

EM_GA-RS: Expectation Maximization and GA-based Movie Recommender System

Asha K N¹, R Rajkumar^{2*}

¹School of Computing Science & Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India.

²School of Computing Science & Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India.
cuashin@gmail.com, *vitrajkumar@gmail.com

Abstract

This work introduced a novel approach for the movie recommender system using a machine learning approach. This work introduces a clustering-based approach to introduce a recommender system (RS). The conventional clustering approaches suffer from the clustering error issue, which leads to degraded performance. Hence, to overcome this issue, we developed an expectation-maximization-based clustering approach. However, due to imbalanced data, the performance of RS is degraded due to multicollinearity issues. Hence, we incorporate PCA (Principal Component Analysis) based dimensionality reduction model to improve the performance. Finally, we aim to reduce the error; thus, a Genetic Algorithm (GA) is included to achieve the optimal clusters and assign the suitable recommendation. The experimental study is carried out on publically available movie datasets performance of the proposed approach is measured in terms of MSE (Mean Squared Error) and Root Mean Squared Error (RMSE). The comparative study shows that the proposed approach achieves better performance when compared with a state-of-art movie recommendation system.

Keywords: Recommender system, machine learning, PCA, GA, IoT, Signal Strength, deep reinforcement learning, digital technology, Segmentation, Clustering.

Received on 23 February 2022, accepted on 12 July 2022, published on 13 July 2022

Copyright © 2022 Asha K N *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [Creative Commons Attribution license](#), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi: 10.4108/eetsis.vi.1947

²Corresponding author. Email: vitrajkumar@gmail.com

1. Introduction

Nowadays, we live in a society where we come across several types of information in daily life scenarios. Moreover, the demand for Internet-based applications is increasing rapidly, such as online books, movies, images, web pages, etc. Due to this increased information, accessing the relevant information becomes a tedious task that can be addressed by developing a support mechanism for decision making. The decision-making process helps provide the relevant information according to the user's interest.

Based on this concept, recommender systems (RS) have attracted the research community and various e-commerce and Internet-based businesses [1]. The recommender system is an automated system that collects the information from the user and provides the decision to

access the relevant information. Generally, the information is collected explicitly (based on the user's rating for different types of items or products) or implicitly (based on the user behaviour during interaction with RS) [2]. The RS ideas are widely adopted to overcome the retrieval challenge where users insert the query, and the retrieval systems provide relevant information. These systems are widely adopted in e-commerce [3, 8], e-learning [4], book recommendation [5], and movie recommendation [6] in real-time applications.

The online storage of multimedia data increases, such as YouTube, Netflix, etc., where recommending a suitable video or movie becomes a challenging task. Several techniques have been introduced to develop an efficient and robust system for movie recommendation. These schemes include collaborative filtering [6], content-based

filtering [7], demographic [9], knowledge-based [10] and hybrid filtering [11]. Recently, collaborative filtering schemes have gained huge attraction for recommendation systems. These methods use the user's historical interaction and provide the recommendation for the desired application. However, these techniques fail to achieve the desired performance for sparse interaction, such as on online shopping platforms.

Moreover, these techniques do not recommend a product that is not rated previously [12]. Content-based recommendation system works on the ratings of the user in the past. Machine learning, optimization techniques, and evolutionary computations have been widely adopted for recommendation systems. Such as Hassan et al. [13] introduced a neural network-based approach for improving the accuracy of the recommender system, Kuang et al. [14] introduced a fuzzy rule model for e-commerce RS, and Kothari et al. [15] introduced a support vector machine (SVM) recommender system. Moreover, various optimization schemes are also incorporated to improve the RS performance.

Katarya et al. [16] improved collaborative filtering (CF) by incorporating particle swarm optimization (PSO), in [17] CF is improved using cuckoo search optimization, in [18] grey wolf optimization, in [19] genetic algorithm based CF, and in [20] evolutionary scheme for collaborative filtering are developed. However, the existing schemes suffer from various challenges such as cold-start problems, accuracy, scalability, privacy, etc. In order to overcome these issues, hybrid filtering schemes have been introduced recently. As mentioned earlier, these filterings utilize the filtering and devise a new filtering scheme. Paradarami et al. [21] introduced a hybrid recommendation system using a neural network scheme.

Similarly, various hybrid recommender systems are introduced, such as Hyper [22], hybrid RS based on context feature relationship [23] and fuzzy logic systems [24], etc. these recommender systems have several advantages over conventional schemes. However, computational complexity and accuracy remain challenging tasks in this field. Hence, in this work, we focus on developing a novel hybrid approach for a movie recommendation system. The main contributions of the work are as follows:

- (a) We introduce an expectation-maximization clustering scheme for data clustering.
- (b) We introduced PCA based regression model to reduce the multicollinearity
- (c) A genetic algorithm-based optimization scheme is introduced to reduce the clustering error and improve prediction.

The rest of the article is organized as follows: In section II, we describe the proposed system; in section III, we introduce the experimental analysis and comparative

performance. Finally, in section V we introduce the proposed system's concluding remarks and future work.

2. Proposed Model

This section introduces the proposed solution for a hybrid filtering approach for the movie recommendation system. First of all, we define a problem with the movie recommendation system. Later, we introduce the Expectation maximization clustering scheme to obtain the clusters. In the next phase, we find the nearest neighbour clusters to estimate the expected maximization.

2.1. Problem formulation

Let us consider that total N_u users have rated N_m movies as $r_{1:N_u}$. This matrix is considered a training matrix, which is denoted as a set of triples as $(user, movie, rating)$. The movie rating matrix can be given as follows:

$$\mathcal{A}_{mu} = \begin{cases} R_{m,u}, & \text{rating of user } u \text{ for movie } m \\ ? & \text{if no rating is provided} \end{cases} \quad (1)$$

The user movie rating matrix is denoted as $\mathbb{A} \in \mathcal{R}^{N_m \times N_u}$. In this work, our main aim is to predict the user's rating, recommend the new movies from the list based on these predictions, and predict the rating to replace the "?" from the dataset. The movie ratings are given as 1 to 5 integer values. We focus on reducing the RMSE (Root Mean Square) for the given test set during prediction. The RMSE can be expressed as:

$$RMSE = \sqrt{\frac{1}{|S_{test}|} \sum_{(m,u) \in S_{test}} (R_{m,u} - P_{m,u})^2} \quad (2)$$

Where, $(m,u) \in S_{test}$ test matrix, $R_{m,u}$ denotes the actual rating in the training set and $P_{m,u}$ is the predicted rating from the test set. Similarly, we try to improve the prediction accuracy using proposed hybrid filtering.

2.2. Proposed Solution

According to the proposed approach, we first consider the movie dataset and apply the expectation-maximization approach to perform the data clustering in the initial stage. After achieving the clusters, we identify the nearest clusters and estimate the rating. However, the estimated rating is processed through a genetic algorithm to achieve the optimized ranking. Finally, the achieved rankings are used for recommending the movies based on the list of movies.

2.2.1. Expectation Maximization (EM) Clustering

Several recommendation systems have used K-means clustering, an instance of the Expectation Maximization

(EM) algorithm. Generally, this algorithm is used for density estimation based on the distance parameter. The movie datasets contain irregularities, i.e., not all users need to give a rating to all the movies; hence there is a probability of achieving missing values in the data, which is difficult to handle using K-means clustering. However, missing value imputation methods can be incorporated, but we avoid the missing value imputation mechanism and adapted the Gaussian mixture model-based iterative clustering approach due to increased complexity. This scheme starts with an initial estimate and iterates to find the maximum likelihood for the input parameters.

As a result, our user modelling objectives differ from those generally assumed in recommender systems, such as enhancing accuracy or related metrics like objective functions. Our intentions also differ from past work in interaction, allowing the customer modelling, which focuses on the ability to follow the recipient's goals as the conversation develops but does not often preserve models over several talks. Our assumption is that by employing an inconspicuously derived user model to guide the platform's intuitive search for items to propose, we can improve efficiency and efficacy. Our method ensures that there is a huge repository of items to choose from and that describing these items involves a multitude of attributes. When the information is tiny, or the objects are simple to explain, lesser techniques may be sufficient.

Moreover, the EM approach is more accurate and requires less computational time when compared to K-Means [25].

Let us consider that total N points are introduced in the given data as $\{x_i\}_{i=1}^N$. We aim to assign these data points into k clusters. In this work, we model the cluster as a random variable assigned as $z = j$ with the multi-distribution probability, which satisfy $\sum \vartheta_j = 1$ such that

$$\vartheta_j = p(z = j), \forall j, j = 1, \dots, k \quad (3)$$

The Gaussian distribution is $p(x|z = j) \sim N(\mu_j, \sigma_j^2 I_j)$ where I is an identity matrix of order j , μ_j denotes the mean, variance as $\Sigma_j = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_j)$ and ϑ_j is the distribution function:

$$\begin{aligned} \varphi &= \left\{ \mu_j, \sum_j \vartheta_j \right\}_{j=1}^k p(x|\varphi) \\ &= \sum_{z=1}^k p(x|z, \varphi) p(z|\varphi) \vartheta_j \end{aligned} \quad (4)$$

z is a hidden variable; the log-likelihood of the complete data can be given as:

$$l(\varphi, D) = \log \prod_{i=1}^N \sum_{j=1}^k \vartheta_j \exp \left[-\frac{\|x_i - \mu_j\|^2}{2\sigma_j^2} \right] \quad (5)$$

Here, D introduces the input data, and the conditional expectation of $z = j$ is represented as:

$$\begin{aligned} w_j &= p(z = j|x, \varphi) = \frac{p(x|z = j, \theta) p(z = j|\vartheta_j)}{p(x|\varphi)} \\ &= \frac{\vartheta_j N(x_i|\mu_j, \Sigma_j)}{\sum_{i=1}^k \mu_j N(x_i|\mu_j, \Sigma_j)} \end{aligned} \quad (6)$$

Each data point is used for constructing the w_j , for any given x_i , w_j can be expressed as follows:

$$w_{ij} = \frac{\vartheta_j N(x_i|\mu_j, \Sigma_j)}{\sum_{i=1}^k \vartheta_j N(x_i|\mu_j, \Sigma_j)} \quad (7)$$

In order to achieve the final estimates, we have initial random mean and variance models as $\mu_j^{(0)}, \Sigma_j^{(0)}, \vartheta_j^{(0)}$.

Later we apply cosine similarity to identify the neighbouring clusters. This can be expressed as:

$$C_{sim}(u_i, u_k) = \frac{\sum_{j=1}^l (r_{ij})(r_{kj})}{\sqrt{\sum_{j=1}^l (r_{ij})^2 \sum_{j=1}^l (r_{kj})^2}} \quad (8)$$

Where u_k and u_j are the users in the cluster, r denotes their corresponding rating. The selected clusters are processed, and the difference in their ratings is computed. The smallest centroid of the cluster is assigned as the current rating of the user.

2.2.2. Multicollinearity reduction scheme

Generally, the clustering scheme suffers from the multicollinearity issue, which reduces the accuracy of the prediction system [26]. The PCA (Principle Component Analysis) is considered the most suitable scheme [27] to reduce multicollinearity. We incorporate a regression model with PCA to overcome multicollinearity in this work. Let us consider a model given as

$$Y = X\beta + \varepsilon \quad (9)$$

Where X is the data obtained from the maximum-likelihood model, Y is the response of regression [28] PCA, β denotes the constant value, and ε denotes the random error. The regression model of the principal component can be given as:

$$\begin{aligned} Y &= Z\alpha + \varepsilon \\ X &= CX, \alpha = C\beta \end{aligned} \quad (10)$$

C is a matrix that satisfies the condition as $C'(XX')C = \Lambda$ and $CC' = C'C = 1$, Λ is a diagonal matrix. Later, we apply an optimization scheme to achieve a reliable rating to improve the system performance.

2.2.3. GA Optimization scheme

This section describes the Genetic algorithm [29] for the optimization scheme to improve the prediction accuracy. This scheme [30] is a population-based stochastic approach that includes several steps such as parameter initialization, selection operator, crossover, mutation operator and finally, achieves the optimal solution.

Algorithm 1 shows the pseudo-code for the genetic algorithm [31-33].

```

Input: cluster data, number of clusters, GA parameters
such as maximum iterations=20, size of population
=50, Crossover probability  $P_c$ , Mutation probability  $P_m$ 
Output: optimal cluster according to the fitness value
(distance)
Step 1: Initialize the GA parameters and define the
fitness function as

$$f(chromosome) = \sum_{x_j \in X} \min(dist(C_i, x_j))$$

Step 2: generate initial population: generate an initial
random population as  $P(0)$  for each cluster.
Step 3: while ( $t < T_{max}$ )
    For  $i = 1: Total Population$ 
         $f_i = f(chromosome_i)$ 
    End
// Operator selection
Step 4: Apply Roulette Selection
//Crossover operator
Step 5: select  $q \in [1, k]$  randomly
    Perform crossover between parent chromosomes
using  $P_c$  and generate child chromosomes
    Parent chromosomes are  $C_{i1}, C_{i2}$  and child
chromosomes are  $C'_{i1}, C'_{i2}$ . These chromosomes can be
represented as

$$C'_{i1} = [C_{i1}(1:q, :); C_{i2}(q+1:end, :)]$$


$$C'_{i2} = [C_{i2}(1:q, :); C_{i1}(q+1:end, :)]$$

    The new child chromosome will replace the parent-
child and be considered a parent for the next iteration.
// Mutation Operation
Step 6: Mutate centre of chromosome with the given
mutation probability
Step 7: find the best chromosome and compare the
fitness.
Step 8: select the best-fit outcome chromosome as the
optimal solution.
    
```

3. Results and Analysis

In this section, we introduce the experimental analysis using the proposed approach and compare the obtained performance with existing techniques [34].

3.1. Dataset description

In order to evaluate the performance of the proposed model, we have considered several publically available datasets. *MovieLens dataset*: This has the data rating of different movies. This dataset contains 6040 datasets and a rating for 3952 movies. These ratings are provided on a scale of 1 to 5 stars. We have considered the users who have provided a minimum of 20 ratings to any movies from this dataset [35-39].

Yahoo movie dataset: in this dataset also, the ratings are provided on a scale of 1 to 5 [40]. The training dataset includes 7642 users, 11915 movies and 211231 ratings for movies. The test data includes 2309 users, 2380 movies and a total of 10136 ratings [41].

Netflix dataset: in this dataset, a total of 480,000 customers are identified using a unique ID for each user [42-45]. There are over 17000 movies considered to extract the rating data where each user’s unique ID and rating are organized for recommendation purposes [46]. This dataset was collected from October 1998 to December 2005, and 100 million ratings were achieved [47-49]. Table 1 shows a general arrangement of these datasets [50].

Table 1. Movie data description

Movie ID/Name	Customer ID	Rating	Title	Year of Release	Date
1...N	1...M	1 to 5	Movie Name	Year	Release date

3.2. Performance measurement

The proposed model is experimented with using MATLAB simulation tool running under 4GHz processor, 8GB RAM and 64-bit Windows 8 operating system [51-54]. The performance of the proposed model is computed in terms of Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) and compared with the existing techniques [55-58].

The mean absolute error is computed as:

$$\begin{aligned}
 &MAE(Predicted, Actual) \\
 &= \sum_{i=1}^N \left| \frac{predicted_{u,i} - actual_{u,i}}{N} \right| \tag{11}
 \end{aligned}$$

Where N denotes the number of users, actual and predicted ratings are denoted as *Predicted* and *actual*.

Similarly, RMSE can be computed as follows:

$$\begin{aligned}
 &RMSE(Predicted, Actual) \\
 &= \sqrt{\sum \frac{(predicted\ rating - actual\ rating)(predicted\ rating - actual\ rating)}{n}} \tag{12}
 \end{aligned}$$

First of all, we measured the performance for a varied number of clusters and compared the MAE and RMSE performance with the existing K-means and cuckoo search-based optimization [17]; [59-62].

Table 2. MAE performance for varied cluster numbers for Movielens dataset

Number of Cluster	Mean Absolute Error [17]	Mean Absolute Error (Proposed)
-------------------	--------------------------	--------------------------------

4	0.825	0.751
8	0.806	0.705
10	0.795	0.681
12	0.790	0.668
16	0.781	0.65
20	0.776	0.701
28	0.768	0.6842
32	0.761	0.714
36	0.754	0.703
40	0.744	0.697

The above-given table 2 shows a comparative performance in terms of MAE [63]. In order to achieve this performance, we have varied the cluster size from 4 to 40 [64]. In this work, we have obtained the average MAE as 0.78 0.695 using the cuckoo search and proposed approach, respectively [65-67]. Table 3 compares the performance with existing techniques for varied cluster sizes.

Table 3. MAE comparative performance

Approach Cluster	5	10	15	20	25	30	35	40
PCA-GAKM [26]	0.79	0.77	0.77	0.78	0.781	0.785	0.786	0.788
PCA-SOM[26]	0.82	0.79	0.79	0.79	0.8	0.805	0.806	0.8
SOM-CLUSTER[26]	0.819	0.81	0.81	0.81	0.81	0.805	0.81	0.81
UPCC[26]	0.825	0.825	0.824	0.828	0.824	0.824	0.825	0.825
k-means cluster[26]	0.825	0.82	0.818	0.815	0.815	0.814	0.813	0.812
PCA-k-means[26]	0.85	0.845	0.841	0.84	0.84	0.84	0.84	0.84
GAKM CLUSTER[26]	0.815	0.805	0.804	0.804	0.804	0.804	0.803	0.803
ABC-KM[26]	0.773	0.764	0.764	0.771	0.78	0.784	0.782	0.787
Proposed Approach	0.759	0.681	0.661	0.701	0.702	0.701	0.703	0.703

Similarly, we computed the MAE performance for the Yahoo movie and Netflix datasets. The obtained performance is introduced in table 4.

Table 4. Comparative MAE performance using the proposed approach.

Number of Cluster	Yahoo Movie	Netflix
4	0.802	0.856
8	0.855	0.802
10	0.741	0.789
12	0.71	0.756
16	0.689	0.712
20	0.652	0.687
28	0.613	0.613
32	0.591	0.601
36	0.587	0.612
40	0.579	0.589

This study aims to achieve the MAE performance of 0.681 and 0.701 for the Yahoo Movie and Netflix dataset using the proposed approach. Similarly, we measure the performance in terms of RMSE for the movielens dataset,

and the obtained performance is introduced in below-given table 5.

Table 5. RMSE Performance

Number of Cluster	RMSE [17]	RMSE (Proposed)
4	1.3065	0.751
8	1.2975	0.705
10	1.2939	0.681
12	1.288	0.668
16	1.285	0.65
20	1.28	0.701
28	1.276	0.6842
32	1.271	0.714
36	1.266	0.703
40	1.255	0.697

Similarly, we have measured the RMSE performance for Yahoo and Netflix datasets as given in table 6.

Table 6. RMSE performance for Yahoo and Netflix data

Number of Cluster	Yahoo Movie	Netflix
4	1.32	1.28
8	1.38	1.34
10	1.33	1.33
12	1.28	1.34
16	1.22	1.32
20	1.26	1.3
28	1.29	1.28
32	1.32	1.25
36	1.33	1.32
40	1.3	1.33

According to this experiment, we achieved RMSE values like 1.303 and 1.309 for Yahoo Movie and Netflix datasets using the proposed approach. These experiments show that the proposed approach achieves better accuracy for movie recommendations using hybridized modelling.

4. Conclusion

In this work, we have introduced a hybrid movie recommender system. According to the proposed approach, we first apply an EM clustering scheme for grouping the rating data, and later, the nearest cluster is selected, and a rating is assigned for each user based on the centroid. In the next stage, we apply a PCA-based regression model followed by a genetic algorithm to minimize the clustering error and achieve better rating prediction accuracy based on the optimal rating. Top neighbours are selected for achieving the recommended movies. The experimental study shows that the proposed approach achieves better performance in terms of RMSE and MAE

References

- [1] Ricci F, Rokach L, Shapira B. Recommender systems: introduction and challenges. In *Recommender systems handbook 2015* (pp. 1-34). Springer, Boston, MA.
- [2] Akcayol MA, Utku A, Aydoğan E, Mutlu B. A weighted multi-attribute-based recommender system using extended user behavior analysis. *Electronic Commerce Research and Applications*. 2018 Mar 1; 28:86-93.
- [3] Jiang L, Cheng Y, Yang L, Li J, Yan H, Wang X. A trust-based collaborative filtering algorithm for E-commerce recommendation system. *Journal of Ambient Intelligence and Humanized Computing*. 2018:1-2.
- [4] Lu J, Wu D, Mao M, Wang W, Zhang G. Recommender system application developments: a survey. *Decision Support Systems*. 2015 Jun 1; 74:12-32.
- [5] Vaz PC, Martins de Matos D, Martins B, Calado P. Improving a hybrid literary book recommendation system through author ranking. In *Proceedings of the 12th ACM/IEEE-CS joint conference on Digital Libraries 2012 Jun 10* (pp. 387-388). ACM.
- [6] Wang Z, Yu X, Feng N, Wang Z. An improved collaborative movie recommendation system using computational intelligence. *Journal of Visual Languages & Computing*. 2014 Dec 1; 25(6):667-75.
- [7] Deldjoo Y, Elahi M, Cremonesi P, Garzotto F, Piazzolla P, Quadrana M. Content-based video recommendation system based on stylistic visual features. *Journal on Data Semantics*. 2016 Jun 1; 5(2):99-113.
- [8] Xu H, Zhang R, Lin C, Gan W. Construction of e-commerce recommendation system based on semantic annotation of ontology and user preference. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2014; 12(3):2028-35.
- [9] Al-Shamri MY. User profiling approaches for demographic recommender systems. *Knowledge-Based Systems*. 2016 May 15; 100:175-87.
- [10] Carrer-Neto W, Hernández-Alcaraz ML, Valencia-García R, García-Sánchez F. Social knowledge-based recommender system. Application to the movies domain. *Expert Systems with Applications*. 2012 Sep 15;39(12):10990-1000.
- [11] Wei S, Zheng X, Chen D, Chen C. A hybrid approach for movie recommendation via tags and ratings. *Electronic Commerce Research and Applications*. 2016 Jul 1;18:83-94.
- [12] Zhang F, Yuan NJ, Lian D, Xie X, Ma WY. Collaborative knowledge base embedding for recommender systems. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining 2016 Aug 13* (pp. 353-362). ACM.
- [13] Hassan M, Hamada M. A neural networks approach for improving the accuracy of multi-criteria recommender systems. *Applied Sciences*. 2017;7(9):868.
- [14] Kuang G, Li Y. Using fuzzy association rules to design e-commerce personalized recommendation system. *TELKOMNIKA Indonesian J. Elec. Engin*. 2014 Feb;12(2):1519-27.
- [15] Kothari AA, Patel WD. A novel approach towards context based recommendations using support vector machine methodology. *Procedia Computer Science*. 2015 Jan 1;57:1171-8.
- [16] Katarya R, Verma OP. A collaborative recommender system enhanced with particle swarm optimization technique. *Multimedia Tools and Applications*. 2016 Aug 1;75(15):9225-39.
- [17] Katarya R, Verma OP. An effective collaborative movie recommender system with cuckoo search. *Egyptian Informatics Journal*. 2017 Jul 1;18(2):105-12.
- [18] Katarya R, Verma OP. Recommender system with grey wolf optimizer and FCM. *Neural Computing and Applications*. 2018 Sep 1;30(5):1679-87.
- [19] Ar Y, Bostanci E. A genetic algorithm solution to the collaborative filtering problem. *Expert Systems with Applications*. 2016 Nov 1;61:122-8.
- [20] da Silva EQ, Camilo-Junior CG, Pascoal LM, Rosa TC. An evolutionary approach for combining results of recommender systems techniques based on collaborative filtering. *Expert Systems with Applications*. 2016 Jul 1;53:204-18.
- [21] Paradarami TK, Bastian ND, Wightman JL. A hybrid recommender system using artificial neural networks. *Expert Systems with Applications*. 2017 Oct 15;83:300-13.
- [22] Kouki P, Fakhraei S, Foulds J, Eirinaki M, Getoor L. Hyper: A flexible and extensible probabilistic framework for hybrid recommender systems. In *Proceedings of the 9th ACM Conference on Recommender Systems 2015 Sep 16* (pp. 99-106). ACM.
- [23] Aslanian E, Radmanesh M, Jalili M. Hybrid recommender systems based on content feature relationship. *IEEE Transactions on Industrial Informatics*. 2016 Nov 21.
- [24] Kermany NR, Alizadeh SH. A hybrid multi-criteria recommender system using ontology and neuro-fuzzy techniques. *Electronic Commerce Research and Applications*. 2017 Jan 1;21:50-64.
- [25] Jung YG, Kang MS, Heo J. Clustering performance comparison using K-means and expectation maximization algorithms. *Biotechnology & Biotechnological Equipment*. 2014 Nov 14;28(sup1):S44-8.
- [26] Katarya, R. (2018). Movie recommender system with metaheuristic artificial bee. *Neural Computing and Applications*. doi:10.1007/s00521-017-3338-4.
- [27] Rangarajan S, Liu H and Wang H. Web service QoS prediction using improved software source code metrics. *PLoS ONE* 15(1): e0226867, 2020. <https://doi.org/10.1371/journal.pone.0226867>
- [28] Jiahua Du, Jia Rong, Hua Wang and Yanchun Zhang. Helpfulness prediction for online reviews with explicit content-rating interaction. *International Conference on Web Information Systems Engineering*, pp.795-809, 2020.
- [29] JiaoYin, MingJianTang, JinliCaoa, HuaWang, and MingshanYou. A Real-time Dynamic Concept Adaptive Learning Algorithm for Exploitability Prediction. *Neurocomputing*, 472, pp.252-265, 2022.
- [30] JiaoYin, MingJianTang, JinliCaoa, HuaWang, MingshanYou and Yongzheng Lin. Vulnerability exploitation time prediction: an integrated framework for dynamic imbalanced learning. *World Wide Web*, 25(1), pp.401-423, 2022.
- [31] A, V. V. ., T, S. ., S, S. N. ., Rajest, D. S. S. . IoT-Based Automated Oxygen Pumping System for Acute Asthma Patients. *European Journal of Life Safety and Stability* (2660-9630), 19 (7), 8-34, 2022.
- [32] A. Bibi et al., "Skin Lesion Segmentation and Classification Using Conventional and Deep Learning Based Framework," 2022.
- [33] A. H. El-Gamal, R. R. Mostafa, and N. A. Hikal, "Load Balancing Enhanced Technique for Static Task Scheduling in Cloud Computing Environments," in *Internet of*

- Things—Applications and Future: Springer, 2020, pp. 411-430.
- [34] Aakanksha Singhal and D.K. Sharma, "New Generalized 'Useful' Entropies using Weighted Quasi-Linear Mean for Efficient Networking," *Mobile Networks and Applications*, no., pp. 1–11, 2022.
- [35] Adil, M., Ali, J., Attique, M., Jadoon, M. M., Abbas, S., Alotaibi, S. R., ... & Farouk, A. Three Byte-Based Mutual Authentication Scheme for Autonomous Internet of Vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 2021.
- [36] Adil, M., Khan, M. K., Jamjoom, M., & Farouk, A. MHADBOR: AI-enabled Administrative Distance based Opportunistic Load Balancing Scheme for an Agriculture Internet of Things Network. *IEEE Micro*, 2021.
- [37] Anantha Krishnan Venkatesan , Umashankar Subramaniam , Mahajan Sagar Bhaskar , O. V. Gnana Swathika , Sanjeevikumar Padmanaban , Dhafer J. Almakhlis , and Massimo Mitolo, "Small-Signal Stability Analysis for Microgrids Under Uncertainty Using MALANN Control Technique", *IEEE Systems Journal*, Vol. 15, No. 3, pp. 3797-3807, September 2021.
- [38] Aoudni, Y., Donald, C., Farouk, A., Sahay, K. B., Babu, D. V., Tripathi, V., & Dhablya, D. Cloud security based attack detection using transductive learning integrated with Hidden Markov Model. *Pattern Recognition Letters*, 157, 16-26, 2022.
- [39] D Datta, S Mishra, SS Rajest, "Quantification of tolerance limits of engineering system using uncertainty modeling for sustainable energy" *International Journal of Intelligent Networks*, Vol.1, 2020, pp.1-8, 2020.
- [40] D. K. Sharma, N. C. Singh, D. A. Noola, A. N. Doss, and J. Sivakumar, "A review on various cryptographic techniques & algorithms," *Materials Today: Proceedings*, 2021.
- [41] D. Kumar, S. Kumar, and R. Bansal. "Multi-objective multi-join query optimisation using modified grey wolf optimisation." *International Journal of Advanced Intelligence Paradigms*, vol.17, no.1-2, pp. 67-79, 2020.
- [42] Adil, M., Khan, M. K., Jadoon, M. M., Attique, M., Song, H., & Farouk, A. An AI-enabled Hybrid lightweight Authentication Scheme for Intelligent IoMT based Cyber-Physical Systems. *IEEE Transactions on Network Science and Engineering*, 2022.
- [43] D. Kumar, S. Kumar, R. Bansal and P.Singla. "A Survey to Nature Inspired Soft Computing." *International Journal of Information System Modeling and Design*, vol. 8, no. 2, pp.112-133, 2017.
- [44] Anantha Krishnan.V and N. Senthil Kumar, "Real-Time Simulation Analysis of LM Algorithm-Based NN For The Control of VSC In Grid Connected PV-Diesel Microgrid Using OP4500 RT-Lab Simulator", *International Journal of Power and Energy Systems*, Acta Press, Vol. 42, No. 10, pp. 1-10.
- [45] D. Kumar, D.Mehrotra, and R. Bansal , "Metaheuristic Policies for Discovery Task Programming Matters in Cloud Computing." *Proceedings of the 4th International Conference on Computing Communication and Automation (ICCCA) 2018*, pp. 1-5, 2018.
- [46] P. Rajesh, C. Naveen, Anantha Krishan Venkatesan, and Francis H. Shajin, "An optimization technique for battery energy storage with wind turbine generator integration in unbalanced radial distribution network", *Journal of Energy Storage*, Vo. 43, pp 1-12, 2021.
- [47] R. M. Adnan, R. R. Mostafa, A. R. M. Islam, A. D. Gorgij, A. Kuriqi, and O. J. W. Kisi, "Improving Drought Modeling Using Hybrid Random Vector Functional Link Methods," vol. 13, no. 23, p. 3379, 2021.
- [48] S. Suman Rajest, Bhopendra Singh, P. Kavitha, R. Regin, K. Praghash, S. Sujatha, "Optimized Node Clustering based on Received Signal Strength with Particle Ordered-filter Routing Used in VANET" *Webology*, Vol.17, No.2, pp. 262-277, 2020.
- [49] D. Kumar, D.Mehrotra, and R. Bansal. "Query Optimization in Crowd-Sourcing Using Multi-Objective Ant Lion Optimizer." *International Journal of Information Technology and Web Engineering*, vol. 14, no. 4, pp. 50-63, 2019.
- [50] E. T. Elkabbash, R. R. Mostafa, and S. I. J. P. o. Barakat, "Android malware classification based on random vector functional link and artificial Jellyfish Search optimizer," vol. 16, no. 11, p. e0260232, 2021.
- [51] Farouk, A., Alahmadi, A., Ghose, S., & Mashatan, A. Blockchain platform for industrial healthcare: Vision and future opportunities. *Computer Communications*, 154, 223-235, 2020.
- [52] G. S. Bagale, Vandadi, V. R., Singh, D., Sharma, D. K., Garlapati, D. V. K., Bommiseti, R. K., Gupta, R. K., Setsiawan, R., Subramaniaswamy, V., & Sengan, S., "Small and medium-sized enterprises' contribution in digital technology," *Annals of Operations Research*, pp. 1–24, 2021.
- [53] Adil, M., Attique, M., Khan, M. M., Ali, J., Farouk, A., & Song, H. HOPCTP: A Robust Channel Categorization Data Preservation Scheme for Industrial Healthcare Internet of Things. *IEEE Transactions on Industrial Informatics*, 2022.
- [54] H. Tao et al., "Training and testing data division influence on hybrid machine learning model process: application of river flow forecasting," vol. 2020, 2020.
- [55] I. M. El-Hasnony, R. R. Mostafa, M. Elhoseny, and S. I. J. T. o. E. T. T. Barakat, "Leveraging mist and fog for big data analytics in IoT environment," vol. 32, no. 7, p. e4057, 2021.
- [56] I. Puthige et al., "Safest Route Detection via Danger Index Calculation and K-Means Clustering," 2021.
- [57] Jayakumar P., Suman Rajest S., Aravind B.R. An Empirical Study on the Effectiveness of Online Teaching and Learning Outcomes with Regard to LSRW Skills in COVID-19 Pandemic. *Technologies, Artificial Intelligence and the Future of Learning Post-COVID-19. Studies in Computational Intelligence*, vol 1019. Springer, Cham. 2022.
- [58] K.K.D. Ramesh, G. Kiran Kumar, K. Swapna, Debabrata Datta, and S. Suman Rajest, "A Review of Medical Image Segmentation Algorithms", *EAI Endorsed Transactions on Pervasive Health and Technology*, 2021.
- [59] Leo Willyanto Santoso, Bhopendra Singh, S. Suman Rajest, R. Regin, Karrar Hameed Kadhim, "A Genetic Programming Approach to Binary Classification Problem" *EAI Endorsed Transactions on Energy*, Vol.8, no. 31, pp. 1-8, 2021.
- [60] Adil, M., Song, H., Ali, J., Jan, M. A., Attique, M., Abbas, S., & Farouk, A. EnhancedAODV: A Robust Three Phase Priority-based Traffic Load Balancing Scheme for Internet of Things. *IEEE Internet of Things Journal*, 2021.
- [61] Mendonça, R. V., Silva, J. C., Rosa, R. L., Saadi, M., Rodriguez, D. Z., & Farouk, A. A lightweight intelligent intrusion detection system for industrial internet of things using deep learning algorithm. *Expert Systems*, e12917, 2021.

- [62] N. M. Ashraf, R. R. Mostafa, R. H. Sakr, and M. Z. Rashad, "Optimizing hyperparameters of deep reinforcement learning for autonomous driving based on whale optimization algorithm," *Plos one*, vol. 16, no. 6, p. e0252754, 2021.
- [63] Nandakumar, K., Vinod, V., Akbar Batcha, S. M., Sharma, D. K., Elangovan, M., Poonia, A., Mudlappa Basavaraju, S., Dogiwal, S. R., Dadheech, P., & Sengan, S., "Securing data in transit using data-in-transit defender architecture for cloud communication," *Soft Computing*, vol. 25, no. 18, pp. 12343–12356, 2021.
- [64] Adil, M., Jan, M. A., Mastorakis, S., Song, H., Jadoon, M. M., Abbas, S., & Farouk, A. Hash-MAC-DSDV: Mutual Authentication for Intelligent IoT-Based Cyber-Physical Systems. *IEEE Internet of Things Journal*, 2021.
- [65] Neffati, O. S., Sengan, S., Thangavelu, K. D., Kumar, S. D., Setiawan, R., Elangovan, M., Mani, D., & Velayutham, P., "Migrating from traditional grid to smart grid in smart cities promoted in developing country," *Sustainable Energy Technologies and Assessments*, vol. 45, p. 101125, 2021.
- [66] Adil, M., Jan, M. A., Mastorakis, S., Song, H., Jadoon, M. M., Abbas, S., & Farouk, A. Hash-MAC-DSDV: Mutual Authentication for Intelligent IoT-Based Cyber-Physical Systems. *IEEE Internet of Things Journal*, 2021.
- [67] Zhu, F., Zhang, C., Zheng, Z., & Farouk, A. Practical Network Coding Technologies and Softwarization in Wireless Networks. *IEEE Internet of Things Journal*, 8(7), 5211-5218, 2021.