-R/blob/master/SAHeart.csv
[33] https://github.com/ieee8023/COVID-chestxray-dataset

Author Profile

Indrani Kumari Sahu received the MCA from IGNOU and MBA degree from Berhampur University in 2002. She is with the MCA department of Khallikote (Auto) Govt. College, Berhampur, Odisha, India. She is continuing her research in the field of rough set theory, healthcare diagnosis systems as a research scholar in P.G. Department of Computer Science, Berhampur University, Odisha, India. To her repute, she holds certified UGC-NET qualification. She is a life time member of Computer Society of India.

Dr. Susant Kumar Das is with Berhamp-ur University, Odisha, India as Reader, senior grade in the PG Dept. of Computer Application. In his 23 years of post PhD academic carrier, most of his students and scholars have benefited in the courses of PG DCA, MCA, MTech, PhD and Post Docs. He supervised one Post Doc, fourteen PhDs, and several MCA/MTech dissertation thesis. To his credit he lead as head of the department, actively directed self financing course of MTech(CS), and coordinating the spoken tutorial project in consultation with IIT Bombay (funded by MHRD, Govt. of India). He contributes his research expertise as invited speaker in several international conferences, seminars and has numerous publications in the field of network security, software engineering and medical diagnostic systems. Recently, he felicitated by Dept. of Mathematics, Maharshi Dayanand University Rohtak, Haryana, India.