GLASS DETECTION FOR DIGITAL FORENSICS USING UNSUPERVISED CLUSTERING ALGORITHM

Glass Detection for Digital Forensics Using Unsupervised Clustering Algorithm

Arman Rasool Faridi*

Department of Computer Science, Aligarh Muslim University, Aligarh, India

***Abstract** – In many studies attempting to segment various types of populations, the Fuzzy C-means (FCM) algorithm has been found to be one of the most effective algorithms, and it can thus be used to assist forensic experts in identifying different types of glasses discovered at a crime scene. This paper suggests a method for creating clusters of different types of glasses from their properties that is based on the FCM algorithm. The proposed method aims to create clusters based on various factors such as refractive index and composition. New values can be plotted on the graph after clusters have formed, and the type of glass located at the crime scene can be calculated depending on the orientation of the projection.*

Keywords — Digital Forensics, Glass Detection, Fuzzy C-Means, Clustering, Segmentation, Matlab

- x -

I.   INTRODUCTION

The glass industry includes containers (bottles and jars), flat glass (for architecture and transportation glazing), glass fiber (for reinforcement and insulation), domestic glass (kitchen and tableware), and specialized glasses (for a host of scientific and industrial uses). Based on the market, different manufacturing processes are used [1], [2]. Technological and commercial advances in glass manufacture have resulted in modern products, and manufacturing methods have greatly increased in terms of output volumes, glass quality, and the amount of peripheral procedures for handling glass over the past few years. While there are a variety of glass formulations, they all fall into one of a few buckets, making classification easier. [3].

A formulation is developed to satisfy the manufacturing process' specifications, as well as the properties required for the final product's use and production costs. Catalogues from manufacturers show a vast variety of products, each with its own structure. There are some minor variations in detail, the compositions in the high tonnage sectors (container, flat, and domestic) are mostly comparable. Glass is made in many technologically advanced nations, and there is much international exchange. Glass will be used for a long time because it is such a solid material, as demonstrated by church windows [4].

Glass samples from a specific location or occurrence can thus have a readily determined structure, whether ancient or contemporary, domestic or international. A fusion material that has not crystallised and has cooled to a rigid state is known as glass. As a consequence,

the core of glass is amorphous or non-crystalline [5]. Glasses are supercooled liquids with a unique set of properties, including clarity with or without colour, longevity, electrical and thermal resistance, a wide variety of thermal expansions, as well as hardness, rigidity, and resilience.

Inorganic products, many of which are naturally occurring oxide minerals, such as silica powder, are melted together, and chemical reactions occur to create glasses. As a result, commercial glasses are silicates. Other additives are used to maximise manufacture and output based on the combination of properties available in the final product [6]. In this paper, a discussion on the Identification of different glasses is done that can be used for forensic examination.

The most definitive way of proving the existence of a similar origin between glass samples will continue to be physical matching of moderate-sized bits, where possible [12]. Individualization attempts based on physical property measurements and refined chemical analysis methods will never achieve the degree of trust provided by physical matches. There's also no need to rely on libraries to make those connections. Physical matching of small particles can be accomplished using computerized methods capable of recognizing and comparing geometric shapes.

In glass examinations, we may see a continued shift away from advanced methods of physical property measurement and toward a greater reliance on trace elemental analysis methods. Physical property calculations used to provide excellent discrimination

between glass samples from different sources, but this has changed in two respects. Although glass samples vary less than in the past, methods for calculating small differences in trace elemental composition have advanced by orders of magnitude. Two significant factors are responsible for the decrease in variance among glass samples [7][8]. The majority of glass manufacturing has been concentrated in the hands of a small group of large-volume manufacturers. This has been supplemented by increased control of output tolerances, resulting in a more standardised product.

Soft clustering (also known as fuzzy clustering or soft k-means) is a clustering method in which each data point is assigned to several clusters. Clustering, also known as cluster analysis, is a method of grouping data points into clusters with the goal of making occurrences in the same cluster as identical as possible when individuals in different clusters are as dissimilar as possible. Similarity tests are used to classify clusters. Width, connectivity, and strength are all considerations in similarity tests. Based on the data or the software, different similarity measurements may be used [9].

The fuzzy c-means technique is used to group various types of glasses based on their refractive index and composition in this article. At first history of fuzzy c-means is given in this article, as well as an explanation of the dataset. Following that, we went over how to do the experiment in MATLAB, followed by a discussion of the findings and a conclusion.

II. BACKGROUND

A. Clustering algorithms

Clustering is a mechanism that divides a given data set into homogeneous classes based on given characteristics so that identical objects are held together while dissimilar objects are separated. It is the most significant unsupervised learning problem. It is concerned with the discovery of structure in a collection of unlabeled results.

Algorithms for clustering can be divided into two categories. These are Unsupervised linear clustering and another one is Unsupervised non-linear clustering. The most famous algorithms for Unsupervised linear clustering includes k-means clustering, Hierarchical clustering, Gaussian(EM) clustering and Fuzzy c-means algorithm.

One of the most popular unsupervised learning algorithms for solving the well-known clustering problem is k-means. The algorithm follows a basic and straightforward method for classifying a given data set using a fixed number of clusters (assume k clusters). The key concept is to create k cores, one for each cluster. Since different locations provide different results, these centres should be strategically located. As a result, it is preferable to position them as far away from each other as possible. The next move is to

connect each point in a data set with the centre that is closest to it. The first stage is completed, and an early group age is completed where no points are pending. Next, recalculate k new centroids as the barycenters of the clusters generated in the previous stage. Once k new centroids are created, then the same data set points need to be rebound to the nearest new core [10].

There are two kinds of hierarchical clustering algorithms, i.e. Agglomerative Hierarchical Clustering Algorithm (AGNES agglomerative nesting), and another one is Divisive Hierarchical Clustering Algorithm, (DIANA divisive hierarchical clustering) [11]. These algorithms are precisely the opposite of each other. Agglomerative Hierarchical Clustering algorithm groups data points one by one using the nearest distance estimate of all pairwise distances between them. The distance between the data points is recalculated once more, but which distance should be taken into account once the groups have been formed? There are several options for this. These include Single linkage or single-nearest-distance, the distance between the centroid and the origin etc. We continue to group the data in this manner before one cluster emerges. We can now quantify the number of clusters that should be present based on the dendrogram graph.

The Gaussian(EM) clustering algorithm assumes that there are 'n' Gaussian centres and then tries to fit the data into those centres by anticipating the groups of all data points and then optimising the maximum probability of Gaussian centres [12]. On the basis of the distance between the cluster centre and the data point, the fuzzy c-means clustering algorithm assigns membership to each data point belonging to each cluster centre. The closer the data is to the cluster center, the more likely it is to belong to the cluster center. Obviously, the number of each data point's membership could equal one. Membership and cluster centers are modified for each iteration. Fuzzy-c is a clustering algorithm that is used in this article [9].

B. Description of the dataset

The dataset that we used consists of six different glasses with nine features. Different types of glasses include Float Window, Non-float Window, Float Vehicle window, Non-Float Vehicle window, Containers, Table wares and Headlamps.

Float glass is a layer of glass formed by floating molten glass on a bed of molten metal, usually tin, but other low-melting-point alloys have been used in the past. This process creates a layer with a constant thickness and incredibly smooth surfaces [13].

Float glass is used in modern windows. Soda-lime glass makes up the majority of float glass, while speciality borosilicate and flat panel show glass are still made in small quantities [14]. The float glass process is also known as the Pilkington process, after

Arman Rasool Faridi*

the British glassmaker Pilkington, who invented it in the 1950s at their St Helens, Merseyside, plant [15].

Float glass can be used for vehicle windows as well. Numerous experimental and computational tests of the properties of float glass for windows have been performed. For the front windshield of a vehicle, laminated float glass was widely used. In a predictive strength model for monolithic and laminated glass, some researchers have discussed the issue of degree dependence in laminated float glass [16].

According to a report using static modelling, float glass increased variability between 1.29-1.58, while hard floating glass reduced variability between 0.63-0.76. Delamination was improved by increasing the thickness between layers from 0.38 to 0.76 mm, but it was reduced by increasing the thickness between layers to higher values [17].

Container glass is a form of glass that is used to make glass bottles, pots, drink ware, and bowls. Flat glass (used for mirrors, glass doors, translucent walls, and windshields) and glass fiber are two types of container glass (used for thermal insulation, fiberglass composites, and optical communication). Flat glass contains less magnesium oxide and sodium oxide, while container glass contains more silica, calcium oxide, and aluminum oxide. Most of container glass is blown and pressed soda-lime glass, with some laboratory glassware consisting of borosilicate glass [18].

Glasses used for table setting, preparing meals, eating, and decoration. It may be soda-lime – silica, lead or barium crystal, or Borosilicate glass, depending on the application and consistency specifications. Glasses made of soda, lime, and silica are the most popular kind for regular glass wares that do not need high-temperature tolerance or a gleaming appearance. When glassware is in close contact with fire or can be used within an oven, borosilicate glassware is typically used for high-temperature resistance.

A continuous flow process in a very long tank furnace running at about 1500°C produces headlight glass. Raw materials are loaded into the melting chamber in batches. The viscous mass flows under a fireclay float after melting and is continuously drawn off from the working end to be fed to shaping machines, which press it into the lens or reflector configurations required. The method is vulnerable to leaving striations, bubbles, or pieces of foreign material in the glass, resulting in a non-uniform physical and optical property in the moulded object. The refractive index varies significantly from one location to the next [19].

The refractive index is one of the characteristics. In optics, a substance's refractive index (also called refraction index or index of refraction) is a dimensionless number that describes how easily light passes through it. The speed of light in a substance decreases as the refractive index of the material rises.

The refractive index of a substance determines how much light is bent or refracted as it passes through it. Glasses are made up of a variety of materials. We calculate the weight of these elements based on their oxides. Magnesium, Sodium, Aluminum, Potassium, Calcium, Silicon, Barium, and Iron are among them.

III. EXPERIMENTAL SETUP

Data on glasses were collected from a well-known repository that has already been verified. It is made up of various data files. One file contains a summary of the data, while the other contains the whole data collection. Nine records for each kind of glass were taken, each containing all of the characteristics. The initial data was in the format seen in table 1 and had more than nine entries.

From the original data, the Id number was removed. The type of glass field is used in such a way that we can categories data. Sample data is shown in figure 1.

Table 1: The Original Data Fields

S. No.	Fields
1	Id number (Serial Number)
2	RI (Refractive Index)
3	Na (Sodium)
4	Mg (Magnesium)
5	Al (Aluminum)
6	Si (Silicon)
7	K (Potassium)
8	Ca (Calcium)
9	Ba (Barium)
10	Fe (Iron)
11	Type of glass

Fig. 1 Snapshot of the dataset

This data was transformed into a JSON file, and then in MATLAB, the file name was set, and the JSON file was read. MATLAB was used to write the code.

filename='glassDataset.json';

Following that, the index will be derived for various glasses based on the last field from the data.

buildingWindowsFloat **=** glassesDataset**(:,**10**)==**1**;**

buildingWindowsNFloat **=** glassesDataset**(:,**10**)==**2**;**

vehicleWindowsFloat **=** glassesDataset**(:,**10**)==**3**;**

containersIndex **=** glassesDataset**(:,**10**)==**5**;**

tablewareIndex **=** glassesDataset**(:,**10**)==**6**;**

headlampIndex **=** glassesDataset**(:,**10**)==**7**;**

The last field will be deleted and normalize the data until we have indexes.

glassesDataset=glassesDataset**(:,**1:9**);**

[data_num,~]=size(glassesDataset);

normalizedData=(glassesDataset - ones(data_num,1)*min(glassesDataset))./(ones(data_num,1)*(max(glassesDataset)-min(glassesDataset)));

Following that, the data fields are segregated using the indexes discovered earlier :

buildingWindowsFloat **=** glassesDataset**(**buildingWindowsFloat,**:);**

buildingWindowsNFloat **=** glassesDataset**(**buildingWindowsNFloat,**:);**

vehicleWindowsFloat **=** glassesDataset**(**vehicleWindowsFloat,**:);**

containersIndex **=** glassesDataset**(**containersIndex,**:);**

tablewareIndex **=** glassesDataset**(**tablewareIndex,**:);**

headlampIndex **=** glassesDataset**(**headlampIndex,**:);**

Following that, extra rows are removed, and only nine rows are chosen for all diseases :

buildingWindowsFloat **=** buildingWindowsFloat**(**1:9,**:);**

buildingWindowsNFloat **=** buildingWindowsNFloat**(**1:9,**:);**

vehicleWindowsFloat **=** vehicleWindowsFloat**(**1:9,**:);**

containersIndex **=** containersIndex**(**1:9,**:);**

tablewareIndex **=** tablewareIndex**(**1:9,**:);**

headlampIndex **=** headlampIndex**(**1:9,**:);**

Following separation, now graphs of different functions can be ploted, but first we must establish characteristics and combinations of various characteristics:

Characteristics = {'RI', 'Mg', 'Na', 'Si','AI','K','Ca', 'Fe', 'Ba'};

[~,s]=size(Characteristics**);**

totalPairs**=**s***(**s-1**)/**2**;**

current**=**1**;**

xyCoordinates=zeros**(**totalPairs,2**);**

for col **=** 1:**(**s-1**)**

for row **=** col**+**1:s

xyCoordinates**(**current,1**)=**col**;**

xyCoordinates**(**current,2**)=**row**;**

current**=**current**+**1**;**

end

end

As a consequence, the total number of pairs between different features is equal to

$$P = F * \frac{(F - 1)}{2}$$

Where P = Number of pairs and F= Number of features

After the pairs have been made, subplots for each pair will be made:

for i **=** 1 **:** totalPairs

x **=** xyCoordinates**(**i, 1**);**

y **=** xyCoordinates**(**i, 2**);**

subplot**(**6,6,i**)** ;

plot **([**buildingWindowsFloat**(:,**x**)** buildingWindowsNFloat**(:,** x**)** vehicleWindowsFloat**(:,** x**)** containersIndex**(:,**x**)** tablewareIndex**(:,**x**)** headlampIndex**(:,**x**)],**...

[buildingWindowsFloat**(:,**y**)** buildingWindowsNFloat**(:,**y**)** vehicleWindowsFloat**(:,**y**)** containersIndex**(:,**y**)** tablewareIndex**(:,**y**)** headlampIndex**(:,**y**)],** '.','MarkerSize',5**)**

xlabel**(**Characteristics**{**x**})**

ylabel**(**Characteristics**{**y**})**

end

Arman Rasool Faridi*

These graphs now include data points between various features, and now the fuzzy c means algorithm will be used to determine the centres for each cluster. We must first set the parameters of the fuzzy c means algorithm before calling it.

numberOfClusters **=** 6**;**

exponent **=** 2.0**;**

maxIter **=** 100**;**

minImprove **=** 1e-6**;**

clusteringOptions **= [**exponent maxIter minImprove true**];**

Where numberOfClusters denotes the number of clusters desired, exponent denotes the matrix's exponent, maxIter denotes the maximum number of iterations needed, and minImprove denotes the minimum amount of change required. Finally, we combine all of the clustering options into a single vector of variables.

Following that, the fuzzy c means algorithm is invoked, and the centers are collected in a variable:

[centers,U]   = fcm(glassesDataset,numberOfClusters,clusteringOptions);

Now, once we get the centers, we map them in the subplots from earlier so we can see whether the centers are suitable or not.

for i **=** 1**:**totalPairs

subplot**(**6**,**6**,**i**) ;**

for j **=** 1**:**numberOfClusters

x **=** xyCoordinates**(**i**,**1**);**

y **=** xyCoordinates**(**i**,**2**);**   text**(**centers**(**j**,** x**),**centers**(**j**,** y**),**int2str**(**j**));**

end

end

RESULTS & CONCLUSION

Sample results are shown in Appendix 1. These figures provide plots for all of the variations that can be created using test features or characteristics. Different colours reflect different types of glasses, and as we can see, the centres are generated using a fuzzy c means algorithm and are based on the various colours or diseases. Some outcomes are more closely related than others. As can be seen, all of the plots containing Barium or Iron do not yield consistent findings, making it impossible to identify clusters. However, on all remaining relations, it is simple to locate the cluster and determine where the new point would belong.

In the results provided, it can be seen that this can be a way to determine the type of glass without any further test. We give details about the elements, and then the new point is mapped, allowing us to determine what kind of glass it is. The point that is closest to the centre indicates the style of glass is more likely. In the future, we will be able to create an algorithm that would simply take the features and show the probability of the glass depending on them.

REFERENCES

[1]   P. S. Rogers (1968). "Inorganic glass-forming systems." JSTOR.

[2]   P. F. James (1975). "Liquid-phase separation in glass-forming systems," J. Mater. Sci., vol. 10, no. 10, pp. 1802–1825.

[3]   I. W. Donald (1993). "Preparation, properties and chemistry of glass-and glass-ceramic-to-metal seals and coatings," J. Mater. Sci., vol. 28, no. 11, pp. 2841–2886.

[4]   H. Rawson (1988). "Glass and its history of service," IEE Proc. A (Physical Sci. Meas. Instrumentation, Manag. Educ. Rev., vol. 135, no. 6, pp. 325–345.

[5]   J. Jackle (1986). "Models of the glass transition," Reports Prog. Phys., vol. 49, no. 2, p. 171, 1986.

[6]   V. S. Minaev, S. P. Timoshenkov, S. A. Oblozhko, and P. V Rodionov (2004). "Glass formation ability: Is the Rawson's" liquidus temperature effect" always effective?," J. Optoelectron. Adv. Mater., vol. 6, pp. 791–798.

[7]   R. Saferstein (2007). Criminalistics: An introduction to forensic science. Pearson Prentice Hall Upper Saddle River, NJ.

[8]   G. Zadora (2007). "Glass analysis for forensic purposes—a comparison of classification methods," J. Chemom. A J. Chemom. Soc., vol. 21, no. 5–6, pp. 174–186.

[9]   J. C. Bezdek, R. Ehrlich, and W. Full (1984). "FCM: The fuzzy c-means clustering algorithm," Comput. Geosci., vol. 10, no. 2–3, pp. 191–203.

Arman Rasool Faridi*

[10]   J. MacQueen (1967). "Some methods for classification and analysis of multivariate observations."

[11]   L. Rokach and O. Maimon (2005). "Clustering Methods," in Data Mining and Knowledge Discovery Handbook, O. Maimon and L. Rokach, Eds. Boston, MA: Springer US, pp. 321–352.

[12]   A. W. Moore (2001). "Clustering with gaussian mixtures," Sch. Comput. Sci. Carnegie Mellon Univ. www. cs. C. edu/~ awm awm@ cs. C. edu.

[13]   O. Uusitalo and T. Mikkola, "Revisiting the case of float glass," Eur. J. Innov. Manag., 2010.

[14]   F. Angeli et. al. (2012). "Effect of temperature and thermal history on borosilicate glass structure," Phys. Rev. B, vol. 85, no. 5, p. 54110.

[15]   J. B. L. L. B. DAVID (2007). "Elusive decisions: A case study of intuitive strategic decision Making in the exploitation of the Pilkington Float glass process, 1952-1987," Citeseer.

[16]   S. De Pauw (2010). "Experimental and numerical study of impact on window glass fitted with safety window film," Ghent University.

[17]   Q. Wang, Z. yi, J. Sun, J. Wen, and S. Dembele (2011). "Temperature and Thermal Stress Simulation of Window Glass Exposed to Fire," Procedia Eng., vol. 11, pp. 452–460.

[18]   T. P. Seward III and T. Vascott (2005). High temperature glass melt property database for process modeling. Wiley-American Ceramic Society.

[19]   R. S. Greene and D. Q. Burd (1949). "Headlight glass as evidence," J. Crim. L. Criminol., vol. 40, p. 85.

APPENDIX 1

— Cluster 1 — Cluster 2 — Cluster 3 — Cluster 4 — Cluster 5 — Cluster 6

Arman Rasool Faridi*
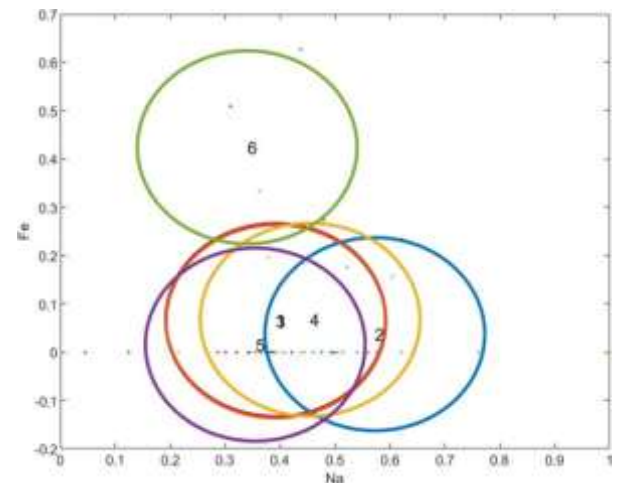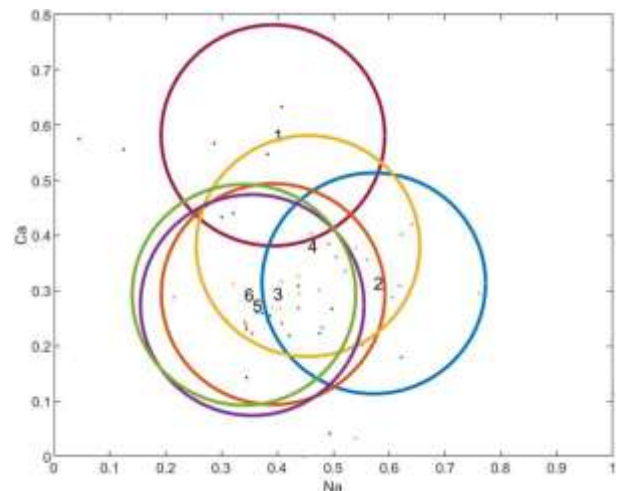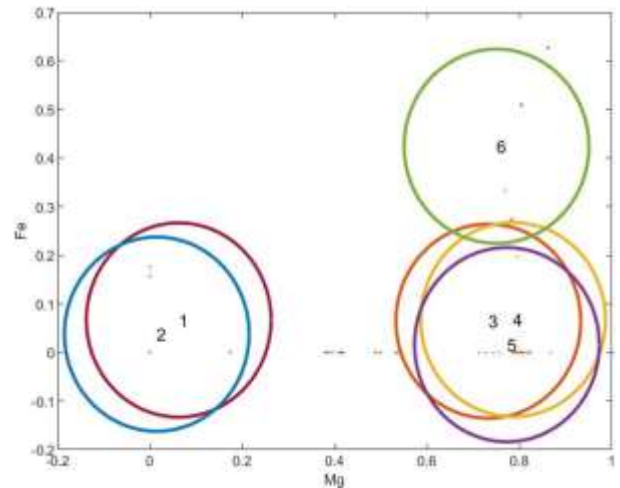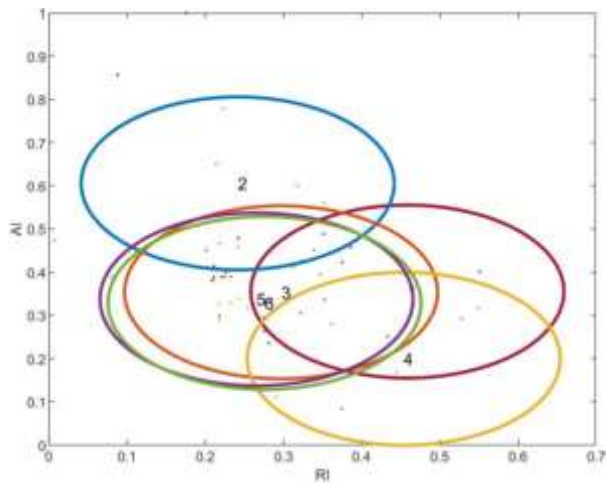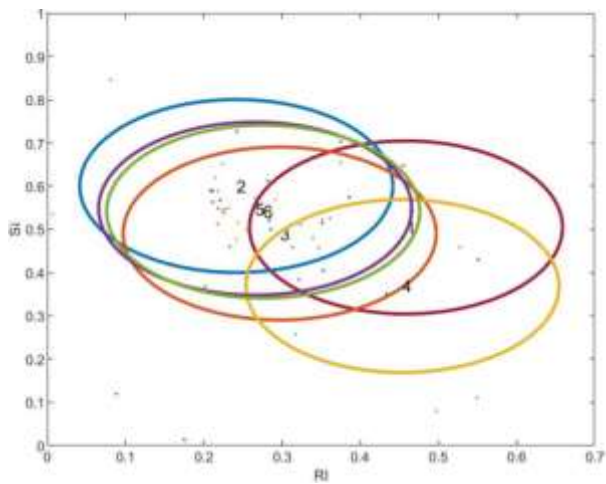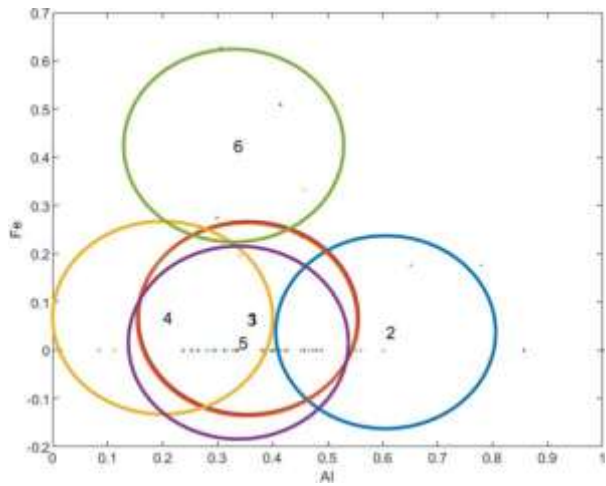
Corresponding Author

Arman Rasool Faridi*

Department of Computer Science, Aligarh Muslim University, Aligarh, India

arman.faridi@gmail.com