3rd International Conference on Scientific and Academic Research

December 25-26, 2023 : Konya, Turkey

AS-Proceedings
https://alls-academy.com/index.php

https://www.icsarconf.com/

Harris hawks optimization algorithm with a linear regression model to estimate the number of the accidents, injuries and fatalities in Türkiye

Murat Aslan*

[1] *Department of Computer Engineering, Faculty of Engineering, Şırnak University, Şırnak, Türkiye*

(murataslan@sirnak.edu.tr) Email of the corresponding author

Abstract – Estimating the number of accidents, injuries and fatalities is one of the most important issue for individuals and society. This is because traffic accidents cause a lot of serious consequences to people, such as personal injuries, damage to property, and loss of time. Therefore, in this study, a new model is developed using the Harris hawks optimization algorithm (HHO) and a linear regression model for estimating the number of accidents, injuries and fatalities problem of Türkiye. In the experiments, Türkiye's observed number of accidents, injuries and fatalities, population and number of vehicles historical records taken from Turkish statistical institute (TUIK) for the years between 2002 and 2022 have been used. The historical data records of Türkiye for the years between 2002 and 2022 will be used in this study for the first time. Three different linear regression models have been proposed to solve the problems that are the subject of this study. This is because the number of fatalities, the number of injuries, and the number of accidents are different output parameters, and it is necessary to have a linear regression model for each output parameter separately. In addition, the number of vehicles and population indicators are used as input parameters of the all proposed models. In the experiments, the HHO algorithm is implemented on each problem (number of fatalities, number of injuries, and number of accidents) with 30 independent runs. And, the experimental results are reported in terms of the best fitness values of the 30 runs, the errors, and the relative errors. According to the experimental results, HHO method based proposed models produces comparable performance on the problem dealt with this study in terms of solution quality and robustness.

Keywords –Estimation Problems, Harris Hawks Optimization Algorithm, Traffic Accidents, Linear Model.

I. INTRODUCTION

Traffic accidents are one of the inevitable consequences of traveling on the road. Traffic accidents cause some serious economic impacts on individuals and society. Some of the more tangible personal costs include personal injury, property damage, and lost time [1]. Road traffic injuries are a major public health problem. According to the World Health Organization (WHO), road traffic injuries are responsible for approximately 1.3 million deaths each year. An additional 20 million to 50 million people suffer from non-fatal injuries, and many more are disabled as a result of their injuries [2]. For this reason, a collective effort is needed to develop an impactful and sustainable road injury prevention strategy [3].

Road is the most preferred means of transporting in Türkiye. It is used for the transportation of both passengers and goods. More than 90% of all transportation is carried out by road. The number of vehicles and traffic accidents has increased due to population growth, economic development and rising standards of living [4]. There are thousands of people affected by traffic accidents according to the historical data of Türkiye taken from the Turkish statistical institute (TUIK) [5]. As it is the case all over the world, road traffic accidents are one of the main causes of death in Türkiye. Consequently, traffic congestion and deaths represent a significant cost to the economy [6, 7].

There are many different approaches to solving estimation problems that have been proposed in the literature by researchers over the last few decades [4, 7-16]. Akgüngör and Doğan [4] have proposed two different model based on artificial neural network (ANN) and genetic algorithm (GA) to estimate the number of accidents, fatalities, and injuries in Ankara, Türkiye with using the historical data records of Türkiye between 1986 and 2005. In another study, Özkış and Tahir [8] have proposed some different models based on particle swarm optimization (PSO) and artificial bee colony algorithm (ABC) methods with linear regression model to solve the estimated number of accidents, injuries and fatalities in Türkiye. In the experiments, they have used the historical data records of Türkiye between 2002 and 2014 which were taken from TUIK.

The harris hawks optimization (HHO) method is one of the newly proposed population-based stochastic optimization methods by Heidari et al. [17] for the solution of continuous optimization problems. HHO method is used for solving different kinds of problems in recent years such as estimation problem [10], feature selection problem [18-20], combined heat and power economic dispatch problem [21], community detection problem [22] and so on. In this study, the original HHO algorithm is used with a linear regression model in order to estimate the number of the accidents, injuries and fatalities in Türkiye for the years (2002-2022), which are taken from TUIK [5]. In addition, the historical data records of Türkiye for the years between 2002 and 2022 will be used in this study for the first time.

The paper is organized as follows: The basic HHO algorithm is explained in Section II. The linear regression model for forecasting the number of the accidents, injuries and fatalities in Türkiye is explained in Section III. The experiments of HHO method are addressed in Section IV. And, the conclusions are presented in Section V.

II. Harris Hawks Optimization (HHO)

HHO algorithm is a population-based optimization algorithm that is inspired by the hunting style of the Harris Hawks and uses a collaborative treatment [17, 23]. This strategy involves hawks attacking from different directions to confuse prey. Depending on the dynamics of the scenario and the prey's escape technique, Harris hawks can elicit a variety of hunting patterns [24]. The HHO algorithm contains phases of exploration and exploitation that are assisted by different strategies based on the surprise attack, the nature of the discovery of a prey, and the attacking phenomenon of the Harris hawk [22]. HHO has a number of important techniques to use in the exploration phase. Furthermore, HHO has a strong strategy for the transition from exploration to exploitation [25]. The algorithmic steps of the HHO are presented in Fig 1.

Procedure HHO (population size $Npop$, maximum iterations $MaxIt$, X)
The initialization of algorithm
Initialize randomly population (X) in range of the search space
Set iteration counter iter=1
Searching process in solution space
while iter < $MaxIt$
Evaluate fitness values of hawks
Set X_{best} as the position of the best hawk
For i=1: $Npop$ (for each X_i)
Update the initial energy(E_0)
And the jump strength J $E_0 = 2rand() - 1$; J = $2(1 - rand)$;
Update the E
if($|E| \geq 1$)
Update the location vector using exploration equation
end
if ($|E| < 1$)
if($|r| \geq 0.5$ and $|E| \geq 0.5$)
Update the location vector using soft besiege
else If ($|r| \geq 0.5$ and $|E| < 0.5$)
Update the location vector using hard besiege
else If ($|r| < 0.5$ and $|E| \geq 0.5$)
Update the location vector using soft besiege with progressive rapid dives
else If ($|r| < 0.5$ and $|E| < 0.5$)
Update the location vector using hard besiege with progressive rapid dives
end if
end if
end for
iter=iter+1;
end while
Testing the termination condition
return X_{best}
end procedure

Fig. 1 The pseudo-code of basic HHO method [17]

III. Linear Regression Model

Population growth and the corresponding increase in the number of vehicles are important factors in road traffic accidents, so these two parameters can be used to develop estimation models. The result of the estimation can help to improve road safety and prevent future problems

[8]. Therefore, in this study, population and number of vehicle are used as input parameters of linear estimation model. The linear regression model used in this study is given in Eq. (1) [8].

$$E_{\text{linear}} = W_1 + W_2.P + W_3.V \quad (1)$$

Where W_1, W_2 and W_3 represent the weight coefficients of the proposes model. W_1 indicates an independent weight coefficient, and W_2 and W_3 are the weight coefficients of population (P) indicator and number of vehicle (V) indicator, respectively. In this study, Eq. (1) is used for forecasting three different output indicators such as number of accidents, number of injuries and number of fatalities. So, there different estimation models are obtained for estimation process. The fitness function between the observed performance and the estimated performance is computed using Eq. (2) [14].

$$\min \quad f(v) = \sum_{h=1}^{H} \left(E_h^{observed} - E_h^{estimated} \right) \quad (2)$$

Where, h shows a historical record for a year and $h = 1, 2, \dots, H$. In here, H represents the total number of years. E_h^{observed} and $E_h^{\text{estimated}}$ show the estimated performance and observed performance of hth year, respectively.

IV. EXPERIMENTS

In experiments, HHO method is implemented on the historical records of Türkiye taken from TUIK for the years (2002-2022) with a linear regression model. In the proposed estimation model, population (P) and number of vehicle (V) are used as input indicators, and number of accidents, number of injuries and number of fatalities are used as output indicators. It should be noted that input parameters are used separately for each estimation model. Thus, three different linear regression models have been proposed for the estimation of the number of accidents, the number of injuries, and the number of fatalities. HHO algorithm was coded in MATLAB platform. Furthermore, the proposed method is implemented on the problem addressed in this study with 30 independent runs for each case. The number of population ($Npop$) and the stopping criteria ($MaxIt$) are selected as 100 and 5×10^3, respectively. Experiments are presented as the best of 30 runs, the number of errors, and the relative error. The algorithmic parameters of the HHO method are selected according to the original study. The historical data records of Türkiye for the years 2002-2022 are presented in Table 1.

Table 1. The historical data records of traffic accidents values and the population size of Türkiye for the years between 2002 and 2022

Year	Output Parameters			Input Parameters	
	Number of accidents (10^5)	Number of injuries (10^5)	Number of fatalities (10^3)	Mid-year Population (10^6)	Number of vehicle (10^6)
2002	4.39777	1.16412	4.093	66.003	8.65517
2003	4.55637	1.18214	3.946	66.795	8.903843
2004	5.37352	1.36437	4.427	67.599	10.236357
2005	6.20789	1.54086	4.505	68.435	11.145826
2006	7.28755	1.6908	4.633	69.295	12.227393
2007	8.25561	1.89057	5.007	70.158	13.022945
2008	9.5012	1.84468	4.236	71.052	13.765395
2009	10.53346	2.0138	4.324	72.039	14.3167
2010	11.06201	2.11496	4.045	73.142	15.095603
2011	12.28928	2.38074	3.835	74.224	16.089528
2012	12.96634	2.68079	3.75	75.176	17.033413
2013	12.07354	2.74829	3.685	76.148	17.939447
2014	11.9901	2.85059	3.524	77.182	18.828721
2015	13.13359	3.04421	7.530	78.218	19.994472
2016	11.82491	3.03812	7.30	79.278	21.090424
2017	12.02716	3.00383	7.427	80.313	22.218945
2018	12.29364	3.07071	6.675	81.407	22.865921
2019	11.68144	2.83234	5.473	82.579	23.156975
2020	9.83808	2.26266	4.866	83.385	24.144857
2021	11.86353	2.74615	5.362	84.147	25.249119
2022	12.32957	2.88696	5.229	84.98	26.482847

Table 2 shows the experimental results of HHO method with linear regression models for the number of accidents, number of injuries and number of fatalities.

Table 2. The experimental results of HHO method for three different estimation models

Estimation Model	Weight Coefficients			Criterion		
	W_1	W_2	W_3	Best Fitness Value	Total Error	Total Relative Error (%)
for estimation the number of accidents	43.03	-0.71	1.191	59.76	30.39	347.49
for estimation the number of injuries	6.456	-0.11	0.219	1.89	5.17	230.23
for estimation the number of fatalities	60.87	-1.03	1.248	19.78	16.70	337.06

The mathematical formulation of the developed linear regression models by HHO algorithm are given in Eq. (3), Eq. (4) and Eq. (5). The models are created with the weight coefficients presented in Table 2.

$$HHO_{\text{for number of accidents}} = 43.03 - 0.71.P + 1.191.V \quad (3)$$

$$HHO_{\text{for number of injuries}} = 6.456 - 0.11.P + 0.219.V \quad (4)$$

$$HHO_{\text{for number of fatalities}} = 60.87 - 1.03.P + 1.248.V \quad (5)$$

Table 3 shows the observed number of accidents and the estimated number of accidents obtained by the HHO method for the years (2002-2022). In addition, the error values, which refer to the difference between the observed number of accidents and the estimated number of accidents, and the relative error results are presented in Table 3. When Table 3 is analyzed, it can be seen that HHO method generally estimates similar number of accidents when compared with observed number of accidents in terms of error and relative error criteria for each year listed in Table 3.

Table 3. The experimental results of HHO method for the number of accidents indicator

Year	Observed number of accidents	Estimated number of accidents	Error	Relative Error (%)
2002	4.39777	6.43905	2.04	46.42
2003	4.55637	6.17241	1.62	35.47
2004	5.37352	7.18810	1.81	33.77
2005	6.20789	7.67720	1.47	23.67
2006	7.28755	8.35422	1.07	14.64
2007	8.25561	8.68846	0.43	5.24
2008	9.50120	8.93743	-0.56	-5.93
2009	10.53346	8.89266	-1.64	-15.58
2010	11.06201	9.03653	-2.03	-18.31
2011	12.28928	9.45141	-2.84	-23.09
2012	12.96634	9.89907	-3.07	-23.66
2013	12.07354	10.28744	-1.79	-14.79
2014	11.99010	10.61179	-1.38	-11.50
2015	13.13359	11.26400	-1.87	-14.24
2016	11.82491	11.81603	-0.01	-0.08
2017	12.02716	12.42461	0.40	3.30
2018	12.29364	12.41775	0.12	1.01
2019	11.68144	11.93156	0.25	2.14
2020	9.83808	12.53537	2.70	27.42
2021	11.86353	13.30905	1.45	12.18
2022	12.32957	14.18648	1.86	15.06

Fig. 2 shows the observed and estimated number of accidents for the years (2002-2022). According to Fig. 2, HHO method is obtained higher number of accidents than obtained number of accidents for the years (2002-2006) and (2020-2022). Moreover, for the years (2007-2008) and (2016-2019), HHO is obtained similar number of accidents to observed number of accidents.

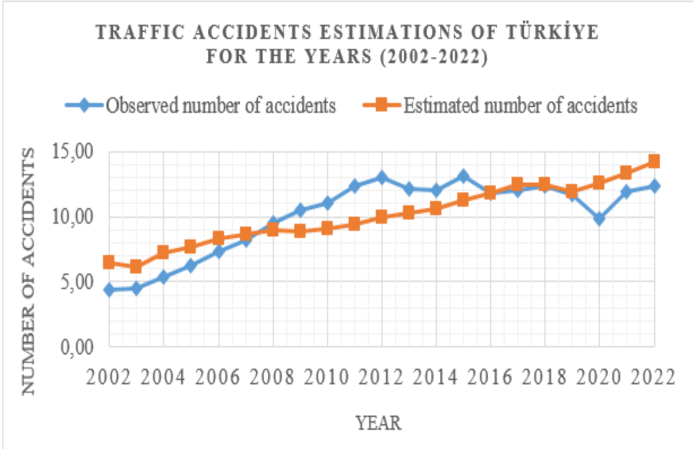

Fig. 2 The observed and estimated number of accidents results of Türkiye for the years 2002-2022

For the years (2002-2022), the observed number of injuries and the estimated number of injuries obtained by the HHO method are shown in Table 4. Furthermore, Table 4 shows the error scores that refer to the difference between the observed and the estimated number of injuries, and the relative error results. Analysis of Table 4 shows that the HHO method generally estimates a similar number of injuries to the observed number of injuries in relation to the error and relative error criteria for each year listed in Table 4.

Table 4. The experimental results of HHO method for the number of injuries indicator

Year	Observed number of injuries	Estimated number of injuries	Error	Relative Error (%)
2002	1.16412	1.40086	0.24	20.34
2003	1.18214	1.37187	0.19	16.05
2004	1.36437	1.57864	0.21	15.70
2005	1.54086	1.68953	0.15	9.65
2006	1.69080	1.83552	0.14	8.56
2007	1.89057	1.91866	0.03	1.49
2008	1.84468	1.98692	0.14	7.71
2009	2.01380	2.00358	-0.01	-0.51
2010	2.11496	2.05781	-0.06	-2.70
2011	2.38074	2.16127	-0.22	-9.22
2012	2.68079	2.26748	-0.41	-15.42
2013	2.74829	2.36330	-0.38	-14.01
2014	2.85059	2.44893	-0.40	-14.09
2015	3.04421	2.59481	-0.45	-14.76
2016	3.03812	2.72290	-0.32	-10.38
2017	3.00383	2.86074	-0.14	-4.76

2018	3.07071	2.88707	-0.18	-5.98
2019	2.83234	2.82734	0.00	-0.18
2020	2.26266	2.95854	0.70	30.75
2021	2.74615	3.11982	0.37	13.61
2022	2.88696	3.30193	0.41	14.37

Fig. 3 presents the observed and estimated number of injuries for the years (2002 - 2022). According to Fig.3, the observed number of injuries for the years (2012-2016) are higher than the number of injuries obtained by HHO method. However, HHO method obtains higher experiment results than observed results for the year between 2020 and 2022. In addition, HHO method obtains similar number of injuries to the observed number of injuries for the years (2002-2010) and (2017-2019).

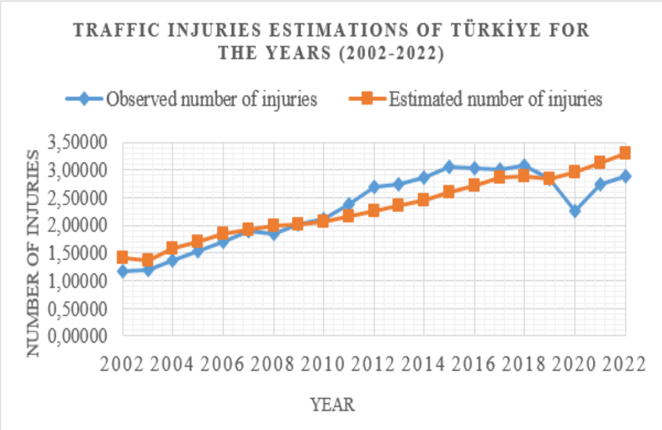

Fig. 3 The observed and estimated number of injuries results of Türkiye for the years 2002-2022

Table 5. The experimental results of HHO method for the number of fatalities indicator

Year	Observed number of fatalities	Estimated number of fatalities	Error	Relative Error (%)
2002	4.093	3.781	-0.31	-7.62
2003	3.946	3.277	-0.67	-16.95
2004	4.427	4.114	-0.31	-7.08
2005	4.505	4.389	-0.12	-2.58
2006	4.633	4.855	0.22	4.78
2007	5.007	4.960	-0.05	-0.94
2008	4.236	4.967	0.73	17.26
2009	4.324	4.640	0.32	7.32
2010	4.045	4.478	0.43	10.71
2011	3.835	4.606	0.77	20.10
2012	3.750	4.805	1.06	28.13
2013	3.685	4.936	1.25	33.96
2014	3.524	4.983	1.46	41.40
2015	7.530	5.373	-2.16	-28.65
2016	7.300	5.650	-1.65	-22.60
2017	7.427	5.995	-1.43	-19.29
2018	6.675	5.677	-1.00	-14.95
2019	5.473	4.835	-0.64	-11.66
2020	4.866	5.239	0.37	7.66
2021	5.362	5.834	0.47	8.80
2022	5.229	6.517	1.29	24.63

The observed number of fatalities and the estimated number of fatalities obtained by the HHO method for the years (2002-2022) are presented in Table 5. When Table 5 is analyzed, it can be seen that the HHO method generally estimates a similar number of fatalities to the observed number of fatalities for each year given in Table 5.

The observed and estimated number of fatalities for the years (2002 - 2022) are presented in Fig. 4. It is seen from Fig.4, HHO method obtains similar number of fatalities to the observed number of fatalities for the years (2002-2010) and (2019-2020). In addition, for the years (2011-2014) and (2021-2022), the estimated number of fatalities obtained by HHO method are higher than the observed number of fatalities. However, the HHO method produces fewer number of fatalities than observed number of fatalities for the year between 2015 and 2018.

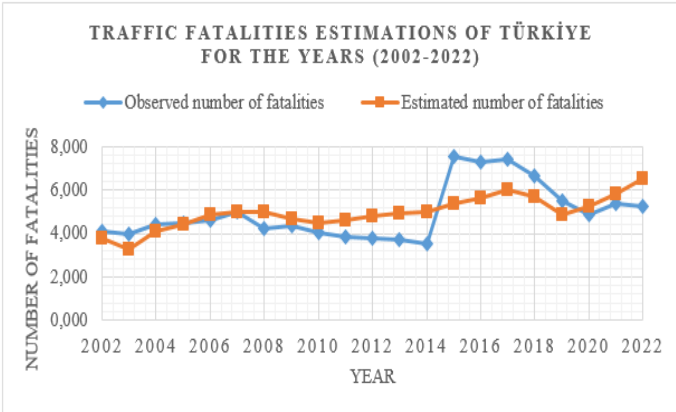

Fig. 4 The observed and estimated number of fatalities results of Türkiye for the years 2002-2022

V. CONCLUSION

Road traffic accidents have a wide range of serious consequences for the individuals involved. For this reason, the assessment of the number of accidents, injuries and fatalities is one of the major issues for individuals and societies. Therefore, this study focuses on using the Harris hawks optimization (HHO) algorithm and a linear regression model to create some appropriate estimation models to estimate the number of accidents, injuries, and fatalities in Türkiye. In the experiments, Türkiye's observed number of accidents, injuries and fatalities, population and number of vehicles historical records taken from Turkish statistical institute (TUIK) for the years between 2002 and 2022 have been used. The estimations models are created with HHO method and linear regression model for each output parameter (the number of accidents, the number injuries, and the number fatalities). For estimation the number of accidents, the HHO method based model obtains the best fitness value, total error value and total relative error value as 59.76, 30.39 and 347.49, respectively. For estimation the number of injuries, HHO method obtains the best fitness value, total error value and total relative error value as 1.89, 5.17 and 230.23, respectively. And lastly, for estimation the number of fatalities, HHO method obtains the best fitness value, total error value and total relative error value as 19.78, 16.70 and 337.06, respectively. According to the experimental results, it can be seen that the proposed method is obtained competitive and effective experimental results for estimating the number of accidents, the number injuries, and the number fatalities for Türkiye.

REFERENCES

[1] P. Ong and H.-G. Sung, Exploratory Study of Spatial Variation in Car Insurance Premiums, Traffic Volume and Vehice Accidents, (2003).

[2] World Health Organization, Global Status Report on Road Safety: Supporting A Decade of Action, in: World Health Organization., https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries, Geneva, . 2023.

[3] P. Margaret, R. Scurfield, D. Sleet, D. Mohan, A.A. Hyder, E. Jarawan and C. Mathers, World Report on Road Traffic Injury Prevention, in, https://www.who.int/publications/i/item/world-report-on-road-traffic-injury-prevention, 2004.

[4] A.P. Akgüngör and E. Doğan, An artificial intelligent approach to traffic accident estimation: Model development and application, Transport, 24 (2009) 135-142.

[5] National Statistics, http://www.tuik.gov.tr, in, 2022.

[6] O. ÖZTÜRK and E. Cenker, Motorlu taşıt satışlarının trafik kazaları üzerine olan etkileri, SDÜ Tıp Fakültesi Dergisi, 13 (2006) 12-15.

[7] O. Ercan, G. Erhan and K. Zubeyde, Estimation of the costs of traffic accidents in Turkey: An evaluation in terms of the insurance and financial system, Yaşar Üniversitesi E-Dergisi, 9 (2014) 5649-5673.

[8] A. Özkış and T. Sağ, Model development for estimation of traffic accidents with metaheuristic algorithms, in: International Conference on Engineering Technologies (ICENTE'18), Konya. , 2018.

[9] D. Özdemir, S. Dörterler and D. Aydın, A new modified artificial bee colony algorithm for energy demand forecasting problem, Neural Computing and Applications, 34 (2022) 17455-17471.

[10] M.F. Tefek and M. Arslan, Türkiye'de Trafik Sigorta Primlerinin Harris Şahinleri Algoritması ile Tahmini, Avrupa Bilim ve Teknoloji Dergisi, (2022) 711-715.

[11] M.F. Tefek and M. Arslan, Highway accident number estimation in Turkey with Jaya algorithm, Neural Computing and Applications, 34 (2022) 5367-5381.

[12] M. Aslan, Archimedes optimization algorithm based approaches for solving energy demand estimation problem: a case study of Turkey, Neural Computing and Applications, 35 (2023) 19627-19649.

[13] M.S. Kıran, E. Özceylan, M. Gündüz and T. Paksoy, A novel hybrid approach based on particle swarm optimization and ant colony algorithm to forecast energy demand of Turkey, Energy conversion and management, 53 (2012) 75-83.

[14] M. Aslan and M. Beşkirli, Realization of Turkey's energy demand forecast with the improved arithmetic optimization algorithm, Energy Reports, 8 (2022) 18-32.

[15] T.O. Blinova, Analysis of possibility of using neural network to forecast passenger traffic flows in Russia, Aviation, 11 (2007) 28-34.

[16] Y.S. Murat and H. Ceylan, Use of artificial neural networks for transport energy demand modeling, Energy policy, 34 (2006) 3165-3172.

[17] A.A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja and H. Chen, Harris hawks optimization: Algorithm and applications, Future generation computer systems, 97 (2019) 849-872.

[18] R. Al-Wajih, S.J. Abdulkadir, N. Aziz, Q. Al-Tashi and N. Talpur, Hybrid binary grey wolf with Harris hawks optimizer for feature selection, IEEE Access, 9 (2021) 31662-31677.

[19] L. Peng, Z. Cai, A.A. Heidari, L. Zhang and H. Chen, Hierarchical Harris hawks optimizer for feature selection, Journal of Advanced Research, (2023).

[20] X. Wang, X. Dong, Y. Zhang and H. Chen, Crisscross Harris hawks optimizer for global tasks and feature selection, Journal of Bionic Engineering, 20 (2023) 1153-1174.

[21] M. Ramachandran, S. Mirjalili, M. Nazari-Heris, D.S. Parvathysankar, A. Sundaram and C.A.R.C. Gnanakkan, A hybrid grasshopper optimization algorithm and harris hawks optimizer for combined heat and power economic dispatch problem, Engineering Applications of Artificial Intelligence, 111 (2022) 104753.

[22] I. Koc, A fast community detection algorithm based on coot bird metaheuristic optimizer in social networks, Engineering Applications of Artificial Intelligence, 114 (2022) 105202.

[23] H. Moayedi, A. Osouli, H. Nguyen and A.S.A. Rashid, A novel Harris hawks' optimization and k-fold cross-

validation predicting slope stability, Engineering with Computers, 37 (2021) 369-379.
[24] E.H. Houssein, M.E. Hosney, D. Oliva, W.M. Mohamed and M. Hassaballah, A novel hybrid Harris hawks optimization and support vector machines for drug design and discovery, Computers & Chemical Engineering, 133 (2020) 106656.
[25] V.K. Kamboj, A. Nandi, A. Bhadoria and S. Sehgal, An intensify Harris Hawks optimizer for numerical and engineering optimization problems, Applied Soft Computing, 89 (2020) 106018.