A Framework for Feature Selection using Data Value Metric and Genetic Algorithm

Ojie Deborah Voke
Department of Software
Engineering
Faculty of Computing
University of Delta, Agbor,
Delta State

Akazue Maureen, PhD
Department of Computer
Science
Faculty of Science
Delta State University, Abraka

Imianvan Anthony
Department of Computer
Science
Faculty of Physical Science
University of Benin, Benin City

ABSTRACT

Most organizations analyse input data to develop an accurate description or model using the features present in the data. There have been huge amount of generated data in the big data ecosystem which demand better and efficient ways to distil high utility or value from it so as to compliment decision makers in making recommendations and decisions. Before applying classification algorithm, relevant features are selected by a suitable feature selection algorithm. Data Value Metric (DVM) is an information theoretic measure based on the notion of mutual information which has been shown to be a good metric to validate the quality and utility of data in a big data ecosystem and in traditional data. Data Value Metric (DVM) suffers from local minima and loss of diversity in the population since it is using forward selection search strategy, however, hybridizing it with Genetic Algorithm is hoped to overcome the problem of local minima as there would be a blend of evolutionary search to ensure a balance between exploration and exploitation of the search space. This paper proposed the hybrid model of Genetic Algorithm and Data Value Metric (DVM) as an information theoretic metric for quantifying the quality and utility for feature selection which can be applied to traditional data.

Keywords
Genetic Algorithm, Data Value Metric, feature selection, search, local minima.

1. INTRODUCTION
The accessibility and penetration of computers and associated computing devices have progressively made data accumulation available. This huge amount of accumulated data, which are passive, are not processed in a timely manner to make appropriate decisions and recommendations, hence data mining has become an essential aspect of decision making process to ensure high utility or value driven process in the course of decision and recommendations. An important step in the data mining is data pre-processing, as the feature of decisions is based on the value of data. Enhancing the medical database improve the feature of medical diagnosis. Data pre-processing steps are data cleaning, data integration, data transformation and data reduction (feature subset selection). A few attributes of data sets possibly redundant as their information may be contained in other attributes. More attributes can affect the computation time for the diagnosis accuracy. Some data in the dataset may not be useful for diagnosis and thus can be eliminated before learning. The goal of feature choosing is to find a least set of attributes so that the resulting probability distribution of the data classes is as close as likely to the original distribution obtained by all attributes (Adam et al, 2012)[14]. Irrelevant, redundant, or noisy data can be removing through data reduction process. This reduction gives speed on data mining algorithm, and improving mining performance such as predictive accuracy and result Comprehensibility. Feature set reduction is based on feature relevance and redundancy with respect to the goal. A feature is usually categorized as: 1) strongly relevant, 2) weakly relevant, but not redundant, 3) Irrelevant, and 4) redundant. A strongly relevant Feature is always necessary for an optimal feature subset; it cannot be removed without affecting the original conditional target distribution (Yu and Liu, 2004).[15][4] Weakly relevant feature may not always be necessary for an optimal subset; this may depend on certain conditions. Irrelevant features are not necessary to include at all. Redundant features are those that are weakly relevant but can be completely replaced with a set of other features such that the target distribution is not disturbed (the set of other features is called Markov blanket of a feature). Redundancy is thus always inspected in multivariate case (when examining feature subset), whereas relevance is established for individual features. The aim of feature selection is to maximize relevance and minimize redundancy. It usually includes finding a feature subset consisting of only relevant features.

Good, relevant data and artificial intelligence (Machine Learning) are two wings on a bird with complimentary abilities and work synergistically to enhance decision support and other allied

benefits depending on its areas of application. Feature selection is a major phase in the machine learning whose action produce good relevant data/feature set that are fed into artificial intelligence learning systems and are outputted as models that works on data to arrive at a better utility for the

data. How much data/feature set would be sufficient for an artificial intelligence or machine learning algorithm to derive a model? If there is such phenomenon, there should be trade-off point of a number of data suitable for a particular algorithm giving the deluge of data in the big data community. Due to the deluge of data and information as well as it multisource capability in big data ecosystem, there is an inherent or potential drawback of uncertainties, noises and inherent redundancies. These uncertainties, noises, redundancies and its negative effect have necessitated the need to select a representative set of features high packed with relevant data with minimal redundancy and noises to ensure high classification accuracy or in general high performance. More complex dependencies involving several features might exist but not correctly reflected. In some other cases, we can expect that summing across all the features, some spurious dependencies might amplify the dependencies score thus

producing a less accurate output. There is therefore a need to Iteratively select the best features representation that will produce the set of features without the problem of local minima which inherently present in data value metric as proposed by **Noshad et al. (2021)[12]** as well as to present an elaborated method to score the dependencies between features. Genetic algorithm being a randomized search strategy avoid local minima hence the need to hybridize genetic algorithm and data value metric algorithm so as to investigate its performance.

This paper work proposes the use of genetic algorithm for the search strategy while a data value metric value would be used as the information metric for the fitness score of the genetic algorithm which will ensure that there is no local minima in the search space as indicated for sequential search techniques.

1.1 Statement of the Problem
The data value metric introduced by **Noshad et al. (2021)**[12] is a promising metric for feature selection; however, selecting the best features using data value metric will suffer from local minima since it is using a sequential forward selection search strategy. The feature space is an essential component in ensuring quality out of the data and doing that in a linear manner would be time consuming and also adds more complexity. Fitness function is the driving force, which plays an important role in selecting the fittest individual in every iteration of an algorithm.(katoch et al. , 2021). It is in the light of this, genetic algorithms which have been shown to avoid local minima is proposed as a search strategy in data value metric algorithm instead of forward selection search strategy.

The resulting model would be tested using big data dataset and normal dataset to assay its performance in terms of feature selection and overall model performance. Assume there are so many local minima, then, the question is:

i. Whether we can expect to get global minima from the data using only data Value metric that will be computationally economical giving the fact that it is using sequential forward selection search strategy?

ii. What will be the effect of the local minima on the classification performance if Data Value Metric Algorithm is used for feature selection?

iii. Would there be an improvement in the classification performance if genetic algorithm is hybridized into the Data Value Metric?

In general, such questions are difficult to answer solely by considering a particular measure of performance on a given dataset.

2. REVIEW OF LITERATURE
In real-world applications, data is collected to the granular level. This has been carried out over many years in the belief that more data means more useful information for processing. With the increase in number of devices worldwide, there has been a surge in the availability of data in a way that storing, handling, and processing data has become difficult. Additionally, the data collected is most often not pre-processed, and hence contains redundant and irrelevant data. More attributes can affect the computational time for the diagnosis accuracy the goal of feature selection is to find the least set of attributes that the resulting probability distribution of the data classes is as close as likely to the original distribution obtained by all attributes. Dimensionality reduction techniques have been adopted to reduce the vast dimensions of data to smaller dimensions (Leskovec et al[1], 2014). The most popular dimensionality reduction techniques combine features to reduce the dimension. Feature selection is one such dimensionality reduction which selects features from the feature set without making modifications to them.

2.1 Categories of feature selection:
i. Filter method: - This method involves ranking features using suitable criteria such that the highly-ranked features are picked for application (Chandrashekar and Sahin, 2014).[2][10] The idea is to filter out lower ranked features. The most important factor in this method is determining the rank or relevance of a feature. A couple of ranking methods are (Chandrasekhar and Sahin, 2014)[2][10].

(a) Correlation criteria: - This is used to detect linear dependencies between features. It is measure using Pearson correlation coefficient.

(b) Mutual information: - This measure is used to measure the dependency between features. A value of 0 implies that 2 features are independent. The advantages of this method are that it is simple to compute and that it doesn't rely on learning algorithms. The drawback is that the features selected may not be guaranteed to be non-redundant [2][10]

ii. Wrapper method: - This method depends on use of classification to determine a feature subset. Exhaustive search methods may be able to arrive at the most optimal result but they can be computationally intensive for large datasets. Therefore, 2 types of wrapper methods may be used [2][10].
(a) Sequential Search Algorithms: - These algorithms add or remove features until a target optimization function is obtained. Sequential Forward Search algorithm starts with an empty set and adds features as and when they qualify. Sequential Backward Search algorithms start with the entire feature set and progressively eliminate ones that do not meet the performance criteria. it can lead to local minima.
(b) Heuristic Search Algorithms: - Genetic algorithms can be used to select features, wherein a chromosome represents the inclusion/exclusion of the set of features. Although this proves to be a convenient method for the selecting features, the main drawback is that the entire model must be built and evaluated for each feature subset considered.

(c) Embedded methods: - This method tries to compensate for the drawbacks of filter and wrapper methods. It involves algorithms that have in-built feature selection methods. This combines the step of selecting features and determining performance into one step (Chandrashekar and Sahin, 2014).[2][10]

2.2 Genetic algorithm
Evolutionary computation was developed with the idea that it could be used as a tool for optimization and solutions to problems could be evolved using operators of natural selection. Early methods involved representing tasks as finite-state machines and performing mutation by randomly changing the state diagrams. John Holland invented Genetic Algorithms (GA) which was a population-based algorithm. The goal was to study the process of evolution and design a framework that would apply to different applications (Mitchel, 1998).A GA is a heuristic search algorithm based on the concepts of natural selection and genetics. The idea is to mimic biological processes such as survival of the fittest, to evolve a solution for a problem. GA is a method of evolving a population of chromosomes to new populations using

selection along with operations such as crossover and mutation [11]. Each chromosome consists of genes. Selection operators choose individuals from the population that are the fittest, while crossover and mutation mimic biological processes responsible for introducing diversity to the population. While selection is an exploitation process, crossover and mutation are exploration processes. Evolutionary algorithms are most suitable for problems that involve a large search space i.e. many possible solutions. Other problems require that new solutions are produced at each stage, to explore new options or they involve complex solutions, that can be processed by hand [11]. GAs, like the process of evolution, depends on the fittest organisms/solutions to survive. The fitness of an organism/solution is determined based on the problem at hand, and it is a factor which continuously evolves. Given this, the parameters of a GA are:

(i) A population of individuals/chromosomes – each chromosome is a possible solution to the problem at hand. The population is modified or replaced over n iterations of the algorithm.

(ii) Fitness function – each chromosome is assigned a fitness value/score which indicates how close the solution represented by the chromosome is, as compared to the expected result.

(iii) Selection criteria – the fitter the chromosome, the higher the chance it has of being selected.

(iv) Crossover operator – to create a new chromosome, sub-sequences of 2 chromosomes are exchanged at a randomly chosen locus point.

(v) Mutation operator – to create a new chromosome, random bits in the chromosome are flipped.

2.3 Data Value Metric and related works

There are several studies that have proposed metrics for assessing or quantifying the information gain of a given dataset. One of such metrics is the value of information(VOI) analysis which is a decision-theoretical statistical frame work representing the expected increased inference accuracy or reduction in loss based on prospective information(Jackson et al 2019).The basic three types of VoI methods include :

(1) Inferential and modeling cases for linear objective functions under simplified parameter distribution restrictions, which limits their broad practical applicability (Madan et al,2001)

(2) Methods for estimating the expected value of partial perfect information (EVPPI) involving partitioning of the parameter space into smaller subsets and assuming constant and optimal inference over the local neighbourhoods, within subsets (strong et al , 2013) ; and

(3) Gaussian process regression methods approximating the expected inference (Strong et al, 2013).More specifically, for a particular parameter φ , the EVPPI is the expected inferential gain, or reduction in loss, when φ is perfectly estimated. As the perfect φ is unknown in advance, this reduction of loss expectation is taken over the entire parameter space $\varphi \in \_$:

i. $\text{EVPPI}(\varphi) = E\theta \ (L(d*, \ \theta)) \ - \ E\varphi(E\theta|\varphi(L(d*\varphi, \theta)))$…………………………………..3.1

Where *d*is decision, inference or action, d∗φ is the optimal inference obtained when φ is known, θ is the model parameter vector, *E* is the expectation, and L(d, θ) is the likelihood function . Note that VoI techniques are mainly suitable for specific types of problems, such as evidence synthesis in the

context of decision theory. Further, their Computational complexity tends to be high and require nested Monte Carlo procedures. Another relevant study utilizes a unique decomposition of the differences (errors) between theoretical (population) parameters and their sample-driven estimates (statistics) into three independent components. If θ and ˆ θ represent a theoretical characteristic of interest (e.g., population mean) and its sample-based parameter estimate (e.g., sample arithmetic average), respectively, then, the error can be canonically decomposed as:

θ -θˆ(error)= A(Data Quality) +B(Data Quantity)+ C(Inference Problem Complexity)…   3.2

Bayes error rate is another metric that quantifies the intrinsic classification limits. In classification problems, the Bayes error rate represents the minimal classification error achieved by any classifier (Wang et al, 2005). The Bayes error rate only depends on the distributions of the classes and characterizes the minimum achievable error of any classifier.

Data Value Metric for feature selection by Noshad et al (2021)[12]

The pseudo code below (Algorithm 1) outlines the computational implementation strategy we employ in the DVM package for evaluating the DVM. The metric captures the relative analytical value of the dataset relative to the computational complexity of the supervised prediction, classification, or regression problem. In practice, the regularization term, *R(g)*, is estimated according to the known algorithmic complexity

Input : Data sets **X** , **Y** , model *g*, parameters β , λ
For *a random split* (**X′**i , $\_$ Ẍ i) *of* **X** do
Train *g* based on (**X′**i , **Y′**i)
$\_$**T**i ← g($\_$Ẍ i)

Fi ← I($\_$**T** i;**Y**i)−βI(Ẍ i;$\_$**T**i |$\_$**Y**i)
I(**X**;**Y**)
$\_$**D**← 1
M_M
i=1 Fi − $\_$R(g)
Output : $\_$D

Algorithm 1: DVM calculation for supervised problems.
Since DVM can be used to measure the quality of a feature set *T*, it can also serve as a feature selection method. In this section, we demonstrate a heuristic algorithm for sequential feature selection based on DVM values. For a classification problem, the feature selection is defined as follows. Based on an initial feature set, choose a smaller set of features that yields a minimum prediction error. Let X = {X1, ...,Xd} denote the *d* initial features. The objective is to select a smaller set of *r* features with maximum DVM score. One specific approach is based on a forward selection involving *r* iterative steps. At each step, we select a feature from the initial feature set, {X1, ...,Xd} , which increases DVM score the most. For a given (initial or intermediate) feature set F , DVM{F} represents the DVM score corresponding to that specific feature set F . The pseudocode implementing this strategy for DVM-based feature selection is given in Algorithm 2.

Unsupervised inference
We can extend the definition of DVM for supervised problems to unsupervised clustering models. In the unsupervised problems, we don't have explicit outcomes to evaluate the model performance.

Input: Input dataset, $\mathbf{X} = \{X_1, ..., X_N\}$
Labels, $\mathbf{Y} = \{Y_1, ..., Y_N\}$
Desired number of output features, *r*
F := φ , R := {1, ..., r}
for*each* i ∈R do
f←j∈R−F (DVM{F} − DVM{F ∪Xj})
Add *f* into F
Output: F

Algorithm 2.

Input: Data sets $\mathbf{X}$,$\mathbf{Y}$, model *g*, parameters β and _
for*a random split* ($\mathbf{X}$′i ,$\mathbf{X}$″, _$\mathbf{X}$i) *of* $\mathbf{X}$ do
Apply unsupervised model *g* based on _$\mathbf{X}$iˆ$\mathbf{Y}$
i← g($\mathbf{X}$′i)
_$\mathbf{T}$i ← g($\mathbf{X}$′i)
Fi ← I(_$\mathbf{T}$ i;ˆ$\mathbf{Y}$ i)−βI(_$\mathbf{X}$i;_$\mathbf{T}$i |ˆ$\mathbf{Y}$ i)
I($\mathbf{X}$;ˆ$\mathbf{Y}$)_D← 1
T_T
i=1 Fi − _R(g)
Output: _D
Algorithm 3: DVM calculation for unsupervised problems.

3. METHODOLOGY
Selecting a proper fitness function is an essential aspect of the effectiveness of using genetic algorithm for feature selection. Many existing system have employed different fitness functions in selecting relevant features yet there is still room for improvement so as to handle the problem of selcti0n the right, usable and informative features.

3.1 Analysis of Existing System
A detailed review and analysis of existing system of Noshad et al. (2021)[12] was carried out in order to bring to fore areas to improve on in order to use it as a feature selection algorithm for traditional data. We reviewed the following:-
i. The approaches and methods used in the existing system in

the search space in terms of exploration and exploitation

ii. The fitness function used in the genetic algorithm

iii. The suitability and interdependency of Data Value Metric as a fitness function in the genetic algorithm

3.2 Architecture of the Existing systems using Genetic Algorithm for Feature Selection
A typical algorithm for genetic algorithm used in existing s for feature is shown in Figure 1.

Input :   Population Size n: Maximum Number of iterations, Max   **Output:**   Global best solution Y_{bt}

Begin:
 Generate initial population of n chromosomes yi=(1,2,3,4,5..n)
 Set iteration counter t=0
 Compute the fitness of each chromosome
 While(t<Max)
 Select a pair of chromosome from the initial population based on fitness
 Apply crossover operator on the selected pair with cross over probability
 Apply mutation on the offspring with mutation probability
 Replace old population with newly generated population
 Increment the iteration t by 1
 end while
 Return the best solution Y_{bt}
 end

Fig 1: Classical Genetic Algorithm
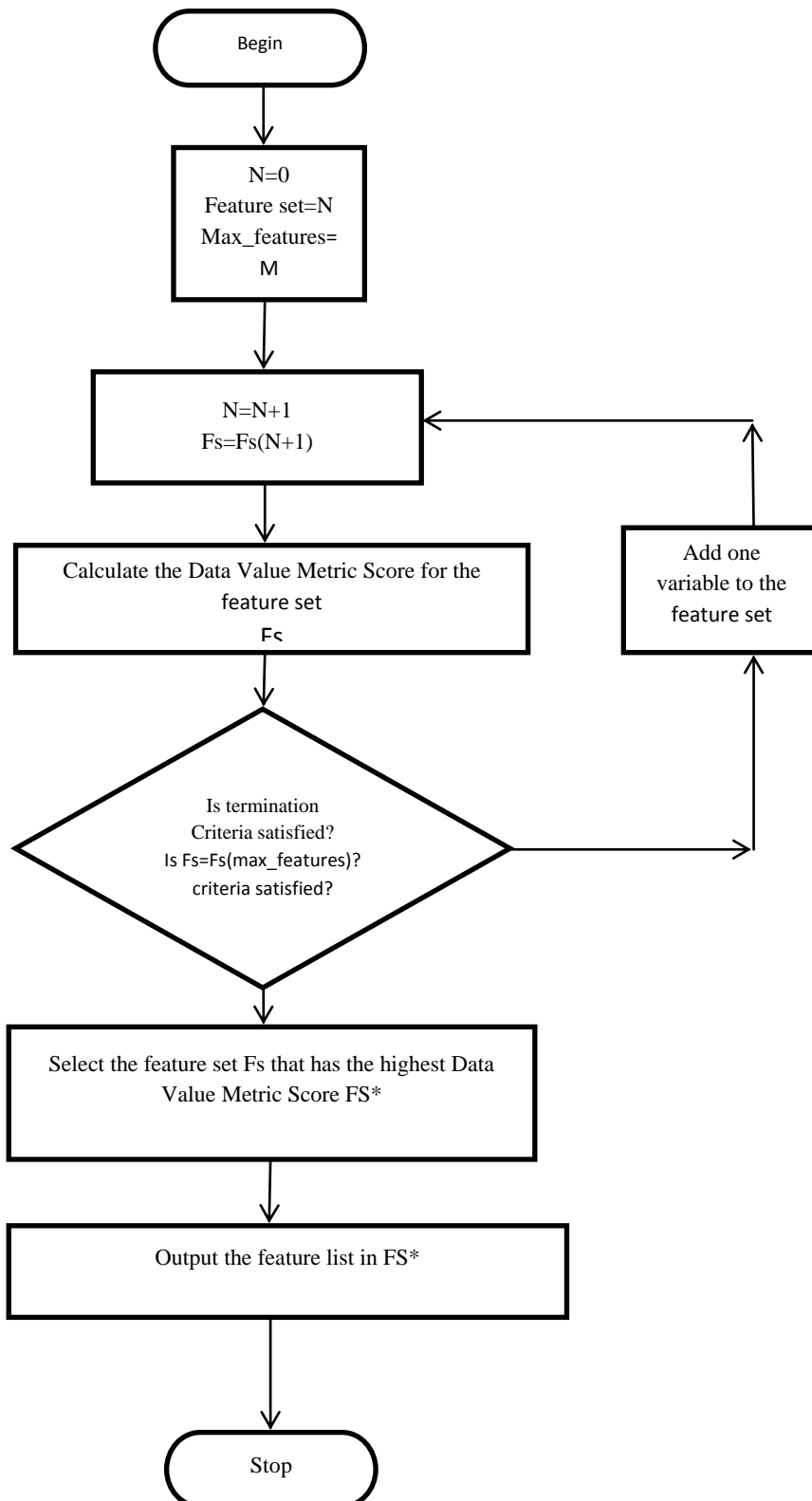
Fig. 2: Existing System of Data Value Metric (Noshad et al. , 2021)

3.3 Limitation of the Existing System of Noshad et. al (2021)[12] and Genetic Algorithm

The use Data Value Metric for feature selection in both supervised and unsupervised machine tasks was proposed By .[12] These methods made use of sequential forward selection search strategy thereby prone to local minima which

ultimately have negative impact on the performance of the machine algorithm when outputted features are used.

i.   There is problem of local minima which is inherently present sequential search strategy which is used in the Data Value Metric algorithm sequential

ii.   there is need to tackle the problem of some

redundant features inherent in Mutual information as the features selected may not be guaranteed to be non- redundant.[2]

```
                    ┌─────────┐
                    │  Begin  │
                    └─────────┘
                         │
  ┌──────────────────────────────────────────────┐
  │ Initial Population: population size, Max. of   │
  │ generation, cross over, mutation               │
  │ and reproduction probabilities                 │
  └──────────────────────────────────────────────┘
                         │
  ┌──────────────────────────────────────────────┐
  │ Randomly generate initial population of        │
  │ features set G=1                               │
  └──────────────────────────────────────────────┘
                         │
  ┌──────────────────────────────────────────────┐
  │ Calculate data value metric score for each     │
  │ individual feature set in the current          │
  │ population as the fitness value                │
  └──────────────────────────────────────────────┘
                         │
               ◇ Is termination criteria satisfied?
                         │
               ◇ Is Gen= Max generation
                         │
  ┌──────────────────────────────────────────────┐
  │ Select parents based on their scores           │
  └──────────────────────────────────────────────┘
                         │
  ┌──────────────────────────────────────────────┐
  │ Reproduce children by application of genetic   │
  │ algorithm operators: selection, cross over     │
  │ and mutaion                                    │
  └──────────────────────────────────────────────┘
                         │
  ┌──────────────────────────────────────────────┐
  │ Gen=Gen +1                                     │
  └──────────────────────────────────────────────┘
                         │
  ┌──────────────────────────────────────────────┐
  │ Replace the current population with children   │
  │ to form the next generation                    │
  └──────────────────────────────────────────────┘

  ┌───────────────────┐
  │ Selected feature  │
  │ set for machine   │
  │ learning          │
  └───────────────────┘

                    ┌─────────┐
                    │  Stop   │
                    └─────────┘
```
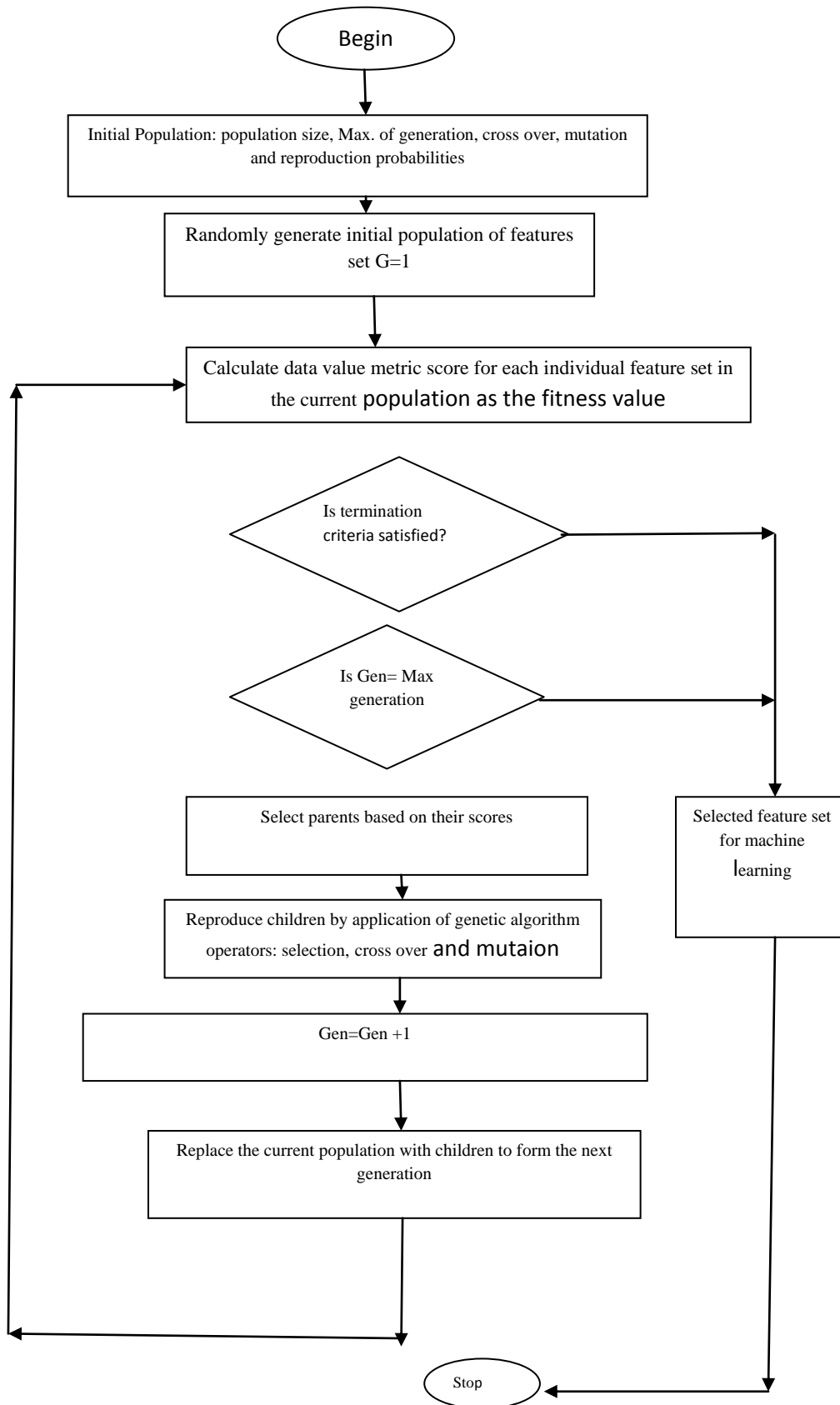
Fig. 3: Proposed Genetic Algorithm based Data Value Metric feature selection for System

3.4 Brief Description of the Components of the Proposed System

3.4.1 Data Value Metric (DVM)

It is a new information theoretical measure proposed by Noshad et al. (2021)[12] that quantifies the useful information content of large heterogeneous and traditional datasets. The DVM make use of data analytical value (utility) and model complexity. It can be used to find out if by appending, expanding, or augmenting a dataset may be beneficial in specific application domains. Subject to the choices of data analytic, inferential, or forecasting techniques employed to interrogate the data, DVM quantifies the information boost, or degradation, associated with increasing the data size or expanding the richness of its features. DVM is defined as a mixture of fidelity and a regularization terms. The fidelity captures the usefulness of the sample data specifically in the context of the inferential task. The regularization term represents the computational complexity of the corresponding inferential method. Inspired by the concept of information bottleneck in deep learning, the fidelity term depends on the performance of the corresponding supervised or unsupervised model. DVM captures effectively the balance between analytical-value and algorithmic-complexity. Changes in the DVM expose the tradeoffs between algorithmic complexity and data analytical value in terms of the sample-size and the feature-richness of a dataset. DVM values may be used to determine the size and characteristics of the data to optimize the relative utility of various supervised or unsupervised algorithms.

3.4.2 Genetic Algorithm (GA)

A GA is a heuristic search algorithm based on the concepts of natural selection and genetics. The idea is to mimic biological processes such as survival of the fittest, to evolve a solution for a problem. GA is a method of evolving a population of chromosomes to new populations using selection along with operations such as crossover and mutation (Mitchel, 1998). Each chromosome consists of genes. Selection operators choose individuals from the population that are the fittest, while crossover and mutation mimic biological processes responsible for introducing diversity to the population. While selection is an exploitation process, crossover and mutation are exploration processes. Evolutionary algorithms are most suitable for problems that involve a large search space i.e. many possible solutions. Other problems require that new solutions are produced at each stage, to explore new options or they involve complex solutions, that can be processed by hand (Mitchel, 1998)[11]. GAs, like the process of evolution, depends on the fittest organisms/solutions to survive. The fitness of an organism/solution is determined based on the problem at hand, and it is a factor which continuously evolves

4. CONCLUSION

There are efforts to collect data, however, not all data is of equal quality or equally informative. Effective data reparation which feature selection comes under is a panacea for a high performing machine learning algorithm.This work presents a hybrid model of Data Value Metric and Genetic Algorithm for feature selection in a supervised learning environment. The feature spaces of the data are represented as the initial population with values. The feature space is randomly selected to form the first generation which is sent into the evaluation metric where the Data Value Metric of that feature set is calculated. The Data Value Metric is used to quantify the amount of information present in the data. The Genetic

Algorithm is applied to the features space of the data until a suitable high Data Value Metric is met. Those feature set that produced the highest Data Value Metric is subsequently used for machine learning endeavours which is hope to produce better classification performance. This is necessary because information or data might suffer from data redundancy and noises as well as local minima if only data value Metric algorithm is used through sequential forward selection search strategy.

5. FUTURE WORK

In the paper, analysis of the existing systems was carried out and some problems were highlight for improvement. The proposed framework combine the power of genetic algorithm in handling local minima with the enormous benefit of quantifying the value of data using Data Value Metric which acts as the fitness function of the genetic algorithm. The search strategy of both algorithms will blend into each other to provide exploration and exploitation in the search space. Future work will delve into the implementation procedure of the framework for the for feature selection using well known publicly available datasets where implementation and evaluation will be provided. The necessary interface and performance metrics for the system will also be provided.

6. REFERENCES

[1] LeskovecJ., Rajaraman A. and David J. (2014),. Dimensionality Reduction. in *Mining of Massive Datasets*, New York, Cambridge University Press, pp. 415-447.

[2] Chandrashekar G. and Sahin F.(2014). A survey on feature selection methods. *Computers and Electrical Engineering,* vol. 40, no. 1, pp. 16-28, 2014.

[3] Lin C. H., ChenH. Y. and. WuY. S (2014). Study of image retrieval and classification based on adaptive features using genetic algorithm feature selection," *Expert Systems with Applications,* vol. 41, no. 15, pp. 6611-6621.

[4] Lu J., Zhao T. and ZhangY., (2008). Feature selection based-on genetic algorithm for imageannotation," *Knowledge-Based Systems,* vol. 21, no. 8, pp. 887-891.

[5] Kushwaha P. And Welekar R.(2016). Feature Selection for Image Retrieval based on Genetic Algorithm. *International Journal of Interactive Multimedia and Artificial Intelligence,* vol. 4, no. 16, pp. 16-21.

[6] Huang H., WuY., ChanY. and LinC.(2010). Study on image feature selection: A genetic algorithm approach," in *IET International Conference on Frontier Computing. Theory, Technologies and Applications*, Taichung.

[7] Gómez F. and Quesada A.(2017). Genetic algorithms for feature selection in Data Analytics.[Online]. Available: https://www.neuraldesigner.com/blog/genetic_algorithms _for_ feature _selection. [Accessed 05 12 2021].

[8] Sivasankar E. and Rajesh R. S. (2012). Design and development of efficient feature Selection and classification techniques for Clinical decision support system," Shodhganga, Tirunalveli.

[9] LiangL., PengJ. and YangB.(2012). Image Feature Selection Based on Genetic Algorithm. In *International Conference on Information Engineering and Applications*, Chongqing

[10] SahinerB., ChanH., WeiD., Petrick N., Helvie M. A., Adler D. D. and Goodsitt M. (1996).Image feature selection by a genetic algorithm: Application to classification of mass and Normal breast tissue. *Medical Physics,* vol. 23, no. 10, pp. 1671-1684.

[11] M. Mitchell, An Introduction to Genetic Algorithms, Cambridge, MA: MIT Press, 1998.

[12] Noshed M, Choi, sun y, hero, A and Iyo, D (20201)a dta value metric for quantifying information . Retrieved 12/8/22, at the journal of bigdata. Springeropen .com>...

[13] S. Khuri, *Genetic Algorithms,* Helsinki, 2017.

[14] Adam Woznica, Phong Nguyen, Alexandros Kalousis (2012) .Model mining for robust feature selection., KDD '12 Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM New York, NY, USA, PP 913-921, 2012.

[15] Yu Land LiuH.(2004). Efficient Feature Selection via Analysis of Relevance and Redundancy,. J. Mach. Learn. Res., vol. 5, pp.1205–1224.

[16] Sourabh Katoch, Sumit Singh Chauhan and Vogay Kumar(2021). A review on genetic algorithm : Past, Present and Future. *Multimedia tools and Applications.* 80:8091-8026.