

# A Deep Reinforcement Learning for Medical Trajectory Data Clustering Type 2 Diabetes

Arunkumar K (✉ [arunjntua@gmail.com](mailto:arunjntua@gmail.com))

JNTUA CEA: Jawaharlal Nehru Technological University Anantapur College of Engineering  
Ananthapuramu

Vasundra S

JNTUA CEA: Jawaharlal Nehru Technological University Anantapur College of Engineering  
Ananthapuramu

---

## Research Article

**Keywords:** Deep Learning, Multi-Attributes, Trajectory data, Type 2 Diabetics

**Posted Date:** September 13th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-888442/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# **A Deep Reinforcement Learning for Medical Trajectory Data Clustering Type 2 Diabetes**

**K.Arun Kumar**, Research Scholar, Dept of CSE, JNTUA, Ananthapuramu - 515002, India.  
[arunkumar.kanagana@gmail.com](mailto:arunkumar.kanagana@gmail.com) (corresponding author)

**S. Vasundra**, Professor, Dept of CSE, JNTUACEA, Ananthapuramu - 515002, India.  
[Vasundras.cse@jntua.ac.in](mailto:Vasundras.cse@jntua.ac.in)

Abstract: Deep Reinforcement learning is incorporated in trajectory data clustering to investigate the trajectories gathered from medical information's. Generally Trajectory mining determines the patterns in data, detects anomalies, and does informative clustering, location prediction, and classification. The main intent of Medical trajectory data clustering is identifying the trajectories with identical patterns for better patient treatment outcomes. Medical trajectory data stored in a multidimensional format which is further processed using the machine learning and deep learning architectures. Machine learning approaches employed to mine trajectory data and identifying the future treatment is a complicated task. To deal with this, the deep learning approaches in trajectory mining concentrate to eliminate the computational complexity on type 2 diabetic's data. To overcome this problem, deep reinforcement learning for medical trajectory data clustering approach is proposed that is a combination of various strategies to flexible adapt to changes of the trajectory data. After the proposed pre-processing and feature transformation, features are clustered on basis of the weights of the model with lesser efforts and the proposed clustering plays a key role in the process of multi-attribute trajectory data investigation. The proposed deep learning methodology is more suitable for clustering the multi-attribute trajectory with fewer complexity computations than existing machine learning based methods. The experimental results also states that the results of deep reinforcement learning are promising than the other approaches with respect to precision, Recall and F Measure respectively.

Keywords: Deep Learning, Multi-Attributes, Trajectory data, Type 2 Diabetics

## **1. Introduction**

A trajectory is a collection of ordered information about moving objects. Each point of a trajectory represents the attributes of a moving object or on-going event and it generally has time as well as location attributes. Processing or analysing of trajectory data provides rich knowledge that is used to make decisions regarding the context where the information is collected. Recent trajectory data have added dimensions in addition to time

and location attributes. This added dimensionality is challenging while performing the mining process on such data and needs additional efforts.

When a trajectory data source has more than two attributes, it is called multi-attribute trajectory data. The trajectories with multiple attributes provide additional information regarding the data context in addition to primary attribute information like location and time. The additional attributes provide more contexts to a moving object or entity. If this additional data is analysed along with the added attributes diverse knowledge about the objects under study can be acquired. Most of the previous work concentrated on two-attribute trajectory data where most application contexts are related to vehicle moving data, transportation, and traffic analysis. Today many sequencing datasets relating to medical diagnostics, money markets, and more have a trajectory nature. These types of trajectory data have high potential to proceed with and can provide rich knowledge when it is cleverly mined.

Similarity metrics play a vital role in data grouping or clustering. There are varieties of similarity measures for normal data clustering which may not be extended for trajectory data clustering. A comparison of two trajectories needs the set of points in the trajectory, the order of these points, and the dimension set of each point. When a comparison is made about each of these aspects the comparison provides the similarity among the trajectories. Based on the similarity the trajectory data can be analysed to know the behaviour of the objects represented by the trajectories. The recent past research on trajectory data similarity suggested kinds of trajectory similarity measures where each type of similarity measure has its limitations. In this work, better measures of trajectory similarity and a methodology for trajectory clustering is proposed. Designing of a finer clustering strategy that can deal with multi-attribute trajectory data is the primary goal of this work.

Most of the previous research on trajectory data analysis and mining concentrated on data with two dimensions that represent Spatio-temporal information. It is possible to have more dimensional information of a trajectory point like the colour of an object, speed of a vehicle, blood pressure of a person, heart rate of a person, and temperature of a moving object, and so on. Consideration of such additional dimensional issues into trajectory data analysis and mining will provide useful knowledge for decision making concerning the data context. Dealing with the additional dimensions of multi-attribute trajectory is a new challenge concerning storage, computation, and interpretation. Therefore, there is a need to develop models to deal with multiple attributes of trajectory

data.

To deal with multi-attribute trajectories with less effort than the researchers' previous experiments, this Work attempts a point-by-point study of trajectory data. In this proposed approach, the deep learning architecture based on reinforcement learning has been employed to cluster the trajectory medical data on basis of future outcome of the disease and treatment. To find estimate the effectiveness of the clustering of the trajectory data, Euclidean distance is computed on the outcomes of the cluster. Collection trajectories can be of uniform or non-uniform length. Some grouping models insist on uniform length [1]. In this paper trajectory data with uniform length is considered.

An empirical data about diabetic patients is collected. Laboratory diagnosis and medical demographics of patients over a period formed as trajectories. Data like age and dietary control level of every individual diabetic patient is concentrated in this work.

### **Key Contribution**

- Introduction of trajectory marking by grouping data points from all trajectories, which liberalizes the complexity of a set of matches by multiple attributes.
- Proposal of deep reinforcement learning trajectory data that can provide more useful grouping information. The problem of computation complexity is resolved, and it provides results with higher accuracy and efficiency.
- Mining of a significant medical trajectory data to predict the future outcome of the disease is an interesting insights about the patients of diabetes to guide treatment alternatives.

This paper is organized into five sections. Followed by this introductory section, the literature review is given in the next section. The third section detailed about the proposed methodology with assumptions made for the modeling. In the fourth section a description of dataset used in this work followed by results and discussion is presented. The final section concluded the total work and significance of this experimental study.

## **2. Literature Review**

Trajectory data analysis and mining playing a vital role in analyzing various real time data contexts. In trajectory data mining similarity measures play a vital role. Many models suggested in literature require unified trajectories with little information loss due to fusion. Euclidean distance is such a measure for calculating the distance between two unified trajectories [2]. A larger measure that requires uniformity in length trajectories is the

Bhattacharya distance. It is based on the distribution of the sample data [3]. Some other distance measurements that do not need the same length include the Hausdorff distance, Frechet distance, LCSS distance and DTW distance [4] [5]. In [6] the authors introduced the concept of statistical discriminate sub-trajectories, in which the sub-trajectory is discriminatory with others, if it is similar to many sub-trajectories of one class and different from the sub-trajectories of the other group. Trajectory similarity measurement techniques fall into two categories [7] [8]. To deal with a similarity search problem containing multiple attribute some authors [9][10] proposed a tree index structure embedded with large-scale and shift operations which are faster than the standard multidimensional data structure .It is also needed to combine trajectories and their corresponding semantic representations [11]. For data grouping, a density and network-based

technique has been proposed [12], in which the trajectories are divided into segments using an angular difference and an algorithm with a minimum length of description. For trajectories of unequal length, a distance measure called distance edition [13] has been proposed, in which interpolation for corrections is applied. Efforts have also been made to find residues in the trajectory data and grouping for recommendations [14][15][16][17]. A density and density-based network has been proposed to group data [18] in which the trajectories are segmented using angular difference algorithm and minimum description length. More efforts on trajectory data mining can be found in [19] [20][21][22].

Nature of input, classes of data, dimensionality, processing time, and accuracy are some of the main issues to deal with in data [23]. Most of the existing measures of trajectory similarity worked for two main location and time and limited to vehicle movements. There may be more measured trajectory information, such as the color of the object, vehicle speed, human blood pressure, human heart rate, and temperature of the moving object, etc. Addressing the additional dimensions of the multi-attribute trajectory is a new challenge in terms of storage, computation, and interpretation. Existing machine learning and data extraction algorithms for data trajectory analysis are somewhat limited. Therefore, models need to be developed to handle multiple attributes of trajectory data. Most of the trajectory analysis and performance studies have focused on two-dimensional data representing space-time information.

### **3. Research Methodology**

#### **3.1 Problem Statement**

Data relating to vehicle movement, share market, periodic medical reports, and customer buying patterns, are some examples of multi-attribute trajectory data. Multi-attribute trajectory data requires a lot of computational effort to process, store, and analyze the data, where every trajectory data is a combination of multiple information stored in multiple dimensions. The study in the recent and past literature has shown various procedures for dealing with the trajectory data and they mainly focus on time and space. In the real-world, the trajectory attribute can represent semantic issues also besides time and space. Additional dimensions in trajectory analysis certainly provide useful insights for decision-making regarding the perspective of the data. At the same time higher dimensions leads to higher complexity in studying and processing.

Processing of trajectory data with deep learning architecture is the main intent of

the methodology proposed in this paper. The proposed technique uses a trajectory clustering after pre-processing and feature extraction for trajectory grouping.

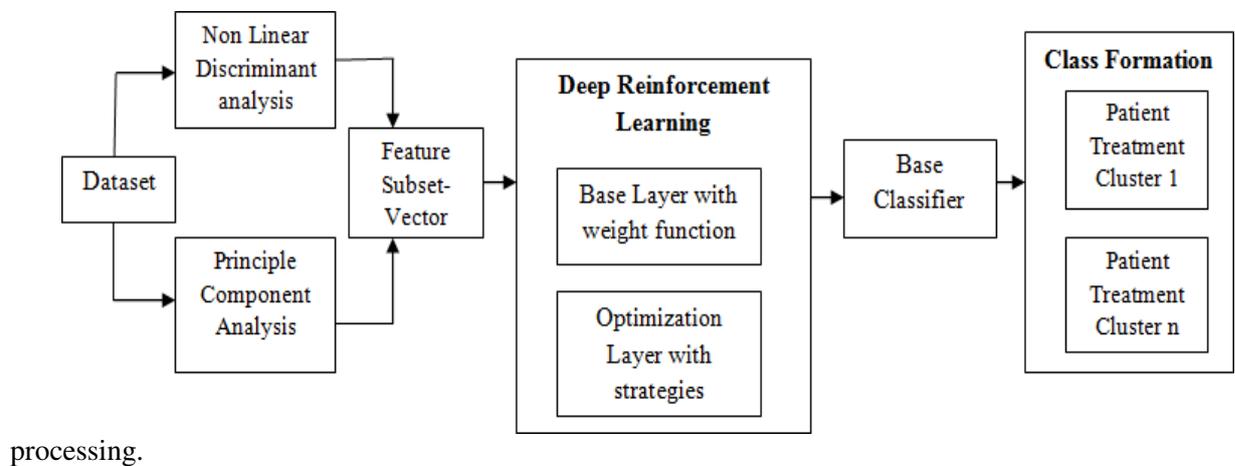
### **3.2 Deep Reinforcement Learning – Trajectory Clustering**

A two-level similarity approach is used to complete the trajectory analysis and grouping. In the pre-processing phase, points of all trajectories will undergo a grouping process where a set of similar points are assigned with a common label. Every point is labelled according to the group and the labelling procedure reduces the complexity of data processing, which are in the form of multi-dimensionality. Then the labelled trajectories are more flexible to measure similarities among them. Using this deep reinforcement learning, future outcome of the disease is effectively computed on the collected medical trajectory dataset.

Suppose in a trajectory dataset, there is 'N' number of trajectories where each data point of a trajectory has 'D' dimensions where the number of points in a trajectory is 'M' and the value of M is the constant for all trajectories. The primary goal of the proposed methodology is to suggest a future treatment pattern among trajectories of patient treatment. Then a trajectory dataset is grouped into meaningful clusters using the strategies of the reinforcement learning on basis of weight function.

### **3.3 Data Pre-Processing**

In the pre-processing phase, trajectories in the dataset will undergo a scheme of grouping process where a set of similar points are assigned with a common label. At first, all points of all trajectories undergo a clustering process which provides clustering of points. At this stage, every point represents a trajectory and a cluster. Using this relationship each point is labeled with the respective cluster name. Now the trajectory data have labeled trajectories. The conversion of trajectory information with multidimensional points on trajectory information is represented with a point cluster. Clustering provides the trajectory data in labelled sequences and the labelling makes the trajectory data more computable for similarity finding than the raw point information. The complexity of multi dimensionality of trajectory points comes down with this labelled and clustered pre-



**Figure 1: Proposed Architecture**

### 3.4 Clustering

Deep reinforcement learning is employed to learn the discriminant information in the feature vector to cluster the patient treatment trajectories. The model is capable of handling mapping of the trajectories in the form of vector with requiring adaptation of the weight function for cluster generation composing the strategies. Strategies based model is developed to determine cluster instance to the trajectory formed. Non linear learning finds the optimal solutions to the patient trajectory data on basis of action selection strategies.

Patient Treatment prediction function for feature vector is given by action selection strategies is

Where State of the instance is  $S_t$

$E$  is the expected instance of the trajectory and  $r$  is a reward function

$$NL(s_t, a_t) = T(s_t, a_t) + \alpha [w + \gamma \max_a T(s_{t+1}, a) - T(s_t, a_t)]$$

Prediction function is delay scalar feedback and it determines the shortest path

$$NL(s_t, a_t) = E \{ w_{t+1} + \gamma w_{t+2} + \gamma^2 w_{t+3} + \dots \}$$

on basis of state action pair on NN to approximate. Further instance is a probability of being selected that is directly proportionate to their fitness

### 3.5 Base Classifier using Reward function

Base classifier on trajectory learning using reinforcement learning will be established using weighted aggregation of majority trajectory data [13]. In this base learner maintains the weight feature vector which combines linearly with local cluster and global cluster prediction vector composed of the feature instance of the cluster using reinforcement learning on Non Linear learning. Weighted majority scheme determines the base learner to particular type of instance cluster to determine the treatment on basis of

$$NL(s, a) = T(s, a) + \alpha [W + \gamma \max_{a'} T(s', a') - T(s, a)]$$

patient. Treatment determination is as follows

#### Algorithm 1: Deep Reinforcement learning

Input: Trajectory Instance of the medical data  $(x_t, y_t)$  at time  $t$ , Cluster Set  $C_t$ , Feature subset  $F_t$  Output: Trajectory formation

Process Initialize  $T(s, a)$

arbitrarily Repeat (for each

patient):

Initialize  $s$

Repeat (for each step of feature instance  $F_t$ ):

Choose  $a$  from  $s$  using action selection strategies derived from *Non linear weight* for function

$$NL(s, a) = T(s, a) + \alpha [r + \gamma \max_{a'} T(s', a') - T(s, a)]$$

$s \leftarrow s'$ ;

Until  $s$  is terminal

Assign the Label  $C_E$  to the feature instance  $F_t$  as treatment option for the patient.

## 4. Experimentation, Results and Discussion

### 4.1 Data collection plan

Time-series /trajectory of diabetic data is represented as a tuple of diabetic patient profile. Interval of diabetic test is usually irregular, as it depends on the condition of a patient. However, in the progression of multistage matching, it is essential to signify time-series as a set of data points with a persistent interval to signify the time span by the amount of data points.

The examination data of a patient is considered with TS different time series of investigation and the time –series is signified by  $t_i^{\text{th}}$  examination by  $in_i(TS)$ , where  $t_i$  belongs to TI. The result of trajectory information is  $TC(TS)$  is signified as,

$$TC(TS) = \{in_1(TS), in_2(TS), \dots, in_{TI}(TS)\}$$

The observation scale is signified by  $\sigma$  and the time series of  $t_i^{\text{th}}$  investigation is derived using the Lindeberg's kernel at a scale of  $\sigma$  and the  $IN_i(TS, \sigma)$  is equated as,

$$IN_{t_i}(TS, \sigma) = \sum_{N=-\infty}^{\infty} E^{-\sigma T I_N(\sigma)} IN_{t_i}(TS - N)$$

The modified Bessels function is signified as  $J_0(x)$  and in the order of  $N$ . This convolution is incorporated in all investigation and in the scale of  $\sigma$ , the trajectory information  $CN(TS, \sigma)$  is retrieved as,

$$CN(TS, \sigma) = \{IN_1(TS, \sigma), IN_2(TS, \sigma), \dots, IN_{TI}(TS, \sigma)\}$$

The alteration in the value of, the trajectory at diverse observation scale is retrieved and the trajectory curvature is signified as  $TI=2$  and the cluster is signified as  $K$ ,

$$K(TS, \sigma) = \frac{IN_1' IN_2'' + IN_1'' IN_2'}{(IN_1'^2 + IN_2''^2)^{3/2}}$$

The  $IN_{t_i}''$  and  $IN_{t_i}'$  signify the derivatives of first and second order  $IN_{t_i}(TS, \sigma)$  is equated as follows,

$$IN_1^i(TS, \sigma) = \sum_{N=-\infty}^{\infty} -\frac{N}{\sigma} E^{-\sigma T I_N(\sigma)} IN_{ti}(TS - N)$$

$$IN_1^r(TS, \sigma) = \sum_{N=-\infty}^{\infty} -\frac{1}{\sigma} \left( \frac{N^2}{\sigma} - 1 \right) E^{-\sigma T I_N(\sigma)} IN_{ti}(TS - N)$$

#### 4.2 Distance Metrics used

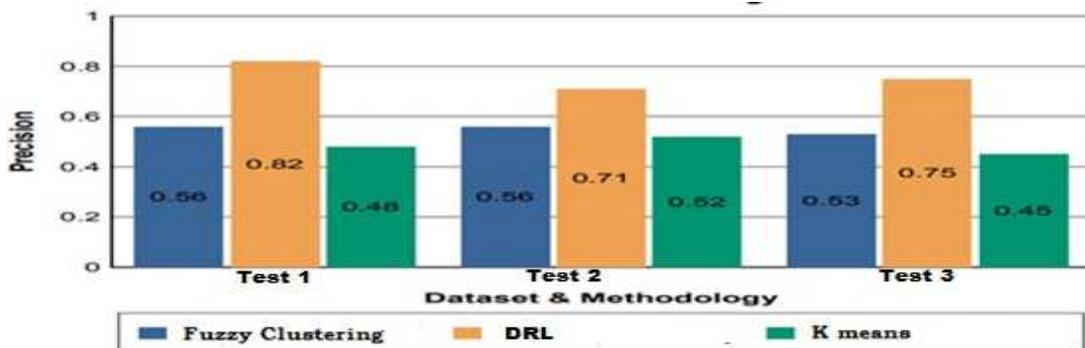
- **Euclidean distance:** The traditional Euclidean distance is used to cluster the trajectories. To find the similarity between two trajectories, the complete Euclidean distance between the points of the two trajectories is calculated.
- **Positional distance:** The trajectory points are labeled through data clustering of all the points representing all the trajectories. Now the point trajectories became labeled sequences. Longest common Sub Sequence method is used to find the distance between two trajectories.
- **Semantic Distance:** The attribute values are categorized based on the ranges of values. These semantic categories are compared between two trajectories to find the distance.

#### 4.3 Evaluation

The proposed architecture for patient treatment trajectory mining has been evaluated against the following performance measures against traditional machine learning. In this work, proposed model is evaluated using 10 fold validation with types as test data 1, test data 2, test data 3 to compute the performance of prediction results generated on dataset mentioned. The performance evaluation of the proposed Deep Reinforcement Learning model with PCA and NLDA depends on the process of base layer and optimization layer.

##### • Precision

Precision is a referred as measure of Positive predictive value in the clustre. It is further represented as the fraction of similar instances among the each cluster groups generated using the model. Figure 2 represents the performance evaluation of the proposed architecture in terms of precision on dataset.



**Figure 2: Performance analysis of the methodology on aspect of Precision**

Performance measures are suitable for determining the feasibility of proposed architecture on treatment trajectory generation. Effectiveness is achieved due to optimization of the reward function with strategies

$$\text{Precision} = \frac{\text{True positive}}{\text{True positive} + \text{False Positive}} \quad \frac{\text{True positive}}{\text{True positive} + \text{False Positive}}$$

Precision is computed using True positive and true negative. True positive is a number of similar points in the pattern generated and false negative is number of real dissimilar points in the pattern[15]. Mostly a good performance is also characterized by Euclidean distance[16] for the data points. It can be calculated using recall measure.

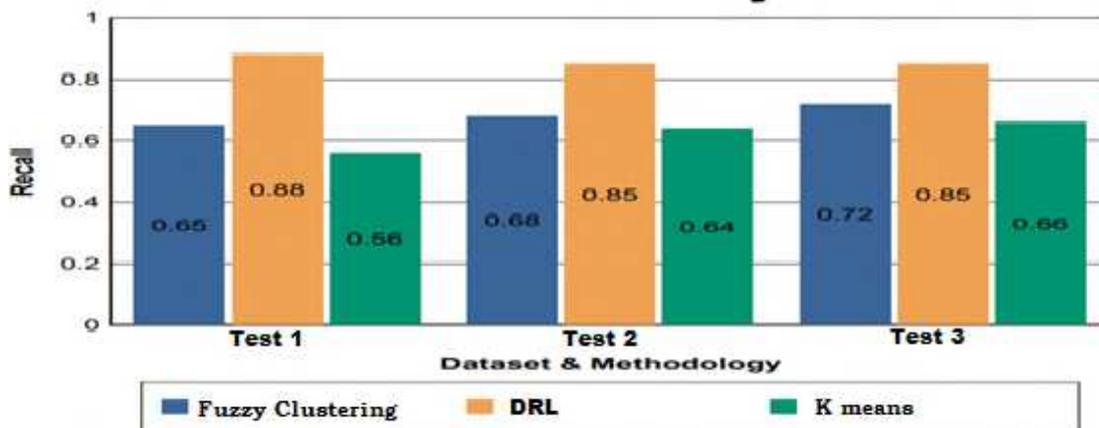
- Recall**

$$\frac{\text{True positive}}{\text{True positive} + \text{False negative}} \quad \frac{\text{True positive}}{\text{True positive} + \text{False negative}}$$

Recall is the part of similar instance of the dataset that have been extracted over the total amount of relevant instance of the dataset[17]. The recall is the part of the similar documents that are successfully classified into the exact classes.

Recall =

True positive is a number of similar data points in the pattern and false negative is number of similar data points in the pattern extracted. Figure 3 represents the performance evaluation of the proposed architecture on recall measure along state of art approaches.



**Figure 2: Performance analysis of the methodology on aspect of Recall**

- F measure**

It is the number of correct class predictions to the incoming data among total number

of predictions to whole category of data[18].

Accuracy is given by

$$\frac{\text{True positive} + \text{True Negative}}{\text{True positive} + \text{True Negative} + \text{false positive} + \text{False negative}}$$

The different instance of the patient treatment pattern generated may have different treatment trajectory pattern. The performance of the proposed model in terms of f measure against state of art approaches for treatment trajectory formation[19]. However, after a certain point, this data vector is diminished because of curse of dimensionality. The available trajectory for patient treatment on basis fitness computation normoglycemia attenuates the negative effect of prior exposure. Table 1 presents the performance value of the technique for cluster analysis on test category 1.

**Table 1: Performance Evaluation of the Technique to Test 1 category**

Technique	Precision	Recall	F measure
K means	0.45	0.66	0.54
Fuzzy Clustering	0.53	0.72	0.61
Deep reinforcement learning based trajectory Clustering - Proposed	0.75	0.85	0.80

Table 1 presents the performance value of the technique for cluster analysis on test category 2 of the data set used to trajectory clustering on future treatment mining.

**Table 2: Performance Evaluation of the Technique to Test 2 data category**

Technique	Precision	Recall	F measure
K Means	0.46	0.72	0.55
Fuzzy Clustering	0.52	0.79	0.61
Deep Reinforcement learning based trajectory Clustering - Proposed	0.79	0.86	0.80

Table 2 presents the performance value of the technique for cluster analysis on test

category 2 of the dataset for trajectory clustering. . It is observed that the proposed method is always better when compared to clustering results of state of art approaches

**Table 3: Performance Evaluation of the Technique to Test 3 data category**

Technique	Precision	Recall	F measure	Cumulative Gain	Mean Average Precision
K Means	0.46	0.66	0.64	135	0.64
Fuzzy Clustering	0.54	0.72	0.63	145	0.73
Deep Reinforcement Learning based trajectory clustering – Proposed	0.77	0.85	0.84	185	0.92

#### 4.5. Clustering results

In the context of medical data, similarity measures are compared using the Euclidean, positional, and semantic distances with its variants. The execution time and the purity of the cluster is estimated in various contexts of data and methods. The proposed methods showed better clustering than the existing methods. From the above tables it can be observed that the proposed methods are providing clustering with better accuracy and similarity. The execution time is also better. One more observation is that trajectory clustering with traditional similarity measures are differentiating trajectories in large scale and as a result low grouping is observed wherever the proposed similarity measures can provide more meaningful clustering.

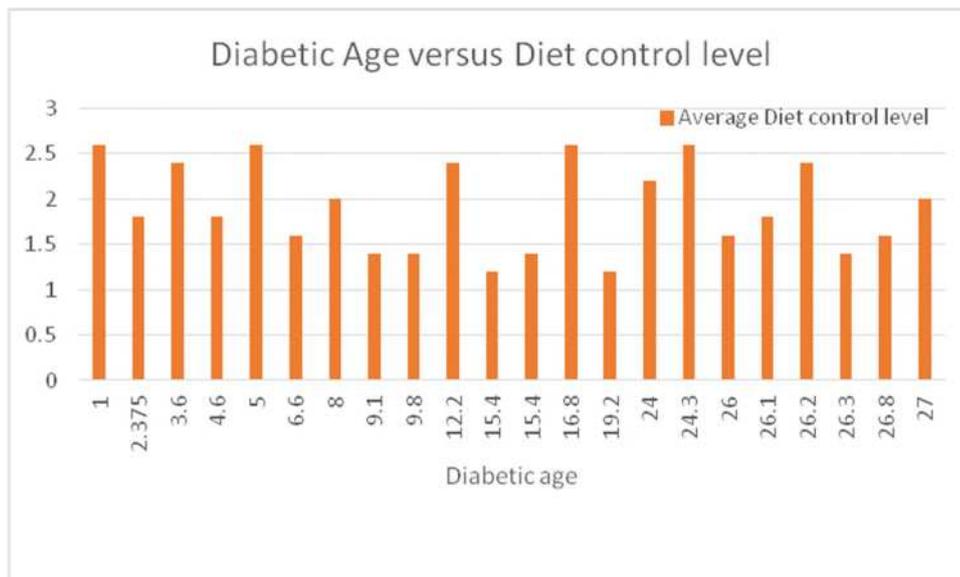
Diabetic patient dataset after analysis through proposed methods produced interesting results which are presented in Table 5. Clusters are representing groups of patients as per their diabetic age and level of diet control. This segregation helps the patients and doctors to take corrective measures to fight with diabetes.

Table 4. Significant clusters formed after the data of 1000 trajectories

Cluster Count	Average diabetic age	Average Diet control level
1	2.375	1.8
2	5	2.6
3	27	2
4	26	1.6
5	26.3	1.4

6	9.8	1.4
7	26.8	1.6
8	4.6	1.8
9	24.3	2.6
10	12.2	2.4
11	8	2
12	26.2	2.4
13	9.1	1.4
14	26.1	1.8
15	19.2	1.2
16	15.4	1.2
17	16.8	2.6
18	3.6	2.4
19	15.4	1.4
20	24	2.2
21	1	2.6
22	6.6	1.6

An analysis of clustering summary is presented through a graph in Figure.5. Here we can observe that clusters representing different diabetic age groups are not strictly associated with diet control levels. Therefore diabetic age is a key characteristic to assess a patient. This finding helps doctors to take better decisions in treatment and to give suggestions to patients.



## **Figure 5: Disease Characterization**

### **Conclusion**

A deep reinforcement learning trajectory data clustering model is developed with suitable and multi-dimensional patient treatment medical data. The complexity of higher dimensionality of trajectory data is handled by a two-phase clustering strategy. Initially, multi-dimensional points belong to all trajectories undergo a primary clustering process that assigns a cluster label for every trajectory point. The labeling of trajectories is based on the clustering done by using feature extraction and Non linear learning strategies. Labeled trajectories are finally clustered again to get informative groups for decision making. Proposed methodologies are tested on a diabetes dataset where every trajectory signifies a patient diagnostic history that is acquired from the nearby laboratories and hospitals. The investigation indicates that the proposed technique is faster than the existing trajectory grouping approaches and offers a new direction for liberal trajectory analysis. The proposed model is deep clustering process that tends to a better performance with lesser computational time and gives better treatment direction in case of medical data context under consideration.

### **Declaration:**

Ethics Approval and Consent to Participate:

No participation of humans takes place in this implementation process

Human and Animal Rights:

No violation of Human and Animal Rights is involved.

Funding: No funding is involved in this work.

Conflict of Interest: Conflict of Interest is not applicable in this work.

Authorship contributions:

There is no authorship contribution

### **Acknowledgment:**

No Acknowledgments

## Reference

- [1] T. Zhang, H. Lu, S. Z. Li, Learning semantic scene models by object classification and trajectory clustering, in Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, IEEE, 2009, pp. 1940-1947.
- [2] M. Nanni, D. Pedreschi, Time-focused clustering of trajectories of moving objects, *Journal of Intelligent Information Systems* 27 (3) (2006) 267-289.
- [3] X. Li, W. Hu, W. Hu, A coarse-to-fine strategy for vehicle motion trajectory clustering, in 18th International Conference on Pattern Recognition (ICPR'06), Vol. 1, IEEE, 2006, pp. 591-594.
- [4] M.-Y. Liu, O. Tuzel, S. Ramalingam, R. Chellappa, Entropy-rate clustering: Cluster analysis via maximizing a submodular function subject to a matroid constraint, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36 (1) (2014) 99-112.
- [5] B. Morris, M. Trivedi, Learning trajectory patterns by clustering: Experimental studies and comparative evaluation, in Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, IEEE, 2009, pp. 312-319.
- [6] Nguyen Le Duy, Vo Sakuma, Takuto et.al," Statistically Discriminative Sub-trajectory Mining",2019.
- [7] PAN XIAO, MA ANG, ZHANG JIAWEI, AND WU LEI, "Approximate Similarity Measurements on Multi-Attributes Trajectories Data", *IEEE Access*, a special section on big data learning and discovery, date of publication December 24, 2018.
- [8] C. Zhang, J. Han, L. Shou, J. Lu, and T. F. La Porta. Splitter: Mining Fine-Grained Sequential Patterns in Semantic Trajectories. *PVLDB*, 7(9):769–780, 2014.
- [9] T. Kahveci , A. Singh, and A. Gurel, "Similarity searching for multi-attribute sequences", *SSDBM '02 Proceedings of the 14th International Conference on Scientific and Statistical Database Management*, page 175, IEEE computer society, 2002.
- [10] L. Chen, G. Cong, C. S. Jensen, and D. Wu. Spatial keyword query processing: An experimental evaluation. *PVLDB*, 6(3):217–228, 2013
- [11] JianqiuXu and Ralf HartmutGüting, "Query and Animate Multi-attribute Trajectory Data", *CIKM'17*, Nov 2017, Singapore © 2017 Association for Computing Machinery.

- [12] Yingchi Mao, ZhongHaishi, Qi Hai, and Ping Ping, "An Adaptive Trajectory Clustering Method Based on Grid and Density in Mobile Pattern Analysis" *Sensors (Basel)* 2017 Sep; 17(9): 2013. Published online 2017
- [13] S Ranu, P Deepak, AD Telang, P Deshpande, S Raghavan," Indexing and matching trajectories under inconsistent sampling rates".2015 IEEE 31st International Conference on Data Engineering, 999-1010

- [14] Yong Geat.al, "Top-Eye: top-k evolving trajectory outlier detection".CIKM '10 Proceedings of the 19th ACM international conference on Information and knowledge management, Pages 1733-1736.
- [15] Zheng, Y., Zhang, L., Ma, Z., Xie, X., and Ma, W.-Y. 2011. Recommending friends and locations based on individual location history. ACM Trans. Web 5, 1, Article 5 (February 2011)
- [16] Lu CT., Lei PR., PengWC., Su IJ. (2011) A Framework of Mining Semantic Regions from Trajectories. In: Yu J.X., Kim M.H., Unland R. (eds) Database Systems for Advanced Applications. DASFAA 2011. Lecture Notes in Computer Science, vol 6587. Springer, Berlin, Heidelberg.
- [17] Hassan MM, Curley SA, Li D, et al. Association of diabetes duration and diabetes treatment with the risk of hepatocellular carcinoma. Cancer. 2010;116(8):1938-1946. doi:10.1002/cncr.24982.
- [18] Yingchi Mao, Zhong Haishi, Qi Hai, and Ping Ping, "An Adaptive Trajectory Clustering Method Based on Grid and Density in Mobile Pattern Analysis" Sensors (Basel) 2017 Sep; 17(9): 2013. Published online 2017
- [19] S Ranu, P Deepak, AD Telang, P Deshpande, S Raghavan," Indexing and matching trajectories under inconsistent sampling rates".2015 IEEE 31st International Conference on Data Engineering, 999-1010
- [20] Yong Geat.al, "Top-Eye: top-k evolving trajectory outlier detection".CIKM '10 Proceedings of the 19th ACM international conference on Information and knowledge management, Pages 1733-1736.
- [21] Lu CT., Lei PR., Peng WC., Su IJ. (2011) A Framework of Mining Semantic Regions from Trajectories. In: Yu J.X., Kim M.H., Unland R. (eds) Database Systems for Advanced Applications. DASFAA 2011. Lecture Notes in Computer Science, vol 6587. Springer, Berlin, Heidelberg.
- [22] Ramakrishna, S. S., & Anuradha, T. (2018). AN EFFECTIVE FRAMEWORK FOR DATA CLUSTERING USING IMPROVED K-MEANS APPROACH. International Journal of Advanced Research in Computer Science, 9(2).



*K.Arun kumar research scholar in Department of Computer Science and Engineering from JNTUA, Ananthapuramu, Andhra Pradesh. he received M.Tech Department of Computer Science and Engineering, JNTUACEA Research interests are in the areas of Wireless Networks, Data Mining,*



*Dr.S Vasundra, received her Ph.D degree in Computer Science and Engineering from JNTUA, Ananthapuramu, Andhra Pradesh. She is currently extending her services as Professor in the Department of Computer Science and Engineering, JNTUA. She has good contribution in many reputed journals. Her research interests are in the areas of Wireless Networks, Network Security, Data Mining, Internet of Things and Cloud Computing.*