An Approach for Requirements Engineering Analysis using Conceptual Mapping in Healthcare Domain

Aya Radwan, A. Abdo, Sayed Abdel Gaber
Faculty of Computers and Artificial Intelligence, Helwan University, Cairo, Egypt

Abstract—**Healthcare systems aim to achieve the best possible support for patient care and to provide good medical care. Good analysis of requirements is essential to avoid any crises. Elicitation of healthcare systems requirements is an emerging and critical phase. It is a challenging task to deal with constraints from the stakeholders and restrictions of the legal issues. In this research, an approach "Conceptual Mapping for non-functional Health care Requirements; CMHR" is proposed to perform an analysis and to evaluate the relationship between the clinical non-functional requirements of medical devices (ventilators as an example in this research) according to the following five attributes: prioritization of requirements, suitability, feasibility, achievability, and risky. Requirements are automatically clustered using the K-means++ algorithm to find out the optimal number of clusters. Requirements are then clustered to visualize the concept map. Clustering is applied on different combinations of the attributes to sort the requirements and to visualize them. Label names are assigned to the classes of requirements to assign each requirement to the appropriate class. Consequently, a prediction of a new requirement can be figured automatically. The approach achieved less rework, fast delivery of the project with good quality, and achieved a higher level of user satisfaction.**

Keywords—Conceptual mapping; healthcare systems; clustering; requirement engineering analysis

I. INTRODUCTION

Healthcare systems aim to achieve the best possible support for patient care and to provide good medical care. software engineering is an emerging field for healthcare systems. Clinical requirements are more complex comparing to other types of requirements as they should be in specific terms. Software requirement engineers should concentrate when eliciting requirements to avoid making mistakes that could be disastrous in healthcare [1]. Software Requirement analysis; also known as software requirement engineering, aims to ensure that systems meet the needs of stakeholders while also determining user expectations. Requirement analysis consists of activities that are important to the software development process [2]. This process is composed of the following activities [2]: elicitation, modeling and analysis, assurance, management, and Evolution. First, the elicitation phase is called software requirements elicitation, which is a challenging task to understand and analyze software requirements specifications. Its objective is discovering stakeholders' needs, and understand the context [2]. It's important to perform this step carefully, to minimize the changes that may occur in order to save software development time [3]. Second, requirements analysis and negotiation; requirements are identified, and system modeling is

performed. Third, requirement specification while the requirement should be documented in a format of Software Requirements. Fourth, validate the requirements to ensure that requirements meet the needs of the stakeholders. Finally, Requirements management is the last activity to manage all requirements-related activities [4] [5]. Further, the requirements are written in natural text format, requirements should be thoroughly analyzed to generate the software requirements specifications needed to validate and verify the final product. Software requirements are divided into Functional and non-Functional requirements [6]. Functional requirements are about the software service and its functional behavior, while non-functional requirements focus on the performance and the quality of requirements that are unrelated to software functionality [7].

The healthcare system is an emerging field. Projects should be well analyzed to avoid any crises. The consequences of poor requirement analysis lead to project failure, consuming effort and time, and repeating tasks and processes. In this research, an approach is presented to analyze non-functional requirements through the use of unsupervised learning to arrange the related requirements into clusters. This approach visualizes and sorts the nonfunctional requirements through conceptual mapping. Furthermore, the approach can predict the class of the new requirements.

The rest of this paper is organized as follows: Section 2 explains the background of some related concepts, Section 3 illustrates the literature review, Section 4 discusses the proposed methodology, and finally Section 5 explains the experimental results.

II. BACKGROUND

Machine learning consists of three different types, which are supervised learning, unsupervised learning, and reinforcement learning [8]. In this research, the k-means clustering method is used to group similar data into one group. The related data is combined to make it easier to find relationships between data.

The requirement engineering consists of a series of activities, [4] that begin with requirement elicitation and inception to meet the needs and desires of the stakeholders. Then requirements analysis and negotiation; requirements are identified, and system modeling is performed, which necessitates that the product is completely modeled and designed prior to the construction. Then requirement specification; requirement should be documented in a format of SRS. While the validation of the requirements to ensure

that they meet the needs of the stakeholders is the next activity. Finally, Requirements management is the last activity to manage all requirements-related activities. [5].

Conceptual mapping is a diagram that visualizes the semantic relations between concepts [9]. Conceptual maps help to give a consistent evaluation. This process mainly composed of the following steps; first, make brainstorming then generate the statement of needs, after that sorting and rating these statements, then represent the statements on the map that's could be called as conceptual mapping analysis. Finally, they make interpretation for the map and utilize it [10][11]. Conceptual mapping creates a structure of clustered concepts and provides a visual or graphical representation of data.

III. LITERATURE REVIEW

Al-Dahmash et al. in [12], proposed SEMHTA, a methodology for healthcare to build system and application with reliable and protected software. It stands for software engineering methodology for healthcare applications development, this methodology relies on what developers do when building a different system. There was no consideration for the elicitation phase. Gausepohl et al. in [13], proposed a methodology in a healthcare system that focused on storytelling in the elicitation phase for medical device requirements. Results contributed similar quantity and breadth of information in significantly less time. Participants contributed more distinct context-of-use information with an emphasis on the social context using ontology technics. However, this methodology presents the elicitation of medical device requirements rather than a clinical requirement. Kaiya et. al. [14] did similar work to enhance requirement elicitation using ontology technical web mining and lightweight natural language processing, but this wasn't applied with healthcare systems. Widya et al. in [15], proposed a methodology for eliciting requirements in the eHealth domain. Developers developed a scenario that reflects the treatment protocol and it works with telemedicine treatment. Martin et al. in [16], proposed a user center designed approach in the clinical systems to keep users in the cycle of development. They performed semi-structured open-ended interviews to investigate the clinical need for the device as well as the supposed effects for patients and clinical users. Regarding approaches in requirements engineering in the non-clinical systems: Laporti et al. in [17], proposed the Athena approach that applied requirements elicitation by grouping storytelling from users to be merged in one story. Stories then transformed to scenarios then to use-cases. Andreas et al. in [18] explore and define a requirement engineering methodology for machine learning systems. This methodology incorporates additional types of standards, such as special legal requirements, and explains the ability and freedom from prejudice, in order to enhance the requirement engineering process. They improve the machine learning models by taking too many decisions. Kamal Rudin et al. [19] [20] have created a simple method for gathering consistent requirements from the client stakeholders. A library pattern that supports various application domains is developed to store the essential requirements following the essential use-case. The library pattern, on the other hand, is not concerned with the use of

healthcare requirements. Hamzeh Eyal Salman in [21], designed an approach to cluster functional requirements automatically based on semantic measure Using Agglomerative Hierarchical Clustering (AHC) by grouping similar functional requirements into clusters. Results achieved high performance according to a well-known measure and didn't apply yet for the clinical system. Zeng Zhen in [22], applied the TA-ART algorithm ta generate an automatic concept map for text, by making text analysis and set records into clusters. Nadiah Daud in [23] created a methodology to enhance requirements elicitation. This methodology is composed of three phases; 1) the analysis: they collect requirements to analyze it or read the literature and gather requirements to analyze gaps and requirements. 2) they go to the design and development. 3) they make an evaluation and test the results. Rebecca Orsi in [10] presented a study using concept maps analysis by running more than multiple cluster analyses to describe a quantitative validity analysis. She used the R statistical software methods and packages to represent four clustering methods.

IV. PROPOSED APPROACH

This research presents a novel approach; "Conceptual mapping for health care non-functional requirements, (CMHR)". This approach is applied to requirement engineering to enhance the analysis process of the requirements. CMHR is composed of three phases as illustrated in Fig. 1. The first phase is Requirements Gathering. In this phase, the non-functional requirements are extracted from the software requirements specifications. The domain experts rank requirements by the identified attributes (suitability, priority, stability, risky, achievable) from 1 to 5. The second phase is Data Preprocessing. In this phase, the data pass through preprocessing phase to prepare the data for the next step and to ensure that the data is cleaned. The third phase is Machine Learning and visualization. In this phase, the approach starts to perform clustering using the k-means++ algorithm to perform the conceptual mapping analysis. In this phase, The Elbow method is used to calculate the number of clusters, then clustering the related requirements to each other based on similarity according to the identified attributes. Then, a visualization is performed through a matplotlib plot, and the data is exported to CSV file. Now labeling of the clusters and requirements can be performed. The benefit of labeling the data is to make a classification for the requirements and split it into the training and testing set. Finally, predict the new requirements classification. The details of these phases are decelerating in the following diagram.

Phase 1: Requirements Gathering. This phase consists of two main steps. Step1: Extract non-functional requirements from SRS. The structured requirements are extracted and the non-functional requirements are being selected to analyze it. In this research, 104 non-functional requirements are collected from SCRIBD1 website about ventilators as a medical device. Table I illustrates a sample of these requirements.
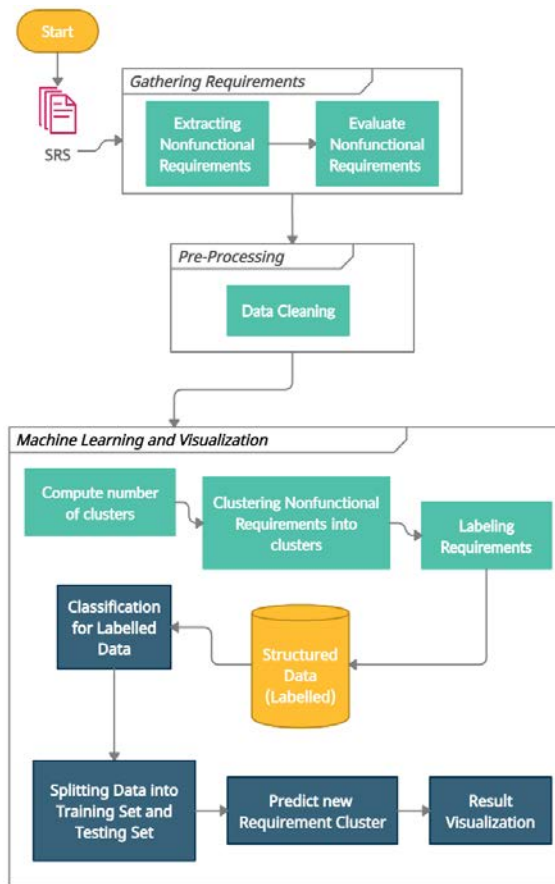

Fig. 1.   Conceptual Mapping for Health Care non-Functional Requirements: CMHR.

TABLE I.   A Sample of Non-functional Requirements for Ventilators Dataset

12.1 For devices which incorporate software or which are medical software in themselves, the software must be validated according to the state of the art taking into account the principles of development lifecycle, risk management, validation and verification.
12.2. Devices where the safety of the patients depends on an internal power supply must be equipped with a means of determining the state of the power supply.
12.3. Devices where the safety of the patients depends on an external power supply must include an alarm system to signal any power failure.
12.4. Devices intended to monitor one or more clinical parameters of a patient must be equipped with appropriate alarm systems to alert the user of situations which could lead to death or severe deterioration of the patient's state of health.
12.5. Devices must be designed and manufactured in such a way as to minimize the risks of creating electromagnetic fields which could impair the operation of other devices or equipment in the usual environment.
12.6. Devices must be designed and manufactured in such a way as to avoid, as far as possible, the risk of accidental electric shocks during normal use and in single fault condition, provided the devices are installed correctly.
12.7.1. Devices must be designed and manufactured in such a way as to protect the patient and user against mechanical risks connected with, for example, resistance, stability and moving parts.
12.7.2. Devices must be designed and manufactured in such a way as to reduce to the lowest possible level the risks arising from vibration generated by the devices, taking account of technical progress and of the means available for limiting vibrations, particularly at source, unless the vibrations are part of the specified performance.

Step 2: Requirements Evaluation. In this step, a group of qualified scientists evaluates the non-functional requirements from the SRS document with attributes. The selected attributes are suitability, priority, stability, risky, achievable. Then the following questions are asked about every single requirement to a domain expert.

- Suitability: How much this requirement is complete, correct, and appropriate?

- Priority: How much this requirement is important?

- Stability: Does the requirement has reached a certain level of stability?

- Risk: What's at stake if the requirement isn't implemented?

- Achievable: How far this requirement could be achieved?

The mean of the answers of the survey is taken and be considered during analysis.

Phase 2: Pre-processing. Data cleaning is the process; which implemented at this phase. The kernel-based random method is used. This method finds the relations between concepts or data to perform data cleaning and normalization [24]. Data cleaning detects errors, expels mistakes and irregularities from information in arrange to progress the quality of information [25]. So, this step, try to fill the missing data by setting the most relevant value by calculating the mean value using machine learning.

Phase 3: Machine Learning and Visualization. This phase consists of four main steps. Step1: Compute the number of clusters using the ELBOW method to determine the optimal number of clusters for the data. It uses the k-means++ algorithm and applies a range of values for the k from 1 to 10 on the dataset to calculate the WCSS. WCSS stands for the Within-Cluster-Sum-of-Squares. It's the sum of squares of the distance of each data in all clusters to their centroid point using the following formula:

$$\text{WCSS} = \sum_{C_k}^{C_n} \left(\sum_{d_i \ in \ C_i}^{d_m} distance(d_i, C_k)^2 \right)$$

Where,

C is the cluster centroids. And d is the data point in each Cluster.   (1)

Step 2: Clustering similar Non-Functional Requirement: The k-means++ algorithm used to present multidimensional scaling by measuring similarity, which can be defined as the distance between various datasets; calculating the sum of the distance between point and centroid point. After calculating the distance, the group of related clusters is combined then the requirements could be visualized among the selected attributes throw the multidimensional scaling of the conceptual map and can be exported through CSV file. Requirements could be dynamically sorted based on any two selected attributes. Also, the map could fit for three attributes is developed.

Step 3: Labeling Requirements: Data labeling is the method of identifying the raw data. Data here is ready to be

labeled after performing clustering; each requirement can be assigned with a label name. Assume there are four clusters, so a label will be assigned to each cluster based on the data in that cluster.

Step 4: Classification of labeled data: Data now in the shape of multi-label classes. Since the requirements are labeled in the last step, so the data is ready to be classified. The Naïve Bayes technique is applied in this research to make a classification for the labeled data. Naïve Bayes split the data into a training set and testing set in aim to learn the machine.

Step 5: Predicting the new requirement cluster: This is the final step. The developed approach can predict the classification of new requirements or for an edit of an existed requirement.

V. Experimental Study and Evaluation

The CMHR approach is applied to non-functional requirements of a ventilator device. First, 104 requirements were extracted from the SRS. Then the domain experts evaluated each requirement based on the defined attributes as it was explained in the previous section. The mean average of the evaluation was calculated and assigned to each requirement. Then the file is converted into CSV file. A sample of the data is presented in Table II.

The second step performs the preprocessing phase by using the sklearn library in python to complete the missing data. The data is ready now to perform the machine learning phase. This phase starts with computing the number of clusters using the ELBOW method to determine the optimal number of clusters, as shown in Fig. 2.

Fig. 2 depicts the result of the Elbow method. The clustering is applied based on prioritization and suitability. And figure showed the result of the optimal clusters is five.

Fig. 3 depicts the result of the Elbow method. Prioritization and achievability are used to cluster the requirements. The figure showed that the optimal clusters is four.

The next step is to cluster similar non-functional requirements using the k-means++ algorithm. The multidimensional scaling of the requirements is presented in Fig. 4 and Fig. 5.

In Fig. 4, Requirements are sorted and visualized based on prioritization and suitability attributes. Requirements are clustered into the five clusters as the Elbow method showed.

In Fig. 5, Requirements are sorted and visualized based on prioritization and achievability attributes. Requirements are clustered into four clusters as the Elbow method showed.

CMHR approach is implemented dynamically. It offers clustering the requirements based on three attributes and makes a 3-Dimensions representation. For example, Assum the clustering is made based on three attributes prioritization, suitability, and achievability.

Fig. 6 presents the result of the Elbow method. The clustering is applied based on prioritization, achievability and suitability. The figure showed that the optimal clusters is five.

Going through the next step to cluster similar non-functional requirements, the multidimensional scaling of the requirements is presented using the k-means++ algorithm in Fig. 7.

TABLE II. Ranking of Non-functional Requirements for Ventilators (Sample of the CSV File)

| Requirements	Suitable	Priority	Achievable	Feasible	Risky
For devices which incorporate software or which are medical software in themselves, the software must be validated according to the state of the art taking into account the principles of development lifecycle, risk management, validation and verification.	4.1	3.7	4.5	4.96	4.9
Devices where the safety of the patients depends on an internal power supply must be equipped with a means of determining the state of the power supply.	4.05	3.7	4.35	4.9	4.87
Devices where the safety of the patients depends on an external power supply must include an alarm system to signal any power failure.	4.78	4.56	4.3	4.74	4.3
Devices intended to monitor one or more clinical parameters of a patient must be equipped with appropriate alarm systems to alert the user of situations which could lead to death or severe deterioration of the patient's state of health.	3.82	3.48	3.88	4.36	3.66
Devices must be designed and manufactured in such a way as to minimize the risks of creating electromagnetic fields which could impair the operation of other devices or equipment in the usual environment.	3.88	3.89	4.9	3.8	3.8
Devices must be designed and manufactured in such a way as to avoid, as far as possible, the risk of accidental electric shocks during normal use and in single fault condition, provided the devices are installed correctly.	4.52	1.72	4.87	3.1	4.67
Devices must be designed and manufactured in such a way as to protect the patient and user against mechanical risks connected with, for example, resistance, stability and moving parts.	4.85	4.35	2.61	2.6	1.4
Devices must be designed and manufactured in such a way as to reduce to the lowest possible level the risks arising from vibration generated by the devices, taking account of technical progress and of the means available for limiting vibrations, particularly at source, unless the vibrations are part of the specified performance.	2.1	4.3	4.1	3.8	1.5
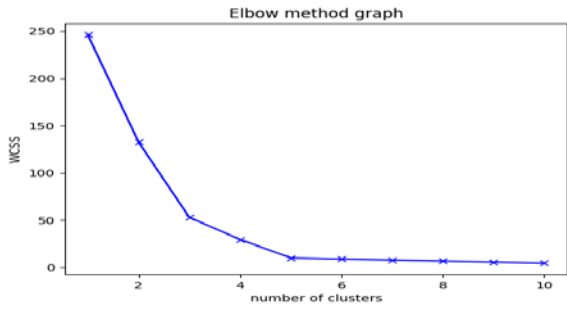
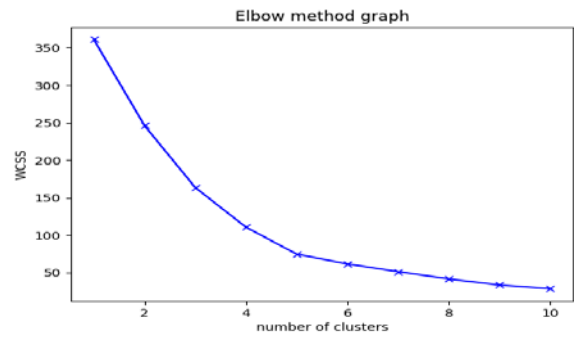Fig. 2.   Elbow Method Graph for Prioritization and Suitability Attributes.
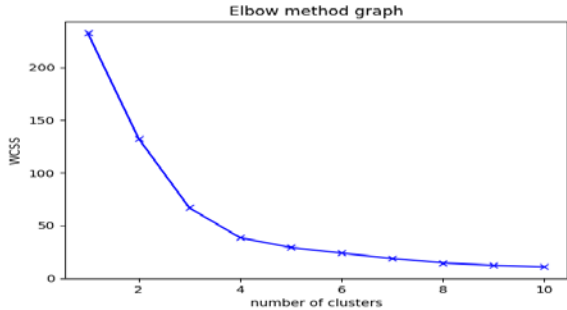

Fig. 3.   Elbow Method Graph Prioritization and Achievability Attributes.

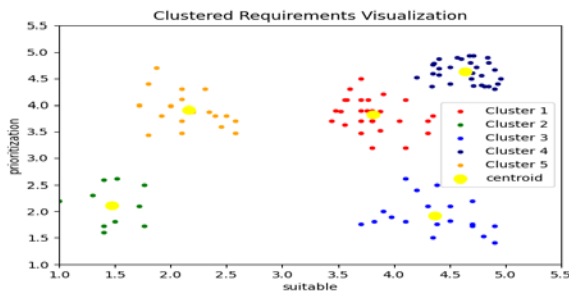Fig. 4.   2D Visualization for Five Clustered Requirements.

Fig. 5.   2D Visualization for Four Clustered Requirements.

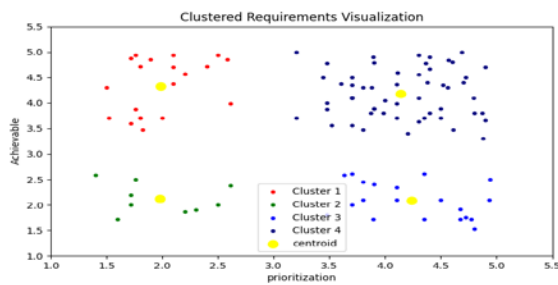Fig. 6.   Elbow Method Graph for Assigned Three Attributes Prioritization, Achievability and Suitability.

Fig. 7.   3D Visualization for Clustered Requirements.

In Fig. 7, K-means++ algorithm runs based on the evaluation of three attributes, which are suitability, prioritization, and achievability, Requirements are clustered into five clusters as the Elbow method showed. The Requirements sorted and visualized through 3-Dimensions matplotlib.

Silhouette method is applied to evaluate the clustering. This technique used to determine the performance and how much the clustering technique that has been used is correct. [26] Silhouette scope determines the measures between every object in the cluster and the other objects from other clusters. [26] Silhouette coefficient range from -1 to 1, where 1 implies that clusters are well separated and the objects are related to each other at the same cluster, while 0 implies that clusters are not separated well and distance between clusters are too close. -1 implies that objects and clusters are wrongly appointed. [27] Silhouette score coefficient is calculated for requirements to determine the optimal number of clusters. It tested from 2 clusters to 6 clusters and the result is presented here in Fig. 8 to 12.
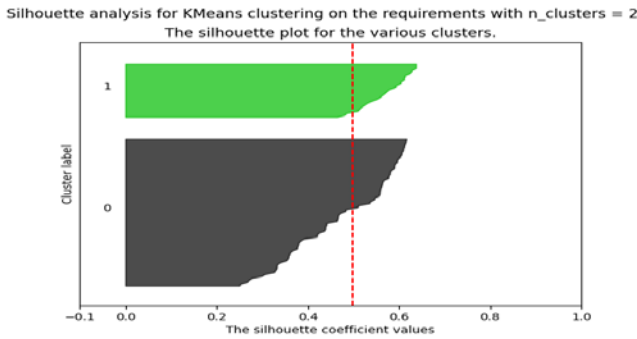
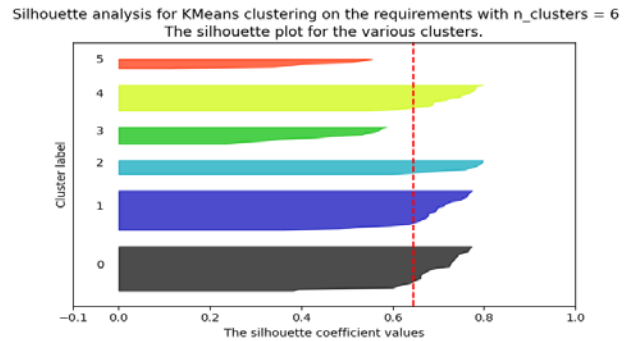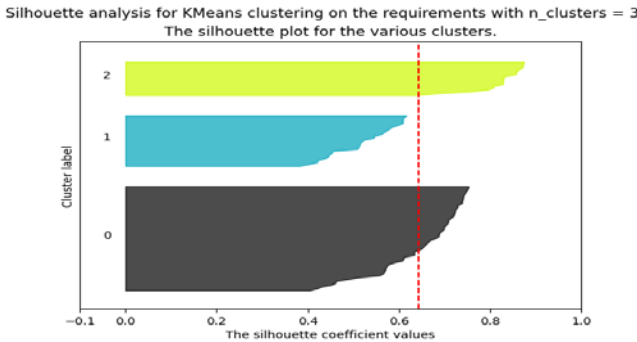Fig. 8.   The Silhouette Coefficient Value for Two Clusters.

Fig. 9.   The Silhouette Coefficient Value for Three Clusters.
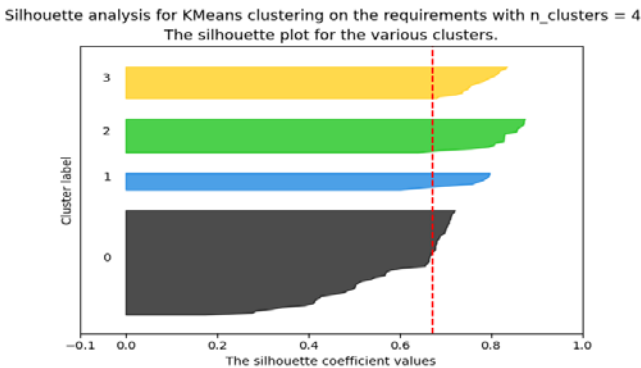

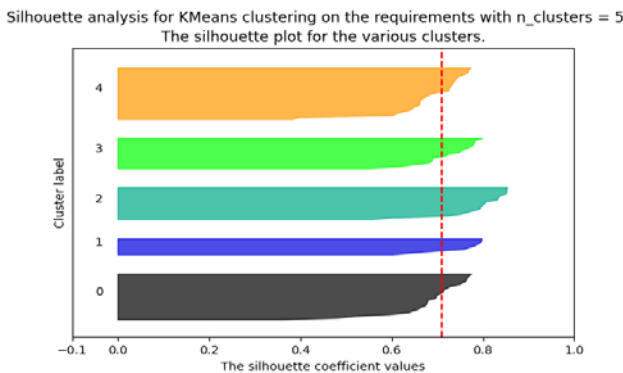Fig. 10. The Silhouette Coefficient Value for Four Clusters.

Fig. 11. The Silhouette Coefficient Value for Five Clusters.

Fig. 12. The Silhouette Coefficient Value for Six Clusters.

Fig. 8 to 12 shows the silhouette analysis for Kmeans clustering. The cross column on the x axis shows the silhouette score as it's explained in Table III. The thickness of the silhouette plot represents the related requirements for each cluster. Fig. 11 shows the silhouette score is 0.71.

In Table III, shows silhouette score for the different number of clusters.

Fig. 4 showed that there are five clusters. Labeling of these clusters showed in Table IV.

Since the requirements were labeled then the data can be classified. Data were split into a training and testing set; with 80% for training and 20% for testing. The data splitting is to confirm the predicting of the class of any new element. The number of tested requirements is 21. The Confusion matrix is a visual evaluation technique used in machine learning. It presents the actual class result associated with the prediction class result [28]. The result of the confusion matrix is presented in Table V and the accuracy score was 1 for the tested cases.

TABLE III.   AVERAGE SILHOUETTE SCORE FOR NUMBER OF CLUSTERS

Number of clusters	Silhouette score
2	0.49
3	0.64
4	0.67
5	0.71
6	0.64

TABLE IV.   LABELING THE CLUSTERS

Cluster Number	Label Name
1	High Priority High Suitability
2	Very Low Priority Very Low Suitability
3	High Suitability Very Low Priority
4	Very High Priority Very High Suitability
5	High Priority Low Suitability

TABLE V.   CONFUSION MATRIX FOR THE TESTING SET

Class 1	Class 2	Class 3	Class 4	Class 5
8	0	0	0	0
0	3	0	0	0
0	0	6	0	0
0	0	0	2	0
0	0	0	0	2

In Table V, 21 requirements had been tested with the classification technique Naïve Bayas. Results are shown in the table that eight requirements belong to class 1. Three requirements belong to class 2. Six requirements belong to class 3. Two requirements belong to class 4. Two requirements belong to class 5.

The classification report is used to assess the accuracy of a classification algorithm's predictions. The classification report presents the precision, recall, f1-score.

Precision = (TP/(TP+FP))

Recall = (TP/(TP+FN)).

The classification report shows that the precision is one for the five classes. The recall is one for the five classes. The F1-Score is one for the five classes.

VI. CONCLUSION AND FUTURE WORK

In this research; a novel approach "Conceptual mapping for health care non-functional requirements; CMHR" is applied to enhance requirement engineering analysis of healthcare systems. The approach focuses on ranking requirements using a set of attributes and calculates the mean value of each requirement for every attribute. CMHR is applied using the K-means++ algorithm to gather the related requirements into one group according to its semantic similarity. Semantic similarity is captured using conceptual mapping of the requirements. The CMHR approach can be extended to any different number of clusters. Clusters can be defined based on two or three attributes to sort the data based on the attributes and visualize it. Assigning labels to the clusters is essential to identify every cluster and requirement, in order to make a classification. CMHR offers an automatic classification for a new requirement. CMHR is applied to medical device requirements and generated five clusters of the requirements depending on prioritization and suitability with a silhouette score 0.71. Also, generated four clusters depending on prioritization and achievability. It generated five clusters based on three attributes. CMHR can work dynamically between any two or three attributes. In future work, CMHR will be applied on more than project to consume the time will be saved to finish the project. More attributes will be found to enhance CMHR approach and to find relations between requirements. Swarm intelligence can be applied to get more enhancement for the proposed approach.

REFERENCES

[1] Safiah Sidek 1, Massila Kamalrudin 2, Nadia Daud3 Eliciting Functional Requirements in Clinical Domain, International Research Journal of Engineering and Technology (IRJET), vol. 4, No. 06, 2017 e-ISSN: 2395 -0056.

[2] Amel Bennaceur, Thein Tun, Yijun Yu, Bashar Nuseibeh, Requirements Engineering Handbook of Software Engineering, April 2018.

[3] Sharon McGee, Des Greer, towards an understanding of the causes and effects of software requirements change: two case studies, Springer, 2012.

[4] Muhammad Azeem Akbar , Ahmed Alsanad, Sajjad Mahmood, Abeer Abdulaziz Alsanad, and Abdu Gumaeia, systematic study to improve the requirements engineering process in the domain of global software development, IEEE Access, vol.8, 2020, pp 53374.

[5] Danial Mendez Fernandez, Stefan Wagner, Naming the pain in requirements engineering: A design for a global family of surveys and first results from Germany, Information and Software Technology, 2013.

[6] Sommerville, I. Software Engineering, 9th ed.; Addison-Wesley Publishing Company: Boston, MA, USA, 2010.

[7] Vivek Shukla, Dhirendra Pandey and Raj Shree, Requirements Engineering: A Survey, Communications on Applied Electronics (CAE) –, 2015, ISSN: 2394-4714.

[8] Uday Shankar Shanthamallu, Andreas Spanias, Cihan Tepedelenlioglu, and Mike Stanley, A Brief Survey of Machine Learning Methods and their Sensor and IoT Applications, 2017, 8th International Conference on Information, Intelligence, Systems & Applications (IISA).

[9] Beat A. Schwendimann, Concept maps, In R. Gunstone (Ed.), Encyclopedia of science education, 2014.

[10] Rebeca Orsi, Use of multiple cluster analysis methods to explore the validity of a community outcomes concept map, Evaluation and program, 2016.

[11] Thomas J. Waltz, Byron Powell, et al., Use of concept mapping to characterize relationships among implementation strategies and assess their feasibility and importance: results from the Expert Recommendations for Implementing Change (ERIC) study, Implementation Science 10(109):1-8, 2015.

[12] Abdullah Al-Dahmash and Samir. El-Masri, A New Proposed Software Engineering Methodology for Healthcare Applications Development vol 3, No. 3, 2013, pp.1566–157.

[13] Kim Gausepohl, Woodrow W. Winchester, James D. Arthur, & Tonya Smith-Jackso, Using Storytelling to Elicit Design Guidance for Medical Devices. Ergonomics in Design: The Quarterly of Human Factors Applications, 19(2), 2011, pp.19–24.

[14] Haruhiko Kaiya, Yuutarou Shimizu et al., Enhancing Domain Knowledge for Requirements Elicitation with Web Mining. 2010, 17th Asia Pacific Software Engineering Conference, pp.3–12.

[15] Widya R G A Bults, et al., Scenario-based requirements elicitation in a pain-tele treatment application. ICSOFT 2009 - 4th International Conference on Software and Data Technologies, Proceedings, 2(July), 2009, pp.406–413.

[16] Jennifer L. Martin, Daniel J. Clark et al., A user-centred approach to requirements elicitation in medical device development: a case study from an industry perspective." Applied ergonomics, 43(1), 2012, pp.184–90.

[17] Vivian Laporti, Marcos. R. S. Borges, and Vanessa. Braganholo, "Athena: A collaborative approach to requirements elicitation. Computers in Industry", 60(6), 2009. pp.367–380.

[18] Andreas Vogelsang, Markus Borg, Requirements engineering for machine learning: perspectives from data scientists, international requirements engineering conference workshops, 2019, ISBN:978-1-7281-5165-6.

[19] Massila KamalRudin, John Grundy and John G. Hosking, Tool support for essential use cases to better capture software requirements." Proceedings of the IEEE/ACM international conference on Automated software engineering - ASE '10, 2010, p.255.

[20] Massila KamalRudin, John Grundy, "Generating essential user interface prototypes to validate requirements." 2011 26th IEEE/ACM International Conference on Automated Software Engineering, 2011, pp.564–567.

[21] Hamzeh Eyal Salman, Mustafa Hammad, Abdelhak-Djamel Seriai and Ahed Al-Sbou, Semantic Clustering of Functional Requirements Using Agglomerative Hierarchical Clustering, MDPI, 2018.

[22] Zengzhen Shao and Yancong Li, Research on a new automatic generation algorithm of concept map based on text analysis and association rules mining, Springer-Verlag GmbH Germany, part of Springer Nature, 2018.

[23] Nadiah Daud, Massila kamalrudin, Safiah Sidek, Rosmiza Wahida, Characterization of Healthcare Domain Library to Capture Accurate Requirements of Healthcare Software, e-Proceeding of Software Engineering Postgraduates Workshop, 2013.

[24] Salvador Garcia, Julian Luengo, Francisco Herrera, Data Preprocessing in Data Mining, Springer, vol. 72, 2015, ISSN 1868-4394.

[25] Jason Brownlee, Data Preparation for machine learning, 2020, Online book.

[26] Godwin Ogbuabor and Ugwoke, Clustering algorithm for a healthcare dataset using silhouette score value, International Journal of Computer Science & Information Technology (IJCSIT) Vol 10, No 2, 2018, DOI:10.5121.

[27] Meshal Shutaywi, Nezamoddin n. Kachoui, silhouette analysis for performance evaluation in machine learning with applications to clustering, MDPI, vol. 23, no.2, 2021, pp 759.

[28] Jianfeng Xu, Yuanjian Zhang, Duoqian Miao, Three-way confusion matrix for classification: A measure driven view, Information Sciences, 2019.