Detection of HIV by using Rough Set and Homotopy Analysis Method

Subrata Kumar Nayak[1], S K Pradhan[2], Sujogya Mishra[3], S. Pradhan[4], P. K. Pattnaik[5]

[1]MCA Department, Utkal University ,Research Scholar

Bhubaneswar,

subratnayaksap@gmail.com

[2]MCA Department, Utkal University,

Bhubaneswar

sateesh1960@gmail.com

[3]OUTR,

Bhubaneswar

sujogya123@gmail.com

[4]RBVRR Women's College,

Hyderabad

sipalipradhan06@gmail.com

[5]Department of Mathematics

Odisha University of Technology and Research,

Bhubaneswar

papun.pattnaik@gmail.com

Abstract— The significant objective of this research is to recognize how to calculate the classification process using rough set theory (RST) for the Human immunodeficiency virus infection and acquired immune deficiency syndrome (HIV & AIDS) symptoms dataset. RST has a multi-dimensional concept with multiple approaches. In this paper, our main objective is to find the symptoms of (HIV & AIDS) using basic RST and Homotopy Analysis Method (HAM) to validate our claim using statistical techniques. We prefer RST & HAM over other soft computing techniques and Mathematical Modelling as both RST and Homotopy Analysis (HAM) because RST can handle vague and imprecise data efficiently, and HAM is a suitable technique for finding analytical solutions. We have used the chi-squared test to validate our claim.

Keywords- RST; Soft Computing; HIV; AIDS; HAM.

I. INTRODUCTION

Determining, treating, calculating and analyzing evidence is among the major significant results of our age. The representation of information depends upon how the information is interpreted. Enhancing the representation of the information needs an automated technique. When the data size is large enough, there are significant chances of repetitions of information. The major objective of the RST concept is to significantly analyse vague and imprecise data. RST was proposed by Polish mathematician Pawlak [1]. Several studies are going on uncertainty and vagueness. Currently, several researchers were implemented RST to get several results with significant accuracy. Das *et al.* [2] proposed a novel technique using RST to predict heart diseases. Several researchers depend upon machine learning tools to estimate the pattern and analyze various models in this context. Das *et al.* [3] discussed the detection of Pneumonia using machine learning tools. RST hybrid with various machine learning tools useful in predicting various meteorological phenomena have been discussed in this research. Das *et al.* [4] developed a novel model by combining

RST and Random Forest Model (RFM) to predict wind speed in this context. Das *et al.* [5] designed a model for predicting rainfall with significant accuracy. Recognizing the pattern of diseases is also an important application of soft computing in this context. Mishra *et al.* [6] discussed the symptoms of COVID-19 by using soft computing techniques. Nayak *et al.* [7,8] discussed using RST and SVM to predict symptoms of malaria and cardiac arrest. Soft computing technique has useful application in social science. In this context, RST was employed by Mishra et al. [9] to study the mental models of Odisha workers in government and non-governmental organisations. The primary cause of the lethal, incurable disease known as AIDS is the virus HIV, which weakens the human immune system. Despite the fact that there is no treatment for HIV/AIDS, medications can be used to control the infection and slow the disease's development. International organisations [10] are working to encourage access to care and prevention measures in underdeveloped nations. According to a WHO estimate, the disease has been contracted by 84.2 million people since the epidemic's start and has claimed the lives of 40.1

1460

million. Mohameda and Mohammed [11] designed a model to predict the activity of anti-HIV molecules using soft computing techniques. To find information that is concealed within a data set, particularly one that is large and complex. True or false was the starting point for traditional information interpretations, but as knowledge increased, people started searching for more nuanced answers that went beyond these limitations. In the course of this expansion, Akduman *et al.* [12] discussed the enhancement of the production of B_{12}. Andonie *et al.* [13] discussed potential HIV-1 using a soft computing approach, and Akduman *et al.* [14] discussed the emotional intelligence of generation Majdi *et al.* [15] discussed the use of fuzzy logic in attribute reduction. Chen *et al.* [16] discussed alternative therapy to treat HIV. Several machine tools available to analyse give a concrete result regarding the types and symptoms of HIV, the process of predicting also depends upon its symptoms. Over the years, several types of research have been carried out with significant accuracy. Ayele *et al.* [17] designed a mathematical model for optimal control and applied their concept to the population of Ethiopia. Bentout *et al.* [18] designed an age-structured model for COVID-19. Mahmoudi *et al.* [19] discussed the classifications of several time series models with significant error margins and fuzzy clustering. Hussain *et al.* [20] discussed data dependency's stability and error detection. Alrabaiah *et al.* [21] discussed a comparative study of COVID-19 using a modified SEIR model. Mohiuddine *et al.* [22] discussed the fractional order of the difference operator. Wang *et al.* [23] discussed ant colony optimization applications on schedule. Zhang *et al.* [24] designed a model of prostate cancer using magnetic resonance imaging. Several mathematical models like SIB and SEIR were designed to predict several pandemics Hntsa, and Kahsay [25] designed a model for controlling cholera. Uncertainty, vagueness and inaccurate data are always challenging for physicians and data scientists to counter. Ghosh and Mukhopadhyay [26] discussed the precise factor in a rule-based system finding useful information or selecting a significant feature from an inaccurate data set is always a challenge in this context. Shen and Jensen [27] designed a model which was useful in retrieving important information based on Fuzzy and RST. RST also has useful applications in data mining. Bal [28] used RST as symbolic data mining to a complete decision table. Fuzzy logic had wide application in rule generation and helped find optimal results in certain analyses in this context. Wang [29] discussed rule generation knowledge from the instance. Several researchers also work on this context's uncertainty, impreciseness, and vagueness. Riza *et al.* [30] discussed the implementation of RST and Fuzzy Rough sets using R-Package. There were various documents available, for instance, Li [31]. Busse *et al.* [32] discussed the application of RST in data mining and applied it to design a set covering method to find optimal reduct. There was contemporary research in the field of impreciseness

uncertainty several software are available to find optimal reduct. Uusitalo *et al.* [33] discussed the uncertainty of various deterministic designs. Barnoi and Tarantola [34] discussed various applications of the probabilistic framework to analyze uncertainty. Martyn *et al.* [35] discussed structural errors among various diagnosis modifications between hydrological designs. Gregory *et al.* [36] discussed the importance of multicriteria choice-making in the case of general decision-making. Several researchers also work on risk management using various forms of decision analysis. In this context, Bonano *et al.* [37] designed an alternative tool for risk management. Bose *et al.* [38] discussed the utility of multiple attributes combination in group decision making. Several papers related to vague data analysis using several soft computing techniques. Bazan and Skowron *et al.* [39] discussed RST and approximation and analysis of vagueness in this line of research. Butz *et al.* [40] designed an efficient algorithm for the RST flow graph. There are several correlation techniques available in which we can group the data set according to their dissimilarities. Correlation is a statistical tool that explains the relation between two pair variables. Correlation measured by correlation coefficients -1,0,1. Suppose the correlation coefficients near to -1 are negatively correlated. Similarly, the coefficients near 0 will not correlate with the pairs, and if the coefficients are near 1, then both pairs are strongly correlated. In this work, we are using the correlation technique to find dissimilar groups.

II. BACKGROUND

The background of data analysis depends upon various factors like vagueness, impreciseness, and complicated medical data structure. Data analysis is required to counter these anomalies. We use tools like RST, Correlation and Regression techniques to make the impreciseness precise. In the subsequent section, we will briefly discuss the above said.

A. Correlation and Regression

- Correlation is a data analysis useful in finding the relation between two groups of objects moving concerning a particular relation. The correlation was measured concerning correlation coefficients. The correlation coefficients are represented by r = 0, -1, 1 as No correlation, Negative correlation, and Positive correlation.

- Regression: Regression investigates the relationships between variables, or more simply, how changes in one variable affect changes in another, or cause and effect. It implies that one or more elements will determine the outcome. For instance, regression shows how two variables interact, whereas correlation tells how two variables are related. Consider how different crops may

_____

benefit from additional rainfall as an example, just as they might wilt or not grow during a drought.

● Fundamental of Rough Set Theory: RST is a tool which is a useful tool to handle uncertainty based on the concept of Upper Approximation, Lower Approximation, Boundary region, Reduct, and Core.

B. Significance of Rough Set Theory

The significance of RST is useful in data cleanness, preciseness, data analysis, and systematically deriving knowledge from raw data.

● Indiscernibility: The method of reducing information set represented in the form of a table to keep a single representative.

● Approximation: Basically, an approximation based on the concept of how accurately a group is being defined concerning a target set. RST is classified into two classes, i.e., Upper Approximation and Lower Approximation.

● Upper Approximation: Upper approximation consists of all possible cases from which we can retrieve certain consequences. If RST is defined as K(x), then, the lower approximation is defined as,

$$\overline{K}(X) = \bigcup \{L \in U/R : L \bigcap X \neq \varphi\}$$

● Lower Approximation: Lower Approximation consists of all possible cases from which we can retrieve certain consequences $\underline{K}(x) = \bigcup \{L \in U/R : L \subset X\}$.

● Boundary region: Boundary region (BNR) is defined as

$$BNR = \overline{K}(X) - \underline{K}(x)$$ and when BNR $= \varphi$, the set is called a crisp set.

C. Basic structure of RST

In general, RST was represented as an Information Table as described below. The table given below has records <1,2,3,4>, <a,b,c> are conditional attributes, d is the decision attribute. <11,22> are the values of conditional attributes, <1,0> is the values of the decision attributes.

TABLE I. DESCRIPTION OF RST INFORMATION

Records	a	b	c	d
1	11	11	11	1
2	22	11	22	1
3	11	11	11	0
4	11	11	22	1

D. Fundamental Steps towards Data Analysis

We collected various diseases like cardiac problems, which include pulmonary tuberculosis and other cardiac problems. We collected nearly 500,000 data on diseases from various places in our state. This entire data is described in the following table.

TABLE 2. INITIAL DATA

District	Cardiac Arrest	Pulmonary Tuberculosis	HIV	Pneumonia	Lungs infection	General malaise	Total
Anugul	15000	15000	25000	30000	5000	10000	100000
Dhenkanal	30000	30000	10000	15000	5000	10000	100000
Puri	15000	150000	25000	25000	10000	10000	100000
Khordha	30000	30000	10000	10000	10000	10000	100000
Jajpur	15000	15000	150000	10000	25000	20000	100000

E. Algorithm to find reductreduct

Algorithm to find the Indiscernibility

Step-1 Consider the data set as <U, C, D> where U is the records, C is the conditional attributes, and D is the decision attributes.

Step-2 for i = 1 to n-1 (where n is the number of conditional attributes)

Step-3 if
{
Some of the values of conditional attributes are the same and can be called Indiscernible.
}
else
Reduct found

Step-4 end if

Step-5 end for

Step-6 Continue steps 1 to 5 until no conditional attribute is left for analysis. So, the correlation technique has been used in the data set to find the number of dissimilar records. The correlation for the two-group data set was taken for 10 years in one instance and 9 years duration for another instance given.
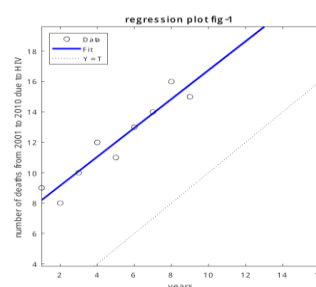

Fig. 1. Regression plot for population

1462

_____

The figure-1 describes the regression between years and death took over 9 years. Similarly, the next figure represents data randomly taken over 5 years, as described below.
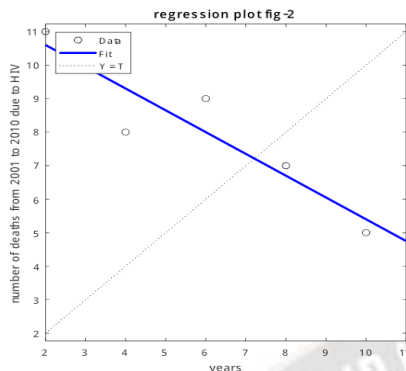

Fig. 2. Regression plot of sample

Using the above correlation techniques, 6- different records are found, which are given as follows.

F.   Renaming of Conditional attributes

We have renamed the conditional attributes as t for cardiac arrest, u for Pulmonary Tuberculosis, HIV as v, Pneumonia as w, and Lungs infection as x, and do not consider General malaise as this is a common syndrome for all. $<t,u,v,w,x>$ consider conditional attributes and their values significant and insignificant as a and b. Similarly, l is the decision attribute and its values as agree and disagree as 1 and 2. The entire correlation table is given below as follows.

TABLE III. INITIAL DATA TABLE AFTER CORRELATION

K	t	u	v	w	x	l
K_1	a	a	b	b	a	1
K_2	a	a	a	b	a	2
K_3	b	b	a	a	a	2
K_4	b	b	b	a	a	2
K_5	b	b	b	a	b	2
K_6	a	b	a	a	b	2

Using the Reduct algorithm, the following indiscernibility is calculated, e.g. IND(t) = {$<K_1,K_2, K_6>,<K_3,K_4,K_5>$}. Similarly, we are finding indiscernibility for u, v, w, and x, and we find indiscernibility for combining two attributes, three attributes, and 4 attributes, resulting in the following reduced set.

TABLE IV. REDUCE TABLE

Sl. No.	Reduct
1	(t,u,v,x),
2	(t,u,v,w),
3	(t,v,x,w),
4	(u,v,x,w),
5	(u,v,w)}

Core in RST define as ∩ <Reduct>, In this work, Core={(t,u,v,x)} ∩ {(t,u,v,w)} ∩ {(t,v,x,w)} ∩ {(u,v,x,w)} ∩

$\{(u,v,w)\}= v$. In this correlation approach, v is considered as HIV. In the next section, the strength of RST has been used to find the section of people vulnerable to HIV.

2.1.6 Algorithm finding the reduct using strength of RST

Step-1 For a given target set S and given set of conditional attributes P and decision attribute R where S is the subset. (We have considered Boolean values for the target set, i.e., a subset of the decision attribute).

Step-2 The values of conditional attributes count are calculated, and then we find the count of the target set.

Step-3 for (j = 1 ; j <= n,j^{++})

We find the Z-values next, where n is the number of conditional attributes.

Z- values are the ratio of the count of conditional attributes to the count of the target set concerning their values.

If Z >60, we accept

else

reject

end if

end for

We collected 10,000 HIV cases only explained in Table V. We have renamed the conditional attributes as homosexual, lesbian, unprotected sex, drug addiction (those who use injection), and shaving without proper restriction as 1 2 3 4 5 respectively and its values are significant and insignificant renamed as k and l. We have renamed the decision attributes as d, positive as 1 and negative as 2 .

TABLE V. HIV INFORMATION TABLE AFTER CORRELATION

M	1	2	3	4	5	d
M_1	K	k	l	l	K	1
M_2	K	k	k	l	K	1
M_3	L	l	k	k	K	1
M_4	L	l	l	k	K	2
M_5	L	l	l	k	L	2
M_6	K	l	k	k	L	2

We have considered our target set as 1 taken from the decision attributes set. Next, using the algorithm, M = <M_1, M_2, M_3, M_4, M_5, M_6 > are the 6-dissimilar records. Finding strength $1(k)_1$ represents attribute is 1 value of the attribute is k target is 1 = 2/3 = 66%. Similarly, $2(k)_1$ = 100% ,$3(k)_1$ = 66%, $4(k)_1$ = 25%, $5(k)_1$ = 80%, so attribute 4 has been dropped from the HIV information Table.

TABLE VI. REDUCE HIV INFORMATION TABLE AFTER CORRELATION

M	1	2	3	5	d
M_1	k	k	l	k	1
M_2	k	k	k	k	1
M_3	l	l	k	k	1
M_4	l	l	l	k	2
M_5	l	l	l	l	2
M_6	k	l	k	l	2

From Table VI, we have developed a set of rules as follows:

1463

_____

Rule-1 if 1 value is k, 2 value is k, 3 value is l, and 5 value is k then the decision is 1

Rule-2 if value <1,2,3,5> is k the decision is 1

Rule-3 if value of <1,2> is l and value of <3,5> is k then decision is 1

Rule-4 if value of <1,2,3> is l and value of <5> is k then decision is 2

Rule-5 if values of <1,2,3,4,5> is l then decision is 2

Rule-6 if values of <1,3> is k and values <2,5> is l then decision is 2

As the problem domain is very large and the diseases spread very fast, like they can be transmitted by various means, prediction of HIV is always challenging by conventional soft computing techniques. So, in the next section, we have formulated a mathematical model to find the growth of HIV infection efficiently due to several other parameters. This model is based on the HAM model and is effectively used for classification.

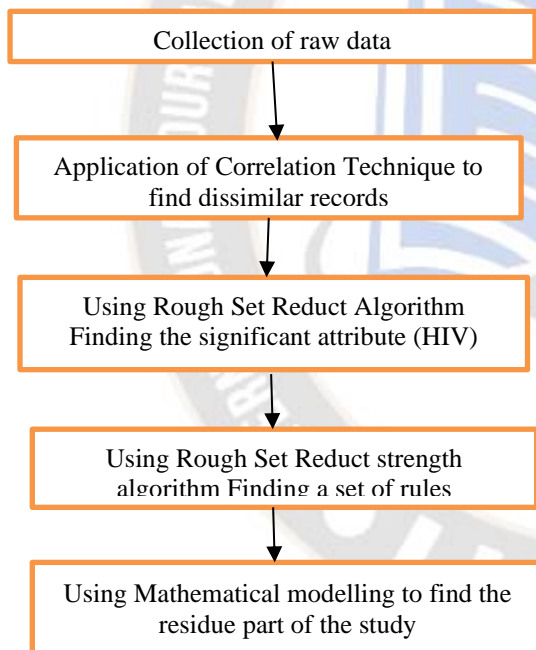## III. DETECTION OF HIV BY USING MATHEMATICAL MODELLING AND ROUGH SET

Collection of raw data

↓

Application of Correlation Technique to find dissimilar records

↓

Using Rough Set Reduct Algorithm Finding the significant attribute (HIV)

↓

Using Rough Set Reduct strength algorithm Finding a set of rules

↓

Using Mathematical modelling to find the residue part of the study

Fig. 3. Flow chart for Mathematical modelling

A. *Application of the HAM for solving a model for HIV infection*

Several nonlinear mathematical models have been created to explain HIV infection. Perelson created a model for HIV infection in the human immune system in 1989 [41]. The populations of uninfected cells, infected cells, and free virus particles are the three variables in this model of viral transmission. In order to account for the four components of uninfected cells, latently infected cells, actively infected cells, and free virus particles, Perelson et al. [42] enlarged the model

of [41] and created a new model. Four ordinary differential equations might be used to define the proposed model. It was demonstrated that the model may imitate several of the clinically recognised symptoms of AIDS. The model provided in [42] was condensed by Culshaw and Ruan into a set of three ordinary differential equations under the premise that any infected cell may produce the virus. The Culshaw and Ruan model for HIV infection is approximately resolved by the homotopy analysis approach (HAM), which is introduced and improved in this paper. The model is presented below,

$$\left.\begin{array}{l} \dfrac{dT}{dt} = s - \mu_T T + g_r T\left(1 - \dfrac{T+1}{T_{max}}\right) - k_1 VT, \\[3mm] \dfrac{dI}{dt} = k_1' VT - \mu_I I. \\[3mm] \dfrac{dV}{dt} = N\mu_b I - k_1 VT - \mu_V Vm, \end{array}\right\}$$

(1)

where $T(t)$, the concentration of healthy cells, $I(t)$, infected cells, and $V(t)$ free HIV at a time t. $\mu_T, \mu_I, \mu_B, \mu_V$ Represent natural and blanket death rates, Lytic and death rates, respectively. k_1, k_1' represent the rate of infected cells with the virus and the rate of infected cells becoming active, g_r growth rate N, number of virions produced T_{max}, maximal concentration, s source term, T_0 and cell concentration for HIV-negative persons.

B. *Solution Methodology*

The semi-analytical homotopy analysis technique (HAM) for the resolution of linear/nonlinear ordinary/partial differential equations was initially introduced by Liao Shijun in 1992. The homotopy analysis approach produces a convergent series solution for linear/nonlinear systems by drawing on the notion of the Homotopy from topology. In this work, a nonlinear model has been considered, and this model's solution is calculated using HAM. This is possible by managing the system's nonlinearities with a homotopy-Maclaurin series. The HAM is an analytical approximation method developed for the age of computers, and its goal is to compute using functions instead of numbers. The homotopy approach is one of the most well-known methods for solving nonlinear equations that many academics have looked into [43-46]. Pattnaik and his coresearchers [47-58] in different work had used different techniques, such as numerical methods like the 4th order Runge-

_____

Kutta method with the shooting technique, and analytical methods like Laplace Transform method etc., to solve highly nonlinear differential equations. They have used similarity transformation methods in their research to convert higher-order nonlinear partial differential equations to highly nonlinear ordinary differential equations. This motivates us to use a novel method such as HAM to find analytical expressions of the so-called HIV model. However, just like other nonlinear analytical methods, analysis methodologies have their own set of limitations.

According to Liao, $\mathcal{N}[f(t)] = 0$ (2)

$\mathcal{N}$: nonlinear differential operator, with a linear operator, $\mathcal{L}$ which satisfies,

$\mathcal{L}[f_0(t)] = 0$ (3)

Here, choices of the linear operator are vital as the solution should satisfy the initial and boundary conditions.
Fundamentals of Homotopy analysis given as follows

$H(\phi(t,r);r) = (1-r)\mathcal{L}[\phi(t,r) - f_0(t)] + q\mathcal{N}[\phi(t,r)]$ (4)

with homotopy parameter (r).

For $r = 0$,

$H(\phi(t,0);0) = \mathcal{L}[\phi(t,0) - f_0(t)] = 0$ (5)

i.e., $\phi(t,0) = f_0(t)$

and for $r = 1$,

$H(\phi(t,1);1) = \mathcal{N}[\phi(t,1)] = 0$ (6)

with $\phi(t,1)$ is the solution $\mathcal{N}[\phi(t,1)]$.

Now introducing the auxiliary parameter $\hbar$ and the auxiliary function $H(t)$,

$H(\phi(t,r);r) = (1-r)\mathcal{L}[\phi(t,r) - f_0(t)] + r\hbar H(t)\mathcal{N}[\phi(t,r)]$ (7)

For $H = 0$, generalized Homotopy can be obtained. The use of the auxiliary parameter $\hbar$ should be adjusted to get the convergence of the solution. Auxiliary function $H(t)$ is used to ensure homotopy results in the solution process.
The procedure can be defined as follows:
1. A solution of the form can be assumed as,

$\phi(t,1) = f_0(t) + \sum_{m=1}^{\infty} \frac{1}{m!} \left(\frac{\partial^m \phi(t,r)}{\partial r^m} \right)_{r=0} r^m$ (8)

In particular, Eqn. (8) Taylor series expansion of the solution about the embedding parameter r.
2. Individual terms in the series solution can be calculated using

$v_m(t) = \frac{1}{m!} \left(\frac{\partial^m \phi(t,r)}{\partial r^m} \right)_{r=0}$, which can be differentiated to

get the generalized homotopy $m-$ times r. So, deformation equations can take the form,

$L[v_m(t) - \chi_m v_{m-1}(t)] = \hbar H(t) R_m(\overline{v}_{m-1})$ subject to

$v_m(0) = 0$ (9)

where, $R_m(\overline{v}_{m-1}) = \frac{1}{(m-1)!} \left(\frac{\partial^{m-1} N[\phi(t,r)]}{\partial r^{m-1}} \right)_{q=0}$ (10)

$\chi_m = \begin{cases} 0, & m \le 1 \\ 1, & \text{otherwise} \end{cases}$ (11)

$v_m(t)$ is then the solution to (8) and

$\overline{v}_n = \{v_0(t), v_1(t), \dots, v_n(t)\}$ (12)

3. The choice $\hbar$ is determined by plotting the lth derivative, $t = 0$ which shows an essentially horizontal variation on some interval. Choice $\hbar$ can be adjusted from this interval to get the convergent solution.

C.   Algorithms
Step 1: Guess the initial choice.
Step 2: Guess the linear operator.

Step 3: Find a homogeneous solution f^h using a linear operator.

Step 4: Compute a particular solution f_m^p.
Step 5: Then, the total solution is the sum of the homogeneous

solution and particular solution, $f = f^{\hbar} + f_m^p$
Step 6: Compute a particular solution by using initial conditions.

Step 7: Adjust control convergence parameter $(\hbar)$ for convergent solution.

Step 8: Put the value $(\hbar)$ in step 6 to find the final solution.
3.4   Solution of HIV model
To achieve numerical solution for the above problems this studies uses HAM,

_____

$$T(t) = T_0, I(t) = I_0, V(t) = V_0, \quad at \quad t = 0 \tag{13}$$

$$\mathcal{L}_T = \frac{dT}{dt}, \mathcal{L}_I = \frac{dI}{dt}, \mathcal{L}_V = \frac{dV}{dt}, \tag{14}$$

with the following properties:

$$\mathcal{L}_T(c_1) = \mathcal{L}_I(c_2) = \mathcal{L}_V(c_3) = 0 \tag{15}$$

where c_1, c_2 c_3 and constants.

Now defining the nonlinear parameter as

$$\mathcal{N}_T\left[\tilde{\tilde{T}}(t;r), \tilde{\tilde{I}}(t;r), \tilde{\tilde{V}}(t;r)\right] = \begin{cases} \dfrac{\partial \tilde{\tilde{T}}(t;r)}{\partial t} - \left(g_r - \mu_T\right)\tilde{\tilde{T}}(t;r) + \dfrac{g_r}{T_{\max}}\left(\tilde{\tilde{T}}(t;r)\right)^2 \\ -s + \dfrac{g_r}{T_{\max}}\tilde{\tilde{T}}(t;r)\tilde{\tilde{I}}(t;r) + K_1\tilde{\tilde{V}}(t;r)\tilde{\tilde{T}}(t;r) \end{cases} \tag{16}$$

$$\mathcal{N}_I\left[\tilde{\tilde{T}}(t;r), \tilde{\tilde{I}}(t;r), \tilde{\tilde{V}}(t;r)\right] = \frac{\partial \tilde{\tilde{I}}(t;r)}{\partial t} - K_1'\tilde{\tilde{V}}(t;r)\tilde{\tilde{T}}(t;r) + \mu_I\tilde{\tilde{I}}(t;r), \tag{17}$$

$$\mathcal{N}_V\left[\tilde{\tilde{T}}(t;r), \tilde{\tilde{I}}(t;r), \tilde{\tilde{V}}(t;r)\right] = \frac{\partial \tilde{\tilde{V}}(t;r)}{\partial t} - n\mu_b\tilde{\tilde{T}}(t;r) + K_1\tilde{\tilde{V}}(t;r)\tilde{\tilde{T}}(t;r) + \mu_V\tilde{\tilde{V}}(t;r), \tag{18}$$

we can obtain,

$$(1-r)\mathcal{L}_T\left[\tilde{\tilde{T}}(t;r) - T_0(t)\right] = r\hbar H_T(t)\mathcal{N}_T\left[\tilde{\tilde{T}}(t;r), \tilde{\tilde{I}}(t;r), \tilde{\tilde{V}}(t;r)\right], \tag{19}$$

$$(1-r)\mathcal{L}_I\left[\tilde{\tilde{I}}(t;r) - I_0(t)\right] = r\hbar H_I(t)\mathcal{N}_I\left[\tilde{\tilde{T}}(t;r), \tilde{\tilde{I}}(t;r), \tilde{\tilde{V}}(t;r)\right], \tag{20}$$

$$(1-r)\mathcal{L}_V\left[\tilde{\tilde{V}}(t;r) - V_0(t)\right] = r\hbar H_V(t)\mathcal{N}_V\left[\tilde{\tilde{T}}(t;r), \tilde{\tilde{I}}(t;r), \tilde{\tilde{V}}(t;r)\right], \tag{21}$$

The homotopy equations can be,

$$\left.\begin{matrix} \tilde{\tilde{T}} = T_0, \tilde{\tilde{I}} = I_0, \tilde{\tilde{V}} = V_0, \quad for \quad r = 0 \\ \tilde{\tilde{T}} = T, \tilde{\tilde{I}} = I, \tilde{\tilde{V}} = V, \quad for \quad r = 1 \end{matrix}\right\} \tag{22}$$

where T, I, V are the exact solutions of model (1).

Using the Taylor series respect to r, one can write,

$$\tilde{\tilde{T}}(t;r) = T_0(t) + \sum_{m=1}^{\infty}\frac{1}{m!}\left(\frac{\partial^m\tilde{\tilde{T}}(t;r)}{\partial r^m}\right)_{r=0}r^m, \tag{23}$$

$$\tilde{\tilde{I}}(t;r) = I_0(t) + \sum_{m=1}^{\infty}\frac{1}{m!}\left(\frac{\partial^m\tilde{\tilde{I}}(t;r)}{\partial r^m}\right)_{r=0}r^m, \tag{24}$$

$$\tilde{\tilde{V}}(t;r) = V_0(t) + \sum_{m=1}^{\infty}\frac{1}{m!}\left(\frac{\partial^m\tilde{\tilde{V}}(t;r)}{\partial r^m}\right)_{r=0}r^m, \tag{25}$$

All the above series are convergent with some particular values of an auxiliary parameter $\hbar$. Differentiating (19) - (21) m-times concerning r and dividing by $m!$ then setting $r = 0$, one can get,

$$\mathcal{L}_T\left[T_m(t) - \chi_m T_{m-1}(t)\right] = \hbar_T R_{m,T}\left(\vec{T}_{m-1}, \vec{I}_{m-1}, \vec{V}_{m-1}\right)H_T(t), \tag{26}$$

$$\mathcal{L}_I\left[I_m(t) - \chi_m I_{m-1}(t)\right] = \hbar_I R_{m,I}\left(\vec{T}_{m-1}, \vec{I}_{m-1}, \vec{V}_{m-1}\right)H_I(t), \tag{27}$$

$$\mathcal{L}_V\left[V_m(t) - \chi_m V_{m-1}(t)\right] = \hbar_V R_{m,T}\left(\vec{T}_{m-1}, \vec{I}_{m-1}, \vec{V}_{m-1}\right)H_V(t), \tag{28}$$

where

$$T_m = \frac{1}{m!}\left(\frac{\partial^m\tilde{\tilde{T}}(t;r)}{\partial r^m}\right)_{q=0}, I_m = \frac{1}{m!}\left(\frac{\partial^m\tilde{\tilde{I}}(t;r)}{\partial r^m}\right)_{r=0}, V_m = \frac{1}{m!}\left(\frac{\partial^m\tilde{\tilde{V}}(t;r)}{\partial r^m}\right)_{r=0} \tag{29}$$

with initial conditions

$$T_m(0) = I_m(0) = V_m(0) = 0, \tag{30}$$

where

$$R_{m,T}(t) = \begin{cases} \dfrac{dT_{m-1}(t)}{dt} + \left(\mu_T - g_r\right)T_{m-1}(t) + \dfrac{g_r}{T_{\max}}\sum_{i=0}^{m-1}T_i(t)T_{m-1-i} \\ + \dfrac{g_r}{T_{\max}}\sum_{i=0}^{m-1}T_i(t)I_{m-1-i}(t) + K_1\sum_{i=0}^{m-1}T_i(t)V_{m-1-i}(t) - \left(1-\chi_m\right)s \end{cases} \tag{31}$$

$$R_{m,I}(t) = \frac{dI_{m-1}(t)}{dt} - K_1'\sum_{i=0}^{m-1}V_i(t)V_{m-1-i}(t) - \left(1-\chi_m\right)s + \mu_I I_{m-1}(t), \tag{32}$$

$$R_{m,V}(t) = \frac{dI_{m-1}(t)}{dt} - N\mu_b I_{m-1}(t) + \sum_{i=0}^{m-1}V_i(t)T_{m-1-i}(t) + \mu_I V_{m-1}(t), \tag{33}$$

Using $H_T(t) = H_I(t) = H_V(t) = 1$, the solution of the mth-order deformation Eqs. (31) - (33).

For $m \geq 1$ becomes,

_____

$$T_m(t) = \chi_m T_{m-1}(t) + \hbar_T \int_0^t R_{m,T}(\tau) d\tau, \qquad (34)$$

$$I_m(t) = \chi_m I_{m-1}(t) + \hbar_I \int_0^t R_{m,I}(\tau) d\tau, \qquad (35)$$

$$V_m(t) = \chi_m V_{m-1}(t) + \hbar_V \int_0^t R_{m,V}(\tau) d\tau, \qquad (36)$$

Now, three-term approximations for T_t, I_t and V_t are expressed below.

$$T(0) = 1000, T(1) = -2.002400 \times 10^2 \hbar_T t,$$

$$T(2) = \hbar_T \left(10t + 290880 \hbar_T t^2\right) - 2\hbar_T^2 \left(2.2400t + 999197.6 t^2\right) - 2.673071 \times 10^9 \hbar_T^3 t^3$$

Now, the iterations I_i are,

$$I(0) = 0, I(1) = 200000 \hbar_I t,$$

$$I(2) = \hbar_I^2 \left(200000t - 2.60000 t^2\right) - 2.002400 \times 10^{10} \hbar_T \hbar_I t^2$$

Now, the iterations V_i are,

$$V(0) = 1000, V(1) = -2424 \hbar_V t,$$

$$V(2) = 1000 \hbar_V t^2 \left(2.402880 \times 10^{10} \hbar_T + 1.200 \hbar_I\right) - \hbar_V^2 \left(2424t - 2908.800 t^2\right)$$
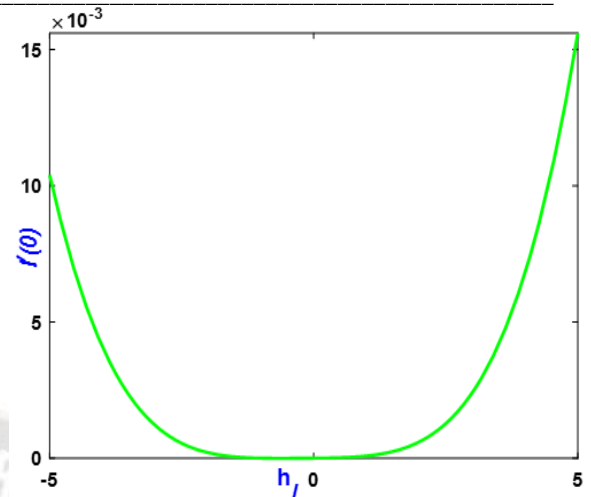

Fig. 5. $\hbar_I$ curve of infected cells at time t

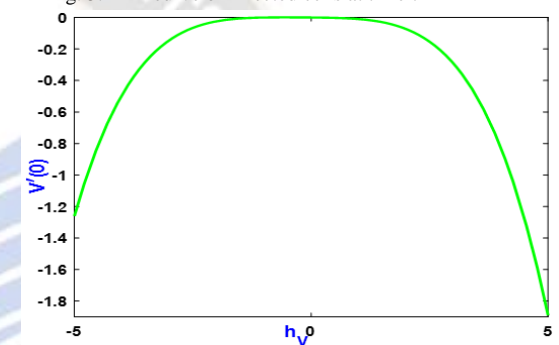

Fig. 6. $\hbar_V$ curve of virus-free cells at time t

From the above graph, we calculate the control convergence parameters

$$\hbar_T = 0.835851, \ \hbar_I = -0.836583, \ \hbar_V = -0.838172 \quad ,$$

values.

Using all these control convergence parameters, one can get,

$$T = \begin{cases} 1000 - 16.93757t + 113.7495 t^2 - 5.187718 t^3 + 8.755704 \times 10^6 t^4 \\ + 3.471499 \times 10^{10} t^5 + 7.981955 \times 10^{11} t^6 + 4.021058 \times 10^{13} t^7 \\ + 5.910754 \times 10^{18} t^8 - 9.854850 \times 10^{23} t^9 - 1.504150 \times 10^{27} t^{10} \\ + 1.227423 \times 10^{32} t^{11} + 1.509346 \times 10^{37} t^{12} - 9.479148 \times 10^{43} t^{13} \\ - 5.975036 \times 10^{48} t^{14} + 3.773786 \times 10^{53} t^{15} \end{cases}$$

$$(37)$$

$$I = \begin{cases} -4.647852 \times 10^6 t - 2.572950 \times 10^6 t^2 + 2.040757 \times 10^7 t^3 \\ -7.17647 \times 10^9 t^4 - 1.184875 \times 10^{17} t^5 + 2.209217 \times 10^{22} t^6 \\ +1.027699 \times 10^{27} t^7 - 1.217037 \times 10^{32} t^8 \end{cases}$$

$$(38)$$

Fig. 4. $\hbar_T$ curve of healthy cells at time t

_____

$$V = \begin{cases} 1000 + 5.597767 \times 10^4 t + 4.127067 \times 10^3 t^3 - 3.023513 \times 10^3 t^3 \\ +1.006857 \times 10^3 t^4 + 1.484094 \times 10^{16} t^5 - 2.656095 \times 10^{22} t^6 \\ -1.235581 \times 10^{27} t^7 + 1.463219 \times 10^{32} t^8 \end{cases}$$
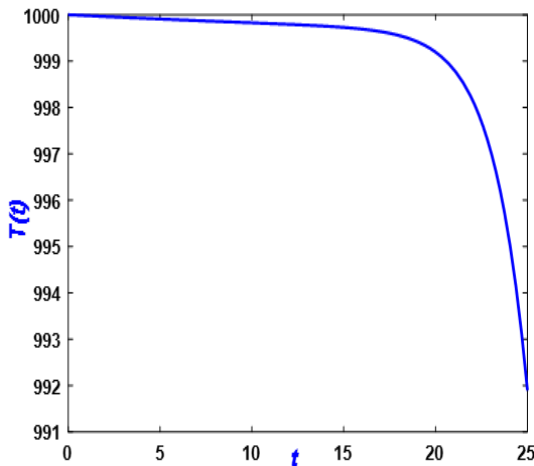
$$(39)$$

Fig. 7. Execution of Concentration of Healthy Cells with time t
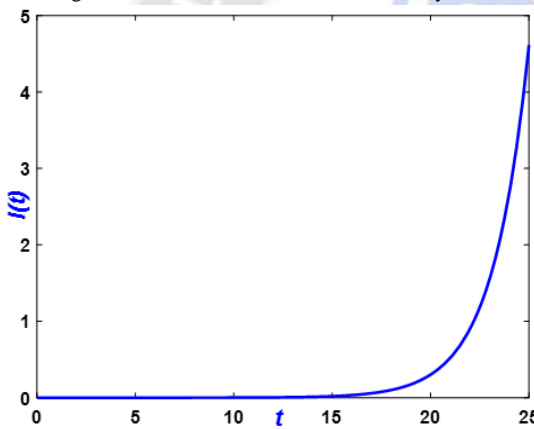

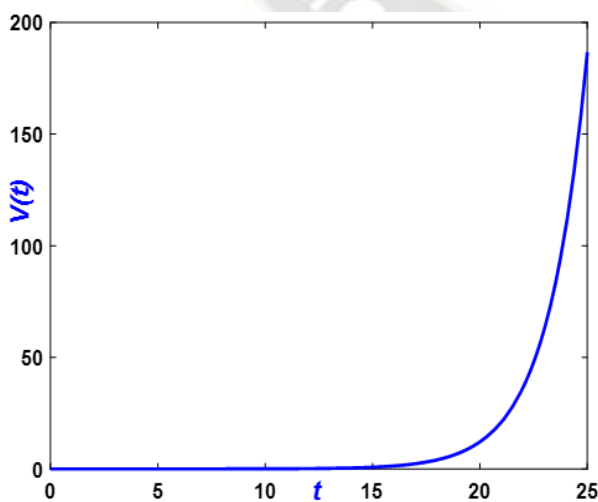Fig. 8. Execution of Infected cells with time t

Fig. 9. Execution of free HIV with time t

IV. STATISTICAL VALIDATION

We have used a one-dimensional statistical test to validate our claim with an observed sample taken from various parts of our state.

H_0: All the analytical solutions using the HAM method and RST have a significant error margin.

H_a: All the analytical solutions using the HAM method and RST have marginal error.

Here H_0 and H_a are null hypothesis and alternate hypothesis..

The samples are in numeric form.
Observed samples: 10,25,25,20,20,25,25,25,15,10
Expected value: 50, 10, 10,10,10,10,40,50,5,5

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 162.625,\text{ with a p-value of } 0.001,$$

V. CONCLUSION AND FUTURE SCOPE

We have used two standard techniques, RST & HAM. Using RST, we have concluded that HIV may spread rapidly in future. Further this study consider 100,000 records using two correlation techniques, one for the population and another for the sample; we have got 6 dissimilar records applying the RST (Reduct) technique we got HIV is the most significant medical problem that needs to be taken care of. Then we use RST strength to find a set of symptoms for HIV. Although we have got HIV symptoms by using RST but still needs further analysis as RST was useful for inaccurate data but did not give any analytical solution to overcome this problem, we are using HAM.

REFERENCES

[1] Pawlak Z. (1982): Rough Set, Int. J. of Comp. & Info. Sci., 11, 341–356.

[2] Das S., Pradhan S.K., Mishra S., Pradhan S., Pattnaik P.K. (2022): Prediction of Heart Diseases Using Soft Computing Technique, Lect. Notes in Net. and Sys., 431, 155–167.

[3] Das S., Pradhan S., Mishra S., Pradhan S., Pattnaik P.K. (2022): A Machine Learning based Approach for Detection of Pneumonia by Analyzing Chest X-Ray Images, Proceedings of the 2022 9th Int. Conf. on Comp. for Sustainable Global Dev., INDIACom 2022, 177–183

[4] Das R., Mishra J., Mishra S., Pattnaik P.K., Das S. (2022): Mathematical Modelling using Rough Set and Random Forest Model to Predict Wind Speed Proceedings of the 2022 9th Int. Conf. on Comp. for Sustainable Global Dev., INDIACom 2022, INDIACom 2022, 207–213.

[5] Das R., Mishra J., Mishra S., Pattnaik P.K. (2022): Design of mathematical model for rainfall prediction, J. of Int. Math., 25(3), 587–613

[6] Mishra S., Mohmaed A., Pattnaik P.K., Muduli K., Ahmad T.S.T. (2022): Soft Computing Techniques to Identify the Symptoms for COVID-19, Lect. Notes on Data Engg. and Comm. Tech., 86, 283–293.

1468

[7]   Nayak S.K., Pradhan S.K., Mishra S., Pradhan S., Pattnaik P.K. (2021): Rough set technique to predict symptoms for malaria, Proceedings of the 2021 8th Int. Conf. on Comp. for Sustainable Global Dev., INDIACom 2022, INDIACom 2021, 312–317, 9441312

[8]   Nayak S.K., Pradhan S.K., Mishra S., Pradhan S., Pattnaik P.K. (2022): Prediction of Cardiac Arrest Using Support Vector Machine and Rough Set Proceedings of the 2022 9th Int. Conf. on Comp. for Sustainable Global Dev., INDIACom 2022, INDIACom 2022, 164–172.

[9]   Mishra S., Mohanty S.P., Pradhan S.K. (2016): Proceedings of the 10th INDIACom; 2016 3rd Int. Conf. on Comp. for Sustainable Global Dev., INDIACom 2022, INDIACom 2016, 2018–2022, 7724620

[10]   The Global Health Observatory, Explore a world of health data

[11]   Mohameda M., Mohammed R. (2019): Modelling and prediction of activity anti-HIV molecules using soft computing techniques, 1, http://dx.doi.org/10.5958/j.0975-8070.1.2.011

[12]   Andaman N., Lightfoot J.W., Röseler W., Witte H., Lo W., Rödelsperger C., Sommer, R.J. Bacterial vitamin B12 production enhances nematode predatory behaviour, ISME, https://doi.org/10.1038/s41396-020-0626-2

[13]   Andonie R., Fabry-Asztalos L., Abdul-Wahid S., Collar C.J., Salim N. (2006): An Integrated Soft Computing Approach for Predicting Biological Activity of Potential HIV-1 Protease Inhibitors, 2006 Int. Joint Conf. on Neural Net., Sheraton Vancouver Wall Centre Hotel, Vancouver, BC, Canada, 16-21.

[14]   Akduman G., Hatipoğlu Z., Yüksekbilgili Z. (2015): A Research about Emotional Intelligence on Generations, Int. J. of Adv. Multi. Res. and Rev., 3(4).

[15]   Majdi M., Abdullah S., Jaddi N.S. (2015): Fuzzy Population-Based Meta-Heuristic Approaches for Attribute Reduction in Rough Set Theory, World Academy of Science, Engineering and Technology, Int. J. of Comp., Elec., Auto., Control and Inf. Engg., 9(12).

[16]   Chen S.B., Rajaee F., Yousefpour A., Alcaraz R., Chu Y.M. Mez-Aguilar J.G.G. Bekiros, Aly A.A., Jahanshahi, H. (2021): Antiretroviral therapy of HIV infection using a novel optimal type-2 fuzzy control strategy, Alex. Engg. J., 60(1), 1545-1555.

[17]   Ayele T.K., Goufo E.F.D., Mugisha S., (2021): Mathematical modelling of HIV/AIDS with optimal control: A case study in Ethiopia, Results in Phy.26, 104263,

[18]   Bentout S., Tridane A., Djilali S., Touaoula T.M., (2020): Age-Structured Modeling of COVID-Epidemic in the USA, UAE and Algeria, Alex. Engg. J., https://doi.org/10.1016/J.AEJ.2020.08.053

[19]   Mahmoudi M.R., Baleanu D., Qasem S.N., Mosavi A., Band S.S., Fuzzy clustering to classify several time series models with fractional Brownian motion errors, Alex. Engg. J., https://doi.org/10.1016/J.AEJ.2020.10.037

[20]   Ali A.H.D., Karapinar E., (2020): Stability data dependency and errors estimation for a general iteration method, Alex. Engg. J., https://doi.org/10.1016/J.AEJ.2020.10.002

[21]   Alrabaiah H., Arfan M., Shah K., Mahariq I., Ullah A.A., (2021): Comparative study of spreading of novel corona virus disease by using fractional order modified SEIR model. Alex. Engg. J., 60(1), 573-585.

[22]   Mohiuddin S.A., Raj K., Mursaleen M., Alotaibi A., (2020): Linear isomorphic spaces of fractional-order difference operators, Alex. Engg. J., https://doi.org/10.1016/J.AEJ.2020.10.039

[23]   Wang Y., Chen J., Ning W., Shimei Lin H.Y., Wang Z., Pang G., Chen C. (2021) A time-sensitive network scheduling algorithm based on improved ant colony optimization, Alex. Engg. J., 60(1), 107-114.

[24]   Zhang Li., Li L., Tang M., Huan Yi., Zhang X., Zhe X. (2021): A new approach to diagnosing prostate cancer through magnetic resonance imaging, Alex. Engg. J., https://doi.org/10.1016/J.AEJ.2020.10.018

[25]   Hntsa K.H., Kahsay B.N. (2020): Analysis of Cholera Epidemic Controlling Using Mathematical Modelling, Int. J. of Math. And Math. Sci., 2020, art. ID: 7369204, 13 pages, https://doi.org/10.1155/2020/7369204

[26]   Ghosh J., Mukhopadhyay S. (2011): Role of Certainty factors in generating Rough-Fuzzy Rule, Int. J. of Comp. Sci., Engg. and Appl., 1(6).

[27]   Shen Q., Jensen R. (2004): Selecting informative features with fuzzy-rough sets and its application for complex systems monitoring, Pattern Recognition, 37(7), 1351-1363.

[28]   Bal M. (2013): Rough Sets Theory as Symbolic Data Mining Method: An Application on Complete Decision Table, Info. Sci. Lett., 2(1), 35-47, http://dx.doi.org/10.12785/isl/020105

[29]   Wang L.X. (1992): Generating Fuzzy Rules by Learning from Examples, IEEE Transaction on Systems, Man, Cybernetics, 22(6).

[30]   Rizaa L.S., Januszb A., Bergmeira C., Cornelisa C., Herreraa F., S´lezakb D., Beníteza J.M. (2014): Implementing algorithms of rough set theory and fuzzy rough set theory in the R package ''RoughSets'', Info. Sci., 287, 68-69. https://doi.org/10.1016/J.INS.2014.07.029

[31]   Xin Li. (2006) Finding Optimal Reduct for Rough Sets by Using a Decision Tree Learning Algorithm, Master's Projects. 125, San José State University. DOI: https://doi.org/10.31979/etd.nuhk-2kpp

[32]   Busse J.W.G., Wojciech R. (2010): A Local Version of the MLEM2 Algorithm for Rule Induction, Funda. Info.,100, 1-18. doi-:10.3233/FI-2010-272

[33]   Uusitalo L., Lehikoinen A., Helle I., Myrberg K. (2015): An overview of methods to evaluate uncertainty of deterministic models in decision support, Env. Modelling & Software, 63, 24-31.

[34]   Baroni G., Tarantola S., (2014): A General Probabilistic Framework for uncertainty and global sensitivity analysis of deterministic models: A hydrological case study, Env. Modelling & Software, 51, 26-34.

[35]   Clark M.P., Slater A.G., Rupp D.E., Woods R.A., Vrugt J.A., Gupta H.V., Wagener T., Hay L.E., (2008): Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models, Water Resources Res.,44(12), https://doi.org/10.1029/2007WR006735

1469

[36] Kiker G.A., Bridges T.S., Varghese A., Seager T.P., Linkov I., (2009): Application of Multicriteria Decision Analysis in Environmental Decision Making, Inte. Env. Asse. and Mgmt., 1(2), 95–108.

[37] Bonano E.G., Apostolakis G.E., Salter P.F., Ghassemi A., Jennings, S. (2000): Application of risk assessment and decision analysis to the evaluation, ranking and selection of environmental remediation alternatives, J. of Haza. Mat., 71(1-3), 35-57. DOI: 10.1016/s0304-3894, (99) 00071-0.

[38] Bose U., Davey A.M., Olson D.L. (1997): Multi-attribute utility methods in group decision-making: Past applications and potential for inclusion in GDSS. Int. J. of Mgmt. Sci., 25, 691–706.

[39] Bazan J., Skowron A., Swiniarski R. (2006): Rough Sets and Vague Concept Approximation: From Sample Approximation to Adaptive Learning, LNCS, Trans. on Rough Sets V, 39-62.

[40] Butz C.J., Yan W., Yang B. (2006): An Efficient Algorithm for Inference in Rough Set Flow Graphs, LNCS, Trans. on Rough Sets V, 102-122.

[41] Perelson A.S. (1989): Modelling the interaction of the immune system with HIV, in: C. Castillo-Chavez (Ed.), Mathe. and Stat. Appr. to AIDS Epid., LNBM, 83, 350-370.

[42] Perelson A.S. (1993): Kirschner, D.E., Boer, R. De., Dynamics of HIV infection of CD4+ T-cells, Math. Biosci., 114, 81.

[43] Liao S. (2005): Comparison between the homotopy analysis method and homotopy perturbation method, Appl. Math. and Comp., 169, 1186-1194.

[44] Liao S., Pop I. (2004) Explicit analytic solution for similarity boundary layer equation, Int. J. of Heat and Mass Tran., 47, 75-85.

[45] Liao S., Fu Z., Liu S. (2005): Periodic solutions for a class of coupled nonlinear partial differential equations, Phys. Letters A, 336, 75-179.

[46] Mishra A.K., Pattnaik P.K., Mishra S.R., Senapati N. (2021): Dissipative heat energy on Cu and Al_2O_3 ethylene–glycol-based nanofluid flow over a heated semi-infinite vertical plate. J Therm Anal Calorim, 145(1), 129–137.

[47] Mohanty B., Jena S., Pattnaik P.K. (2019): MHD nanofluid flow over stretching/shrinking surface in the presence of heat radiation using a numerical method, Int. J. on Emgg. Techn., 10(2b), 119-125.

[48] Pattnaik P.K., Biswal T. (2015): Analytical solution of MHD free convective flow through porous media with time-dependent temperature and concentration, Walailak J. of Sci & Tech, 12(9), 749-762.

[49] Mishra S.R., Pattnaik P.K., Bhatti M.M., Abbas T. (2017) Analysis of heat and mass transfer with MHD and chemical reaction effects on viscoelastic fluid over a stretching sheet, Ind. J. of Phy., 2017, 91(10), 1219-1227.

[50] Pattnaik P.K., Mishra S.R., Barik A.K., Mishra A.K. (2020): Influence of chemical reaction on magnetohydrodynamic flow over an exponentially stretching sheet: a numerical study, Int. J. of Fluid Mech. Res., 47(2), 1–12.

[51] Pattnaik P.K., Jena S., Dei A., Sahu G. (2019): Impact of chemical reaction on micropolar fluid past a stretching sheet, JP J. of Heat and Mass Tran., 18(1), 207-223.

[52] Pattnaik P.K., Mishra S.R., Mahanthesh B., Gireesha B.J. Gorji, M.R. (2020) Heat transport of nano-Micropolar fluid with an exponential heat source on a convectively heated elongated plate using Numerical computation, Mult. Modelling in Mat. and Str., 16(5), 1295-1312. DOI: 10.1108/MMMS-12-2018-0222

[53] Barik A.K., Mishra S.K., Mishra S.R., Pattnaik P.K. (2020): Multiple Slip effects on MHD nanofluid flow over an inclined, radiative and chemically reacting stretching sheet by means of FDM, Heat tra. Asian res., 49(1), 477-501. DOI: 10.1002/htj.21622

[54] Mishra S.R., Pattnaik P.K., Dash G.C. (2015): Effect of heat source and double stratification on MHD free convection in a micropolar fluid, Alex. Engg. J., 54(3), 681-689.

[55] Mishra S.R., Mahanthesh B., Mackolil J., Pattnaik P.K. (2021): Nonlinear radiation and cross-diffusion effects on the micropolar nano liquid flow past a stretching sheet with an exponential heat source, Heat Tran., 50(4), 3530–3546.

[56] Ali B., Pattnaik P.K., Naqvi R.A., Waqas H., Hussain, S. (2021): Brownian motion and thermophoresis effects on bioconvection of rotating Maxwell nanofluid over a Riga plate with Arrhenius activation energy and Cattaneo-Christov heat flux theory, Ther. Sci. and Engg. Pro., 2021, 23, 100863.

[57] Pattnaik P.K., Mohapatra D.K., Mishra S.R. (2021): Influence of Velocity Slip on the MHD Flow of a Micropolar Fluid Over a Stretching Surface, Lect. Notes in Mech. Engg., 307–321.