International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
Article Received: 10 September 2023 Revised: 20 October 2023 Accepted: 30 October 2023

Optimizations Based Feature Selection Method for Disease Survival Prediction

Beschi I S[1], Dr. S. Prakash Kumar[2]
[1]Research Scholar, PG & Department of Computer Science, Maruthupandiyar College (Affiliated to Bharathidasan University, Tiruchirappalli), Vallam, Thanjavur, Tamilnadu, India.
[2]Assistant Professor, PG & Department of Computer Science, Maruthupandiyar College (Affiliated to Bharathidasan University, Tiruchirappalli), Vallam, Thanjavur, Tamilnadu, India.

Abstract

In the realm of survival prediction, identifying relevant features plays a pivotal role in enhancing model accuracy and interpretability. This research proposes a novel feature selection method that leverages the synergies between the Whale Optimization Algorithm (WOA) and Genetic Algorithm (GA) to optimize the selection process. The WOA, inspired by the social behavior of humpback whales, is employed to explore the solution space efficiently, while the GA, inspired by the process of natural selection, is used for refining and evolving potential feature subsets. The proposed hybrid algorithm, termed WOA-GA, introduces a dynamic framework that adaptively adjusts the exploration-exploitation trade-off during the search process. The WOA's exploration capabilities are harnessed in the early stages to efficiently traverse the solution space, while the GA's exploitation capabilities are employed later to fine-tune and evolve promising feature subsets. The synergistic combination of these two optimization techniques aims to mitigate the limitations of each individual algorithm and capitalize on their complementary strengths.

Keywords: Survival prediction, Feature Selection, Whale Optimization Algorithm, Genetic Algorithm, Machine Learning, Classification.

1. INTRODUCTION

Over the last few years, healthcare data has become more complex for the reason that large amount of data are being available lately, along with the rapid change of technologies and mobile applications and new diseases have discovered [1][2][3][4]. Therefore, healthcare sectors have believed that healthcare data analytics tools are really important subject in order to manage a large amount of complex data. The purpose of Artificial Intelligence is to make computers more useful in solving problematic healthcare challenges and by using computers we can interpret data which is obtained by diagnosis of various chronic diseases like Alzheimer, Diabetes, Cardiovascular diseases and various types of cancers like breast cancer, colon cancer etc [5] [6] [7].

Artificial intelligence (AI) and related technologies are increasingly prevalent in business and society, and are beginning to be applied to healthcare [8] [9] [10]. These technologies have the potential to transform many aspects of patient care, as well as administrative processes within provider, payer and pharmaceutical organizations. Today, algorithms are already outperforming radiologists at spotting malignant tumours, and guiding researchers in how to construct cohorts for costly clinical trials.

However, for a variety of reasons, we believe that it will be many years before AI replaces humans for broad medical process domains [11] [12] [13]. Through this research work, various Machine Learning and Deep Learning approaches will be used to enhance the prediction of disease survival [23] [24]. In this research work, considered a dataset with maximum common diseases (diabetes, obesity, Hepatitis, Kidney Disease, HIV) features to predict the survival rates for the considered diseases [14].

2. GENETIC ALGORITHM

Holland proposed the GA in 1975, based on Darwin's biological evolution theory [15]. This algorithm shows a flection of natural selection process where people with the best fitness are selected for reproduction to generate the next-generation children based on an initial population of chromosomes (i.e., solutions). The offspring inherit the parent's characteristics and even pass them down to the next generation. If the parents are in better shape, the children will be in better shape as well, which means they will have a better chance of surviving than their parents [16]. The GA flowchart is shown in Figure 1.

On this foundation, in GA, the journey toward the optimal solution begins with a randomly initialised population of chromosomes. The size of the beginning population is determined by the nature and complexity of the task at hand, and it remains constant throughout the algorithm's iterations. A fitness function examines the values assigned to each chromosome. As a result, parent chromosomes are chosen from the chromosomes with the highest fitness values when compared to other chromosomes. Crossover and mutation operators are used to achieve this.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
Article Received: 10 September 2023 Revised: 20 October 2023 Accepted: 30 October 2023

The crossover operator swaps portions of one chromosome with those of another chromosome at random [20] [21]. As a result, rather than closely matching either of the parent chromosomes, the offspring receives significant characteristics from each of them. This operator sets the tone for delivering higher-quality results. When the mutation operator is applied to a chromosome, the value of one or more genes in a portion of the offspring chromosomes is randomly modified. Following that, the fitness function is used to evaluate the newly generated solutions, and the procedure is repeated until the stopping requirement is reached, at which time the best solution is presented.

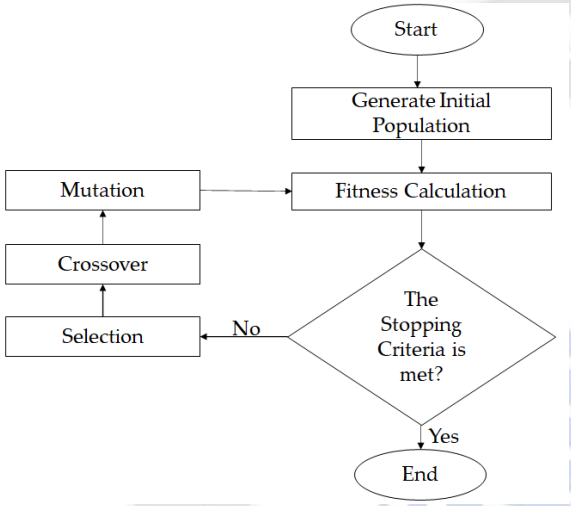

Figure 1: Flowchart of Genetic Algorithm

3. WHALE OPTIMIZATION ALGORITHM

The mathematical model of WOA [17][18] is based on the humpback whales' unique hunting technique:

3.1 Encircling Prey

The objective victim in WOA is the current best agent. Because actual humpback whales can detect and surround the location of a victim, the remaining whales in the population will attempt to improve their site in the direction of the best search agent using the following equations:

$$\vec{D} = \left|\vec{C}.\overrightarrow{X^*}(t) - \vec{C}(t)\right| \quad (1)$$

$$\vec{D}\,(t+1) = \overrightarrow{X^*}(t) - \vec{A}.\vec{D} \quad (2)$$

Where t is the current iteration, $\vec{D}$(t) is a vector indicating position, $\overrightarrow{X^*}(t)$ is a vector reflecting the location of the best solution obtained so far, and $\vec{A}$ and $\vec{C}$ are obtained by:

$$\vec{A} = 2\vec{a}.\vec{r} - \vec{a} \quad (3)$$

$$\vec{C} = 2.\vec{r} \quad (4)$$

Where $\vec{r}$ is an arbitrary vector in the range [0,1], and $\vec{a}$ decreases linearly from 2 to 0 according to the equation:

$$\vec{a} = 2 - \frac{2t}{MaxIter} \quad (5)$$

Where MaxIter is the total iterations.

3.2 Shrinking Encircling Method

Throughout the rounds, the value of $\vec{a}$ in equation (3) is lowered from 2 to 0. As a result, $\vec{A}$ is an arbitrary value in the range [-1,1], and every whale's new location is somewhere between its native location and the current best whale's location.

3.3 Spiral updating position

The space between the whale at X and the victim at X' is calculated here. The following equation is used to represent genuine whales' helix-shaped motion:

$$\vec{X}(t+1) = \overrightarrow{D'}.e^{bl}.\cos(2\pi l) + \overrightarrow{X^*}(t) \quad (6)$$

Where $\vec{D} = \left|\overrightarrow{X^*}(t) - \vec{X}(t)\right|$ denotes the distance between the ith whale and the victim, b is a constant, and l is an arbitrary value in the range [-1,1].

It is assumed that half of the people will choose the spiral model or the diminishing encircling mechanism. The following equation can be used to express this:

$$\vec{X}\,(t+1) = \begin{cases} \overrightarrow{X^*}(t) - \vec{A}.\vec{D} & p<0.5 \\ \overrightarrow{D'}.e^{bl}.\cos(2\pi l) + \overrightarrow{X^*} & p\geq 0.5 \end{cases} \quad (7)$$

Where p is an arbitrary number in [0,1].

3.4 Exploration Phase

Humpback whales search for their prey at random, as seen by their proximity to one another. During this phase, a randomly selected whale, not the current best whale, refreshes the location of a searching whale. In order to execute a global search, the criterion in this phase changes to $|\vec{A}| > 1$:

$$\vec{D} = \left|\vec{C}.\overrightarrow{X_{rand}} - \vec{X}\right| \quad (8)$$

$$\vec{X}\,(t+1) = \overrightarrow{X_{rand}} - \vec{A}.\vec{D} \quad (9)$$

4. WHALE GENETIC OPTIMIZATION FEATURE SELECTION METHOD

Whale Genetic Optimization Feature Selection (WGOFS) method is a feature selection wrapper that combines GA and WOA methods. After WOA is completed, GA is utilised to discover the best option. In each iteration of the WGOFS technique, a fitness function is used to evaluate

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
Article Received: 10 September 2023 *Revised:* 20 October 2023 *Accepted:* 30 October 2023

each search agent, and then the desired feature subset is chosen. In WGOFS, the selected feature subset is represented by the current solution $\vec{X}(t)$. The corresponding features for the elements in $\vec{X}(t)$ that are higher than 0.5 are chosen. The other functionalities are also removed. The fitness function is constructed as, in order to incorporate classification accuracy and specified feature dimension at the same time:

$$FF = \alpha\gamma + (1-\alpha)\frac{W}{N} \quad (10)$$

Where $\alpha \in [0,1]$ equals a balance between the relevance of classification accuracy and the dimension of the specified characteristic. γ is the classification error rate for a given classifier. W is the number of selected features, while N denotes the total number of features. It's worth noting that, according to this definition of fitness, a smaller fitness equals a better solution.

One of the most widely used selection procedures is tournament selection. Two random search agents are chosen first in the tournament selection process. Then a number between 0 and 1 is generated at random. Following that, this number is compared to a given probability (mostly is defined as 0.5). The solution with the better fitness value is accepted when the random number is bigger than the preset probability. Aside from that, the mediocre answer is approved. The weak solution has a better probability of being chosen in a tournament. Tournament selection replaces randomly selecting a search agent to update the position of the next move because it improves the ability to explore the feature space.

Equations (2) and (9) are employed in the WOA to calculate the next step from a randomly chosen solution and the current best solution, respectively. Instead of (2) and (9) in WGOFS, the mutation and crossover operators are used to determine the position of the next step to increase feature space exploration. The mutation rate y is computed using the following formula:

$$y = 0.9 + \frac{-0.9 \times (t-1)}{M-1} \quad (11)$$

Where M is the maximum number of iterations and t denotes the number of iterations currently in progress. As the number of iterations rises, the mutation rate y drops linearly from 0.9 to 0. The crossover operation is carried out between the current solution $\vec{X}(t)$ and the mutation resulting solution. The following is the formula for this operation:

$$\vec{X}(t+1)_d = \begin{cases} \vec{X}(t)_d & \mathcal{Z} < 0.5 \\ \overrightarrow{X_d^{Mut}} & \mathcal{Z} \geq 0.5 \end{cases} \quad (12)$$

The result of the mutation process is $\overrightarrow{X^{Mut}}$, and $\vec{X}(t+1)$ is the newly created solution. Furthermore, in the range [0,1], $\mathcal{Z}$ is a random number, and $\vec{X}(t+1)_d$ is the dth dimension in $\vec{X}(t+1)$.

Algorithm: Whale Genetic Optimization Feature Selection (WGOFS) Method

Input: Number of iterations M, and Number of search agents n.
Output: Position of the best search agent $\overrightarrow{X^*}$.
Step 1: Generate initial population $\vec{X}$
Step 2: Calculate the fitness of each agent using equation (10).
Step 3: Find the current best search agent $\overrightarrow{X^*}$
Step 4: for $t = 1, 2, \ldots, M$ do
  Step 4.1: for i = 1, 2, . . . , n do
    Step 4.1.1: Randomly generate l, r, and p.
    Step 4.1.2: Calculate parameters a, A, and C using (3)(4)(5).
    Step 4.1.3: if p<0.5 then
      If |A| < 1 then
        Employ mutation operation to get $\overrightarrow{X^{Mut}}$ from $\overrightarrow{X^*}$ (best solution) with the rate y using (11)
        Apply crossover between $\overrightarrow{X^{Mut}}$ and $\vec{X}(t)$ to get the new position of $\vec{X}(t+1)$ as the output of crossover using (12).
      Else
        Select a search agent $\overrightarrow{X_{rand}}$ using the tournament selection.
        Employ mutation operation to get $\overrightarrow{X^{Mut}}$ from $\overrightarrow{X_{rand}}$ (best solution) with ratio y using (11).
        Apply crossover between $\overrightarrow{X^{Mut}}$ and $\vec{X}(t)$ to get the new position to $\vec{X}(t+1)$ as the output of the crossover using (12).
      End if
    Else
      Update the position of current solution using (6)
    End if
  Step 4.1.4: End for
  Step 4.2: GA is used to find the best solution around the current best search chromosome.
  Step 4.3: Check if all the chromosomes are in the search space.
  Step 4.4: Calculate the fitness of each agent (chromosome) using (10)
  Step 4.5: Once a better solution appears, update $\overrightarrow{X^*}$.
Step 5: end for
Step 6: Return $\overrightarrow{X^*}$

5. RESULT AND DISCUSSION

5.1 Evaluation Metrics

Table 1 depicts the evaluation metrics for analyzing the performance of the Proposed WGOFS method, and existing feature selection methods using three different classification techniques. The disease survival dataset is considered from the repository [19].

Table 1: Performance Metrics

Metrics	Equation
Accuracy	$\frac{TP+TN}{TP+FN+TN+FP}$
True Positive Rate (TPR) (Sensitivity or Recall)	$\frac{TP}{TP+FN}$
False Positive Rate (FPR)	$\frac{FP}{FP+TN}$
Precision	$\frac{TP}{TP+FP}$
True Negative Rate (Specificity)	1- False Positive Rate (FPR)
Miss Rate	1-True Positive Rate (TPR)
False Discovery Rate	1- Precision

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
Article Received: 10 September 2023 Revised: 20 October 2023 Accepted: 30 October 2023

5.2 Performance Analysis

The performance of the proposed WGOFS method is evaluated with the existing techniques like Whale Optimization Algorithm (WOA), Animal Migration Optimization (AMO), Genetic Algorithm (GA), and Particle Swarm Optimization (PSO) for the given performance metrics using Artificial Neural Network (ANN), Gradient Boosting Tree (GBT), and Random Forest (RF) [22] for the given performance metrics.

Table 2 depicts the Classification Accuracy (in %) obtained by the Proposed WGOFS method, WOA, GA, AMO, and PSO based Feature selection methods using GBT, ANN and RF classification methods. From the table 2, it is shown that the proposed WGOFS method gives more accuracy than the other feature selection methods.

Table 2: Classification Accuracy (in %) obtained by the Proposed WGOFS method, WOA, GA, AMO, and PSO based Feature selection methods using GBT, ANN and RF classification methods

Feature Selection Methods	Classification Accuracy (in %) by Classification Techniques		
	ANN	GBT	RF
Original dataset	55.38	45.63	43.32
WOA	73.85	63.81	58.45
GA	74.99	71.86	68.02
AMO	69.74	65.95	63.54
PSO	68.68	62.65	61.45
Proposed WGOFS method	95.78	92.29	89.63

Table 3 depicts the True Positive Rate (in %) obtained by the Proposed WGOFS method, WOA, GA, AMO, and PSO based Feature selection methods using GBT, ANN and RF classification methods. From the table 3, it is shown that the proposed WGOFS method gives more TPR than the other feature selection methods.

Table 3: True Positive Rate (in %) obtained by the Proposed WGOFS method, WOA, GA, AMO, and PSO based Feature selection methods using GBT, ANN and RF classification methods

Feature Selection Methods	True Positive Rate (in %) by Classification Techniques		
	ANN	GBT	RF
Original dataset	54.49	44.54	42.23
WOA	74.96	64.92	59.56
GA	75.81	72.95	69.13
AMO	68.86	64.86	62.63
PSO	67.77	61.56	60.53
Proposed WGOFS method	95.59	91.38	89.72

Table 4 depicts the False Positive Rate (in %) obtained by the Proposed WGOFS method, WOA, GA, AMO, and PSO based Feature selection methods using GBT, ANN and RF classification methods. From the table 4, it is shown that the proposed WGOFS method gives reduced FPR than the other feature selection methods.

Table 4: False Positive Rate (in %) obtained by the Proposed WGOFS method, WOA, GA, AMO, and PSO based Feature selection methods using GBT, ANN and RF classification methods

Feature Selection Methods	False Positive Rate (in %) by Classification Techniques		
	ANN	GBT	RF
Original dataset	53.61	64.17	65.69
WOA	27.53	33.62	34.47
GA	22.42	30.18	33.47
AMO	38.82	44.51	45.84
PSO	41.72	47.34	48.73
Proposed WGOFS method	5.94	6.41	9.54

Table 5 depicts the Precision (in %) obtained by the Proposed WGOFS method, WOA, GA, AMO, and PSO based Feature selection methods using GBT, ANN and RF classification methods. From the table 5, it is shown that the proposed WGOFS method gives improved precision than the other feature selection methods.

Table 5: Precision (in %) obtained by the Proposed WGOFS method, WOA, GA, AMO, and PSO based Feature selection methods using GBT, ANN and RF classification methods

Feature Selection Methods	Precision (in %) by Classification Techniques		
	ANN	GBT	RF
Original dataset	66.81	53.92	46.76
WOA	78.72	69.82	67.81
GA	79.25	71.38	62.74
AMO	65.88	62.76	58.97
PSO	60.52	61.53	57.85
Proposed WGOFS method	96.52	90.53	80.66

Table 6 depicts the Specificity (in %) obtained by the Proposed WGOFS method, WOA, GA, AMO, and PSO based Feature selection methods using GBT, ANN and RF classification methods. From the table 6, it is shown that the proposed WGOFS method gives improved specificity than the other feature selection methods.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
Article Received: 10 September 2023 Revised: 20 October 2023 Accepted: 30 October 2023

Table 6: Specificity (in %) obtained by the Proposed WGOFS method, WOA, GA, AMO, and PSO based Feature selection methods using GBT, ANN and RF classification methods

Feature Selection Methods	Specificity (in %) by Classification Techniques		
	ANN	GBT	RF
Original dataset	46.39	35.83	34.31
WOA	72.47	66.38	65.53
GA	77.58	69.82	66.53
AMO	61.18	55.49	54.16
PSO	58.28	52.66	51.27
Proposed WGOFS method	94.06	93.59	90.46

Table 7 depicts the Miss Rate (in %) obtained by the Proposed WGOFS method, WOA, GA, AMO, and PSO based Feature selection methods using GBT, ANN and RF classification methods. From the table 7, it is shown that the proposed WGOFS method gives reduced miss rate than the other feature selection methods.

Table 7: Miss Rate (in %) obtained by the Proposed WGOFS method, WOA, GA, AMO, and PSO based Feature selection methods using GBT, ANN and RF classification methods

Feature Selection Methods	Miss Rate (in %) by Classification Techniques		
	ANN	GBT	RF
Original dataset	45.51	55.46	57.77
WOA	25.04	35.08	40.44
GA	24.19	27.05	30.87
AMO	31.14	35.14	37.37
PSO	32.23	38.44	39.47
Proposed WGOFS method	4.41	8.62	10.28

Table 8 depicts the False Discovery Rate (in %) obtained by the Proposed WGOFS method, WOA, GA, AMO, and PSO based Feature selection methods using GBT, ANN and RF classification methods. From the table 8, it is shown that the proposed WGOFS method gives reduced miss rate than the other feature selection methods.

Table 8: False Discovery Rate (in %) obtained by the Proposed WGOFS method, WOA, GA, AMO, and PSO based Feature selection methods using GBT, ANN and RF classification methods

Feature Selection Methods	False Discovery Rate (in %) by Classification Techniques		
	ANN	GBT	RF
Original dataset	33.19	46.08	53.24
WOA	21.28	30.18	32.19
GA	20.75	28.62	37.26
AMO	34.12	37.24	41.03
PSO	39.48	38.47	42.15
Proposed WGOFS method	3.48	9.47	19.34

6. CONCLUSION

Through this research work, optimization techniques-based feature selection is proposed to enhance the disease prediction accuracy of the classification. Whale Optimization Algorithm and Genetic Algorithm is hybridized to extract the most pre-dominant features from the disease's datasets. The accuracy of the proposed Genetic Whale Optimization Feature Selection method is evaluated with various metrics using three different classifiers like GBT, ANN and RF. From the results obtained, it is shown that the proposed GWOFS method gives better result with GBT classifier in terms of Accuracy, TPR, FPR, Specificity, Miss Rate, False discovery rate than other feature selection techniques and classifiers.

REFERENCES

[1] Almansour, Njoud Abdullah, et al. "Neural network and support vector machine for the prediction of chronic kidney disease: A comparative study." Computers in biology and medicine 109 (2019): 101-111.

[2] Belić, Minja, et al. "Artificial intelligence for assisting diagnostics and assessment of Parkinson's disease—A review." Clinical neurology and neurosurgery 184 (2019): 105442.

[3] Liang, Huiying, et al. "Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence." Nature medicine 25.3 (2019): 433-438.

[4] Wu, Chieh-Chen, et al. "Prediction of fatty liver disease using machine learning algorithms." Computer methods and programs in biomedicine 170 (2019): 23-29.

[5] Jo, Taeho, Kwangsik Nho, and Andrew J. Saykin. "Deep learning in Alzheimer's disease: diagnostic classification and prognostic prediction using neuroimaging data." Frontiers in aging neuroscience 11 (2019): 220.

[6] Ngiam, Kee Yuan, and Wei Khor. "Big data and machine learning algorithms for health-care delivery." The Lancet Oncology 20.5 (2019): e262-e273.

[7] Kawakami, Eiryo, et al. "Application of artificial intelligence for preoperative diagnostic and prognostic prediction in epithelial ovarian cancer based on blood biomarkers." Clinical Cancer Research 25.10 (2019): 3006-3015.

[8] Abdar, Moloud, et al. "A new machine learning technique for an accurate diagnosis of coronary artery disease." Computer methods and programs in biomedicine 179 (2019): 104992.

[9] Kwon, Joon-myoung, et al. "Artificial intelligence algorithm for predicting mortality of patients with acute heart failure." PloS one 14.7 (2019): e0219302.

[10] Kiely, David G., et al. "Utilising artificial intelligence to determine patients at risk of a rare disease: idiopathic pulmonary arterial hypertension." Pulmonary Circulation 9.4 (2019): 2045894019890549.

[11] Makino, Masaki, et al. "Artificial intelligence predicts the progression of diabetic kidney disease using big data machine learning." Scientific reports 9.1 (2019): 1-9.

[12] Shamai, Gil, et al. "Artificial intelligence algorithms to assess hormonal status from tissue microarrays in patients with breast cancer." JAMA network open 2.7 (2019): e197700-e197700.

[13] Ding, Yiming, et al. "A deep learning model to predict a diagnosis of Alzheimer disease by using 18F-FDG PET of the brain." Radiology 290.2 (2019): 456-464.

[14] Shen, Jiayi, et al. "Artificial intelligence versus clinicians in disease diagnosis: Systematic review." JMIR medical informatics 7.3 (2019): e10010.

[15] Kanwal, Samina, et al. "An effective classification algorithm for heart disease prediction with genetic algorithm for feature selection." *2021 Mohammad Ali Jinnah University International Conference on Computing (MAJICC)*. IEEE, 2021.

[16] Anbarasi, M., E. Anupriya, and N. C. S. N. Iyengar. "Enhanced prediction of heart disease with feature subset selection using genetic algorithm." *International Journal of Engineering Science and Technology* 2.10 (2010): 5370-5376.

[17] Zamani, Hoda, and Mohammad-Hossein Nadimi-Shahraki. "Feature selection based on whale optimization algorithm for diseases diagnosis." *International Journal of Computer Science and Information Security* 14.9 (2016): 1243.

[18] Nadimi-Shahraki, Mohammad H., Hoda Zamani, and Seyedali Mirjalili. "Enhanced whale optimization algorithm for medical feature selection: A COVID-19 case study." *Computers in biology and medicine* 148 (2022): 105858.

[19] https://archive.ics.uci.edu/ml/datasets/HCC+Survival

[20] Priyadharshini, D., Poornappriya, T.S., & Gopinath, R., A fuzzy MCDM approach for measuring the business impact of employee selection, International Journal of Management (IJM), 11(7), 1769-1775 (2020).

[21] Poornappriya, T.S., Gopinath, R., Application of Machine Learning Techniques for Improving Learning Disabilities, International Journal of Electrical Engineering and Technology (IJEET), 11(10), 392-402 (2020).

[22] Poornappriya, T.S., Selvi, V., Evolutionary Optimization of Artificial Neural Network for Diagnosing Autism Spectrum Disorder, International Journal of Electrical Engineering and Technology (IJEET), 11(7), 47-61 (2020).

[23] Poornappriya, T. S., and M. Durairaj. "High relevancy low redundancy vague set based feature selection method for telecom dataset." *Journal of Intelligent & Fuzzy Systems* 37.5 (2019): 6743-6760.

[24] Durairaj, M., and T. S. Poornappriya. "Why feature selection in data mining is prominent? A survey." *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications: AISGSC 2019*. Springer International Publishing, 2020.