

COVID-19 World Vaccine Adverse Reactions Based on Machine Learning Clustering Algorithm

1st Lozan M. Abdulrahman
Researcher

Duhok Polytechnic University
Duhok, Iraq

lozan.abdulrahman@dpu.edu.krd

2nd Adnan Mohsin Abdulazeez
Research Center

Duhok Polytechnic University
Duhok, Iraq

adnan.mohsin@dpu.edu.krd

3rd Dathar A. Hasan
Researcher

Duhok Polytechnic University
Duhok, Iraq

dathar.hasan@dpu.edu.krd

<https://doi.org/10.48161/qaj.v1n2a59>

Abstract— In December 2019, a novel coronavirus, now named SARS-CoV-2, caused a series of acute atypical respiratory diseases in Wuhan, Hubei province, China. It triggered several acute atypical respiratory diseases. COVID-19 was the name given to the virus's disease. The infection is human-to-human transmissible, and it has triggered a global pandemic. Vaccines against COVID-19 are an essential global intervention to control the current pandemic situation and in a fairly short time, several vaccines have been developed to try to control the situation but have also led to consequences in the form of adverse reactions. Clustering algorithms have been used in computational intelligence and digital analysis, which is one of the areas that has taken this into account. Clustering can be described as a method of grouping similar data into one population or cluster and separating unrelated data into another. For clustering COVID-19 vaccine adverse reactions datasets, a variety of clustering algorithms are used. The objective of this paper is to use the clustering algorithms used in the case of COVID-19 Vaccine Adverse Reactions datasets, demonstrating how these algorithms help to provide accuracy for clustering the COVID-19 Vaccine Adverse Reactions. This study compared four clustering algorithms using the WEKA tool. Furthermore, it details the datasets in terms of different variables of precision, cluster case, number of iterations, time, and present the findings of these papers, and which clustering algorithms used and the accuracy of these algorithms. It is found that the clustering algorithm k-means is used widely in different types of the COVID-19 vaccine adverse reactions datasets with high accuracy.

Keywords— COVID-19 Vaccine Reactions, Machine Learning, Clustering Algorithm, VARSE, Weka.

I. INTRODUCTION

Nowadays, machine learning (ML) used in every area of computational work where algorithms are designed, and performance is increased [1,2]. In the last years, learning from unbalanced data sets has become a critical problem in machine learning and is frequently found in several applications such as computer security [3,4], engineering [5,6], remote sensing [7], biomedicine [8]. Corona virus-19 disease (COVID-19) has now spread worldwide to become a global pandemic affecting millions of people and claimed the lives of people worldwide [9,10]. COVID-19 symptoms may

vary from one individual to another. Vaccination will be an effective weapon in the battle against the COVID-19 pandemic, considering the shortage of specific therapy and the virus's rapid dissemination [11]. More recently, on December 11, 2020, and December 18, 2020, respectively, the US Food and Drug Administration (FDA), granted emergency authorization to the Pfizer/BioNTech and Moderna COVID-19 vaccines. These two COVID-19 vaccines were developed quickly to benefit humanity and arrest the rise in the number of COVID-19 cases [12,13]. From the time when the COVID-19 genome was released in early 2020 until these two vaccines received EUA status, less than one year passed [14]. The fastest any vaccine had previously been developed. There have been some concerns about the potential negative effects of these vaccines and therefore a reason why some people should avoid taking them [15,16,17,18]. Above all, it has been observed that the COVID-19 vaccine can induce mild side effects such as discomfort, redness, or swelling at the injection site, fever, weakness, headache, muscle aches, nausea, vomiting, scratching, chills, muscle stiffness, and joint pain after the first or second dosage, and can also trigger an anaphylactic shock in extreme cases. However, more severe side effects such as (Bell's palsy), facial inflammation, and swelling of the mouth, face, or tongue consistent with anaphylaxis have been recorded in certain instances [19]. Machine learning, like some other technologies, played a crucial role in determining the virus's triggers and conditions [20]. It was an attempt to clean up the noisy data that had scattered across the world in order to educate biological areas where research was attempting to understand how the virus resides beyond the human body and the effects of various factors such as climate, population, and on COVID-19 spread. Furthermore, the cleansing findings that can be used to maintain power the vaccines to help combat the latest pandemic would be global, some have been created to fight it, but several have brought on side effects. Clustering is often a valuable function for learning data. Uncontrolled clustering is known as the segmentation of data into clusters that contain the same data, mainly to make homogeneous groups [21,22]. Clustering algorithms are used in a variety of applications, including image segmentation, data cleaning, and exploratory analysis, content processing, web page grouping, consumer

segmentation, and science and engineering analysis [23,24]. Clustering can also be used as a clustering process to classify pattern groups for supervised classification later on. Clustering is a subject of active study in many areas, including analytics, pattern detection, machine learning, and data mining. Clustering, as previously mentioned, is a method for categorizing data into related and dissimilar groups. That is homogeneous clusters. Simply placed, a clustering algorithm divides a wide number of data points that share all features into a smaller number of groups [25]. Clustering is the practice of grouping items into classes that are related in any way. The k-means algorithm, hierarchical, and assumption and maximization algorithms are often used in data mining. This study aimed to address automated machine learning (ML) cluster algorithms for the adverse Reactions vaccine. To achieve this aim, we need a particular dataset of the corona vaccine adverse reactions to apply in the Waikato environment for the knowledge analysis (WEKA) platform tool. The rest of this paper is structured as follows: The related work is in section two, methodology in section three, experimental result and discussion in section four, finally conclusion in section five.

II. LITERATURE REVIEW

The COVID-19 pandemic created a great deal of concern and problems globally and therefore many studies have been done during the years 2020 and 2021 to try to investigate the connection between the COVID-19 disease and the vaccine outcomes. In this section, we highlight some of these studies where they used algorithms found in Machine learning.

Carrieri et al. [26] presented Vaccine hesitancy as an exaggerated danger to the next Vaccine Trust Crusade. used machine-learning algorithms to forecast populations at high VH risk that can effectively be used by policymakers to focus on regional indicators. highlighted strategy with infant vaccination drives in six 408 Italian municipalities for seven non-obligatory vaccinations. The region in the receiver operating characteristics (ROC) curve is related to a battery of machine learning models. The Random Forest algorithm better forecasts areas with a high VH chance, increased by 24 percent in terms of accuracy to an uncertain baseline level. The waste recycling proportion and the work rate are the most potent predictors of high HV among the area-level indicators. This will help politicians work on prevention measures at the privacy stage.

Siddiqui et al. [27] K-means clustering was used to test his results based on the presence of a correlation between the heat and the three instances of COVID-19 disease (complete, suspected and death cases). The datasets for this study were gathered from the WHO for various regions of China. When they used K-means clustering, they discovered that the clustering of 30 classes had happened, and the key result was that heat was not only a factor in the distribution of the COVID-19 pandemic. They proposed that other causes could have played an effective role in the spread of COVID-19 by analyzing the impact of temperatures in the three cases of the disease (confirmed, presumed, and death).

Ong et al. [28] Presented Vaxign and the recently developed reverse vaccine ML-technology machine learning tools to

predict vaccine candidates for COVID-19. Vaxign-ML also projected that the S, NSP3, and NSP8 proteins have good immunostimulatory properties. nsp3 was discovered to be more conserved in SARS-2, SARS-like, and MERS-like viruses and 15 and non-human primates than in SARS-CoV and humans and livestock. The protein was shown to bind promiscuous MHC class I and class II molecules, and the linear B-cell recognition sites were located on the protein's surface. identified several successful and noninvasive targets for our projected vaccine production. created an algorithm named Vaxi-ML that increases the prediction accuracy of it by 59%.

Carrillo [29] proposed a framework for categorizing countries based on the COVID-19 degree of impact by using (K-means). This method makes use of epidemic incidence figures, environmental quality, health sector forecasts, and economic crisis. The paper is based on an open-access dataset from approximately 155 countries. Using this number of countries (155), a three-PCA (principal component analysis) model grouped the countries into five categories based on the number of people infected with COVID-19. The report did not take into account the number of deaths or even the fatality rate. They used the ANOVA test to compare the effects of the K-means clustering. According to the comparison findings, the model of five to six clustering is (p0.001). The suggested procedure is only used in verified situations.

Pinter et al. [30] proposed hybrid machine learning to forecast COVID-19, and they use Hungarian data to illustrate its promise. The multi-layered perceptron-imperialist competitive algorithm (MLP-CIA) and the hybrid computer training methods of adaptive network-dependent fuzzy infusion (ANFIS) estimate the time series of contaminated people and the rate of mortality. The models estimate that the epidemic and general morality would decline dramatically by the end of May. Appropriate findings affirm the model's consistency which will be validated over 9 days. The model is supposed to remain accurate as long as there is no major disturbance. They have initial benchmarking to show the scope for future study of machine learning.

Gupta et al. [31] focused on clustering social network messages to determine whether or not the temperatures would impact the distribution of COVID-19. The datasets used consisted of 166,005 English tweets shared between January 23 and June 22, 2020. They used a machine-learning algorithm to clear the information from repetitive terms and topics in tweets. They deleted (4803) repetitive tweets and grouped the remaining tweets (23,752). They clustered the tweets using (K-means) strategies. The clustering groups were K= 25, so each group was associated with one of the 20 keywords' outputs. Cluster 10 was discovered to be concerned with the cold weather impact on COVID-19 distribution. Cluster 24 addressed the effects of elevated temperatures on the COVID-19, which discussed the various climates and the virus's general spread. Cluster 6 likened COVID-19 to the influenza virus, clusters 20 and 22 grouped tweets on how COVID-19 is sluggish to propagate and how social isolation influenced COVID-19 spread, and clusters 11, 14, 18, and 21 referenced related topics about science experts and Trump's contributions on Twitter about COVID-

19. The findings of this clustering phase revealed that 19,33.5 percent had no impact, 26.1 percent affected, and 40.4 percent had an unknown effect of weather on COVID.

Nguyen et al. [32] The hierarchical approach and DBSCAN were used to discover the relationship between COVID-19, bat CoVs, and pangolin CoVs by clustering the COVID-19 virus genus and the bat and pangolin virus genus sequences. The researcher used 334 genus sequences that were grouped into 12 categories of 12 major virus classes. The hierarchical methods verified that the cut-off (C) parameter is a significant variable that functions as a cutoff and that changing the value of C during the experiment is useful. In comparison to this approach, DBSCAN requires the identification of the key point as well as the definition of the minimum number of neighbour parameters. According to the findings of both algorithms' sets 2 and 8, SARS-CoV-2 refers to the Batcoronavirus genus.

III. METHODOLOGY

The methodology outlines many of the measures involved in doing a comparative study of clustering algorithms.

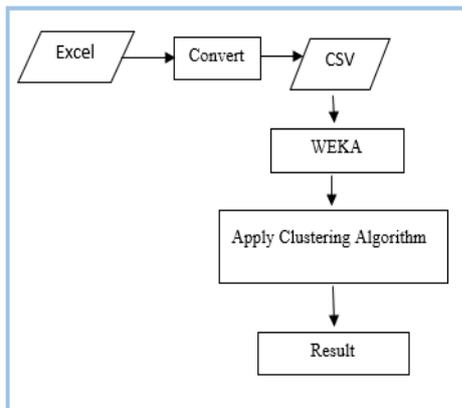


Fig.1. Flowchart of Research Methodology.

Step 1. Choose the clustering algorithms that are used in the comparative study.

Step 2. Select the Vaccine Adverse Event Reporting System (VAERS) Dataset in .CSV file format from (HHS.gov).

Step 3. Data to be loaded into WEKA is as follows: Load the data file into the program for further study.

Step 4. Apply clustering algorithms: On datasets, use all-clustering algorithms (algorithms K-means, expectation-maximization, heretical, and density-based).

Step 5. Save the results: The findings are tabulated after all algorithms have been performed and compared depending on the number of iterations, the amount of squared error, the time it takes to construct clusters and the correctly clustered output.

Step 6. Map the graph: Reflect the conclusions graphically.

A. Clustering Algorithms

A clustering algorithm divides a data set into many classes, with the similarity inside each group being greater than the similarity within groups. Furthermore, it seems that the bulk of the data obtained in several problems has certain underlying properties that contribute themselves too natural

groupings. Clustering algorithms are commonly used for data structure and categorization, as well as data compression and model creation [33],34,35].

- K-means clustering algorithm

k-means algorithm is a clear partition process. It separated (N) data objects into (K) cluster sets to obtain low cross similarity and high intracluster specificity. The mean diameter of a cluster's points, also known as the cluster's centroid, is used to evaluate cluster similarity [36,37,38]. It goes like this: first, choose **k** points at random as the cluster's mean (center). Following that, all artefacts are allocated to the **k** clusters with the shortest Distance measure between their centroids and objects. Mean is revised before it has been allocated to all of the artefacts. This update is continuing until the allocation is secure.

Steps:

Step 1: Get started: As initial centers, select k data items at random from Data collection D. K is the number of clusters.

Step 2: Repetition:

a) Assume that each cluster is a centroid.

b) Evaluate the difference between all of the data points, as well as the centroids.

c) Allocate di to the cluster that is nearest to you as a data object.

Step 3: By each cluster j (1=j=k), create an update. Calculate the cluster center once more.

Step 4: Repeat until there is no difference in the cluster's core.

Step 5: Finish.

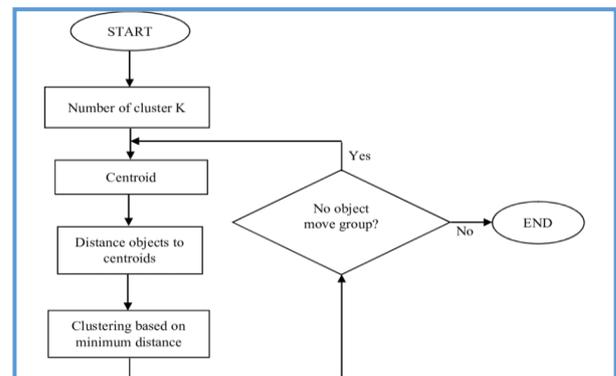


Fig. 2. K-means Clustering Process.

- Hierarchical clustering algorithm

By creating a hierarchy of clusters, also known as a dendrogram, the hierarchical Clustering approach combined or separates identical data items [39,40]. The hierarchical Clustering approach creates clusters in a step-by-step manner. Agglomerative and Divisive algorithms are the two kinds of hierarchical clustering algorithms. Agglomerative hierarchical clustering algorithm: Agglomerative hierarchical clustering is a bottom-up approach that starts from each person within a cluster. And, before all of the information is in any cluster or the user specifies otherwise, it blends the closest similar clusters by following some similarity criterion in the subsequent iteration. [41,42,43].

Steps:

Step 1: Get started: Assign a cluster number equivalent to the number of objects.

Step 2: Repetition: When the number of clusters is set to 1 or when the consumer specifies the number of clusters.

- a) Calculate the shortest inters cluster size.
- b) Combine the clusters with the smallest inter-cluster distance.

Step 3: Finish.

Divisive hierarchical clustering algorithm: This is a hierarchical clustering algorithm, as opposed to a bottom-up approach. The method of splitting hierarchical clustering begins with a single cluster comprising all data items. Following that, with each iteration, it separates into clusters by meeting certain resemblance requirements, before every piece of information entity creates its cluster or the stopping criteria are reached [44,45].

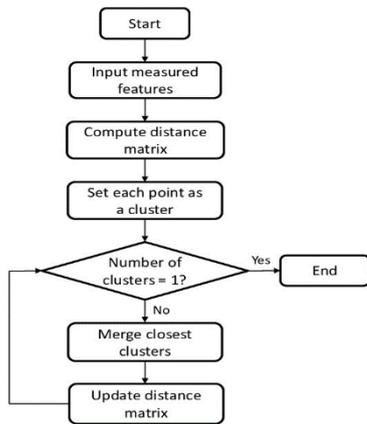


Fig. 3. Flowchart Agglomerative Hierarchical Clustering.

- Expectation and maximization (EM) algorithm

EM an unphased algorithm for estimating the maximum likelihood values of variables in unstructured statistical models. When it runs the expectation stage, it will calculate variables that maximize the assumptions discovered during the E phase; therefore, it must perform a maximization step in order to search for new ones. These parameter projections are then used in the next E steps to evaluate the distributions of the latent variables [46].

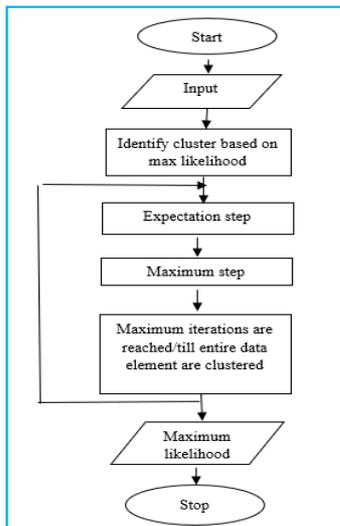


Fig. 4. EM Clustering Algorithm Process.

- Density-Based Algorithm

A cluster is a compact area pointed quote isolated from densely dense regions on the low-density side. When the clusters are irregular, there might be different ways to implement this cluster algorithm. It locates key artefacts, that is, objects with dense neighborhoods. It shapes dense regions as clusters by linking core objects and their surroundings. Clusters are made up of the most tightly linked points which can track noise and be used to find outliers [47,48].

Steps:

Step 1: Choose a random point r .

Step 2: Get the r neighborhood using ϵ .

Step 3: If the neighborhood's population density exceeds when the level is met, the clustering phase begins. Otherwise, the argument is labelled as noise.

Step 4: Continue until all of the points have been covered.

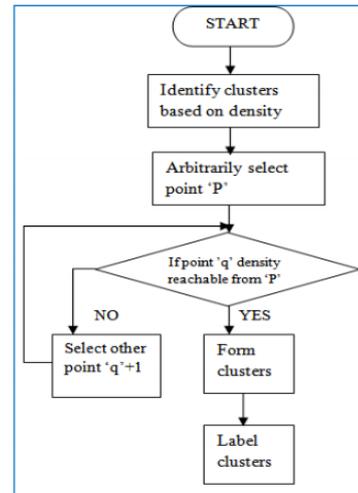


Fig. 5. Flowchart for Density-based algorithm process

B. Dataset

The COVID-19 World Vaccine Adverse Reactions dataset was used to do the comparison analysis. The datasets were downloaded from the website (HHS.gov). The data on Vaccine Adverse Reactions is related to consumer knowledge and contains 7 characteristics and 5471 cases. Data aggregation, data extraction, data discretization, vector transformation, and dealing with lost values are some of the techniques used in data analysis. The "VAERS" data is included in the paper as a.csv file. Table 1 shows the data set's architecture, with a description of each feature and the associated data form and restriction values.

TABLE 1. DESCRIPTION OF ATTRIBUTES USED IN COVID-19 WORLD VACCINE ADVERSE REACTIONS DATASET

Attribute	Explanation	Data Types	Constraint
VAERS_ID	VAERS Identification Number	Number	Not null
VAX_TYPE	Administered Vaccine Type	Char and Number	Not null
VAX_MANU	Vaccine Manufacturer	Char	Not null
VAX_LOT	Manufacturer's Vaccine Lot	Char and Number	Not null and Null
VAX_DOSE	Number of previous doses administered	Char	Not null and Null
VAX_ROUTE	Vaccination Route	Char	Not null and Null
VAX_NAME	Vaccination Name	Char and Number	Not Null

C. Weka

Weka is a Java-based development implement the expression was conceptualized and developed at the University of Waikato in New Zealand [49]. It is mainly a set of data mining machine learning algorithms in the form of WEKA. WEKA is open source and user-friendly with a user interface that helps get up and running quickly running. The algorithms may be added to the datasets directly or named from your java code. Data preprocessing, grouping, regression, clustering, correlation law, and visualization are all resources used in WEKA. The data collected is stored in the Feature File Format (.arff) and.csv format in the WEKA tool [49]. WEKA allows it simple to evaluate various response solutions using the same test approach and determine which one is best for the given challenge. It's written in Java and can operate on virtually every computer system. WEKA offers learning algorithm implementations that can easily execute any dataset. It also provides a range of methods for converting data sets. WEKA is a machine-learning algorithm development platform that is open-source [50].

Clustering in Weka: clustering is a pattern in Weka that separates the records in the database into classes. The properties of the groups in the same category are identical [51]. Differences within classes should be as large as possible, whereas differences within the same category should be as minimal as possible. Since there is no predefined level, it is classified as unsupervised learning [52,53]. Portioning methods, hierarchical techniques, density-based techniques, grid-based techniques, model-based techniques, clustering high-dimensional results, constraint-based clustering, and outlier study are all used in cluster analysis. Clustering can be done using a variety of algorithms [54,55].

IV. RESULT AND DISCUSSION

After briefly presenting clustering algorithms, we'll pass on a discussion of these algorithms focus on a case study. Using COVID-19 World Vaccine Adverse Reactions data, the experimental findings of each of the four clustering algorithms are presented in this section. Table 2 displays the effects of the procedure. For simpler evaluation and analysis, the simulation outcome is divided into many sub-items.

TABLE 2: CLUSTERING ALGORITHMS RESULT FOR COVID -19 WORLD VACCINE ADVERSE REACTIONS DATASET FOR TIME

Algorithms	No. of Clusters	Cluster instance	Time Take cluster (Second)	Log likelihood	Within cluster sum of squared error
K-means	2	3685 1786	0.01	-	6369.450
Hierarchical	2	5470 47	2.45	-	-
EM	2	3718175 3	0.17	-17.4045	-
Density based	2	3716175 5	0.03	-17.40902	6369.450

This paper focuses according to the time it takes to form the number of clusters, cluster instance accuracy, and the number of iterations while doing comparative analysis. The K-means algorithm requires the shortest time (0.01) seconds and has the best accuracy (67 %). A density-based algorithm distributes cluster instances more accurately, but it requires longer (0.03) seconds and has a lower precision (67%) than K-means.

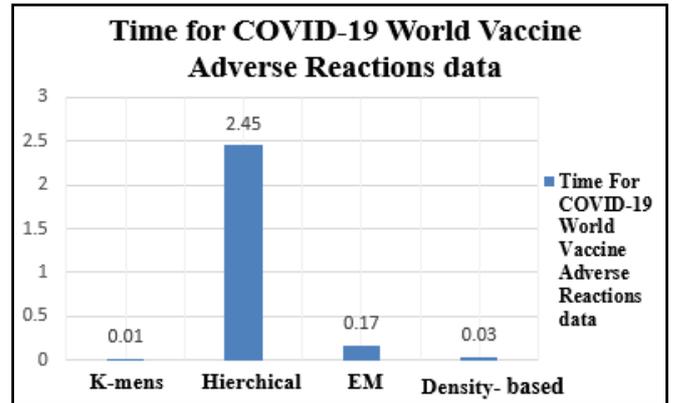


Fig. 6. k-means, density-based, EM, and hierarchical Clustering both require time to cluster datasets

From the analysis of clustering results shown in Table 3 we see the trends that Covid-19 (Pfizer BioNTech) has the highest accuracy of 67%, and Covid-19 (Moderna) 33% at different accuracy. Here, it becomes clear that the vaccine called Pfizer BioNTech has the most adverse reaction of all others in the world.

Table 3. K-means cluster results on Covid-19 world vaccine adverse reactions

Cluster ID	Population world	Vaccine Manufacturer	Vaccination Name	Accuracy
2	3685	Pfizer/BioNTech	Covid19(covid19 (Pfizer BioNTech))	67%
	1786	Moderna	Covid19(covid19 (Moderna))	33%

The k-means algorithm was found to be the most efficient for broad datasets of COVID-19 World Vaccine Adverse Reactions results. The consumer must select the number of clusters in the k-means, Hierarchical, and EM clustering algorithms. However, to render it density-based, the consumer does not need to choose the number of clusters; instead, a fitting collection of clusters is generated automatically. The density-based clustering algorithm is not sufficient for data with large density variance. Huge datasets were not supported by the hierarchical clustering algorithm. For noisy datasets, a hierarchical clustering algorithm is much more sensitive. To ensure more stable clusters, the EM clustering algorithm provides log-likelihood values for clusters. The EM clustering algorithm is an extension of the K-means algorithm that maximizes the assumption utilizing further repetition to satisfy the created clusters from the K-means algorithm. While this is a complicated algorithm, it can be parallelized to achieve the best results. It is preferred to use the K-Means algorithm or construct a density-based algorithm for a broad dataset. If the dataset includes more noisy results, the hierarchical clustering algorithm is the better alternative. For data with low variance, density-based clustering is preferable. Finally, we discover that k-mean is a quicker and safer algorithm than the others we tested.

CONCLUSION

Vaccines against covid-19 (and its emerging variants) are a critical global intervention in the present pandemic situation. Vaccines also cause side effects. Occurring Adverse Event

Following Immunization (AEFI) for most doses of vaccines. Therefore, we retrieved data sets on covid-19 world vaccine adverse reactions and have processed this information using four algorithms found in machine learning. It is evident that machine learning can significantly improve the development of the covid-19 pandemic. Machine learning unsupervised type clustering algorithm is used to accurately assess the issues with the vaccines. In this paper, a comparison of the clustering algorithms k-means, expectation-maximization, heretical, and density-based was carried out. The covid-19 world vaccine adverse reaction data were compared using the Weka tool, with the comparable outcomes in table and graph format the study is carried out on a time- and accuracy-based basis. After we applied our dataset on weka software the results show that the simple k-means clusters type clustering algorithms with a lower time and higher accuracy than most other algorithms.

REFERENCES

- [1] Abdulqader, D. M., Abdulazeez, A. M., & Zeebaree, D. Q. (2020). Machine Learning Supervised Algorithms of Gene Selection: A Review. *Machine Learning*, 62(03).
- [2] Chicho, B. T., Abdulazeez, A. M., Zeebaree, D. Q., & Zebari, D. A. (2021). Machine Learning Classifiers Based Classification For IRIS Recognition. *Qubahan Academic Journal*, 1(2), 106-118.
- [3] Ibrahim, I., & Abdulazeez, A. (2021). The Role of Machine Learning Algorithms for Diagnosing Diseases. *Journal of Applied Science and Technology Trends*, 2(01), 10-19 Conference on Advanced Science and Engineering (ICOASE) (pp. 106-111). IEEE.
- [4] Abdullah, D. M., & Abdulazeez, A. M. (2021). Machine Learning Applications based on SVM Classification A Review. *Qubahan Academic Journal*, 1(2), 81-90.
- [5] S. I. Saleem, A. M. Abdulazeez and Z. Orman, "A new segmentation framework for arabic handwritten text using machine learning techniques," *Computers, Materials & Continua*, vol. 68, no.2, pp. 2727-2754, 2021.
- [6]. Ahmed, O., & Brifceni, A. (2019, April). Gene Expression Classification Based on Deep Learning. In 2019 4th Scientific International Conference Najaf (SICN) (pp. 145-149). IEEE.
- [7] Zeebaree, D. Q., Haron, H., Abdulazeez, A. M., & Zebari, D. A. (2019, April). Trainable model based on new uniform LBP feature to identify the risk of the breast cancer. In 2019 International
- [8] Zeebaree, D. Q., Haron, H., & Abdulazeez, A. M. (2018, October). Gene selection and classification of microarray data using convolutional neural network. In 2018 International Conference on Advanced Science and Engineering (ICOASE) (pp. 145-150). IEEE.
- [9] Dashraath, P., Wong, J. L. J., Lim, M. X. K., Lim, L. M., Li, S., Biswas, A., ... & Su, L. L. (2020). Coronavirus disease 2019 (COVID-19) pandemic and pregnancy. *American journal of obstetrics and gynecology*, 222(6), 521-531.
- [10] Abdulkareem, N. M., Abdulazeez, A. M., Zeebaree, D. Q., & Hasan, D. A. (2021). COVID-19 World Vaccination Progress Using Machine Learning Classification Algorithms. *Qubahan Academic Journal*, 1(2), 100-105.
- [11] Pfefferbaum, B., & North, C. S. (2020). Mental health and the Covid-19 pandemic. *New England Journal of Medicine*, 383(6), 510-512.
- [12] Ong, E., Wong, M. U., Huffman, A., & He, Y. (2020). COVID-19 coronavirus vaccine design using reverse vaccinology and machine learning. *Frontiers in immunology*, 11, 1581.
- [13] Kannan, S., Subbaram, K., Ali, S., & Kannan, H. (2020). The role of artificial intelligence and machine learning techniques: Race for covid-19 vaccine. *Archives of Clinical Infectious Diseases*, 15(2).
- [14] MacKenna, B., Curtis, H. J., Morton, C. E., Inglesby, P., Walker, A. J., Morley, J., ... & Goldacre, B. (2021). Trends, regional variation, and clinical characteristics of COVID-19 vaccine recipients: a retrospective cohort study in 23.4 million patients using OpenSAFELY. *medRxiv*.
- [15] Wedlund, L., & Kvedar, J. (2021). New machine learning model predicts who may benefit most from COVID-19 vaccination.
- [16] Kannan, S., Subbaram, K., Ali, S., & Kannan, H. (2020). The role of artificial intelligence and machine learning techniques: Race for covid-19 vaccine. *Archives of Clinical Infectious Diseases*, 15(2).
- [17] D. Q. Zeebaree, A. M. Abdulazeez, D. A. Zebari, H. Haron and H. Nuzly, "Multi-level fusion in ultrasound for cancer detection based on uniform lbp features," *Computers, Materials & Continua*, vol. 66, no.3, pp. 3363-3382, 2021.
- [18] Othman, G., & Zeebaree, D. Q. (2020). The Applications of Discrete Wavelet Transform in Image Processing: A Review. *Journal of Soft Computing and Data Mining*, 1(2), 31-43.
- [19] Zawbaa, H., El-Gendy, A., Saeed, H., Osama, H., Ali, A. M., Gomaa, D., ... & Abdelrahim, M. E. (2021). A study of the possible factors affecting COVID-19 spread, severity and mortality and the effect of social distancing on these factors: Machine learning forecasting model. *International Journal of Clinical Practice*, e14116.
- [20] Müller, M., & Salathé, M. (2020). Addressing machine learning concept drift reveals declining vaccine sentiment during the COVID-19 pandemic. *arXiv preprint arXiv:2012.02197*.
- [21] Charbuty, B., & Abdulazeez, A. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*, 2(01), 20-28.
- [22] Adeen, N., Abdulazeez, M., & Zeebaree, D. Systematic Review of Unsupervised Genomic Clustering Algorithms Techniques for High Dimensional Datasets.
- [23] Mahmood, M. R., & Abdulazeez, A. M. (2019, April). Different model for hand gesture recognition with a novel line feature extraction. In 2019 International Conference on Advanced Science and Engineering (ICOASE) (pp. 52-57). IEEE.
- [24] Armya, R. E., & Abdulazeez, A. M. (2021). Medical Images Segmentation Based on Unsupervised Algorithms: A Review. *Qubahan Academic Journal*, 1(2), 71-80.
- [25] Jahwar, A. F., & Abdulazeez, A. M. (2020). Meta-Heuristic Algorithms For K-Means Clustering: A Review. *PalArch's Journal of Archaeology of Egypt/Egyptology*, 17(7), 12002-12020.
- [26] Carrieri, V., Lagravinese, R., & Resce, G. (2021). Predicting vaccine hesitancy from area-level indicators: A machine learning approach. *medRxiv*.
- [27] Siddiqui, M. K., Morales-Menendez, R., Gupta, P. K., Iqbal, H. M., Hussain, F., Khatoun, K., & Ahmad, S. (2020). Correlation between temperature and COVID-19 (suspected, confirmed and death) cases based on machine learning analysis. *J Pure Appl Microbiol*, 14(suppl 1), 1017-1024.
- [28] Ong, E., Wong, M. U., Huffman, A., & He, Y. (2020). COVID-19 coronavirus vaccine design using reverse vaccinology and machine learning. *Frontiers in immunology*, 11, 1581.
- [29] Carrillo-Larco, R. M., & Castillo-Cara, M. (2020). Using country-level variables to classify countries according to the number of confirmed COVID-19 cases: An unsupervised machine learning approach. *Wellcome Open Research*, 5.
- [30] Pinter, G., Felde, I., Mosavi, A., Ghamisi, P., & Gloaguen, R. (2020). COVID-19 pandemic prediction for Hungary; a hybrid machine learning approach. *Mathematics*, 8(6), 890.
- [31] Gupta, M., Bansal, A., Jain, B., Rochelle, J., Oak, A., & Jalali, M. S. (2021). Whether the weather will help us weather the COVID-19 pandemic: Using machine learning to measure twitter users' perceptions. *International journal of medical informatics*, 145, 104340.
- [32] Nguyen, T. T., Abdelrazek, M., Nguyen, D. T., Aryal, S., Nguyen, D. T., & Khatami, A. (2020). Origin of novel coronavirus (COVID-19): a computational biology study using artificial intelligence. *BioRxiv*.
- [33] Rodriguez, M. Z., Comin, C. H., Casanova, D., Bruno, O. M., Amancio, D. R., Costa, L. D. F., & Rodrigues, F. A. (2019). Clustering algorithms: A comparative approach. *PLoS one*, 14(1), e0210236.
- [34] Jahwar, A., & Ahmed, N. (2021). Swarm Intelligence Algorithms in Gene Selection Profile Based on Classification of Microarray Data: A Review. *Journal of Applied Science and Technology Trends*, 2(01), 01-09.
- [35] Muhammad, M., Zeebaree, D., Abdulazeez, A. M., Saeed, J., & Zebari, D. A. (2020). A Review on Region of Interest Segmentation Based on Clustering Techniques for Breast Cancer Ultrasound Images. *J. Appl. Sci. Technol. Trends*, 1(3), 78-91.
- [36] Fränti, P., & Sieranoja, S. (2018). K-means properties on six clustering benchmark datasets. *Applied Intelligence*, 48(12), 4743-4759.

- [37] Zeebaree, D. Q., Haron, H., Abdulazeez, A. M., & Zeebaree, S. R. (2017). Combination of K-means clustering with Genetic Algorithm: A review. *International Journal of Applied Engineering Research*, 12(24), 14238-14245.
- [38] Sadeeq, H., Abdulazeez, A., Kako, N., & Abraham, A. (2017, April). A Novel Hybrid Bird Mating Optimizer with Differential Evolution for Engineering Design Optimization Problems. In *International Conference of Reliable Information and Communication Technology* (pp. 522-534). Springer, Cham.
- [39] Cohen-Addad, V., Kanade, V., Mallmann-Trenn, F., & Mathieu, C. (2019). Hierarchical clustering: Objective functions and algorithms. *Journal of the ACM (JACM)*, 66(4), 1-42.
- [40] Haji, S. H., & Abdulazeez, A. M. (2021). COMPARISON OF OPTIMIZATION TECHNIQUES BASED ON GRADIENT DESCENT ALGORITHM: A REVIEW. *PalArch's Journal of Archaeology of Egypt/Egyptology*, 18(4), 2715-2743.
- [41] Naeem, A., Rehman, M., Anjum, M., & Asif, M. (2019). Development of an efficient hierarchical clustering analysis using an agglomerative clustering algorithm. *Current Science*, 117(6), 1045.
- [42] Omar, N., Abdulazeez, A. M., Sengur, A., & Al-Ali, S. G. S. (2020). Fused faster RCNNs for efficient detection of the license plates. *Indonesian Journal of Electrical Engineering and Computer Science*, 19(2), 974-982.
- [43] Zebari, Dilovan Asaad, Diyar Qader Zeebaree, Adnan Mohsin Abdulazeez, Habibollah Haron, and Haza Nuzly Abdull Hamed. "Improved Threshold Based and Trainable Fully Automated Segmentation for Breast Cancer Boundary and Pectoral Muscle in Mammogram Images." *IEEE Access* 8 (2020): 203097-203116.
- [44] Ishizaka, A., Lokman, B., & Tasiou, M. (2020). A Stochastic Multi-criteria divisive hierarchical clustering algorithm. *Omega*, 102370.
- [45] Adnan Mohsin Abdulazeez, Sulaiman, M. A., Zeebaree, Diyar Qader. (2020). Evaluating Data Mining Classification Methods Performance in Internet of Things Applications. *Journal of Soft Computing and Data Mining*, 1(2), 11-25
- [46] Sammaknejad, N., Zhao, Y., & Huang, B. (2019). A review of the expectation maximization algorithm in data-driven process identification. *Journal of process control*, 73, 123-136.
- [47] Mouchet, J., & Bégaud, B. (2018). Central demyelinating diseases after vaccination against hepatitis B virus: a disproportionality analysis within the VAERS database. *Drug safety*, 41(8), 767-774.
- [48] Babichev, S., Durnyak, B., Pikh, I., & Senkivskyy, V. (2019, May). An evaluation of the objective clustering inductive technology effectiveness implemented using density-based and agglomerative hierarchical clustering algorithms. In *International Scientific Conference "Intellectual Systems of Decision Making and Problem of Computational Intelligence"* (pp. 532-553). Springer, Cham.
- [49] Attwal, K. P. S., & Dhiman, A. S. (2020). Exploring data mining tool-Weka and using Weka to build and evaluate predictive models. *Advances and Applications in Mathematical Sciences*, 19(6), 451-469.
- [50] Mohammed, D. Y., & Karabatak, M. (2018, March). Terrorist attacks in Turkey: An evaluate of terrorist acts that occurred in 2016. In *2018 6th International Symposium on Digital Forensic and Security (ISDFS)* (pp. 1-3). IEEE.
- [51] Aksu, G., & Dogan, N. (2019). An analysis program used in data mining: WEKA. *J Meas Eval Educ Psychol*, 10, 80-95.
- [52] Kodati, S., Vivekanandam, R., & Ravi, G. (2019). Comparative analysis of clustering algorithms with heart disease datasets using data mining Weka tool. In *Soft computing and signal processing* (pp. 111-117). Springer, Singapore.
- [53] Abdulazeez, A. M., Hajy, D. M., Zeebaree, D. Q., & Zebari, D. A. (2021). Robust watermarking scheme based LWT and SVD using artificial bee colony optimization. *Indonesian Journal of Electrical Engineering and Computer Science*, 21(2), 1218-1229
- [54] Zebari, R., Abdulazeez, A., Zeebaree, D., Zebari, D., & Saeed, J. (2020). A Comprehensive Review of Dimensionality Reduction Techniques for Feature Selection and Feature Extraction. *Journal of Applied Science and Technology Trends*, 1(2), 56-70.
- [55] Zeebaree, D. Q., Haron, H., Abdulazeez, A. M., & Zebari, D. A. (2019, April). Machine learning and Region Growing for Breast Cancer Segmentation. In *2019 International Conference on Advanced Science and Engineering (ICOASE)* (pp. 88-93). IEEE.