**Supplementary Information for the paper Rough Set Based Feature Selection Model for Diabetic Retinopathy Classification**

**1. Introduction of Rough Set Theory:**

According to the Rough Set Theory theoretical idea, each item in the world of discourse is believed to be specific knowledge or learning (RST). The objects categorized by similar information are ambiguous. If the items are ambiguous or imprecise, the set is rough. A limit district of a (unpleasant) set expresses ambiguity as articles that can’t be classified with certainty or acknowledged as having a place with the set in RST. The fundamental ideas of Rough Set Theory are exhibited through a little model.

* 1. **Relative indiscernibility and decision feature:**

Let us consider, a decision system DS as DS = (U, F) where U is the universe, a finite set of objects, U = {P1, P2, … Pn} and F is the set of features such that F = C ∪ D and C ∩ D = ∅ where C and D are the set of conditional features/features and the decision features/features, respectively.

For every attribute a ∈ F signifies an information function: fa: U → Va , where Va signifies the set of values of a, named as domain of attributes. Each subset of attributes S regulates an indiscernibility relation over U, and is rep- resented as IND (S), which can be defined by Equation 1.

I N D(S) = {(p, y) ∈ U 2 |∀a ∈ S, a(p) = a(y)} (1)

If (p, y) IND(S ), then x and y are indiscernible with respect to feature set S. IND (S ) is called the S -indiscernibility relation and the equivalence classes of the S -indiscernibility relation are denoted by [p]S. Here relative indiscernibility relation gives relative indiscernible objects based on each conditional feature, relative to decision feature D. Every conditional feature Fi of C determines n relative indiscernibility relation (RIR) over U relative to D, and is denoted as RIRD S (Fi), which can be defined by Equation 2(shown in Table 2).

RIRD (Fi) = {(p,y)X = …(2)

To illustrate the concept of relative indiscernibility, a sample medical dataset is considered in Table 1 with eight patient objects {p 1, p 2, p 3, p 4, p 5, p 6, p 7, p 8}, four conditional features {d, a, h, e} and one decision feature/feature (D) with their respective values.

Table 1: A sample decision system with eight objects, one decision (class) feature and four conditional features

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Objects** | **Pressure**  **(d)** | **Age**  **(a)** | **Heredity**  **(h)** | **Eye sight**  **problem**  **(e)** | **Decision**  **(D)** |
| p1 | High | Middle | Yes | Major | Y |
| p2 | High | Junior | Yes | Minor | N |
| p3 | Medium | Junior | Yes | Less Major | N |
| p4 | Low | Senior | Yes | Minor | N |
| p5 | Low | Middle | Yes | Minor | N |
| p6 | Low | Senior | Yes | Major | N |
| p7 | High | Senior | No | Less Major | Y |
| p8 | Medium | Junior | No | Major | N |

**2. Feature Similarity Measurement**

A feature Fi is connected to another feature Fj if under their respective relative indiscernible relations, they convince the same equivalence classes of objects. Because this is unusual, attribute similarity is computed by integrating the Similarity measurement factor, which determines how similar two characteristics are. In this case, a feature Fi is said to be similar to another feature Fj with the degree of similarity y and is denoted by Fi →Fj If the chance of generating the same equivalence classes of objects under their respective relative indiscernible relations is high enough, , ( , 100) percent, then Fj , where is computed using Equation 3.

**=** **…….(3)**

The specifics for calculating similarity measurement for feature similarity Fi →Fj (Fi ≠ Fj ) is described in algorithm SFACTOR below. Table-1 lists feature similarity and its similarity factor for all features in Table 1 to demonstrate the feature similarity computation procedure.

The computation of of each feature similarity using Equation 3 can be understood by Table 3, in which similarity d → a in RST row of Table 3 is considered, where, UD/d = {{ p1, p7}, {p2}, {p3, p8}, {p4}, {p5, p6}} and U D/a = {{p1}, { p5}, {p2, p3, p8}, {p4, p7}, {p6}}.

3. Generation of Feature Similarly Graph:

The Feature Similarity Graph is a weighted directed graph that is produced based on the similarity of characteristics (FSG), G = (V, E), is a directed graph with V and E as the set of vertices and edges, respectively. The “SFACTOR” algorithm, as described above, computes the similarity factor for each pair of conditional features (Fi, Fj). Higher the similarity factor of Fi to Fj is higher means that the relative indiscernibility relations RIRD(Fi) and RIRD(Fj) produce more similar equivalence classes. This implies that both the features Fi and Fj have almost similar classification power and so Fi → Fj is considered as strong similarity of Fi to Fj. FSG is the complete directed (shown in Fig 1) graph where each and feature are connected to each other and the edge between a pair of features is the similarity factor. Table 2 shows the decision-making system: For each feature, the left column shows the equivalence classes determined by the indiscernibility relation IND (P) described in Equation 1. For each conditional feature, the right column shows the equivalence classes determined by the relative indiscernibility relation RIRD (Fi) described in Equation 2.

4. Computation of reduct from Feature Similarly Graph:

The Following a greedy approach, significant nodes(features) are selected from the Feature Similarly Graph (FSG). Several centrality metrics of graph are available such as out-degree, in-degree, weighted out-degree, weighted in-degree, PageRank, eigen value centrality and . The greed algorithm identifies a node (feature) n having highest centrality measure such as highest in-degree or highest page rank value. Once any node (feature) is included in the reduct set, it’s adjacent edges are removed from the graph as it’s dependent node is included in reduct set. This process continues until and unless all the edges are removed from the graph. For each above-mentioned centrality metrics similar greedy approach is followed and eight different reduct sets have been generated. Each reduct set consists of a sub set of features from the Feature Similarly Graph (FSG).

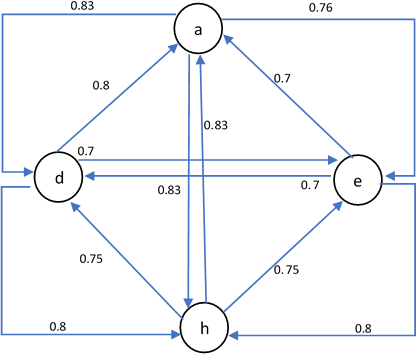


Fig. 1: Feature similarity graph for the dataset in Table 1

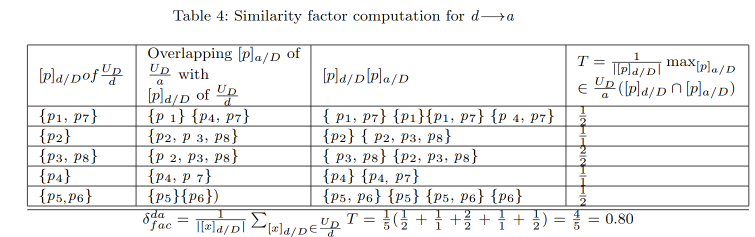
Table 2: Table 1 shows the decision-making system: For each feature, the left column shows the equivalence classes determined by the indiscernibility relation IND (P) described in Equation 1. For each conditional feature, the right column shows the equivalence classes determined by the relative indiscernibility relation RIRD(Fi) described in Equation 2.

|  |  |
| --- | --- |
| **Here, equivalence classes by**  **Indiscernibility relation IND (P) defined in Equation (1) for each feature are:** | **The equivalence classes by relative indiscernibility relation RIRD(Fi) defined in Equation (2) for each**  **conditional feature are:** |
| U/D = {{p1, p4, p7}, {p2, p3, p5, p6, p8}} |  |
| U/d = {{p1, p2, p7}, {p3, p8}, {p4, p5, p6}} | RIRD(d) = {{p1, p7}, {p2}, {p3, p8},{p4}, {p5, p6}} |
| U/a = {{p1, p5}, {p2, p3, p8}, {p4, p6, p7}} | RIRD(a) = {{p1}, {p5}, {p2, p3, p8},{p4, p7}, {p6}} |
| U/h = {{p1, p2, p3, p4, p5, p6}, {p7, p8}} | RIRD(h) = {{p1, p4}, {p2, p3, p5, p6},{p7}, {p8}} |
| U/e = {{p1, p6, p8}, {p2, p4, p5}, {p3, p7}} | RIRD(e) = {{p1}, {p6, p8}, {p2, p5}} |

Table 3: Degree of similarity of all pair of features

|  |  |  |  |
| --- | --- | --- | --- |
| **Feature Similarity**  **(Fi 🡪 Fj)** | **Equivalence Classes by RIRD (Fi) (UD/Fi)** | **Equivalence Classes by RIRD (Fj) (UD/Fj)** | **Similarity factor Fi to Fj ()** |
| d🡪 a | {p1,p7} , {p2},{p3,p8},{p4},{p5,p6} | {p1},{p5} , {p2,p3,p8},{p4,p7},{p6} |  |
| d🡪 h | {p1,p7} , {p2},{p3,p8},{p4},{p5,p6} | {p1, p4},{p2,p3,p5, p6},{p7},{ p8 } |  |
| d 🡪 e | {p1,p7} , {p2},{p3,p8},{p4},{p5,p6} | {p1},{ p6 ,p8 },{p2,p5 },{p4},{ p3},{ p7} |  |
| a 🡪 d | {p1},{p5} , {p2,p3,p8},{p4,p7},{p6} | {p1,p7} , {p2},{p3,p8},{p4},{p5,p6} |  |
| a 🡪 h | {p1},{p5} , {p2,p3,p8},{p4,p7},{p6} | {p1, p4},{p2,p3,p5, p6},{p7},{p8} |  |
| a 🡪 e | {p1},{p5} , {p2,p3,p8},{p4,p7},{p6} | {p1},{p6 ,p8 },{p2,p5 },{p4},{p3},{p7} |  |
| h 🡪 d | {p1, p4},{p2,p3,p5, p6},{p7},{ p8 } | {p1,p7} , {p2},{p3,p8},{p4},{p5,p6} |  |
| h 🡪 a | {p1, p4},{p2,p3,p5, p6},{p7},{ p8 } | {p1},{p5} , {p2,p3,p8},{p4,p7},{p6} |  |
| h 🡪 e | {p1, p4},{P2,p3,p5, p6},{p7},{ p8 } | {p1},{p6 ,p8 },{p2,p5 },{p4},{p3},{p7} |  |
| e 🡪 d | {p1},{ p6 ,p8 },{p2,p5 },{p4},{ p3},{ p7} | {p1,p7} , {p2},{p3,p8},{p4},{p5,p6} |  |
| e 🡪 a | {p1},{ p6 ,p8 },{p2,p5 },{p4},{ p3},{ p7} | {p1},{p5} , {p2,p3,p8},{p4,p7},{p6} |  |
| e 🡪 h | {p1},{ p6 ,p8 },{p2,p5 },{p4},{ p3},{ p7} | {p1, p4},{p2,p3,p5, p6},{p7},{p8 } |  |

Table 4: Similarity factor computation for d 🡪 a



End if End for

End for

**Algorithm 1: SFACTOR (Fi,Fj)** */\* Similarity factor computation for Fi* 🡪 *Fj \*/*

**Input:** Feature Fi and Fj.

**Output:** Similarity factor

Begin

For every conditional attribute Fi do

Determine relative indiscernibility RIRD (Fi) using Equation 2 RIRD (Fi) induces equivalence classes UD/Fi = [𝒑]𝑭𝒊/𝑫

End for

*/\* similarity measurement of Fi to Fj \*/*

For each [𝒑] €

𝑼

𝑫

𝒊

do

𝑫

𝑭𝒊

max\_overlap=0

For each [𝒑] €

𝑼

𝑫

𝒋

𝑫

𝑭𝒋

overlap = **|** 𝜹 𝒊 ∩ 𝜹 𝒋 |

𝑫 𝑫

if ((overlap > max\_overlap)) then

max\_overlap = overlap

𝜹𝑑𝑎

𝒇𝒂𝒄

= 𝜹𝑑𝑎 + 𝒎𝒂𝒙\_𝒐𝒗𝒆𝒓𝒍𝒂𝒑

𝒇𝒂𝒄

|[𝒑] 𝒋 |

𝑫

𝛿𝑑𝑎 =

𝛿𝑑𝑎

𝑓𝑎𝑐

𝑓𝑎𝑐

| 𝑈𝐷 |

𝐹𝑖

End

**5. Standard feature selection approaches:**

1. **Correlation Feature Selection (CFS):**

This method measures a reduct’s quality by calculating the predictive power of each individual characteristic as well as the amount of redundancy among them. Subsets of characteristics with a high degree of interconnection with the categorization are chosen.

## **Consistency Subset Evaluation (CON):**

This approach predicts the effectiveness of reduct using the level of reliability/consistency in the values of the class. The subset evaluator is usually used in incorporation with search methods- such as exhaustive search, random search, best first, rank search, and so on - which finds the least subset with uniformity same as that of complete set of features. Best First search is used here for the experiments.

**6. Experimental Dataset Description:**

**a. Graphical features from medical imaging:**

These includes Fovea, Blood vessel, Optic distance, Exudate number, Blot hemorrhages, Bifurcation between two blood vessel, Edema, Kapur entropy, Shannon entropy, LBP entropy, Renyi’s entropy, Micro aneurysm, Exudates Area and LBP energy.

**b. Medical historical features:**

Aside from diabetic imaging features, we consider some aspects of the patient’s background, such as age, genetic information, and duration of diabetes. Spots or dark lines floating in your field of vision (floaters), fluctuating vision, Blurred vision, dark or vision loss, colour vision impairment and empty areas in vision. The later section describes the efficiency of the proposed rough set-based feature reduction strategy using experimental datasets.

Table 5: List of features for the diabetic dataset

|  |  |  |
| --- | --- | --- |
| **Sl.** | **Features** | **Brief Description** |
| **Graphical features from medical imaging** | | |
| 1 | Optic distance | The point where the optic nerve strands depart the retina is known as the optic separation or Optic Nerve Head. |
| 2 | Fovea | The area’s cone cells are connected to ganglion cells,  resulting in fovea vision, which is the retina’s sharpest centre vision. |
| 3 | Blood vessel | When delicate veins rupture beneath the tissue that covers the white of the eye (conjunctiva), resulting in eye redness, it’s possible that you have a sub-conjunctival channel. |
| 4 | Blot hemorrhages | They are not illness components without anyone else’s input; rather, they’re indications of visual and fundamental diseases |
| 5 | Exudate number | Exudate refers to any fluid that leaks from the circulatory system into wounds or regions of disturbance. |
| 6 | Edema | It is the advancement of uid in the macula, a district in the point of convergence of the retina |
| 7 | Bifurcation between two blood vessel | Bifurcation is the split of a vein into at least two branches. |
| 8 | Shannon entropy | Shannon entropy is a standard measure of the amount of information in a communication. |
| 9 | Kapur entropy | It’s utilised to come up with a good picture segmentation approach. Different pictures are chosen for trials in order to evaluate the performance of the suggested method. |
| 10 | Renyi’s entropy | Renyi’s entropy detects collectibles, including eye glimmers, in the regular course of things. |
| 11 | LBP entropy | LBP is an unrivaled textural depiction chairman who requires an immaterial proportion of calculation and is especially impenetrable to light impedance |
| 12 | LBP energy | LBP energy is often low in glaucoma, giving it an alpha LBP characteristic, indicating that it has a coarser textural range than the average class. |
| 13 | Microaneurysm | MA is a actually little aneurysm, or swelling, which is placed in the side of a vein. Microaneurysms are at times found in the retina of the eye  in people with diabetes. |
| 14 | Exudates Area | Hard exudates are the yellow dots. It’s the lipid residues from shattered vessels. |
| **Medical historical features** | | |
| 15 | Age | Age of the patient |
| 16 | Genetic information | Genetic information regarding parental diseases |
| 17 | Duration of diabetics | How long patient is suffering from the disease |
| 18 | Spots or dark strings floating in your vision (floaters) | Specific eye problem of floating vision |
| 19 | Blurred vision | Blurriness in vision of the patient |
| 20 | Fluctuating vision | Fluctuation in vision of the patient |
| 21 | Impaired color vision | Problem in identification of objects |
| 22 | Dark or empty areas in your vision | Visualization of spots, specks, and other things in dark bits |
| 23 | Vision loss | Loss of vision of the patient |