International Journal of Public Health Science (IJPHS)
Vol. 13, No.2, June 2024, pp. 794~803
ISSN: 2252-8806, DOI: 10.11591/ijphs.v13i2.23322

Validation on selected breast cancer drugs of physicochemical features by using machine learning models

Vuddagiri Mutya Naga Sri Surya Venkata Krishna Rao Gupta[1], Chitta Venkata Phani Krishna[2], Konakanchi Venkata Subrahmanya Srirama Murthy[3], Reddy Shiva Shankar[3]
[1]Faculty of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India
[2]Faculty of Computer Science and Engineering, Teegala Krishna Reddy Engineering College, Hyderabad, Telangana, India
[3]Faculty of Computer Science and Engineering, Sagi Rama Krishnam Raju Engineering College (A), Bhimavaram, Andhra Pradesh, India

Article Info

Article history:

Received Apr 23, 2023
Revised Jun 29, 2023
Accepted Jul 13, 2023

Keywords:

Cluster validation
Clustering
Combinational therapy
Fuzzy clustering
Hopkins statistic
Physico chemical properties

ABSTRACT

Breast cancer is one of the leading causes of death among females today. The elbow approach determines the ideal number of clusters after determining that the Dataset is highly cluster able with the Hopkins statistic. Three distinct groups with distinct differences were produced using the dataset's proposed expectation maximization fuzzy k-means clustering algorithm (PEMFKM). Different fuzzy clustering techniques, such as fuzzy k-means (FKM), fuzzy k-means with entropy (FKM.ENT), fuzzy k-means with entropy and noise (FKM.ENT.NOISE), Gustafson and Kessel - like fuzzy k-means (FKM.GK), Gustafson and Kessel - like fuzzy k-means with entropy regularization (FKM.GK.ENT), Gustafson and Kessel - like fuzzy k-means with entropy regularization and noise (FKM.GK.ENT.NOISE), and PEMFKM, are evaluated. The partition coefficient (PC), partition entropy (PE), and Modified partition coefficient index (MPC) index values are better for FKM.GK than the suggested PEMFKM method. When compared to the FKM.GK method, the index values for the proposed PEMFKM algorithm have superior results for the parameters Silhouette (SIL), Xie and Beni index (XB), and fuzzy silhouette index (SIL.F). The results shows that the PEMFKM algorithm will provide better clusters and that the drugs in a given cluster may be combined for use in combination therapy for breast cancer treatment.

Corresponding Author:

Vuddagiri Mutya Naga Sri Surya Venkata Krishna Rao Gupta
Faculty of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation
Andhra Pradesh, India
Email: guptavkrao@gmail.com

1. INTRODUCTION

Breast cancer is characterized by the uncontrolled proliferation of cancerous cells in the breast. The cancerous cells may spread to other organs if the patient is not treated regularlywith proper medicine. According to the National Cancer Institute, around 25-30% of females will be diagnosed with breast cancer throughout their lifetimes [1]. Despite the fact that its five year survival rate [2] has improved from 75% to 90% between 1980 and 2013, it is still a concerning statistic. Some data-driven AI algorithms are beneficial in predicting breast cancer [3].

The NCCN, the National Comprehensive Cancer Network, recognizes the significance of early cancer detection and treatment. Cancer-related pain is first referred to be the multi-factorial dreadful emotional occurrence of a cognitive psychological,and emotional experience. Cancer Distress or discomfort

may vary from common emotions like vulnerability, dissatisfaction, and doubts to issues that might develop into impairments like melancholy, social isolation, panic attacks, anxiety, and religious crises [4]. The illness has been treated incorrectly by forcing all of the underlying reasons, and cancer should be diagnosed and treated according to the associated symptoms to improve humankind's quality of life [5]. A significant obstacle remains in helping people understand the gravity of cancer and its effects on patients. Susceptibility, melancholy, and transition worries necessitate engagement or assistance. Cancer patients are frequently checked or evaluated at clinical facilities to increase their survival chances [4].

Even though 37% of breast cancer survivors must visit an oncologist two or more times in the first year after stage1 or main therapy [6], this is no reason to avoid checkups. It is a novel approach to identifying breast cancer patients who are most at risk for the disease's recurrence and the subsequent shift in their chances of survival. For instance, participation in preventative cognitive behavioural therapy (CBT) has reduced depressive symptoms and anxiety in cancer patients by 50% [7]. Nanocarriers will improve the therapeutic effectiveness of anti-cancer medications. Breast cancer treatment includes various modalities, including chemotherapy, radiation therapy, and surgical removal of cancerous tissue [8]. Current medical practice favours conserving breast tissue while treating breast cancer.

Breast cancer therapy is a major priority but poses a significant challenge to the medical community. Two of the most pressing issues are slowing the disease's progression and reducing the impact of malignant cells. Therefore, combination therapy will be more successful than single-medication treatment.

2. RELATED WORK

Cirkovic *et al.* [9] emphasized estimating the survival rate and the decline of cancer cells in breast cancer patients. They evaluated the breast cancer data set maintained by the Clinical Center of Kragujevac to train the algorithm. A total of six classifiers are acquainted using this Dataset, with 20 characteristics chosen for training. In predicting whether or not a cancer patient would make it through treatment, the Naïve Bayes (NB) classifier and support vector machine (SVM) performed best, while the artificial neural networks (ANN) performed best regarding recurrence. Analyzing machine learning (ML) models requires the use of the accuracy (AC), sensitivity (SENS), and specificity (SPEC) metrics. ANN were used by Ayer *et al.* [10] to estimate the likelihood of developing breast cancer. They created a three-layer feed-forward ANN with a thousand training iterations. Breast cancer development was predicted using a recurrent neural network (RNN) by Appaji *et al.* [11]. They accessed the breast cancer data archived at UC Irvine for their training purposes. The RNN employed the ReLU activation function, and after the necessary processes of data pre-processing, model creation, and result analysis, it achieved a 97% F1 score.

Goyal *et al.* [12] researched to track a cancer tumor’s progression inside a patient's body. For their training purposes, they used the Wisconsin diagnostic breast cancer (WDBC) dataset. The SVM also outperformed competing models in terms of both specificity and sensitivity. Posterior probability estimate through neural networks (NN) is offered by Zhang [13], who describes the procedure of calculating the new probability Value by considering an event's updated information. Explained how different types of NN classifier’s function and then zeroed focus on the training of NNs. Bouguessa *et al.* [14] focused on analyzing the behavior of data present in crossing groups. Fuzzy covariance (FC) was the method of choice for this job. The membership matrix was first generated via FC, which allowed for a reduction in the proposed algorithm's time complexity. The algorithm's cluster validity has been improved thanks to a new validation function they devised Fuzzy maximum likely hood algorithm (FML). ML allows for a higher degree of precision to be reached. When it comes to detecting breast cancer, the latest version of K-NN has been employed by Ahmad *et al.* [15].

Akay [16] suggested an SVM-like model for breast cancer diagnosis, and they dubbed the model that was constructed using SVM as BCRSVM ('breast cancer recurrence prediction based on SVM). He trained his model using the Wisconsin breast cancer dataset (WBCD). Instead of considering all of the Dataset's properties, he computed the F score and selected the features with the highest F score. McDonald *et al.* [17] advanced diagnostics include molecular imaging and genomic expression profiles enhance tumor characterization. These diagnostics, together with modern surgical and radiation procedures, provide a multidisciplinary approach to recurrence and treatment-associated morbidity. Early breast cancer diagnosis is crucial to therapy. Imaging is one of the key breast cancer diagnostic platforms that may give useful data [18]. Reddy *et al.* [19] concentrated on breast cancer and gestational diabetes in patients by using various datasets with various ML techniques [20].

3. METHOD

The significant steps involved in selection of breast cancer drugs of physiochemical properties are Data Collection, Dataset Description, Models used. The organization of this work is shown in Figure 1; it includes the selection of breast cancer drugs, computing the values of identified attributes, cluster formation with variant fuzzy algorithms, and then applying the proposed expectation maximization fuzzy k-means clustering algorithm (PEMFKM) algorithm on the drugs to form the clusters. The selected drugs from the Drug Bank database are shown in the Table 1.

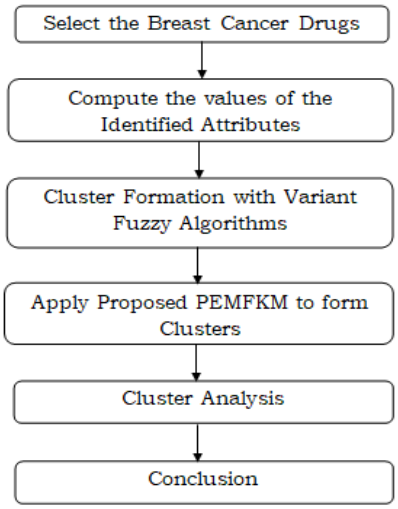

Figure 1. Organization of this work

Table 1. Selected breast cancer drugs

SNO	Drugs	SNO	Drugs	SNO	Drugs
1	Paclitaxel	7	Raloxifene	13	Ixabepilone
2	Everolimus	8	Fulvestrant	14	Tamoxifen
3	Pamidronate	9	Letrozole	15	Docetaxel
4	Anastrozole	10	Eribulin	16	Lapatinib
5	Exemestane	11	Palbociclib	17	Capecitabine
6	Epirubicin	12	Gefitinib	-	---

3.1. Objectives

To determine the best number of clusters 'k' the Elbow approach applied on the food and drug administration (FDA)approved seventeen breast cancer medicines based on physico-chemical properties. Then apply (PEMFKM) clustering technique to the physico chemical properties of seventeen FDA-approved breast cancer drugs to form well-defined clusters. To validate the efficiency of clusters reliability parameters like Partition Coefficient index (PC), partition entropy index (PE), Modified partition coefficient index (MPC), Silhouette (SIL), Xie and Beni index (XB), and fuzzy silhouette index (SIL.F) are used.

3.2. Computing the values of the attribute

3.2.1. Molecular weight (MW)

The MW of a material or molecule can be determined using its chemical formula, and the values are taken from the periodic table. The total sum of the atomic weights of each atom in the molecule and the computation of the drugs was shown in Table 2.

For example: Consider the Drug Paclitaxel

Paclitaxel Molecular Formula: $C_{47}H_{51}NO_{14}$ and MW =853.9 g/mol

Molecular Weight=47*(12.0107)+51*(1.0078)+(14.0067)+14*(15.999)

=853.8934

Log P: The **Log P value** of the composite indicates the drugs' ability to reach the body's goal tissue. Every identified compounds are in Log P because the **Log P**>0 (or **P**>1)

$$P = \frac{concentrationofdruginoctolphase}{concentrationofdruginwaterphase}$$

$$\text{Log (P)} = \log \frac{[c]octonal}{[c]water}$$

Hydrogen Acceptor count: H-bond acceptors are found from the general rule. The Lipinski violations are just counted for oxygen and nitrogen. Acceptor atoms are defined as a lone-pair electrons.
Hydrogen donor count: A molecule that supplies the hydrogen atom of a hydrogen bond. Donor atoms are always connected to a minimum of one H atom.
Rotatable bond count: RBN is the number of free-rotating bonds. The high rotational energy barrier excludes amide carbon-nitrogen (C–N) bonds.

Table 2. Computation of MW for the selected breast cancer drugs

Drug	Formula	C 12.011	H 1.008	N 14.007	0 15.999	F 18.998	I 126.904	P 30.97	S 32.07	CI 35.45	MW
Paclitaxel	$C_{47}H_{51}NO_{14}$	47	51	1	14	0	0	0	0	0	853.9
Everolimus	$C_{53}H_{83}NO_{14}$	53	83	1	14	0	0	0	0	0	958.3
Pamidronate	$C_{53}H_{83}NO_{14}$	3	11	1	7	0	0	2	0	0	235.1
Anastrozole	$C_{17}H_{19}N_5$	17	19	5	0	0	0	0	0	0	293.4
Exemestane	$C_{20}H_{24}O_2$	20	24	0	2	0	0	0	0	0	296.4
Epirubicin	$C_{27}H_{29}NO_{11}$	27	29	1	11	0	0	0	0	0	543.5
Raloxifene	$C_{28}H_{27}NO_4S$	28	27	0	3	5	0	0	1	0	538.6
Fulvestrant	$C_{32}H_{47}F_5O_3S$	32	47	0	3	5	0	0	1	0	606.8
Letrozole	$C_{17}H_{11}N_5$	17	11	5	0	0	0	0	0	0	285.3
Eribulin	$C_{40}H_{59}NO_{11}$	40	59	1	11	0	0	0	0	0	729.9
Palbociclib	$C_{24}H_{29}N_7O_2$	24	29	7	2	0	0	0	0	0	447.5
Gefitinib	$C_{22}H_{24}ClFN_4O$	22	24	4	3	1	0	0	0	1	447
Ixabepilone	$C_{27}H_{42}N_2O_5S$	27	42	2	5	0	0	0	1	0	506.7
Tamoxifen	C26H29NO	26	29	1	1	0	0	0	0	0	371.5
Docetaxel	$C_{43}H_{53}NO_{14}$	43	53	1	14	0	0	0	0	0	807.9
Lapatinib	$C_{29}H_{26}ClFN_4O_4S$	29	26	4	4	1	0	0	1	1	581.2
Capecitabine	$C_{15}H_{22}FN_3O_6$	15	22	3	6	1	0	0	0	0	359.4

4. CLUSTER FORMATION USING VARIANT FUZZY CLUSTERING ALGORITHMS

In this step, the well-known k-means procedure is augmented with variant fuzzy algorithms [21], then applies to the physicochemical properties of the drugs to construct meaningful clusters. This approach uses entropy-regularized fuzzy k-means clustering [22]. Figure 2 shows how this method clusters the Dataset, moving certain items from one cluster to another. Entropy regularization removes m. High-membership objects are outliers [23]. Figure 3 shows the results of a misclassification by Fuzzy K-Means using entropy regularization and the noise cluster approach. This approach finds non-spherical clusters using GKFK [Gustafson and Kessel-like fuzzy k-means clustering [24]. Figure 4 shows that this algorithm cannot identify better cluster groups or centroids when applied to the Dataset.

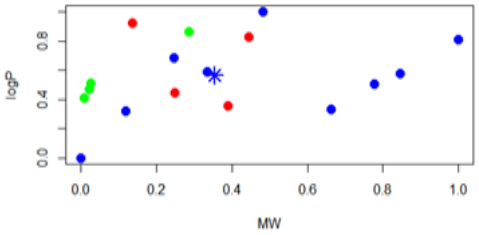

Figure 2. Entropy regularization produces 3 clusters with few data points in others

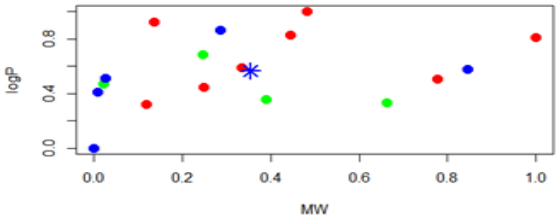

Figure 3. Three noise clusters with entropy regularization

This algorithm uses GKFK with entropy regularization [25]. Figure 5 shows the clustering procedure used to the specified Dataset if standardization is stand=1. This tool could not find better clusters for discrete data. This application uses the GKFK approach with entropy regularization. This technique failed to find non-spherical clusters. The natural FKM technique is the only one that produces three cluster solutions.

4.1. PEMFKM algorithm

We are introducing a new PEMFKM algorithm which uses Hopkins statistics to determine whether data is clusterable or not; then, the elbow method is applied to know the optimal number of clusters for the breast cancer drugs dataset, which later forms the clusters. The proposed algorithm is shown in Figure 6.

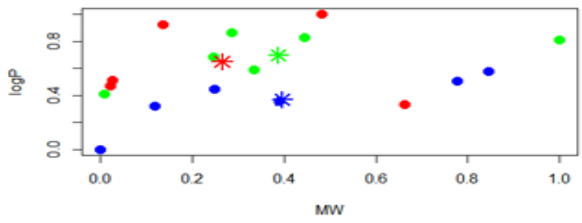

Figure 4. Showing distributed data using Gustafson and Kessel-like fuzzy k-means clustering

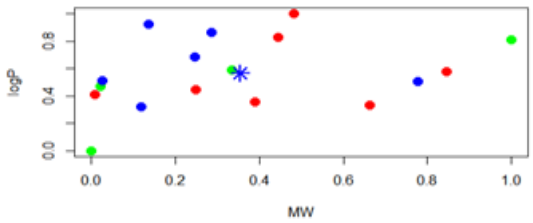

Figure 5. Entropy regularization showing diverged clusters

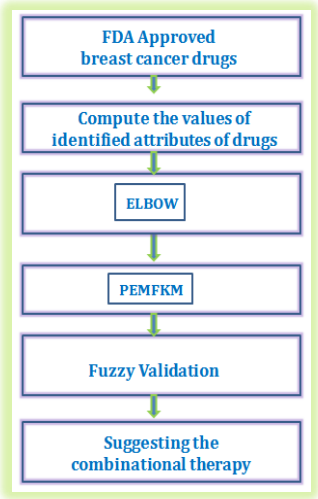

Figure 6. Proposed model

4.2. Algorithm for proposed model

The dataset was used to test the PEMFKM algorithm with m=2 to 5. The membership numbers for m=2 and m=3 are essentially identical except for the cluster centroid placements displayed in Figure 7. Three clusters when m=3. Table 3 lists breast cancer medications and their cluster memberships. In Figure 7 shows the plots for m=2 and m=3.

Algorithm 1:

STEP 1.
a. Calculate drug features.
b. Recognize drug features.

Extraction: Elbow technique to determine optimal cluster number

STEP 2.
1. K-means clustering is calculated for 1-10 clusters.
2. Each iteration computes WSS, the total within-cluster sum of squares.
3. The number of clusters and total of squares were plotted on an X-Y plot.
4. Find the graph's knee and accurate cluster count.

Construct distinct clusters.

i. After pre-processing find first mean predictions by reading the K
ii. EM increases chances.

iii. Randomize point coefficients between iterations (A).
iv. Consider threshold (B).
v. If A<=B, go to step (ii).
vi. Stop.

Fuzzy validation

STEP 3. Our dataset is used to test cluster efficiency using with various FKM algorithms and PEMFKM. F, XB.

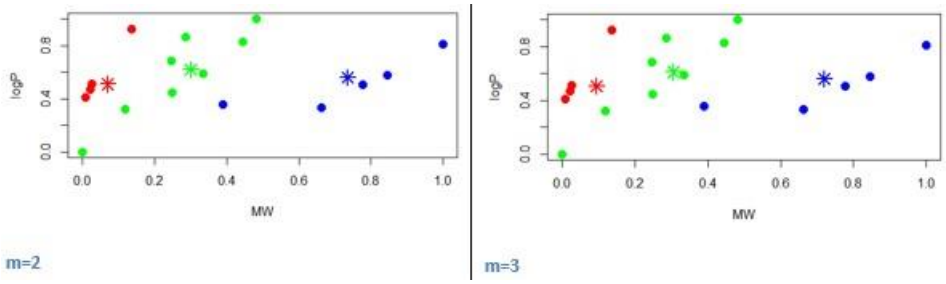

Figure 7. Cluster plots for memberships 2 and 3

Table 3. Membership degree of drugs to respective clusters using the PEMFKM algorithm

Object	Drug name	Cluster	Membership degree	Object	Drug name	Cluster	Membership degree
1	Paclitaxel	1	0.6643790	10	Eribulin	1	0.3942318
2	Everolimus	1	0.5010083	11	Palbociclib	2	0.5358908
3	Pamidronate	2	0.3618738	12	Gefitinib	2	0.5502121
4	Anastrozole	3	0.6132759	13	Ixabepilone	2	0.4650885
5	Exemestane	3	0.6011592	14	Tamoxifen	3	0.4509424
6	Epirubicin	1	0.4154320	15	docetaxel	1	0.7004234
7	Raloxifene	2	0.5215889	16	Lapatinib	2	0.4556951
8	Fulvestrant	2	0.3836216	17	Capecitabine	2	0.4317693
9	Letrozole	3	0.6818887				

Thus, the proposed PEMFKM algorithm can cluster 3 sets with clear differentiation when m=3 for the given Dataset, as shown in Figure 8. PEMFKM, like FKM, can produce three better cluster solutions. Thus, before trying allied algorithms, test all options.

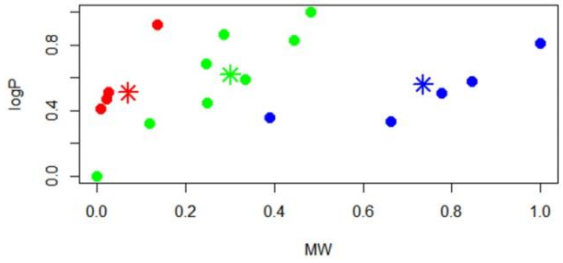

Figure 8. PEMFKM clustering algorithm could produce better clusters when m=3

5. CLUSTER ANALYSIS

Fuzzy clustering methods [26], [27] group related objects in the dataset to segment it into fuzzy partitions. Some external and internal validation indices are optional. Here are the internal validation indexes [28]: k is reached when PC is the greatest optimal number of clusters [26], [27].PE: for ideal clusters, PE should be minimum [26], [27]. MPC maximizes the clusters, *nk* [26], [27]. SIL: for the best number of clusters *nk*, SIL should be maximum [27], [29]. SIL.F: For the optimal number of clusters *nk*, the SIL.F index value must be maximal [27], [29]. The best number of clusters is *nk* when the XB is low [30].

The breast cancer medication dataset uses fuzzy algorithms like PEMFKM. Table 4 displays validation metrics values. Table 4 displays FKM results within limitations. Compared to PEMFKM and other fuzzy clustering algorithms, FKM-GK has higher validity index values for PC, PE, and MPC. PEMFKM has superior index values for SIL, SIL.F, and XB than the variation FKM-GK method.

Table 4. Various validation index values of fuzzy algorithms

Algorithm/ Parameter	FKM	FKM ENT	FKM ENT NOISE	FKM GK	FKM GK ENT	FKM GK ENT NOISE	PEMFKM
PC	0.4034	0.3333	0.3333	0.9999	0.3333	0.4569	0.4144
PE	0.9968	1.0986	1.0986	0.0000	1.0986	0.8563	0.9396
MPC	0.1051	0.0000	0.0006	0.9999	0.0000	0.1854	0.1073
SIL	0.3603	ND	ND	0.0038	ND	-0.2482	0.3726
SIL.F	0.4621	ND	ND	0.0038	ND	-0.0033	0.4852
XB	0.2041	1.1626	1.5915	2.7973	6.4494	5.7436	0.1921

Compare Figures 2-8 careful study of both graphs shows that PEMFKM's 3 clusters are more convincing than fkm's. PEMFKM separated clusters with well-defined data points better. And so, the drugs in the same cluster may have substantial connection qualities. Combining clustering medications for combination therapy is conceivable after chemical and enzyme research reveals their functional activity.

The PEMFKM technique is compared to several fuzzy clustering algorithms in Figure 9, using validation parameters PC, PE, MPC, SIL, SIL.F, and XB. Figure 9 demonstrate that PEMFKM's PC value is comparable to other index values. Figure 9 shows variation fuzzy algorithm SIL.F values. This graph shows that PEMFKM's SIL.F value is superior than FKM.GK and FKM.GK.ENT.NOISE methods and comparable to FKM. FKM, FKM.ENT, and FKM.ENT.NOISE algorithms have lower XB values than PEMFKM.

Figure 10 was drawn as per the values of Table 4. The preceding graphs demonstrate that the form-gk algorithm validity index values for the first three indices, PC, PE, and MPC, are within limitations for the PEMFKM method. PEMFKM passes SIL, SIL.F, and XB validity index tests better than FKM-GK. Chemical trials on breast cancer patients [31]–[34]. Our study clusters medications by physicochemical features to recommend combinational breast cancer treatment. Combinational therapy may use cluster medicines.

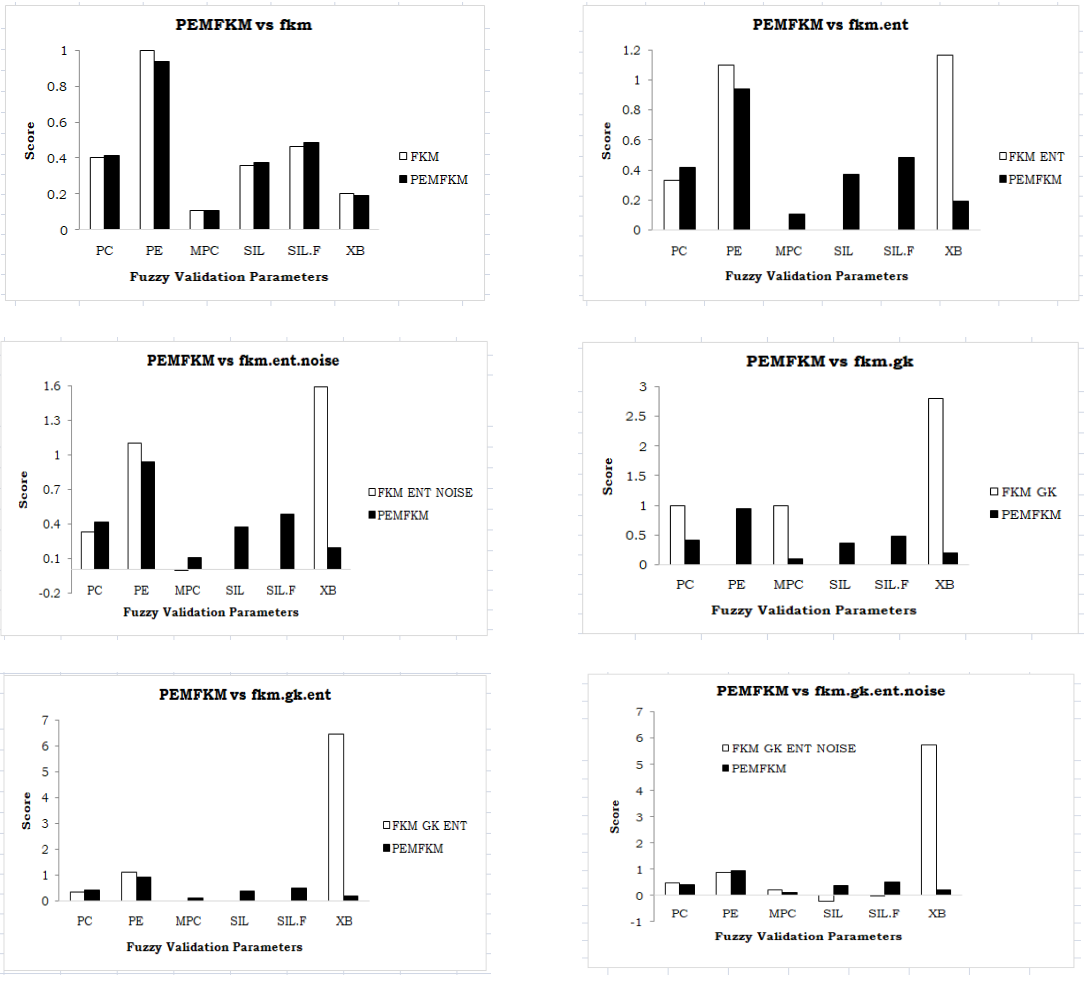

Figure 9. Various validation index values of various algorithms

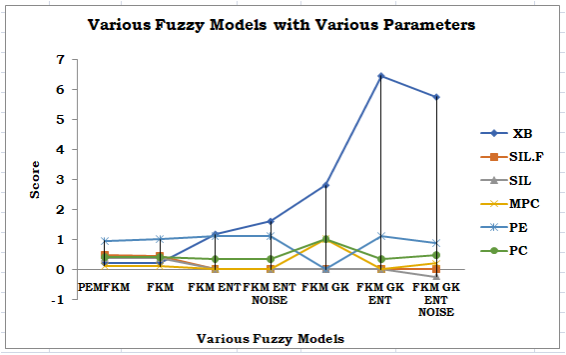

Figure 10. Graph shows the various parameters with various fuzzy models

6. DISCUSSION

Unsupervised learning of novel medications from correct world data is costly. Applying supervised learning to unsupervised process output can optimize software functioning. Clusters will automatically add new medication output parameters. The Drugs dataset is validated using the PEMFKM algorithm and various fuzzy clustering approaches. PEMFKM builds three clear clusters from these data and graphs and outperforms FKM-GK for validation indices SIL, SIL.F, and XB.

The graphs demonstrate that FKM-GK algorithm validity index values are within the bounds for the first three indices (PC, PE, and MPC) compared to PEMFKM algorithm values. PEMFKM outperforms FKM-GK in SIL, SIL.F, and XB validity index. Combinational chemotherapy improves efficacy, reduces medication dose, and delays illness start [31], [32]. Combinational breast cancer treatment reduces tumor development, cancer stem cells, and apoptosis [33].

7. CONCLUSION

The physicochemical parameters of the Selected Breast Cancer Drugs dataset are clustered using the PEMFKM algorithm and variant fuzzy clustering algorithms. The proposed PEMFKM algorithm constitute three distinct clusters. The existing clustering techniques and PEMFKM techniques are tested using cluster validation metrics such PC, PE, MPC, SIL, SIL.F, XB. PEFKM outperforms FKM-GK for SIL, SIL.F, and XB validation indices. Thus, the PEMFKM algorithm will build better clusters for the selected medications, allowing them to be paired for breast cancer combinational therapy.

In this work, how the selected brest cancer drugs are grouped to clusters are discussed, to acquire more accuracy in this grouping of objects to clusters, machine learning algorithms will be helpful. Another essential task is when a new drug to cure breast cancer arrives, a model or framework is needed to map it to the appropriate cluster. So, our future work will be improving the accuracy of clusters and mathematical model to map the newly approved drugs.

REFERENCES

[1] C. E. DeSantis et al., "Breast cancer statistics, 2019," *CA: A Cancer Journal for Clinicians*, vol. 69, no. 6, pp. 438–451, Nov. 2019, doi: 10.3322/caac.21583.

[2] P. Hallquist Viale, RN, MS, CNS, ANP, "The American cancer society's facts & figures: 2020 edition," *Journal of the Advanced Practitioner in Oncology*, vol. 11, no. 2, Mar. 2020, doi: 10.6004/jadpro.2020.11.2.1.

[3] J. A. Cruz and D. S. Wishart, "Applications of machine learning in cancer prediction and prognosis," *Cancer Informatics*, vol. 2, pp. 59–77, 2006.

[4] J. C. Holland, "Distress management," *NCCN Guidelines for Patients*, 2016. https://www.nccn.org/patients/guidelines/content/PDF/distress-patient.pdf (accessed May 08, 2016).

[5] N. E. Adler, *Cancer care for the whole patient: Meeting psychosocial health needs*. Washington, D.C.: National Academies Press, 2008. doi: 10.17226/11993.

[6] E. Grunfeld, D. C. Hodgson, M. E. Del Giudice, and R. Moineddin, "Population-based longitudinal study of follow-up care for breast cancer survivors," *Journal of Oncology Practice*, vol. 6, no. 4, pp. 174–181, Jul. 2010, doi: 10.1200/JOP.200009.

[7] C. Pitceathly, P. Maguire, I. Fletcher, M. Parle, B. Tomenson, and F. Creed, "Can a brief psychological intervention prevent anxiety or depressive disorders in cancer patients? A randomised controlled trial," *Annals of Oncology*, vol. 20, no. 5, pp. 928–934, May 2009, doi: 10.1093/annonc/mdn708.

[8] M. I. Nounou, F. ElAmrawy, N. Ahmed, K. Abdelraouf, S. Goda, and H. Syed-Sha-Qhattal, "Breast cancer: Conventional diagnosis and treatment modalities and recent patents and technologies," *Breast Cancer: Basic and Clinical Research*, vol. 9s2, Jan. 2015, doi: 10.4137/BCBCR.S29420.

[9] B. R. A. Cirkovic, A. M. Cvetkovic, S. M. Ninkovic, and N. D. Filipovic, "Prediction models for estimation of survival rate and relapse for breast cancer patients," in *2015 IEEE 15th International Conference on Bioinformatics and Bioengineering (BIBE)*, Nov. 2015, pp. 1–6. doi: 10.1109/BIBE.2015.7367658.

[10] T. Ayer, O. Alagoz, J. Chhatwal, J. W. Shavlik, C. E. Kahn, and E. S. Burnside, "Breast cancer risk estimation with artificial neural networks revisited," *Cancer*, vol. 116, no. 14, pp. 3310–3321, Jul. 2010, doi: 10.1002/cncr.25081.
[11] S. V. Appaji, R. S. Shankar, K. V. S. Murthy, and C. S. Rao, "Breast cancer disease prediction with recurrent neural networks (RNN)," *International Journal of Industrial Engineering & Production Research*, vol. 31, no. 3, pp. 379–386, 2020.
[12] K. Goyal, P. Sodhi, P. Aggarwal, and M. Kumar, "Comparative analysis of machine learning algorithms for breast cancer prognosis," in *Proceedings of 2nd International Conference on Communication, Computing and Networking*, 2019, pp. 727–734. doi: 10.1007/978-981-13-1217-5_73.
[13] G. P. Zhang, "Neural networks for classification: A survey," *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, vol. 30, no. 4, pp. 451–462, 2000, doi: 10.1109/5326.897072.
[14] M. Bouguessa, S. Wang, and H. Sun, "An objective approach to cluster validation," *Pattern Recognition Letters*, vol. 27, no. 13, pp. 1419–1430, Oct. 2006, doi: 10.1016/j.patrec.2006.01.015.
[15] L. G. Ahmad, A. T. Eshlaghy, A. Poorebrahimi, M. Ebrahimi, and A. R. Razavi, "Using three machine learning techniques for predicting breast cancer recurrence," *Journal of Health & Medical Informatics*, vol. 04, no. 02, 2013, doi: 10.4172/2157-7420.1000124.
[16] M. F. Akay, "Support vector machines combined with feature selection for breast cancer diagnosis," *Expert Systems with Applications*, vol. 36, no. 2, pp. 3240–3247, Mar. 2009, doi: 10.1016/j.eswa.2008.01.009.
[17] E. S. McDonald, A. S. Clark, J. Tchou, P. Zhang, and G. M. Freedman, "Clinical diagnosis and management of breast cancer," *Journal of Nuclear Medicine*, vol. 57, pp. 9S-16S, Feb. 2016, doi: 10.2967/jnumed.115.157834.
[18] S. H. Jafari *et al.*, "Breast cancer diagnosis: Imaging techniques and biochemical markers," *Journal of Cellular Physiology*, vol. 233, no. 7, pp. 5200–5213, Jul. 2018, doi: 10.1002/jcp.26379.
[19] S. Shankar Reddy, N. Pilli, P. Voosala, and S. Ravi Chigurupati, "A comparative study to predict breast cancer using machine learning techniques," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 27, no. 1, pp. 171–180, Jul. 2022, doi: 10.11591/ijeecs.v27.i1.pp171-180.
[20] R. S. Shankar, V. M. Gupta, K. V. S. S. Murthy, and C. S. Rao, "Breast cancer data classification using machine learning mechanisms," *Indian Journal of Public Health Research & Development*, vol. 10, no. 5, pp. 214–220, 2019.
[21] M.-S. Yang, "A survey of fuzzy clustering," *Mathematical and Computer Modelling*, vol. 18, no. 11, pp. 1–16, Dec. 1993, doi: 10.1016/0895-7177(93)90202-A.
[22] J. Zhou, L. Chen, C. L. P. Chen, Y. Zhang, and H.-X. Li, "Fuzzy clustering with the entropy of attribute weights," *Neurocomputing*, vol. 198, pp. 125–134, Jul. 2016, doi: 10.1016/j.neucom.2015.09.127.
[23] M. Zarinbal, M. H. Fazel Zarandi, and I. B. Turksen, "Relative entropy collaborative fuzzy clustering method," *Pattern Recognition*, vol. 48, no. 3, pp. 933–940, Mar. 2015, doi: 10.1016/j.patcog.2014.09.018.
[24] G. Grekousis and H. Thomas, "Comparison of two fuzzy algorithms in geodemographic segmentation analysis: The fuzzy C-means and gustafson–kessel methods," *Applied Geography*, vol. 34, pp. 125–136, May 2012, doi: 10.1016/j.apgeog.2011.11.004.
[25] C. Puri and N. Kumar, "An entropy based method for overlapping subspace clustering," *Procedia Computer Science*, vol. 122, pp. 276–283, 2017, doi: 10.1016/j.procs.2017.11.370.
[26] K.-L. Wu and M.-S. Yang, "A cluster validity index for fuzzy clustering," *Pattern Recognition Letters*, vol. 26, no. 9, pp. 1275–1291, Jul. 2005, doi: 10.1016/j.patrec.2004.11.022.
[27] A. Starczewski and A. Krzyżak, "Performance evaluation of the silhouette index," in *ICAISC 2015: Artificial Intelligence and Soft Computing*, 2015, pp. 49–58. doi: 10.1007/978-3-319-19369-4_5.
[28] Y. Liu, Z. Li, H. Xiong, X. Gao, and J. Wu, "Understanding of internal clustering validation measures," in *2010 IEEE International Conference on Data Mining*, Dec. 2010, pp. 911–916. doi: 10.1109/ICDM.2010.35.
[29] C. Subbalakshmi, G. R. Krishna, S. K. M. Rao, and P. V. Rao, "A method to find optimum number of clusters based on fuzzy silhouette on dynamic data set," *Procedia Computer Science*, vol. 46, pp. 346–353, 2015, doi: 10.1016/j.procs.2015.02.030.
[30] M. K. Pakhira, S. Bandyopadhyay, and U. Maulik, "Validity index for crisp and fuzzy clusters," *Pattern Recognition*, vol. 37, no. 3, pp. 487–501, Mar. 2004, doi: 10.1016/j.patcog.2003.06.005.
[31] T.-C. Chou, "Theoretical basis, experimental design, and computerized simulation of synergism and antagonism in drug combination studies," *Pharmacological Reviews*, vol. 58, no. 3, pp. 621–681, Sep. 2006, doi: 10.1124/pr.58.3.10.
[32] J. Foucquier and M. Guedj, "Analysis of drug combinations: current methodological landscape," *Pharmacology Research & Perspectives*, vol. 3, no. 3, Jun. 2015, doi: 10.1002/prp2.149.
[33] R. B. Mokhtari *et al.*, "Combination therapy in combating cancer," *Oncotarget*, vol. 8, no. 23, pp. 38022–38043, Jun. 2017, doi: 10.18632/oncotarget.16723.
[34] F. A. Fisusi and E. O. Akala, "Drug combinations in breast cancer therapy," *Pharmaceutical Nanotechnology*, vol. 7, no. 1, pp. 3–23, May 2019, doi: 10.2174/2211738507666190122111224.

BIOGRAPHIES OF AUTHORS

Vuddagiri Mutya Naga Sri Surya Venkata Krishna Rao Gupta holds Master of Science (M.Sc.), M.Tech Degree in Computer Science and Technology, Ph.D in Computer Science and Engineering, besides several professional certificates and skills. He is currently Professor in the Department of Computer Science Engineering at SRKR Engineering College, Bhimavaram. Andhra Pradesh, India. His research areas of interest include Data mining, Image Processing, Medical Mining and Machine Learning, He has written 30+ International Journals and conferences. He can be contacted at email: guptavkrao@gmail.com.

Chitta Venkata Phani Krishna Professor in CSE Department at Teegala Krishna Reddy Engineering College, Hyderabad has completed Ph.D. in 2013 from Acharya Nagarjuna University in the field of Software Engineering. His Research Area of Interest are – Data Science, Artificial Intelligence, and Block Chain. He has guided two scholars and currently guiding 6 scholars from various universities. He has 60 publications in reputed international journals. He is an author for three books and published 9 patents. He received 7 awards for his academic excellence. He can be contacted at email: phanik16@gmail.com.

Konakanchi Venkata Subrahmanya Srirama Murthy is an Assistant Professor in the Department of Computer Science and Engineering in Sagi Rama Krishnam Raju Engineering College, Bhimavaram, Andhra Pradesh, India. He received B.Tech and M.Tech Degrees from Andhra University, Visakhapatnam. He is pursuing Ph.D in Computer Science and Engineering at KLU. His Research areas are Machine Learning, Edge Computing, Natural Language Processing and Image Processing. He can be contacted at email: kvssrmurthy75@gmail.com.

Reddy Shiva Shankar is an Assistant Professor at the Department of Computer Science and Engineering in Sagi Rama Krishnam Raju Engineering College, Bhimavaram, Andhrapradesh, INDIA. He is pursuing PhD degree in Computer Science and Engineering with a specialization in Medical Mining and Machine Learning. His research areas are Image Processing, Medical Mining, Machine Learning, Deep Learning and Pattern Recognition. He published 50+ papers in International Journals and Conferences. S.S. Reddy has filed 05 patents. His research interests include Image Processing, Medical Mining, Machine Learning, Deep Learning and Pattern Recognition. He can be contacted at email: shiva.csesrkr@gmail.com.