

# A Technique to Identify Key Players that Helps to Improve Businesses Using Multilayer Social Network Analysis

Aftab Farooq, Usman Akram, Gulraiz Javaid Joyia, and Chaudhry Naeem Akbar

**Abstract**—Social network analysis has gained importance in the current era of online networks. Companies nowadays are using the social network analysis tools and techniques to enhance their business positions. Analyzing the social networks provide the massive amount of useful information for the business. In this paper, we have analyzed the online social networks of leading mobile phone company Huawei. The importance of any individual (Node) can be determined by the participation and behavior of the node in the OSN. Centrality measure is the key factor to visualize any social network. Centrality in term includes Density, Closeness Centrality, Degree Centrality, Eigenvector Centrality and some other metrics such as Coefficient Clustering, Page Rank. These metrics can be used to determine the most important node (key player) in a network. Visualization of the social network is a complicated task as it is expanded every minute as thousands of users join the social network daily. We have put an effort to analyze the social network and determine the influential node with the calculation of above-mentioned metrics. The overall aim of our research is to improve the business by identifying the most influential node. The experimental results and a detailed quantitative analysis show that this is the more efficient and effective way to detect the influential nodes in an online social network.

**Index Terms**—Social network analysis (SNA), data mining, community detection, online social network (OSN), online business communities, centrality measures, performance, efficiency, business improvement, crawling.

## I. INTRODUCTION

The number of users and the content shared on the social network by these users shows the significance of the networks on the internet. About 2/3 of the internet users visits the social websites weekly this testifies the importance of the social media [1]. Online shopping websites have been evolved after the emergence of the web 2.0. These days customers can interact on the social media sites and share the thoughts by sharing pictures and videos, post comments, provide reviews, as well as provide or get advice. A social network consists of a group of users which are also called social actors. The visual representation of the social network provides vertices and edges. Users (nodes) are represented by the vertices and the relation between the nodes is represented by the edges between the vertices.

The term key player refers to the nodes which are most active in the social network, with respect to the business community the key player or key customer is the one which is the most prominent user in their business. In other words, the

key customer is one which is most beneficial for the company in the terms of their business. Various methods have been defined in the modern era to visualize and analyze the social networks. In a complex social network, the user is defined by hundreds of attributes. The complete set of user attributes is of no use, some of the attributes are incomplete or do not have relevant information with respect to our research are discarded and the only the metrics that are related to our analysis are kept under consideration. Fig. 1 shows the visual representation of the social network.

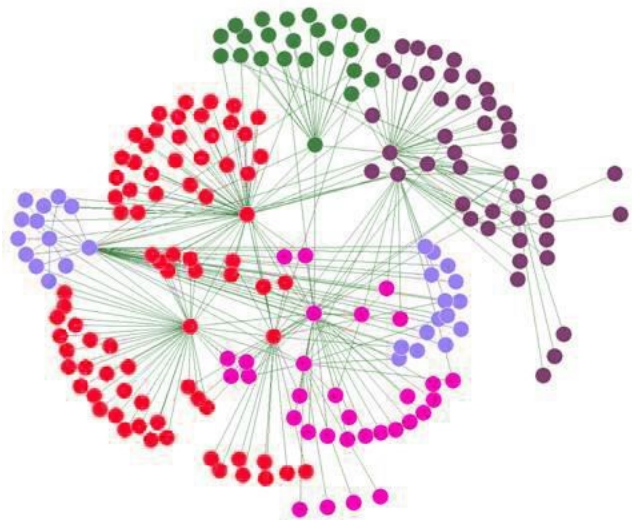


Fig. 1. Showing nodes and their relationships.

Detection of the key player in a social network is an old and well-known problem, it is no different than finding a needle in a haystack. To detect the key players in any social media centrality measure is important.

Social network analysis is being used in the modern era for various applications some of them include; fraud detection, terrorist activities in covert networks, military surveillance for enemy activities. The private companies are using the SNA methods to analyze the customer behaviors and relative users for marketing campaigns [2].

## II. LITERATURE REVIEW

The Researchers have used the social network analysis techniques and methods to analyze the different social networks for broad applications. Most of the research has been carried out on organized criminal networks to detect the most influential nodes (leader of the criminal network or Disseminator).

But in the recent years, the business communities have started using the SNA methodologies to enhance their business by detecting the key players, developing the

Manuscript received February 20, 2018; revised May 30, 2018.

The authors are with the National University of Sciences and Technology Islamabad, Pakistan (e-mail: aftabfarooq2012@gmail.com, usmanakram@gmail.com, ingrgulraiz@gmail.com, chmnaeemakbar@yahoo.com).

marketing campaigns, detecting behaviors of their customers, predicting the customer's churns, developing recommender systems etc.

Leidys del Carmen measured the clustering and association rules with a famous CRISP-DM method to analyze the behaviors of the customers of the fashion industry in the Instagram social network, which provided the industry with the handful of important information regarding their products and their trends and likeness among their customers [3]. The number of new users joins the social network daily, few of them are new while the some of them are the individuals which are already present in the social network. Pasquale De Moe *et al.* [4] defined a new approach to predict, whether the new user is already a node in the social network by comparing the social events and user interest.

Microblogging websites are gaining more and more attention of the people around the globe. Twitter is the most commonly used microblogging forum, in which users interact and show their expression (sentiments) towards any event or person by tweeting. Agarwal *et al.* proposed a framework to analyze the tweets and classify them into 3 main categories i.e. positive, negative or neutral on the basis of sentiment [5].

The diverse amount of research has been carried out in the area of social network analysis, a new approach to study the viral nature of the content on the social media is defined by Mills [6] which is executed on the basis of four elements: nexus, integration, spread ability, and propagativity, Abhishek *et al.* [7] presented the new system to calculate the virality of the information on the online social media.

To detect the key player various network metrics are kept under consideration, which includes Degree, Geodesic Distance (GD), Eigenvector Centrality (EC), Closeness Centrality (CC), Clustering Coefficient (CC), Degree Centrality (DC), Graph Density (GD), Betweenness Centrality (BC) and Page Rank. These metrics can be calculated with the wide range of the applications that are available i.e. Gephi, Cytoscape, KeyLines, Linkurious, Neoclipse, Polinode, Wandora, Socilyzer.

### III. PROPOSED SCHEME

#### A. Multilayer Dataset

The multi-layered system is defined by the many names e.g. Multi-relational, Multi-dimensional term, multi-slice, multilevel and multi-type networks. Although the Multi-layered network is most commonly used. In social networks, accounts of the individuals are present rather than themselves. The most significant part of this analysis is to map the nodes of two different layer in multilayer network. Multiple accounts of one node in SN are merged or one account is said to be official and the other ones are ignored.

Fig. 2 depicts the three layers of the social network. The three layers include one layer of Facebook, Instagram, and Twitter.

#### B. Dataset Description

The data is crawled with the help of APIs from the online social forums of Huawei Mobile PK i.e. Facebook, Instagram and Twitter. The data consist of 210 individuals interacting with each other in the comments section. There are three

layers of the social network i.e. Huawei Instagram Page, Huawei Facebook Page, Huawei Twitter Page. The dataset consists of directed relationships. Anyone who replies to someone's in the comments is considered to be in a relationship in with the one who commented. It is obvious that the person who commented is also in relation with the one who replied e.g. if A is in relation to B, then B is also in relation to A.

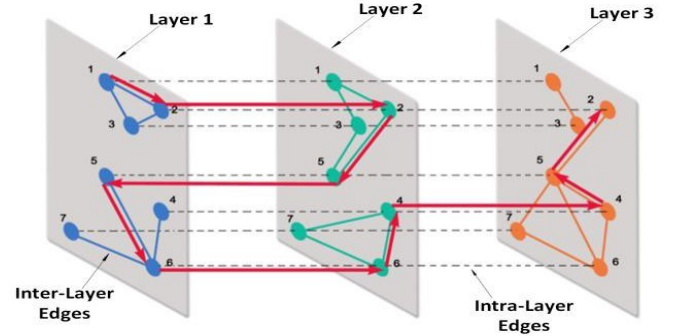


Fig. 2. Structure of the Multilayer social network.

The Gephi Graph Visualization tool was used to represent the dataset extracted from the Huawei online forums. The dataset consists of 425 nodes and 17842 edges showing the relationships between the nodes. Fig. 3 below shows the graphical representation of the dataset generated by the tool.



Fig. 3. Graphical Representation of multilayer social network.

#### C. Degree Centrality (DC)

Degree Centrality defines the direct contact of an individual. The higher direct contacts mean higher degree centrality and vice versa [8]. As mentioned earlier, DC is the measure of the direct nodes or links of the selected node. The Degree Centrality is directly proportional to the connectivity, higher the DC, well connected the node is in the network. The formula to calculate the degree centrality is given as:

$$Dv = aiv \frac{n}{i=1} \quad (1)$$

where  $\sigma x$  is DC (Degree Centrality) of individual  $x$  and

calculated with the help of adjacency matrix  $A$ .

### 1) Centrality

Determining the key player (influential individual) in the social media/network is the commonly used application of the graph theory. The extent to which a node is in the center of a network is defined by centrality.

Various techniques are used to define the centrality of a node. The most crucial measures to define the centrality are Betweenness, Degree and Closeness Centrality.

### 2) Betweenness Centrality (BC)

Two nodes in the network that are not directly connected don't mean they do not interfere. Entities which have not been in direct relation to each other may communicate through other nodes in the network, especially through the nodes that act as a bridge on the paths. The nodes Betweenness Centrality is said to be high if it is present in many other nodes. We can represent Betweenness Centrality as:

$$C_B v = \frac{\sigma_{st}(v)}{\sigma_{st}} \quad (2)$$

where  $\sigma_{st}$  represents the all the shortest distances from a node  $s$  to a node  $t$  and  $\sigma_{st}(v)$  is the count of the paths that pass through  $v$ .

### 3) Closeness Centrality (CC)

The only shortcoming of the degree centrality is it only considers the direct attached nodes and links to the account. While Closeness Centrality calculates the shortest path of the selected node and other nodes in the nodes in the network. We can also say Closeness Centrality is the aggregate distance between the individual and social media. The CC is termed as the inverse of two nodes distant [9], [10] that is:

$$C_x = \frac{1}{\sum_y d(y,x)} \quad (3)$$

where  $d(y,x)$  is the measured distance between the  $x$  and  $y$ . Closeness Centrality (CC) can also be defined as a time taken to disperse the information from one node to the whole network.

### 4) Eigenvector Centrality (EC)

Measuring importance of a node in a social media refers to an Eigenvector Centrality. It is proportional to the sum of all centralities of node's neighbors. This causes the node to gain high importance either by connecting to the nodes that are highly important nodes themselves or by connecting with many other nodes. Eigenvector Centrality (EC) prefers to pinpoint the most connected intermediary [11]. To attain the high Eigenvector Centrality the node has to connect itself to other nodes which also have higher eigenvector [12] Google's Page Rank and Katz centrality are the examples of the eigenvector centrality.

Vertex  $v$  whose Eigenvector Centrality is given as [13], [14]:

$$ECv = v_x = \frac{1}{\lambda_{\max}(A)} \sum_{j=1}^n a_{jx} v_j \quad (4)$$

whereas  $v = (v_1, 2, \dots, v_n)^T$  refers an eigenvector for the maximum eigenvalue  $\lambda_{\max}(A)$  of the adjacent matrix  $A$ .

### 5) Clustering Coefficient

Clustering Coefficient measures the likeliness of the nodes which are associates among each other. The Node who has high clustering coefficient indicates greater 'cliquishness'. In a graphical theory, a clustering coefficient measures the degree to which all participating nodes in a graph cluster together. We can calculate the average local clustering coefficients of all vertices  $n$  [15], [16].

$$C = \frac{1}{n} \sum_{i=1}^n C_i \quad (5)$$

## IV. RESULTS AND DISCUSSIONS

The detailed results and quantitative analysis of our proposed schema are discussed in this portion of the paper. In order to evaluate our proposed schema, we have measured the following metrics; In Degree, Out Degree, Degree, Weighted Degree, Weighted Out degree, Weighted In Degree, Eccentricity, Closeness Centrality, Harmonic Closeness Centrality, Betweenness Centrality, Page Rank, Clustering, Eigenvector Centrality to pinout the influential nodes from the Huawei online social network.

Below Fig. 4 depicts the graphical information of complete network metrics. To explain the network metrics this Fig. 4 is divided into six different sub figures, which are; Degree, Closeness Centrality (CC), Betweenness Centrality (BC), Eigenvector Centrality (EC), Clustering Coefficient (CC) and Page Rank. Each one is labeled and has different parameters on x and y-axis.

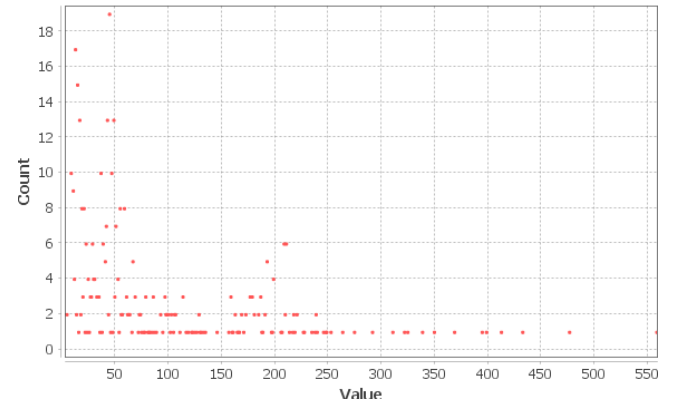


Fig. 4(a). Degree distribution.

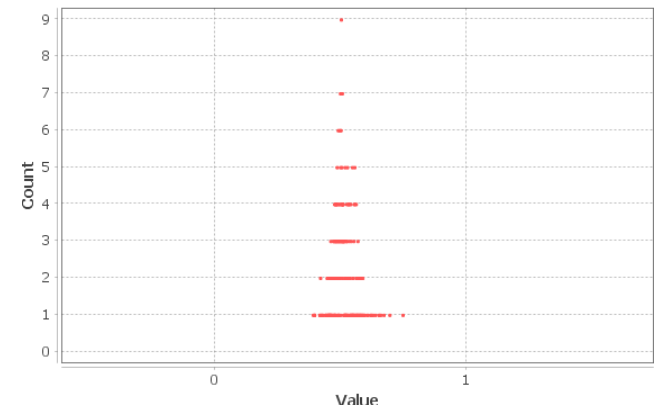


Fig. 4(b). Closeness centrality distribution.



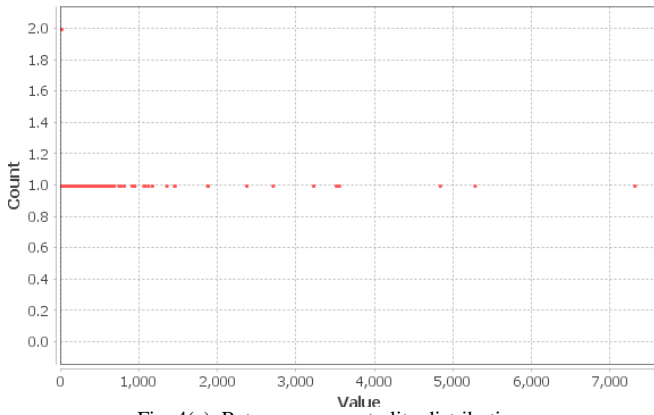


Fig. 4(c). Betweenness centrality distribution.

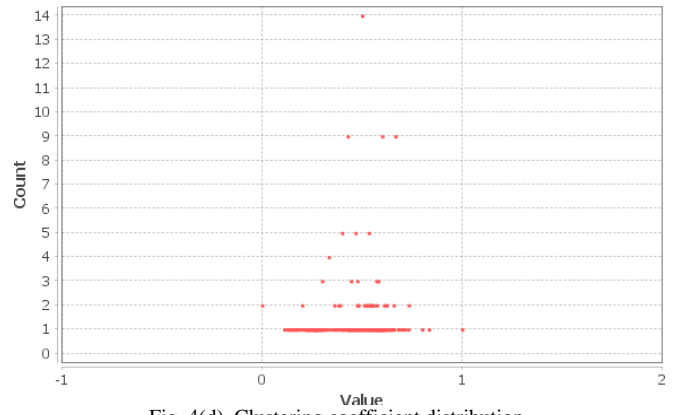


Fig. 4(d). Clustering coefficient distribution.

	Nodes	In Degree	Out Degree	Degree	Weighted Degree	Weighted Outdegree	Eccentricity	Closness Centrality	Betweenness Centrality	Page Rank	Clustering	Eigenvector Centrality
1	Huawei Mobile PK	280	278	558	372	370	2	0.746479	7606.121602	0.01428	0.132463	1
2	Akhtar Munir	238	238	476	302	302	3	0.693944	7302.736934	0.01356	0.110307	0.791561
3	Shoaib Khan	216	216	432	286	286	2	0.670886	4825.111827	0.011494	0.151852	0.81348
4	Malik Farooq	206	206	412	271	271	3	0.658385	5267.208474	0.011758	0.12754	0.729131
5	Fariddu Khan	199	199	398	257	257	3	0.652308	3213.798345	0.010136	0.151109	0.760531
6	Faizan Ahmed	197	197	394	255	255	2	0.651306	3539.023666	0.010393	0.170569	0.776908
7	Jinsar Ali Shar	184	184	368	260	260	3	0.637594	2697.679448	0.009397	0.198384	0.773886
8	Owais Arain	175	174	349	225	224	3	0.62908	3502.052713	0.009814	0.160394	0.672641
9	Abdul Rafah	169	169	338	216	215	3	0.623529	1440.091334	0.008246	0.171932	0.675855
10	Abdullah Butt	162	162	324	225	227	2	0.618978	1860.758942	0.008268	0.235598	0.737042
11	Muhammad Khan	161	160	321	235	233	2	0.617176	1446.838935	0.007706	0.209161	0.693661
12	Zeeshan Khan	155	155	310	193	193	3	0.606581	2361.361249	0.008309	0.176707	0.635817
13	Sameed Shahzad	147	144	291	207	201	3	0.603129	901.942705	0.00696	0.229056	0.666003
14	Syed Muhammad Shah	137	137	274	172	172	3	0.592179	1342.736337	0.006983	0.182589	0.562645
15	Muhammad Ali	131	132	263	141	142	2	0.593007	569.541635	0.006309	0.277522	0.636844
16	Muhammad Raza Ansari	126	126	252	160	160	2	0.587258	599.361453	0.006039	0.268444	0.610443
17	Uzair Abbasi	124	123	247	129	128	3	0.586445	558.476472	0.006029	0.240762	0.569877
18	Hassan Bhatti	123	122	245	146	144	3	0.584022	655.56675	0.005942	0.220955	0.545237
19	Muhammad Shahzain	122	126	248	163	170	3	0.586445	733.091133	0.006127	0.233846	0.54217
20	Farhan Rana	121	118	239	135	131	2	0.584828	598.587676	0.005943	0.288285	0.585166
21	Freaky Alee	119	119	238	121	121	3	0.577657	508.877852	0.005758	0.239567	0.521411
22	Mohammad Ata Farid	119	119	238	120	120	3	0.576087	518.225612	0.005685	0.242843	0.525626

Fig. 5. Metrics calculation of Huawei network.

All the graphs (Sub Figures) were produced using the Gephi graph visualization tool. The graphs show the Network Metrics of Huawei Dataset extracted from the online social forums of the Huawei.

Fig. 4(a) shows the degree distribution (edges attached) to any node. Nodes are represented on x-axis and number of direct contacts on the y-axis. The higher the number of nodes attached the higher is the degree. Fig 4(b) depicts the Closeness Centrality Distribution of the nodes present in the network. The x-axis portrays the nodes while y-axis reveals the closeness path to the whole network. Fig. 4(c) illustrates the Betweenness Centrality Distribution of the node. Nodes are represented on the x-axis and the number of betweenness nodes on the y-axis. Fig. 4(d) illustrates the Clustering Coefficient Distribution of the node.

Fig. 5 shows that Huawei Mobile PK (Row no 1) is the Key Player (Influential Node) in whole network and feature values of its metrics is high as compared to other nodes. The degree of this node is 280, Betweenness Centrality is 7606.121, Closeness Centrality is 0.7464, Eigenvector Centrality is 0.7769, Page Rank is 0.0142 and Clustering Coefficient is 0.132. Network The Above figure shows the top 3 influential nodes in the whole social network of Huawei. The most influential node is Huawei Mobile PK followed by Akhtar Munir and Shoaib Khan respectively.

## V. CONCLUSION

Every day number of people are joining the social networks so detection of influential nodes in such a network

is not easy. Thus our paper proposed a schema to identify the most influential nodes that are based on network measures that include Degree, Betweenness Centrality, Closeness Centrality, Eigenvector Centrality, Page Rank and Clustering Coefficient. The proposed method has been tested using a test case. The experimental results and a detailed quantitative analysis show that this proposed schema is a more efficient way for identifying influential nodes in a large OSNs.

## REFERENCES

- [1] N. Wire, "Social networking's new global footprint," Nielsen Online Report, 2009.
- [2] C. Canali, S. Casolari, and R. Lancellotti, "A quantitative methodology to identify relevant users in social networks," in *Proc. of IEEE International Workshop on Business Applications of Social Network Analysis*, vol. 50, no. 22, pp. 1-8, 2011.
- [3] L. D. C. C. Chinchilla and K. A. R. Ferreira, "Analysis of the behavior of customers in the social networks using data mining techniques," in *Proc. of 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 623-625, 2016.
- [4] P. D. Meo *et al.*, "Finding similar users on Facebook," *Information Science Reference*, pp. 304-323, 2012.
- [5] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, "Sentiment analysis of Twitter data," in *Proc. of the Workshop on Languages in Social Media*, pp. 30-38, 2011.
- [6] J. Mills, "Virality in social media: The SPIN framework," *Journal of Public Affairs*, vol. 12, no. 2, pp. 162-169, 2012.
- [7] Abhishek, S. Akshay, M. Dharmaparakash, G. Rajiv, and G. Utkarsh, "Quantifying virality of information in online social networks," *International Journal of Social Networking and Virtual Communities*, vol. 1, no. 1, pp. 19-30, 2012.
- [8] M. D. Laat, V. Lally, L. Lippinen, and R. J. Simons, "Investigating patterns of interaction in networked learning and computer-supported collaborative learning: A role for social network analysis,"

*International Journal of Computer-Supported Collaborative Learning*, vol. 2, no. 1, pp. 87-103, 2007. Doi: 10.1007/s11412-007-9006-4

- [9] G. Sabidussi, "The centrality index of a graph," *Psychometrika*, vol. 31, pp. 581-603, doi: 10.1007/bf02289527.
- [10] C. Dunne and B. Shneiderman, "Improving graph drawing readability by incorporating readability metrics: A software tool for network analysis," University of Maryland, Human-Computer Interaction Lab Tech Report (HCIL), 2009.
- [11] T. Opsahl, F. Agneessens, and J. Skvoretz, "Node centrality in weighted networks: Generalizing degree and shortest paths," *Social Networks*, vol. 32, no. 3, pp. 245-251, 2010.
- [12] S. P. Borgatti, "Centrality and network flow," *Social Networks*, vol. 27, no. 1, pp. 55-71, 2005.
- [13] P. Bonacich and P. Lloyd, "Eigenvector-like measures of centrality for asymmetric relations," *Social Networks*, vol. 23, no. 3, pp. 191-201, 2001.
- [14] J. Kim and J. G. Lee, "Community detection in multi-layer graphs: A survey," *SIGMOD Rec.*, vol. 44, no. 3, pp. 37-48, 2015.
- [15] M. Ahsan, T. Singh, and M. Kumari, "Influential node detection in the social network during community detection," in *Proc. of 2015 International Conference on Cognitive Computing and Information Processing (CCIP)*, 2015.
- [16] A. Farooq, G. Javaid, M. Uzair, and U. Akram, "Detection of influential nodes using social network analysis based on network metrics", in *Proc. of Computing, Mathematics and Engineering Technologies*, doi: 10.1109/ICOMET.2018.8346372.



**Aftab Farooq** was born in Punjab Province, Pakistan in 1991. He received the B.S. degree in Computer Sciences from the University Dr. A. Q. Khan Institute of Computer Science and Information Technology (KICSIT) at Kahota in 2014 and he is currently the research scholar at College of Electrical and Mechanical Engineering (EME), NUST, Islamabad, Pakistan. His research interests include Data Engineering, and Network Security.



**Muhammad Usman Akram** is an assistant professor at College of Electrical & Mechanical Engineering, National University of Sciences & Technology, Pakistan. He holds a PhD degree in Computer Engineering with specialization in medical image analysis. His research areas are image/signal processing, biometrics, medical image analysis and pattern recognition.



**Gulraiz Javaid Joyia** was born in Punjab Province, Pakistan in 1989. He received the B.S. degree in computer engineering from the University College of Engineering and Technology (I.U.B), Bahawalpur in 2013 and he is currently the research scholar at College of Electrical and Mechanical Engineering (EME), NUST, Islamabad, Pakistan. His research interests include IoT, bioinformatics, data engineering, requirement engineering and computer networks security.



engineering.

**Chaudhry Muhammad Naeem Akbar** was born in Punjab Province, Pakistan in 1975. He received the BE degree in Electrical Engineering and MS in Computer Engineering from College of Electrical and Mechanical Engineering (EME), NUST, Islamabad, Pakistan. Presently he is a student of Ph.D. and also is a research scholar in computer engineering at College of Electrical and Mechanical Engineering (EME), NUST. His research interests include digital image processing, CPR and data