Metaheuristic Algorithms with Applications

www.maa.reapress.com

Metaheur. Algor. Appl. Vol. 1, No. 1(2024) 88–108.

Paper Type: Original Article

Development of the Efforts-Based Outlook Learning Models Using Information Set Theory

Madasu Hanmandlu[1], **Jyotsana Grover**[2,*]

[1] Department of Electrical Engineering, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India; mhmandlu@ee.iitd.ac.in.

[2] Birla Institute of Technology and Science, Pilani, India; jyotsana.grover@pilani.bits-pilani.ac.in.

Citation:

Received: 24 June 2023 *Revised: 22 September 2023* *Accepted: 10 November 2023*	Hanmandlu, M., & Grover, J. (2024). Development of the efforts-based outlook learning models using information set theory. *Metaheuristic algorithms with applications, 1*(1), 88-108.

Abstract

This paper presents the conceptualization of efforts in the formulation of Effort-Based Outlook Learning Models (EB-OLMs), named as Efforts that Disregard Outcome (EDO), Efforts to Raise Outcome (ERO), and Efforts to Steel Outcome (ESO). The efforts are put in to achieve a goal or an objective by a contender for success. The outlook of each contender is modelled by the way efforts are contemplated. We have used some of the elements of the information set theory in the construction of EB-OLMs. The superiority of these models over some well-known Meta-heuristic techniques is demonstrated on some standard optimization functions. The design of both the classifier and predictor showcases the applicability of EB-OLMS in learning their parameters through two case studies. The first case study presents the formulation of a classifier for the detection of retinal diseases, whereas the second case study is that of a predictor for the prediction of heart diseases.

Keywords: Efforts that disregard outcome, Efforts to raise outcome, Efforts to steel outcome, Effort-based outlook learning model, Information set, Pervasive information set.

1 | Introduction

Learning is an ongoing process in humans from childhood onwards by interaction with the mother in the early years and then with the outside world via communication. The initial efforts on human learning in [1] include the proposition of Teacher-Learner Based Optimization (TLBO) where a student learns from his/her teacher and takes help from his/her friends, development of the JAYA algorithm by improving TLBO in [2], and formulation of three metaphor-less algorithms for solving the optimization problems in [3] towards the betterment of their previous methods.

Corresponding Author: mhmandlu@ee.iitd.ac.in

https://doi.org/10.48313/maa.vi.57

An evolutionary learning model, called Human Effort for Achieving Goals (HEFAG) in [4], is leveraged upon the information concept to learn its parameters. In this model, a contender for success engages self-effort exceeding the average effort while competing with the best achiever at the emulation phase, but updates this effort by an incremental effort from any two fellow contenders while cooperating with them at the boosting phase.

As an extension of the above, two more learning models clubbed under the title called Competitive-Cooperative Learning Models (CCLMs) in [5] are based on a higher form of information set concept, such as Hanman Transform (HT) and pervasive information values, elaborated later.

In the first approach, the competition with the best achiever is represented as the difference between the HT values of a contender and the best achiever, whereas that between the pervasive information values in the second approach. Both approaches use the difference between the HT values of two randomly selected fellow contenders to augment the first difference in the update equation of the learning model.

Recently, a prudent learning model was developed by Grover and Hanmandlu [6] in which competition is done with both the best and the least achievers, thus solely favouring the competition as a success mantra. In this work, we would like to investigate how each type of effort would lead to a different kind of outlook that has a bearing on the learning behaviour. We will also show the utility of outlook learning in solving several problems. The motivation for this work stems from the desire to utilize the certainty/uncertainty representation as the backbone of modelling the variables of any kind having their own distributions amenable to fit with the mathematical models.

The objectives of this paper are: 1) to expose the concepts of information sets, 2) to derive the HT divergence and pervasive HT, 3) formulation of three efforts-based learning models, 4) to illustrate the power of certainty/uncertainty representation on two case studies, and 5) to deliberate on the suitability of the information set concepts for the modelling of certain modules.

The rest of the paper is structured as follows: Section 1 presents the characterization of an outlook resulting from three types of efforts. The rudiments of the information set theory are covered in Section 3, and the formulation of the Efforts-Based Outlook Learning Models (EB-OLMs) is given in Section 4. The design of an information set-based classifier, regression-based predictor, and DeepInfoNet-based GAN, including an exposition of rough information sets, is the concern of Section 5. Two case studies, with the first on the detection of retinal diseases using the classifier and the second on the prediction of heart diseases using the predictor, are discussed in Section 6, where their parameters are learned using the proposed EB-OLMs. The conclusions are relegated to Section 7.

2|The Characterization of Efforts-Based Outlook Learning Models

The activities are of different kinds, such as business, economic, sports, religious, community, etc., and these are performed through efforts by an individual or a group to achieve a purpose or an objective. For any activity, say, a learning activity, the efforts have to be put in to achieve a goal. The way the efforts are made while performing an activity reflects the attitude of a person towards life. Hence, the efforts mirror the human outlook. As a matter of fact, we need not associate any outlook with the efforts related to an activity. In that case, we would be disregarding the variability in the efforts. But we want to see this variability in one's efforts as a reflection of his/her outlook.

In this, we focus on the efforts of a contender aimed at an activity that involves finding a solution to a problem. We assume a group or population of contenders each trying to find an optimal solution by optimizing an objective function. We propose to categorize the efforts of each contender into three types, with each type leading to one kind of outlook. It is opined that a contender can display any one of the three kinds of outlooks depending on how he/she performs in the efforts to achieve the goal. The first type of efforts that aim at gaining his/her own outcome without ever bothering about the outcomes of others leads

to an outlook termed as "Efforts that Disregard Outcome (EDO)" abbreviated as EDO, and the contender with this outlook mode is therefore termed as EDOY.

The second type of efforts is the result of possessing the competitive spirit in a contender. In this mode, the contender compares his/her outcome of efforts with those of other contenders and then tries to raise his/her outcome through increased efforts to score over the best performing contender, called the best achiever. This type of effort is termed as "Efforts to Raise Output (ERO)" abbreviated as ERO, and the contender displaying this outlook mode is termed as EROY.

The third kind of effort prefers stealing the outcome of another contender instead of earning his/her own outcome, as the contender seeks an easy way to achieve a goal. Hence, his/her efforts are directed in eyeing others' outcomes rather than exerting them. So, the efforts are termed as "Efforts to Steal Outcome (ESO)" abbreviated as ESO, and the contender in this outlook mode is termed as ESOY.

A contender can be in any outlook at any time depending on what triggers in him/her on seeing the outcome of his/her efforts. But success weds a contender when he/she cherishes all kinds of outlooks. Moreover, while expressing an outlook as a learning component, we need the mathematical support that is forthcoming from information set theory. We utilize all three learning outlook components together to minimize the loss function to reach the goal.

Moreover, we intend to seek a population-based evolutionary learning in which several contenders participate to achieve the goal. There exist several strategies of species such as ants, birds, bacteria, particles, animals, etc. to explore the solution paths in the search space, and the final solution is reached by attaining the minimal loss function value. As humans are gifted with better skills than those of species because of their communication abilities, we prefer to deal with their outlooks towards learning.

The loss functions consisting of the unknown parameters are problem-specific but domain independent, and they are framed to reflect what is to be achieved from a problem. In our case, the problem is to achieve a goal by an appropriate choice of efforts. The goal is achieved if and only if the loss function attains the minimum value. The loss function is taken as a function of the error between what is expected and what is obtained from efforts, called an outcome. However, depending on the complexity of a loss function that grows in proportion to the number of unknown parameters, its minimization through learning of these parameters by the classical optimization techniques becomes difficult. So, we go in for evolutionary learning methods that employ a set of populations to search the solution space.

To meet our objective, the number of populations is the same as the number of contenders. The efforts, i.e., the unknown parameters, in the loss function are learned by minimizing it.

3 | A Brief Note on Information Sets

Based on the above concepts, the representation of either certainty or uncertainty associated with the attribute, termed the Information Source (IS) values, rules gains prominence because of its problem-solving abilities. The choice of either certainty or uncertainty representation is important for a particular application. The importance of this choice will be clear in the case studies.

A set of IS values, for the sake of generalization in the terminology of the information set theory, is modeled by a mathematical function called the Membership Function (MF) by assuming a distribution which could be statistical, like Gaussian /generalized Gaussian, or random, like exponential. In a fuzzy set, a pair comprising an attribute value and the corresponding MF value is an element. The MF value connects the attribute value with a degree of association to the set. As such, the complement MF is deprived of association to the set. This drawback is mitigated in the information set constituting a set of information values, which results from the certainty/uncertainty representation of IS values through the Hanman-Anirban (HA) entropy function [7], by which the two components of each pair are merged into a product called an information value, thus invalidating the degree of association. The significance of the variable is judged by how much certainty/uncertainty it owes to the information set. The HA entropy function is the offshoot of the Pal and

Pal (PP) entropy function [8], and this being information theoretic can deal with both the probabilistic, P_x and possibilistic IS values, I_x and is defined as:

$$H^{HA} = \frac{1}{n}\sum_{i=1}^{n}\{I_{xi}e^{-(aI_{xi}^3+bI_{xi}^2+cI_{xi}+d)}\}, \quad (1)$$

with substitution of $a = 0,\ b = \frac{1}{2\sigma_x^2},\ c = \bar{I}/\sigma_x^2$ and $d = \bar{I}^2/2\sigma_x^2$ for its parameters based on the statistical parameters: $\bar{I}$ (the mean) and σ_x^2 (the variance), *Eq. 1* becomes:

$$H^{HA} = \frac{1}{n}\sum_{i=1}^{n} I_{xi}\mu_{xi}, \quad (2)$$

where $\mu_{xi} = e^{-\frac{(I_{xi}-\bar{I})^2}{2\sigma_x^2}}$ is the Gaussian MF derived from the polynomial exponential gain function and the product $I_{xi}\mu_{xi}$ is the information value with the set denoted by: $H^{HA}(I) = \{I_{xi}\mu_{xi}\}$.

The Mamta-Hanman (MH) entropy function in [9] generalizes the HA entropy function by modifying both the exponential gain function and IS values as follows:

$$H_p^{MH} = \frac{1}{n}\sum_{i=1}^{n} I_{xi}^{\gamma}e^{-(cI_{xi}^{\alpha}+d)^{\beta}}, \quad (3)$$

where α, β, γ, c, and d are the real constant parameters. Representing *Eq. 3* similar to *Eq. 2*, we get, $H_p^{MH} = \frac{1}{n}\sum_{i=1}^{n} I_{xi}^{\gamma}\mu_{xi}^{\beta}$, where μ_{xi}^{β} is the generalized Gaussian MF that takes different shapes for the values of β in the range (0, 5).

The gain function need not be from the same set of IS values, I_{yi} , it could be from another set, I_{yi}. To represent this, we make use of the Adaptive Heterogeneous (AH) HA entropy function, given by

$$H_{ah}^{HA} = \frac{1}{n}\sum_{i=1}^{n}\{I_{xi}e^{-\left(a_yI_{yi}^3+b_yI_{yi}^2+c_yI_{yi}+d_y\right)}\}. \quad (4)$$

By choosing $a_y = 0, b_y = \frac{1}{2\sigma_y^2}, c_y = \frac{-\bar{y}}{\sigma_y^2}$ and $d_y = \frac{-\bar{y}^2}{2\sigma_y^2}$ where $\bar{y}$ is the mean, and σ_y^2 is the variance in *Eq. (4)*, we obtain,

$$H_h^{HA} = \frac{1}{n}\sum_{i=1}^{n} I_{xi}\mu_{yi}, \quad (5)$$

where μ_{yi}, the Gaussian MF defined based on the distribution of $\{I_{yi}\}$ is evaluating another distribution $\{I_{xi}\}$. Unlike in a fuzzy set, an MF here is empowered to act as an agent that can work on probabilities, assume a negative role as the complement, and become an evaluator, thereby having an extended scope outside the four walls of degree of association.

Representation of higher-level certainty/uncertainty representation: a brief note on the representation of certainty/ uncertainty is the need of the hour. Any variable having the IS values in a set can have either a statistical distribution (i.e., it can be fitted with a Gaussian MF), or a random distribution (i.e., it can be fitted with an exponential function). Application HA entropy converts the IS values into the information values, the sum of which yields the former case, whereas uncertainty in the latter case. The original impetus for the uncertainty representation is provided in [10]. For the higher-level representation of uncertainty, the possibilistic transforms based on the adaptive HA and MH entropy functions only provide recourse. These transforms are vital to the development of not only the learning models but also the design of a classifier and predictor in this paper.

To derive the HT, we invoke the adaptive HA entropy function in which the parameters are variables. This is defined as [11],

$$H_a^{HA} = \frac{1}{n}\sum_{i=1}^{n}\left\{I_{xi}e^{-(a_x I_{xi}^3 + b_x I_{xi}^2 + c_x I_{xi} + d_x)}\right\}. \quad (6)$$

For $a_x = b_x = d_x = 0$ and $c_x = \mu_{xi}$, *Eq. 6* becomes HT derived for the first time in [12]:

$$H_T^{HA} = \frac{1}{n}\sum_{i=1}^{n} I_{xi}e^{-\mu_{xi}I_{xi}}. \quad (7)$$

The corresponding MH transform is obtained as $H_T^{MH} = \frac{1}{n}\sum_{i=1}^{n} I_{xi}e^{-I_{xi}^{\gamma}\mu_{xi}^{\beta}}$ from the adaptive version of the MH entropy function [13] on replacing c with $c_x = \mu_{xi}$ and d with $d_x = 0$ in *Eq. (3)*. This is a higher-level representation of uncertainty because the exponential gain function is a function of the information values, $\mu_{xi}I_{xi}$, which leads to either certainty or uncertainty depending on whether the agent μ_{xi} is from a statistical or random distribution. We will now derive some extensions of HT necessary for the learning model.

Outcome-based HT: while dealing with an optimization problem involving the learning of the unknown parameters, denoted by a weight vector, w_i termed as an effort vector using an evolutionary and population-based learning. We invariably employ several competitors, each of which finds a fitness/outcome by applying w_i on an objective function, $f_k = F(rs, w_k)$, where rs is the resource and w_k is the effort vector of the k^{th} competitor. There are as many outcomes as the number of competitors. We can easily determine the outcome based on the k^{th} MF, μ_{fk} based on f_k and $f_{av.}$ Then, the outcome-based HT can be written as:

$$H_{fT}^{HA} = \frac{1}{n}\sum_{i=1}^{n} w_i e^{-\mu_{fi}w_i}. \quad (8)$$

We will be using this formula in the EB-OLM discussed later.

Derivation of the divergent HT and HT ratio: to this effect, consider two ISs I_{xi} and I_{yi} with the corresponding MFs, μ_{xi} and μ_{yi}. The difference between their information is computed from:

$$\Delta H_i = I_{yi}\mu_{yi} - I_{xi}\mu_{xi}. \quad (9)$$

We will now incorporate ΔH_i from *Eq. (9)* by setting a_x, b_x to 0, $c_x = -\mu_x$ and $d_x = \mu_{yi}I_{yi}$ into *Eq. (6)* to result in what we call the HT divergence that gives the uncertainty because of the external agent, μ_{yi}:

$$H_{\Delta T}^{H} = \frac{1}{n}\sum_{i=1}^{n}\left\{I_{xi}e^{-(\mu_{yi}I_{yi} - \mu_{xi}I_{xi})}\right\}. \quad (10)$$

Note that the parameters $a_x, b_x$and c_x contribute to the internal agent, μ_{xi} whereas d_x, to the external agent, μ_{yi}.

Dividing the above ΔH_i by $I_{yi}\mu_{yi}$, we get the change in the information ratio as:

$$\frac{\Delta H_i}{I_{yi}\mu_{yi}} = 1 - \frac{I_{xi}\mu_{xi}}{I_{yi}\mu_{yi}}. \quad (11)$$

By taking the ratio between two information values from *Eq. (11)*, setting a_x and b_x to 0, $d_x = -1$ and $c_x = \mu_{xi}/(I_{yi}\mu_{yi})$ and then using the logarithmic approximation in *Eq. (6)* convert it into the Kullback-Leibler (KL) divergence as follows:

$$H_{RT}^{H} = \frac{1}{n}\sum_{i=1}^{n} I_{xi}e^{-\left(\frac{\mu_{xi}I_{xi}}{\mu_{yi}I_{yi}} - 1\right)} = -\frac{1}{n}\sum_{i=1}^{n} I_{xi}\log\left(\frac{\mu_{xi}I_{xi}}{\mu_{yi}I_{yi}}\right). \quad (12)$$

We will now show how a non-Membership Function (nMF) can have a role in ameliorating the inaccurate fuzzy modelling. This development has led to the emergence of a pervasive Membership Function (pMF).

Pervasive information set: when the fuzzy model, MF, i.e. μ_{xi}, is unable to model the distribution satisfactorily, then the intuitionistic fuzzy set due to Atanassov [14] comes to our rescue by introducing an nMF, i.e. v_{xi} , to correct for the deficiency in MF. How much their sum value deviates from 1 evokes the concept of hesitancy, given by:

$$h_{xi} = 1 - (\mu_{xi} + v_{xi}) = 1 - P_{xi}. \quad \textbf{(13)}$$

In order to vary the values of hesitancy, a hesitancy degree, γ is taken as the power on both μ_{xi} and v_{xi}.

$$\gamma.h_{xi} = 1 - (\mu_{xi}^{\gamma} + v_{xi}^{\gamma}) = 1\text{-}P_{xi}^{\gamma}. \quad \textbf{(14)}$$

We reveal the importance of *Eqs. (13)* and *(14)* in passing, as their differencing has led to the birth of incremental hesitancy features that are used in the design of a color-based detector for detection of sign boards from the world scenes in [15]. In this work, we are concerned with the pMF, denoted by P_{xi} and the variable pervasive MF, denoted by P_{xi}^{γ}. They are the sum of MF and nMF. The pervasive information set denoted by $\{I_{xi}. P_{xi}^{\gamma}\}$ extends the information set to encompass the intuitionistic fuzzy set by borrowing terms like nMF, hesitancy degree, and hesitancy function into its fold. To get over the unavailability of v_{xi}, we replenish it with the Yager complement of μ_{xi}, $(1 - \mu_{xi}^{s})^{1/s}$, where s is a scale parameter to fine-tune its value. Replacing μ_{xi}^{γ} with P_{xi}^{γ} converts the HT into the pervasive Hanman Transform (pHT), $H_{PT}^{HA} = \frac{1}{n}\sum_{i=1}^{n} I_{xi} e^{-P_{xi}^{\gamma} I_{xi}}$. This is of interest in the formulation of the Outlook Learning Models (OLMs).

4 | Formulation of Outlook Learning Models

The three kinds of outlook are proposed depending on how a contender for success conceptualizes the efforts with regard to an outcome. The three perceived outlooks are due to the EDO, ERO, and ESO. We make an important assumption that all the contenders make efforts under these three independent categories, so that their influences are additive. The non-additive case is not taken up here for the sake of simplicity in modelling.

The contender with EDO characteristic is termed as EDOY, and that with ERO characteristic is termed as EROY. Lastly, the contender with ESO characteristics is termed ESOY. As we are attempting to learn, let us pinpoint their specific characteristics suitable for learning.

The learning component in EDO mode doesn't depend on the outcome of one's efforts. Whatever one does is aimed at gaining experience sans competition with others. In the parlance of information set theory, EDOY gauges his/her efforts based on how much they have contributed to the output. In ERO mode, EROY compares his/her performance in terms of the outcome of his/her efforts with that of the best achiever.

In the ESO model, ESOY is bent on usurping the fruits of others' labour or simply saying, they snatch away the outcomes of others' efforts.

Now the ground is set for the formulation of EB-OLM. Let F (x, w) be a function of resource values x and the unknown parameters as the IS values in the learning model. The three components of the learning model enter into a race to learn these parameters by optimizing an objective function.

Let us first focus on the component of EDO. EDOY doesn't compete with anybody and doesn't expect an outcome commensurate with his/her efforts. We will use the pervasive HT to update the efforts of EDOY.

The component of EFO strives to improve his/her outcome as he/she competes with the best performer by making efforts. We can use either HT divergence or HT ratio to update the efforts of EROY.

The component of ESO tries to steal the fruits of others' efforts. This behavior is opposite to the EDO mode that shuns all sorts of external inducements.

In our deliberations on learning, we are concerned with the representation of certainty in the efforts, which arises from fitting their distribution with a suitable MF. The efforts of EDOY, EROY and ESOY are denoted by "w" with three arguments.

The first one denotes the t^{th} iteration, the second one denotes the k^{th} contender; it could be one of the three kinds of outlook: EDOY, EROY and ESOY, denoted by D, R and S as subscripts on w respectively, and the third one denotes the l^{th} effort. Accordingly, the weight/effort vectors of EDO, EfgvRO and ESO are w_D (t, k, l), w_R (t, k, l) and w_S (t, k, l). The overall weight vector isw2(t, k, l).

As an EROY competes with others, we assume it to be the best-performing contender or best achiever, whose effort is denoted by w (t, best, l). To evaluate the outcome of any contender, all the efforts are needed.

As we are interested in developing the learning model, our objective is to learn the efforts of the k^{th} contender, that is destined to play a role as per the respective outlook perspective.

We now formulate the OLM learning accrued by a particular type of effort. Let us model the component of OLM earning in EDO mode in which an expectation of any sort is abhorred. As EDOY seeks knowledge, we associate two features with this mode. These include: 1) the first is reinforcement of his/her current efforts based on the previous outcomes, and 2) the second is the use of the pervasive information in the domain of activity due to the internal (self) and the external agents.

Notations used in the learning models: in this, x is considered here as a resource rather than an IS, I_{xi}. The weight vector w is dependent on the manoeuvrable iteration t, contender k, and effort l. It is therefore denoted by w (t, k, l). When all efforts, l = 1,..., L, are used, then the weight/effort vector is indicated by w(t, k). The outcome of the k^{th} contender, symbolized by f_k= F(rs, w(t, k) banks upon on the resources rs and efforts l of the k^{th} contender at iteration t.

4.1|Efforts that Disregard Outcome-Component

As an EDOY leverages his/her past outcomes to improve upon the present efforts, a reinforcement factor is devised as a sigmoid of an average of the (t-1) elapsed values of outcome from i =0 to i=t-1, computed as:

$$\gamma(t,k)=\frac{1}{1+e^{-\Delta F(t,k)}}, \tag{15}$$

where the average is given by

$$\Delta F(t,k)=\frac{1}{t-1}\sum_{i=1}^{t-1}F(rs,w(i,k))-F(rs,w(i-1,k))=\frac{1}{t-1}\sum_{i=1}^{t-1}f(i,k)-f(i-1,k).$$

The above reinforcement factor is chosen to be simple, as it is not associated with a reward function based on some policy. Here, its role is to give weightage to the weight vector of EDO. For instance, we could give more weightage to the latest fitness values and the least to the past values. Note that we have suppressed the abbreviations rs and w in f. The pHT is modelled as pHT weighted with the reinforcement factor that accounts for the past outcomes, given by:

$$w_D(t+1,k,l)=\gamma(t,k).w_D(t,k,l)e^{-\mu_P(t,k)w_D(t,k,l)}, \tag{16}$$

where $\mu_P(t,k)=\mu_D(t,k)+v_D(t,k)$ and $\mu_D(t,k)=e^{-|f(t,k)-f(t-1,k)|}$.

4.2|Efforts to Raise Outcome-Component

In this, the HT divergence based on the difference between two information values of the best achiever, $\mu(t,best)w_R(t,best,l)$ and the k th contender $\mu(t,k)w_R(t,k,l)$ is computed using *Eq. (16)* as:

$$w_R(t+1,k,l)=w_R(t,k,l)e^{-(\mu(t,best)w_R(t,best,l)-\mu(t,k)w_R(t,k,l))}. \tag{17}$$

As shown in *Eq. (12)*, the above HT divergence can also be written as :

$$w_R(t+1,k,l) = w_R(t,k,l).\log(\frac{\mu_R(t,k)w_R(t,k,l)}{\mu_R(t,best)w_R(t,best,l)}), \quad \textbf{(18)}$$

where the outcome-based MFs of the best achiever and the k[th] contender are:

$$\mu_R(t,best) = e^{\frac{-(f_{best}-f_{av})^2}{2\sigma_f^2}},$$

and

$$\mu_R(t,k) = e^{\frac{-(f_k-f_{av})^2}{2\sigma_f^2}}.$$

With f_{best}=F(w(t, best)), f_k=F(w(t, k)), and f_{av} as the average outcome.

It may be noticed that we have associated the advanced concept called pervasive information and reinforcement with the EDO component of learning, as EDOY is entrusted with a plan at his disposal for the next outcome, so it is vested with the ability to utilize the past and present information. This capability is a rare kind of outlook of a contender, whereas an inquisitive higher form of information called transform is bestowed with the ERO component of learning. As ESO has a tendency to steal, we have to brace for a new concept next.

4.3 | Efforts to Steel Outcome-Component

We presume that the characteristic of ESOY is to steal others' fruit (i.e., outcome). As per the information set, MF gives an association to a concept, whereas the nMF has no allegiance to the concept. As we know, any person can develop the habit of stealing the outcome of others' efforts (nMF) in case he/she doesn't want to earn his/her outcome by his/her efforts (MF) or never wants to exert. Then he/she directs his/her efforts to steal; it can be presumed that ESOY eyes the other's outcome. In that case, the component of k[th] ESOY is modelled as:

$$w_S(t+1,k,l) = v_S(t,k,l).f(t,q), \quad \textbf{(19)}$$

where we make use of the product of the nMF of the k[th] ESOY, v_s (t, k, l), and the outcome of the q[th] contender, f(t, q), to represent the usurping of the q[th] contender's outcome by the k[th] contender that exerts none of its efforts. If an outcome is not produced by one's efforts, the claim to it is said to be false. Hence, his/her MF cannot be used, but only nMF.

It may be noted that $v_S(t,k,l)$is taken as the Yager complement of $\mu_S(t,k,l)$ that is given by

$$\mu_S(t,k,l) = e^{-\left\{\frac{(w_S(t,k,l)-w_{av}(t,k))^2}{2\sigma_{w_S}^2}\right\}}, \quad \textbf{(20)}$$

where the mean denoted by w_{av} (t, k) and the variance by $\sigma_{w_S}^2$ computed based on the l[th] efforts of all contenders.

The Efforts Based Outlook Learning Model (EBO-LM): having derived all three components of learning, the weight vector is updated as:

$$w(t+1,k,l) = w(t,k,l) + r_1 w_D(t+1,k,l) + r_2\, w_R(t+1,k,l) + r_3\, w_S(t+1,k,l), \quad \textbf{(21)}$$

where r_1, r_2 and r_3 are the random numbers. When t=0, w (0, k, l) = 0, whereas the other three weights are initialized with random values.

We have refrained from discussing several ways by which the variants of the learning models can be generated, as that would distract us from our avowed objectives. So, we present an algorithm for the implementation of EB-OLM.

4.4 | An Algorithm for Effort-Based Outlook Learning Models

Step 1. Choose an objective function to be minimized along with the input data. Specify the number of weight vectors (populations) and the number of iterations.

Step 2. Assume random weight vectors. Let Iter be equal to 1.

Step 3. Compute the outcomes due to the randomized weight vectors and find the outcome MFs.

Step 4. Compute the reinforcement factor using *Eq. (15)* and evaluate the EDO learning component using *Eq. (22).*

Step 5. Compute the learning component of EROY using *Eq. (17)* or *Eq. (18)* by evaluating the outcome-based MFs of the best achiever and the k^{th} contender.

Step 6. Compute the learning component of ESOY using *Eq. (19)* and find the nMF of ESOY using the effort-based MF using *Eq. (20).*

Step 7. Update the previous weight vectors using three learning components in *Eq. (21).*

Step 8. Repeat *Steps 3-7* until we achieve convergence in the weight vectors or the minimum objective function value.

In the above algorithm, we have used all three components in the updating equation of the learning model. For the algorithms in [3], three model equations need not be applied. Depending on the objective function, we need one or two model equations, and these are applied sequentially one after another. Very rarely do we need the third model equation. Based on this procedure, the above algorithm is modified.

4.5 | Modified Algorithm of Effort-Based Outlook Learning Models

Step 1. Find the reinforcement factor and evaluate the ENO learning component.

Step 2. Update the model equation with only the ENO component. If the objective function is reduced drastically, go to *Step 3*.

Step 3. Evaluate the outcome of the membership best performer and also of an EROY. Compute the ERO learning component.

Step 4. Update the model equation from *Step 4* with the ERO component.

Step 5. If the objective function is reduced drastically, then go to *Step 3*.

Step 6. Find the nMF of ESOY using either the outcome-based or the effort-based MF and compute the ESO learning component.

Step 7. Update the model equation from *Step 8* with the ENO component.

Step 8. Repeat *Steps 3-10* until we achieve convergence in the weight vectors or the minimum objective function value.

5 | Applications of Effort-Based Outlook Learning Models

We will briefly state some applications of this model for learning the parameters. They include:

I. The design of a classifier.

II. The formulation of regression and forecasting models.

III. The development of new neural networks.

5.1 | The Design of a Classifier

The design of classifiers based on information sets can be seen in [16]. Assume that the Feature Vectors (FVs) of a dataset are normalized, and then they are split up into the training and test sets. The number of classes is assumed to be l. The length of an FV is m, and the number of FVs in a class is n. Let f_{tr} (i, j) and f_{te}(j) with i = 1,.., n and j = 1, ..., m be the training and test FVs be abbreviated as tFVs and teFVs respectively. Our objective is to find the statistics of all features in a class. For this, we compute the mean and the variance of each FV across all tFVs of a class to form an information set by fitting the Gaussian MF. By this procedure, we get m MFs yielding m information sets, with each set having n information values. Let $\mu_{tr}(i,j)$ be the MF vectors of $f_{tr}(i,j)$. Next, we evaluate their HT from:

$$H^{H}_{T,tr}(i,j) = f_{tr}(i,j)e^{-(f_{tr}(i,j).\mu_{tr}(i,j))}. \quad (22)$$

Approach 1 (Without learning). In this approach, HT is applied to tFVs of a class by finding the MF vector that requires an average tFV and a vector of variances. The FV giving the minimum HF value is termed the support vector of the class based on the certainty concept. All other FVs with higher HF values possess more certainty to the class, and hence they are well within the boundary of a class. But the support vector being on the boundary of the class is most critical. This concept was used in the design of an optimal classifier using the prudent learning model in Grover and Hanmandlu [6]. The support vector for the l^{th} class is computed from:

$$H^{H}_{T,tr}(i,l) = \left\{ \sum_{j}^{m} f_{tr}(i,j)e^{-(f_{tr}(i,j).\mu_{tr}(i,j))} \right\}_{l}; i = 1,..,N_{tr}; l = 1,..,N_c, \quad (23)$$

where N_{tr} is the number of tFVs of each class and N_c is the number of classes. It may be noted that the subscript l on the r.h.s. of *Eq. (23)* indicates that the operations in the curly bracket are performed on tFVs of class l. By applying Min_i on $H^{H}_{T,tr}(i,l)$ fetches us the minimum value of HT for the specific index im, given by

$$H^{H}_{T,tr}(im,l) = Min_i\{H^{H}_{T,tr}(i,l)\} = \sum_{j}^{m} f_{tr}(im,j)e^{-(f_{tr}(im,j).\mu_{tr}(im,j))}|_{l}. \quad (24)$$

This Equation leads to a support vector for each l, $f_{tr}(im,j)e^{-(f_{tr}(im,j).\mu_{tr}(im,j))}$.

While testing to find the label of a teFV, we will invoke the HT divergence. As we don't have statistics governing the teFV, we will use the statistical parameters (i.e., mean and variance) of tFVs to compute MF, $\mu_{te}(j)$ of a teFV. To make it compatible with μ_{tr}, the test MF is updated w.r.t. each support vector of class l from:

$$\hat{\mu}_{te}(j) = \mu_{te}(j)\{e^{-(f_{tr}(im,j).\mu_{tr}(im,j))}|_{l}, \quad (25)$$

with the help of $\hat{\mu}_{te}(j)$, we can frame the HT divergence as:

$$H^{H}_{\Delta T}(l) = \sum_{j}^{m} f_{tr}(im,j)\{e^{-(f_{tr}(im,j).\mu_{tr}(im,j))}|_{l} - e^{-(f_{te}(j).\hat{\mu}_{te}(j))}\}. \quad (26)$$

The identity (Id) of the unknown class is found by determining the infimum out of all classes from:

$$Id = Inf_l\{H^{H}_{\Delta T}(l)\}. \quad (27)$$

Approach 2 (With learning). As we had both FVs and MF vectors, there was no need for a learning model in the above classifier. But to see the power of the learning model, we shall introduce a power β on $f_{tr}(i,j)$ and initialize its values randomly. We create as many values as the number of tFVs of a class and form an effort vector. We aim to find the support vector of each class for learning β. So, we compute the uncertainty representation of each tFV of the l^{th} class using the HT divergence as:

$$H_{\Delta T}^{H}(i,l) = \sum_{j}^{m} f_{tr}^{\beta}(i,j)\{e^{-(f_{tr}(i,j).\mu_{tr}(i,j))}|_l - e^{-\left(f_{te}^{\beta}(j).\mu_{te}(j)\right)}\}. \quad (28)$$

Let $H_{\Delta T}^{H}(i,l)$ be considered as the fitness/outcome value, $f(t,k,le)$, where t denotes the iteration; k denotes the i^{th} contender, and le denotes the effort/weight vector. We invoke the learning model by providing it a set of fitness values and effort vectors. The learning model delivers the minimum value of $H_{\Delta T}^{H}(im,l)$ for the index im from tFV. Next, identifying the class label of teFV using *Eqs. (25)-(27)* is a cake walk based on the HT divergence.

5.2 | The Design of A Forecaster

Forecasting is an important problem in many applications such as forecasting of weather, demand, rain, stock market, election, etc. In this work, we address this problem under the framework of information set theory. For simplicity, we consider two categories of variables: endogenous and exogenous. Both of these types may refer to any quantities depending on an application. We would like to adopt the window-based approach in which we fix the time slot that could be a month, a year, or several years. In this time slot, the data is procured either day-wise or week-wise, and consideration of the lead time varies accordingly. Let us denote I_y as an independent or endogenous variable and I_x as an external or exogenous variable. The past values of these variables are assumed to form fuzzy sets. Let H_y^H be the information set of endogenous variables I_y, and H_{x1}^H, H_{x2}^H,.., H_{xq}^H be the information sets of exogenous variables I_{x1}, I_{x2},.., I_{xq} respectively. These variables take mostly real values, but a few of them may have truth/binary values; then we consider their probabilities.

To represent the past values, an integer of 1 is subtracted from an instant, k. The number of past values to be considered for each exogenous variable depends on how much influence it has on the endogenous variable. To determine this dependency, one way is to perform correlation analysis between the endogenous variable and the exogenous variables. The choice of the number of past values of the endogenous variable can be determined by performing clustering on the values of the endogenous variable or can be arrived at by a trial-and-error process. Alternatively, we can also go in for the structure determination approaches of fuzzy modeling. In the present study, we intend to develop an information set-based regression model linking an endogenous variable with the exogenous variables.

The proposed regression model gives a relation between the present endogenous value and the past information of values of the endogenous and the exogenous variables. This model is mathematically expressed as:

$$y_f(k/k-l) = \sum_{l}^{ly} w_y.H_y^H(k-l) + \sum_{l}^{lx} w_{x1}.H_{x1}^H(k-l) + \sum_{l}^{lx} w_{x2}.H_{x2}^H(k-l) + \cdots + \sum_{l}^{lx} w_{xq}.H_{xq}^H(k-l), \quad (29)$$

where the information values are taken as $H_y^H(k-l) = I_y(k-l).\mu_y(k-l)$ and $H_{xi}^H(k-l) = I_{xi}(k-l).\mu_{xi}(k-l)$. The MFs are assumed to be triangular or trapezoidal. We can also use pMF to improve the regression model. The unknown weight parameters in *Eq. (29)* are learned using OLM. In case the forecasting error is

to be utilized as another exogenous variable, it can be included as an additional term in the model, shown as under:

$$I_{yf}(k+1/k) = I_{yf}(k/k-1) + w_e.e_f(k/k-1), \quad \textbf{(30)}$$

where $e_f(k/k-1) = I_y(k) - I_{yf}(k/k-1)$.

In this context, development of the information set-based regression model can be seen in [17].

5.3|The Design of Generative-Adversarial Network Using DeepInfoNet

The concepts of EB-OLM can be easily ported to deep learning neural networks in creating new architectures from them. We have already made a step in this direction by creating HanmanNets through the modification of ResNet [18]. Inspired by this work, DeepInfoNet is evolved from two handcrafted Convolutional Neural Networks (CNNs). We intend to design the information set-based GAN architecture using DeepInfoNet.

A brief description of this net is as follows: the specialty of DeepInfoNet is that it is built block-wise with the scaffolding of two convolutional layers and one pooling layer in every block. A specific number of kernels /filters of size 3x3 is engaged in a convolutional layer, whereas a stride of 2x2 is made in the max-pooling layer. Each kernel performs a convolution operation on an input image, and this is followed by batch normalization.

The features extracted from feature maps of the first block are examined first. If they fail to meet the requisite classification accuracies, then these features are concatenated with those from the feature maps of the second block for the purpose of reevaluation. If the concatenated features succumb to the same fate, accessing the feature maps of the third block is the last and final recourse.

Let us allot one network for the generator DeepInfoNet catering to the training and another for the adversary DeepInfoNet catering to the testing. A kernel function convolving with the input image at every pixel location incarnates into an MF matrix. Let NF denote the number of feature maps, also referring to the number of kernels/filters, and xy, a pixel location in a feature map. The pixel intensity in a feature map μ'_{xy} is designated as the MF, whereas $\mu'_{n,xy}$ is the nMF, with both having the superscript apostrophe ('). Resizing of the input image Ixy to that of the feature map makes it appear as I'_{xy}. The conglomeration of the pixel intensities from all feature maps at a location xy gives rise to MF as $\mu'_{xy,j}$ and to nMF as $\mu'_{nxy,j}$, j=1 to NF.

To show how to generate $f_{tr}(i,j)$ and $f_{te}(i,j)$ in the context of GAN, we define the MH intuitionistic fuzzy entropy function as:

$$H_F^{MH} = \frac{1}{n^2}\sum_{x=1}^{n}\sum_{y=1}^{n} \mu_{xy}^{\gamma} e^{-(c1(.)\mu_{xy}^{\alpha}+d1(.))^{\beta}} + \mu_{n,xy}^{\gamma} e^{-(c2(.)\mu_{n,xy}^{\alpha}+d2(.))^{\beta}}. \quad \textbf{(31)}$$

First of all, we set $\beta = \gamma$ and $\alpha = 1$ in *Eq. (31)*, and this substitution leads to:

$$H_F^{MH} = \frac{1}{n^2}\sum_{x=1}^{n}\sum_{y=1}^{n} \mu_{xy}^{\gamma} e^{-(c1(.)\mu_{xy}+d1(.))^{\gamma}} + \mu_{n,xy}^{\gamma} e^{-(c2(.)\mu_{n,xy}+d2(.))^{\gamma}}. \quad \textbf{(32)}$$

We derive two functions; the first one with different gain functions while the second one with the same gain functions. For the first one, the substitution is: c1(.)=c2(.)=1 and d1(.)=d2(.)=0 and but for the second one, $c1(.) = \mu_{n,xy}$ and $c2(.) = \mu_{xy}$ with d1(.)=d2(.)=0. The resulting intuitionistic fuzzy functions are:

$$H_{F1}^{MH} = \frac{1}{n^2}\sum_{x=1}^{n}\sum_{y=1}^{n} \left\{\mu_{xy}^{\gamma} e^{-\mu_{xy}^{\gamma}} + \mu_{n,xy}^{\gamma} e^{-\mu_{n,xy}^{\gamma}}\right\}, \quad \textbf{(33)}$$

$$H_{F2}^{MH} = \frac{1}{n^2}\sum_{x=1}^{n}\sum_{y=1}^{n}\left\{P_{xy}^{\gamma}e^{-(\mu_{xy}\mu_{n,xy}))^{\gamma}},\right\} \quad (34)$$

where $P_{xy}^{\gamma} = \mu_{xy}^{\gamma} + \mu_{n,xy}^{\gamma}$. The bracketed terms can also be called the embedded MFs. To reflect that the below are derived from the feature maps, we place an apostrophe on all the functions:

$$H_{F1}^{\prime MH} = \frac{1}{NF^2}\sum_{x=1}^{NF}\sum_{y=1}^{NF}\left\{\mu_{xy}^{\prime\gamma}e^{-\mu_{xy}^{\prime\gamma}} + \mu^{\prime\gamma}{}_{n,xy}e^{-\mu_{n,xy}^{\prime\gamma}}\right\}, \quad (35)$$

$$H_{F2}^{\prime MH} = \frac{1}{NF^2}\sum_{x=1}^{NF}\sum_{y=1}^{NF}\left\{P_{xy}^{\prime\gamma}e^{-(\mu_{xy}^{\prime}\mu_{n,xy}^{\prime}))^{\gamma}}\right\}. \quad (36)$$

In moving from a fuzzy set to an intuitionistic fuzzy set, $\mu_{xy}^{\prime}$ gains the bracketed terms ($H_{F1,xy}^{\prime MH}$ and $H_{F2,xy}^{\prime MH}$), and this is the handiwork of the Intuitionistic MH fuzzy entropy function that has imparted structures to it. Let $I_{xy}^{\prime}$ be the input image with its size reduced to that of the MF matrix. The information values could be of three kinds: $I_{xy}^{\prime}\mu_{xy}^{\prime}$, $I_{xy}^{\prime}H_{F1,xy}^{\prime MH}$ and $I_{xy}^{\prime}H_{F2,xy}^{\prime MH}$. We can choose any one of these as the DeepInfoNet feature, as they are tapped from the feature map of this architecture.

For validation, the training and test sets have to be separated, and the corresponding DeepInfoNet feature sets are assigned to Gen and Adv, which stand for the Generator and Adversary modules of GAN. So, we can designate DeepInfoNet as GAN-DeepInfoNet or simply GAN-InfoNet and name its modules as Gen-InfoNet to deal with $f_{tr}(i,j)$ and similarly Adv-InfoNet, to deal with $f_{te}(j)$. The FVs and MFs involved in the GAN-InfoNets are elaborated as:

$$f_{tr}(i,j) = I_{tr,xy}^{\prime i};\ f_{te}(j) = I_{te,xy}^{\prime}.$$

$\mu_{tr}(i,j) = \{\mu_{tr,xy}^{\prime i}, H_{Ftr1,xy}^{\prime i,MH}, H_{Ftr2,xy}^{\prime i,MH}\}$ and $\mu_{te}(j) = \left\{\mu_{te,xy}^{\prime}, H_{Fte1,xy}^{\prime\ MH}, H_{Fte2,xy}^{\prime\ MH}\right)$, where we have denoted the i^{th} IS vector and MF vector by the superscript, i and j correspond to xy. We have to make sure that the tFVs of each class must be separated out. It is now imperative to make an analogy with the classifier above in assigning the roles to the two modules. The task of Gen-Net is to determine the support vector of each class using *Eq. (36),* and that of Adv-Net is to use the support vectors furnished by the Gen-Net and apply the criterion function to find the class label of the tFV using *Eqs, (25)* to *(27)*. The value of γ is a design parameter, chosen to be less than 1 by trial-and-error till we achieve the desired performance. Here the services of the learning model are no longer needed.

If γ is to be learned, we go in for the learning model by adopting the procedure outlined above. Here, γ replaces β, all other things remaining the same. This GAN-InfoNet is designed to act as a classifier, and depending on the application, we can alter its architecture.

5.4|The Rough Information Sets

Along the lines of rough fuzzy sets, we would like to create rough information sets. The concept of a rough fuzzy set is very useful if the IS/attribute values display two different distributions. In the analysis of a texture image, we can carve out both the smooth and rough constituents by decomposing it into an approximation image (i.e., smooth) and the composite image (i.e., rough) from it. The composite image is constructed from the diagonal and off-diagonal sub-bands arising from decomposition. It is easy to infer that by applying DeepInfoNet on these images, we can get the two types of MF matrices from the respective feature maps representing smooth and rough distributions. These, along with the pixel intensities of the reduced input images, help generate smooth-DeepInfoNet and rough-DeepInfoNet features.

The MF, $\mu'^{\gamma}_{n,xy}$ and the nMF, $\mu'^{\gamma}_{n,xy}$ function values belong to the smooth and rough MF values, respectively, with γ acting as a roughness factor. The correct choice of γ is very important for the delineation of smooth and rough MF values. The resulting information values formed from both the smooth and rough values constitute rough information sets. The roughness factor delineated as the ratio of smooth certainty information divided by the rough certainty information is given by:

$$RF = \frac{\sum_{x=1}^{n1}\sum_{y=1}^{n1} I'_{s,xy}\mu_{xy}^{\prime\gamma}}{\sum_{x=1}^{n2}\sum_{y=1}^{n2} I'_{r,xy}\mu'^{\gamma}_{n,xy}}. \quad (37)$$

5.5 | The Other Possible Applications

In addition to the two case studies in the medical domain, some of the applications of the EB-OLM could focus on different types of forecasting like weather, stock market, sales, etc. It can be applied to solve industrial problems such as decision-making, identification of fake news, recommender systems, cybersecurity, sentiment analysis, etc. The miscellaneous applications include areas of Biometrics, Image processing, and pattern recognition.

We can also cash in on the learning abilities of EDOY and EROY in the design of classifiers for the classification of deep information set features extracted from the feature maps of any deep learning neural network architecture by assigning the roles of objective functions to the transforms, i.e., high-level entropy functions used for the uncertainty representation of features.

6 | Results of Implementation of Effort-Based Outlook Learning Models

6.1 | A Comparison of Effort-Based Outlook Learning Models with Other Meta-Heuristic Techniques

This comparison is undertaken on the following benchmark functions for optimization to learn their parameters:

I. Powell function f1(x).
II. Rosenbrock function f2(x).
III. Dixon and Prince function f3(x).
IV. Schwefe126 function f4(x).
V. Rastrigin function f5(x).
VI. Griewank function f6(x).
VII. Qing function f7(x).
VIII. Alpine01 function f8(x).
IX. Ackleley function f9(x).
X. Perm function f10(x).

All these functions have the same global minimum of zero. The different metaheuristic techniques chosen for comparison are: Whale Optimization Algorithm (WOA), Artificial Fish Swarm Algorithm (AFSA), Krill Herd (KH), HEFAG [4], CCLM [5], Black Widow Optimization (BWO), Red Deer Algorithm (RDA), and Sandpiper algorithm. Here, we have not considered the Genetic Algorithm (GA) for comparison, as it requires several parameters such as the number of chromosomes, fitness function, population size, crossover and mutation operators, and the number of iterations. The correct choice of these parameters is fraught with

many difficulties. In contrast, in the proposed learning model, the population size (i.e., the number of contenders) is set to 30. *Table 1* shows the results of optimization obtained with the meta-heuristic techniques on 10 benchmark functions having 50 parameters each. We have carried out 24 runs, and each run has 1000 iterations. We have also computed the mean, standard deviation, and the median over the 24 runs. As can be noticed from this table, the EB-OLM surpasses other meta-heuristic techniques after convergence is attained.

Figs. 1-3 show the implementation of different meta-heuristic techniques like EB-OLM, WOA, LOA, GBC, and RDA on Powell, Rastrigin, and Grievank functions, respectively. It is clear from these figures that the values of these functions converge to zero with EB-OLM.

To showcase EB-OLM, the two real-life applications taken up include: 1) detection of retinal diseases from retinal images, and 2) prediction of heart diseases from ECGs of patients.

F	Metrics	LOA[14]	WOA [12]	AFSA [15]	KH[11]	BAT [13]	GBC [9]	HEFAG [1]	RDA	Sandpiper	BWO	HT-CCLM	EB-OLM
f_1	best	1.75e-6	8.34e-5	2.93e-16	5.42e-12	7.31e-4	9.26e-2	6.28e-9	3.31e-14	2.57e-13	1.95e-14	1.23e-16	3.45e-20
	median	5.89e-5	8.18e-4	2.37e-15	8.26e-11	5.25e-3	4.74e-11	4.53e-8	3.2e-12	6.2e-11	9.65e-13	3.62e-15	2.51e-19
	mean	8.22e-5	4.21e-4	1.73e-15	3.84e-11	2.75e-3	9.56e-1	1.17e-8	5.6e-12	7.19e-13	2.93e-10	8.78e-15	2.69e-18
	std. dev.	3.79e-4	1.68e-4	6.87e-14	1.36e-10	8.45e-5	5.84e-2	1.63e-9	1.26e-09	9.46e-08	7.34e-07	6.71e-15	6.72e-17
f_2	best	2.61e-6	8.12e-4	1.43e-17	7.24e-8	7.35e-1	9.26e-4	6.17e-12	5.43e-16	9.12e-18	3.03e-12	3.58e-25	2.49e-26
	median	1.59e-5	8.58e-3	2.87e-16	4.76e-7	7.75e-1	5.14e-3	6.53e-10	8.45e-13	9.36e-14	1.45e-11	5.42e-24	3.71e-26
	mean	9.91e-5	3.92e-4	8.73e-16	5.54e-6	7.55e-1	6.76e-3	2.77e-9	8.32e-15	9.2e-14	6.21e-8	3.28e-24	4.39e-27
	std. dev.	2.54e7	5.24e-4	1.06e-15	5.13e-6	5.36e-2	9.91e-4	3.22e-9	4.53e-7	3.68e-13	4.42e-7	2.74e-25	1.65e-26
f_3	best	1.29e-5	7.26e-3	6.42e-17	6.96e-14	4.54e-01	5.63e-4	7.82e-14	2.14e-16	5.12e-15	6.18e-14	5.16e-27	3.73e-29
	median	6.05e-4	7.35e-2	4.48e-16	7.93e-13	7.04e-01	8.96e-3	1.75e-13	9.25e-13	8.23e-12	9.16e-11	7.2e-26	3.4e-28
	mean	8.94e-4	6.75e-2	9.24e-15	4.15e-13	7.67e-01	6.29e-3	6.63e-12	7.21e-10	5.89e-13	4.32e-13	9.25e-26	2.38e-28
	std. dev.	7.02e-5	1.38e-3	1.01e-10	6.96e-16	5.68e-01	2.39e-4	6.95e-11	4.02e-12	6.94e-13	3.72e-11	8.1e-27	4.6e-29
f_4	best	2.15e-01	1.24e+01	5.63e-7	4.52e-3	4.31e-01	2.89e-1	7.78e-4	9.68e-8	2.41e-7	9.73e-12	3.27e-16	5.65e-18
	median	4.29e-01	3.87e+01	4.28e-7	8.58e-2	2.57-01	4.65e-1	9.26e-4	9.61e-7	5.36e-6	8.20e-10	2.34e-15	4.7e3-17
	mean	5.51e-01	5.82e+01	5.53e-7	8.54e-2	6.75e-01	7.46e-1	9.47e-4	2.27e-6	4.79e-5	9.24e-10	1.68e-15	8.99e-17
	std. dev.	9.93e-02	6.89e+01	7.73e-4	3.61e-3	3.56e-02	8.43e-2	7.32e-5	2.96e-7	4.27e-7	1.85e-8	7.25e-16	5.14e-17
f_5	best	2.11e-04	4.72e-05	3.93e-20	6.24e-8	1.45e-02	8.16e-3	6.37e-16	4.53e-13	1.15e-13	2.38e-12	2.88e-28	3.95e-29
	median	8.79e-03	8.88e-04	8.67e-19	6.46e-7	7.95e-2	3.74e-2	9.53e-15	1.23e-12	1.48e-12	5.11e-11	3.41e-27	2.72e-29
	mean	6.81e-03	7.72e-04	5.43e-19	5.74e-7	3.95e-2	5.16e-2	6.27e-15	4.13e-12	5.75e-12	5.52e-11	4.58e-27	5.69e-28
	std. dev.	4.99e-04	9.28e-04	1.27e-19	8.36e-8	4.98e-3	1.15e-2	1.54e-15	7.06e-10	7.18e-10	4.74e11	3.63e-29	1.72e-30
f_6	best	2.21e-05	5.82e-06	2.83e-24	8.34e-10	9.15e-03	1.26e-04	9.27e-9	5.41e-22	5.01e-21	6.81e-22	7.58e-30	3.49e-32
	median	8.59e-04	9.38e-05	7.57e-24	8.66e-09	4.23e-02	2.35e-03	7.64e-8	7.38e-20	7.79e-20	3.48e-20	3.51e-29	1.62e-31
	mean	3.4e-04	9.5e-05	2.6e-24	2.7e-09	6.3e-02	1.1e-03	3.6e-8	7.11e-20	6.89e-20	2.46e-19	1.4e-29	2.7e-31
	std. dev.	2.09e-04	6.85e-05	2.37e-27	9.46e-09	4.55e-04	5.46e-03	8.76e-8	2.07e-19	5.48e-18	9.63e-18	8.37e-29	8.59e-31
f_7	best	6.91e-06	8.81e-9	4.76e-26	6.17e-7	7.54e-3	2.15e-4	2.83e-6	4.06e-30	6.26e-30	3.01e-29	6.15e-34	2.32e-36
	median	3.49e-5	4.18e-8	6.77e-25	5.62e-6	7.54e-2	5.46e-3	5.37e-5	8.47e-18	6.26e-17	5.94e-17	4.13e-33	1.72e-34
	mean	4.96e-5	1.82e-8	2.73e-25	6.65e-6	4.58e-2	2.43e-3	6.27e-5	9.25e-16	3.57e-16	7.81e-16	7.91e-33	7.25e-35
	std. dev.	6.51e-7	3.42e-8	7.53e-25	5.94e-6	1.85e-2	1.69e-5	1.79e-5	7.78e-17	4.42e-17	6.11e-17	2.68e-33	3.99e-35
f_8	best	2.93e-03	8.83e-04	1.79e-14	6.63e-16	3.58e-01	6.46e-03	3.35e-13	1.68e-20	1.48e-19	5.77e-18	5.22e-26	9.12e-28
	median	6.13e-02	3.23e-3	9.36e-13	5.49e-15	4.52-01	6.65e-02	4.76e-12	8.01e-17	9.4e-17	2.16e-16	2.85e-25	6.97e-27
	mean	6.95e-02	4.86e-2	3.74e-12	7.69e-15	3.54e-01	9.43e-02	1.35e-12	1.75e-16	3.06e-16	9.22e-17	1.27e-25	4.15e-27
	std. dev.	3.31e-02	5.74e-03	7.38e-13	5.74e-15	7.57e-01	7.62e-03	5.71e-12	5.22e-15	1.13e-16	4.03e-17	7.83e-26	3.49e-27
f_9	best	4.89e-04	1.58e-10	2.77e-20	4.96e-11	8.45e-4	6.25e-3	2.73e-12	5.22e-22	1.13e-21	4.03e-23	1.42e-30	2.33e-32
	median	8.41e-03	8.82e-09	5.93e-19	9.74e-10	8.75e-3	3.46e-2	1.17e-11	6.41e-20	3.36e-20	2.37e-18	3.48e-29	4.79e-30
	mean	7.97e-03	6.58e-09	8.27e-19	9.56e-10	9.85e-3	9.94e-2	8.53e-11	7.26e-19	7.73e-19	5.05e19	1.62e-29	8.12e-31
	std. dev.	9.14e-03	5.24e-09	1.39e-21	3.44e-12	4.52e-4	9.69e-3	3.72e-13	1.7e-18	4.98e-19	1.95e-18	3.48e-28	2.99e-29
f_{10}	best	6.29e-02	6.48e-06	7.57e-20	1.76e-7	7.25e-02	1.45e-01	2.74e-10	9.65e-22	1.04e-21	7.97e-20	4.83e-29	5.42e-31
	median	1.71e-1	9.22e-05	5.23e-19	4.45e-6	4.74e-01	8.96e-01	1.87e-10	9.5e-20	5.41e-20	5.4e-19	5.98e-29	9.29e-30
	mean	9.69e-1	3.78e-5	4.37e-19	5.46e-6	8.85e-01	6.14e-01	6.23e-10	4.75e-19	1.44e-19	9.12e-18	1.52e-29	3.41e-30
	std. dev.	1.13e-2	8.22e-5	7.39e-19	6.48e-21	4.95e-6	7.86e-03	9.17e-01	2.13e-17	2.65e-17	3.15e-17	4.38e-26	7.39e-30

Fig. 1. A comparison of different meta-heuristic techniques.

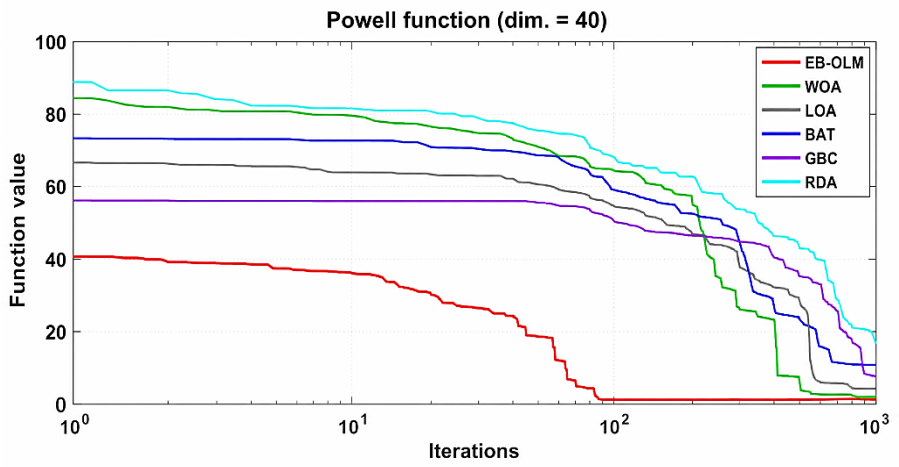

Fig. 2. Comparison of different meta-heuristic techniques on Powell function (dim.=40).

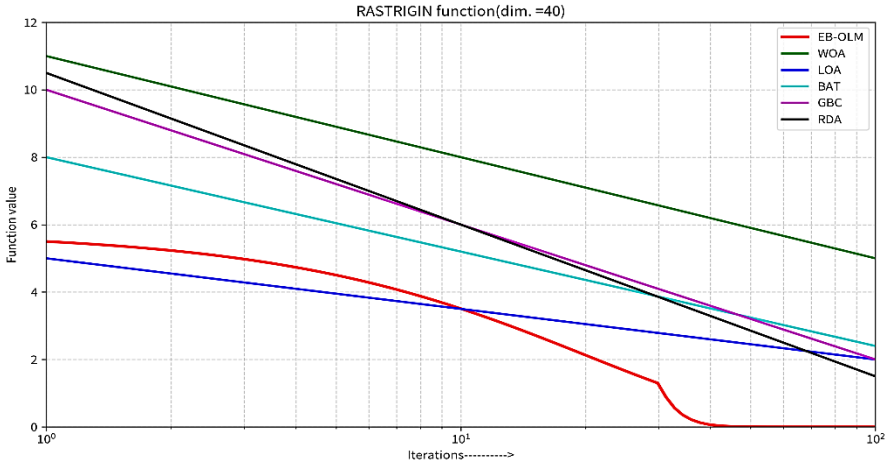

Fig. 3. Comparison of different meta-heuristic techniques on Rastrigin function (dim.=40).

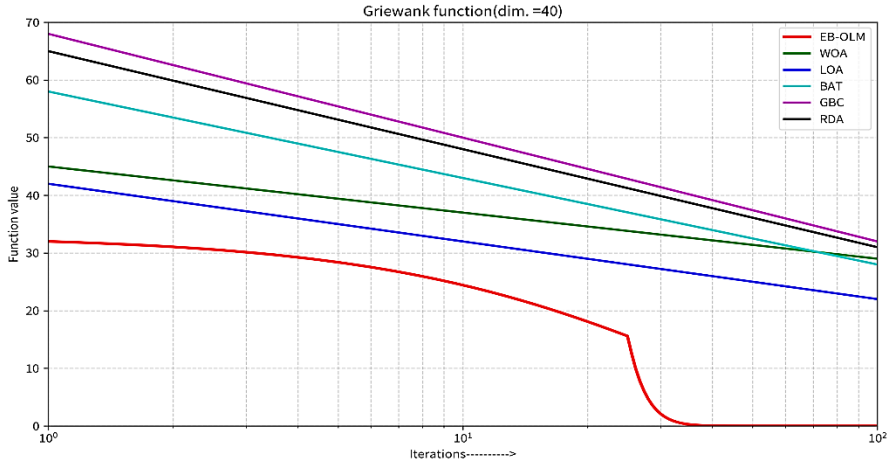

Fig. 4. Comparison of different meta-heuristic techniques on Griewank function (dim.=40).

Case Study 1 (Detection of retinal diseases). The retinal images are of high resolution, captured under a variety of imaging conditions. A clinician rates the presence of Diabetic Retinopathy (DR) in an image on a scale of 0 to 4, with the following categories/classes: 0 - No DR, 1 – Mild DR, 2– Moderate DR, 3 –Severe DR, and 4 - Proliferative DR. A total of 88, 702 images exists and out of these 54,000 are set aside for the training and the rest for the testing. A typical retinal image for each category of DR is shown in *Fig. 5*.

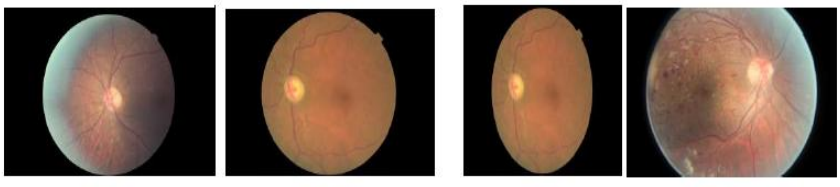

Fig. 5. DR images; a. Mild DR, b. Moderate DR, c. Severe DR, and d. Proliferative DR.

In this study, the probabilistic feature called the hybrid Hanman-Rényi-Tsallis transform derived in [16] is found to be most suitable, as under:

$$H_k^{HRT} = \sum_{x=1}^{n} P_x\{e^{\cdot\alpha.\log(P_x)/(1-\alpha)} + e^{(P_x - P_x{}^{\alpha})/(\alpha-1)}\}, \quad \textbf{(38)}$$

where P_x, the probability of the pixel intensity is computed from the histogram of a window in a retinal image, and α is set at 1.6. The size of the window is selected as 15x15 after experimentation. While learning the parameter β of the classifier using EB-OLM, the population size is fixed at 40 and the number of runs at 24. Within each run, 1000 iterations are executed. The widely used classifiers like Random Forest (RF), CNN, Naïve Bayes classifier, and K-Nearest Neighbour are compared in *Table 2* for the detection of DR.

The proposed classifier is shown to outperform the well-known classifiers compared. As evident from *Table 2*, the proposed methodology gives an accuracy of 97.5% and outperforms other classifiers.

Table 1. Comparison of the performance of different classifiers.

Classifier	Accuracy
RF	89.8%
Naïve Bayes	88.7%
SVM	94%
CNN –Transfer learning	96%
Proposed methodology	97.5%

Case Study 2 (Prediction of heart diseases). The status of a heart, whether it is healthy or defective, is dealt with in [19]. The 13 attributes in the heart dataset include: age, sex, cp: chest pain type; trestbps: resting blood pressure; chol: serum cholesterol; fbs: fasting blood pressure; restecg: resting electrocardiographic results; thalach: maximum heart rate; exang: exercise-induced angina; oldpeak: ST depression induced by an exercise relative to the resting; slope: the slope of the peak exercise ST segment; ca: number of major vessels colored by fluoroscopy; thal: thalassemia.

In this dataset, the variables such as sex, fbs, rest ecg, and exang are binary, whereas slope takes values (0, 1,2), and the variables cp, ca, and thal take integer values (0, 1, 2, 3). So we find the probability of each integer variable separately and then compute the probabilistic uncertainty using the Hanman-Tsallis transform [17], given by:

$$H_x^{HTs} = P_x e^{(P_x - P_x{}^{\alpha})/(\alpha-1)}. \quad \textbf{(39)}$$

This Equation is derived from the Tsallis entropy function [20], defined as:

$$H^T = \frac{\sum_{i=1}^{n} p_{xi} - \sum_{i=1}^{n} p_{xi}{}^{\alpha}}{\alpha - 1}. \quad \textbf{(40)}$$

One term on the r.h.s. of *Eq. (40)* can be written as:

$$H_x^T = \frac{p_{xi} - p_{xi}{}^{\alpha}}{\alpha - 1}. \quad \textbf{(41)}$$

Using H_x^T as the exponential gain function of HT leads to the Hanman-Tsallis transform in *Eq. (39)*.

The variables age, trestbps, chol, thalach, and oldpeak are considered as features denoted by Q_x where x stands for each of these variables numbering q=13; hence we fit the Gaussian MF, μ_x by computing the mean and variance of the distribution underlying each feature. Here, we make use of the possibilistic Rényi entropy function, $\overline{\mu_x}e^{(1-Q_x)}$ derived by Bhatia and Hanmandlu [21] from the probabilistic Rényi entropy function [22]. As the sigmoid function representing the uncertainty in the distribution/variation of each feature is found to be more effective than the possibilistic entropy function, it is given by:

$$H_x^{RS} = \frac{1}{1 + e^{-\overline{\mu_x}e^{(1-Q_x)}}}, \tag{42}$$

where $\overline{\mu_x}$ is its complement MF. The weights (i.e., efforts) to be learned are those of 13 features.

Implementation aspects: before implementing the prediction using *Eq. (41)*, we need to initialize all the weights with random values. Taking each sample at a time, we have to plug in the information values of features. If a feature is an integer, we have to get its probabilistic uncertainty, but if it is real, then its possibility certainty. Using the initial weights and the information values, the output is computed. By invoking EB-OLM, we get the updated values of the weights. In the next iteration, the computed output is considered as the 14th feature, and then it is weighted to serve as an additional input to improve prediction, thus making the number of weights equal to 14. The output weight is assumed to be a random number. This procedure is repeated until all the training samples are utilized. *Table 3* gives a comparison of our methodology with literature detectors on the prediction of heart diseases and outperforms other classifiers.

Table 2. Comparison of the performance of different classifiers.

Classifier	Accuracy
RF	89%
Naïve Bayes	88%
SVM	92%
CNN –Transfer learning	91%
Proposed methodology	97.5%

We now take a respite to deliberate on the contributions made in this paper. These include: 1) formulation of three efforts-based learning models, 2) development of design models for classifier, predictor, and GAN architecture, and 3) Presentation of two case studies for the vindication of the certainty/uncertainty representation.

7 | Conclusions

This paper introduces the concept of Efforts-Based Outlook (EBO) by proposing the three types of efforts, viz., ENO, ERO, and ESO, to be associated with the three kinds of outlooks. We have tried to devise a learning model for each kind of outlook. A contender for success attempting to achieve a goal can be in any one of three outlooks, viz., ENOY, EROY, and ESOY, depending on the type of efforts employed by the contender. ENOY does not expect anything from the outcome of his/her efforts but uses it to improve his/her performance. EROY competes with the best performer, aiming at better fruits, as attaining excellence is the motto. ESOY tries to steal the outcome of others' efforts. Thus, this work has cashed in on one form of human behaviour, but there are several manifestations of behavioural forms that come into play when other endeavours are considered.

The entire modelling is done at the backdrop of the information set theory, in which we have exploited both the basic and higher-level certainty representation through the information sets. As ENOY solely depends on efforts to achieve a goal, we have used an advanced pervasive information set in modelling this component of outlook learning. EROY has a competitive spirit to excel others, and the HT divergence is poised to be adept for modelling this component of learning. An ESOY eyes the outcome of others' efforts, and to model this component of learning, we have used the basic information but with an appropriate modification. The EB-OLM is compared with WOA, LOA, GBC and RDA for the optimization of three standard functions

such as Powell, Rastrigin and Grievank functions. The fastest convergence to the correct parameter values is achieved with the EB-OLM.

We have attempted the design of both classifier and predictor under the aegis of the certainty/uncertainty representation and learned their parameters with the help of EB-OLM. Some applications of EB-OLM, classifier, and predictor are suggested for future work. We have also given the formulation for the design of a Generative-Adversarial network using the handcrafted deep learning network architecture called DeepInfoNet and suggested how to create rough information sets. These applications emanate from the utilitarian aspects of the EBO learning. These developments are the only beginning of the journey, and their outreach can pave the way for solutions to many problems.

Two case studies comprising DR and heart disease prediction are showcased to demonstrate the power of the designs of the classifier and the predictor, respectively, with the aid of the proposed EB-OLM, and their performance is shown to exceed that of the classical classifiers. The versatility of three modes of learning can be utilized in framing different architectures for deep learning neural networks, thereby opening a gateway to the benefits of the information set theory that underlies the learning models.

Lastly, it may be noted that the learning components of outlook differ as per the types of efforts displayed while accomplishing an activity related to achieving a goal. Hence. the three kinds of effort-based outlook learning models are shown to outperform other approaches in two case studies.

As regards the shortcomings of the learning models, selection of the population size and fixation of the number of iterations are critical for the convergence of the parameters to be learned.

The learning models, case studies, and the design of classifiers and predictors show the versatility of information set theory. However, the formulations of GAN and rough information sets are not implemented, and we are in search of suitable applications to validate them. The future work is to embark on the development of information set-based Large Language Models (LLMs) in the context of querying and reasoning.

Authors' Contributions

All aspects of the research and manuscript preparation were carried out by the author. The author has read and approved the final version of the manuscript.

Funding

Not applicable.

Data Availability

All data supporting the reported findings in this research paper are provided within the manuscript.

Conflict of Interest

The author declares that they do not have any conflict of interest.

Consent for Publication

The author confirms consent for the publication of this work

Ethics Approval and Consent to Participate

This article does not contain any studies with human participants performed by the author.

References

[1] Rao, R. V., Savsani, V. J., & Vakharia, D. P. (2011). Teaching-learning-based optimization: A novel method for constrained mechanical design optimization problems. *Computer-aided design*, *43*(3), 303–315. https://doi.org/10.1016/j.cad.2010.12.015

[2] Rao, R., & others. (2016). Jaya: A simple and new optimization algorithm for solving constrained and unconstrained optimization problems. *International journal of industrial engineering computations*, *7*(1), 19–34. https://doi.org/10.5267/j.ijiec.2015.8.004

[3] Rao, R. V. (2020). Rao algorithms: Three metaphor-less simple algorithms for solving optimization problems. *International journal of industrial engineering computations*, *11*(1), 107-130. https://doi.org/10.5267/j.ijiec.2019.6.002

[4] Grover, J., & Hanmandlu, M. (2018). New evolutionary optimization method based on information sets. *Applied intelligence*, *48*(10), 3394–3410. https://doi.org/10.1007/s10489-018-1154-x

[5] Grover, J., & Hanmandlu, M. (2021). Novel competitive-cooperative learning models (CCLMs) based on higher order information sets. *Applied intelligence*, *51*(3), 1513–1530. https://doi.org/10.1007/s10489-020-01881-3

[6] Grover, J., & Hanmandlu, M. (2021). Development of an optimal entropy classifier and prudent learning model. *IEEE transactions on artificial intelligence*, *3*(2), 164–175. https://doi.org/10.1109/TAI.2021.3117491

[7] Hanmandlu, M., & Das, A. (2011). Content-based image retrieval by information theoretic measure. *Defence science journal*, *61*(5), 415–430. https://doi.org/10.14429/dsj.61.1177

[8] Pal, N. R., & Pal, S. K. (1991). Entropy: A new definition and its applications. *IEEE transactions on systems, man, and cybernetics*, *21*(5), 1260–1270. https://doi.org/10.1109/21.120079

[9] Hanmandlu, M., & others. (2014). A new entropy function and a classifier for thermal face recognition. *Engineering applications of artificial intelligence*, *36*, 269–286. https://doi.org/10.1016/j.engappai.2014.06.028

[10] Aggarwal, M., & Hanmandlu, M. (2015). Representing uncertainty with information sets. *IEEE transactions on fuzzy systems*, *24*(1), 1–15. https://doi.org/10.1109/TFUZZ.2015.2417593

[11] Hanmandlu, M., Bansal, M., & Vasikarla, S. (2020). An introduction to information sets with an application to iris based authentication. *Journal of modern physics*, *11*(1), 122–144. https://doi.org/10.4236/jmp.2020.111008

[12] Sayeed, F., & Hanmandlu, M. (2017). Properties of information sets and information processing with an application to face recognition. *Knowledge and information systems*, *52*(2), 485–507. https://doi.org/10.1007/s10115-016-1017-x

[13] Hanmandlu, M., & Singhal, S. (2017). Face recognition under pose and illumination variations using the combination of information set and PLPP features. *Applied soft computing*, *53*, 396–406. https://doi.org/10.1016/j.asoc.2017.01.014

[14] Atanassov, K. T. (1999). Intuitionistic fuzzy sets. In *Intuitionistic fuzzy sets: Theory and applications* (pp. 1–137). Springer. https://doi.org/10.1007/978-3-7908-1870-3_1

[15] Ansari, F., & Madasu, H. (2024). Color thresholding detection and recognition of the road signs using the information set theory. *Journal of modern physics*, *15*(11), 1646–1678. https://doi.org/10.4236/jmp.2024.1511072

[16] Asthana, P., Hanmandlu, M., & Vashisth, S. (2022). Classification of brain tumor from magnetic resonance images using probabilistic features and possibilistic Hanman-Shannon transform classifier. *International journal of imaging systems and technology*, *32*(1), 280–294. https://doi.org/10.1002/ima.22619

[17] Asthana, P., Hanmandlu, M., & Vashisth, S. (2022). Brain tumor detection and patient survival prediction using U-Net and regression model. *International journal of imaging systems and technology*, *32*(5), 1801–1814. https://doi.org/10.1002/ima.22735

[18] Dabass, J., Hanmandlu, M., Vig, R., & Vasikarla, S. (2024). Mammogram classification with HanmanNets using hanman transform classifier. *Journal of modern physics*, *15*(7), 1045–1067. https://ui.adsabs.harvard.edu/link_gateway/2024JMPh...15.1045D/doi:10.4236/jmp.2024.157045

[19] Chellammal, S., & Sharmila, R. (2019). Recommendation of attributes for heart disease prediction using correlation measure. *International journal of recent technology and engineering (IJRTE)*, *8*(23), 870–875. https://doi.org/10.35940/ijrte.B1163.0782S319

[20] Tsallis, C., Mendes, R., & Plastino, A. R. (1998). The role of constraints within generalized nonextensive statistics. *Physica a: Statistical mechanics and its applications*, *261*(3–4), 534–554. https://doi.org/10.1016/S0378-4371(98)00437-3

[21] Bhatia, A., & Hanmandlu, M. (2018). Keystroke dynamics based authentication using possibilistic Rényi entropy features and composite fuzzy classifier. *Journal of modern physics*, *9*(02), 112. https://doi.org/10.4236/jmp.2018.92008

[22] Rényi, A. (1961). On measures of entropy and information. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability, volume 1: Contributions to the theory of statistics* (Vol. 4, pp. 547-562). University of California Press. https://projecteuclid.org/ebook/Download?urlid=bsmsp/1200512181&isFullBook=false