BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

Journal of Computer Science Research

Volume 4 | Issue 4 | October 2022 | ISSN 2630-5151（Online）

BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

Volume 4 Issue 4 • October 2022 • ISSN 2630-5151 (Online)

Journal of Computer Science Research

BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

Volume 4 | Issue 4 | October 2022 | Page1-35
Journal of Computer Science Research

Contents

Articles

BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

Journal of Computer Science Research
https://ojs.bilpublishing.com/index.php/jcsr

ARTICLE

Study, Analysis and Comparison between Amazon A10 and A11 Search Algorithm

Subhradeep Maitra[1] Laxminarayan Sahoo[1*] Kalishankar Tiwary[2]

1. Department of Computer and Information Science, Raiganj University, Raiganj, 733134, India
2. Department of Mathematics, Raiganj University, Raiganj, 733134, India

ARTICLE INFO

Article history
Received: 26 September 2022
Revised: 20 October 2022
Accepted: 26 October 2022
Published Online: 3 November 2022

Keywords:
Amazon business
Algorithm
Page rank
Visibility
Search engine algorithm
E-commerce

ABSTRACT

The entirety of Amazon's sales being powered by Amazon Search, one of the leading e-commerce platforms around the globe. As a result, even slight boosts in appropriateness can have a major impact on profits as well as the shopping experience of millions of users. Throughout the beginning, Amazon's product search engine was made up of a number of manually adjusted ranking processes that made use of a limited number of input features. Since that time, a significant amount has transpired. Many people overlook the fact that Amazon is a search engine, and even the biggest one for e-commerce. It is indeed time to begin treating Amazon truly as the top e-commerce search engine across the globe because it currently serves 54% of all product queries. In this paper, the authors have considered two most important Amazon search engine algorithms viz. A10 and A11 and comparative study has been discussed.

1. Introduction

Amazon is basically a world-based e-commerce company which has over 2 billion of users/clients directly and indirectly involves into this e-commerce platform [1-3]. Customers have convenient access to information about a product through the product detail page, which also makes it easier for customers to compare different offers that are currently available. On the product detail page, the Featured Offer will present the offer that it has determined the customer is most likely to select after conducting an in-depth comparison of all of the available offers. Amazon determines whether an offer qualifies as a Featured Offer by taking into account a number of factors, including the price, the delivery speed, whether the offer is eligible for Prime, and the performance of the seller. As a direct consequence of this, there might be more than one Featured Offer available for a product, or there might be none at

*Corresponding Author:
Laxminarayan Sahoo,
Department of Computer and Information Science, Raiganj University, Raiganj, 733134, India;
Email: lxsahoo@gmail.com

DOI: https://doi.org/10.30564/jcsr.v4i4.5111

all. Customers can easily begin their shopping experience by browsing all of the available offers from the product detail page. The vast majority of customers who compare these offers end up selecting a Featured Offer as their preferred option. Please click here for further details regarding the Featured Offer. The requirements listed in this section, which are located there, must be adhered to in an order so that one can be eligible for the Featured Offer. We reserve the right to remove a seller's eligibility for the Featured Offer or to not display a Featured Offer at all in order to prevent abuse or for any other reason that helps us maintain a positive experience for our customers. The provision of assistance to customers in the course of their product searches is the primary purpose of the Amazon Algorithm. In order to provide users with the most relevant information possible, the system performs an analysis on a variety of businesses and the products that are offered on the website by those businesses. The A9 Algorithm and the A10 Algorithm [4] are very similar to one another and share a lot of similarities. The primary distinction between the two lies in the level of significance attached to the various criteria used for ranking [5]. It is imperative to keep the customer in mind at all times rather than focusing on the algorithm; doing so will allow you to gain an advantage over your competitors. Because the primary objective of Amazon does not change, regardless of whether the algorithm is "A9" or "A10": to satisfy the needs of the customer by providing an exceptional experience when making digital purchases. This objective is not dependent on which algorithm is used. It is difficult to be successful in the ever-increasing competition on Amazon because there are more than 2 million sellers expanding their businesses there. Amazon's goal is to increase the number of new sellers and encourage healthy competition at all times in order to foster innovation. Advertising is one of the key strategies that must be utilized in order to dominate Amazon; the other part of this strategy is to remain current with the Amazon A9 algorithm. However, despite the fact that some people have started referring to it as the A10 algorithm, it is extremely comparable to the A9 algorithm. However, the most important distinction lies in the importance placed on various parameters. Amazon began providing customer searches with a greater degree of relevance after the most recent update. Amazon discovered that customers tend to look further than what sellers are trying to promote in their listings. Therefore, Amazon is making adjustments to its A10 algorithm in order to make the search results more pertinent to the customer. Table 1 contains information about A9 in more detail. The A10 algorithm is the most powerful incarnation of the searching and ranking algorithm used by Amazon.com. A10, which taking over just for Amazon's "A9" algorithm in 2020, provides more emphasis to a lot of ranking factors, enhancing its ability to match customer queries with appropriate products. Here, a comparison study of the two most significant Amazon search engine [6-9] algorithms, A10 and A11, has been presented.

Table 1. Information about Amazon A9 Algorithm

	A9
Type	Search Engine
Industry	Communications
Founded	2004
Founder	Udi Manber
Headquarters	Palo Alto, California
Parent	Amazon
Website	https://www.amazon.com/

2. Definition and Technical Implementation of Amazon A9 and A10 Algorithms

This section has discussed in brief the definition and technical implementation of the Amazon A9 and A10 algorithms.

2.1 Define Amazon A9 Algorithm

The Amazon A9 algorithm is a ranking system [10] that determines where your products will be displayed in the various search results on Amazon. It makes it easier for customers to find the product that best meets their needs by using the search keywords. This helps Amazon sellers gain traffic for their products, which in turn leads to a higher ranking of the products, which in turn helps Amazon sellers make more money. The Google algorithm [11,12] does not take into account the number of sales conversions that the Amazon A9 algorithm does. Because of this, products that have a longer history of sales will rank higher and better than comparable products that have a shorter history of sales. This is due to the fact that Amazon's A9 algorithm will promote the products that have a higher probability of being sold in comparison to other products. When determining where a product should be positioned in the list of results returned by an Amazon search, the A9 algorithm takes into account a wide variety of factors, including relevancy, reviews, keywords, sales history, product promotions, search terms, availability of the product, price, and many others. When determining the ranking of a seller's products, it is at the very centre of what Amazon does.

2.2 Technical Implementation of Amazon A9 Algorithm

The Amazon A9 algorithm is what customers use to

determine where a particular brand stands in the list of search results for their query. It presents the purchasers with the results that are pertinent to the key terms that they have entered into the search box. In addition to this, it includes the results that are recommended as well as other products that may pique the buyers' interest enough to convince them to purchase them. It provides the results based on the customer's previous search history as well as their previous purchasing history. This allows for the products to rank higher in the search results when they are searched for, which in turn increases the likelihood that they will be purchased, which in turn increases sales of the given product and brand. Because the Amazon A9 algorithm is based on the relevancy of search terms, key search terms play an important role in the process of finding results that are relevant to the item that customers are interested in purchasing.

2.3 Define Amazon A10 Algorithm

The Amazon A10 algorithm is the most recent iteration of the company's search engine. Because of this, the algorithm's overall operation has been modified as a result. The A10 algorithm used by Amazon will guide customers straight to the results they are looking for, bypassing any recommended or other search products in the process. The products that are the most pertinent, popular, and highly ranked are highlighted for the purpose of increasing the likelihood of sales for the vendors, as well as the likelihood of buyers discovering the product that best suits their needs. The product that appeals to buyers more than the others does naturally better in rankings than the products that do not. Not only this, but the method of ranking is now determined not by advertisements but rather by the clicks that are generated organically by the content. Introduction one of Amazon's newest features is called Seller's authority. This feature uses Amazon's buy box or product listing to establish a seller's credibility with customers. The duration of the seller's Amazon activity, the number of feedback and ratings received, the number of products sold, and other factors all contribute to the seller's authority. Parallel to this, organic sales are those that come from products that customers not only view but also buy, as opposed to sales that come from paid advertisements. Parallel to this, organic sales are more valuable than paid advertisements. This is due to the fact that Amazon has been placing a greater emphasis on a secure organic SEO strategy for sales rather than a pay-per-click strategy. Aside from that, traffic that comes to Amazon from other websites is referred to as "offsite sourced traffic". For example, Instagram, Facebook, YouTube blogs, TikTok, etc., it is important to make sure that the traffic is relevant because doing so will only increase the views of the page, not the sales. It is important to make sure that the traffic is relevant because it will only increase the views of the page. The internal sales are made up of products that come highly recommended and are frequently purchased. Because these are not the products that are returned by a search, but rather products that are shown to buyers in a separate section in order to boost the product's position in search results, simply put, click-through rates measure the number of times prospective customers click on a company's products during their initial research phase. It is only taken into consideration once they have not only been clicked on but also purchased afterward. Because of this, the products require not only a descriptive name but also a primary image of the highest possible quality.

2.4 Technical Implementation of Amazon A10 Algorithm

The Amazon A10 algorithm is an improved and more up-to-date version of the Amazon A9 algorithm, as was previously mentioned. Because of this, there have been many modifications made to the way the algorithm as a whole works. However, there are still some processes that have not been altered in any way. There are a few that play larger roles and others that play smaller roles in these, such as the pay-per-click (PPC) advertising model used by Amazon and the profitability of the products. The Amazon A10 algorithm has caused an increase in product visibility [13], which enables customers to have a better understanding of the items they are considering purchasing. The Amazon A10 algorithm has been modified a great deal in the past, and additional modifications are planned for the future as well. Because of this, both buyers and sellers are now aware of any potentially fraudulent products that are on Amazon's marketplace [14-16], and they can take immediate action to remove these products from the website and search engine. It is essential for any kind of seller on the Amazon web page to have a working knowledge of the Amazon A10 algorithm if they want their products to have the best chance of ranking higher in the Amazon search results.

3. How does the Amazon Search Algorithm Work

Here, we have discussed how does the Amazon search algorithm work. Each and every product ranking is determined by the Amazon algorithm, which examines the product listings, compares them to online customers' search terms, and then displays the most relevant ones on the top page of the search results. In order to better under-

stand the ranking process, Amazon must take into consideration the points listed below.

1) Amazon emphasizes boosting Revenue Per Customer (RPC) at all times.

2) Every transaction a customer takes on Amazon is recorded.

3) The Amazon A9 Algorithm correlates the tracked data mentioned in point 2 to the objective indicated in point 1 by using a series of mathematical operations.

Relevancy, Conversion Rate, and Product Authority are the three main components of the Amazon A9 algorithm, which determines how well a product ranks on Amazon.com. Figure 1 provides how does Amazon A9 algorithm works.

Every Amazon product needs to be sufficiently relevant to be given a higher search ranking. It is the initial solution to the question of how to rank highly on Amazon.com. The components that need to be taken into account are (i) the title, (ii) the bullet points, (iii) the descriptions, (iv) the brand and manufacturing number, (v) the category and sub-category, (vi) the search term, and (vii) the product specifications. Different factors of A10 algorithm have been shown in Figure 2. Better conversion rates depend on different factors, and one can raise them by paying attention to the following factors which are (i) Sales (ii) Questions and Answers (iii) Product Image (iv) Pricing (v) Variant Products' Placement (vi) Bounce Rate and (vii) Customer Review. When a customer searches for a product on Amazon.com, the product details for the Computer Network Book are shown in Figure 3.

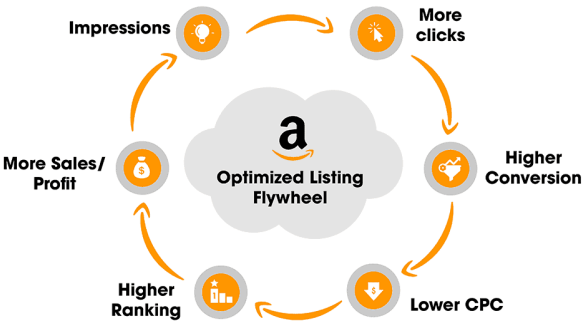


Figure 1. Work flow of Amazon A9 Algorithm

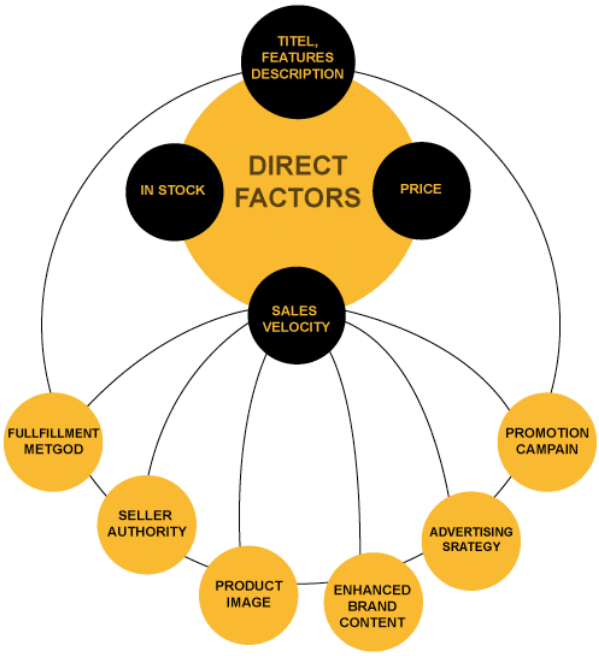


Figure 2. Factors of A10 Algorithm

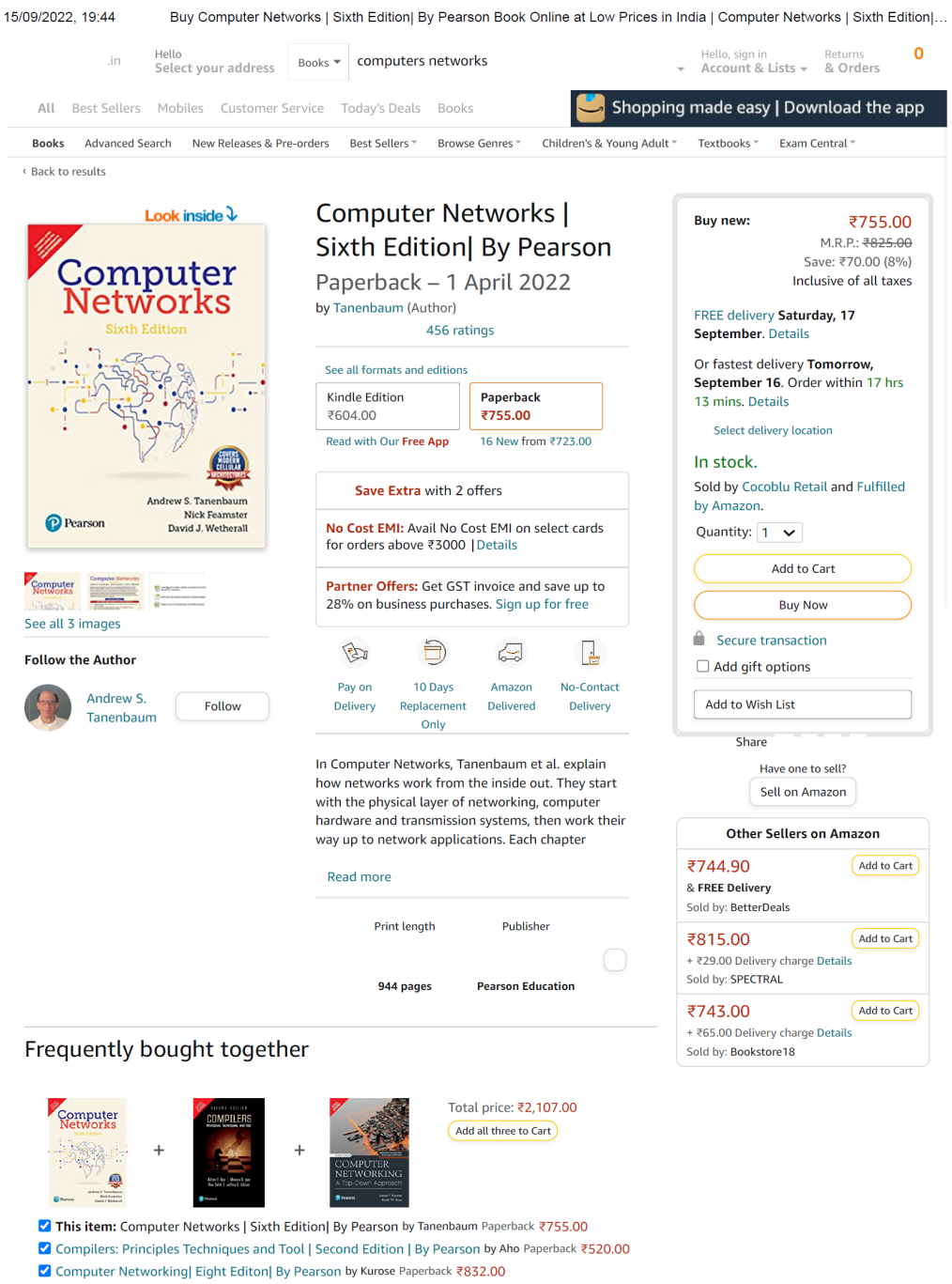


Figure 3. Details of Computer Networks Book by A. S. Tanenbaum, N. Feamster and D.J. Wetherall

4. Amazon A9 Algorithm vs. A10 Algorithm

Both the A9 algorithm and the A10 algorithm that Amazon uses are very similar to one another because they both assist Amazon sellers in increasing the amount of traffic that is directed to their websites and the number of sales that are made using their products. However, there are some distinctions between the two when it comes to the activities that the algorithms perform. Amazon A9 is an older Amazon algorithm that helps sellers on Amazon achieve a higher ranking with search keywords, pay-per-click (PPC), and Amazon's paid advertisements. Amazon A9 is an algorithm that was developed by Amazon. On the other hand, the Amazon A10 algorithm raises a seller's product to a higher position in the list of search results based on the number of times that result is clicked. It

ranks the products that are most viewed or clicked on the website, along with the products that are most bought and the items that are most likely to be purchased. In addition, it displays the products of your competitors directly below the product that your brand offers. This is the point in the programming where the Amazon A9 and A10 algorithms diverge from one another. The first is for pay-per-click advertising and paid advertisements, and the second is for unpaid, organic selling that is determined by the number of views and purchases. In any case, both of these results come from the same search engine that is used by Amazon. This search engine is designed to locate the appropriate product for customers and the appropriate rank of products for sellers on the Amazon web page.

5. Conclusions

The primary objective of the Amazon Algorithm is to provide support for customers during the course of their product searches. The system analyses numerous businesses and the products they offer on the website to provide users with the most pertinent information possible. There is a lot of overlap between the A9 Algorithm and the A10 Algorithm. The level of importance placed on various criteria for ranking is the primary difference between the two. Instead of focusing on the algorithm, it is imperative to keep the customer in mind at all times; doing so will allow you to gain a competitive advantage. Because the primary goal of Amazon remains the same, regardless of whether the algorithm is "A9" or "A10": to fulfil the requirements of the customer by delivering an outstanding experience when making digital purchases.

Author Contributions

S. Maitra, L. Sahoo and K.S. Tiwary formulated and studied the problem. S. Maitra, L. Sahoo and K.S. Tiwary wrote the first draft of the manuscript. All authors edited the manuscript and approved the final version.

Financial Support

This study was not supported by any financial grants.

Conflict of Interest

The authors declare no conflict of interest.

Ethical Approval

This article does not contain any studies involving animals performed by the authors.

References

[1] Amazon e-Market Place (www.amazon.com).

[2] Amazon Product availability vs other e-commerce website such as flipkart, snapdeal, myntra etc refers to the google search engine: www.google.com.

[3] Belsare, S., Patil, S., 2012. Study and Evaluation of user's behaviour in e-commerce Using Data. Research Journal of Recent Sciences. 1(ISC-2011), 375-387.

[4] Algorithm study of Amazon A9 and Amazon A10 refers to: https://www.pagemaker.io/post/how-to-crack-amazons-a10-algorithm-in-2022.

[5] Sorokina, D., Cantú-Paz, E., 2016. Amazon Search: The Joy of Ranking Products, SIGIR '16 July 17-21, 2016, Pisa, Italy c 2016 Copyright held by the owner/author(s).
DOI: http://dx.doi.org/10.1145/2911451.2926725

[6] Sullivan, D., 2002. How search engines work. SEARCH ENGINE Watch. http//www. search enginewatch. com/webmasters/work. html (last Updated. June 26, 2001) (on file with New York Univ. J. Legis. Public Policy).

[7] Cho, J., Roy, S., 2004. Impact of search engines on page popularity. Proceeding 13th International conference on World Wide Web. 20-29.

[8] Seymour, T., Frantsvog, D., Kumar, S., 2011. History of search engines. International Journal of Management & Information Systems. 15(4), 47-58.

[9] Maitra, S., Sahoo, L., Tiwary, K.S., 2022. Methods and Strategies for search engine Optimization. COJ Rob Artificial Intel. 2(2), 1-7.

[10] Robertson, S., 2019. A Brief History of Search Results Ranking. IEEE Annals of the History of Computing. 41, 22-28.

[11] Green, E.A., Markey, K.L., Kreider, M., 2016. Subject matter context search engine. Google Patents.

[12] Luh, C.J., Yang, S.A., Huang, T.L.D., 2016. Estimating Google's search engine ranking function from a search engine optimization perspective. Online Information Review. 40(2), 1-29.

[13] Shahzad, A., Nawi, N.M., Hamid, N.A., et al., 2017. The Impact of Search Engine Optimization on The Visibility of Research Paper and Citations. International Journal of Informatics Visualization. 1(4-2), 195-198.

[14] Raorane, A.A., Kulkarni, R.V., 2012. Association Rule – Extracting Knowledge Using Market Basket Analysis. Research Journal of Recent Sciences. 1(2), 19-27.

[15] Bhandari, R.S., Bansal, A., 2018. Impact of Search engine optimization as a marketing Tool. Jindal Journal of Business research. 7(1), 24-36.

[16] Poongkode, J.P.S., Nirosha, V., 2014. A Study on Various Search Engine Optimization Techniques. International Journal of Innovative Research in Computer and Communication Engineering. 2(11), 6738-6742.

BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

Journal of Computer Science Research
https://ojs.bilpublishing.com/index.php/jcsr

ARTICLE

Computational Experience Optimization of Colors in Complex Fractal Images in Carpet Design Using the Simplex Method

Fakhriddin Nuraliev[1] Oybek Narzulloyev[1*] Saida Tastanova[2]

1. Tashkent University of Information Technologies Named after Muhammad, Al-Khwarizimi, Tashkent, 100084, Uzbekistan
2. Almaty No.110 school, Kazakhstan

ARTICLE INFO

Article history
Received: 29 October 2022
Revised: 18 November 2022
Accepted: 24 November 2022
Published Online: 30 November 2022

Keywords:
Fractal
Optimization methods
Simplex-method
Geometric transformations
Computer graphics
Pascal's triangle

ABSTRACT

This article proposes a new approach based on linear programming optimization to solve the problem of determining the color of a complex fractal carpet pattern. The principle is aimed at finding suitable dyes for mixing and their exact concentrations, which, when applied correctly, gives the desired color. The objective function and all constraints of the model are expressed linearly according to the solution variables. Carpet design has become an emerging technological field known for its creativity, science and technology. Many carpet design concepts have been analyzed in terms of color, contrast, brightness, as well as other mathematical concepts such as geometric changes and formulas. These concepts represent a common process in the carpet industry. This article discusses the use of complex fractal images in carpet design and simplex optimization in color selection.

1. Introduction

At present, special attention is paid to the study of the mathematical aspects of the theory of fractals, as well as methods for describing natural processes and phenomena using the ideas of the theory of fractals. Especially when constructing fractal equations, fractal theories, methods and systems of computer graphics are widely used. Therefore, one of the important tasks is the development of geometric models and algorithms for objects with a fractal structure, as well as methods for their implementation. Much attention is paid to solving theoretical and practical problems in the development of technologies for the use of fractal geometric shapes in the design of urban planning and light industry.

The development of technologies for applying fractal

*Corresponding Author:
Oybek Narzulloyev,
Tashkent University of Information Technologies Named after Muhammad, Al-Khwarizimi, Tashkent, 100084, Uzbekistan;
Email: oybek.88.13@gmail.com

DOI: https://doi.org/10.30564/jcsr.v4i4.5215

geometric shapes in the field, creating textures and design remains one of the most important problems. In particular, interest in studying the processes that serve the fractal geometry of nature has led to the emergence of new scientific areas in biology, physics, mathematics, computer graphics, light industry, electronics, radio engineering, astrophysics, materials science, medicine and other sciences. As you know, the development of pattern design is carried out mainly by hand, including fractal patterns, which are quite difficult or impossible to draw by hand, are drawn using computer graphics technologies.

The main attention is paid to new areas of application of fractal analysis of engineering problems. This work is devoted to collecting new results on the application of fractals in technology, both from a theoretical and numerical point of view.

A number of monographs and scientific articles are devoted to the problems of geometric modeling of objects with a fractal structure using analytical methods. B. B. Mandelbrot, G. M. Julia, G. O. Peitgen, R. P. Tylor, B., B.A.Bondarenko, Sh.A.Nazirov [1,2] and other scientists are striving to expand the scope of fractal geometry, including all over the world, by applying them in practice, in the fields of radio engineering and radar, from predicting the value of securities in the market to new discoveries in theoretical physics. In the Republic, Academician B.A. Bondarenko [1,2] contributed to the construction of generalized Pascal triangles, Pascal pyramids and equations of their fractals based on the theory of binomial basic polynomials with arithmetic properties [3].

In 1980, the famous researcher Benoit Mandelbrot discovered the principle according to which the whole world of such structures is formed in some unexpected way, and in 1984 he developed the concept of a fractal [4].

The science of fractals is very young, as they began to emerge with the development of computer technology. So much remains to be learned and so much remains to be explored.

The main reason why fractals are used in various sciences is that they sometimes represent the real world better than traditional physics or mathematics [5].

The study of fractals has expanded in such a short space of time that it is applied to over 200 fields of art and design. In industry, fractals are used to compress images by reducing data redundancy, creating an ideal platform for textile design [1].

The monograph [6] states that fractal structures or geometry currently play a key role in all models of natural and industrial processes that demonstrate the formation of rough surfaces and interfaces. It is noted that computer modeling, analytical theories and experiments have led to significant progress in modeling these phenomena in the wild. In particular, many problems originating in engineering, physics or biology are characterized both by the presence of different time and space scales, and by the presence of contacts between different components through (irregular) interfaces that often connect media with different characteristics. Thus, this work is devoted to collecting new results on the application of fractals in technology, both from a theoretical and numerical point of view.

Methods for determining the optimal sequence of technological operations belong to the category of problems that can be solved by objectively placing binary decisions about the analysis or synthesis of processes, taking into account certain optimization criteria, as well as the corresponding conditions and accuracy requirements or the capabilities available in the production process.

The charm of similarity inherent in fractal elements is associated with many textile patterns. The little subtleties that lead to the decorative variability these designs create is the design. Jane Barnes was the first woman to use fractals. She redefined the fashion textile trend and defined the use of textile software to create designer fabrics [7].

Another of the most important stages of carpet production continues to introduce advanced technologies and new stages of carpet production. One important process takes center stage in carpet design: dyeing to add color and beauty to patterns produced for contemporary carpets. Carpets and fibers are dyed in many different processes at different stages of production, from fibre, yarn or carpet, depending on the use of the product, the economics of the process and the market demand for the color. In this paper, which discusses the technology of carpet dyeing has evolved into advanced processes in the carpet industry [8].

In the carpet industry, carpet and its design are of great importance. Design and production represent the history and experience of generations, changing from culture to culture, from artist to artist. In the modern technological world, fractal design carpets have become another wave of new ideas in design.

The quality of the performance and environmental friendliness of paint in the production of carpets and textiles is one of the important stages in the carpet industry. The use of non-toxic and environmentally friendly natural dyes for textiles has received much attention due to the strengthening of environmental protection to avoid some dangerous synthetic dyes. Therefore, we used natural dyes for flowers in the production of carpets [9]. Natural dyes and their extraction methods have been the subject of several recent studies. Natural dyes are renewable, biodegradable, and they can replace synthetic dyes already used in

the textile industry, as they will lead to cleaner and more sustainable processes.

At the present time, also special attention to the carpet production plays an important role in the optimal choice of colors, taking into account the cost of production.

The carpet industry research partnership involved 13 & 9 Design, a Fractals Research consulting firm founded by the Mohawk Group and Taylor. Two of Richard Taylor's graduate students, Julian Smith and Conor Rowland, worked on the project. At the end of this project, the global flooring and carpet market is expected to reach $450 billion by 2025 [10].

2. Research Methods

A method for constructing images of a fractal form, taking into account the theory of algebraic structures and prime numbers based on Pascal's triangle and geometric transformations.

Using the method of the theory of binomial polynomials, methods and algorithms for visualizing images in a fractal form have been developed, taking into account algebraic structures based on Pascal's triangle and the theory of prime numbers based on Mod *p*.

Binomial coefficients are the simplest combinatorial objects and are defined as the number of separate combinations of m elements, except for k:

$$(1+\tilde{o})^n = \sum_{k=0}^{n} \binom{n}{k} x^m, \tag{1}$$

The general formula for the degree x of a generalized Pascal triangle of the sth order is written as follows:

$$\left(1+x+x^2+\cdots+x^{s-1}\right)^n = \sum_{k=0}^{(s-1)n} \binom{n}{k}_s x^n,\ s \geq 2. \tag{2}$$

For s=2, the binomial coefficient will be as follows:

$$\binom{n}{k}_2 = \binom{n}{k},\ \binom{n+1}{k+1} = \binom{n}{k} + \binom{n}{k+1}. \tag{3}$$

Pascal's triangle has many interesting properties, and one of them follows from considering two of all modules. The elements in Pascal's triangle can be colored by marking them as follows, as shown in Figure 1, all odd numbers are yellow, all even numbers are red.

The points of the sides of the triangle in the initial element of Pascal's triangle are located in the first quadrant in the coordinate system. In this case, the following relationship is established between the point located in the coordinate system and the elements of Pascal's triangle:

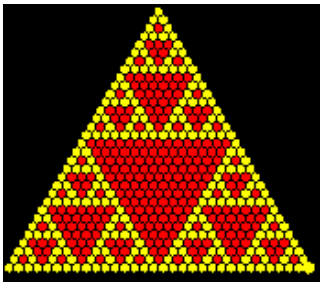


Figure 1. Fractal based on Pascal's triangle Mod=2.

$$\binom{n}{k} = \frac{x!}{y!(x-y)!} \tag{4}$$

where $n = x$, $k = y$.

In the coordinate system, the recurrent relation between geometric shapes (triangle, quadrilateral, hexagon) in the elements of Pascal's triangle is written as follows:

$$\binom{x+1}{y+1} = \binom{x}{y} + \binom{x}{y+1} \tag{5}$$

The formula for determining the elements of Pascal's triangle when dividing it by Mod *p* is written as follows:

$$\binom{n}{m}_p = \binom{n-1}{m-1}_p + \binom{n-1}{m}_p \tag{6}$$

When reducing the elements of Pascal's triangle to prime numbers, when in particular Mod *p* consists of prime numbers, the appearance of Pascal's triangle changes by coloring depending on its division into prime numbers (Figure 2).

Further, based on the method of binomial polynomials, an object with a complex fractal structure was created, taking into account algebraic structures and the theory of prime numbers based on Pascal's triangle. At the first step, we used a transformation in the form of a reflection of computer graphics. Using the geometric reflection transformation, we have:

$$\begin{aligned} x' &= x; \quad x' = -x; \\ y' &= -y.\text{'} \quad y' = y. \end{aligned} \tag{7}$$

Using this formula, the corresponding results are obtained. Applying the geometric transformation of reflection and displacement, the fractal has the form shown in Figure 3a.

Taking into account the geometric transformation of displacement, we have:

$$\begin{aligned} x' &= x + \lambda, \\ y' &= y + \mu. \end{aligned} \tag{8}$$

Using the formula for the geometric transformation of

reflection and displacement, it can obtain the following result (Figures 3b and 3c). Also using the rotation (rotation) geometric transformation formula:

$$\begin{aligned} x' &= x\cos\phi - y\sin\phi, \\ y' &= x\sin\phi + y\cos\phi, \end{aligned} \tag{9}$$

as well as scaling (if necessary):

$$\begin{aligned} x' &= \alpha x, \\ y' &= \delta y, \end{aligned} \tag{10}$$

It can get the following result attached in Figure 3d.

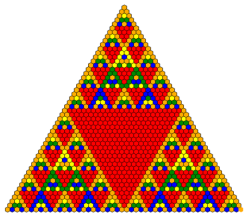
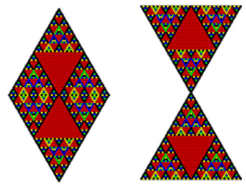
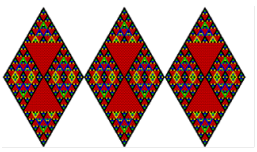


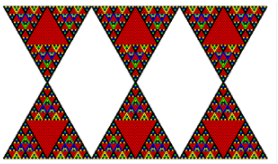
Figure 2. Fractal based on Pascal's triangle Mod=5.



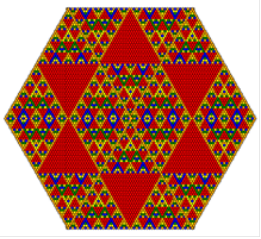
a)



b)



c)



d)

Figure 3. Images of complex fractal structures based on Pascal's triangle according to Mod=5.

3. Method for Optimizing the Color of Fractal Structure Images

The choice of color variants of paints for the manufacture of a particular carpet product is a complex issue, since it is necessary, on the one hand, to ensure the specified accuracy of dyeing, on the other hand, to be able to produce carpets at the lowest cost and at the same time maintain the highest productivity. To solve these problems, in general, methods of mathematical programming and optimization are used.

In real problems, connections inevitably arise between objective functions, criteria:

1) Criteria can be consistent with each other;
2) Criteria may conflict with each other;
3) Criteria can be independent.

Preliminary expert evaluation of selected criteria [11] will allow solving the problem of multicriteria optimization in the simplest, but sometimes most effective ways.

Methods for determining the optimal sequence of technological operations belong to the category of problems that can be solved by objectively placing binary decisions about the analysis or synthesis of processes, taking into account certain optimization criteria, as well as the corresponding conditions and accuracy requirements or the capabilities available in the production process.

In the carpet and textile industry, Pascal's triangle and formulas are used to create patterns, optimization methods for choosing colors and estimating the cost of a product, calculating its cost.

Therefore, it is necessary to develop procedures and algorithms aimed at finding the optimal structure of the synthesized object. These procedures are usually based on the use of mathematical programming methods (mainly discrete programming), sequential and iterative synthesis algorithms, network and graphical design models, as well as heuristic methods of decision theory.

One of the optimization methods is the genetic method [12]. This method has also been considered in the textile industry. This article discusses the genetic algorithm in the textile industry. The genetic algorithm can contain a large number of answers and find the best of them by getting feedback from the tasks. In the jacquard knitting system, several designs of different colors can be made. However, many patterns can be unattractive and lack beauty. According to the customer, it is not easy to choose interesting and stylish models from such a variety of designs. An interactive genetic algorithm can be used to optimize colors for price and select ideal patterns based on desired user feedback. We can use it to optimize color when producing rugs with complex designs.

At present, technology, considered as the science of methods and means of converting materials, or in general the production processes used to produce a product, has evolved recently thanks to approaches aimed at systematizing scientific knowledge, developing strictly reasoned methods capable of developing rational solutions for building modern technological systems. Now, in many areas of industry, the technological preparation of the manufacturing industry is forced to solve many problems and complications in a short time, a situation that often requires summing up decisions in project activities that are aimed only at execution with obviously unfavorable results [13].

Finding the optimal solution in many important tasks involves the analysis and selection of an element in a certain set of feasible solutions, in which the enterprise will receive the maximum profit. Here, "tolerable plan" means a plan that can realistically be carried out taking into account all the capabilities of the enterprise, i.e. subject to restrictions on material, energy, human and similar resources. In this case, the optimality criterion is the achievement of the maximum objective function - income. Thus, the problem under consideration is the problem of finding the maximum of the objective function on the set of feasible solutions. The latter is determined by a number of restrictions on the variables of the problem, which are given in the form of inequalities and equalities. Problems of this kind are usually called resource allocation problems.

In carpet design, it is formed using complex fractal images and parameters. The fractal image uses y_i types of colors, of which $i=\overline{1,p}$. Natural colors are used in the production of carpets. Natural colors have x_j prices, $j=\overline{1,m}$. m is the number of pieces in the price set of natural color (Table 1). The buyer offers $S_{\max}$ and $S_{\min}$ prices for the carpet and can choose k color in the fractal image, when choosing $k<p$. k values of $S_{\max}$ and $S_{\min}$ are recalculated using the following formulas.

$$S^1_{\min}=S_{\min}-\sum_{l=1}^{k}x_l y_l, l=\overline{1,k}, \tag{11}$$

$$S^1_{\max}=S_{\max}-\sum_{l=1}^{k}x_l y_l, l=\overline{1,k}. \tag{12}$$

Here x_i is the price of recognition of colors chosen by the customer and natural colors are excluded from the price list, y_i indicates how many times the selected colors are used in the fractal image. The carpet manufacturer must find a x_i way to keep costs to a minimum with $i=\overline{1,p}$, $p_1=p-k$. This optimization problem can be solved by the simplex method, for which the objective function and conditions are written as follows.

$$\begin{aligned} &F=\sum_{i=1}^{p_1}x_i y_i \to \min \\ &x_{i\max}\geq x_i \geq x_{i\min} \\ &S^1_{\max}\geq \sum_{i=1}^{p_1}x_i y_i \geq S^1_{\max},\ i=\overline{1,p_1},\ p_1=\overline{1,p-k} \end{aligned} \tag{13}$$

Computational algorithm:

Step 1: The values n and p are given instead;

Step 2: The k value is set;

Step 3: The values $S_{\min}$ and $S_{\max}$ are given;

Step 4: Found $y_i, i=\overline{1,p}$;

Step 5: k colors are selected. $p_1=p-k$, then $y_i, i=\overline{1,p_1}$ will be. (S_{max} and $S_{\min}$ are found by formulas (11) and (12));

Step 6: According to the formula (13) the values of x_i are found;

Step 7: If there are equal cases in x_i, the first value is the current value and the second value is the price of the next color that is more expensive.

4. Computational Experiment

Carpet company manufactures and sells carpet products. For the production of carpets, natural dyes are used.

A fractal pattern with 7 colors of a certain size was chosen and the cost of painting the carpet was calculated. The number of each color in Pascal's triangle, consisting of seven colors, taken for example 489, 263, 66, 140, 97, 84 and 136. Let's imagine that the client chose the main 2 colors and the number of 489 and 263. And he designated an intermediate amount for the carpet. This is ours there will be a maximum and a minimum. $\max=75,000.00$ and $\min=65,000.00$.

$$S'_{\min}=65,000.00-489*50-263*40=30,040.00,$$

$$S'_{\max}=75,000.00-489*50-263*40=40,040.00.$$

In the process of solving the problem, we use the values given in Table 1 for each color.

Let's select 7 colors from the preset paints CEDAR YELLOW EXTRACT Myrobalan, MALLOW GOLD EXTRACT Pomegranate Peel, ESTEBIO INDIGO Indigo, CANDY ORANGE EXTRACT Annatto, WINE RED EXTRACT 4001 Lac, ONION PEEL EXTRACT Onion Skin, SUN YELLOW EXTRACT Marigold) from the Table 1.

$\delta_1=50$ c.u. and $\delta_2=40$ c.u.

Let's denote y_1, y_2, y_3, y_4, y_5 the corresponding amount of paint for carpets.

Table 1. Prices for natural dyes

S. NO	Product	ENGLISH NAME/ COMMO	Prices/ Kg/c.u.
1	MICHIGAN BROWN EXTRACT	Black Catechu/Kattha	11
2	CEDAR YELLOW EXTRACT	Myrobalan	24
3	MALLOW GOLD EXTRACT	Pomegranate Peel	26.5
4	GARNET BROWN EXTRACT	Bark of Acacia	35
5	GALLNUT EXTRACT Tannin Grade	Aleppo Oak/Oak	37
6	CEDAR GRAY EXTRACT	Myrobalan	37
7	GALLNUT EXTRACT Dyeing Grade	Aleppo Oak/Oak	39
8	ESTEBIO INDIGO	Indigo	40
9	NUT BROWN	Areca Nut	42
10	CANDY ORANGE EXTRACT	Annatto	47
11	WINE RED EXTRACT 4001	Lac	50
12	APSRA YELLOW EXTRACT	Himalayan Rhubarb	50.5
13	TURKEY RED EXTRACT RT	Madder	61
14	ONION PEEL EXTRACT	Onion Skin	78
15	SUN YELLOW EXTRACT	Marigold	85

Constraints that task variables must satisfy:

$$x_1, x_2, x_3, x_4, x_5 \geq 0$$

$$n = 50(32, 98, ...) \text{ line}$$

$$P = 7(m)(2, 3, 5, 7.....) \text{ fuzzy numbers}$$

Next, we solve the optimization problem for the remaining three colors for. The remaining colors are selected by the simplex optimization method from the remaining colors of the tables remaining by 13 colors.

Then calculate the maximum and minimum:

$$S'_{\min} = 65,000.00 - 489*50 - 263*40 = 30,040.00,$$

$$S'_{\max} = 75,000.00 - 489*50 - 263*40 = 40,040.00$$

We select Indigo Blue (ESTEBIO INDIGO) and Wine Red (WINE RED EXTRACT 4001, Lac) from the color table as preset colors.

Target function of the task:

$$F = \sum_{i=1}^{p_1} x_i y_i \to \min.$$

Denote F -income from the sale of carpets, then the objective function of the problem is written as follows:

$$F = 66x_1 + 140x_2 + 97x_3 + 84x_4 + 136x_5 \to \min.$$

Thus, the task is to find $\min F = 66x_1 + 140x_2 + 97x_3 +$, $84x_4 + 136x_5 \to \min$ under the constraints:

$$11 \leq x_1 \leq 85$$
$$11 \leq x_2 \leq 85$$
$$11 \leq x_3 \leq 85$$
$$11 \leq x_4 \leq 85$$
$$11 \leq x_5 \leq 85$$

$$66x_1 + 140x_2 + 97x_3 + 84x_4 + 136x_5 \leq 40.030$$
$$66x_1 + 140x_2 + 97x_3 + 84x_4 + 136x_5 \geq 30.030.$$

Using the simplex method to solve this optimization problem, it get the following results:

$x_1 = 47.00;\quad x_2 = 78.00;\quad x_3 = 26.50;\quad x_4 = 24.00;\quad x_5 = 85.00;$

$$F(x_1, x_2, x_3, x_4, x_5) = 30,168.50.$$

So, the cost of the flowers that used for the carpet, if it is equal to $x_1 = 47.00$ c.u.; $x_2 = 78.00$ c.u., $x_3 = 26.50$ c.u.; $x_4 = 24.00$ c.u.; $x_5 = 85.00$ c.u; $x_6 = 50.00$ c.u, $x_7 = 40.00$ c.u ; carpet $F = 65,138.50$ c.u. equal to the total cost of production.

Below is Pascal's triangle and hexagon for the design of the carpet, obtained from the simplex by optimization at the best prices of the 7 color paints that we used for the carpet (Figures 4, 5)

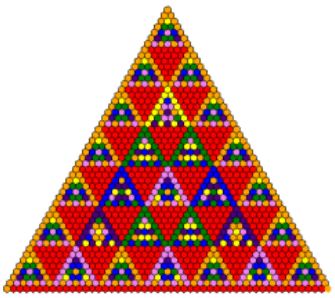


Figure 4. Images of complex fractal structures based on Pascal's triangle by Mod=7 for a carpet obtained from a simplex by the optimization method.

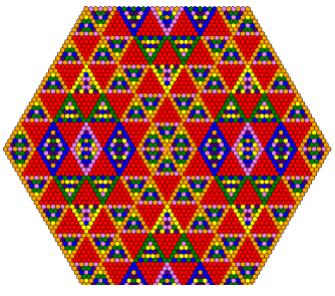


Figure 5. Images of complex fractal structures based on Pascal's triangle by Mod=7 for a carpet obtained from a simplex by the optimization method.

5. Discussion

It has found that fractals can only be used in the exact sciences, carpet design and much more.

Nowadays, a great potential for increasing the level of profitability and profitability of industrial enterprises producing carpets lies in the use of mathematical programming methods in the development of the main production plan of the factory. At the same time, the profit of the enterprise can be increased not only from the sold carpets, but also by reducing costs as a result of a more rational mixing of individual components and color dyes in the manufacture of carpet products.

Therefore, in the production of carpet products of a given size, it is advisable to use mathematical programming methods to select the optimal set of colors, which can be done at the lowest cost and maximum benefit.

A great potential for increasing the level of profitability and profitability of industrial enterprises producing carpets lies in the use of mathematical programming methods in the development of the main production plan of the factory. At the same time, the profit of the enterprise can be increased not only from the sold carpets, but also by reducing costs as a result of a more rational mixing of individual components and color dyes in the manufacture of carpet products.

6. Conclusions

Using the method of the theory of binomial polynomials with arithmetic properties, an algorithm for the method of visualizing images in fractal form has been developed, taking into account the theory of algebraic structures and prime numbers based on Pascal's triangle. This method and algorithm made it possible to visualize fractal shapes based on Pascal's triangle.

An algorithm for constructing complex images of a fractal structure using two-dimensional geometric transformations in the space of computer graphics was developed. Based on this method and algorithm, triangular, square and hexagonal spiral fractals, as well as complex fractal shapes based on Pascal's triangle, were created.

The current widespread use of complex fractal shapes in the carpet industry, the creation of its mathematical model, as well as the use of linear programming methods and optimization methods, can move to a new innovative modern stage in the carpet industry by evaluating the cost of production and creating cost optimization.

Therefore, in the production of carpet products of a given size, it is advisable to use mathematical programming methods to select the optimal set of colors, which can be done at the lowest cost and maximum benefit.

Based on this, an algorithm was developed to improve the efficiency of the carpet factory using complex fractal image forms and the simplex method was applied. In addition, natural colors were used for carpets of a certain size, and prices were taken from India. The optimal price of carpet colors selected on the basis of this algorithm is determined at the request of the customer. As an example, 50 rows and 7 color numbers from Pascal's 7-color triangle were used. We had 5 colors and let the buyer choose 2 colors so that we can find the most optimal price using the maximum and minimum amounts set by the buyer to the best of their opportunities.

One of the optimization methods is the genetic method [14]. This method has also been considered in the textile industry. This article discusses the genetic algorithm in the textile industry. The genetic algorithm can contain a large number of answers and find the best of them by getting feedback from the tasks. In the jacquard knitting system, several designs of different colors can be made. However, many patterns can be unattractive and lack beauty. According to the customer, it is not easy to choose interesting and stylish models from such a variety of designs. An interactive genetic algorithm can be used to optimize colors for price and select ideal patterns based on desired user feedback. We can use it to optimize color when producing carpets with complex designs.

Conflict of Interest

There is no conflict of interest.

References

[1] Bondarenko, B.A., 2010. Generalized Pascal Triangles and Pyramids, their Fractals, Graphs, and Applications – USA, Santa Clara: Fibonacci Associations, The Third Edition. pp. 296.

[2] Bondarenko, B.A., 1991. Generalized triangles and Pascal pyramids, their fractals, graphs and applications. Discrete Mathematics. 3(3), 159-160.

[3] Nazirov, Sh.A., Anarova, Sh.A., Nuraliyev, F.M., 2017. Fundamentals of Fractal Theory, Tashkent: Navruz. Monograph. pp. 128.

[4] Mandelbrot, B., 2002. Fractal geometry of nature. Moscow: Institute of Computer Research.

[5] Anarova, Sh.A., Nuraliev, F.M., Narzulloev, O.M., 2019. Construction of the equation of fractals structure based on the rvachev r-functions theories. Journal of Physics Conference Series. 1260(072001), 1-9.

[6] Nikolaev, E.V., 2016. Fractal Manifesto. Collection of articles / Per. From English, French St. Petersburg: "Strata". pp. 156.

[7] Ponder, C., Gregory, B., Griffing, E., et al., 2019. Life Cycle Comparison of Carpet Dyeing Processes. Journal of Advanced Manufacturing and Processing. 1(e10012), 1-6.
DOI: https://doi.org/10.1002/amp2.10012

[8] Smith, J.H., Rowland, K., Moslehi, S., et al., 2020. Relaxing Floors: The Fractal Fluency of the Built Environment. Nonlinear Dynamics, Psychology and Life Sciences. 24, 127-141.

[9] Nuraliev, F.M., Anarova, Sh.A., Narzulloev, O.M., 2019. Mathematical and software of fractal structures from combinatorial numbers. 2019 International Conference on Information Science and Communications Technologies (ICISCT), Tashkent, Uzbekistan. pp. 1-4.

[10] Malea, C.I., Nitu, E.L., 2020. Optimization of the technological process and equipment of complex profiled parts. IOP Conference Series: Materials Science and Engineering. 916(012058), 1-13.

[11] Zak, Yu.A., 2014. Applied problems of multiobjective optimization. Economics. pp. 455.

[12] Ghoreishian, S.M., Maleknia, L., Mirzapour, H. et al., 2013. Antibacterial properties and color fastness of silk fabric dyed with turmeric extract. Fibers and Polymers. pp. 201-207.

[13] Alevras, D., Padberg, M.W., 2001. Linear Optimization and Extensions: Problems and Solutions PDF. New York: Springer. pp. 451.

[14] Dariush, S., Mehdi, H., Hamed, A., et al., 2014. Jacquard pattern optimizing in weft knitted fabrics via interactive genetic algorithm. Fashion and Textiles. pp. 1-9.

BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

Journal of Computer Science Research
https://ojs.bilpublishing.com/index.php/jcsr

ARTICLE

Bidirectional Recurrent Nets for ECG Signal Compression

Eman AL-Saidi Khalil El Hindi*

Computer Science Department, King Saud University, Riyadh, Saudi Arabia

ARTICLE INFO

Article history
Received: 25 October 2022
Revised: 23 November 2022
Accepted: 28 November 2022
Published Online: 30 November 2022

Keywords:
BERT model
Convolutional neural networks (CNN)
Data compression
Deep learning
ECG diagnosis

ABSTRACT

Electrocardiogram (ECG) is a commonly used tool in biological diagnosis of heart diseases. ECG allows the representation of electrical signals which cause heart muscles to contract and relax. Recently, accurate deep learning methods have been developed to overcome manual diagnosis in terms of time and effort. However, most of current automatic medical diagnosis use long electrocardiogram (ECG) signals to inspect different types of heart arrhythmia. Therefore, ECG signal files tend to require large storage to store and may cause significant overhead when exchanged over a computer network. This raises the need to come up with effective compression methods for ECG signals. In this work, the authors investigate using BERT (Bidirectional Encoder Representations from Transformers) model, which is a bidirectional neural network that was originally designed for natural language. The authors evaluate the model with respect to its compression ratio and information preservation, and measure information preservation in terms of the of the accuracy of a convolutional neural network in classifying the decompressed signal. The results show that the method can achieve up to 83% saving in storage. Also, the classification accuracy of the decompressed signals is around 92.41%. Furthermore, the method enables the user to balance the compression ratio and the required accuracy of the CNN classifiers.

1. Introduction

Recently, cardiovascular diseases (CVDs) have become one of the main causes of mortality in the world. The number of people died from CVDs is estimated at 17.9 million in 2019 according to the world health organization (WHO) [1]. ECG signal represents the electrical signals changes during a cardiac cycle. It shows heart muscle polarization and depolarization, and it is widely used by automatic system to inspect different types of heart arrhythmia. ECG consists of three main components: the P wave, which represents atria depolarization; QRS complex, which represents ventricles depolarization; and the T wave, which represents ventricles repolarization [2]. Figure 1 illustrates the components of an ECG signal. ECG signals need to be long to obtain an accurate diagnosis of a patient condition. Therefore, developing a practical automated system for dealing with ECG signals requires

*Corresponding Author:
Khalil El Hindi,
Computer Science Department, King Saud University, Riyadh, Saudi Arabia;
Email: khindi@ksu.edu.sa

DOI: https://doi.org/10.30564/jcsr.v4i4.5204

compressing these signals before storing or transferring them over a computer network [3].

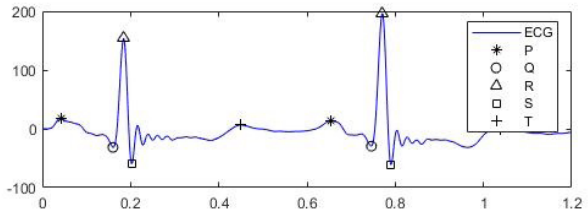


Figure 1. ECG components

The main challenge is to compress an ECG signal in such a way that retains its discriminating features. This is a major challenge especially if we want to attain a high compression ratio, at the same time. It is therefore important to evaluate compression performance in terms of both high compression ratio and quality of decompressed signals to ensure that the reduction in size does not cause a serious loss in diagnostic information.

Deep learning methods have proved to be good predictors in many domains, and according to information theory, good predictors can be utilized as effective compressors [4]. Many existing compressors attempt to learn from data and perform prediction-based compression. Deep convolutional auto-encoders have been used to compress ECG. The encoder section of the suggested model reduces ECG signal to low-dimensional vectors, and the signals are reconstructed by a decoder [5]. Also, auto-encoders with long short-term memory (LSTM) were proposed for ECG signal compression, where LSTM network helps to reconstruct the original data from its compressed file [6], a deep compression approaches can learn to compress different ECG records automatically and perform well in terms of compression ratio and quality of reconstructed signal [7].

In this paper, we introduce a novel Bert-based ECG compressor for ECG. BERT was originally designed for natural language applications. We design a Tiny BERT that is pre-trained on unlabeled ECG beats and then fine-tuned to predict next beats based on previous ones. We exploit the BERT predictor as a compressor for ECG signal. Our method is faster than transformational methods, as it can learn automatically from different ECG records without any need to analyze the domains. It achieves a competitive compression ratio while retaining much of diagnosis information. This is evident as we can use the reconstructed (decompressed) signal to classify it using 1D-CNN with a high accuracy.

We can summarize the main features of our compressor as follow:

i. The proposed method does not perform any analysis of the domain to generate the compressed file. The model is pre-trained to learn ECG beats using the MIT-BIH data-set, and then the model is fine-tuned (trained) for different ECG records to predict next beats based on a number of saved beats and thus act as a compressor for different ECG signals.

ii. The proposed model performance -in terms of its ability to retain diagnostic information- has been validated by using 1D-CNN to diagnose heart problems using the decompressed ECG signals as input data. Our empirical work shows that the classification accuracy of decompressed signal using 1D-CNN is very close to the accuracy of the original ECG signals.

iii. Our system can be used in real time clinical diagnosis, where ECG can be recorded, compressed and then transformed through different channels to other parties who can decompress it and use for classification.

The organization of this paper is as follow: in Section 2, we review some related works. Then in Section 3, we describe the proposed method. Section 4 evaluates our compression algorithm and describes ECG data-set. In Section 5, we present and analyze the results. Section 6 is the conclusion.

2. Related Works

In this section, we describe the different methods for compressing ECG signals and we review transformers in general and the BERT model in particular.

2.1. Compression Methods for ECG Signals

ECG signal represents Cardiac electrophysiology, and it has been widely used in cardiac medical diagnosis and has the advantages of real time. ECG recording is long and there is a need to compress the recordings during storage and transmission phases to make an efficient use of channels.

Many automatic compression techniques have been proposed in the literature to compress ECG signals and they can be classified into three main classes of parameter extraction methods, transformation methods and direct methods [7-15]. These three classes differ mainly in processing domain. Direct methods analyze time domain features to extract samples from ECG to act as compressed file, and then the original signal is reconstructed by an inverse process. Direct methods require time and efforts in designing an appropriate compressing algorithm signals through the extraction of a subset of samples from the original signal sample set, they are usually fast because they use heuristic analysis [7]. Transformation domain methods transform ECG signal to frequency domain or other domains

to analyze energy distribution and compress ECG signal, then an inverse transformation is performed to reconstruct the original signal [14-18]. Parameter extraction method is based on extracted optimized and discriminated features from original signal to reduce the size of ECG signal [4].

In time-domain based methods, a subset of representative samples is extracted from ECG time domain. Usually this task is tedious and time consuming; the reconstruction method is performed by an inverse process. Kumar et al. [19] developed an improved Amplitude Zone Time Epoch Coding (AZTEC) over existing (AZTEC). Statistical parameters of ECG signal are extracted and adapted in-accordance with the nature of signal by calculating a threshold value. The proposed method optimized a tradeoff between ECG information and data reduction. The performance of method was evaluated on the basis of compression ratio and reached a value of ranges between 2.76 and 9.91. L. Zheng et al. [7] proposed a decomposition method utilized singular value decomposition (SVD) to decompose ECG signals and compress it; the ECG signal in the proposed method is divided into segments by QRS peak detection of length n, each segment represents a part of ECG morphology. At end, the ECG signal is represented by a matrix composed of the normalized ECG segments to be compressed. SVD is applied on the matrix to compress ECG signal then the decompressed data are passed to a convolutional neural network (CNN) and supporting vector machine (SVM) for ECG classification. They obtained an average accuracy exceeds 96%.

In transformation-domains based methods, ECG signal is transformed to another domain and the energy distribution is analyzed to compress ECG in the new domain. S. M. Ahmeda et al. [20] used four different discrete wavelet transforms, then the wavelet coefficients are linearly predicted and errors are minimized. A compression ratio of 20 to 1 is reached after encoding the residual sequence obtained using linear prediction. J. Chen et al. [21] presented a new ECG compression technique based on adaptive quantization strategy and orthogonal wavelet transform, by which a user-specified percent root mean square difference (PRD) of the reproduced ECG segments is guaranteed at the minimum entropy.

H. M. Tun et al. [22] proposed transform ECG compression technique, and the method optimized compression ratio by removing redundancy in ECG signal. They used different types of wavelet to decompose ECG signal. The decomposed signals are compressed by applying global and local thresholding and run-length encoding. The performances of different types of wavelet are evaluated in terms of compression ratio (CR) and percent root mean square difference (PRD). M. Elgendi et al. [15] transformed ECG signal to frequency domain using discrete cosine transform to compact ECG signal energy to lower frequency coefficients, then coefficients normalization process is performed. Finally, the normalized and rounded coefficients are encoded using Huffman encoding and run length encoding to compress data.

K.C. Hung et al. [23] proposed a genetic algorithm (GA) to optimize the quantization of wavelet ECG signals to compress data, where GA define stationary relationship property to control quantization scales of multi-resolution levels by using a single variable. Quantization scheme with linear distortion characteristic that does not depend on ECG signal is derived to reduce error. C. K. Jha et al. [24] proposed a method that used adjustable parameters to compress ECG depending on the tunable Q-wavelet transform that reduce the energy of ECG signal to lower transform coefficients, quantization and rounding of coefficients is performed to discard small values and then apply encoded using run-length coding. ECG records of the MIT-BIH arrhythmia were used to measure algorithm performance. The average compression performance obtained in terms of CR is 20.61. Also, they used support vector machine to measure the quality of reconstructed signals, results reported accuracy of 98.35%. It indicates that the proposed method compressed ECG signal efficiently and preserved diagnostic information. S. K. Mukhopadhyay et al. [25] proposed a compression technique of ECG signal, first the ECG signals was de-noised and threshold, then ECG beats were extracted and arranged into (2D) matrix, the generated matrix was decomposed by (SVD), the number of singular matrices are truncated and encoded into ASCII characters, the LLACE compressed left truncated singular matrix coefficients. Performance of method is expressed in terms of subjective measure, where results indicated that quality of reconstructed ECG is very good. Discrete wavelet transform was applied to first intrinsic mode and four sifting functions. Then quantization and rounding were performed on the transformed coefficients. ECG compressed file was generated by running length encoding. The proposed methods outperformed existing methods in terms of quality of reconstructed data [26]. P. Kanjalkar et al. [28] applied wavelet transform to the ECG signal which reduces energy to a smaller number of transform coefficients. The coefficients are quantized and all small-value coefficients were discarded while others were kept according to the formulated quantized output interval. An average value is assigned to some coefficients, and then the coefficients are encoded using modified run-length coding. The average compression ratio was 17.18 for 84 ECG records. In most cases, direct methods are better than transform methods with respect to system

complexity and error control mechanism, but transform methods usually achieve higher compression ratios and elaborate noise contained in original ECG signals.

Deep learned features systems have outperformed hand-engineered features in the field of ECG arrhythmia classification systems, and it is known from theory of information that a good predictor can act as a good compressor. So, many of new suggested compressors attempt to learn models for the data and accomplish compression using prediction. Neural networks show excellent results in many predictions tasks. Recurrent neural networks were used as data compressor and the results show that deep learning models can be good alternatives for traditional compression techniques [4].

2.2 Transformers and the BERT Model

BERT, which stands for Bidirectional Encoder Representations, is based on transformers and its main principle-attention, which understands the textual relationship between different input words by determining the importance of each word with respect to others in a sentence or which words are relevant to come together. BERT produces language model through encoding rather than decoding the encoded information, so using an encoder of a transformer is sufficient. Contrasted to directional models such as RNN and LSTM which process each input sequentially (left to right or right to left). Transformer and BERT are non-directional, because both do not perform sequential ordering; they take in the whole sentence as input. The non-directional property allows models to learn the textual information of a word with respect to all other words in the sentence. BERT has been used in language representation models and it shows good results in natural language processing tasks such as text classification, sentence classification, and semantic similarity between pairs of sentences, question answering tasks, and text summarization. The novelty of BERT is that it acts as language model that uses attention to solve tasks which contains sequences without using recurrent connections, which delay the training process, so we felt motivated to use it as a language model for ECG signals.

As we mentioned earlier BERT is a transformer-based model, it composed of encoder only, while the transformer is composed of an encoder and decoder. BERT uses different hyper-parameters such as attention heads to achieve the best performance than typical transformers.

A transformer is a deep learning model that weights the significance of each input through adopting self-attention mechanism; it is designed to process the entire sequential input data at once. It has Encoder-decoder architecture, as can be seen in Figure 2 [29].

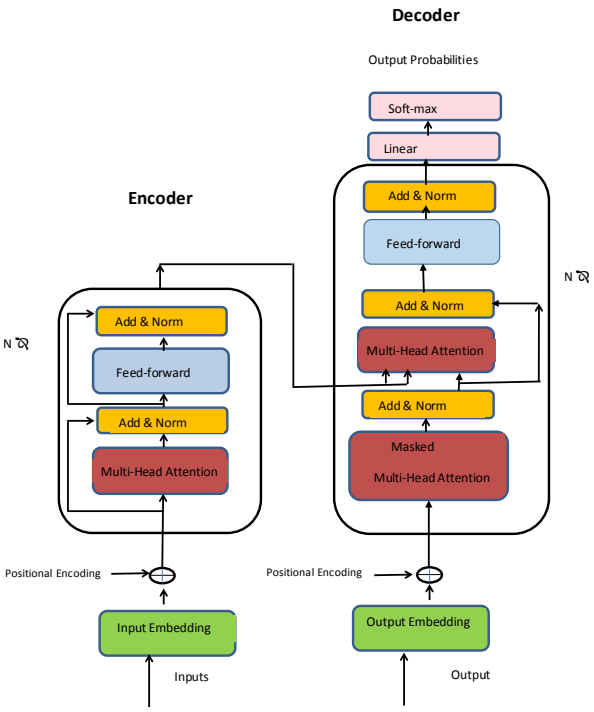


Figure 2. The transformer model architecture

Encoder: Encoder makes up of a stack of a number of identical encoding layers and each layer's input is the output of the previous layer. Each encoding layer generates encoding representation of each input, which contains information about relevancy of each input to others and passes it to the next encoding layer as input. Each layer has multi-head self-attention sub layer, and fully connected feed-forward network sub layer. A residual connection followed by layer normalization is performed around each of the sub-layers.

Decoder: The decoder is also composed of a stack of a number of identical layers; each decoder layer performs the opposite of the corresponding encoder, taking all the encoding representation of inputs to generate an output sequence. To accomplish this, the decoder has a third sub-layer, which performs multi-head attention over the output of the encoder stack. Similar to encoder a residual connection followed by layer normalization is performed around each of the sub-layers.

Encoder and decoder layers use attention mechanism. We give a brief description of self-attention mechanism.

Self-attention mechanism: The transformer building blocks are attention units that produce embedding for every input. Each embedding contains information of other relevant inputs, each weighted by its attention weight. Attention in transformers is based on scaled dot production attention, which is performed with the definition of three concepts: queries (Q), keys (K), and values (V). Keys represent input; queries represent requests for keys

of significant information while checking all available keys. And values represent the retrieved values of queries matched keys. The definition of attention layer is given in Equation (1):

$$\boldsymbol{Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d^k}}\right)V} \quad (1)$$

where d is the dimension of the input vector.

For each an attention unit the transformer model needs to learn three weight matrices; the query weights W_Q, the key weights W_K, and the value weights W_V. One set of (W_Q, W_K, W_V) is called attention head, and each layer has multiple attention heads. Each attention head attends to the inputs that are relevant to each input, the transformer can do head attention for different types of relevance. For each set of weights (i), the head definition can be seen in Equation (2):

$$\boldsymbol{Head_i = Attention\left(QW_Q^i, KW_K^i, VW_V^i\right)} \quad (2)$$

3. Proposed BERT-based Compression (BBC) Method

In this section, we describe how to use BERT as a data compression method for ECG signals. The main idea is to save only the beats that BERT cannot predict and that are not predicted with good accuracy by the BERT model. First of all, we need to train the BERT model to understand different types of ECG arrhythmia. To accomplish this task, we pertained BERT to predict the next beat based the previous n beats. We conducted several experiments to determine n, which is the number of beats that we need to predict the next beat with good accuracy. The best accuracy was obtained when we set n to 10.

We then used a number of saved beats to predict the same number of beats; we tried four numbers of beats, 25, 20, 15 and 10 with 32 for different data sets. After performing pre-training stage, our Bert model can now understand different types of arrhythmia types and can be used in a fine-tuning phase to compress different ECG signal. It is worth mentioning that we used Tiny-BERT which is a multi-layer bidirectional transformer encoder based on implementation [29]. In our model architecture, L, H and A refer to number of layers (Transformer Blocks), Hidden size, and number of self-attention heads, respectively. Since four types of beats only, which is much smaller than the number of words in a natural language, we used Tiny-BERT [29] instead of the full BERT architecture. Tiny-BERT is very small compared to word language diversity. We trained number of tiny models with a differing numbers of self-attention, namely 6, 4, and 2. We also experimented with different number of layers namely 4, 3, and 2). The hidden size was set to 128 in all models. The best number of layers was 2 regardless the number of self-attention heads. We used a model of self-attention of 4 and number of layers of 2 with hidden size of 128. We call the chosen model CBERT (for Compression BERT).

Next, we train (fine-tune) CBERT to predict the next block of beats based on the current block of beats. A block has a fixed size of n beats. We set n=20 during this stage.

We use the trained CBERT to compress the ECG signal file as follows; we start with the first block of n beats in the ECG signal file and store it on the output file. We also use this block as input to BERT to predict the next block; also of n beats. We then find the maximum number of consecutive beats p in the predicted block with a good classification accuracy, i.e., above a certain threshold t. We replace the n-p beats that where not predicted correctly in the predicted block with the corresponding beats in the ECG file; we store the block in the output file as use it as input to BERT in the next iteration. The process continues until the end of the ECG signal file. Algorithm-1 describes the compressed method in details.

The decompression method is straight forward. We use the fine-tuned BERT to decompress the ECG signal file as follows; we a start with the first block of n beats in the ECG compressed file and store it on the ECG output file. We also use this block as input to BERT to predict the next block; also of n beats. We concatenate first predicted p beats with n-p in the corresponding compressed file, and store the block in the output file as use it as input to BERT in the next iteration. Process continues until the end of the ECG compressed file. Algorithm-2 describes the decompressed method in details.

Algorithm-1. Bert-Based-Compression

Input: *ECG signal file*

: n a block consists of n beats ;

: t is an accuracy threshold

Output: *compressed ECG signal file*

Let vector V be the first block of n beats in the ECG signal

Store V on the output file (the compressed file)

While not end of ECG signal file do

    Let P be the block of n beats predicted by BERT given V as the input block

    Let p first consecutive beats in P with c

    Replace the remaining n – p beats in P with the corresponding beats in the ECG signal file

    Let V=P

    Remove the first p beats from P

    insert the integer number p at the beginning of P

    Sotre P in the output file

end

Algorithm-2. Bert-Based-Decompression

Input: *ECG compressed file*

: n represents a fixed number of beats

Output: *ECG signal file*

1 *Let vector V be the first block of n beats in the ECG compressed file*

2 *Store V on the output file (ECG signal file)*

3 *While not end of ECG compressed signal file do*

5 *Read p which is the number of correctly predicted beats*

6 *Use CBERT to predict p beats, based on V*

7 *Read n-p beats from the input file file*

8 *Let V be the concatenation of the predicted beats with the read beats*

9 *Store V on the decompressed output file*

10 *end*

Our empirical work showed that CBERT gives better results than BERT; this is probably because our problem is much simpler than the natural language processing application that BERT was designed for. The full BERT seems to over-fit our training data.

4. Evaluating the Proposed Bert-Based Compression Method

In this section, we describe the empirical results we obtained when using our Bert-Based compression method to compress some real world ECG signal files. Data accuracy is very important to help cardiologists making an accurate diagnosis. It allows quick treatments and can save patients life. So, while compressing ECG signal during transmission and storage, the performance of ECG compression technique must evaluate. The most important issue is to verify quality of the reconstructed ECG signal and preserve the morphological features which are significant to obtain high classification accuracy. Therefore, evaluate BBC not only with respect to the compression ratio but also to the quality of the decompressed file with respect to the diagnostic information it retains. This is done be using the decompressed ECG as input to a pre-trained 1D-CNN to classify it. The closer the obtained accuracy to the accuracy of the 1D-CNN classifier on the original ECG signal, the better the compression method.

Compression ratio (CR) is defined as the ratio of the original signal size and compressed signal size. The CR provides information about how compression algorithm can remove redundant data. The higher the CR is, the less space is used to save it. CR is used as an indicator of how much data has been compressed and thus save space. Also, we use accuracy and F1- measure to measure the performance of reconstructed ECG Signal. The definitions of these measures are given by Equations (3), (4) and (5), respectively.

$$\textbf{\textit{Space saving}}\ (\textbf{\textit{Reduction rate}})\ (\boldsymbol{S}) = \left(\mathbf{1} - \frac{S_c}{S_o}\right) * \mathbf{100}\% \quad (3)$$

$$\textbf{\textit{Accuracy}}\ (\textbf{\textit{Acc}}) = \frac{\boldsymbol{TP+TN}}{\boldsymbol{TotalSamplesNumber}} \quad (4)$$

$$\boldsymbol{F1} = \frac{\boldsymbol{2*precision*recall}}{\boldsymbol{precision+recall}} \quad (5)$$

where, S_C and S_O represent the size of compressed file and the original file, respectively. True Positive (TP) refers to number of normal beats correctly classified; True Negative (TN) refers to number of abnormal beats classified as abnormal beats, False Negative (FN) refers to number of abnormal beats classified as normal beats, False Positive (FP) refers to number of normal beats classified as abnormal beats, $\text{Precision} = \frac{TP}{TP+FP}$, $\text{Recall} = \frac{TP}{TP+FN}$.

ECG Data

We acquired benchmark data-sets, which are available from public domains such as MIT-BIH arrhythmia data-set. MIT-BIH is the most commonly used database that was developed to act as an objective evaluation tool for different arrhythmia classification systems; it consists of 48 records of ECG data, each record is 30 minutes long, 23 records of the 48 records were chosen randomly from a large set of long-term ECG Holter recordings to represent variations of ECG data, while the remaining records were chosen specially to represent complex arrhythmia such as supra-ventricular arrhythmia that may encounter arrhythmia classification systems. The recordings were digitized at 360 samples per second per channel with 11-bit resolution over a 10 mV range. Each ECG record contains two leads, each lead is recorded by placing electrodes on classification places on the skin. MIT-BIH data-set was pre-processed and digitized, a set of rhythm labels and beat labels were added to each record, during the early usage of the data-set; some of these labels were revised and corrected several times.

5. Discussion and Results

We acquired data from ECG signal and pre-trained Tiny-Bert models, then fine-tuned multiple ones for different ECG records. In this section, we discuss the results for different cases related to multiple ECG records. Each ECG record is fine-tuned using a pre-trained defined model. During the compression process, we monitor average predication accuracy using BBC algorithm, which ensures that predication accuracy does not go beyond specific threshold. We present three different threshold values of 98%, 95%, and 90% and calculate the efficiency of com-

pression ratio in terms of saving space of three different ECG records donated as (DS1, DS2, and DS3). Each data-set represent 1-hour ECG recording of patient extracted from MIT-BIH database and each data-set has four different types of beats, Normal beat, Left bundle branch block beat, Right bundle branch block beat and Bundle branch block beat We address different model size with a differing number of layers and attention heads to give more insight into the effect of model size on the performance of the model during compression.

We calculate the compression performance of each record in terms of saved space (S). We run different models with various multi head attention and encoders (layers) number. The results are presented in Tables 1, 2 and 3.

Then we calculated the average saved space for each multi-head attention number to clarify the effect of multi-head attention number on different datasets. The results are represented in Table 4. Based on all results in Table 4, it can be observed that there is no significant evidence that number of multi-head attention has strong effect on S. Furthermore, the classification accuracy of all cases differs in range of 0-2.62% which gives an indication that the number of multi heads attention has a small effect on model performance. In addition, we calculate the average accuracy of saved space per number of layers. The results are indicated in Table 5. In most cases, using two layers has the highest performance in term of saving space. To give more insight into our compression technique, a visualization of the reduction rate percentage in each block of the compressed ECG file of DS1 is given in Figure 3. As we can see from Figure 3, the reduction rate in blocks after the first block is between 75% and 85% which gives an indicator of a high compression ratio and thus a good compression technique.

Table 1. Save space with different data-sets and threshold of 95%

Threshold 95%	**Multi heads attention number**								
Layers number	6	4	2	6	4	2	6	4	2
	DS1			DS2			DS3		
4	81.42%	80.76%	81.12%	79.25%	81.22%	81.25%	80.42%	81.87%	80.35%
3	82.87%	83.77%	83.96%	82.14%	82.35%	82.40%	82.54%	80.76%	81.12%
2	83.76%	83.91%	82.31%	83.11%	83.13%	83.25%	82.77%	81.14%	83.22%

Table 2. Save space with diffrent dat-sets and threshold of 90%

Threshold 90%	**Multi heads attention number**								
Layers number	6	4	2	6	4	2	6	4	2
	DS1			DS2			DS3		
4	82.54%	80.43%	81.10%	83.55%	81.75%	81.25%	80.63%	80.44%	81.51%
3	83.11%	82.67%	83.43%	84.14%	82.68%	82.78%	83.48%	83.89%	83.65%
2	83.56%	83.88%	83.34%	84.77%	83.44%	83.56%	81.84%	81.75%	83.23%

Table 3. Save space with different data-sets and threshold of 98%

Threshold 98%	**Multi head attention number**								
Layers number	6	4	2	6	4	2	6	4	2
	DS1			DS2			DS3		
4	80.05%	80.46%	80.50%	77.35%	80.27%	82.44%	81.55%	81.22%	81.25%
3	81.67%	82.75%	83.11%	80.13%	82.35%	81.60%	82.31%	80.51%	81.12%
2	83.11%	83.88%	84.17%	82.40%	82.44%	83.14%	81.78%	80.34%	83.15%

Table 4. Average Save space and classifcation accuracy for diffrent head attention number

Threshold 95%	**Multi head attention number**								
	6	4	2	6	4	2	6	4	2
	DS1			DS2			DS3		
S	82.6%	82.81%	82.46%	81.5%	82.23%	82.3%	81.91%	81.25%	81.56%
Acc	92.78%	91.65%	92.88%	93.22%	93.13%	93.14%	92.67%	93.13%	93.71%
Threshold 98%									
S	81.61%	82.36%	82.59%	79.69%	81.68%	82.19%	81.88%	80.69%	81.84%
Acc	93.73%	92.98%	92.24%	93.67%	92.56%	92.42%	93.43%	92.84%	92.13%
Threshold 90%									
S	83.07%	82.81%	82.46%	84.15%	82.62%	82.55%	81.98%	82.02%	82.79%
Acc	91.44%	91.54%	91.64%	92.11%	91.11%	91.13%	91.99%	91.10%	91.13%

Table 5. Average save space for different layers number

Threshold 95%									
Layers number	4	3	2	4	3	2	4	3	2
	DS1			DS2			DS3		
S	81.1%	83.53%	83.32%	80.57%	82.29%	83.16%	80.88%	81.47%	82.37%
Threshold 98%									
S	80.33%	82.51%	83.72%	80.02%	81.36%	82.66%	81.34%	81.31%	81.74%
Threshold 90%									
S	81.47%	83.61%	83.26%	82.18%	83.2%	83.95%	80.86%	83.67%	82.27%

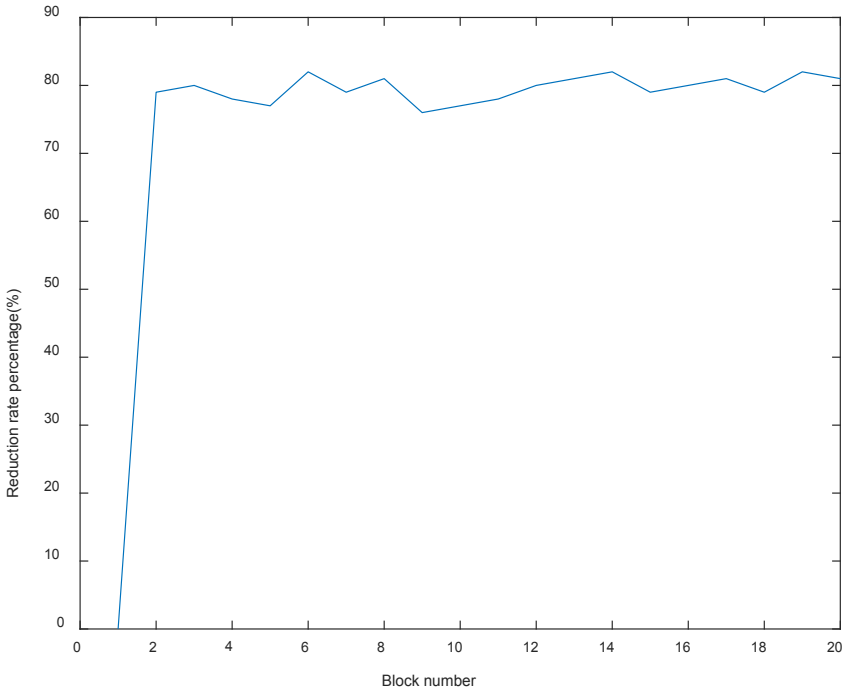


Figure 3. Reduction rate percentage (%) of DS1

To measure the quality of reconstructed ECG signal, we calculated the average classification accuracy of the reconstructed ECG with respect to original files. This process is performed by passing it to 1D-CNN and compared the classification accuracy with respect to original files. The average accuracy of all cases of reconstructed original files reached 92.41% which is less by approximate 2% than our obtained classification accuracy of ECG signal using ensemble model which prove the robustness of our compressor model and prove that it can be used in real time diagnosis. The accuracy is presented in Table 6. Also, we obtained an average F1-score of 0.8266 as indicated in Table 7. In addition that we have achieved a high classification accuracy of the reconstructed ECG with respect to original files, a visualization of an original ECG signal block (before compression) and the reconstructed one (after compression) is shown in Figure 4 and Figure 5 respectively. A key challenge for data compression is to reconstruct the original data that maintains the correct information. If data compression results in the loss of information; it will reduce the performance of the compression technique. According to our data compression it can be easily verified that it is information-preserving.

Our proposed method based on deep learning method. Conventional compression methods like parameter extraction methods, transformation methods and direct methods usually process different domains to extract features. Direct methods analyze time domain features to compress ECG signal. Which require time and efforts [7]. Transformation domain methods transform ECG signal to frequency domain or other domains to compress ECG signal, then an inverse transformation is performed to reconstruct the original signal [14-18]. Parameter extraction method is based on extracted discriminated features from original signal to reduce the size of ECG signal [4]. Since our proposed compression algorithm used deep learning model, it does not perform any analysis of domain to generate compressed file. It acts as an end-to-end model, where there is no need to extract any domain features. Also, it achieved a competitive saving space reached to 83% conventional methods in-terms of save spacing.

In contrast to M.Goyal et al. [4] method which used RNN that represents a shallow concatenation of a left-to-right and a right-to-left model. Our model is a truly bidirectional learning model based on attention mechanism, where the model learns information from left to right and from right to left, which make our model more powerful than their method. Furthermore, our model is designed for ECG signal, which is huge and consume storage over network, so our systems can be used in diagnose systems.

Table 6. The (%) accuracy of different threshold values

Threshold value	95%	98%	90%	Average
Accuracy	92.9	92.88	91.46	92.41

Table 7. The F1 score of different threshold values

Threshold value	95%	98%	90%	Average
F1-score	0.82	0.84	0.82	0.8266

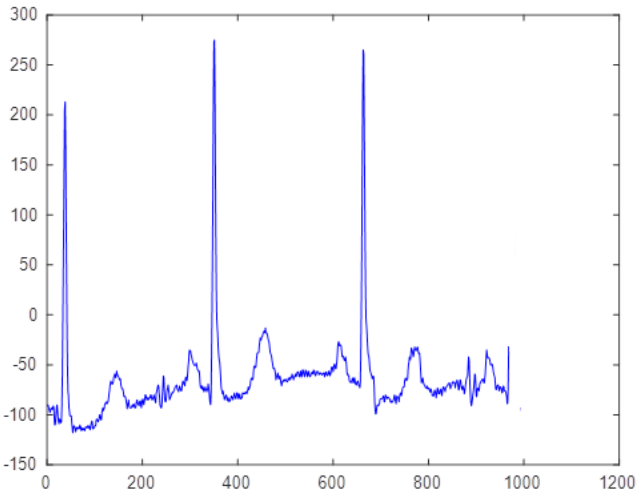


Figure 4. Original ECG signal (before compression)

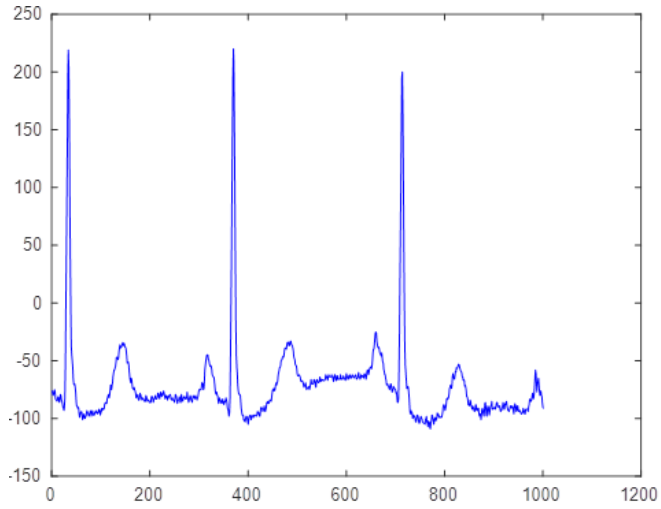


Figure 5. Reconstructed ECG signal (after compression)

6. Conclusions and Future Work

ECG plays an important role in ECG diagnosis; deep learning has become a state of the art ML approach that is widely used in many applications. In this paper, we built a compression system where ECG signal can be recorded and compressed using a novel suggested TinyBERT model, then it can be reconstructed. The quality of the reconstructed ECG is measured by passing it to another deep learning model, 1D-CNN, and achieved a high classification accuracy that is near an ensemble 1D-CNN for classifying 4 types of arrhythmia, proving the effectiveness of our compressor frame. For future work we can apply transfer learning to compress new ECG signals based on pre-trained ECG signal.

Conflict of Interest

There is no conflict of interest.

References

[1] WHO, 2021. Cardiovascular diseases (CVDs).

[2] Aspuru, J., Ochoa-Brust, A., Félix, R., et al., 2019. Segmentation of the ECG signal by means of a linear regression algorithm. Sensors. 19(4), 775.

[3] Eren, H., Webster, J.G., 2018. Telemedicine and Electronic Medicine. CRC Press.

[4] Goyal, M., Tatwawadi, K., Chandak, S., et al., 2018. Deepzip: Lossless data compression using recurrent neural networks. arXiv Prepr. arXiv1811.08162.

[5] Yildirim, O., San Tan, R., Acharya, U.R., 2018. An efficient compression of ECG signals using deep convolutional autoencoders. Cognitive Systems Research. 52, 198-211.

[6] Dasan, E., Panneerselvam, I., 2021. A novel dimensionality reduction approach for ECG signal via convolutional denoising autoencoder with LSTM. Biomedical Signal Processing & Control. 63, 102225.

[7] Zheng, L., Wang, Z., Liang, J., et al., 2021. Effective compression and classification of ECG arrhythmia by singular value decomposition. Biomedical Engineering Advances. 2, 100013.

[8] Rebollo-Neira, L., 2019. Effective high compression of ECG signals at low level distortion. Scientific Reports. 9(1), 1-12.

[9] Zhang, B., Zhao, J., Chen, X., et al., 2017. ECG data compression using a neural network model based on multi-objective optimization. PLoS One. 12(10), e0182500.

[10] Polania, L.F., Plaza, R.I., 2018. Compressed sensing ECG using restricted Boltzmann machines. Biomedical Signal Processing & Control. 45, 237-245.

[11] Abo-Zahhad, M., Ahmed, S.M., Zakaria, A., 2012. An efficient technique for compressing ECG signals using QRS detection, estimation, and 2D DWT coefficients thresholding. Modelling & Simulation in Engineering.

[12] Qian, J., Tiwari, P., Gochhayat, S.P., et al., 2020. A Noble Double-Dictionary-Based ECG Compression Technique for IoTH. IEEE Internet of Things Journal. 7(10), 10160-10170.
DOI: https://orcid.org/10.1109/JIOT.2020.2974678

[13] Rajankar, S.O., Talbar, S.N., 2019. An electrocardiogram signal compression techniques: a comprehensive review. Analog Integrated Circuits & Signal Processing. 98(1), 59-74.

[14] Sahoo, G.K., Ari, S., Patra, S.K., 2015. Performance

evaluation of ECG compression techniques. 2015 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT). pp. 1-5.

[15] Elgendi, M., Mohamed, A., Ward, R., 2017. Efficient ECG compression and QRS detection for e-health applications. Scientific Reports. 7(1), 1-16.

[16] Padhy, S., Sharma, L.N., Dandapat, S., 2016. Multilead ECG data compression using SVD in multiresolution domain. Biomedical Signal Processing & Control. 23, 10-18.

[17] Jha, C.K., Kolekar, M.H., 2017. ECG data compression algorithm for tele-monitoring of cardiac patients. International Journal of Telemedicine & Clinical Practices. 2(1), 31-41.

[18] Chowdhury, M.H., Cheung, R.C.C., 2019. Reconfigurable architecture for multi-lead ecg signal compression with high-frequency noise reduction. Scientific Reports. 9(1), 1-12.

[19] Kumar, V., Saxena, S.C., Giri, V.K., et al., 2005. Improved modified AZTEC technique for ECG data compression: Effect of length of parabolic filter on reconstructed signal. Computers & Electrical Engineering. 31(4-5), 334-344.

[20] Ahmeda, S.M., Abo-Zahhad, M., 2001. A new hybrid algorithm for ECG signal compression based on the wavelet transformation of the linearly predicted error. Medical Engineering & Physics. 23(2), 117-126.

[21] Chen, J., Itoh, S., 1998. A wavelet transform-based ECG compression method guaranteeing desired signal quality. IEEE Transactions on Biomedical Engineering. 45(12), 1414-1419.

[22] Tun, H.M., Moe, W.K., Naing, Z.M., 2017. Analysis on ECG data compression using wavelet transform technique. International Journal of Psychological and Brain Sciences. 2(6), 127-140.

[23] Wu, T.C., Hung, K.C., Liu, J.H., et al., 2013. Wavelet-based ECG data compression optimization with genetic algorithm.

[24] Jha, C.K., Kolekar, M.H., 2021. Tunable Q-wavelet based ECG data compression with validation using cardiac arrhythmia patterns. Biomedical Signal Processing & Control. 66, 102464.

[25] Mukhopadhyay, S.K., Ahmad, M.O., Swamy, M.N.S., 2018. An ECG compression algorithm with guaranteed reconstruction quality based on optimum truncation of singular values and ASCII character encoding. Biomedical Signal Processing & Control. 44, 288-306.

[26] Jha, C.K., Kolekar, M.H., 2021. Empirical mode decomposition and wavelet transform based ECG data compression scheme. IRBM. 42(1), 65-72.

[27] Kolekar, M.H., Jha, C.K., Kumar, P., 2021. ECG Data Compression Using Modified Run Length Encoding of Wavelet Coefficients for Holter Monitoring. IRBM.

[28] Shinde, A.A., Kanjalkar, P., 2011. The comparison of different transform based methods for ECG data compression. 2011 International Conference on Signal Processing, Communication, Computing and Networking Technologies. pp. 332-335.
DOI: https://orcid.org/10.1109/ICSCCN.2011.6024570

[29] Vaswani, A., Shazeer, N., Parmar, N., et al., 2017. Attention is all you need. International Conference on Neural Information Processing Systems. 30.

BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

Journal of Computer Science Research
https://ojs.bilpublishing.com/index.php/jcsr

ARTICLE

CO-BOT: An Intelligent Technique for Designing a Chatbot for Initial COVID-19 Test

Projit Das[1] Ira Nath[1] Avishek Banerjee[2] Laxminarayan Sahoo[3*]

1. Department of CSE, JIS College of Engineering, Kalyani, 741235, India
2. Department of CSBS, Asansol Engineering College, Asansol, 713305, India
3. Department of Computer and Information Science, Raiganj University, Raiganj, 733134, India

ARTICLE INFO

Article history
Received: 28 November 2022
Revised: 15 December 2022
Accepted: 19 December 2022
Published Online: 28 December 2022

Keywords:
Chatbot
COVID
Cloud

ABSTRACT

The coronavirus (nCOV-19), which was discovered, has now spread around the world. However, managing the flow of a large number of cases has proven to be a significant issue for hospitals or healthcare professionals. It is becoming increasingly challenging to speak with a medical expert after the epidemic's initial wave has passed, particularly in rural areas. Thus, it becomes clear that a Chatbot that is well-designed and implemented can assist patients who are located far away by advocating preventive actions, and viral updates in various cities, and minimising the psychological harm brought on by dread. In this study, a sophisticated Chabot's design for diagnosing individuals who have been exposed to COVID-19 is presented, along with recommendations for immediate safety measures. Additionally, when symptoms grow serious, this virtual assistant makes contact with specialised medical professionals.

1. Introduction

To stop the country from seeing an increase in COVID-19 instances, numerous research projects are now being conducted. Before the successful development of these kits, our nation imported medical supplies, including PPE (Personal Protection Kits) and masks, from abroad. Our nation has taken steps to raise awareness of the disease as well as initiatives to combat it. The news and media play an important role in fostering this knowledge by educating the public about the preventive steps that can shield them against infection. Keeping a virus from spreading can be made much harder if people are more aware of the need to take all precautions.

COVID-19 has been quickly spreading around the world. Many people have started to show various viral

*Corresponding Author:
Laxminarayan Sahoo,
Department of Computer and Information Science, Raiganj University, Raiganj, 733134, India
Email: lxsahoo@gmail.com

DOI: https://doi.org/10.30564/jcsr.v4i4.5287

symptoms. When persons with impaired immune systems, the elderly, or those with chronic conditions experience COVID-19 symptoms, it might cause considerable complications [1-4]. Because of the burden, this put on hospitals and medical call centres, people began to become enraged with the COVID-19 virus, especially after multiple false reports about it were spread on social media. For instance, in March 2020, the developing coronavirus resulted in more than 1,400,000 calls to Saudi Arabia's Ministry of Health in a single month [5]. AapkaChikitsak, an Indian Chabot [6], and ANA, a Brazilian Portuguese Chabot [7], both of which were developed to counteract fake news and misinformation and provide trustworthy information regarding COVID-19. Furthermore, several recently released research studies have chosen to use chatbot technology, which is useful in providing telehealth services in the COVID-19 pandemic scenario [8,9] or for numerous goals connected to the healthcare industry, such as medical diagnosis, treatment of drugs, and pathology queries, or even for general medical awareness purposes [10,11].

Using text or voice, a chatbot is computer software that converses with users in normal language, interprets their intentions, and then responds following the organization's business policies and data. ELIZA [12] was among the initial chatbots. It served as an early test case for the Turing Test, which examines whether a machine's intellect can be compared to or is indistinguishable from that of a person. Early consumers felt they were speaking with a person who understood their input since the computer utilised "pattern matching" and replacement methods to offer prepared responses. The software's scripts restricted its functionality. In the middle of the 1960s, Joseph Weizenbaum, a researcher at MIT's Artificial Intelligence Laboratory, developed a pioneering natural language processing method. It was supposedly created to demonstrate how tenuous ties existed between humans and computers at the time. On the other hand, when it was put on computers, people found it quite entertaining. Chatbots have a wide range of uses in the modern world, including online commerce, ticketing, news reporting, meal ordering, etc.

More safety and caution regarding this condition are provided by this chatbot system. The AWS platform, which enables users to forego expert advice, can be used to develop this system. It is designed to determine the user's ailment and deliver pertinent information about the disease. With enhanced access to information regarding diseases, this method is designed to be cost-effective. Only when a Chatbot can diagnose any illness and give consumers the information they need will it be of benefit to them. The suggested system uses a conversational agent to interact with individuals and learn more about their medical issues in order to get an accurate diagnosis. Today's digital environment allows for very high amounts of information interchange per minute, enabling people to conduct business from afar without having to be physically there. Using Human Machine Interaction (HMI) technology, which cut out humans from the process, is used in many services and interaction settings. Despite recent technological advancements, humans will still be an integral part of the contact loop for a very long time. They mostly act as "answers" machines, which is their fundamental issue; they are unable to reliably and effectively react to human inquiries. We propose a novel HMI scheme to address this issue that is based on SPN discourse rather than providing responses to inquiries and is capable of improved communication and engagement with human users.

The objective of the proposed work

The following are the observations of my proposed work:

(i) To determine the possible outcome and analysis COVID report based on some given health conditions.

(ii) This chatbot will store all the responses given by the users and then after analysing the data available doctors or executives call them and discuss their situations.

(iii) The chatbot is used by users to comprehend the facts and news surrounding the sickness, to learn how to protect themselves from the coronavirus, and to help stop the spread of the illness.

(iv) This chatbot gives us information about the comparative situation of COVID-19 trends in different cities.

2. Related Work

There have been several text-based human-computer interaction systems created, such as ELIZA [12], which mimics a psychiatrist, and PARRY [13], which indicates the thoughts of a patient who is paranoid. In two different trials, Raij et al. [14] contrasted virtual human interactions with real human interactions in the context of medical consultation. Their findings demonstrate parallels between real-world and virtual interactions. An article by Fadhils et al. [15] demonstrates the use of intelligent conversational systems to communicate with elderly people in order to gather data and conduct ongoing health condition monitoring, particularly after hospital release. A medical recommendation system particularly created to engage with users and take on the role of a doctor is presented by Amato et al. [16,17]. Pharmabot is a paediatric generic medicine consultant chatbot that Comendador et al. introduce. It is intended to refer to and give relevant data on generic

drugs for kids.

A chatbot that is a personalised conversational assistant was built by Poongothai. M. et al. [18] to aid students in asking inquiries and resolving problems regarding current and completed Internet of Things projects. The advantages of this system include a few characteristics including auto-scaling, real-time notifications, and quick response. Later, their laboratory system was upgraded to include this chatbot.

Ali Parsa founded Babylon Health [19] in 2013, which is a provider of value-based healthcare, that combines an AI-powered stage with virtual clinical procedures for sick persons. Through its website and mobile application, patients are connected with healthcare providers. Users can email queries or photographs to the medical specialist in a manner similar to text messaging. Patients can also video chat with doctors to seek guidance for their health problems. Additionally, users can get referrals to doctors, have their prescriptions for pharmaceuticals delivered to them or forwarded to a pharmacy, and consult with therapists to discuss problems like depression and anxiety. When a physical examination is required, users in London can book health exams with a select group of institutions; nurse visits, however, are only offered at one location.

According to its website, this healthcare chatbot [20] attempts "to assist doctors in their daily work". It functions as an assistant that gives medical users precise data through chat by offering helpful advice and keeping track of a woman's health while she is nursing. Additionally, it assists healthcare professionals in advising women on the appropriate drugs to take while they are breastfeeding. This can be incredibly beneficial for people to search for quick, accurate information anytime it's needed, saving time by having all the information at their fingertips rather than having to spend time looking it up.

Florence [21] chatbot, which functions on Facebook Messenger, Skype, or Kik, is essentially a "personal nurse". She can remind people to take their medication, which could be a useful feature for elderly patients. You simply need to enter the drug's name, how frequently you need to take it, and when in the chat window. Florence then casually reminds you to take the prescription every time you need to. The user's health, including things like body weight, mood, and menstruation, may also be monitored by Florence, who will help them succeed in their goals. The chatbot can also assist you in finding the nearest pharmacy or medical facility if necessary.

To enable objective data reception and transmission in real-time, "Smart Wireless Interactive Healthcare Systems" (SWITCHes) [22] are being developed. In addition to engaging users with customised feedback in an interactive manner through a health chatbot powered by artificial intelligence, the SWITCHes app can also connect users with healthcare professionals who can offer them more precise medical advice depending on the customer's data collected from the app and auxiliary data from medical devices. SWITCH's technology differs from researcher-controlled and commercial apps because of an integrated health chatbot. Users can converse with the health chatbot and receive information in real-time, or they can heed the bot's counsel and adhere to its recommendations for healthy living, which may include diet and exercise regimens. The AI-powered health chatbot may provide real-time information as well as assist users in keeping track of their health issues and staying motivated by providing inspiration and reminders. The healthcare practitioner can give the user more precise medical recommendations based on the user's data collected from the SWITCHes app and the supporting data from medical equipment. A health chatbot powered by artificial intelligence can also communicate with SWITCHes app users to provide personalised feedback.SWITCHes distinguishes itself from researcher-controlled and for-profit apps with its built-in health chatbot.

With the aid of Tidio's [23] conversational AI chatbot, you can improve your customer service and increase sales. You can design your AI bot exactly how you want because it is simple to use. You may also use a visual builder interface to watch how it develops. Your bot will use NLP technology to interact with and serve your customers more effectively. And you'll be accessible to your clients around the clock to ensure that you don't miss any chances to close deals.

3. Proposed Method

Conversational voice and text interfaces can be created using Amazon Lex. Your applications can incorporate chatbots that you create with Amazon Lex. Internally, Amazon Lex uses the same deep learning platform as Amazon Alexa. Amazon Lex's internal technology combines natural language understanding (NLU) and automatic speech recognition (ASR) to comprehend the intent of the text and convert speech to text. As this course progresses, we'll go into more detail on these subjects. Essentially, Amazon Lex frees you from the difficulties of speech recognition and natural language understanding so that you can quickly create really engaging chatbots. A budget-friendly option is Amazon Lex. Like the majority of AWS services, Amazon Lex offers pay-as-you-go pricing and has no upfront expenses. Utilizing a mixture of aliases and versioning, Amazon Lex offers deployment functionalities that let you rapidly roll out your conversational interfaces across numerous environments with-

out any difficulties. You may scale up with Amazon Lex without worrying about bandwidth usage because Amazon Lex does not place restrictions on it. Finally, there is seamless integration between Amazon Lex and a number of other AWS services. Natural language understanding and automatic speech recognition, or ASR, for speech-to-text conversion offer sophisticated ways to transform your voice or text commands into commands that can be carried out. Intents, utterances, slot kinds, slots, and channels are just a few examples of the many child configuration pieces that make up a bot. As we go on, we'll discuss each of these in more detail. Within Amazon Lex, the bot itself serves as the building and deployment unit. Developers are able to create and deploy a variety of bots, each with its own unique set of capabilities. Several child configuration pieces, including intents, utterances, slot types, slots, and channels, to mention a few, make up a bot. In the coming sections, we'll go through each of these in detail. Within Amazon Lex, the bot itself serves as the unit of construction and deployment. The ability to create and deploy numerous bots, each with a unique set of abilities and behaviours, is available to developers. The results that the bot might take are represented by intent. For instance, scheduling an appointment, booking a flight, or reserving lodging are all examples of intentions. A term that accurately conveys what the objective achieves, Utterances. The user says or enters one or more phrases that trigger the purpose, and fulfilment procedure- how the intent was carried out or fulfilled. Aside from the built-in intentions that Amazon Lex offers, you can also use your own custom intents. For each goal, the user may need to give more qualities, also called slots, in order to get the desired result.

The working method of this chatbot is that the user first gives some query-related statement to the chatbot, and the chatbot will check the spelling with its stored intents. After that, if both statements are the same, the chatbot will reply, and then our chatbot will ask for some personal information from the user. The next user asks questions as per their demand. If they want to check COVID status or COVID-related basic questions, or if COVID updates in different cities, our chatbot will reply as per stored value in the database. In the end, the chatbot will show the thank you, and if the user gives some wrong inputs, it will show an error response. Here in Figure 1, we briefly describe the working flow diagram of our chatbot.

Figure 1 describes when a user or patient gives some input or comments as per their needs, then CO-BOT will start replying. Before that, CO-BOT will check the input coming from the user with its database. If it matches the database, it will give a suitable response. CO-BOT will save all the information given by a user for further response.

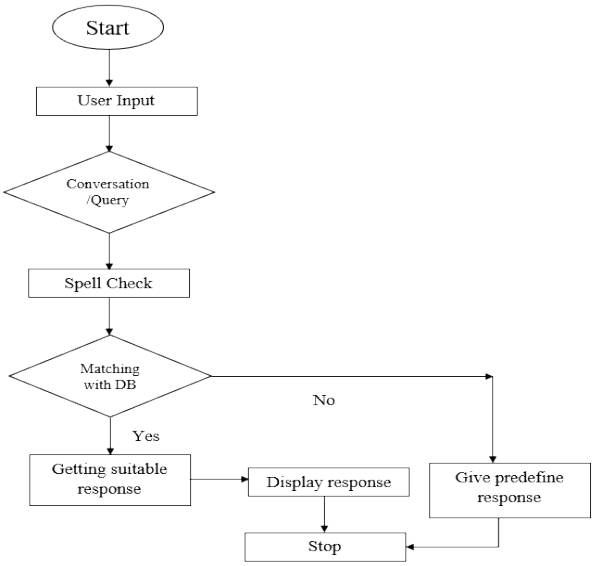


Figure 1. Flow Diagram of COVID Chatbot

In this project, we used Amazon Lex API to create the sample utterances and intent from a chatbot, then connected this Lex with a lambda function and did the validation also. Then create IAM and Cognito for authentication and connect these with S3. Also, connect S3 with a lambda function for static web hosting. Here, by using Figure 2, we have given a clear architecture of our project.

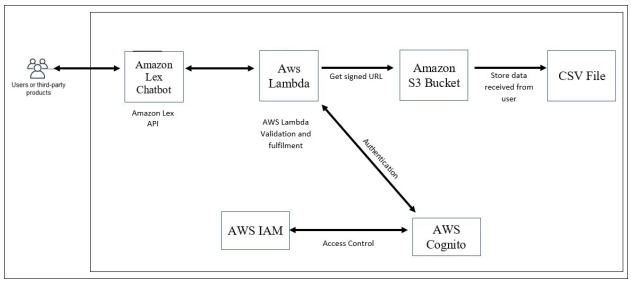


Figure 2. System Architecture of Chatbot

Figure 2 briefly describes the actual working system of CO-BOT. The Amazon Lex API is used to create a bot and configure it with one or more intents that you want to support. Configure the CO-BOT as per the basis of the patient goal, and engages in conversation with the user or patient to elicit information. Next, lex is connected with lambda which allows us to add custom logic to AWS resources such as Amazon SE bucket and also allows quick calculation or processing as the input to other functions. When we connect projects with Amazon S3 organizer, and manage our data in such a way that supports specific use cases, enables cost efficiencies, enforces security and meets compliance requirements. When we connect with

a CSV file, it will store all the responses of a patient or user and use them for future reference. Here, using temporary security credentials, Amazon IAM roles offer a mechanism to access AWS. A role is not associated with a specific user or group, but rather with a set of rights for requesting AWS services. With Amazon Cognito, adding user sign-up and authentication to your mobile and online apps is a snap. With the help of Amazon Cognito, you can use AWS or any service behind Amazon API Gateway to get to the backend resources of your app. Amazon Cognito also gives you temporary security credentials.

3.1 Amazon Lex Component Design

In Amazon Lex, a bot is the main kind of resource. Intent, slot type, alias, and bot channel association are the additional resource kinds in Amazon Lex. You can develop a bot using the model-building API or the Amazon Lex console. You can create a production-ready bot for your application using the graphical user interface offered by the console. You can develop a bot using your own custom software using the model-building API through the AWS CLI if you'd like. You can either integrate a bot into your own application or deploy it on one of the supported platforms after it has been created. When a user interacts with the bot, the client application uses the Amazon Lex runtime API to make queries to the bot. For instance, your client uses one of the runtime API calls to transmit the user's input to Amazon Lex when they say, "I want to buy pizza." Users can enter information orally or in text form. Sample utterances are structured as:

<what is the condition in {city}>
<tell me {city} affection>
<Want to know about covid>

3.2 Lambda Function in AWS

Lambda functions are supported by Amazon Lex so that the bot can get code hooks. These features can accomplish a variety of tasks, including enhancing user engagement with the bot by utilising existing knowledge, validating the user-provided input data, and carrying out the user's goal.

Figure 3 is a snapshot of our lambda function. By using this lambda function, we have done all the condition checking like whether the patient has a fever or not, breathing issues or not, checking 10-digit contact details, etc. and also connected a CSV file to store all the details of the patient's physical health and contact details for future reference. For searching, add more queries, and to control these queries, add some functions that check if certain conditions are met.

```
#LAMBDA FUNCTION CONNECTED WITH A CSV FILE
import boto3
import csv

# call s3 bucket
s3 = boto3.resource('s3')
BUCKET_NAME = 'new-covidbot-bucket-user-data'
bucket = s3.Bucket(BUCKET_NAME) # Enter your bucket name, e.g 'Data'
# key path, e.g.'customer_profile/Reddit_Historical_Data.csv'
key = 'user_data.csv'

def positive():
    return "Thank you for providing the answers!! You Might have covid, no need to worries our executive will be in touch"

def negative():
    return "Thank you for providing the answers!! It seems you do not have Covid, still our executive will contact you on your

def append_data_to_s3(fname, lname, con_no, report):
    # download s3 csv file to lambda tmp folder
    local_file_name = '/tmp/user_data.csv'  #
    s3.Bucket(BUCKET_NAME).download_file(key, local_file_name)
    # list you want to append
    lists = [fname, lname, con_no, report]
    print(lists)
    # write the data into '/tmp' folder
    with open('/tmp/user_data.csv', 'r') as infile:
        reader = list(csv.reader(infile))
        reader = reader[::-1]  # the date is ascending order in file
        reader.insert(0, lists)
    with open('/tmp/user_data.csv', 'w', newline='') as outfile:
        writer = csv.writer(outfile)
        for line in reversed(reader):  # reverse order
            writer.writerow(line)
    # upload file from tmp to s3 key
    bucket.upload_file('/tmp/user_data.csv', key)
```

Figure 3. Sample Lambda function

3.3 Amazon Cognito

AWS's Cognito service makes it simple to integrate user management into online and mobile applications. It works with SAML 2.0 and provisions social identity providers like Facebook, Google, and business identity providers. This is an effective service and hard to grasp at first.

The fact that there are two main components of Amazon Cognito is one of the things that causes the most confusion:

1) User Pools for Amazon Cognito

2) Amazon Cognito Identity Pools, second (aka Federated Identities)

3.3.1 Cognito User Pool

This means an anonymous user of our application (e.g. a mobile or a Single Page Application) can fill out a registration form and then become a registered user. The chosen credentials (i.e., username and password) will be safely stored in the Cognito User Pool.

In this case, Amazon Cognito acts as an Identity Provider (IdP). When this registered user wants to log in, the User Pool will be used as the source of truth to assess the authenticity of provided credentials; if valid, a JSON Web Token (JWT) will be returned.

3.3.2 Cognito Identity Pool

Usually, REST APIs are protected through the use of a token – e.g., a JSON Web Token (JWT) –and that's why Amazon API Gateway with the help of Cognito User Pool, supports this scenario natively. Alas, the vast majority of AWS resources don't support a JWT as a means of authentication! For instance, if our application reads order item 42 directly from DynamoDB, we need an IAM Role that has permission to read data from the Orders table.

3.3.3 AWS Credentials

// Initialize the Amazon Cognito credentials provider

CognitoCachingCredentialsProvidercredentialsProvider = new CognitoCachingCredentialsProvider(getApplication-Context(),
“us-east-1:77511a8f-cb2d-488f-8cd5-233e16e3137c”, // Identity pool ID
Regions.US_EAST_1 // Region
);

3.4 Amazon S3

In order to store and access any amount of data or information from anywhere online, Amazon S3 [24] (Simple Storage Service) provides object storage. It was designed for developers to help with web-scale computing and offers 99.999999999 percent durability and 99.99 percent availability of items. Data file up to 5 terabytes in size can also be stored on it.

Figure 4 describes the web hosting of our project. Using this feature, we can create a bucket and upload a .html file for accessing the chatbot. After enabling static web hosting we can use that chatbot virtually by using a single link.

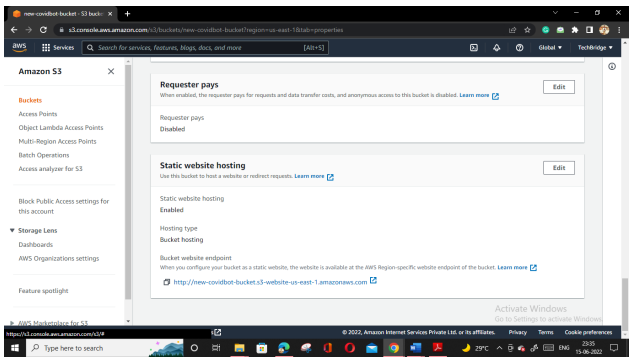


Figure 4. Static Web Hosting from S3

3.5 Data Storing Process in Amazon s3

Objects are used to store data on Amazon S3. With this method, cloud storage is greatly scalable. Different physical disc drives dispersed around a data centre can house various objects. Amazon datacentres use specialised hardware, software, and distributed file systems to offer great scalability. The block storage method implements features like redundancy and versioning. By default, when a file is kept in S3 as an item, it is simultaneously stored in several locations (such as discs, data centres, or availability zones). Control hash sums are frequently checked as part of the S3 service’s data consistency verification. If data corruption is found, redundant data is used to recover the object. S3 buckets are used to store objects. Objects in S3 storage can, by default, be viewed and handled through the web interface [24].

Object URL for error.html
https://webbucket07.s3.amazonaws.com/error.html
Object URL for index.html
https://webbucket07.s3.amazonaws.com/index.html
Bucket Policy

```
{
“Version”: “2012-10-17”,
“Statement”: [
{
“Sid”: “PublicReadGetObject”,
“Effect”: “Allow”,
“Principal”: “*”,
“Action”: “s3:GetObject”,
“Resource”: “arn:aws:s3:::webbucket07/*”
}
]
}
```

Figure 5 shows the bucket to upload HTML files like index.html for the front-end interface of the chatbot and error.html represent the error that occurs interface of the chatbot. From this terminal, we can also enable the web hosting part that helps to access our chatbot project.

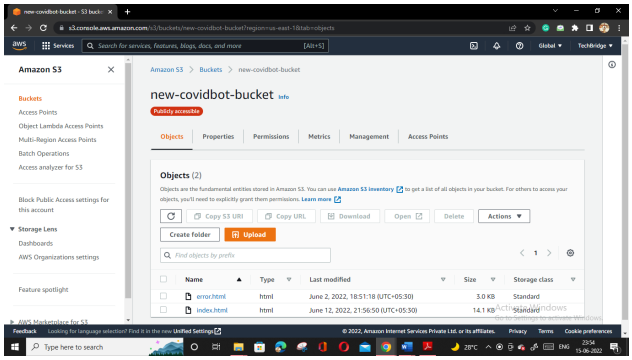


Figure 5. Upload the HTML file to the bucket

For storing data, we have used a CSV file which stores all the records of COVID feedback. Patients will give their health situation by selecting the different conditions from the auto-generated options. On the basis of given feedback, the bot will analyse if COVID has been or not. Then this CSV file will store the name of the patient, contact number, and status of COVID. Later on, the patient will get a call from an executive and discuss COVID over the phone.

Figure 6 shows the CSV which stores all the records of patients and from this given contact number we can contact that patient in the future as per need.

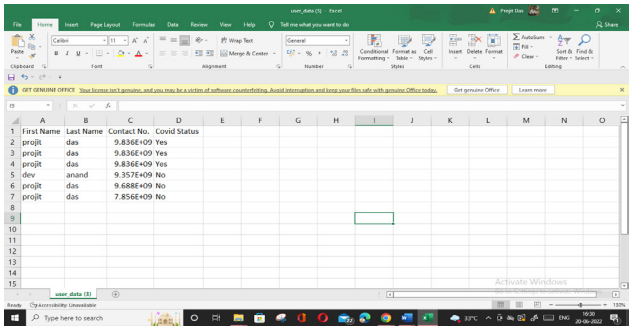


Figure 6. Store user COVID report data

3.6 Valuation of Dataset

Any chatbot should appear normal when reacting to user input, and it must have a clean dataset and long-lasting backend logic for result production. We have created a ten-question COVID-19 basic symptomatic questionnaire at the Telemedicine department at the University of Camerino. Simple Yes/No questions make up the entirety of this survey. The bot delivers a score of one if the user selected "Yes", and a score of zero if the user selected "No". The chatbot anticipates using AIML logic to respond to human input and maintain input that could feed the machine.

Table 1 lists the early symptoms that can be used by the chatbot to determine whether or not a person is infected.

Table 1. Characteristic analysis dataset

No.	Questions	Type
1	Confirm your age.	Enter an int value
2	Do you have a fever?	Yes/No
3	Do you have a cough?	Yes/No
4	Do you have a sore throat	Yes/No
5	Do you have chest congestion or a runny nose?	Yes/No
6	Do you facing any breathing issues	Yes/No
7	Do you have diabetics?	Yes/No
8	Do you have Hypertension?	Yes/No
9	Do you have lung disease?	Yes/No
10	Do you have heart disease?	Yes/No
11	Have you travelled in the past 14 days to any states?	Yes/No
12	Which city do you want to know?	Select any city

4. Result and Discussion

Initially, the chatbot successfully provided the right output in about 40% of cases. However, as the training data improved with continuous interaction with the chatbot, this accuracy improved to 70%. For queries unrelated to the intents described in the chatbot, a generalized response is generated, and such questions are logged to be checked later and included in the datasets. The Amazon S3 connected to the chatbot has successfully monitored the traffic to various sections of the chatbot by capturing the general purpose of the user. This information can be further used to analyse the kinds of people visiting the site and their purpose. Hence, more emphasis can be put on such for performance testing. Two parameters were evaluated: answer delivery delay and question matching accuracy. 1) Answer Delivery Delay: It is defined as the time from the question sent by the user until he/she receives the answer. This latency was constantly under 3 seconds, with an average of 1.76 seconds. 2) Matching Accuracy: In this, we check whether the chatbots work even when the sentences are mistyped, and the proposed system has successfully matched the sentences to the predefined utterances in 70% of the cases.

When a user gives some statements to the chatbot, it will then ask for some information from the user. Here users have to put some basic information to us for validation. The user must respond as required, after which the chatbot will ask for the user's contact information and confirm whether or not the user has COVID. Snapshots of this feature are given below in following Figures 7 and 8. Getting all the information regarding COVID will be stored in a CSV file from where an executive or doctor can call the patient.

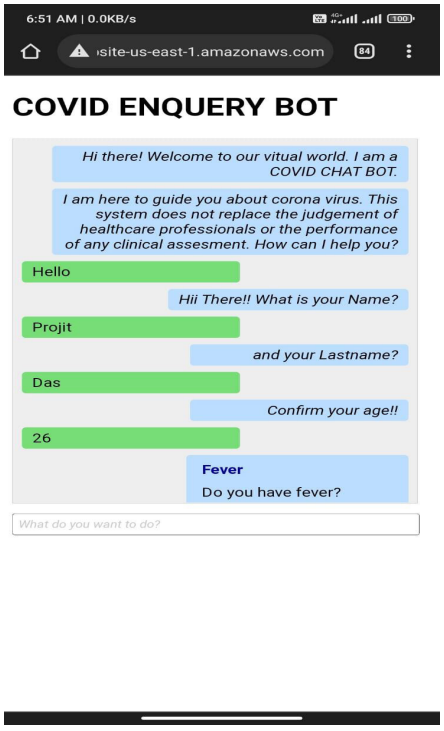


Figure 7. Response for COVID validation

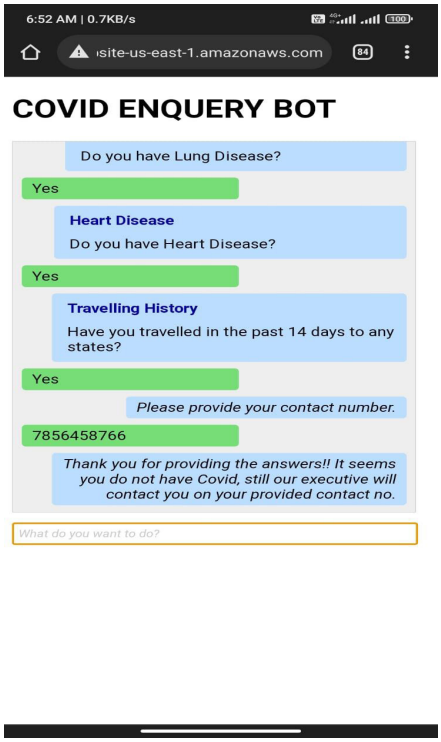


Figure 8. Response for COVID confirmation feedback

Here we have one option if a user wants to know about the information regarding COVID-19, what are the symptoms of COVID-19 and what precautions also? We can also get information regarding rumours and fake news. So, from this chatbot, users can also get information regarding that. Snapshots of this feature are given below; following Figures 9 and 10.

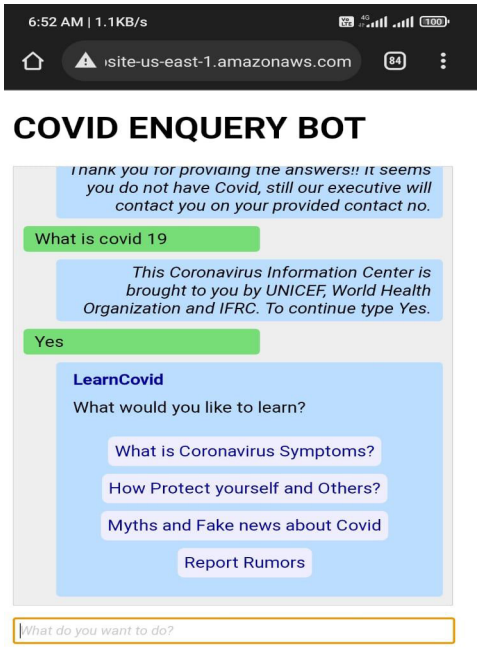


Figure 9. Response for COVID information 1

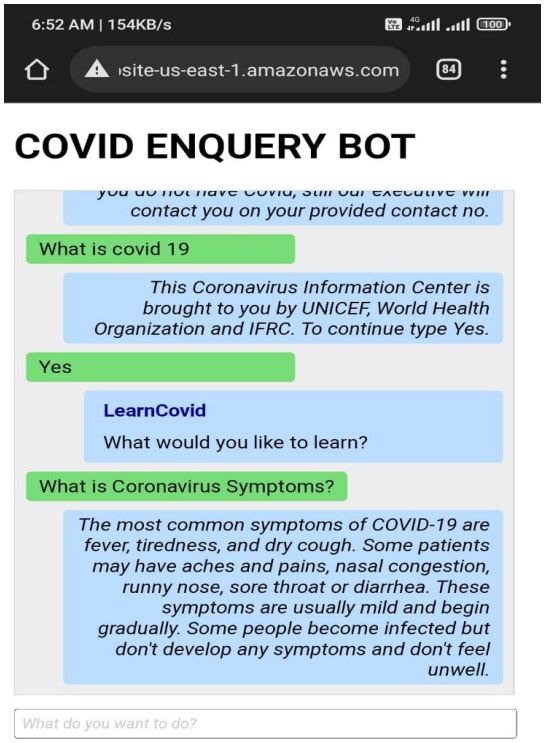


Figure 10. Response for COVID information 2

From this application, we also got some information related to the COVID situation in a different city. What are the COVID rate, and the active situation of a city we can easily get from this chatbot? Figures 11 and 12 describe this part.

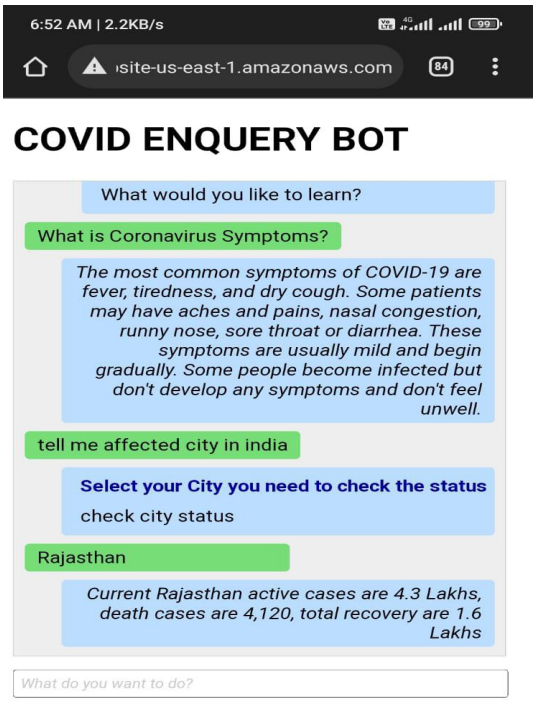


Figure 11. Choose a city to know COVID update

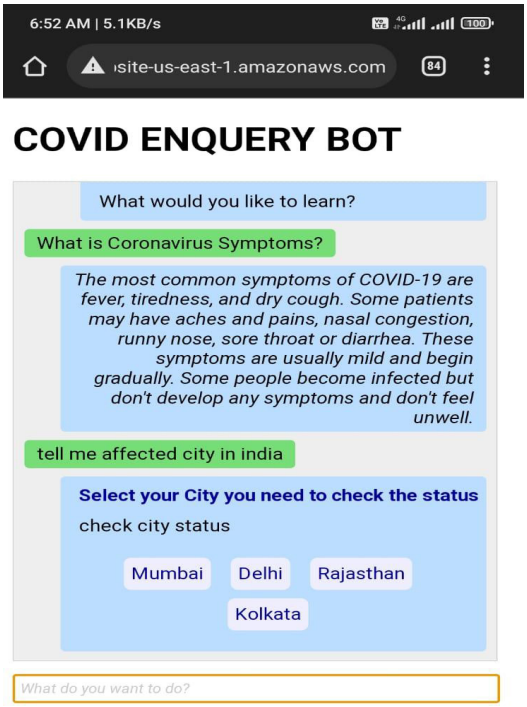


Figure 12. Response from chatbot

5. Conclusions

The design and execution of a clever chatbot application utilized to certify help and e-awareness during the COVID-19 epidemic are described in depth in this study. The purpose of this study is to develop smart chatbots that can provide consumers with information about the coronavirus during pandemics like COVID-19. A chatbot's goal is to start a real-world discussion. In our work, we suggest using the Amazon Lex, Cognito, and S3 approaches provided by Amazon Web Services in order to properly address customer inquiries. Our chatbot, CO-BOT, operates by going through three steps in succession: connecting, understanding, and replying. The user delivers a query via the user interface in plain text during the connection stage. Processing the user's question is enabled by the comprehension phase. Building the chatbot responses is the responsibility of the response stage. An English version is one of the two CO-BOT versions that are used and tested. In addition, CO-BOT can raise awareness of these disorders and offer support. This study takes an innovative approach to raise awareness of COVID-19 symptoms, preventing misunderstanding, taking precautions, providing facts and myths, reducing infection-restricting behaviours, and even assisting in mitigating the pandemic mental health burden. The key data have been compiled with the help of interviews conducted to promote this study. After reviewing the collected data, the researcher will assume that the chatbot has been rated very highly on the basis of user experience. In terms of usability, it was found functional and simple by all the respondents. The chatbot is very accurate, well-formatted, and readily accessible due to the nature of the information delivered.

6. Future Scope

With the aid of the input provided from the interviews, the potential direction of the research study can be concluded. It can be inferred after data is evaluated that additional functions can be incorporated into the system. In future work, we want to compare our application's analytical performance against that of other COVID-19 applications and chatbots. We also intend to improve our chatbot by taking into account some extended versions viz.: 1) Include a module related to speech recognition; 2) Offer a virtual doctor meeting and best practices for a healthy lifestyle; and 3) Improve COVIBOT's data interpretation by incorporating features like graphical and pictorial representation. 4) Link a real-time database to obtain the most recent COVID status.

Conflict of Interests

The authors declare that they have no conflict of interests.

References

[1] Gates, B., 2020. Responding to Covid-19-A once-in-a-century pandemic? The New England Journal of Medicine. 382(18), 1677-1679.

[2] Sohrabi, C., Alsa, Z., O'Neill, N., et al., 2020. World health organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). International Journal of Surgery. 76, 71-76.

[3] Thorp, H.H., 2020. A dangerous rush for vaccines. Science. 369(6506), 885.

[4] Nehme, M., Braillard, O., Alcoba, G., et al., 2020. COVID-19 symptoms: Longitudinal evolution and persistence in outpatient settings. Annals of Internal Medicine. 174(5), 723-725.

[5] Driss, M., Almomani, I., Alahmadi, L., et al., 2022. COVIBOT: A Smart Chatbot for Assistance and E-Awareness during COVID-19 Pandemic. SMARTTECH 2022: The Second International Conference of Smart Systems and Emerging Technologies; Apr 2022; Saudi Arabia. DOI:10.48550/arXiv.2204.07851.

[6] Bharti, U., Bajaj, D., Batra, H., et al., 2020. Medbot: Conversational artificial intelligence-powered chatbot for delivering telehealth after covid-19. 5th Interna-

tional Conference on Communication and Electronics Systems (ICCES); 2020 Jun 10-12 ; Coimbatore, India.USA: IEEE. pp. 870-875. DOI:10.1109/ICCES48766.2020.9137944.

[7] Ferreira, T.C., Marcolino, M.S., Ramos, I., et al., 2020. ANA: A Brazilian Chatbot Assistant about COVID-19. Proceedings of Conference ACL 2020 Workshop on Natural Language Processing for COVID-19 (NLP-COVID); 2020 Jul; Seattle, Washington, United States.

[8] Dennis, A.R., Kim, A., Rahimi, M., et al., 2020. User reactions to COVID-19 screening chatbots from reputable providers. Journal of the American Medical Informatics Association. 27(11), 1727-1731.

[9] Tiwari, V., Verma, L.K., Sharma, P., et al., 2021. Neural network and NLP based chatbot for answering COVID-19 queries. International Journal of Intelligent Engineering Informatics. 9(2), 161-175.

[10] Abd-Alrazaq, A.A., Rababeh, A., Alajlani, M., et al., 2020. Effectiveness and safety of using chatbots to improve mental health: systematic review and meta-analysis. Journal of Medical Internet Research. 22(7), e6021.

[11] Abd-Alrazaq, A.A., Alajlani, M., Ali, N., et al., 2020. Perceptions and opinions of patients about mental health chatbots: Scoping review. Journal of Medical Internet Research. 23(1), e17828.

[12] Li, B., Wisniewski, G., Crabbé, B., 2021. Are Transformers a Modern Version of ELIZA? Observations on French Object Verb Agreement. Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing; 2021 Nov; Punta Cana, Dominican Republic. USA: Association for Computational Linguistics. p.4599-4610. DOI: 10.18653/v1/2021.emnlp-main.377.

[13] Colby, K.M., 1981. Modeling a paranoid mind. Behavioral and Brain Sciences. 4(4), 515-534.

[14] Raij, A.B., Johnsen, K., Dickerson, R.F., et al., 2007. Comparing interpersonal interactions with a virtual human to those with a real human. IEEE Transactions on Visualization and Computer Graphics. 13(3), 443-457.

[15] Fadhil, A., 2018. Beyond patient monitoring: Conversational agents role in telemedicine & healthcare support for home-living elderly individuals [Internet] [Submitted on 2018 Mar 3]. Available from: https://arxiv.org/abs/1803.06000.

[16] Amato, F., Marrone, S., Moscato, V., et al., 2017. Chatbots Meet eHealth: Automatizing Healthcare. WAIAH@ AI* IA. Computer Science. 40-49.

[17] Comendador, B.E.V., Francisco, B.M.B., Medenilla, J.S., et al., 2015. Pharmabot: A pediatric generic medicine consultant chatbot. Journal of Automation and Control Engineering. 3(2).

[18] Josephine, D.D., 2021. Amazon Lex based personalized chatbot for IoT LAB@CIT. International Journal of Advance Scientific Research and Engineering Trends. 6(5).

[19] Armstrong, S., 2018. The apps attempting to transfer NHS 111 online. Technology in Healthcare. DOI: https://doi.org/10.1136/bmj.k156.

[20] Holmes, S., Moorhead, A., Bond, R., et al., 2019. Usability testing of a healthcare chatbot: Can we use conventional methods to assess conversational user interfaces? Proceedings of the 31st European Conference on Cognitive Ergonomics; 2019 Sep 10. p. 207-214. DOI: https://doi.org/10.1145/3335082.3335094.

[21] Gupta, J., Singh, V., Kumar, I., 2021. Florence-a health care chatbot. 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS) ;2021 Mar 19-20; Coimbatore, India. USA: IEEE. p. 504-509.

[22] Rogers, R., 2021. Internet of things-based smart healthcare systems, wireless connected devices, and body sensor networks in covid-19 remote patient monitoring. American Journal of Medical Research. 8(1),71-81.

[23] Waheed, N., Ikram, M., Hashmi, S.S., et al., 2022. An Empirical Assessment of Security and Privacy Risks of Web based-Chatbots [Internet] [Submitted on 2020 May 17]. Available from: https://arxiv.org/abs/2205.08252.

[24] Bornholt, J., Joshi, R., Astrauskas, V., et al., 2021. Using lightweight formal methods to validate a key-value storage node in Amazon S3. Proceedings of the ACM SIGOPS 28th Symposium on Operating Systems Principles; 2021 Oct. pp. 836-850. DOI: https://doi.org/10.1145/3477132.3483540.

BILINGUAL
PUBLISHING CO.
Pioneer of Global Academics Since 1984

Tel.:+65 65881289
E-mail: contact@bilpublishing.com
Website:ojs.bilpublishing.com

2630-5151 04
9 772630 515223