Optimize Machine Learning Programming Algorithms for Sentiment Analysis in Social Media

Manal S.F. Alharbi
Computer Department,
The Higher Institute of Telecommunication and
Navigation Public Authority for Applied Education
and Training (PAAET), Kuwait

El-Sayed M. El-kenawy
Department of Communications and Electronics
Delta Higher Institute of Engineering and
Technology (DHIET)
Mansoura, 35111, Egypt

ABSTRACT
Sentiment analysis (SA) is used to evaluate the polarity of reviews. In SA, the feature selection phase is essential, as the best way to solve all optimization problems not happen yet so this paper proposes a hybrid approach that combines hybrid algorithms [(GWO and (PSO)], its name (GWOPS). To reduce the search space filter features selection. (GWOPS) used for training neural networks (NN) classifiers to select the best features. Data was collected from Twitter on different topics used in our experimental. The proposed algorithm is compared with three well-known optimization algorithms the experiments and comparisons result to evaluate the quality and effectiveness of the (GWOPS)

Keywords
Sentiment analysis, Twitter, Social Media, GWO, PSO

1. INTRODUCTION
Opinion mining or emotional intelligence (EI) is known as Sentiment Analysis [1]. Sentimental analysis can be described as the art of gathering useful perspectives from various social media and online sources, such as social network chats like Twitter, WhatsApp and Facebook, from unstructured and unorganized textual material Online articles and remarks. Opinion mining entails developing rules-based automated systems that interpret data based on a set of predefined rules. Mechanical systems are also generated to perform opinion mining using some of the principles of machine learning. [2]. To construct a hybrid sentimental analysis model, there are several circumstances in which both rule-based and advanced machine learning algorithms are mixed. The sentimental analysis may also be an important way of studying the textual content's positive or harmful purpose. Sentimental analysis has performed as a reliable outlet for offering insightful perspectives on many market-rolling items, fresh innovations, citizens' thoughts on current government-framed policies, etc. Around the same time, sentiment analysis also plays a crucial role in interpreting a single shop's buying experience based on user input posted through the social media platform. [3-5]. The sentimental analysis thus ignores the value of understanding the customer's viewpoint in terms of the shopping experience obtained when buying a product in a specific shop. [6].

Fig 1: Emotional Rating

1.1 Social Media Importance
Global networking and networks may have swept the planet by surprise, and the world's facets have been reduced. Via numerous social media networks such as Twitter, Facebook, WhatsApp, etc., there is a fair bit of communication between people [7]. These conversations would be casual in most situations; the mood and sentiment of the people participating in the debate are reflected in these discussions [8]. This paves the way to consider the behavioral patterns of the individuals participating in the debate. In the study and interpretation of transcript exchange, Sentimental research, also referred to as opinion mining and natural language processing (NLP), plays a key role [9].
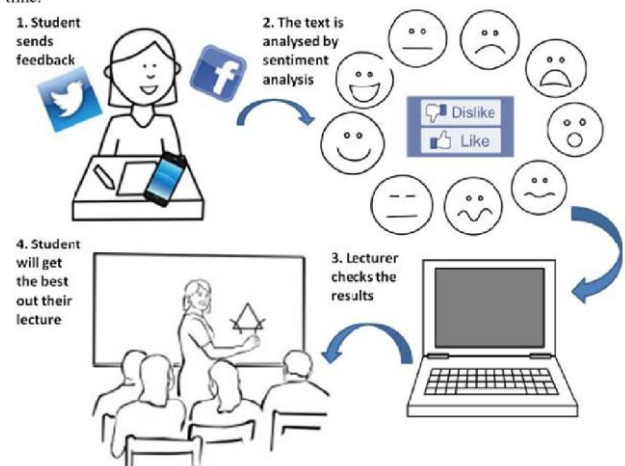

Fig 2: Social Media Networks

1.2 The Feature Selection Problem
Methods of feature selection help you build a useful predictive model [10] in your task. They assist you by selecting features that provide you with good or better accuracy while needing fewer details. Feature filtering

strategies can be used to recognize and delete obsolete and unnecessary data attributes that do not add to the accuracy of a predictive model which can theoretically decrease the model's accuracy. Fewer features are beneficial since the model's complexity is reduced, and it is easier to understand and describe a simpler model [11-13].

Fig 3: Artificial Intelligence

1.3 Machin Learning

Machine learning (ML) is a type of artificial intelligence (AI) that helps software applications to become more effective in predicting outcomes without being directly programmed to do so [14]. Machine learning algorithms use historical data as feedback to simulate new output values. A recommendation engine [15] is an everyday use case for machine learning. Other popular uses include detecting fraud, spam filtering, malware attack detection, business process automation (BPA) and predictive maintenance [16].

1.4 Optimization Algorithm

Today, real-world optimization concerns and higher dimensional search space, are currently sophisticated, so they remain in simple daunting. In several fields, such as architecture, heuristic marketing methods have been used. [17], artificial intelligence [18], company methods [19], mechanics [Twenty], business economics [21], booking [22], transportation [23], incorporated decision-creating [24], and also solutions estimate [25], [26] Marketing pertains to discovering an acceptable ideal solution for a particular problem one of a lot of feasible ones. Marketing activity is usually changed into a hunting problem in a multi-dimensional area. Virtually, the search aims to mitigate or maximize an objective feature that assesses an applicant's quality for a solution that is usually defined by a search region angle. Meta-heuristics are loved by comparative optimization approaches that offer realistic alternatives in a reasonable period. [27]

1.5 Paper Contribution

This paper proposes a hybrid approach that combines modified hybrid algorithms [(GWO) (PSO)], its name (GWOPS). To reduce the search space filter feature selection has been used. (GWOPS) used for training neural networks (NN) classifier to select the best features

2. RELATED WORK

Along with the social networking sites systems including Twitter, Facebook, Instagram and also WhatsApp taking the communication globe through a hurricane, it has come to be crucial that the records living all over these social networking sites systems are going to share information relevant information concerning the point of view, mood, as well as conviction of people over any kind of item, tip or even plans

Previously, a variety of works, have been carried out to analyze the contents of Twitter and also conduct perspective exploration over twitter info. The writers of [28] Using the deep convolution semantic network to analyze the Twitter feed has been proposed. The role collection was included for training right into based CNN and even projected emotions by analyzing Twitter details. When it concerns Twitter data assessment, the preprocessing of tweets has an outstanding obligation to play. Numerous preprocessing approaches to text messages have been tested and compared as a feature of [28] Because the feeling study involves studying and translating various sorts of convictions, multi training class sentiment evaluation is critical and is reviewed in [30] Authors of [31] have established a resource called SENTA that utilizes a design contrast strategy to conduct multiclass view analysis. As gone over in [32], it is actually to comprehend the evolving changes of point of view exploration, the situation, and the powerful celebrations in the Twitter talks. The use of Twitter has sparked even more research work to understand the emotions using Twitter records. One such piece discussed in [Thirty-three] uses a crossbreed platform to conduct view analysis using an inherited algorithm-based approach. The emphasis of the work was actually to enhance the system coming from a scalability point of view. Writers of [34] have looked into social sentiment variations over a given subject matter using a mathematical style to identify the foreground subject matters and advertise successful candidates' successful position. An attractive task relevant to thoughtful analysis is to describe convictions depending on the subject of debate. Writers of [35] have actually delivered a successful topic flexible belief classification over tweets from the above jobs. It can be simply understood that the components of Twitter records provide helpful knowledge about any subject under discussion and impart people's opinion over those subject matters. The main perk and the exceptional performance of machine learning formulas are removing the given problem's extract. It offers full info regarding the data and also produces the designer to think about selecting the appropriate protocol which can be utilized for the concern. Clustering, sorting, regression and rule extraction are a few common groups of artificial intelligence problems. [36]

3. PROPOSED MODEL

3.1 Dataset

Data was gathered from Twitter on various themes, for example, expressions and legislative issues. This twitter corpus contains 2000 tweets audits with 1000 positive tweets and 1000 negative tweets.

Preprocessing Dataset

- Insert Dataset
- Remove null values
- Check data category
- Training data and test data (Splitting)
- Scaling of feature

3.2 PSO

The PSO algorithm copies the insights of real-life bird swarms. Velocity is the location of a particle's change of position. The location particle is transformed during the time. The particle's velocity is arbitrarily accelerated toward its last best place and the best solution for the neighborhood at flight. As shown in the next equations

$$v_i^{k+1} = v_i^k + c_1 r_1 \left(Pbest_i^k + x_i^k \right) + c_2 r_2 \left(g_{best} - x_i^k \right)$$

$$x_i^{k+1} = x_i^k + v_i^{k+1}$$

3.3 GWO

Grey wolves are one of the zenith predators; pinnacle killers go to the highest point of the all-natural pecking order. In the real online Grey wolves typically take better in groups. The standard of team numbers may be actually from 5-- 12. The most fantastic aspect of grey wolves is that they have a strict social driving hierarchy. The innovators got in touch with alphas, are actually a male and also a girl. The alpha is accountable for deciding on meaningful choices for the group like searching, relaxing, etc. The entire group needs to obey the alpha's commands [37]

3.4 GWOPS


```
Initialize the grey wolf population Xi (i = 1, 2, ..., n)
Calculate α, β, and δ
Calculate a as per Eq (3.2)
for each search agent
Update A, and C
Calculate Xα, Xβ, and Xδ
crossover ( Xα , Xβ , Xδ)
        solution (t+1) = Mutate
end for
While (t < Max_iter)
if  t % 2 == 0:
   Calculate fitness(3)
Else
Calculate fitness using Eq (4)
Update A, a, C
Evaluate all particles using the objective function
Update the positions of the three best agents α;β;δ
t=t+1

return α
```

Fig 4: Positive and Negative Keywords

3.5 NN

Neural networks are a set of formulas, created freely after the human brain, that is actually developed to realize designs. They analyze biological data via a kind of maker understanding, labelling or concentration raw input. The techniques they learn are even mathematical, consisting of in vectors. All real-world data, be it pictures, audio, content or time series, have to be actually converted. Neural networks help our company flock and identify [38-41] You can quickly think about all of them as clustering and category level atop the records you save and handle. They help to team unlabeled records according to correlations one of the instance inputs, and also they identify data

The output ct can be calculated by

$$Q(c_t|\{a_i\}_{i \neq t}) = \sigma(W_c^f h_t^f + W_c^b h_t^b + b_c)$$

Where

$$h_t^f = tanh(W_h^f h_{t-1}^f + W_x^f x_t + b_h^f)$$
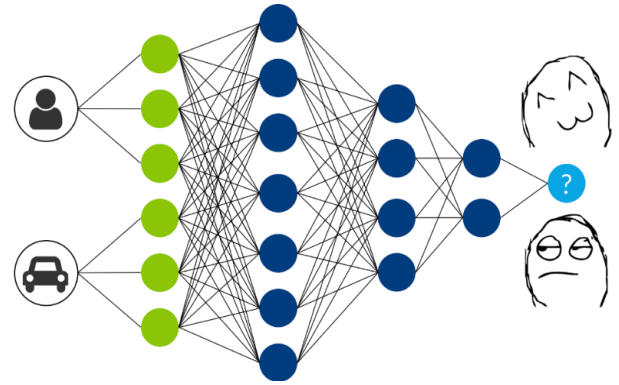
$$h_t^b = tanh(W_h^b h_{t+1}^b + W_a^b a_t + b_h^b)$$

Fig 5: Neural networks

The output is:

$$c = f(a) = \sum_{j=1}^{n} a_j w_j$$

Where n represents the number of inputs. Generally, by doing this array multiplication, the weighted sum can be calculated. Bias is an extra variable that can be used along with the weighted sum of the inputs to the neuron to change the output. A neuron's final performance is:

$$c = f(a) = \sum_{j=1}^{n} a_j w_j + b$$

4. RESULTS

This procedure is rehashed multiple times. Ultimately, the average accuracy, average fitness, and average number of selected features over ten runs are accounted for. The GWOPS calculation is contrasted, and others surely understood calculations counting PSO, GA, and WOA. Besides, in this work, NN classifier is applied to the various investigations as exhibited in the accompanying

Table 1: GWOPS Accuracy

Number of input features		GWOPS%	PSO%	GA%	WOA%
150		93.89	84.54	85.35	88.84
290		92.79	86.61	86.81	87.78
400		91.64	87.12	87.38	86.41

Table 2: Descriptive statistics of GWOPS Accuracy

	GWOPS %	PSO%	GA%	WOA %
Minimum	93.19	82.54	81.35	83.84
25% Percentile	93.89	84.54	85.35	88.84
Median	93.89	84.54	85.35	88.84
75% Percentile	93.89	84.54	85.35	88.84
Maximum	93.99	85.54	86.35	90.84
Range	0.8	3	5	7
Mean	93.85	84.47	85.14	88.63
Std. Deviation	0.191	0.6157	1.122	1.477
Std. Error of Mean	0.05105	0.1646	0.2998	0.3947

Table 3: ANOVA

ANOVA table	SS	DF	MS	F (DFn, DFd)	P value	
Treatment (between columns)	773.6	3	257.9	F (3, 52) = 267.5	P<0.0001	
Residual (within columns)	50.12	52	0.9638			
Total	823.7	55				

Table 4: One sample t test

	GWOPS%	PSO%	GA%	WOA%
Theoretical mean	0	0	0	0
Actual mean	93.85	84.47	85.14	88.63
One sample t test				
t, df	t=1838, df=13	t=513.3, df=13	t=284.0, df=13	t=224.5, df=13
P value (two tailed)	<0.0001	<0.0001	<0.0001	<0.0001
P value summary	****	****	****	****
Significant (alpha=0.05)?	Yes	Yes	Yes	Yes
How big is the discrepancy?				
Discrepancy	93.85	84.47	85.14	88.63
SD of discrepancy	0.191	0.6157	1.122	1.477
SEM of discrepancy	0.05105	0.1646	0.2998	0.3947
95% confidence interval	93.74 to 93.96	84.11 to 84.82	84.49 to 85.78	87.77 to 89.48
R squared (partial eta squared)	1	1	0.9998	0.9997
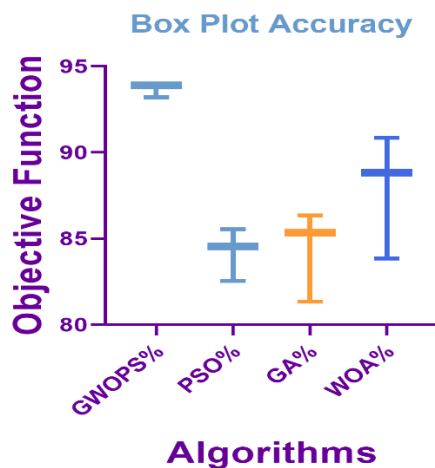
Fig 6: Box Plot Accuracy

5. CONCLUSION AND FUTURE WORK

To add to its cutting edge, this paper concentrated on SA. This is aimed at improving the regular GWOPS algorithm by enhancing the implementation of GWO and enhancing the capacity of the nearby pursuit like transformation and training (NN) classifier to pick the best features to evaluate and rate it. In correlation with WOA, PSO, and GA, we used and assessed the GWOPS, the predominance of GWOPS accuracy results in analysis with various algorithms as shown by robust text design.

6. REFERENCES

[1] Tubishat, M., Abushariah, M.A.M., Idris, N. et al. Improved whale optimization algorithm for feature selection in Arabic sentiment analysis. Appl Intell 49, 1688–1707 (2019) doi:10.1007/s10489-018-1334-8

[2] Bahassine S, Madani A, Kissi M (2016) An improved Chi-sqaure feature selection for Arabic text classification using decision Tree In Intelligent Systems: Theories and Applications (SITA), 2016 th International Conference on (pp. 1–5). IEEE

[3] Cheng, Y., Zhao, S., Cheng, B., Hou, S., Shi, Y., & Chen, J. (2018). Modeling and Optimization for Collaborative Business Process towards IoT Applications. Mobile Information Systems, 2018.

[4] Nouiri, M., Bekrar, A., Jemai, A., Niar, S., & Ammari, A. C. (2018). An effective and distributed particle swarm optimization algorithm for flexible job-shop scheduling problem. Journal of Intelligent Manufacturing, 29(3), 603-615.

[5] Lin, W. Y. (2016). A novel 3D fruit fly optimization algorithm and its applications in economics. Neural Computing and Applications, 27(5), 1391-1413.

[6] Hussain, K., Salleh, M. N. M., Cheng, S., & Shi, Y. (2018). Metaheuristic research: a comprehensive survey. Artificial Intelligence Review, 1-43.

[7] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," Advances in Engineering Software, vol. 69, no. 0, pp. 46 – 61, 2014.

[8] Ibrahim, Abdelhameed, and El-Sayed M. El-kenawy. "Image Segmentation Methods Based on Superpixel Techniques: A Survey." Journal of Computer Science and Information Systems 15, no. 3 (2020).

[9] Ibrahim, Abdelhameed, and El-Sayed M. El-kenawy. "Applications and Datasets for Superpixel Techniques: A Survey." Journal of Computer Science and Information Systems 15, no. 3 (2020).

[10] Eid, Marwa M., El-Sayed M. El-kenawy, and Abdelhameed Ibrahim. "An Advanced Patient Health Monitoring System." Journal of Computer Science and Information Systems 17, no. 11 (2021).

[11] El-kenawy, El-Sayed M., Marwa M. Eid, and Abdelhameed Ibrahim. "Anemia Estimation for COVID-19 Patients Using A Machine Learning Model." Journal of Computer Science and Information Systems 17, no. 11 (2021).

[12] El-kenawy, El-Sayed M., Marwa M. Eid, and Abdelhameed Ibrahim. "Automatic Identification from Noisy Microscopic Images." Journal of Computer Science and Information Systems 17, no. 11 (2021).

[13] M. M. Fouad, A. I. El-Desouky, R. Al-Hajj and E. M. El-Kenawy, "Dynamic Group-based Cooperative Optimization Algorithm," in IEEE Access, doi: 10.1109/ACCESS.2020.3015892.

[14] El-Kenawy, E. S. M. T., El-Desoky, A. I., & Sarhan, A. M. (2014). A bidder strategy system for online auctions trust measurement. International Journal of Strategic Information Technology and Applications (IJSITA), 5(3), 37-47.

[15] Hassib, E. M., El-Desouky, A. I., Labib, L. M., & El-kenawy, E. S. M. (2019). WOA+ BRNN: An imbalanced big data classification framework using Whale optimization and deep neural network. Soft Computing, 1-20.

[16] El-kenawy, E. S. M., El-Desoky, A. I., & Al-rahamawy, M. F. (2012). Distributing Graphic Rendering using Grid Computing with Load Balancing. International Journal of Computer Applications, 975, 888.

[17] El-kenawy, E. S. M. T. (2019). A Machine Learning Model for Hemoglobin Estimation and Anemia Classification. International Journal of Computer Science and Information Security (IJCSIS), 17(2).

[18] Hassib, E. M., El-Desouky, A. I., El-kenawy, E. S. M., & Elghamrawy, S. (2019). An Imbalanced Big Data Mining Framework for Improving Optimization Algorithms Performance. IEEE Access.

[19] H. Hassan, A. I. El-Desouky, A. Ibrahim, E. M. El-kenawy and R. Arnous, (2020) "Enhanced QoS-based Model for Trust Assessment in Cloud Computing Environment," in IEEE Access. doi: 10.1109/ACCESS.2020.2978452

[20] El-Knawy, E. S. M. T., & El-Desoky, A. I. (2016). TRUST MEASUREMENT FOR ONLINE AUCTIONS: PROPOSAL OF NEW MODEL. INTERNATIONAL JOURNAL OF INNOVATIVE COMPUTING INFORMATION AND CONTROL, 12(2), 385-394.

[21] El-sayed, M., El-Desoky, A. I., & Sarhan, A. M. (2014). A bidder behavior learning intelligent system for trust measurement. International Journal of Computer Applications, 89(8).

[22] Arnous, R., & El-Sayed Towfek, M. (2019). El-kenawy and M Saber. A Proposed Routing Protocol for Mobile Ad Hoc Networks. International Journal of Computer Applications, 178(41), 26-30.

[23] El-Sayed Towfek M El-kenawy. Trust Model for Dependable File Exchange in Cloud Computing. International Journal of Computer Applications 180(49):22-27, June 2018.

[24] El-Sayed Towfek M El-kenawy, M Saber and Reham Arnous. An Integrated Framework to Ensure Information Security Over the Internet. International Journal of Computer Applications 178(29):13-15, July 2019

[25] El-kenawy, E. S. M. T. (2018). Solar Radiation Machine Learning Production Depend on Training Neural Networks with Ant Colony Optimization Algorithms. IJARCCE, 7(5). doi: DOI10.17148/IJARCCE.2018.751

[26] E. M. El-Kenawy, M. M. Eid, M. Saber and A. Ibrahim, "MbGWO-SFS: Modified Binary Grey Wolf Optimizer Based on Stochastic Fractal Search for Feature Selection," in IEEE Access, vol. 8, pp. 107635-107649, 2020, doi: 10.1109/ACCESS.2020.3001151.

[27] E.-S. El-Kenawy and M. Eid, "Hybrid gray wolf and particle swarm optimization for feature selection," INTERNATIONAL JOURNAL OF INNOVATIVE COMPUTING INFORMATION AND CONTROL, vol. 16, no. 3, pp. 831–844, 2020.

[28] El-Kenawy, E. S. M., Eid, M., & Ismail, A. H. A New Model for Measuring Customer Utility Trust in Online Auctions. International Journal of Computer Applications, 975, 8887.

[29] E. M. El-Kenawy and M. Saber , "Design and implementation of accurate frequency estimator depend on deep learning" International Journa l of Engineering &Technology (IJET), vol. 9 , Issue 2, PP. 367-377 , 2020 , DOI: 10.14419/ijet.v9i2.30473

[30] El-kenawy, E. S. T., El-Desoky, A. I., & Al-rahamawy, M. F. (2012). Extended max-min scheduling using petri net and load balancing. Int. J. Soft Comput. Eng.(IJSCE), 2(4), 198-203.

[31] Hussien, Hussien Rezk, and M. El-Sayed. "El-Kenawy, and Ali I. El-Desouky." EEG Channel Selection Using A Modified Grey Wolf Optimizer."." European Journal of Electrical Engineering and Computer Science 5: 17-24.

[32] Ibrahim, Abdelhameed, and A. Tharwat. "Biometric authentication methods based on ear and finger knuckle images." International Journal of Computer Science Issues (IJCSI) 11, no. 3 (2014): 134.

[33] Ibrahim, Abdelhameed, Tarek Gaber, Takahiko Horiuchi, Vaclav Snasel, and Aboul Ella Hassanien. "Human thermal face extraction based on superpixel technique." In The 1st International Conference on Advanced Intelligent System and Informatics (AISI2015),

November 28-30, 2015, Beni Suef, Egypt, pp. 163-172. Springer, Cham, 2016.

[34] Ibrahim, Abdelhameed, Mostafa Noshy, Hesham Arafat Ali, and Mahmoud Badawy. "PAPSO: A power-aware VM placement technique based on particle swarm optimization." IEEE Access 8 (2020): 81747-81764.

[35] Ibrahim, Abdelhameed, Takahiko Horiuchi, and Shoji Tominaga. "Illumination-invariant representation for natural color images and its application." In 2012 IEEE Southwest Symposium on Image Analysis and Interpretation, pp. 157-160. IEEE, 2012.

[36] Saber, Mohamed, and Marwa M. Eid. "Low power pseudo-random number generator based on lemniscate chaotic map." International Journal of Electrical & Computer Engineering (2088-8708) 11, no. 1 (2021).

[37] Ibrahim, Abdelhameed, Hesham Arafat Ali, Marwa M. Eid, and El-Sayed M. El-kenawy. "Chaotic Harris Hawks Optimization for Unconstrained Function Optimization." In 2020 16th International Computer Engineering Conference (ICENCO), pp. 153-158. IEEE, 2020.

[38] El-kenawy, El-Sayed M., Seyedali Mirjalili, Abdelhameed Ibrahim, Mohammed Alrahmawy, M. El-Said, Rokaia M. Zaki, and Marwa M. Eid. "Advanced Meta-heuristics, Convolutional Neural Networks, and Feature Selectors for Efficient COVID-19 X-ray Chest Image Classification." IEEE Access (2021).

[39] Moudani, Walid, Grace Zaarour, El-Sayed M. Towfek El-Kenawy, Ali Ibraheem El-Desoky, Amany M. Sarhan, Mona FM Mursi, and Hossam Eldin H. Ahmed. "International Journal of Strategic Information Technology and Applications."

[40] Eid, Marwa M., El-Sayed M. El-kenawy, and Abdelhameed Ibrahim. "A Fast Real-Time Video Encryption/Decryption Technique Based on Hybrid Chaotic Maps." Journal of Computer Science and Information Systems 18, no. 9 (2021).

[41] Eid, Marwa M., El-Sayed M. El-kenawy, and Abdelhameed Ibrahim. "A New Hybrid Video Encryption Technique Based on Chaos Cryptography."