Clustering model for progression of ocular hypertension to primary open angle glaucoma based on rough mereology

Mahmood A. Mahmood1, Mohamed Abd Elfattah2, A. A. Abd El-Aziz1

1 ISSR, Computer Sciences and Information Dept., Cairo University, Cairo - Egypt
mahmoodissr@cu.edu.eg, a.ahmed@cu.edu.eg
2 Faculty of Computers and Information System, Mansoura University, Egypt

Abstract—This paper presents clustering model for progression of ocular hypertension to primary open angle glaucoma using classification approach based on rough mereology. Our model comprises two phases: Clustering and prediction. We start by Clustering that utilises a rough mereology theory for clustering into one of three obtained classes, namely low, mid, and high and generates its association rules, followed by a prediction phase. Experimental results show that the employed rough mereology provides prediction accuracy 92.85%, which is nearest to the prediction accuracy of random forest classifier are 93% but the performance of rough mereology classifier O(n^2) is less than performance of random forest classifier O(n^3 logn), also is better prediction accuracy compared to other machine learning techniques including Bayes net, multi-layer perceptron, radial basis function and naive Bayes tree classifiers.

Keywords-Rough Mereology; Rough Inclusion; Rule Generation; Glaucoma; primary open angle glaucoma; retinal fiber layer; machine learning; pattern classification; random forest classification.

I. INTRODUCTION

Glaucoma is a term describing a group of ocular disorders with multi-factorial etiology united by a clinically characteristic intraocular pressure-associated optic neuropathy[1]. Glaucoma is the second most frequent cause of blindness in the United States with about two million subjects diagnosed and many others undiagnosed, is the leading cause of blindness in African Americans [3], [4] and represents a significant cause of vision-related disability [5].

Glaucoma(Def.) describes a group of eye diseases whose common feature is the gradual loss of optic nerve tissue. Glaucoma diagnosis is one of the most important domains that can benefit from the application of data mining and pattern recognition techniques[2].

The increasing chance of Glaucoma disease is increased by intraocular pressure (IOP), the disease can also exist in the setting of normal IOP [6], another studies have suggested that central corneal thickness, also considered as another risk factor for for development and progression of glaucomatous optic neuropathy.

We need early screening and detection of risk factors which cause of glaucoma to avoid blindness, since glaucoma-induced blindness is irreversible. Primary open-angle glaucoma (POAG), the most common type of glaucoma [8], leads to the destruction of ganglion cells and related axons within the optic nerve [15], [16].

The main contribution in this paper is summarized as follow: a) presents a computer aided diagnosis system(CAD), which interventions of theory of rough mereology in the medical field(Glaucoma eyes). b) we propose prediction model with high accuracy.

The rest of this paper is structured as follows. Section II describes the previous work in this area. Section III presents the proposed approach for Prediction. Section IV presents the experimental result. Finally, Section V addresses conclusions and future work.

II. RELATED WORK

In the earliest studies glaucoma have been defined as the elevation of eye pressure above a certain level and specified six features to classify glaucoma namely symptom, presence of elevated intraocular pressure (IOP), cause of elevated IOP, level of IOP, clinical entity, cause for damage [9].

By 1960 many studies[10], [11], [12]articulated that IOP is not the best discriminant feature that we can depend on to diagnose glaucoma. another studies considered disk characteristics is on of the other entities related to the optic nerve damage to diagnose glaucoma [9], [13].

In[7] proposed an prediction mode using the previous five factors that associate with the glaucoma risk to build a predictive model to estimate the risk that ocular hypertension patients will be primary open angle glaucoma (POAG). While in [14] tested the efficiency of artificial neural network to discriminate between normal and eyes with glaucoma. The average sensitivity of neural network was 65% and a specificity of 71%. Recently, our previous work results shows how the

III. CLUSTERING MODEL BASED ON ROUGH MERELOGY

The Clustering Model based on Rough Mereology

...
among parameters. Then, the granulation mechanism that reflects a given set of inclusion data into collection of granules will be applied via voting by training objects in order to produce the optimal similarity measurement. Algorithm (1) shows the detailed steps of the sample dataset.

**Algorithm 1 Classification Approach based on Rough Mereology**

1: Input: An information System table ($IST$)
2: Output: Rule granule set ($RGS$)
3: Initialize set of Rule sets $SRS = \emptyset$
4: for Each radius from 0 to 1 step 0.1 do
5:     Initialize Rough Inclusion set $RIS = \emptyset$
6:     for Each column in ($IST$) do
7:         Compute rough inclusion and put the values in $RIS$
8:     end for
9:     Initialize Rule generation set $RGS = \emptyset$
10:    for Each row in $RIS$ do
11:        Get the rule according to definition of rough inclusion as shown in [21], and put the rule in $RGS$
12:        Remove duplicates rules from $RGS$
13:    end for
14:    Store $RGS$ in $SRS$
15: end for
16: Initialize $ACCRateset = \emptyset$
17: for Each rules in $SRS$ do
18:     compute Accuracy measure rate and put in $ACCRateset$ as shown in [21]
19: end for
20: Initialize $bestACC$ as the first element in set $ACCRateset$
21: for Each element in $ACCRateset$ do
22:     if element $\in bestACC$ then
23:         $bestACCset = element$
24:     end if
25: end for
26: for Each element in $ACCRateset$ do
27:     if element $= bestACC$ then
28:         Store element in $bestACCset$
29:     end if
30: end for
31: get the corresponding set of rules of $bestACCset$ and put in $RGS$
32: apply voting by object equations as shown in to refine $RGS$
33: Exit

The clustering model used the theory of rough mereology to cluster the dataset into sets of generated rules in three phases as shown in figure (Fig. 1).

The classification model based on rough mereology works in three phases; namely pre-processing, clustering, and voting by objects phases as shown in figure (Fig. 2).

1) **Pre-processing phase**: maps the experienced dataset into a normalized dataset of the range $[0,1]$, where the values 0 and 1 represent the smallest and the largest values in each dataset’s attribute, respectively.

2) **Clustering phase**: uses theory of rough mereology and rough inclusion technique to classify the dataset into sets of granules with different radius.

3) **Voting by objects phase**: applies voting by objects to select the optimized granules of dataset.

**A. Pre-processing phase**

Pre-processing phase takes dataset as a data input in the matrix form, called rating matrix, so that the proposed approach normalize this matrix by finding the largest value and the smallest value in each attribute, which are called $max_a$ and $min_a$, respectively. The produced new matrix, Normalized Matrix, contains values in the range $[0,1]$ and is calculated by the following equation.

$$NormalizedV_a = \frac{V_a - min_a}{max_a - min_a}$$

Where $V_a$ represents the value of an attribute, $min_a$ represents the smallest value of an attribute, $max_a$ represents the largest value of an attribute.

**B. Clustering phase**

Clustering phase received the normalized rating matrix and produced list of sets of granules that describes rule generation according to different radius $r$ in granules. This phase consists of two stages; namely 1) rough mereology and 2) list of granules. For the rough mereology stage, it takes the normalized rating matrix as input and applies the equations of rough mereology with radius $r$. The output of this stage is rough inclusion table that represents similarity measure of each
object in each attribute of the dataset. In the list of granules stage, a set of rough inclusion table is computed by re-applying the first stage of rough mereology ten times with different radius of \( r \) from 0.0 to 1.0 with step 0.1 [21].

C. Voting by objects

Voting by objects phase takes as an input the list of similarity measure tables for each radius resulted from the clustering phase and computes the accuracy rate for each radius as shown in equation (2). Then, the optimum radius that represent the largest accuracy measure is selected, and if there exist more than one largest accuracy measure then the proposed approach generate set of largest accuracy measure. This phase is divided into two stages; namely 1) accuracy measure computation, which uses equation (2) to compute the accuracy for each table representing the similarity measure at radius \( r \) and 2) optimization stage where the table containing the largest accuracy measure at radius \( r \) is selected, so the radius \( r \) called the optimum radius \( r_{opt} \).

\[
AR = \left( \frac{T}{N} \right) \times 100 \tag{2}
\]

Where \( AR \) is the accuracy rate, \( T \) is the number of similar tuples and \( N \) is the total number of tuples in each row [21].

D. Rule Generation

The Apriori algorithm [22] is a classical algorithm that used to generate rule mining by received then minimum support \( \epsilon \), minimum confidence \( \eta \), and dataset then output all association rule.

IV. EXPERIMENTAL RESULT

A. Data Collection

To make our CAD system with high accuracy in detection glaucoma, we must combine between of structural and functional, so practitioners must determine whether the patient has glaucoma or is at risk of developing glaucoma. The clinician must combine many aspects of information, including clinical observations, family history, medical history, diagnostic test results, prior examination findings and other related information [3], [4], [5].

Our CAD system is based on analysis of collected data from randomly selected healthy individuals from a defined catchment area and from glaucoma patients followed at the department of Ophthalmology at Egypt Air Hospital. The study includes 398 participants older than 40 years [7] and the demographic of our dataset shown in figure (Fig. 4).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>40</td>
<td>75</td>
<td>55.93</td>
<td>10.071</td>
</tr>
<tr>
<td>IOP</td>
<td>23</td>
<td>30</td>
<td>26.666</td>
<td>0.149</td>
</tr>
<tr>
<td>CUP/Disc</td>
<td>450</td>
<td>656</td>
<td>574.452</td>
<td>34.826</td>
</tr>
<tr>
<td>CCT</td>
<td>1</td>
<td>2.2</td>
<td>1.3020</td>
<td>0.228</td>
</tr>
<tr>
<td>PSD</td>
<td>-10</td>
<td>-0.06</td>
<td>-2.36</td>
<td>1.471</td>
</tr>
<tr>
<td>RNFL</td>
<td>0</td>
<td>1</td>
<td>0.073</td>
<td>0.26</td>
</tr>
</tbody>
</table>

TABLE I

The Demographic Data

B. Risk Factors for glaucoma development

For building an automatic predictive model requires complex steps which needs acquisition and analysis of data from multiple longitudinal studies. For a disease, there are a group of component causes that collectively lead to an immediate cause to produce the disease. If every group of component causes always contains one particular cause, then that factor constitutes a necessary cause. For each disease, it is also important to define the effect to which the causes lead. In other words, we require a disease definition. Until recently, such a definition for POAG was largely subjective and not subject to rigorous scientific standards [17].

The concept of risk factors is important, because each factor represents a potential target for specific treatment or prevention. Risk factors describe features that may be causal in disease, as they are statistically associated with the disease, and that were present before its occurrence, and thus possibly played an essential role along with other factors in incident disease [18].

Some large, prospective, longitudinal studies have provided evidence regard risk factors for conversion from ocular hypertension to glaucoma. From these, two were randomised clinical trials, the ocular hypertension treatment study (OHTS) [18], and the European glaucoma prevention study (EGPS) [19], which have provided the basis for development and validation of prediction models for glaucoma development. Both studies have evaluated a large number of predictive factors for their potential association with the risk of conversion to glaucoma.

When pooled analysis of OHTS and EGPS data were conducted, five baseline factors were identified as significantly associated with the risk of conversion to glaucoma [20]: central corneal thickness, intraocular pressure (IOP) measurement of the vertical cup/disc ratio of the optic nerve, age and visual-field pattern standard deviation as shown in figure (Fig. 5).

C. Decision Rules

1) if Age and PSD then no Risk (86.05)
2) if IOP then no Risk (94.18)
3) if Cup/Disc and CCT then no Risk (97.14)
4) if RNFLT then no Risk (94.03)
According to the results of decision rules of rough mereology classifier in figure (Fig. 6)[23], rule one presents the Age and PSD parameters shared in the decision of risk value with 13.95%, rule two presents the IOP parameter shared in decision of risk value with 5.82%, rule three presents CUP/DISC and CCT shared in decision of risk value with 2.86%, and rule four presents the RNFLT shared in decision of risk value with 5.97%, so figure (Fig. 7) describe the dependency of risk value on Age and PSD, IOP, CUP/DISC and CCT, and RNFLT in descending order.

Figure 6 shows the confusion matrices and performance measures of different classifiers including naive Bayes tree pattern classification algorithms(NP/Tree), multi-layer perceptron (MLP), Bayes Net, radial basis function (RBF) and random forest. The evaluated measures are precision, recall, F-measure, and area under ROC curve [24]. We can observe from the results.

V. CONCLUSIONS

In this paper, we have presented Clustering model for progression of ocular hypertension to primary open angle glaucoma employing based on rough mereology. Experimental results showed that the random forest approach provided improved prediction results with an overall accuracy of 93% but the complexity time is $O(n^3 \log(n))$, and our classification model give overall accuracy of 92.85% with complexity time $O(n^2)$ compared with 75.6% for Bayes net, 87.2% for multi-layer perceptron, 79% for radial basis function and 82% for Naive Bayes tree classifiers.

REFERENCES


