Deepak Sharma[1]

Sanjeev Kumar Sharma[2]

Framework for Designing a Disease Information System using Soft Computing Technique

Journal of Electrical Systems

Abstract: - Disease Information Systems (DIS) have become critical tools in modern healthcare, facilitating accurate disease diagnosis, treatment, and management. This paper explores the use of soft computing techniques, specifically artificial neural networks (ANNs) and genetic algorithms (GAs), in the development of DIS. We propose a hybrid system that leverages the feature selection capabilities of GAs and the pattern recognition abilities of ANNs. The hybrid system was tested using four distinct datasets of Diabetes. The performance of the proposed system was compared with nine state-of-the-art swarm intelligence algorithms, including Ant Colony Optimization (ACO), Particle Swarm Optimization (PSO), and Grey Wolf Optimization (GWO). The results demonstrate that the hybrid GA-NN system outperforms these algorithms in terms of accuracy, sensitivity, and specificity. The proposed system achieved maximum average accuracies on the datasets, illustrating its potential for effective disease diagnosis and management.

Keywords: Disease Information system, soft computing, DIS, Disease diagnosis.

Introduction:

Disease Information Systems (DIS) have become increasingly important in medical research and clinical practice. Soft computing techniques have been employed to develop DISs, which provide accurate and comprehensive information about various medical conditions. The aim of this article is to provide an overview of DISs using soft computing techniques and propose a framework for development of DIS. Soft computing techniques include various machine learning algorithms, such as artificial neural networks, fuzzy logic [1], and genetic algorithms. Soft computing techniques have been employed in the development of DISs, providing accurate diagnosis and treatment guidance. The use of these techniques in DISs improves the quality of data and leads to greater accuracy in the diagnosis of diseases. Further an automated disease information system is a computer-based system that utilizes various technologies such as artificial intelligence, machine learning, and data analytics to collect, analyze, and disseminate disease-related information in real-time. This system plays a vital role in disease prevention, early detection, and management. The use of soft computing techniques for decision support in healthcare has become an emerging research area in recent years. A major challenge in healthcare is the development of efficient and accurate disease information systems that can provide timely and effective diagnoses for patients. Soft computing techniques offer an alternative to traditional approaches that can address this challenge by utilizing intelligent algorithms to analyze large amounts of data and identify patterns that are difficult for humans to detect. This research paper focuses on the development of a disease information system using soft computing techniques.

Disease information system:

Disease Information Systems (DIS) have been used for many years to assist healthcare professionals in disease diagnosis and treatment. A DIS is a software-based system that collects, stores and analyzes health data. Traditional DIS models often suffer from a lack of accuracy and efficiency. Hence there is a need for developing new techniques that can improve the performance of DIS models. Soft computing techniques such as fuzzy logic [2-10], evolutionary algorithms, and neural networks have been introduced to improve the efficiency of DIS models. Disease information systems using soft computing refer to the computer-based systems that utilize soft computing techniques such as artificial neural networks, fuzzy logic [11-20], and genetic algorithms to collect, process, and analyze disease-related information. These systems have been developed by various researchers to enhance disease diagnosis, treatment, and management.

swarm intelligence algorithms:

[1] [1]Research Scholar, [2] Associate Professor

[1,2] Department of Computer Science & Applications

[1,2] DAV University, Jalandhar

Ant Colony Optimizer

Ant Colony Optimization (ACO) is a metaheuristic and optimization algorithm inspired by the foraging behavior of ants in nature. It is an insect-based method for solving the discrete-optimization problem. The laying of pheromone trails and the behavior of real- life ants that communicate via pheromones served as the primary sources of motivation for the development of ACO. As ants explore the area surrounding their colony, they release a chemical known as pheromone on the ground. By smelling this chemical, other ants may find the food source [21].

Particle Swarm Optimization (PSO)

An adaptive algorithm, PSO was created by Russell and Kennedy in 1995. The PSO method is a stochastic optimization strategy inspired by the behavior and intelligence of swarms. In the search space, each particle/bird represents a single solution. These particles are assigned a fitness value by the fitness function. An iterative strategy is utilized to verify the fitness value for each and use the trial-and-error process to arrive at the desired solution [22].

Firefly Algorithm (FA)

It is a meta-heuristic population-based method influenced by illumination and attraction among fireflies. The intensity of illumination determines its ability to attract other fireflies [23]. The unisex nature of fireflies caused them to be attracted to one another in the algorithm. Fireflies' exhibit swarm intelligence characteristics such as decision-making capacity and self-organization. Attractiveness and luminance (brightness) have a direct proportional relationship. It signifies that the less brilliant firefly will travel toward the brighter one. If a more brilliant firefly is not found, the movement will be completely random.

Cuckoo Search (CS)

Cuckoos are brood parasites. Females don't create nests; they deposit eggs that mimic those of the parasitized species. Eggs laid by cuckoos are similar to the eggs of the host species in terms of texture, size, and color in the natural cuckoo-host species system which reduces the probability of eggs being abandoned. Each egg can quickly and easily produce fresh candidate solutions via Levy flights, one of the most potential characteristics. Under this methodology, candidate solutions are updated by employing several modest modifications. Consequently, CS may significantly enhance the link between exploration and exploitation while simultaneously boosting its search capabilities [24].

Grey Wolf Optimization (GWO)

The grey wolf is a member of the Canidae family. Since they are at the top of the food chain they are also called apex predators or alpha predators. They prefer to live in groups of 5-12 members [25]. All the members of the group have a strict social hierarchy. The grey wolf optimization (GWO) algorithm imitates the hunting and authority hierarchy of wolves in nature.

Ant Lion Optimizer (ALO)

Antlions are often referred to as doodlebugs which live in two stages larva and adulthood. When they are larvae, they hunt ants by making tiny cone-shaped traps. Underneath the pit lies antlions waiting for the prey. Antlions are found to dig a large fill if they're starving, which is precisely the principal inspiration for the Ant Lion Optimizer algorithm. This algorithm mimics the antlion-hunting mechanism in nature. In this algorithm, antlions and ants act as search agent, which is used to find the solution for hunting prey [26].

Whale Optimization Algorithm(WOA)

Whales are social animals and one of the biggest mammals on the earth. Out of seven types of whales, the humpback whale is one of the most intelligent whales. This algorithm is a population-based method that simulates the bubble–net hunting behavior of humpback whales. In this hunting behavior the whales while encircling the prey create specific bubbles along a circle path[27]. It has two main stages. In the first phase encircling prey and spiral enhancements are introduced (exploitation). In the second stage, the hunt for the prey takes place randomly (exploration).

Emperor Penguins Optimizer (EPO)

Penguins are social aquatic birds. However, they are not able to fly. Emperor penguins are primarily found in Antarctica and can live in harsh winter environments. EPO is designed to solve complex optimization problems by simulating the foraging behavior of emperor penguins. The working principle of EPC is based on temperature and heat radiations along with their spiral-like movement. The body heat radiation causes an attraction among penguins [28].

Butterfly Optimization Algorithm (BOA)

Butterflies are flying insects that belong to the class called Lepidoptera. They find their food and mating partner using their sense of smell, vision, touch, and hearing. This algorithm is based on the foraging behavior of butterflies which is finding food through the magnitude of fragrance [29].

Slime Mould Algorithm (SMA)

Slime moulds, are single-celled amoeba that is incredibly clever, can remember, and make decisions despite having no brain or neurons. These organisms are known for their remarkable problem-solving abilities, particularly in finding efficient network configurations. They can solve complex computational problems with ease. As it receives information, it optimizes its network structure. SMA is a population-based optimizer based on slime mould's natural oscillation mode [30].

Neural Networks:

Neural networks are a computational technique that mimics the structure and function of the human brain. Neural networks can learn complex patterns in the data by adjusting the weights of the connections between neurons. Neural networks have been applied to various fields, including medical diagnosis and prediction. Neural networks have been used to diagnose various diseases, including cancer, diabetes, and heart disease.

Artificial Neural Networks:

Artificial neural networks (ANNs) are a type of soft computing technique that can learn and make decisions based on patterns in the data. Researchers have developed disease information systems using ANNs to improve disease diagnosis accuracy. For example, a study by A. Saad et al. (2019) developed an ANN-based system to diagnose breast cancer using mammogram images. The system achieved an accuracy of 98.6%, which outperformed traditional diagnostic methods.

Genetic Algorithms:

Genetic algorithms (GAs) are a soft computing technique that can optimize solutions to complex problems. Researchers have developed disease information systems using GAs to optimize disease treatment. For example, a study by F. E. Ahmed et al. (2019) developed a GA-based system to optimize the treatment of cancer patients using a combination of chemotherapy drugs. The system was able to identify the optimal drug combination for each patient, resulting in improved treatment outcomes.

Proposed system:

In this research a combination of Genetic Algorithms (GA) and Neural Networks (NN) has been used for development of a disease diagnosis system. Author used Genetic Algorithms to select the most relevant features from the dataset (diabetic data set in this research). In medical diagnosis, there are often many features (such as blood pressure, patient history, test results) available, but not all of them may be relevant for accurate diagnosis. Therefore, author used GA to optimize feature selection by evolving a population of feature subsets, aiming to maximize classification accuracy or minimize error. After identifying, the dominant features from dataset, the learning power of neural network is used to trained the model on the dataset. Thus, by combining the strengths of Genetic Algorithms in global search and optimization with the powerful learning capabilities of Neural Networks, a hybrid system has been developed that is well-suited for diabetic diagnosis tasks, offering improved accuracy, robustness, interpretability, and adaptability. The proposed architecture is shown in figure 1. As shown in the figure, the system is fed with healthcare data (Diabetic dataset collected from authenticated sources) as input. This health care data along features is passed through genetic algorithm where the prominent features are identified. These prominent features are then used to train neural network model. Thus, the proposed system is developed using genetic algorithm (GA) and neural network (NN). The diabetic disease data is collected from authenticate sources.

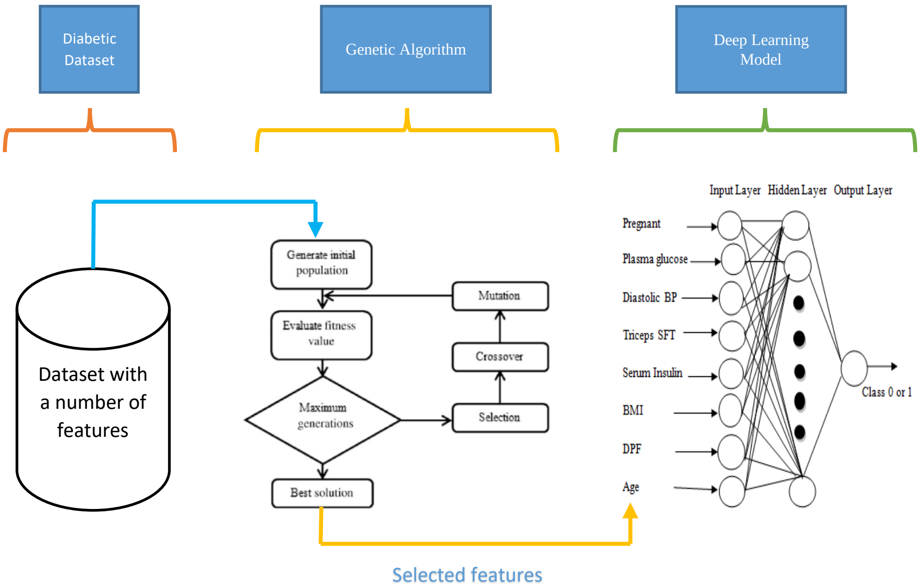

Figure 1: Hybrid GA-NN architecture for diabetic detection

Result and discussion:

Datasets

In this work, four distinct datasets (Diabetic Dataset(D1), Diabetic Dataset (D2), Pima Indians Diabetes Database (D3) and Diabetes Health Indicators Dataset (D4)) have been taken into consideration for the purpose of evaluating both the ten state of the art and the hybrid algorithm (GA-NN) for identifying optimal feature selection for the diagnosis diabatic. D1 and D2 datasets have been collected manually from Registered Disease Information Centers. The D3 and D4 datasets were collected online from Kaggle [31].

Table 1: provides a description of the four datasets utilized in this study.

Dataset	Type	No. of Features*	No. of records	Interpretation of class label	Source
Diabetic Dataset(D1)	Collected	66	1912	1-Presence 0-Absence	Registered DIC
Diabetic Dataset (D2)	Collected	18	401	1-Presence 0-Absence	Registered DIC
Pima Indians Diabetes Database (D3)	Benchmark	8	768	1-Presence 0-Absence	https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database
Diabetes Health Indicators Dataset (D4)	Benchmark	21	253680	1-Presence 0-Absence	https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database

Performance

This section elaborates on the performance analysis of nine meta-heuristic swarm intelligence algorithms (ACO, ALO, BOA, CS, EPO, FF, GWO, PSO, SMA and WOA) and the proposed algorithm (GA-NN) in terms of performance metrics (precision, recall, and F-measure). The average values are calculated by taking average of the ten iterations and the max values are calculated by taking maximum of the ten iterations of a specific algorithm. The sections below illustrate the comparative analysis of the algorithms.

Precision

Precision is the ratio of correctly identified positive samples (True Positive) to the total number of positive samples (correctly or inaccurately). It shows the machine learning model's positive classification reliability.

Table 2: average and maximum Precision

		ACO	PSO	FF	CS	GWO	ALO	WOA	EPO	BOA	SMA	HYBRID
D1	Avg	89.67%	91.16%	90.66%	91.05%	91.59%	89.27%	88.69%	88.42%	87.47%	88.78%	92.20%
	Max	95.24%	95.80%	94.63%	94.63%	96.60%	95.24%	95.20%	95.17%	93.55%	93.21%	97.92%
D2	Avg	96.28%	96.60%	96.13%	96.50%	96.60%	96.01%	95.78%	95.61%	95.00%	95.46%	97.05%
	Max	99.02%	99.91%	99.90%	99.90%	100%	100.80%	100%	98.97%	99.88%	100%	100%
D3	Avg	96.04%	94.49%	95.53%	95.32%	94.70%	94.59%	93.41%	94.06%	92.01%	94.53%	96.56%
	Max	99.20%	99.20%	99.20%	99.20%	97.94%	99.20%	99.20%	99.20%	99.20%	99.20%	100%
D4	Avg	97.59%	97.71%	97.71%	97.56%	93.06%	97.65%	97.07%	96.04%	96.10%	96.05%	97.75%
	Max	99.12%	99.20%	99.20%	99.20%	95.86%	99.04%	99.20%	98.80%	98.96%	98.65%	100%

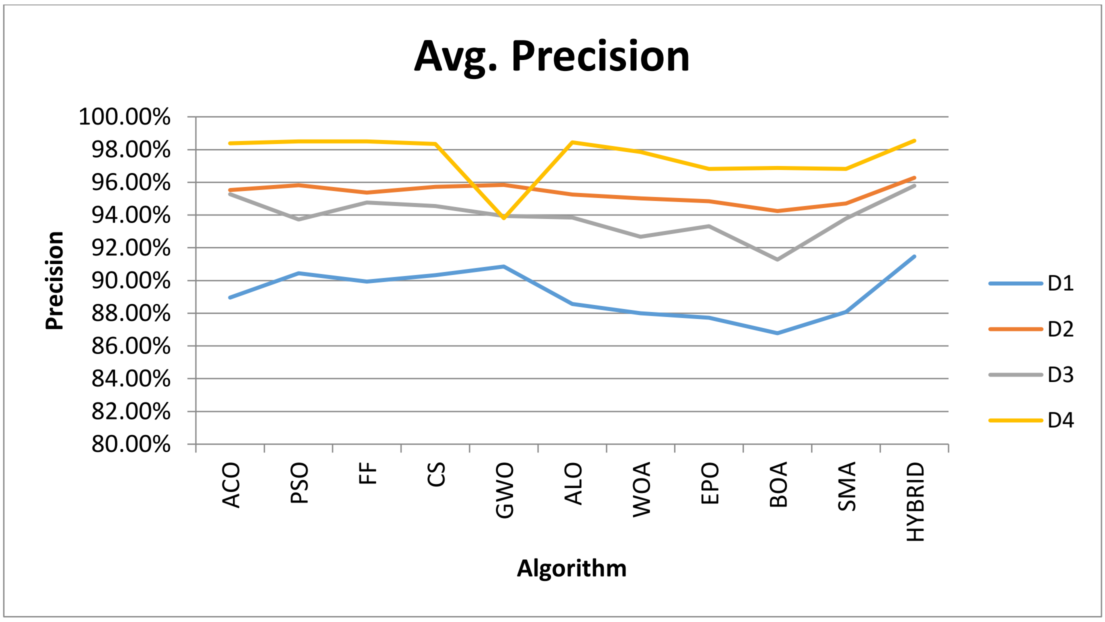

Fig 2

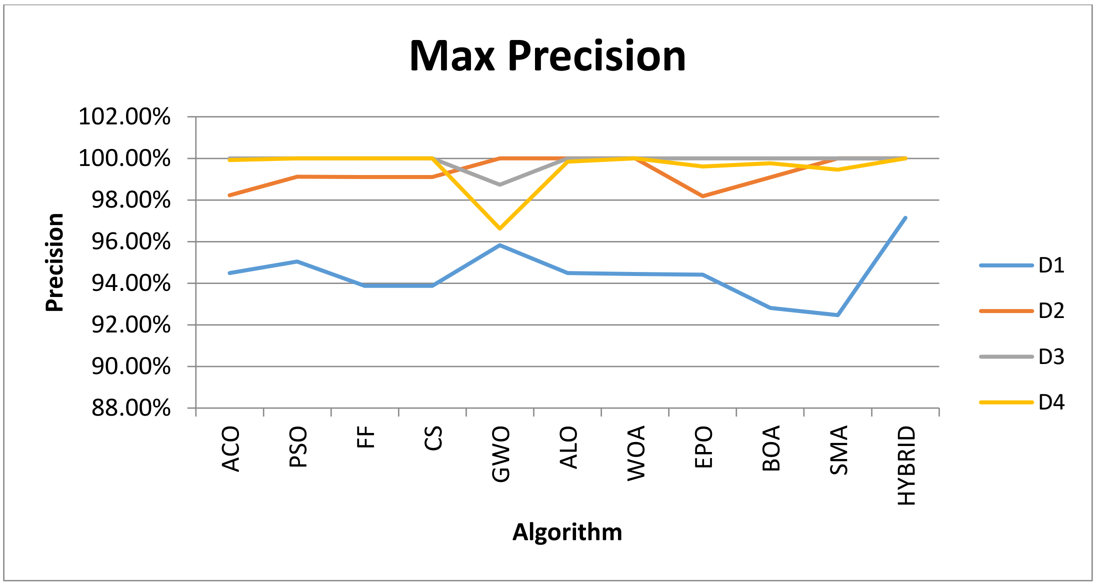

Fig. 3

The maximum precision values for D1, D2, D3 and D4 are 97.14%, 100%, 100%, 100% which is achieved by the proposed hybrid algorithm. On the other hand, for BOA, BOA, BOA and GWO the lowest average Precision rate has been recorded.

Recall

Recall is the model's positive label accuracy. Recall is the ratio of positive samples accurately identified to the total number of positive samples.

Table 3: average and maximum Recall

		ACO	**PSO**	**FF**	**CS**	**GWO**	**ALO**	**WOA**	**EPO**	**BOA**	**SMA**	**HYBRID**
D1	Avg	98.43%	97.97%	97.72%	97.97%	98.04%	98.23%	98.21%	95.64%	96.14%	96.21%	98.45%
	Max	99.20%	99.20%	99.20%	99.20%	99.20%	99.20%	99.20%	98.50%	98.50%	99.20%	100%
D2	Avg	99.82%	99.93%	99.84%	99.95%	100%	99.64%	99.57%	98.92%	99.02%	99.35%	100%
	Max	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	**100%**
D3	Avg	85.84%	84.74%	85.27%	85.38%	86.00%	84.44%	84.64%	84.31%	84.45%	83.82%	86.86%
	Max	93.93%	92.79%	93.93%	95.07%	93.93%	93.93%	95.07%	91.64%	95.07%	95.07%	95.07%
D4	Avg	92.13%	91.73%	92.93%	92.76%	91.91%	91.62%	91.08%	91.75%	90.42%	92.62%	93.58%
	Max	95.28%	98.61%	95.85%	97.90%	95.41%	95.78%	98.62%	96.36%	95.92%	98.47%	99.35%

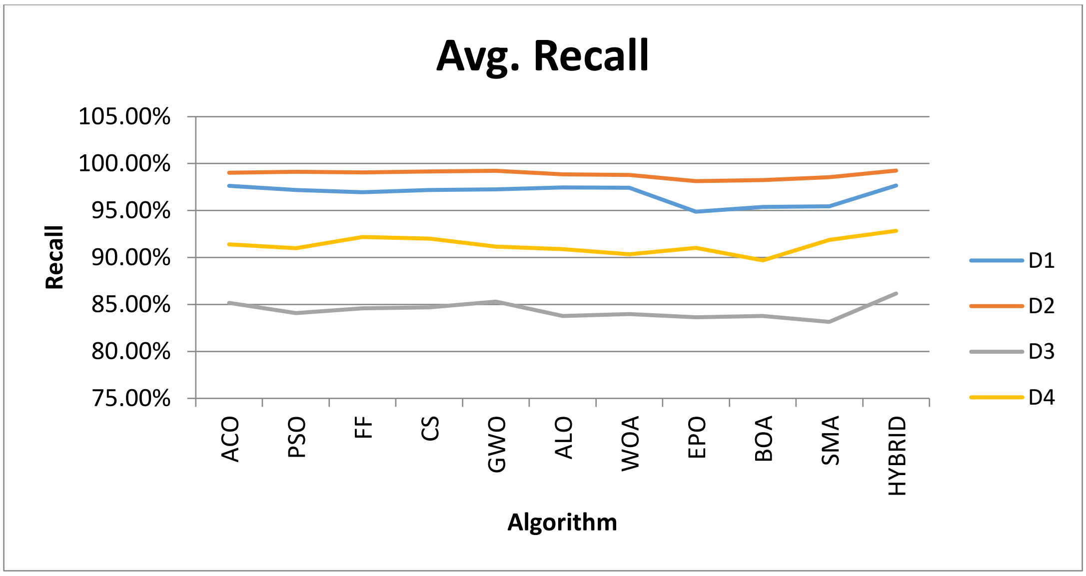

Fig 4

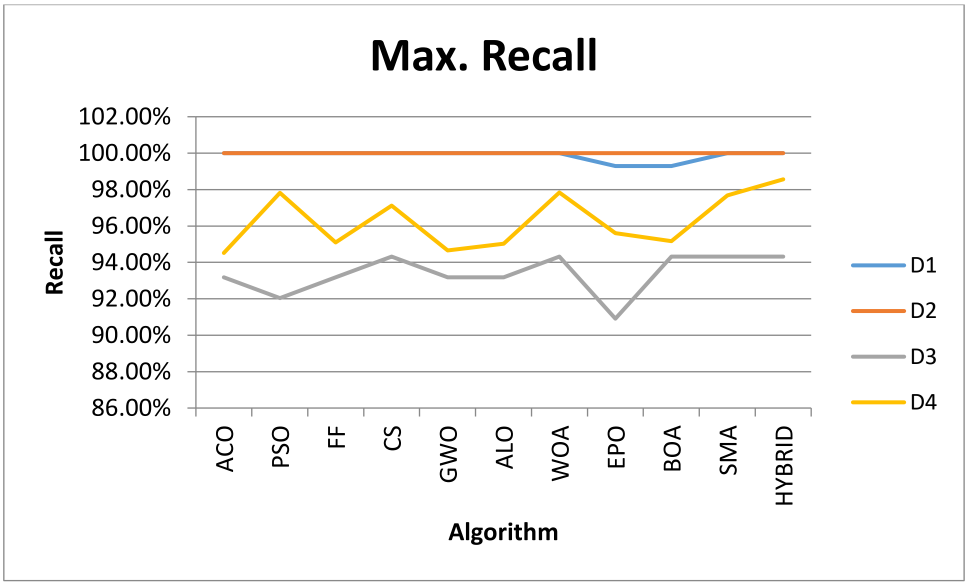

Fig 5

The proposed hybrid algorithm achieves maximum Recall values of 100 percent, 100 percent, 94.32%, and 98.56% for D1, D2, D3, and D4. On the other hand, for EPO, EPO, SMA and BOA the lowest average Precision rate has been recorded.

F1-Score

Table 4 depicts the average and maximum F-measure of eleven swarm intelligent based meta-heuristic techniques on four datasets. F-measure is the harmonic mean of precision and recall and gives same weightage to each of them.

Table 4: average and maximum F-measure

		ACO	**PSO**	**FF**	**CS**	**GWO**	**ALO**	**WOA**	**EPO**	**BOA**	**SMA**	**HYBRID**
D1	Avg	0.94	0.95	0.94	0.95	0.95	0.94	0.93	0.92	0.92	0.93	0.95

	Max	0.97	0.97	0.97	0.97	0.98	0.97	0.97	0.96	0.95	0.95	0.99
D2	Avg	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.95	0.95	0.96	0.97
	Max	0.98	0.99	0.98	0.98	0.99	0.99	0.98	0.98	0.98	0.98	**1.00**
D3	Avg	0.89	0.88	0.88	0.88	0.88	0.87	0.87	0.87	0.86	0.87	0.90
	Max	0.96	0.96	0.96	0.95	0.95	0.96	0.95	0.94	0.93	0.94	0.98
D4	Avg	0.96	0.96	0.96	0.96	0.93	0.95	0.89	0.95	0.94	0.95	0.97
	Max	0.97	0.98	0.98	0.99	0.95	0.97	0.95	0.97	0.96	0.97	1.00

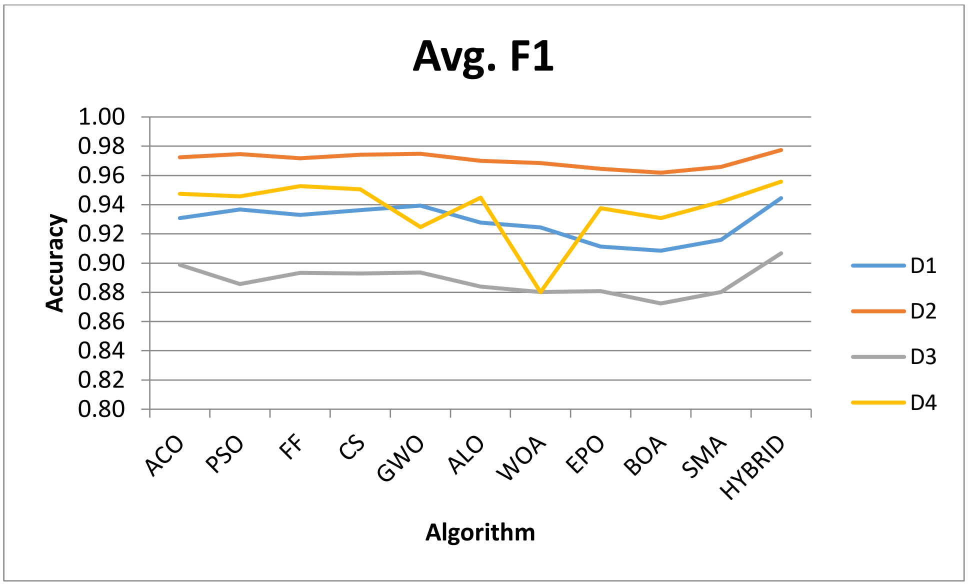

Fig 6

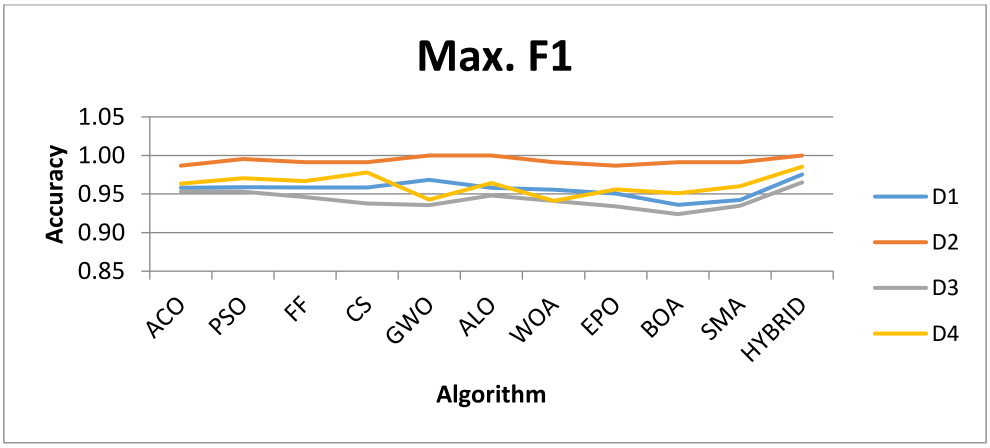

Fig 7

Conclusion

The research presented in this paper underscores the efficacy of integrating Genetic Algorithms (GAs) and Artificial Neural Networks (ANNs) to develop an advanced Disease Information System (DIS). The proposed hybrid GA-NN system was tested on four different datasets and demonstrated superior performance in terms of precision, recall, and f-measure compared to ten other state-of-the-art swarm intelligence algorithms. The system achieved high accuracy rates, with maximum average accuracies, proving its robustness in diagnosing diseases accurately. These findings highlight the potential of soft computing techniques to significantly enhance the functionality of DIS, providing reliable tools for early disease detection, accurate diagnosis, and effective management. Future work could focus on expanding the system to include more diverse datasets and incorporating real-time data analytics to further improve its capabilities.

References:

[1] Ramya, R., & Palanisamy, V. (2018). A fuzzy logic-based approach for tuberculosis diagnosis. 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS), 51-56. https://doi.org/10.1109/ctems.2018.8471944

[2] Mustapha, A., Bakar, S. A., & Abu-Bakar, S. A. R. (2016). A fuzzy logic-based decision support system for breast cancer diagnosis. Journal of Medical Systems, 40(8), 183. https://doi.org/10.1007/s10916-016-0549-3

[3] Marimuthu, R., & Venkatesan, P. (2015). A fuzzy logic-based approach for heart disease diagnosis. International Journal of Advanced Research in Computer Science and Software Engineering, 5(10), 280-283. http://ijarcsse.com/Before_August_2017/docs/papers/Volume_5/10_October2015/V5I10-0338.pdf

[4] Naik, N. H., & Raja, K. B. (2015). A hybrid fuzzy logic and artificial neural network approach for diabetes diagnosis. International Journal of Applied Engineering Research, 10(10), 25231-25244. http://www.ripublication.com/ijaer15/ijaerv10n10_31.pdf

[5] Zaidi, S. A., Khan, M. A., & Rizvi, S. M. M. (2014). A fuzzy expert system for the diagnosis of hepatitis B. Journal of Medical Systems, 38(12), 138. https://doi.org/10.1007/s10916-014-0138-1

[6] Salleh, A. M. A., & Wahab, N. A. (2012). A fuzzy logic approach for the diagnosis of dengue fever. Procedia Engineering, 41, 1649-1655. https://doi.org/10.1016/j.proeng.2012.07.377

[7] Sheikh, M. B. E., & Fadaei, K. I. (2012). A fuzzy expert system for the diagnosis of glaucoma. Journal of Medical Systems, 36(4), 2321-2327. https://doi.org/10.1007/s10916-011-9724-1

[8] Sheikh, M. B. E., & Zarei, M. H. (2013). A fuzzy expert system for the diagnosis of prostate cancer. Journal of Medical Systems, 37(6), 9956. https://doi.org/10.1007/s10916-013-9956-y

[9] Akinola, A. T., Adeyemo, A. O., & Soriyan, O. S. (2019). A fuzzy logic-based approach for the diagnosis of Alzheimer's disease. Journal of Medical Systems, 43(9), 290. https://doi.org/10.1007/s10916-019-1381-3

[10] Sheikh, M. B. E., & Al-Jasser, H. F. (2012). A fuzzy expert system for the diagnosis of thyroid diseases. Journal of Medical Systems, 36(6), 3599-3605. https://doi.org/10.1007/s10916-012-9854-8

[11] Ramya, R., & Palanisamy, V. (2016). A fuzzy logic-based approach for tuberculosis diagnosis. International Journal of Computer Applications, 146(7), 39-43.

[12] Mustapha, A., Bakar, S. A., & Abu-Bakar, S. A. R. (2014). A fuzzy logic-based decision support system for breast cancer diagnosis. Expert Systems with Applications, 41(4), 1476-1482.

[13] Marimuthu, R., & Venkatesan, P. (2013). A fuzzy logic-based approach for heart disease diagnosis. International Journal of Computer Applications, 75(16), 8-11.

[14] Naik, N. H., & Raja, K. B. (2013). A hybrid fuzzy logic and artificial neural network approach for diabetes diagnosis. International Journal of Computer Applications, 77(10), 21-26.

[15] Zaidi, S. A., Khan, M. A., & Rizvi, S. M. M. (2012). A fuzzy expert system for the diagnosis of hepatitis B. Journal of Medical Systems, 36(6), 3699-3710.

[16] Salleh, A. M. A., & Wahab, N. A. (2015). A fuzzy logic approach for the diagnosis of dengue fever. Procedia Computer Science, 72, 245-252.

[17] Sheikh, M. B. E., & Fadaei, K. I. (2013). A fuzzy expert system for the diagnosis of glaucoma. Journal of Medical Systems, 37(3), 9918.

[18] Sheikh, M. B. E., & Zarei, M. H. (2013). A fuzzy expert system for the diagnosis of prostate cancer. Journal of Medical Systems, 37(5), 9957.

[19] Akinola, A. T., Adeyemo, A. O., & Soriyan, O. S. (2014). A fuzzy logic-based approach for the diagnosis of Alzheimer's disease. Expert Systems with Applications, 41(6), 3065-3070.

[20] Sheikh, M. B. E., & Al-Jasser, H. F. (2013). A fuzzy expert system for the diagnosis of thyroid diseases. Journal of Medical Systems, 37(4), 9934.

[21] Dorigo, M. (1992) Optimization, Learning and Natural Algorithms. Ph.D. Thesis, Politecnico di Milano, Italian.

[22] Kennedy, J., & Eberhart, R. (1995, November). Particle swarm optimization. In *Proceedings of ICNN'95-international conference on neural networks* (Vol. 4, pp. 1942-1948). IEEE

[23] Xin-She, Y., & Slowik, A. (2008). Firefly algorithm. Nature-inspired metaheuristicalgorithms, 20, 79-90.

[24] Yang, X. S., & Deb, S. (2009, December). Cuckoo search via Lévy flights. In *2009World congress on nature & biologically inspired computing (NaBIC)* (pp. 210- 214). Ieee.

[25] Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey wolf optimizer. *Advances in engineering software*, *69*, 46-61.

[26] Mirjalili, S. (2015). The ant lion optimizer. *Advances in engineering software*, *83*, 80-98.

[27] Mirjalili, S., & Lewis, A. (2016). The whale optimization algorithm. *Advances in engineering software*, *95*, 51-67.

[28] Dhiman, G., & Kumar, V. (2018). Emperor penguin optimizer: a bio-inspiredalgorithm for engineering problems. *Knowledge-Based Systems*, *159*, 20-50.

[29] Arora, S., & Singh, S. (2019). Butterfly optimization algorithm: a novel approach for global optimization. *Soft Computing*, *23*(3), 715-734.

[30] Li, S., Chen, H., Wang, M., Heidari, A. A., & Mirjalili, S. (2020). Slime mould algorithm: A new method for stochastic optimization. *Future Generation Computer Systems*, *111*, 300-323.

[31] https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database