Journal of Theoretical and Applied Information Technology

30th September 2023. Vol.101. No 18

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DIAGNOSING THE MEDICAL DATA USING ROUGH SET MIN- MAX CLASSIFIER

S. DEVI[1], Dr.V. SASIREKHA[2] , K.VEENA[3]

[1]Assistant Professor, Department of MCA, SSM College of Engineering, Tamil Nadu, India.

[2]Assistant Professor, Department of Computer Science, J.K.K. Nataraja College of Arts and Science, Tamil Nadu, India.

[3]Assistant Professor, Department of Computer Science, J.K.K. Nataraja College of Arts and Science, Tamil Nadu, India.

E-mail: [1]devikrishnamca@gmail.com, [2]sasirekhailangkumaran@gmail.com, [3]veenabharathi44@gmail.com

ABSTRACT

In today's medical disciplines, multiple massive quantities of data are being released, all of which constitute information on patients, diseases, and doctors. Disease The diagnosing process is one of the most important steps that requires a more costly examination. The illness outcome has been successfully predicted using a variety of different methodologies already in existence. However, it is far less capable of managing the huge and intricate medical dataset. A brand new Rough Set Min-Max Classifier (RMMC) is put to use in this approach that has been suggested in order to make illness forecasts. Based on the Euclidean distance measurement, the RMMC model describes the neighborhood connection between the two sets of instance data. This approach makes effective use of rough set theory, and the outcomes of the method's evaluation are examined using three distinct medical datasets. The experimental outcome of the RMMC technique that was presented is compared with the Neighborhood Rough Set Classifier (NRSC) algorithm, which stands for the Neighborhood Rough Set Classifier. The suggested technique achieves 99.42% accuracy in illness prediction, which is confirmed by the k-fold cross-validation, and has thus become a potential tool for diagnosing medical datasets. This is in comparison to the previous method, which only achieved 99.24% accuracy in disease prediction.

Keywords: *Rough Set, RMMC, NRSC, Euclidean Distance, Medical Diagnosis.*

1. INTRODUCTION

In today's highly modern world, the development of the computerized database system lends help to the decision-making and diagnostic processes involved in the medical dataset. The clinical expertise is able to make more informed judgments as a result of the analysis of the medical dataset performed by a knowledge-based system. The clinical dataset consists of a variety of different pieces of information on the patient and their illness. It is simple to develop a diagnostic, perform pattern recognition, and provide a prediction of the required work when using classification algorithms. The sole application of the rough set theory that can be found in the present methodologies is feature selection in medical diagnostics. An approach that works with the information learned from the relationship database is called rough set theory. It is equivalent to the fuzzy theory in a mathematical sense. The structural link of the imprecise data may be established by the use of the rough set technique. It is an extension of the classical sets that take into account their complementarity. When compared to a rough set, a fuzzy set makes use of partial memberships, while a rough set makes use of numerous memberships. This distinction is what distinguishes the fuzzy set from the rough set.

1.1 Rough Set Theory

A basic understanding of the topic at hand may be gained by doing a Set analysis. It helps the process of extracting features from the data that is provided and provides a mathematical tool for assessing the hidden pattern in the data. It is able to reduce unnecessary data and rapidly detect the dependent relationships between the data. As was just said, it does an examination of the database in

order to discover the dependence values of the attributes. Find the value of each variable's dependence after doing some rough processing on each variable. There are certain undecipherable things that have been included in the database and have been replicated on several occasions. These items have the potential to trigger a redundant act, which forces the system to keep certain characteristics consistent according to their neighborhood connection. It is referred to be reduct when a database has a feature set that is comprehensive enough to completely characterize the database. As a consequence, this collection of characteristics is referred to as an adequate set of features to predict the outcome. In order to properly depict the reduct, there are a few criteria that it must have. These traits include being minimum and not being unique.

1.2 Characteristics of Rough Set Theory

- In most cases, the rough set is already quite mature owing to the mathematical structure that underpins it, and previous knowledge is not required in any way.
- Because of its straightforward nature, determining the value is a straightforward endeavor.
- The rough set model is capable of processing many different forms of data, despite the fact that the data is unfinished and imperfect.
- Each idea is broken down into its component parts at varying granularities with the bare minimum of expression.

In most cases, the techniques of data analysis have difficulty resolving certain issues throughout the process of result prediction. These issues include recognizing the interdependence among the characteristics, minimizing the number of redundant attributes, pinpointing the attributes that are most important, and developing a decision rule. The learning idea as well as locating the underlying patterns in the dataset are the topics that are covered by the rough set theory. The element of the boundary line that decides whether an item is part of a set or none at all is the one in charge of making that determination for the system. It is used for a variety of purposes, including the creation of decision rules, data reduction, feature extraction, and feature selection. It does this by determining the comparable classes within the training data. Some techniques of categorization are unable to differentiate between the classes on the basis of their characteristics. The rough set, on the other hand, divided the classes into two distinct groups, such as lower and higher approximation. The data tuples that do not belong to the classes are included in the lower approximation, whereas the higher approximation comprises all of the data tuples that are considered to be part of the data based on the attribute knowledge. It provides the system with the help it needs to choose the significant characteristic from the dataset and decrease the data that is irrelevant. It lessens the amount of computing work that has to be done by the system. The learning for the classification process comes from the attribute set that comprises the preset classes. This learning is driven by the decision attributes. The learning process is carried out by the learning algorithm, which is then applied to the medical dataset. The effectiveness of the algorithm is evaluated by using the testing dataset. An Rough Set Min-Max classifier is used in this investigation to identify the various illnesses present in the dataset. This classifier enables the system to choose just the qualities that are necessary depending on the criteria that are most significant.

The aim of this research is to address the challenges posed by the vast and intricate medical datasets found in today's healthcare field. These datasets contain a wealth of information on patients, diseases, and medical practitioners. The primary focus is on improving the process of disease diagnosis, which often involves costly examinations. While various methods have been developed to predict disease outcomes successfully, they struggle to handle the sheer scale and complexity of medical data.

To overcome these challenges, the researchers introduce a novel approach called the Rough Set Min-Max Classifier (RMMC). This approach utilizes the principles of rough set theory and relies on Euclidean distance measurements to model relationships between different sets of instance data. The central goal is to enhance the accuracy and efficiency of illness forecasting using this new classifier. To evaluate the effectiveness of the RMMC model, the researchers conducted experiments using three distinct medical datasets. They compare the results of the RMMC technique with those obtained from the Neighborhood Rough Set Classifier (NRSC) algorithm, which serves as a benchmark in this context.

The remaining portions of the paper are structured as follows, In the second section, we discussed the work that was done by other researchers in the field. Within this part, the system

architecture of the suggested model is broken down and described. 3. The section discusses the findings of the experimental procedures. 4. A discussion of the conclusion may be found in section 5.

2. RELATED WORK

An article on diagnosing medical data with the use of a rough set was proposed by Manimaran and colleagues [1]. He introduced a rudimentary set-based neural network with the purpose of lowering the amount of time needed for decision calculation. The network is able to accomplish the network gain with the assistance of the rough set, which assists the network in removing the undesirable properties of relations from the dataset. The Wisconsin Breast Cancer Dataset was the one he utilized. Marayam et al. [2] published a study where they advocated utilizing the rough set technique for picture retrieval. He is of the opinion that the dataset including ambiguous and imprecise information should also be used for analytical purposes. The uncertain dataset may provide the rough set with information that is helpful to recover. He examined the accuracy of the rough set in comparison to the results of a variety of different classifiers, including Naive Byes, Support Vector Machine (SVM), and Decision tree. Testing and training procedures are carried out using the Corel Image dataset. Khannan et al. [3] came up with the idea of combining the backpropagation algorithm with an indiscernibility relation technique. For the purpose of illness prediction, he used a clinical dataset. Pahulpreet et al. [4] published a work on machine learning, in which they analyzed a medical database using a variety of categorization methods. The p-value test is used to choose which characteristics to get from the database. A study on the subject of forecasting diabetic Mellitus was given by Baiju et al. [5]. He investigates the publications based on diabetes and the datasets that belong to them. He makes his prognosis based on the Disease Influence Measure (DIM). A work on a rough set perspective was provided by Shusaku et al. [6], and in it the authors said that the author employed both upper and lower approximation while screening and diagnosing patient data. A article on the use of soft computing methods was provided by Pradipta et al. [7]. He is of the opinion that the rough can cope with the uncertainty that is related with the medical data. Analytical methods such as the fuzzy set help limit the amount of data that is erratic and overlaps with other data. A publication written by Hong et al. [10], which represents his work, describes how he partitions brain pictures by utilizing a clustering technique and rough set. A study proposing an overview of the rough set theory was written by Zhang et al. [11]. He went through the fundamental ideas as well as the methodology behind the crude set theory. Udhaya et al [8] published a study on medical diagnostics in which the author used a unique approach of local rough set categorization. Due to the fact that the rough set may include both continuous and decision datasets. He contrasted the outcome of the study with a number of other approaches already in use and made use of five distinct medical datasets. A work on extracting the characteristics from the medical dataset was provided by Devi and colleagues [9]. She used the Epileptic Seizure Recognition Medical Data Set, from which 87 characteristics are retrieved in order to determine which features are regarded to be the most relevant for the analysis process. Concerns were raised by the author over the use of Euclidean distance in the process of developing the neighborhood rough set for illness prediction. In the current models, it is necessary to have an exact rule in order to continue with the algorithm; the rule must be different for each model in order to accurately forecast the outcomes. This results in the present model not being as accurate as it might be.

Ishii et al. [12] propose the Directional Neighborhood Rough Set approach as a solution to address issues related to Generalized Rough Sets. This approach introduces new concepts like information granules, lower and upper approximations, and a three-step classification algorithm. The authors conducted experiments to validate the effectiveness of their approach using real-world machine-learning dataset. Hardani et al. [13] conducted a study with the objective of evaluating the effectiveness of the rough set approach for feature selection in the diagnosis of breast cancer cases. They performed feature selection on the Wisconsin Breast Cancer (Diagnostic) Data Set available from the UCI machine learning repository. The research encompassed various stages, including data pre-processing, feature selection, data randomization, classification, and performance assessment, all aimed at achieving these research goals. A sizable ruleset is set aside just for the purpose of processing the testing and training data. Evaluation of the performance of the present model is done based on the rule. However, the model that is being suggested to forecast the neighborhood association between the qualities makes use of some kind of metric function. It might seem that the Euclidean distance formula is inaccurate, which would result in a failure to provide the required outcome. Implemented a new formula for Euclidean distance to measure the neighborhood

relation, and a tiny ruleset is used for evaluation reasons, both of which are detailed in further detail in section 3. This was done so that the necessary output could be obtained from the training dataset by means of a rough set.

3. PROPOSED METHOD

Take for example a dataset in the medical field that has both continuous and decision features. A crude set analysis, which is shown in Table 1, includes the presentation of some sample data for better understanding the progress being made. A{a1,a2} are examples of conditional characteristics, whereas D{D1, D2} are examples of decision attributes. Together, these 87 features were used by our model to make predictions about the outcome.

Calculate the values of the distance metric for each record of an attribute by using the Euclidean Distance [8]. After doing the calculation to determine the neighborhood connection based on θ(x1) and setting the parameter to 0.1. The neighborhood connection for characteristics changed depending on the value which separated.

Table 1: Sample Dataset of both continuous and discrete data.

Records $X_1 \in U$	a_1	a_2	a_3	a_3	D
X_1	135	190	229	223	4
X_2	386	382	356	331	5
X_3	-32	-39	-47	-37	3
X_4	-105	-101	-96	-92	3
X_5	-9	-65	-98	-102	3
X_6	55	28	18	16	4
X_7	-55	-9	52	111	3
X_8	1	-2	-8	-11	3
X_9	-278	-246	-215	-191	1
X_{10}	8	15	13	3	3

In order to improve overall performance by analyzing the property values in the surrounding area, a new Euclidean formula, equation (1), was put into place.

$$F(xi, xj) = \sum_{k=1}^{n} \sqrt{\left(\left(x_i(k) - x_j(k)\right)^2\right)} \quad (1)$$

The Euclidean distance is calculated for the sample dataset (Table 1), and the condition is satisfied for the new Euclidean distance formula, θ(x1)={x1,x2}.

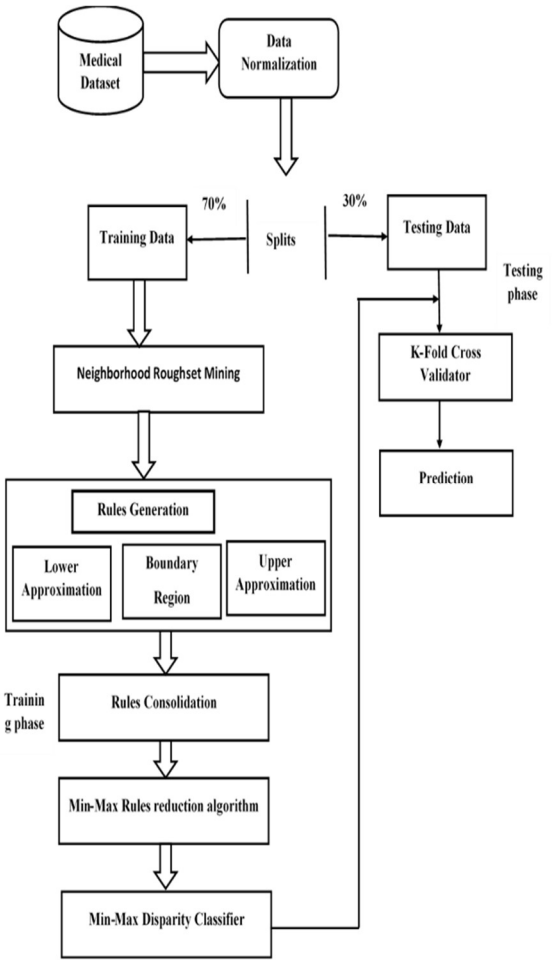

Figure 1: Block Diagram of Proposed Model

Designing the research for proposed Rough Set Min-Max Classifier" is a critical and foundational step in ensuring the study's success and the validity of its findings. A well-thought-out research design serves as a roadmap for conducting the research, collecting and analyzing data, and drawing meaningful conclusions. The suggested system's block diagram may be seen in Figure 1, which can be found here.

The medical dataset is collected from the repository maintained by the University of California, Irvine (UCI). In the beginning, the Normalization process is used to perform the task of changing negative numbers into positive ones. After that, the dataset is partitioned into data for training and data for testing. Calculations of the Neighborhood and Equivalence relations for

attributes are performed on the training data. After that, the "and" operator is used to determine the values of the granules that make up the conditional characteristics. When evaluating the conditional characteristics, the lower approximation is used. When evaluating the decision attributes, however, the higher approximation is used. For both the higher and lower approximation, the computer will automatically construct a rule. The rule that applies to the border area of the data is generated by combining the value derived from the upper and lower approximation. The data are grouped together according to three consolidated rules. A calculation is done based on the cluster rule to determine the minimum value and the maximum value of the conditional characteristics. At long last, the RMMC algorithm is executed, and the minimum-maximum disparity is computed so that the outcome may be evaluated. The outcome is then judged based on how well the test data were applied. For the purpose of determining how accurate the suggested model is, k-fold cross-validation is used.

The Rough Set Min-Max classifier algorithm is explained below,

Algorithm: Rough Set Min-Max classifier

Input: $(U, A \cup D),\ \theta$
Output: Predicted Result (PR)
Step 1: To evaluate the neighborhood relationship among the attributes.

$$f(xi, xj) = \sqrt{\sum_{k=1}^{n} (Xi - Xj)^2}$$

Form the above notation f defined as a metric function of universe U, and (U, f) is the neighborhood relation...$\theta(xi) = \{x | f(xi, xj) \leq \theta, x \in U\}$

Step 2: Construct the equivalence relation for the decision attribute.

$$IND(Dx) = \{x \in \cup \mid \forall d \in Dx, d(x) = d(y)\}$$

Step 3: Apply "Λ" ("and") operator of the neighborhood granules.

$$\theta(Ai \wedge Aj) = \{x | f(Ai, Aj) \leq \theta, Ai \in U, Aj \in U\}$$

Step 4: Construct the neighborhood rough set lower approximation ($\underline{N}D_x$) space for decision attributes $for\ all\ D_x$

$$\underline{N}D_x = \{x_i \mid \theta_B(x_i) \subseteq X, x_i \in \cup)\}$$

Step 5: Construct the neighborhood rough set upper approximation ($\overline{N}\ D_x$) space for decision attributes $for\ all\ D_x$

$$\overline{N}\ D_x = \{x_i \mid \theta_B(x_i) \cap X \neq \emptyset, x_i \in \cup)\}$$

Step 6: Find the boundary region of data set by using neighborhood rough set boundary region ($BNR(D)$)

$$BNR(D) = \overline{N}\ D_x - \underline{N}D_x$$

Step 7: Generate certain rules using Neighborhood rough set based lower approximation ($\underline{N}R$)

$$\underline{N}R = \left\{ f: (\underline{N}D_x \cup) \rightarrow xi \right\}$$

Step 8: Generate the possible rules using Neighborhood rough set based on Upper Approximation ($\overline{N}R$)

$$\overline{N}R = \{ f(\overline{N}\ D_{x1}, \cup) \rightarrow xi \}$$

Step 9: Generate the possible rules using the Neighborhood rough set based Boundary region (BR).

$$BR = \{ f(BNR, \cup) \rightarrow xi \}$$

Step 10: Consolidate the rules by merging lower, upper and boundary rules,

$$R = \underline{N}R + \overline{N}R + BR$$

Step 11: For Every Decision, Attribute applies the Min-Max Rule Reduction algorithm.

$$RR = \{(\min(r), max(r)) \in R \mid \forall Di\}$$

Step 12: Predict Result for test data using below Min Max Mean Disparity Classifier

$$\text{dpv} = \frac{\sum_{i=0}^{n} \left| T_i - \left(\frac{r_{i\,min} + r_{i\,max}}{2} \right) \right|}{N_a(R)}$$

$Ti = i^{th}\ attribute\ in\ test\ data$

$r_{i\,min}$ $- Minimum\ value\ of\ i^{th}\ attribute\ in\ RR$

$r_{i\,max}$ $- Maximum\ value\ of\ i^{th}\ attribute\ in\ RR$

$N_a(R) = No\ of\ attributes\ in\ RR$

The process flow of the new method is shown in figure 2, which can be found here. The RMMC model is evaluated using these three distinct datasets in order to identify the rate of performance it achieves. The value that is read from the input is normalized, which changes any negative values to positive ones. Using a ruleset, the data are consolidated into a single group by means of the

attributes' values and the similarity index. When a Rule cluster is being constructed, the Min-max value of the features being used is computed and then applied to the Min-max discrepancy. At long last, a result has been hypothesized based on the test results. The k-Fold cross-validation method is used in order to determine how accurately the RMMC algorithm performs. The suggested RMMC algorithm's pseudo-code is shown and discussed further below.

Pseudo-code: RMMC Algorithm

Input: d: (A ∪ D) θ: 0.1
Output: PR : Predicted Result

$d_{train} \rightarrow d(7:10)$
$d_{test} \rightarrow d(3:10)$

foreach item **in** d_{train}:

$\theta [x_1 : x_{n-1}]$
← Calculate NHR for Conditional attr

$D[x_n]$
← Calculate ER for decision attr

θ[x1 ∧ . . xn]
← Calculate "
∧ " NHR for Conditional attr
end foreach

foreach r in R:
RRtemp ← Reduce Ruleset for Target

RR ← append (RRtemp)

end foreach

foreach tx **in** d_{text}:
foreach r **in** RR:

dpv ← MMMC(r, tx)
PR[r] ← dpv

end foreach
end foreach

Sort PR in ascending based on disparity
Set result ← select top one from PR
return result

end Procedure.

Variable Definition:

$\theta[x_1 : x_{n-1}]$: Neighborhood Relation array of individual Conditional attributes

$D[x_n]$: Equivalence relation for Decision attributes

$\theta[x_1 \wedge .. xn]$: Neighborhood relation array of combined conditional attributes

$\theta \underline{N} D_x$: Lower Approximation

$\overline{N}\ D_x$; Upper Approximation

$BNR(D)$: Boundary Region

$\underline{NR}$: Lower Rule

$\overline{NR}$: Upper Rule

BR: Boundary Rule

R: Consolidated Rule

RR : Consolidated Reduced Ruleset

dpv : Disparity value

PR: Predicted Result

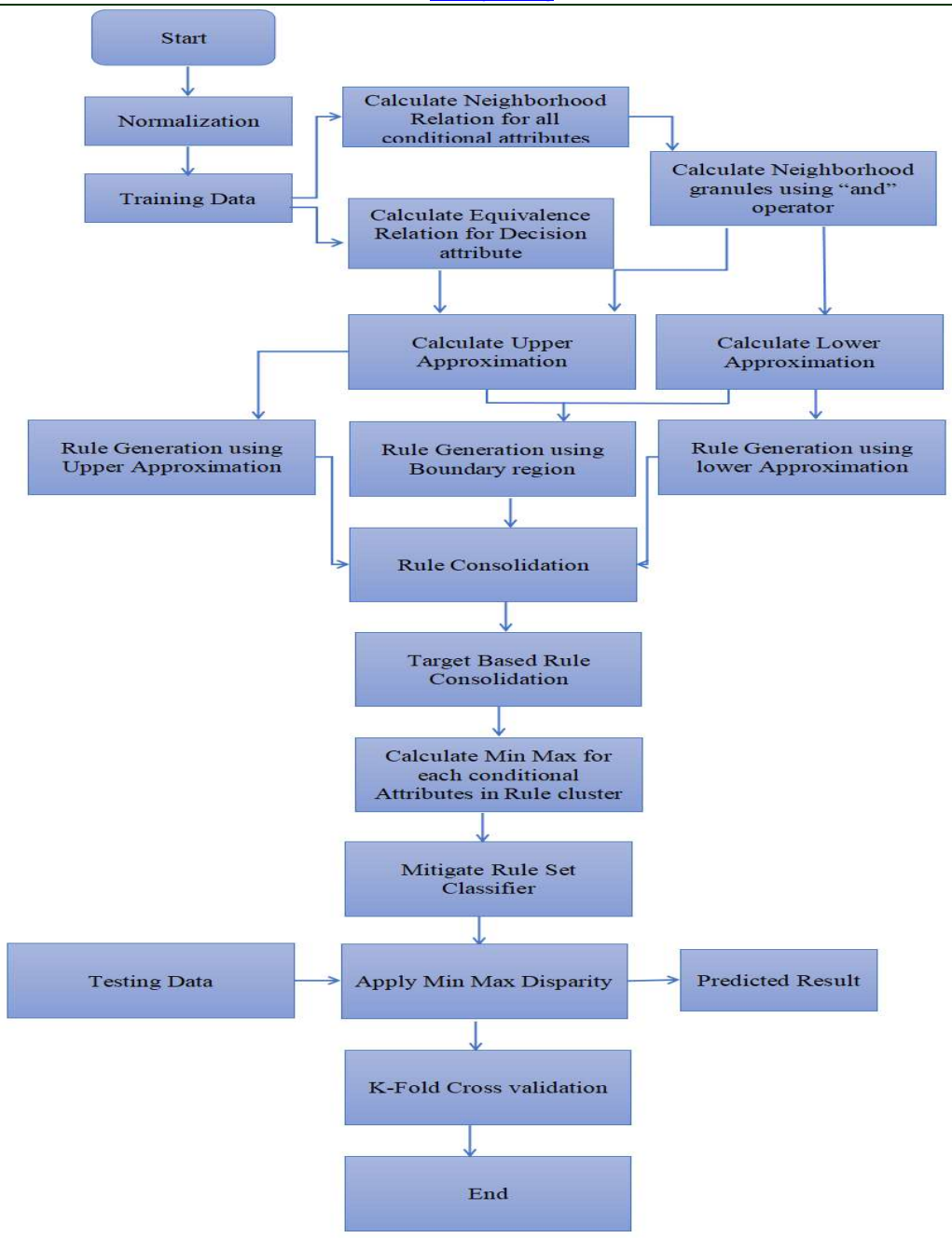

Figure 2: Flow chart of Proposed Model

The pseudo-code variables are explained and represented the procedure of the RMMC algorithm to predict the result. Initially, the $d: (A \cup D)$ is set as 0. 1 the result must achieve this condition. It is considered as the input for the pseudo-code. The output will be predicted results. Then for training and testing, took some data from the data set in the ration of $d_{train} \rightarrow d(7:10)$, $d_{test} \rightarrow d(3:10)$. Consider a dataset contain numerous data where the training process takes 70 percent of the data and for testing 30 percent of the data. Here splitted the data into two parts for training and testing procedure. The Neighborhood relationship is calculated among the individual condition attributes $\theta\ [x_1 : x_{n-1}]$ using for loop condition. Then for the decision attributes equivalence relation also

calculated $D[x_n]$. After applying these conditions an array is developed to hold all the attributes $\theta[x1 \wedge .. xn]$ values. The R is predicted by reducing the ruleset for the target. The lower approximation ($\underline{N}D_x$) and upper approximation ($\overline{N}\ D_x$) is calculated. $BNR(D)$ Based on the values of the attributes boundary region is calculated. The lower rule, upper, and Boundary rules are represented as $\underline{N}R$, $\overline{N}R$,and BR. Consolidation of these three rule emerges the final rule R ($\underline{N}R + \overline{N}R + BR$).

Any kind of data may be processed using the ruleset technique so that the outcome can be predicted, which is one of the many benefits of using this approach. When reducing the number of rules, the goal reduction approach is utilized. Because of this, the ruleset takes up a very little amount of space, which in turn makes the mathematical computation simpler. This provides substantial support for the enhancement of the suggested model's accuracy.

4. RESULTS AND DISCUSSION

The results of the simulation run on the suggested system are provided below utilizing an i5 computer with 4 gigabytes of RAM and the Windows 10 operating system. Python is used as the primary implementation language for the RMMC algorithm's core functionality.

4.1 Dataset

In order to evaluate how well the RMMC algorithm works, two distinct medical datasets from the actual world are employed. Both the Epileptic Seizure Recognition Data Set [14] and the Breast Cancer dataset were retrieved from a repository at the University of California, Irvine. In order to accurately anticipate the outcome, the suggested classification system must first be trained and then tested k times.

4.2 Performance Analysis

Python is used as the programming language for the implementation of the aforementioned RMMC model and assessment of the model's performance. In order to perform data processing, the Epileptic Seizure Recognition Data Set is first imported into the system. Fig. 3 is a representation of the Graphical User Interface (GUI) that the suggested model would have. The dataset is then normalized via a method known as min-max normalization. The dataset is processed such that a normalized value may be obtained. After that, a preliminary set rule that is based on the dataset values is automatically created by the system. It specifies the maximum and minimum values that may be approximated. After the rough set rule has been produced, the database is split into test and training data. Once again, the Rough set rule is constructed depending on the values in the neighborhood. The ruleset is simplified as a result of the consolidation of the regulations. RMMC model's work performance is represented in figure 3, which can be found here. The data from training are taken and put through their paces in the testing phase. The RMMC achieves a higher level of accuracy (99.32%) than the model that was previously used. The findings have shown that the accuracy of the RMMC model is much higher than that of the Neighborhood Rough Set Classifier model. The performance of the proposed model is evaluated using three distinct data sets, including breast cancer, liver, and the Epileptic Seizure Recognition Data Set, as shown in Figure 3. These datasets have the RMMC and NRSC graphs shown for them respectively.

We validated the RMMC suggested model using k-Fold Cross Validation to see how accurate it was. Figure 4 illustrates it for us. As a result of the study, it was determined that the RMMC had an accuracy of 99.2 percent. It is sufficient evidence to demonstrate that the RMMC is one of the potential options that might lead to improved accuracy in illness prediction.

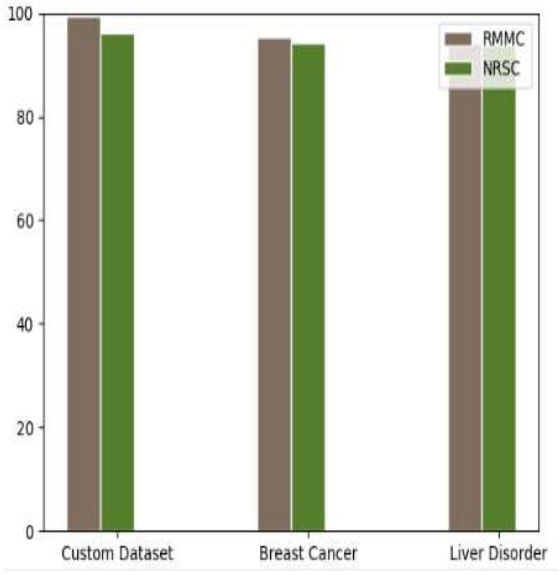

Figure 3: Performance Metric

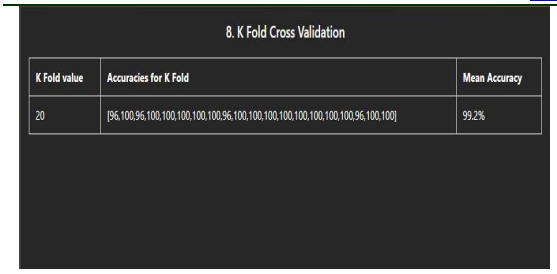

Figure 4: K-Fold Cross Validation

On the basis of the examination of several metrics, the effectiveness of the model that was presented is examined. Data for both the RMMC and NRSC algorithms are computed by using true positive and true negative values as the foundation. Table 2 provides an illustration of the formula for the validation measure.

TABLE 2: Performance Validation Measures

S.No	Validation Measure	Formula
1	Precision	TP/(TP+FP)
2	Recall	TP/(TP+FN)
3	F -measure	(2*Precision*Recall) /(Precision +Recall)
4	Fowlkes Mallows index	$\sqrt{\text{precision*Recall}}$
5	Kulczynski index	(1/2)(Precision +Recall)
6	Rand index	(TP+FN)/(TP+TN+FP+ FN)

Calculations are made for the RMMC algorithm as well as the NRSC method. These calculations include precision, recall, F-measure, Fowlkes Mallows index, Kulczynski index, and the Rand index. According to Table 3, it is shown that the RMMC model accurately predicts more positive outcomes than the NRSC model does. The Rand index determines the degree to which two sets of data are clustered together. The RMMC obtains a score of 0.98, which is representative of the model's quality. In contrast to the NRSC model, the RMMC can simply group together values that are comparable across the index. Thoroughly validating and cleaning collected data can help identify and rectify errors or outliers that may introduce unnecessary variations. Implementing this proposed method with k-fold validation can enhance the internal validity and reliability of our studies and ensure that research findings are not unduly influenced by unnecessary variations.

TABLE 3: Performance Analysis Of The Proposed Classification Algorithm And Other Comparative Algorithms

Medical Dataset	Classification algorithm	Precision	Recall	F -measure	Fowlkes Mallows index	Kulczynski index	Rand index
Epileptic Seizure Recognition Data Set[12]	**RMMC** NRSC	0.996 0.991	0.96 0.94	0.97 0.964	0.978 0.965	0.978 0.965	0.989 0.97
Liver Dataset[8]	**RMMC** NRSC	0.996 0.60	0.96 0.81	0.97 0.72	0.97 0.73	0.97 0.72	0.98 0.42
Breast Cancer Dataset[8]	**RMMC** NRSC	0.87 0.82	0.97 0.88	0.91 0.89	0.923 0.903	0.92 0.91	0.77 0.75

Table 4: Performance Analysis Of The Proposed Method

S.No	Classification Method	Accuracy
1	K-Nearest Neighbour algorithm (KNN)[8]	41.70
2	Support Vector Machine (SVM)[8]	50.14
3	NRSC[8]	96.94
4	**RMMC(Proposed)**	**99.42**

Table 5: Accuracy Comparison

Medical Dataset	Classification algorithm	Accuracy (%)
Epileptic Seizure Recognition Data Set	**RMMC** NRSC	99.42 96.5
Breast Cancer Dataset	**RMMC** NRSC	95.22 94.32
Liver Dataset	**RMMC** NRSC	94.60 93.62

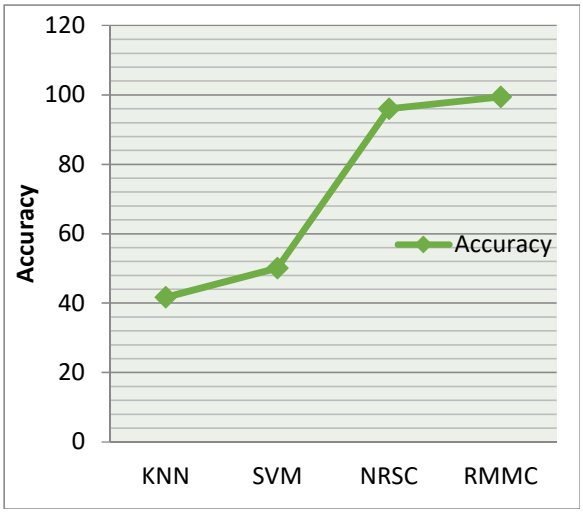

Figure 5: Graphical representation of Accuracy comparison

The accuracy of the already used classification methods is analyzed and contrasted with that of the newly developed algorithm. The RMMC was able to achieve a higher level of accuracy (99.42%) than the NRSC (96.94%), the KNN (41.70%), and the SVM (50.14%), as shown in Table 5. A graph representing the accuracy value has been drawn, and it can be found in figure 5.A very accurate prediction of the illness may be made using the model that was presented.

5. CONCLUSION

In the model that we have suggested, we have applied the RMMC algorithm, which is widely regarded as an effective strategy for addressing the challenge of medical diagnosis. The rough set is able to analyze the complicated information and make an accurate prediction using it. The effectiveness of the RMMC algorithm is measured and compared to that of several other algorithms already in existence. The proposed model is assessed using three distinct medical datasets, and the results reveal that it achieves a 99.42% accuracy rate in illness prediction when compared to the current approach. In the end, we use k-Fold Cross-validation to determine whether or not the model that was provided is accurate. The findings have shown that the suggested model is capable of achieving a greater level of accuracy than models such as NRSC. Therefore, our model, which we have suggested, is quite beneficial for medical professionals and will be of use to them in making judgments about illness prediction. Evaluation of the optimization technique that improves the accuracy of the RMMC result is going to be the focus of a future inquiry. During this period, there will be a decrease in the amount of computational work spent diagnosing. The diagnostic procedures are becoming more accurate, which contributes to an increased level of fidelity within the medical industry.

REFERENCES:

[1] A. Manimaran, V. M. Chandrasekaran, Aishwarya Asesh,"Rough set approach for an efficient medical diagnosis system", published in International Journal Of Pharmacy & Technology, April 2015.

[2] Maryam Shahabi Lotfabadi,Yongzhao Zhan, "Evaluating a Cover based Rough Set Classifier in a Content based Image Retrieval System", published in 14th International conference on Signal-Image Technology & Internet-Based System (SITIS),2018

[3] Khanna Nehemiah, Harichandran, Arputharaj, "Knowledge Mining from Clinical Dataset using Rough Sets and Back propagation Neural Network", published in Computational and Mathematical methods in medicine, 2018.

[4] Pahulpreet Singh, Shriya Arora, "Application of machine learning in disease prediction", published in 4th International conference on computing communication and Automation (ICCCA), 2019.

[5] B.V Baiju, D. John, "Disease influence Measure Based Diabetic Prediction with Medical Dataset using Data Mining", published in 1st International conference on innovation in information and communication technology, 2019.
[6] Shusaku Tsumoto,"Medical Diagnosis: Rough Set View", published in Part of the Studies in Computational Intelligence book series (SCI, volume 708),2017
[7] Pradipta Maji,"Advances in Rough Set Based Hybrid Approaches for Medical Image Analysis", published in International Joint Conference on Rough Sets IJCRS : Rough Sets,2017.
[8] S. Udhaya kumara, H. Hannah Inbarani,"A Novel Neighborhood Rough set Based Classification Approach for Medical Diagnosis", published in the Elsevier, 2015.
[9] S.Devi, V.Sasirekha, "An Experimental Analysis on Rough Set Mean, Median, Mode Method of Dependency Values for Feature Selection in Medical Databases", published Asian Journal of Computer Science and Technology ISSN: 2249-0701 Vol.8 No.S1, 2019, pp. 103-106
[10] Hong Huang, FanzhiMeng,Shaohua Zhou,Feng Jiang,Gunasekaran Manogaran "Brain image segmentation based on FCM clustering algorithm and rough set", published in 10.1109/ACCESS.2019.2893063, IEEE Access, 2019.
[11] Quingug Zhang, Qin Xie, Guoyin Wang, "A survey on rough set theory and its applications", published CAAI Transactions on Intelligence Technology Volume 1,Issues 4, 2016.
[12] Ishii, Y.; Iwao, K.; Kinoshita, T. A New Rough Set Classifier for Numerical Data Based on Reflexive and Antisymmetric Relations. Mach. Learn. Knowl. Extr. 2022, 4, 1065-1087. https://doi.org/10.3390/make4040054
[13] D N K Hardani and H A Nugroho 2020 IOP Conf. Ser.: Mater. Sci. Eng. 771 012017 DOI 10.1088/1757-899X/771/1/012017
[14] https://archive.ics.uci.edu/ml/datasets/Epileptic+Seizure+Recognition
[15] S. Senthil Kumar, H. Hannah Inbarani, Ahmad Taher Azar & Kemal Polat, "Covering-based rough set classification system", Published in New Trends in data pre-processing methods for signal and image classification, 14 June 2016.
[16] Chuan Luo,Tianrui Li,Hongmei Chen, HamidoFujita, "Incremental rough set approach for hierarchical multicriteria classification", published in Information Sciences, March 2018.
[17] Ahmad Taher Azar, H. Hannah Inbarani & K. Renuga Devi, "Improved dominance rough set-based classification system", Published in Neural Computing and Applications, 20 January 2016.
[18] S. Udhaya Kumar & H. Hannah Inbarani, "Neighborhood rough set based ECG signal classification for diagnosis of cardiac diseases", published in soft computing, February 2016.
[19] Fannia Pacheco, Mariela Cerrada, René-Vinicio Sánchez, Diego, "Attribute clustering using rough set theory for feature selection in fault severity classification of rotating machinery", published in Expert system with Application, April 2017
[20] Jothi G, HannahInbarani H., "Hybrid Tolerance Rough Set–Firefly based supervised feature selection for MRI brain tumor image classification", published in Applied Soft Computing,2016.
[21] K.Das, Shampa Sengupta, Siddhartha Bhattacharyya, "A group incremental feature selection for classification using rough set theory based genetic algorithm", published in applied soft computing, April 2018.
[22] Tapash Barman ; Rajesh Ghongade ; Archana Ratnaparkhi, "Rough set based segmentation and classification model for ECG", published in Conference on Advances in Signal Processing (CASP), 2016.