Journal of Advanced Computer Science & Technology, 11 (1) (2022) 1-6

Journal of Advanced Computer Science & Technology

Website: www.sciencepubco.com/index.php/JACST

Research paper

Software engineering's role in visualizing large data as medical datasets

Hind Abdel Amir Sebti [1] *, Safa A. Najim [2]

[1] Sciences College Basrah University
[2] College of CS and IT Basrah University
**Corresponding author E-mail: hindalzbidy2013@gmail.com*

Abstract

Although both 'information visualization' and 'software engineering visualization' are data collection operations, their motivations are clearly distinct. The visualization of software engineering integrates the original design modules and identifies the connections between the main components. The purpose of software engineering is to develop high-quality programs and create appropriate work. The goal of visualization is to get insight into many components of a process we are interested in, such as a scientific simulation or a real-world process, using interactive graphs. Medical imaging techniques are used to detect and comprehend illness progression in the human body. To be able to make a diagnosis concerning it. In health centers and hospitals, managing vast collections of medical pictures is a challenge. Because the image database has grown in size over the last year, meaningful information is needed to assist the specialist in assessing the patient's condition. In this review, we will highlight some of the techniques that we use in data analysis.

Keywords: *Data; Data Visualization; Software Engineering; Medical Imaging.*

1. Introduction

Reliable dimensional reduction algorithms are very well promising nonlinear methods for visualizing various data sets, such as medical pictures [1]. Dimensionality reduction (DR) is a strategy for visualizing high-dimensional data sets by projecting them into a lesser environment where they may be directly viewed [2]. Dimensionality reduction (DR) is a crucial stage in data pre-processing for visualization and knowledge discovery, and it's utilized for a variety of applications including information visualization, noise reduction, and imaging as seen in Fig 1 [2]. Viruses that infect the respiratory system are an example of medical data. The virus is an object has sometime negative impact on the human live. The degree of dangerous depends on the type of virus, where the most dangerous one can kill the human [3]. Each virus has structure differs from other, but others have similarity. Respiratory viruses, especially in newborns and infants, are a prevalent cause of respiratory tract infection (RTI) [4]. Rapid and precise viral infection identification could enhance clinical outcomes while reducing antibiotic use and treatment sessions. The development of diagnostic technologies aids in the accurate diagnosis of viruses. Data visualization is a way of computing that uses graphics and imaging technologies to make complex data visible for the purpose of presenting hidden information[5]. Raw data are files that contain a great deal of information in a variety of fields. Hyperspectral data, pictures, medical, and simulated physical processes are just a few examples. The discovery and analysis of data rely heavily on visualization. Because the relationships between the original data are uncertain, visualization is a straightforward technique to grasp the high dimension. Viruses are widespread creatures that multiply by infecting and infecting other organisms. They can infect a wide spectrum of living organisms, from bacteria to plants and animals, and can be found in all ecosystems. The structure of viruses has two components at the very least [3]. A nucleic genome, made up of double- or single-stranded DNA or RNA, and a capsid are the two components. The capsid is a highly symmetrical structure made up of many copies of a tiny range of proteins that the viral genome encodes [3]. Respiratory- viruses are among the leading causes of disease and mortality in the world today. In the last century, the influenza virus has killed about 100 million people. Children and the elderly bear the brunt of this responsibility. While respiratory viruses do impose a greater burden on people in developing countries, they also cause various issues in wealthier ones [6].
In this paper we will focus on viruses that infect the respiratory tract and the role of software engineering and data visualization in analysis the data.

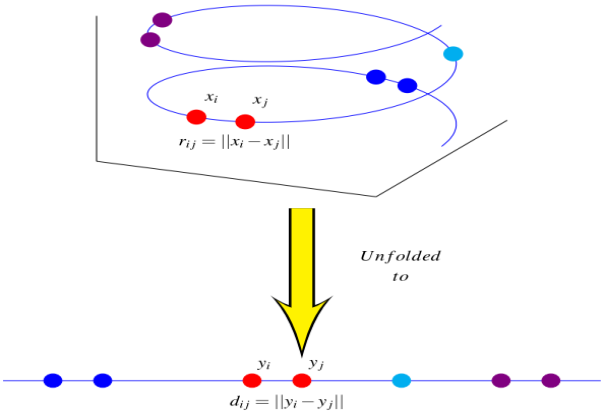

Fig. 1: Spiral Sets of Data in 3 Dimensions Are Collapsed Into A One-Dimensional Straight String Line. in High-Dimensional Data Sets, the Neighborhood Relations Between Points in the Unfolded Space Are Retained Together with Their Corresponding Relations[2].

2. Respiratory tracts

Our bodies are often functioning to combat infection. Viruses that cause human diseases are indeed one this kind of infection that we can receive. Viruses that cause human diseases can influence various sections of the body in different ways. The respiratory tract is one body part that may be infected with a virus. The respiratory tract is the set of tissues and cells that helps us to take in oxygen-rich air and exhale carbon dioxide Figure 1 shows the parts of the respiratory system[7].

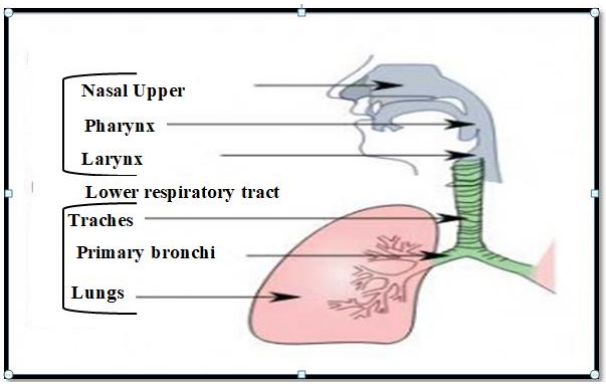

Fig. 2: Upper and Lower Respiratory Tract [8].

3. Types of respiratory tracts viruses

Viruses can infect the respiratory tract in several ways. Any of these viruses that cause human diseases can spread quickly between humans through airborne transmission via droplets and aerosols, resulting in hard outbreaks[9].The viruses are:

3.1. Influenza

Influenza is a virus that causes illness. IAV is a highly infectious virus that can be spread to humans by direct interaction with secretions, aerosolized respiratory droplets, or close contact with pathogenic organisms .Swine influenza, avian influenza (bird flu), influenza B virus and influenza C virus are the most common types of flu[10].

3.2 Sever acute respiratory syndrome (SARS)

As per the WHO, SARS is induced by a coronavirus (CoV) and was first described in 2003. World health organization Coronavirus (CoV) is a bat-borne virus that has targeted humans in China's Guangdong province[11]. The symptoms of SARS are fever, chills, muscle aches, headache, loss of appetite, diarrhea, dry cough, shortness of breath and the hosts are human[12].

3.3. Ebola virus diseases (EVD)

Ebola virus disease (EVD) is a highly contagious and sometimes fatal disease triggered by the Ebola virus (EBOV).
EVD outbreaks usually begin with an initial case of potential zoonotic infection, which is then accompanied by person-to-person transmission through direct interaction, contact with infectious bodily fluids, or contact with polluted fomites. Fever, stomach signs, and multiple organ failure syndromes are all symptoms of EVD[13].

3.4. Middle east respiratory syndrome coronavirus (MERS-COV)

In 2012, the coronavirus (MERS-CoV) was discovered for the first time in Saudi. To today, the (WHO) has collected data of over 2,400 incidents worldwide, including over 850 fatalities. The symptoms for this virus are fever, cough, shortness of breath, vomiting[14].

3.5. Hepatitis a virus

The (HAV) is a type of most frequent causes of acute hepatitis around the world. HAV is spread mainly by feces, tainted food or drink, or direct interaction with an infected human; the symptoms are malaise, loss of appetites, diarrhea and jaundice[15].

3.6. Hepatitis b virus

The (HBV) may be spread by blood or sexual intercourse. Chronic HBV infection puts people at risk for cirrhosis which liver cancer and necessitates medical attention. HBV is extremely contagious, can be spread without detectable blood, and can survive for at least 7 days on environmental surfaces the symptoms are jaundice, fever, vomiting and loss of appetite[16].

3.7. Hepatitis C Virus

HCV is a blood-borne pathogen that is a major public health issue worldwide. Acute HCV infection may lead to chronic infection after virus acquisition, which is linked to a number of morbidities, including liver cirrhosis and cancer[17].

3.8. Measles virus

Measles is a particularly infectious respiratory infection that starts with a fever, cough, coryza, and conjunctivitis, then progresses to a maculopapular rash. Measles is mainly transmitted by airborne delivery, with a 90% secondary attack incidence within nearby susceptible connections[18].

3.9. Mumps virus

The swell of one or so more salivary glands, usually the parotid glands, is a symptom of mumps, a viral infectious disorder. Mumps, despite its reputation as a harmless childhood illness, was the primary cause of viral meningitis and encephalitis in babies, as well as lifelong deafness[19].

3.10. Human Immunodeficiency virus (AIDS)

In the summer of 1981, the outbreak was officially declared to have begun the symptoms of this virus are fever, achy, sick, thrush, sore throat, diarrhea and bleeding from the mouth[20]. HIV is a disease that targets cells in the body that support it resist pathogens, rendering an individual more susceptible to other infections and diseases[21].

3.11. Sever acute respiratory syndrome coronavirus-2(SARS-Cov-2)

In Dec. 2019, a cluster of extreme pneumonia with an unexplained cause occurred in Wuhan, Hubei Province, China. As the causative agent, a new coronavirus known as extreme acute respiratory syndrome coronavirus-2 was extracted from a bottom respiratory tract test. The new SARS-CoV-2 epidemic has been dubbed Coronavirus Virus 2019 (COVID-19) by the (WHO). The hosts of this virus are human, bats and pangolin and the treatment are Hydroxychloroquine and antibiotic azithromycin combination[22].

3.12. Coronavirus

Coronavirus illness 2019 (COVID-19) is a long-term pandemic triggered by the coronavirus that causes extreme (SARS-CoV-2). The outbreak started at the end 2019 in Wuhan, China's capital and capital of Hubei province, and was proclaimed a public health emergency of international significance on January 30, 2020 [23]. Over 48 136 225 COVID-19 incidents have been confirmed worldwide, with over 1 225 913 deaths; over 31 919 360 patients have healed. This epidemic has wreaked havoc on the global economy, forcing trips and athletic activities, as well as national, social, and classical events, to be rescheduled or cancelled. COVID-19 made the world economy to a screeching halt, triggering supply and demand shocks. SARS-CoV-2 causes CORONAVIRUS, a strongly transmissible and pathogenic viral disease[24]. SARS-CoV-2 is taxonomically linked to extreme acute respiratory syndrome (SARS)-like bat viruses that cause human diseases, with approximately 96 percent genome sequence similarities, according to genomic research. As a result, bats could be the main source for human coronavirus (Covid-19)[25].

4. Data visualization

The procedure for converting similarity associations between the points of data into a geometric illustration so that they can be view previously unnoticed information is known as data visualization. The following questions are examined as a result of the findings: Does this visualization aid in the discovery of new information while analyzing vast amounts of data?[2] What's the connection between the original and converted data points? How can the original data be retained in a predicted environment?
One of the most difficult aspects of data visualization is dealing with high dimensionality. Dimension Reduction (DR) is a helpful approach for projecting large dimensional datasets into a low dimensional space that can be directly seen[26]. The use of this approach has two advantages:

4.1. By minimizing the size of the data sets

DR may reduce the amount of storage required

4.2. DR aids in the comprehension of data sets by removing any extraneous characteristics and focusing on the most significant ones[2]

There are various characteristics of large dimensional datasets; however, some of them may be irrelevant to the study of specific information. DR is employed to identify the most essential aspects that allow for analysis and visualization[27]. Data visualization is widely utilized in a variety of tasks, including reporting, managing corporate functions, and monitoring task completion at work. More often, we may use visualization methods to identify things that would be difficult to notice if we only looked at the raw data. Nowadays, data visualization is often used to produce engaging and intriguing visualizations[27].
The data visualization pipeline is importing, filtering, and enriching data, as well as data mapping and rendering, are the four phases of the data visualization pipeline. Fig 3 shows a schematic representation of them[28].

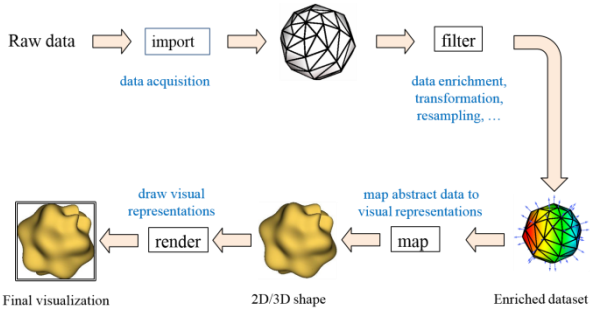

Fig. 3: Visualization Pipeline[28].

5. The benefit of data visualization

There are many benefits of data visualization, but we can summarize the most important of them:

5.1. Process enormous volumes of data quickly

5.2. Obtain vital information in real time

5.3. Interactivity helps you better understand facts

5.4. Examine the connections between business procedures and results

5.6. Use a story with data[2]

Dimension Reduction (DR) is a helpful approach for projecting large dimensional datasets into a low dimensional space that can be directly seen. The use of this approach has two advantages: by minimizing the size of the data sets, DR may reduce the amount of storage required and DR aids in the comprehension of data sets by removing any extraneous characteristics and focusing on the most significant ones[26].
To explain the general idea of DR, let suppose: For a set of n input points $X \subset R^D$, $\phi(X)$ is used to project the D dimensional data points $x_i \in X$ to d dimensional data points yi ∈ Y, where d << D.

$$\phi: R^D \rightarrow R^d \quad (1)$$

$$Xi \rightarrow yi \quad (2)$$

The DR attempts to approximate the pairwise distance, d (xi,xj) with their equivalent in original space r(yi, yj), i.e. r(xi, xj) ≈ d(yi, yj), 1≤ i ≤ n to project X's data point accurately in Y space in Equation 2[29]. The goal of applying dimension reduction (DR) methods is to project large dimensional datasets to a low dimensional space while preserving all of the original datasets information[30].

6. Data visualization applications

6.1. Data visualization was used to analyze the data and extract what is important from the analysis process[31]

6.2. Because color blindness is a disorder centered on colors

data visualization can help with this disease by giving particular techniques for color correction, as new ways based on data visualization concepts have been developed to help identify colors more easily [32].

6.3. We can utilize a way to view medical picture data sets by employing reliable dimensional reduction techniques

Dimensional reduction methods are well-known prospective nonlinear methods for visualizing various data sets, such as medical pictures[1].

7. Software engineering

Software engineering is a crucial technique for developing and testing software. Visualizing big data sets, such as hyperspectral and medical images, is critical for seeing previously hidden information. The visualization of these vast volumes of data should produce a result

that is acceptable to the human sight. It's critical to keep the information from the original data sets in order to discover effective visualization. This study shows how the final visualization of data sets is a simplified representation of the complicated structure of the original data sets [33 - 35].

8. Conclusion

Software engineering is a domain that is concerned with the development of high-quality software that is responsive to market demands and delivered on schedule. To deal with big data, software engineers and developers use software engineering in their applications. The visualization of these big data sets should produce a result that is acceptable to the human sight. It's critical to save the original data sets' information in order to find effective visualizations. Role of data visualization is very important in order to see unseen information. Because of large size of datasets, the requirement to preserve them in structure more accurate is necessary. The DR is one of the most techniques of data visualization, when it is ability to reduce the dimensionality of dataset is clear. Where the relation between viruses is discovered. In the final, we can say data visualization is useful to analyze data and discover invisible information, through find the relation between viruses. The role of data visualization is very important to reach to study the structure of each virus.

References

[1] Safa A. Najim and Mansor W. A., Hybrid Visualization of the Medical Images Data Sets. International Journal of Computer Application 136 (8), 2016. https://doi.org/10.5120/ijca2016908617.

[2] Safa A.Najim , Najim A.A., Lim I.S. and Saeed M., Parallel Faithful Dimensionality Reduction To Enhance The Visualization Of Remote Sensing Imagery. Neurocomputing . 168, pp:560-565, 2015. https://doi.org/10.1016/j.neucom.2015.05.071.

[3] A. Cassedy, A. Parle-McDermott, and R. O'Kennedy, "Virus Detection: A Review of the Current and Emerging Molecular and Immunological Methods," *Front. Mol. Biosci.*, vol. 8, pp. 1–36, 2021, https://doi.org/10.3389/fmolb.2021.637559.

[4] L. Herrero-Uribe, "Viruses, definitions and reality," *Rev. Biol. Trop.*, vol. 59, no. 3, pp. 993–998, 2011, https://doi.org/10.15517/rbt.v0i0.3372.

[5] L. K. Borkenhagen, M. W. Allen, and J. A. Runstadler, "Influenza virus genotype to phenotype predictions through machine learning: a systematic review: Computational Prediction of Influenza Phenotype," *Emerg. Microbes Infect.*, vol. 10, no. 1, pp. 1896–1907, 2021, https://doi.org/10.15517/rbt.v0i0.3372.

[6] H. BJURSTEDT, "Respiratory system.," *Annu. Rev. Physiol.*, vol. 19, pp. 151–174, 1957, https://doi.org/10.1146/annurev.ph.19.030157.001055.

[7] T. Apprentice *et al.*, "Upper & Lower Respiratory System Advice to Anesthetists : Lose anything you like – but NEVER-EVER lose your patient ' s airway !," pp. 1–4, 2021.

[8] J. S. Kutter, M. I. Spronken, P. L. Fraaij, R. A. Fouchier, and S. Herfst, *Transmission routes of respiratory viruses among humans*, vol. 28, no. February 2018. 2018. https://doi.org/10.1016/j.coviro.2018.01.001.

[9] E. S. Bailey *et al.*, "The continual threat of influenza virus infections at the human-animal interface: What is new from a one health perspective?," *Evol. Med. Public Heal.*, vol. 2018, no. 1, pp. 192–198, 2018, https://doi.org/10.1093/emph/eoy013.

[10] S. Law, A. W. Leung, and C. Xu, "Severe acute respiratory syndrome (SARS) and coronavirus disease-2019 (COVID-19): From causes to preventions in Hong Kong," *Int. J. Infect. Dis.*, vol. 94, no. January 2020, pp. 156–163, 2020, https://doi.org/10.1016/j.ijid.2020.03.059.

[11] F. Fleck, "How SARS changed the world in less than six months.," *Bull. World Health Organ.*, vol. 81, no. 8, pp. 625–626, 2003.

[12] D. Malvy, A. K. McElroy, H. de Clerck, S. Günther, and J. van Griensven, "Ebola virus disease," *Lancet*, vol. 393, no. 10174, pp. 936–948, 2019, https://doi.org/10.1016/S0140-6736(18)33132-5.

[13] Z. A. Memish, S. Perlman, M. D. Van Kerkhove, and A. Zumla, "Middle East respiratory syndrome," *Lancet*, vol. 395, no. 10229, pp. 1063–1077, 2020, https://doi.org/10.1016/S0140-6736(19)33221-0.

[14] K. Y. Lin *et al.*, "Hepatitis A virus infection and hepatitis A vaccination in human immunodeficiency virus-positive patients: A review," *World J. Gastroenterol.*, vol. 23, no. 20, pp. 3589–3606, 2017, https://doi.org/10.3748/wjg.v23.i20.3589.

[15] S. Schillie *et al.*, "Prevention of hepatitis B virus infection in the United States: Recommendations of the advisory committee on immunization practices," *MMWR Recomm. Reports*, vol. 67, no. 1, pp. 1–31, 2018, https://doi.org/10.15585/mmwr.rr6701a1.

[16] Z. Al Kanaani, S. Mahmud, S. P. Kouyoumjian, and L. J. Abu-Raddad, "The epidemiology of hepatitis C virus in Pakistan: Systematic review and meta-analyses," *R. Soc. Open Sci.*, vol. 5, no. 4, pp. 1–16, 2018, https://doi.org/10.1098/rsos.180257.

[17] B. M. Laksono, R. D. de Vries, W. P. Duprex, and R. L. de Swart, "Measles pathogenesis, immune suppression and animal models," *Curr. Opin. Virol.*, vol. 41, pp. 31–37, 2020, https://doi.org/10.1016/j.coviro.2020.03.002.

[18] P. Kaaijk *et al.*, "A Third Dose of Measles-Mumps-Rubella Vaccine to Improve Immunity against Mumps in Young Adults," *J. Infect. Dis.*, vol. 221, no. 6, pp. 902–909, 2020, https://doi.org/10.1093/infdis/jiz188.

[19] G. A. Cornia, "Overview of the Impact and Best Practice Responses in Favour of Children in a World Affected by HIV / AIDS," *AIDS, Public Policy and Child Well-being*, no. June, pp. 1–33, 2002.

[20] HIV.gov, "What Are HIV and AIDS ? How Do I Know If I," p. 19, 2020, [Online]. Available: https://www.hiv.gov/hiv-basics/overview/about-hiv-and-aids/what-are-hiv-and-aids.

[21] S. E. Park, "Epidemiology, virology, and clinical features of severe acute respiratory syndrome-coronavirus-2 (SARS-COV-2; Coronavirus Disease-19)," *Korean J. Pediatr.*, vol. 63, no. 4, pp. 119–124, 2020, https://doi.org/10.3345/cep.2020.00493.

[22] D. Kumar, "Corona Virus: A Review of COVID-19," *Eurasian J. Med. Oncol.*, vol. 4, no. 2, pp. 8–25, 2020, doi: 10.14744/ejmo.2020.51418.

[23] M. Naveed *et al.*, "Review of potential risk groups for coronavirus disease 2019 (COVID-19)," *New Microbes New Infect.*, vol. 41, no. 222, p. 100849, 2021, https://doi.org/10.14744/ejmo.2020.51418.

[24] D. Gathering, "COVID-19 Data Analytics," pp. 1–11, 2021.

[25] Safa A. Najim, *Information Visualization by Dimensionality Reduction: A Review. Journal of Advanced Computer Science and Technology. 3(2), pp:101-112, 2014.* https://doi.org/10.14419/jacst.v3i2.2746.

[26] J. P. Cunningham and Z. Ghahramani, "Linear dimensionality reduction: Survey, insights, and generalizations," *J. Mach. Learn. Res.*, vol. 16, pp. 2859–2900, 2015.

[27] A. C. Telea, *Data Visualization: Principles and Practice, Second Edition*. 2014. https://doi.org/10.1201/b17217.

[28] Safa A. Najim and Lim I.S., Trustworthy Dimension Reduction for Visualization Different Data Sets, Information Science. 278, pp:206-220, 2014 https://doi.org/10.1016/j.ins.2014.03.048.

[29] R. Articles, "Introduction to Dimensionality Reduction Last Updated :," pp. 1–6, 2021. https://doi.org/10.1201/9781003190554-1.

[30] Z. H. Majeed, B. Y. Ahmed, and S. A. Najim, "Specifying the visualizing type of multi-dimensional data," *J. Phys. Conf. Ser.*, vol. 1294, no. 3, 2019, https://doi.org/10.1088/1742-6596/1294/3/032033.

[31] B. Y. Ahmed, Z. H. Majeed, and S. A. Najim, "Applying the visualization technique to solve the human color blindness," *AIP Conf. Proc.*, vol. 2290, no. December, 2020, https://doi.org/10.1063/5.0027395.

[32] B. Y., S. A., and W. A., "Role of Software Engineering in Visualizing Large Volume of Hyperspectral and Medical Data Sets," *Int. J. Comput. Appl.*, vol. 176, no. 9, pp. 1–5, 2017, https://doi.org/10.5120/ijca2017915580.

[33] Safa A. Najim and Lim I.S., Wittek P. and Jones M. FSPE: Visualisation of Hyperspectral Imagery Using Faithful Stochastic Proximity Embedding. IEEE Geoscience and Remote Science Letters. 12(1), pp:18-22, 2015 https://doi.org/10.1109/LGRS.2014.2324631.

[34] Safa A. Najim and Lim I.S. (2014). Visualization of Remote Sensing Imagery by Sequential Dimensionality Reduction on Graphics Processing Unit. In Proceedings of the 5th International Conference on Information Visualization Theory and Applications (IVAPP 2014), , pages 71-79, Lisbon, Portugal, 5-8 January.