An analysis of Human Metabolic changes in Covid-19 pandemic scenario with the use of Bio-Informatics data for Thyroid disease using Machine Learning System

Katikireddy Srinivas[1], Dr K V D Kiran[2]
{srinivas.katikireddy@gmail.com[1], kiran_cse@kluniversity.in[2]}

Research Scholar, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India[1], Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India [2]

Abstract. The Recent time the entire world had great experience of human mistakes with respect to environment safety and bio diversity leads to wonderful technological development along with danger symptoms of human life survival by increasing pollution levels to abnormal ranges.Covid-19 pandemic situation handling had shown the negative shade of technology development which ignores environment safety. At this pandemic Covid-19 Medical Emergency situation, Machine learning is inspired to develop an automated drug evaluation system for the patients suffering with routine illness conditions from the existing valid drug bank using a blend of clustering methods. The present pandemic disease covid-19 virus has explored the scope of Medical infrastructures to be improved world wide even in the 21st century. Tele medicine and digital health facilities to be incorporated using the Doctor, Patient, Network and relevant IT tools as the basic stakeholders to offer efficient, cost effective and better health facilities to common public. In this situation it is necessary to utilize software engineering information dependent on the physio synthetic properties and chemical hindrance properties of medication dataset given by standard medication bank storehouse derived from www.drugs.com, www.drugbank.ca and www.malacards.org. The existing grouping strategies are applied as a mix of k-implies, k-medoids, various leveled techniques and Fuzzy k-intends to decide a fitting arrangement of medications from the given drugbank for various disease states of thyroid patients. In this exploration work information groundwork for drug assessment is assuming a critical part, utilized Correspondence Analysis (CA) strategy to make our Cluster framework is productive and compelling. The outcome has shown the investigation as a mix of group techniques fruitful for this bunch framework utilizing graphical introduction and inferred best half and half bunch framework which creates proper medications set reasonable for different sorts of thyroid sicknesses. This machine learning system has shown better and appropriate drugs usage based on Physio Chemical Properties which effects human Metabolic changes in a controlled manner.

Keywords: Machine Learning, Medical Emergency, Covid-19, Telemedicine, Digital Health Clustering, Thyroid.

1 Introduction

Clustering algorithms has been categorized as Exclusive, Overlapping, Hierarchical and Probabilistic Clustering. In exclusive method, data assembled in an elite manner, and if a datum has a place with a clear bunch then it couldn't be incorporated into another group[1]. The clustering by overlapping method utilizes fuzzy sets to group information, so each point may have a place with at least two groups with various degrees of participation. The most for the most part announced and regularly used isolating methods are k-implies, k-medoids, and different assortments.

For K-medoids, a medoid speaks to the most delegate purpose of a gathering of focuses. K-Means clustering [2] is also an iterative clustering procedure, but it predefines the number of clusters that will be in the dataset. PAM stands for "partition around medoids" [3]. The method intends to discover an arrangement of items called k-medoids that are halfway situated in groups. The objective of the algorithmic method is to reduce the object dissimilarities with respect to their nearby selected datum. The structure of k-medoids is nearly similar to that of k-means[4]. The cluster representative is the one data point which is located central in the cluster. Any two objects distance is calculated and the one having minimum dissimilarity when compared to all other objects is chosen as the center.PAM is susceptible but tough to noise as well as outliers than kmeans because k-medoids contemplates marginal distance which isolates it from alternate objects[5].

The observation being classified into groups necessitates few methods for measuring the distance between observations, which means no unsupervised machine learning algorithms can take place without some notion of distances. The selection of distance measure is crucial step in clustering [6]. It characterizes how the likeness of two components (x, y) is computed and it will impact the state of the groups. The most generally utilized and acknowledged technique is Euclidean separation measure. The estimation of separation measures is personally identified with the scale on which estimations are made. Therefore, factors are frequently scaled (i.e. standardized) before estimating the dissimilarities [7]. Generally variables are scaled to have standard deviation one and mean zero. The goal is to make the variables comparable and they will have equal importance in the clustering algorithm. This is especially prescribed when factors are estimated in various scales. The standardized data is a methodology broadly utilized with regards to gene examination before grouping [8].This work presents a hybridized program encompassing both k means and k-medoids algorithm to cluster a dataset of thyroid disease drugs and the program is run to generate data groups based on the algorithm, thereby refining the outcome based on fuzzy k-means.

2 Materials and Methods

2.1 Dataset: Nearly 189 drugs as dataset was utilized where they are reported as thyroid inhibitors, downloaded from Malady cards database [9].It was observed that few drugs come under other disease conditions; however, involved in the dataset because they are known to representation several other diseases including thyroid disease.

2.2 R: Result of information in science understood the criticalness of information mining in the structure of convolution of bio frameworks[11].R program is uninhibitedly accessible programming accessible in a domain utilizing object arranged programming for the most part focused for factual figuring just as designs.

2.3 Hybrid clustering

K-suggests pack technique is least requesting and the most exhaustively utilized parceling framework for segment a dataset into an arrangement of k groups. The system utilizes Euclidean separation evaluates between server farms to pick within and the between-bundle tantamount characteristics[12]. The PAM estimation depends upon the quest for k administrator objects or medoids among the perspective on the dataset. These acknowledgments should address the structure of the information. In the wake of finding a strategy of k medoids, k bundles are made by assigning every wisdom to the closest medoid.

3 Results and Discussion

3.1Cluster Validity: In this work, NbClust was engaged which was integrated with 30 validity indices to examine the numbers of groups in a given dataset. Therefore, from this analysis, the outcome signified that about 13 different index programs suggested three clusters as optimum whereas eleven index programs suggested two groups and 4 indices reported four clusters. As per the majority ruling method, the best output referred to three groups [8].

Hence, it can be established that the optimum numbers of clustering groups, k for the given dataset comprised of various drugs involved in thyroid disease was three cluster results. So, an initial k=3 value was utilized to achieve k-means, PAM as well as hybrid algorithm on the thyroid dataset [13].

3.2 k-means algorithm: The k-means approach is an apportioning issue, wherein the information isolated as gatherings with each redundancy of the calculation [10]. Since the assignments were begun aimlessly, n start = 25 is indicated, which implies that the program will endeavor 25 different arbitrary beginning stages and afterward picked the outcome with lowermost inside bunch disparity(Figure 1).A better group will bring about qualities with least withinss and greater betweenss which further depends on the whole of k groups chose initially. From this time forward, low withinss and high betweenss for k=3 was gotten.
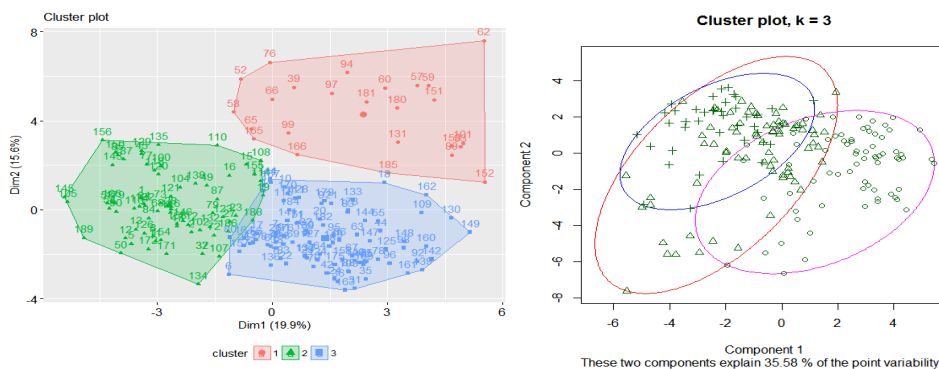
Figure. 1: Output of 3 clusters and cluster centers obtained from kmeans program and K-medoids.

3.3 Partitioning Around Medoids : It was reported that outliers influence the outcome of k-means cluster result which would otherwise influence the task of cluster annotations. Hence, a novel strong algorithm is presented by PAM algorithm, also referred as k-medoids [14].

From both methods, it was observed that some samples have a negative silhouette. This means that they are not in the right cluster. On contrast between k-means versus PAM, k-

means ensued silhouettes around 13 which are negative whereas PAM resulted in 27, respectively.

3.4 Hybrid k means-PAM Clustering Algorithm

Before long, these two standard gathering methods have their own one of a kind inclinations and imprisonments. In this manner, a novel hybrid methodology is executed to mix the best of k-means and PAM grouping. proceeds in three stages. First it figures k beginning medoids as k-bunches on the hidden dataset. By then the PAM bunch centers are determined trailed by handling k-implies by using group focuses as the underlying k [10].

The 3 grouping came about gatherings procured using k=cluster focuses, which are the three gatherings of PAM computation, achieved absolutely three different bundle sizes 77, 25 and 87 independently. Inquisitively, the negative blueprints got from blend system are 11 against 13 from k-suggests alone and 27 by PAM procedure, which prescribes the way that crossbreed method is valuable in expelling information from groups (Table 1).

Table 1: Three cluster groups appeared from kmeans and PAM methods.

	Group 1	Group2	Group3	Size of Hybrid cluster
Group1	59	18	0	**77**
Group2	9	14	2	**25**
Group3	4	44	39	**87**
Size of PAM cluster	**72**	**76**	**41**	

It is worth to take note of that the individual k-means calculation could group dataset as 25, 70 and 94 gatherings while the hybrid k-means_k-medoids brought about comparable cluster of size 25 and remaining being 87 and 77.
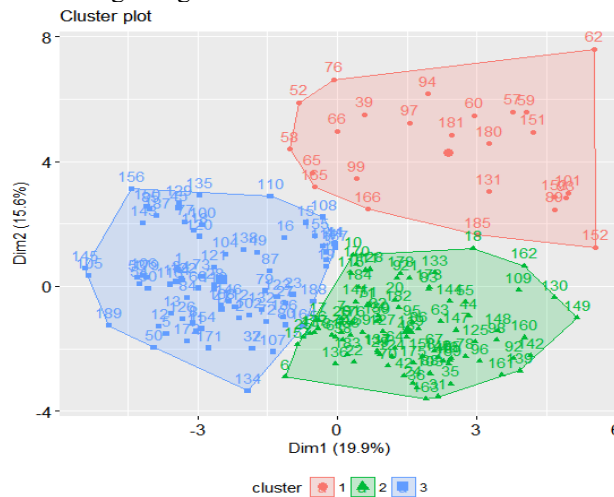

Figure 2: Visual representation of 3 clusters and centers resulted from hybrid k-means_k-medoids method.

It was observed from the plots of hybrid clustering algorithm that the data points at the edge of clusters 2 and 3 were found to be overlapped and efficient clustering was not possible when plots are visualized. Hence, work was initiated to introduce k-means procedure coupled

with fuzzy algorithm [8]. Fuzzy being soft clustering procedure mixed with hard clustering k-means, reported as fuzzy k-means (FKM) algorithm in order to produce meaningful clusters.

3.5 Fuzzy k-means algorithm: A subset including 25 information focuses from the 189 thyroid medication parameter dataset was exposed to fkm calculation and the yield chart is accounted for in Figure 3. It is confirm that the program can bunch 3 sets with clear division.
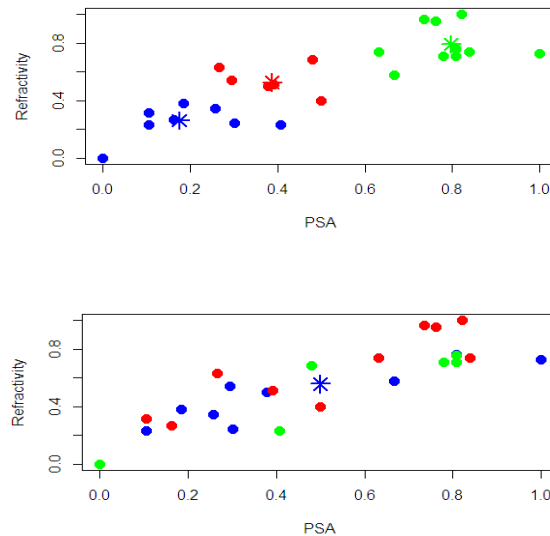
Figure 3: Dataset clusters via FKM algorithm with Entropy regularization.

3.6 Fuzzy k-means via entropy regularization

An energizing stuff in regards to the fluffy k-implies by means of entropy regularization is that the models are gotten as weighted methods with loads proportionate to the enlistment degrees (rather than to the cooperation degrees at the force of m as is for the fluffy k-implies). It is seen from Figure 5 that couple of articles from one bunch showed up in other bunch gatherings.

3.7 Fuzzy k-means via entropy regularization plus noise cluster: The entropy regularization abstained from utilizing the fake fluffiness parameter m. The clamor group is an extra bunch (concerning the k standard groups) to such an extent that items perceived to be anomalies are allotted to it with high enrollment degrees [8].

3.8 Gustafson and Kessel-like fuzzy k-means: The program plays out the Gustafson and Kessel-like soft k-suggests packing computation and is worthwhile to choose gatherings.

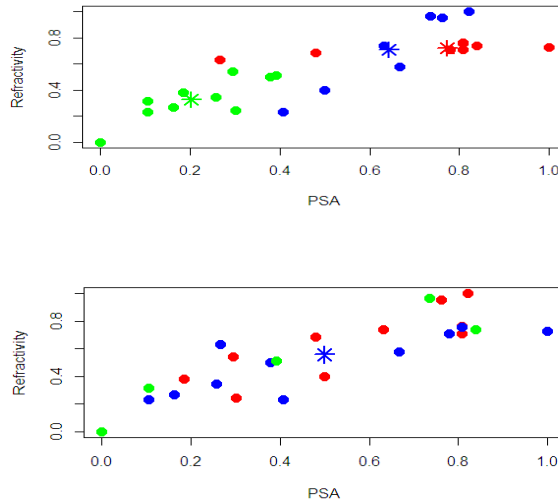**Figure 4:** Gustafson and Kessel - like fuzzy k-means clustering algorithm with Entropy Regularization

3.9 Gustafson and Kessel-like fuzzy k-means via entropy regularization: The program performs the Gustafson and Kessel - like fuzzy k-means clustering algorithm with entropy regularization [21]. The method permits to evade utilizing the artificial fuzziness parameter m. If standardization is set to stand=1, the algorithm runs based on standard data. Figure 5 suggested that the data was discrete and the program unable to identify and cluster better possibilities.

3.10 Gustafson and Kessel-like fuzzy k-means using entropy regularization plus noise cluster: The program runs the Gustafson and Kessel-like fuzzy k-means clusters using entropy regularization and noise cluster which is different from fuzzy k-means, and the method identifies non-spherical clusters.

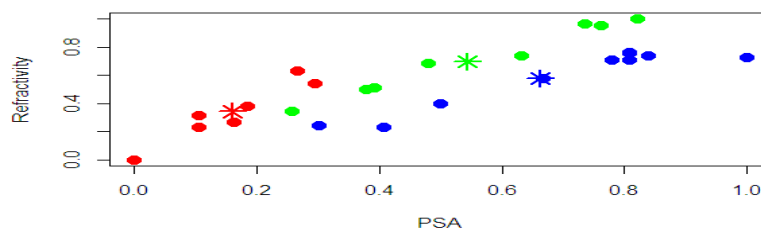**Fig. 5:** Gustafson and Kessel - like fuzzy k-means clustering algorithm with entropy regularization and noise cluster resulted in better clusters.

Of all variations in FKM algorithms presented here, only natural FKM algorithm is able to produce estimated three better cluster solutions. Hence, it should be noted that testing all possibilities should be made before proceeding with allied variations of algorithms.

4 Conclusion

From both individual k-means and k-medoids strategies, it was seen that a few examples announced negative outlines. On correlation between k-means and PAM, the previous brought about 13 negative outlines though PAM technique brought about 27 negative outlines and comparable is the perception with bunch size. In addition, covering of bunches was seen for each situation just as in half breed strategy. Consequently, a lot of six fluffy calculation variations concentrated on a subset of thyroid dataset brought about 3 unmistakable groups by fluffy k-implies followed by Gustafson and Kessel - like fluffy k-implies with entropy regularization and commotion bunch calculation[8]. Table 2 shows the drugs suitable for Hyper Thyroid and Hypo Thyroid patients derived from the mappings of this novel cluster system [10].

Table 2: Cluster System Output – Drugs data set divided into 3 Cluster groups as shown below (Hybrid K-Means K-Medoids Algorithm)

S. No	Drug_ID	Cluster 1 DRUGS	S. No	Drug_ID	Cluster 2 Drugs	S. No	Drug_ID	Cluster 3 Drugs
1	1	Nitroprusside	1	39	Methotrexate	1	6	Sevoflurane
2	2	Propylthiouracil	2	52	Folic Acid	2	7	Etoricoxib
3	3	Hydrocortisone	3	57	Paclitaxel	3	9	Remifentanil
4	4	Prednisone	4	58	Pemetrexed	4	10	Methylprednisolone
5	5	Nitric Oxide	5	59	Everolimus	5	17	Diclofenac
6	8	Propofol	6	60	Sirolimus	6	18	Travoprost
7	11	Menthol	7	62	Octreotide	7	20	Bimatoprost
8	12	Acetylcholine	8	65	Doxycycline	8	21	Latanoprost
9	13	Benzocaine	9	66	Oxytetracycline	9	22	Dexmedetomidine
10	14	Dexamethasone	10	76	Leucovorin	10	24	Tizanidine
11	15	Triamcinolone	11	89	Indinavir	11	25	Clonidine
12	16	Dinoprost Tromethamine	12	93	Ritonavir	12	26	Methyltestosterone
13	19	Timolol	13	94	Azithromycin	13	29	Bupivacaine
14	23	Metoprolol	14	97	Doxorubicin	14	30	Celecoxib
15	27	Estradiol	15	99	Minocycline	15	31	Sertraline
16	28	Guaifenesin	16	101	Vincristine	16	33	Morphine
17	32	Pseudoephedrine	17	131	Etoposide	17	35	Fentanyl

18	34	**Metformin**	**18**	151	**Docetaxel**	**18**	36	**Diazepam**
19	37	**Ephedrine**	**19**	152	**Vinorelbine**	**19**	38	**Xylometazoline**
20	41	**Betamethasone**	**20**	165	**Cefazolin**	**20**	40	**Donepezil**
21	45	**Ribavirin**	**21**	166	**Piperacillin**	**21**	42	**Citalopram**
22	47	**Dopamine**	**22**	180	**Pimecrolimus**	**22**	43	**Exemestane**
23	49	**Entecavir**	**23**	181	**Tacrolimus**	**23**	44	**Amiodarone**
24	50	**Cysteamine**	**24**	185	**Bromocriptine**	**24**	46	**Olanzapine**
25	53	**Niacin**	**25**	153	**Vinblastine**	**25**	48	**Risperidone**
26	54	**Nicotinamide**				**26**	51	**Droperidol**
27	68	**Carboplatin**				**27**	55	**Imatinib Mesylate**
28	72	**Carbimazole**				**28**	56	**Chlorpheniramine**
29	73	**Epinephrine**				**29**	61	**Miconazole**
30	77	**Alendronate**				**30**	63	**Sorafenib**
31	79	**Lacosamide**				**31**	64	**Bexarotene**
32	80	**Carbamazepine**				**32**	67	**Vildagliptin**
33	84	**Caffeine**				**33**	69	**Levonorgestrel**
34	85	**Vitamin C**				**34**	70	**Ropivacaine**
35	87	**Genistein**				**35**	71	**Topiramate**
36	90	**Acetaminophen**				**36**	74	**Esomeprazole**
37	100	**Carbidopa**				**37**	75	**Norgestimate**
38	102	**Cyclophosphamide**				**38**	78	**Ezogabine**
39	104	**Vorinostat**				**39**	81	**Melatonin**
40	105	**Hydroxyurea**				**40**	82	**Ergocalciferol**
41	106	**Fluorouracil**				**41**	83	**Vitamin E**
42	107	**Histamine**				**42**	86	**Cholecalciferol**
43	108	**Bortezomib**				**43**	88	**Midazolam**
44	110	**Fludarabine**				**44**	91	**Lansoprazole**
45	112	**Chlorpropamide**				**45**	92	**Sunitinib**
46	113	**Salicylic acid**				**46**	95	**Loperamide**
47	114	**Glipizide**				**47**	96	**Rosiglitazone**
48	116	**Tolbutamide**				**48**	98	**Ezetimibe**

49	118	Zoledronic acid				49	103	Nitisinone	
50	120	Decitabine				50	109	Dabrafenib	
51	121	Dacarbazine				51	111	Glimepiride	
52	122	Temozolomide				52	115	Glyburide	
53	123	Melphalan				53	117	Desogestrel	
54	126	Valproic Acid				54	119	Pioglitazone	
55	129	Azacitidine				55	124	Tamoxifen	
56	132	Ifosfamide				56	125	Gefitinib	
57	134	Mechlorethamine				57	127	Drospirenone	
58	135	Vidarabine				58	128	Capecitabine	
59	138	Gemcitabine				59	130	Irinotecan	
60	143	Levodopa				60	133	Trametinib	
61	145	Glycine				61	136	Diphenhydramine	
62	146	Tretinoin				62	137	Promethazine	
63	150	Tagatose				63	139	Crizotinib	
64	154	Metronidazole				64	140	Rabeprazole	
65	155	Famotidine				65	141	Lenalidomide	
66	156	Mannitol				66	142	Ponatinib	
67	164	Cocaine				67	144	Calcitriol	
68	167	Ursodeoxycholic acid				68	147	Alfacalcidol	
69	168	Ketorolac				69	148	Tipifarnib	
70	171	Isoflurane				70	149	Lapatinib	
71	172	Desflurane				71	157	Vitamin A	
72	174	Aspirin				72	158	Vatalanib	
73	179	Vigabatrin				73	159	Veliparib	
74	186	Pentoxifylline				74	160	Erlotinib	
75	187	Glucosamine				75	161	Hydroxychloroquine	
76	188	Thiamine				76	162	Bosentan	
77	189	Choline				77	163	Ketamine	
						78	169	Reboxetine	
						79	170	Cortisone acetate	
						80	173	Rocuronium	
						81	175	Anastrozole	
						82	176	Letrozole	

					83	177	**Tropicamide**
					84	178	**Tauroursodeo xycholic acid**
					85	182	**Cyproterone Acetate**
					86	183	**Doxepin**
					87	184	**Lovastatin**

.

References

[1] Hastie T, Tibshirani R, Friedman J. Unsupervised learning. In The elements of statistical learning 2009 (pp. 485-585).Springer New York

[2] www.drugbank.ca

[3] Halkidi M, Batistakis Y, Vazirgiannis M. On clustering validation techniques.Journal of intelligent information systems. 2001 Dec 1;17(2):107-45

[4] http://www.malacards.org/

[5] Banerjee, A. (2004). "Validating clusters using the Hopkins statistic". IEEE International Conference on Fuzzy Systems: 149–153

[6] Ferraro M.B., Giordani P., 2013. A new fuzzy clustering algorithm with entropy regularization. Proceedings of the meeting on Classification and Data Analysis (CLADAG)

[7] Chris Fraley, A. E. Raftery, T. B. Murphy and L. Scrucca (2012)." mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation",Technical Report No. 597, Department of Statistics, University of Washington.

[8] Chris Fraley and A. E. Raftery (2002). "Model-based clustering, discriminant analysis, and density estimation", Journal of the American Statistical Association 97:611:631.

[9] http://www.sthda.com/english/wiki/factoextra-r-package-quick-multivariate-data-analysis-pca-ca-mca-and-visualization-r-software-and-data-mining

[10] Gustafson E.E., Kessel W.C., 1978,"Fuzzy clustering with a fuzzy covariance matrix", Proceedings of the IEEE Conference on Decision and Control, pp. 761-766

[11] www.drugs.com

[12] "The art of R Programming", Norman Matloff, William Pollack, 2013,I edition .

[13] Data mining Concepts and Techniques, Han & Kamber, Morgan Kaufmann(Elsevier)

[14] www.sthda.com

[15] www.datanovia.com