iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** V **Month of publication:** May 2018

DOI: http://doi.org/10.22214/ijraset.2018.5071

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue V, May 2018- Available at www.ijraset.com

Chikungunya Diagnosing using Rough Set Approach

Vinod Kumar Sharma[1], Simarjeet Kaur[2], Tushar Bhatia[3]
[1, 2, 3]*Department of Computer Science, Faculty of Engineering and Technology, Tantia University*

Abstract: ***Chikungunya (CHIK) virus, the same as infectious disease, a significant threat in Tropics and one in all the foremost widespread insect-borne diseases famous nowadays, owing to the presence of Aedes Aegypti mosquitoes. Once a year Chikungunya has been moving the several individuals. During this paper, Rough pure mathematics introduced by Pawlak(1982) is used to extract the choice rule from given information. During this approach call, rules area unit derived from conditional attributes in a rough set analysis in accounting victimization the rule. This important choice of attributes in Chikungunya knowledge set contributes to the great for knowledge unclearness and uncertainty in doubtless reducing knowledge assortment to manage inconsistent issues. The results of a study indicate that victimization this methodology will scale back surplus attributes for the Chikungunya diagnosing supported patient's condition and aim to enhance their detection performance. The deciding factors for Chikungunya area unit found to lead to characteristic itssymptoms. These may be used to predict the presence or absence of unwellness, Chikungunya. The system produces considerable accuracy and is a valuable tool for physicians.***

Keywords: ***Rough Set, Entropy, Call tables, Information gain, Identification and Rule extraction***

I. INTRODUCTION

Chikungunya could be an infective agent unwellness transmitted to humans by the bite of infected Aedes aegypti mosquitoes. Chikungunya virus (CHIKV) could be a member of the genus animal virus, within the family Togaviridae [1]. CHIKV was initially isolated from the blood of a feverish patient in the United Republic of Tanzania in 1953, and has since been known repeatedly in west, central and southern Africa and several areas of Asia, and has been cited because of the reason behind various human epidemics in those areas since that point. The infection circles all through loads of landmasses [1], with the transmission, thought to happen fundamentally amongst mosquitoes and monkeys.

The name Chikungunya springs from scholastic degree African dialect which implies " that curves up" or "stooped stroll" because of the unhealthful torment caused by the disease. Side effects epitomize fever and joint agony.These generally occur 2 to 12 days when exposure. Alternative symptoms could embody a headache, muscle pain, joint swelling, and a rash[3,4]. The general public area unit higher inside a week; but, often the joint pain could last for months. The danger of death is around one in 1,000. The terribly young, old, and people with alternative health issues area unit in danger of a lot of severe unwellness. They primarily bite throughout the day. The virus could flow into inside variety of animals as well as birds and rodents. Identification is by either testing the blood for the virus's RNA or antibodies to the virus. Whereas the unwellness generally happens in Africa and Asia, outbreaks are reported in Europe and also the Americas since the 2000s. In 2014 over 1,000,000 suspected cases occurred. In 2014 it absolutely was occurring in American state Florida within the continental United States however as of 2016 there was no additional domestically non-inheritable case. The term is from the Kimakonde language and means that "to become contorted". Data picked up all through late pandemics recommends that Chikungunya fever could lead on to an incessant area also inferable from the segment of an intense disorder. among the intense area, 2 arranges square measure recognized: Associate in Nursing operator organize all through the initial 5 to seven days, all through that pathology, happens, trailed by a healing stage enduring close to 10 days, throughout those symptoms improve and also the virus can't be detected within the blood. Typically, the unwellness begins with an unforeseen high fever that lasts from many days to per week, and typically up to 10 days. The fever is typically on top of thirty-nine °C (102 °F) and sometimes reaching forty °C (104 °F) and should be biphasic—lasting many days, breaking, then returning. Fever happens with the onset of pathology, and also the level of virus within the blood correlates with the intensity of symptoms within the acute section. When IgM an antibody protein that's a response to the initial exposure to an antigen, seems within the blood, pathology begins to diminish. However, headache, sleep disorder, an extreme degree of exhaustion stay, typically regarding five to seven days.

©IJRASET: All Rights are Reserved

A. *Chikungunya In Asian Country*

Chikungunya happening in Asian country first occurred in 2006. Cases were rumored wide from everywhere the landmass. The most cases were from Maharashtra though' there have been no deaths. In 2007, the most cases of Chikungunya were rumored in Kerala [5]. The cases started disappearing from 2008. As per the present scenario of the subcontinent, it looks that Chikungunya is more and more moving the population. This point around, the toughest HIT looks to be Delhi wherever they're several rumored deaths as a result of Chikungunya. The Delhi government is currently taking initiatives to stop their folks from Chikungunya.

II. ROUGH SET THEORY

Rough set theory, planned in 1982 by a Polish man of science Zdzisław Pawlak [7]. The rough set is taken into account one of the primary non-statistical approaches in information analysis and another intelligent information analysis tool that may be applied to handle unclearness and inconsistencies. Its methodology is bothered with the classification and analysis of inaccurate, unsure or incomplete info and information. (Pawlak, 1982). Rough pure mathematics is comparable to Fuzzy pure mathematics, but the unsure and inexactness during this approach is expressed by a boundary region of a collection, and not by a partial membership as in Fuzzy Set Theory [11]. The elemental conception of Rough pure mathematics is that the approximation of lower and higher areas of a collection, the approximation of areas being the formal classification of data concerning the interest domain. The set generated by lower approximations is characterized by objects that may positively kinda part of an interest set, whereas the higher approximation is characterized by objects that may presumably kinda part of an interest set. Each set outlined through higher and lower approximation is understood as Rough Set[18]. Mainly the rough set analysis works to determine the approximations of concepts.

Rough sets constitute a sound basis for KDD (Knowledge Discovery in Database). It makes one able to get patterns hidden in data by providing mathematical tools. It is used for feature choice, feature extraction, knowledge reduction, call rule generation, and pattern extraction (association rules) etc. It identifies partial or total dependencies in knowledge, eliminates redundant knowledge, offers an approach to null values, missing knowledge, dynamic knowledge et al. Rough set philosophy is supported on the idea that with each object of the universe of discourse some information (data, knowledge) is associated. For example, if objects area unit patients laid low with an explicit disease, symptoms of the disease type information concerning the patients. Objects characterized by a similar info area unit indiscernible (similar) visible to the obtainable information concerning them. The in-discernibility relation generated during this means is that the mathematical basis of rough pure mathematics. Rough pure mathematics has some blessings over different similar theories [23]. For example, it doesn't want any external information and is predicated exclusively on original knowledge. Its results area unit simply graspable even by non-experts. It's attainable to investigate qualitative attributes additionally to quantitative ones. Discovery of dependencies (full or partial) between attributes, removal of redundant knowledge not needed for analysis and generating call rules are different blessings of this theory.

A. *Information Table*

A table is a briefing of information in rows and columns, or probably in an exceedingly a lot of complicated structure. Here associate info table is viewed as a table, consisting of objects (rows) and attributes (columns)[12]. It is utilized in the illustration of data which is able to be used by Rough Set, where each object options a given amount of attributes (Lin. 1997). These objects unit of measurement represented in accordance with the format of the data table, among that rows unit of measurement thought-about objects for analysis and columns as attributes (Wu et al., 2004). Below is shown an associate example of associate info Table I.

Table I Information Table

Patient	Attributes			
	Fever	Joint Pain	Headache	Viral unwellness
P1	High	Yes	No	Yes
P2	High	No	Yes	Yes
P3	Very High	Yes	Yes	Yes
P4	Normal	Yes	No	No
P5	High	No	Yes	No
P6	Very High	Yes	No	Yes

©IJRASET: All Rights are Reserved

III. DECISION TABLE AND DECISION ALGORITHM

Decision tables are a precise yet compact way to model complex rule sets and their corresponding actions. A call table or choice table could be a two-dimensional table that demonstrates the move to be made after a progression of associated decisions. A call table or choice table contains two assortments of traits assigned as the condition characteristic and choice quality [13]. In Table, I, the attributes of fever, joint pain and headache will all be thought-about as condition attributes, whereas the infectious agent unwellness attribute is taken into account a call attribute or decision attribute. Every row of a call table determines a call rule, that specifies the choices (actions) that have to be taken once conditions area unit indicated by condition attributes are satisfied, e.g. In Table I the condition (Fever, high), (Joint pain, yes), (Headache, no) determines the choice (Viral unwellness, yes).

Table one suggests that each patient2 and patient5 experience the ill effects of same side effects since the condition properties of fever, joint agony and cerebral pain have indistinguishable qualities; at the same time, the estimations of call property demonstrates entirely unexpected qualities. This arrangement of tenets unit called either irregularity, non-determinant or clashing. These guidelines unit called consistency, determinant or non-clashing or only, a run the show.

The number of consistency rules, contained within the programing language unit called an element of consistency, that may be denoted by $\gamma(C, D)$, wherever C is that the condition and D the choice. If $\gamma(C, D) = 1$, the choice table is consistent, however, if $\gamma(C, D) \neq$ one the table of call is inconsistent.

Given that Table I, $\gamma(C, D) = 4/6$, that is, the Table I possesses two inconsistent rules (patient2, patient5) and four consistent rules (patient1, patient3, patient4, patient6), inside the universe of six rules for all the Table I

The decision rules are frequently shown as implications within the variety of "if... then..." [14].

To proceed is shown one rule for the implication infectious agent illness:

If
Fever = high and
Joint pain = affirmative and
Headache = low
Then
Chikungunya = affirmative

A set of call rules is selected as call algorithms, as a result of for every multidimensional language, it is related to the choice rule, consisting of all the choice rules that it occurs within the several multidimensional languages. A could also be created a distinction between call rule and call table. a call table could be a knowledge set, whereas a call rule could be an assortment of implications, that is, a logical expression.

IV. ENTROPY

Entropy could be a chance based mostly measure [15,17]. In scientific theory, Entropy is that the life of the uncertainty related to a variable quantity. Entropy is that the randomness collected by associate software system or application to be used in medical or different uses that need random knowledge. This randomness is gathered from pre-existing knowledge or specially provided randomness generators. while not the employment of entropy will have a negative impact on performance and security. Attribute choice in entropy supported minimizing associate info entropy live applied to the examples at a node[17]. Entropy has wide applied to several fields. The entropy live is employed to pick the attributes providing the very best information gain. An information set with some distinct valued condition attributes and one distinct valued call attributes is conferred within the variety of knowledge illustration system J=(U, CD), wherever U={u_1, u_2.....u_s} is the set of information samples, C={ c_1, c_2,..... c_n} is that the set of condition attributes and D= {d} is one elemental set with call attribute or category label attribute. Assume this category label attribute has m distinct values process m distinct categories, d_i (for i=1,2….m), and let s_i be the number of samples of U in class d_i.

The entropy for the subset is given by:

$\text{Entropy}(S) = -\sum_{i=1}^{m} p_i \log_2 p_i$

Where p_i is the probability that an object is in class i. Gain(S, A), an information gain of an example set S on attribute A is defined as, $\text{Gain}(S, A) = \text{Entropy}(S) - \sum((|S_V| / |S|) * \text{Entropy}(S_V))$

Where $\sum$ is each value v of all possible values of attribute A,

S_V= Subset of S for which attribute A has value v,

$|S_V|$= No. of elements in S_V,

|S|=Number of elements in S.

©IJRASET: All Rights are Reserved

V. OBSERVATION USING RST IN DECISION MAKING FOR CHIKUNGUNYA DIAGNOSIS

A. *Information System*

Initially, the information regarding the objects must be determined in the form of data table subject to an analysis using rough set methodology. An information table with seven conditional attributes [12], such as (a) Age distribution; (b) IgM : IgG ratio; (c) Symptoms etc. have been gathered from internet and doctors to identify effective parameters in relation to Chikungunya diagnosis. These attributes shown in Table II. The attributes of age distribution, IgM : IgG ratio, symptoms-I, symptoms-II, seasonal distribution, platelet count and clinical features can all be considered as condition attribute, whereas the viral illness attribute is considered a decision attribute. Each row of decision table determines a decision rule, which specifies the decisions (actions) must be taken when conditions indicated by condition attributes are satisfied. Each column of the mentioned table indicates one of the characteristics (attributes) considered for the diagnosis and the last column implies the decision parameter for Chikungunya fever.

Next, the data should be classified using rough set theory to analyze the information. Therefore each conditional attribute is provided as three classes low, medium and high and the decision level is also classified with low, medium and high conditions as L, M, and H respectively. The arrangement of all attributes has been undertaken to define the specified level and assigning a code to each specified attribute in the rows of the table. Table III shows the relation between the class numbers of conditional attributes of each parameter and its decision attribute. In this stage, all the cases of the conditional attributes should be checked to find the non-deterministic rule.

B. *Selection of Attributes*

Entropy uses a statistical property called information gain to decide the attribute which goes into a decision node. Gain measures how well given attributes separate training examples into targeted classes. The one with the highest information (being most useful for classification) is selected[15]. The calculated information gain for each attribute is shown below:

Gain(a=Age distribution) =0.23462
Gain(b=IgM:IgG ratio)= 0.1131
Gain(c=Symptoms I)= 0.35888
Gain(d=Symptoms II) =0.39425
Gain(e=Seasonal distribution) =0.05389
Gain(f=Platelet count-lakh/cumm) =0.52237
Gain(g=Clinical features)= 0.3108

Table Ii Data Table For Diagnosis Of Chikungunya Fever

S. No.	Conditional Attribute	Classification of individual situations	Chikungunya decision level
A	Age Distribution	1- 0-5 years	M
		2- 5-18 years	H
		3- >18 years	L
B	IgM: IgG Ratio	1- >=1.14	L
		2- <1.14	H
C	Symptoms – I	1- Joint Swelling	L
		2- Neck Pain	M
		3- Joint pain	H
D	Symptoms – II	1- Body ache	H
		2- Nausea	M
		3- Rash & Fatigue	L
E	Seasonal Distribution	1- June –July	M
		2- Aug-Dec	H
		3- Jan-May	L
F	Platelet Count (Lakhs/Cumm)	1- <1,00,000	H
		2- 1 lakh-2 lakh	M
		3- >2 lakh	L
G	Clinical Features	1- Fever	H
		2- Ocular manifestations	M
		3- Haemorrhagic manifestations	L

©IJRASET: All Rights are Reserved

TABLE III
OBSERVATION DATA FOR THE DIAGNOSIS OF CHIKUNGUNYA FEVER

Case	Conditional attribute							Decision attribute
	a	b	c	d	E	f	g	Chikungunya
P1	3	1	3	1	1	1	1	H
P2	2	2	2	3	2	2	1	M
P3	1	1	1	2	3	3	2	H
P4	3	2	2	2	1	2	2	L
P5	2	1	3	1	2	1	3	L
P6	1	2	1	3	3	3	3	L
P7	3	2	2	3	3	2	3	M
P8	2	1	3	1	1	1	3	M
P9	2	2	1	3	2	3	3	L
P10	1	1	2	1	3	2	1	H
P11	1	2	3	2	1	1	2	M
P12	3	1	1	3	1	3	3	L
P13	3	1	2	3	3	2	1	M
P14	2	2	3	1	3	1	2	H
P15	1	1	1	1	2	3	3	H
P16	2	1	2	2	2	1	2	L
P17	3	2	3	2	1	2	1	H
P18	2	1	1	1	1	3	1	H
P19	1	2	2	2	2	1	2	M
P20	3	2	3	3	3	2	3	L

TABLE IV: DATA SET AFTER THE CALCULATION OF INFORMATION GAIN

Case	c	d	f	G
P1	3	1	1	1
P2	2	3	2	1
P3	1	2	3	2
P4	2	2	2	2
P5	3	1	1	3
P6	1	3	3	3
P7	2	3	2	3
P8	3	1	1	3
P9	1	3	3	3
P10	2	1	2	1
P11	3	2	1	2
P12	1	3	3	3
P13	2	3	2	1
P14	3	1	1	2
P15	1	1	3	3
P16	2	2	1	2
P17	3	2	2	1
P18	1	1	3	1
P19	2	2	1	2
P20	3	3	2	3

©IJRASET: All Rights are Reserved

Here platelet count, symptoms I, symptoms II and clinical features have the highest information gain among the seven attributes and hence age distribution, IgM: IgG ratio and seasonal distribution were excluded due to their less importance[3,4]. The data set is shown in Table IV.

Since platelet count has the highest gain, therefore it is used as the decision attribute in the root node.

It has three values, count< 1 lakh; 1 lakh-2 lakh and > 2 lakh. Now with platelet count as root node, it is necessary to decide on the remaining four attributes:

S< 1 lakh =S_1={P1,P5,P8,P11,P14,P16,P19}= cases from table IV

Case	f	c	d	G
P1	1	3	1	1
P5		3	1	3
P8		3	1	3
P11		3	2	2
P14		3	1	2
P16		2	2	2
P19		2	2	2

Symptoms-II has the highest gain after Platelets Count; therefore it is used as the decision node. This process will continue for the remaining values until all data is classified perfectly or we run out of attributes.

C. Deterministic Rules

The decision tree can be built to extract the rules. Table V shows the rules generated from the presented data set.

Table v
Rules generated by the rough set analysis

Rule 1	(d=1,c=3)=>chikangunya=h
Rule 2	(d=2,c=2)=>Chikangunya=M
Rule 3	(d=2,c=3)=>Chikangunya=H
Rule 4	(d=1,g=1)=>Chikangunya=H
Rule 5	(d=1,g=2)=>Chikangunya=H
Rule 6	(d=1,g=3)=>Chikangunya=L
Rule 7	(d=2,g=2)=>Chikangunya=M

VI. CONCLUSIONS

This study, it has discussed the Rough set theory as an approach to diagnosis Chikungunya fever by the elimination of redundant data available in medical database ends with a set of decision rule which can assist the doctor in the elaboration of the patient's diagnosis. It satisfies an excellent capability to detect the minimal possible decision-making algorithm. The Rough set approach to analysis has some important advantages such as (Pawlak, 1997): Data reduction; Evaluates the significance of data; Generates sets of decision rules from data; Facilitates the interpretation of the obtained result. This study paper isn't projected to produce an entire summary of medical data processing however rather describes some areas that appear to be necessary from our purpose of reading for applying rough set in diagnosis for infective agent dataset.

VII. ACKNOWLEDGMENT

Most importantly, acclaims and on account of the God, the Almighty, for his showers of favors all through my exploration work to finish the examination effectively.

I might want to express my profound and true appreciation to my examination director, Dr. Vinod Kumar Sharma M.Sc., Ph.D., Dean Faculty of engineering & technology, Sri Ganganagar engineering college, Tantia University, Sri Ganganagar, Rajasthan for giving me the opportunity to do research and providing invaluable guidance throughout this research. His dynamism, vision, sincerity, and motivation have deeply inspired me. I am to a great degree appreciative to my folks for their adoration, supplications,

©IJRASET: All Rights are Reserved

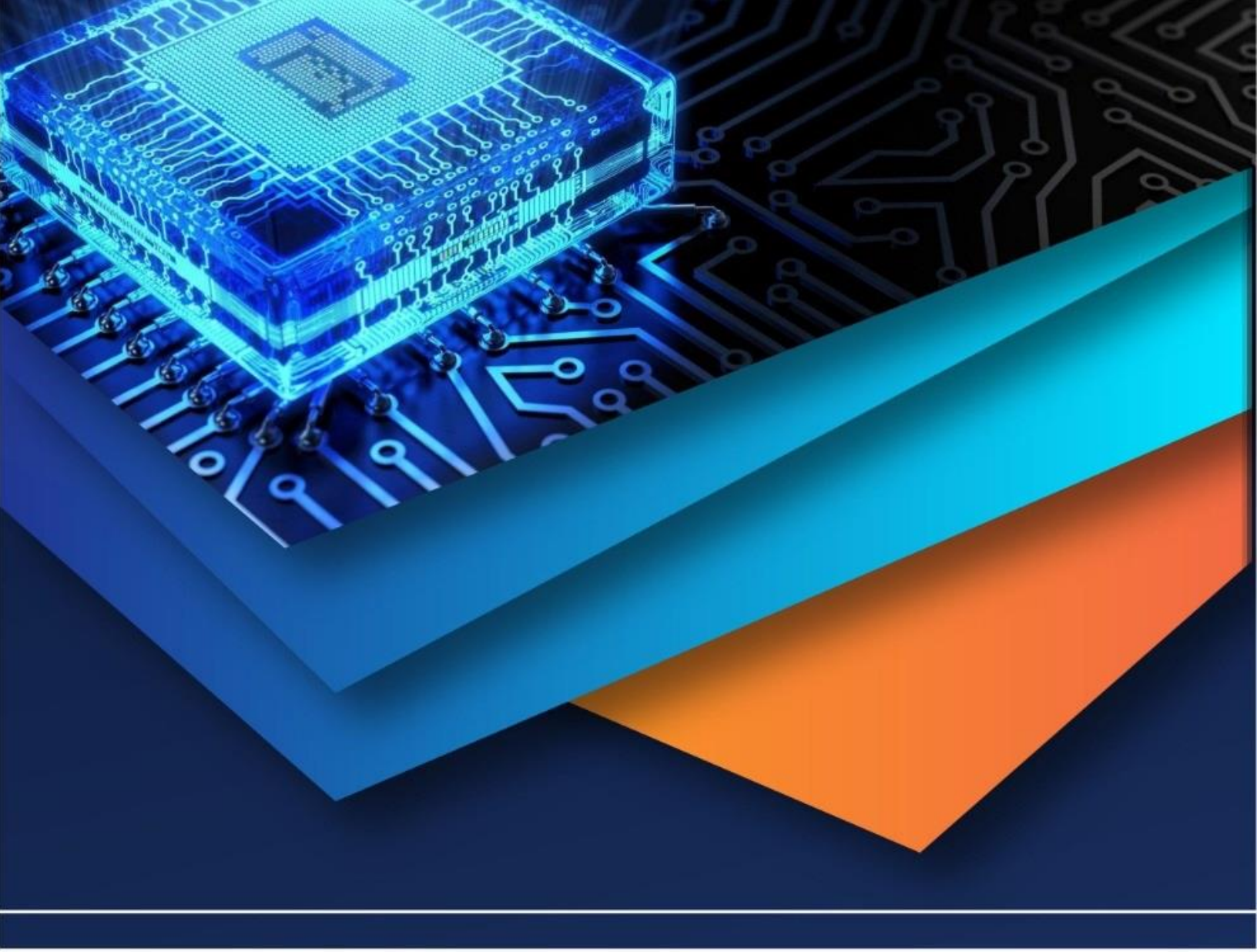
administering to teaching and setting me up for my future. Likewise, I express my gratitude to my sisters, the sibling for their help and significant supplications.

I would like to say thanks to my friends and research colleagues, Tushar Bhatia, Heena Sheetal, and Manpreet Kaur for their constant encouragement. I also thank all the staff of Research section of Tantia University, Sri Ganganagar for their kindness. Finally, my thanks go to all the people who have supported me to complete the research work directly or indirectly.

REFERENCES

[1] Yergolkar P.N., Tandale B.V., Arankalle V.A., Sathe P.S., and Sudeep A. et al.(2006): Chikungunya episodes caused by African genotype, India. Emerg Infect Dis.; 12:1580– 83.
[2] V. Ravi, Re-development of Chikungunya infection in India. Indian J Med Microbiol. 2006; 24: 83– 8
[3] A Swaroop, A Jain, M Kumhar, N Parihar, S Jain. Survey Article: Chikungunya Fever. Diary Indian Academy of Clinical Medicine. 2007;8(2):164– 68
[4] AC Kennedy, J Fleming, L Solomon. Chikungunya viral arthropathy: a clinical portrayal. J Rheumatol. 1980;7(2):231– 36. [PubMed]
[5] P Kaur, P Manickam, MV Murhekar, V Ramachandran, R Ramachandran, HK Raju, et al. Chikungunya outbreak, south India, 2006. Emerg Infect Dis. 2008;14:1623–5.
[6] National Vector Borne Disease Control Programme, Directorate General of Health Services, Ministry of Health and Family Welfare, Government of India, New Delhi. [accessed on April 14, 2010]
[7] Z Pawlak, Rough Sets: Theoretical Aspects of Reasoning with respect to data, Dordrecht, Kluwer tutorialPublishers, 1991.
[8] Z.Pawlak, Rough sets, In T.Y.Lin, N.Cercone (Eds.), Rough Sets and information preparing, Dordrecht: Kluwer tutorialPublisher, (1997), 3– 8.
[9] Jia-Yuarn Guo, Rough set-based way to deal with information preparing, Ph.D. Postulation, 2003.
[10] Z. Pawlak, Rough Set Rudiments, Institute of Computer Science Report, Warsaw University of Technology, Poland, (1996), 1–47
[11] R. Jensen, Q. Shen, Fuzzy-Rough Data Reduction with Ant Colony Optimization, Fuzzy Sets, and Systems, 149 (2005), 5-20.
[12] Janicki, Ryszard; Parnas, David Lorge; Zucker, Jeffery (1997). "Tabular representations in relational documents". In Brink, C.; Kahl, W.; Schmidt, G. Relational Methods in Computer Science. Springer Verlag. ISBN
[13] Fisher, D.L. (1966) "Information, Documentation, and Decision Tables" Comm ACM Vol. 9 No. 1 (Jan. 1966) p. 26– 31.
[14] K. Karimi and H.J. Hamilton (2011), "Age and Interpretation of Temporal Decision Rules", International Journal of Computer Information Systems and Industrial Management Applications, Volume 3
[15] N. F. G. Martin and J. W. Britain, Mathematical Theory of Entropy, Addison-Wesley, Reading, 1981
[16] W. Feller, An Introduction to Probability Theory and Its Applications, Wiley, New York,1957
[17] Carter, Tom (March 2014). An introduction to information theory and entropy (PDF). Santa Fe. Retrieved Aug 2014
[18] R. V. Brown, A. S. Kahr and C. Peterson, Decision Analysis for the Manager, Holt, Rinehart & Winston, New York, 1974.
[19] Simon and Schuster, Grammatical Man, Information, Entropy, Language, and Life, New York, Campbell, Jeremy 198
[20] Q. Shen, Chouchoulas, Rough set-aided keyword reduction for text categorization, Application of Artificial Intelligence, 15 (2001), 843-873.
[21] R.Jensen, Q.Shen, Finding Rough set reducts With Ant Colony Optimization, In Proc. UK Workshop on Machine Intelligence, (2003), 15-22
[22] P. Seethalakshmi, S. Vengataasalam, Attribute choice in Product Mix: A Rough Set Approach, The International Journal of

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)