Cairo University
The 55th Annual International Conference of Data Science

COMPUTER SCIENCE

5-7, Dec ,2022

Index
COMPUTER SCIENCE

Shadow Set: A Brief Overview

Abdelrahman Arab
Faculty of Graduate Studies for
Statistical Research
Cairo University
Cairo, Egypt

Hesham A. Hefny
Faculty of Graduate Studies for
Statistical Research
Cairo University
Cairo, Egypt

Abstract — **Despite the major contribution of the fuzzy set in representing many phenomena in our world characterized by different levels of uncertainty, it suffers from such limitation of excessive precision, where the membership function describes vague phenomena with crisp numbers, the thing that prompted Zadeh, its inventor, to search for solutions through modifying the membership function, however, these solutions added additional level of computational complexity that made them practically highly inapplicable. The shadow set, introduced by Pedrycz, is induced by the fuzzy set as a solution to lower this numerical preciseness of the fuzzy set while preserving, at the same time, the total amount of uncertainty. This paper provides a brief overview of the shadow set covering its definition, the way it is being induced from the fuzzy set, recent modifications regarding the choice of the threshold value, and the operations carried on it, also some application domains have been introduced.**

Keywords—Fuzzy set; shadow set; uncertainty; granular computing

I. INTRODUCTION

Fuzzy sets, introduced by **Zadeh**, succeeded to a great extent in representing diverse vague phenomena. This is entirely done through the usage of a membership function that maps elements from the universe of discourse into a defined concept. Each item has a precise numerical membership value between 0 and 1 indicating to which extent does it belong to the specified concept. An issue arose here concerning the description of a vague phenomenon with these crisp numbers which may be considered an excessive precision [3,14].

This limitation of the fuzzy set is explained by the Canadian professor, **Witold Pedrycz**, who stated that "*Once the membership function has been established (estimated or defined), the concept is described very precisely as the membership values are exact numerical quantities. This seems to raise a certain dilemma of excessive precision in describing imprecise phenomena. In fact, this concern has already sparked a lot of debates starting from the very inception of fuzzy sets*" [16]. As per **Zadeh**, determining the membership function is the biggest challenge for using fuzzy sets, thus trying to handle the excessive precision limitation, he managed to change the membership function to introduce, in 1971, the concept of type-2 fuzzy set as an extension of type-1 fuzzy set. In type-2 fuzzy sets, the membership function is itself a type-1 fuzzy set having its value as a grade of numbers instead of a crisp one, in this case, type-1 fuzzy set is considered a special case of type-2 fuzzy set. However, exploiting type-2 fuzzy sets in applications arose an additional level of complexity which gave way for another extension using the interval-valued fuzzy sets, where the value of the membership function is a sub-interval of [0,1]. Interval-valued fuzzy set, first introduced in 1970, is considered a particular case of type-2 fuzzy set and it took a decade until **Gorzalczany** and **Türksen** came out in 1980 with a complete name and definition of it [8,10, 16].

The shadow set, as well as the fuzzy set, is a particular case of the interval-valued fuzzy set that maps the elements of the universe of discourse into 0,1 and the interval [0,1]. It provides an effective way of dealing with the fuzzy set limitations where it avoids the computation complexity while preserving the uncertainty of the fuzzy set, as **Pedrycz** stated "*shadowed sets help reduce a computational overhead that is associated with the use of fuzzy sets*" [15], it is capable of dealing with real-world problems and interpreting them with more flexibility, for if there is a fuzzy set with millions of elements, inducing a shadow set from that fuzzy set will reduce the complexity by distributing these elements upon three disjoint regions without losing their semantic interpretation [5, 8].

The paper is organized into six sections other than this one, the next section discusses the shadow sets and their induction by the fuzzy set, also some variations of the shadow set are introduced together with different techniques of determining the threshold values. The third section presents some important operations and their properties, the fourth section discusses some of the shadow set applications. The fifth section shed some light on a new approach to a shadow set with a higher approximation degree, while the sixth section states the current challenges facing the shadow set, followed by the final section which is the conclusion.

II. SHADOW SETS

To simplify the computational complexity and the excessive precision introduced by fuzzy sets, **Pedrycz** introduced shadow sets, in 1998, as a three-valued approximation of fuzzy sets. The shadow set, as per **Pedrycz**, "*does not lend itself to precise numerical membership values but relies on basic concepts of truth values (yes-no) and an entire unit interval perceived as a zone of uncertainty*" [16]. Shadow sets use two operations which are; the elevation operation which elevates the membership degree of an element, and the reduction operation which reduces the membership degree of an element. Membership degrees close to the full membership degree (one) are elevated to one, while membership degrees close to zero are reduced to zero. This radical precession from zero to one reduces the vagueness (uncertainty) of the concept being described by the original fuzzy set, thus, in order to restore the vagueness balance, membership degrees that are neither close to one nor close to zero are represented by a third level of approximation around 0.5, where the value 0.5 denotes the maximum uncertainty [11,14,16].

Definition 1 [8,10]:

A shadow set S is defined on the universe of discourse U as:

$$S: U \rightarrow \{0,1,[0,1]\} \tag{1}$$

Each element in U can fall into one of three regions or subsets of U which are defined as follows:

$$\text{Core (S)} = \{x \in U | S(x) = 1\} \tag{2}$$

$$\text{Exclusion (S)} = \{x \in U | S(x) = 0\} \tag{3}$$

$$\text{Shadow (S)} = \{x \in U | S(x) = [0,1]\} \tag{4}$$

Where "core" means that the element is completely included into S, "exclusion" means that the element is completely excluded from S, and "shadow" means that the element's membership to S is completely unknown.

Although the shadow set, as mentioned before, can be considered as a particular case of the interval-valued fuzzy set, they differ from each other in the sense that the shadow set can be induced by a fuzzy set while the interval-valued fuzzy set cannot, this is mainly because the interval-valued fuzzy set is introduced away from the fuzzy set, the fuzzy set is itself a particular case of the interval-valued fuzzy set [8, 16].

A. From fuzzy sets to shadow sets

Shadow sets can be directly induced by fuzzy sets, this is done by determining two threshold values, α (of an α-cut) where $\alpha \in [0,0.5]$, and β where $\beta > \alpha$ and $\beta \in [0.5,1]$, these two values determine the three disjoint regions of the shadow set as shown in Fig. 1. The core region contains elements whose membership values are more than β, the exclusion region contains elements whose membership values are less than α, and the shadow region contains elements whose membership values are between α and β [3,9].

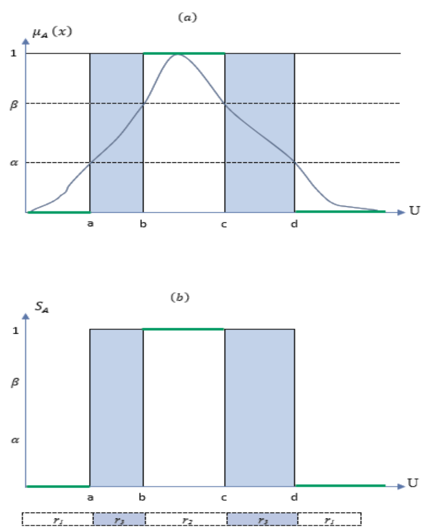


Fig. 1. (a) the fuzzy set a has been divided into three regions. (b) the shadow set S as induced by the fuzzy set A

Definition 2 [6]:

Let A be a fuzzy set, then the shadow set S_A induced by A is defined as:

$$S_A(x) = \begin{Bmatrix} 0 & \mu_A(x) < \alpha \\ 1 & \mu_A(x) > \beta \\ [0,1] & \alpha \le \mu_A(x) \le \beta \end{Bmatrix} \tag{5}$$

Inducing a shadow set by a fuzzy set is carried out according to the principle of uncertainty invariance, proposed in 1987 by **Klir** who stated that "*The principle requires that the amount of uncertainty (and information) be preserved when a representation of uncertainty in one mathematical theory is transformed into its couterpart in another theory*" [17]. However, preserving the whole amount of the uncertainty present in the fuzzy set in the induced shadow set is practically very difficult. The memberships of the elements in the exclusion and the core regions have the values 0 and 1 respectively, this upsets the uncertainty balance where there is an elimination of the uncertainty for the elements belonging to the two regions, they are either fully rejected or fully accepted and the uncertainty situation is gone. Therefore, the uncertainty of the shadow region should be handled in order to restore that uncertainty balance by selecting the optimal threshold values [2, 5, 6, 17].

Fortunately, the uncertainty balance can be explained by the fact that: there is a relation between the size of the shadow region and the amount of uncertainty, the balance between the shadow region and the other two regions (the core and the exclusion regions) can then affect the uncertainty preservation process where any decrease in the two regions increases the size of the shadow region and consequently an increase in the amount of uncertainty preserved [2, 9, 16].

B. Determining the optimal threshold values

Many scientists, including **Pedrycz**, tried to preserve the maximum amount of uncertainty in the induced shadow set by achieving an uncertainty balance between the three regions through the choice of the optimal threshold values, which is the most important task in inducing a shadow set by a fuzzy one, this choice has to be made such that the regions under the membership function are balanced and the uncertainty of the original fuzzy set is fully transferred to the shadow region of the shadow set which is obtained, as per **Pedrycz**, by (6) in case of an infinite discourse of the universe, and (7) in case of a discrete discourse of the universe, where $\mu_A(x)$ is the membership function of the element *x* to the fuzzy set *A*, α and β are the two threshold values, and *V* is the vagueness balance [2,16].

$$V = \left| \int_{-\infty}^{\alpha} \mu_A(x)\,dx + \int_{\beta}^{+\infty} \left(1 - \mu_A(x)\right)dx - \int_{\alpha}^{\beta} \mu_A(x)dx \right| \tag{6}$$

$$V = \left| \sum_{i:\mu_A(x_i) < \propto} \mu_A(x_i) + \sum_{i:\mu_A(x_i) > \beta} \left(1 - \mu_A(x_i)\right) - card\,\{x_i \in U | \alpha \le \mu_A(x_i) \le \beta\} \right| \tag{7}$$

$$V = |uncertainty\ of\ exclusion\ region + uncertainty\ of\ core\ region - uncertainty\ of\ shadowed\ region| \tag{8}$$

Equation (8) can be rewritten as in (9) where $v(r_1)$ is the vagueness of the exclusion region, $v(r_2)$ is the vagueness of the core region, and $v(r_3)$ is the vagueness of the shadow region [9, 16].

$$V = |v(r_1) + v(r_2) - v(r_3)| \tag{9}$$

The balance of vagueness can then be achieved if the three regions are balanced, meaning that the uncertainty of the shadow region is the sum of the uncertainty of both the exclusion and the core regions as shown in (10).

$$v(r_3) = v(r_1) + v(r_2) \tag{10}$$

For simplicity, the two thresholds can be obtained such that $\alpha + \beta = 1$, then $\beta = 1 - \alpha$, only one threshold is obtained in this case which is α then the higher threshold will be accordingly calculated.

The choice of α can then be regarded as an optimization problem (11), where the optimal α is obtained by achieving the most possible balance between the three regions, that is by having V in (9) as close as possible to zero.

$$\alpha(opt) = arg(min_{\alpha}|v(r_1) + v(r_2) - v(r_3)|) \quad (11)$$

Equation (11) can be simplified by (12) as follows

$$\min_{\alpha\in(0,0.5)} V_{(\alpha)} = V_{(\alpha_{opt})} \quad (12)$$

Ibrahim and **William-West** in [3] changed the cardinality of the elements in the shadow region in (7) with the summation of the membership values of these elements as in (13).

$$V = \left| \sum_{i:\mu_A(x_i)<\alpha} \mu_A(x_i) + \sum_{i:\mu_A(x_i)>\beta} \left(1 - \mu_A(x_i)\right) - \sum_{i:\alpha\le\mu_A(x_i)\le\beta} \mu_A(x_i) \right| \quad (13)$$

They claimed that if $\alpha_i \in [\mu_{min}, 0.5]$ represents all the possible thresholds in that period (where μ_{min} is the minimum membership value in the fuzzy set), then it is possible that by taking the average of the uncertainty balance (V) the optimal α can be readily obtained as in (14), where i=1, 2,..., k and k is the number of elements in the shadow region, hence (11) is changed to (15). The threshold obtained this way is supposed to overcome the limitations of the threshold obtained by **Pedrycz**, which they argue that it does not retain the amount of uncertainty preserved in the original fuzzy set [2, 3].

$$avg_{\alpha}V(\alpha) = \frac{\sum_{i=1}^{k} V(\alpha_i)}{k} \quad (14)$$

$$\alpha(opt) = arg\left(avg_{\alpha}V(\alpha)\right) \quad (15)$$

Deng and **Yao** in [11] provided equations for calculating the values of the two thresholds α and β through their proposed decision-theoretic shadow set approach. They propose that if there is an element $x \in U$, it has a possible action decisions $a \in Actions$ where $Actions = \{a_e, a_r, a_{S_\uparrow}, a_{S_\downarrow}\}$, and as per the decision theory, every action has a decision error E_a and a decision cost (loss) λ_a. These values are calculated as per Table 1, where δ is a constant replacing the interval [0,1].

TABLE 1. DECISION ERROR AND COST (LOSS) [11]

Actions	Fuzzy set	Shadowed set	Errors	Loss/ cost
a_e	$\mu_A(x)$	1	$1-\mu_A(x)$	λ_e
a_r	$\mu_A(x)$	0	$\mu_A(x)-0$	λ_r
$a_{s\downarrow}$	$\mu_A(x)\ge\delta$	δ	$\mu_A(x)-\delta$	$\lambda_{s\downarrow}$
$a_{s\uparrow}$	$\mu_A(x)<\delta$	δ	$\delta-\mu_A(x)$	$\lambda_{s\uparrow}$

Then the two thresholds α and β are calculated as follows:

$$\alpha = \frac{\lambda_e + \delta\lambda_{S_\downarrow}}{\lambda_e + \lambda_{S_\downarrow}} \quad (16)$$

$$\beta = \frac{\delta\lambda_{S_\uparrow}}{\lambda_r + \lambda_{S_\uparrow}} \quad (17)$$

Ibrahim et al. in [2] built their proposal based on the fact that: the amount of uncertainty present in the fuzzy set is distributed among the three disjoint regions of the fuzzy set and is equivalent to the cardinality of the shadow set, this fact implies the following equation:

$$V(\alpha) = \left| \sum_{x\in core(S)} \mu_{\varphi_A}(x) + \sum_{x\in exclusion(S)} \mu_{\varphi_A}(x) + \sum_{x\in shadow(S)} \mu_{\varphi_A}(x) - Card\ (shadow(S)) \right| \quad (18)$$

Where φ_A is the uncertainty of the fuzzy set A, S is the shadow sets. Then the optimal threshold is given by:

$$\alpha(opt) = avgmin_{\alpha} V(\alpha) \tag{19}$$

C. Variations of the shadow set

In his paper [15] published in 2002, **Pedrycz** provided a generalized definition of the shadow set in order to increase the range of the applications that the shadow set can serve. He replaced the interval [0,1] with a construct that can be many other things rather than an interval and denoted it by the letter (a), (1) is therefore changed to the following one:

$$A: U \rightarrow \{0,1,a\} \tag{20}$$

Cattaneo and **Ciucci** in [14] replaced the interval [0,1] in the shadow set definition with a constant value 0.5 (Fig. 2) denoting the maximum uncertainty between the complete truth of 1 and the complete falseness of 0, and stated "*This is only a change from a syntactical point of view, without any loss from the semantic point of view*". Then the definition of the shadow set S_A in (5) will be changed to:

$$S_A(x) = \begin{Bmatrix} 0 & \mu_A(x) < \alpha \\ 1 & \mu_A(x) > \beta \\ 0.5 & \alpha \leq \mu_A(x) \leq \beta \end{Bmatrix} \tag{21}$$

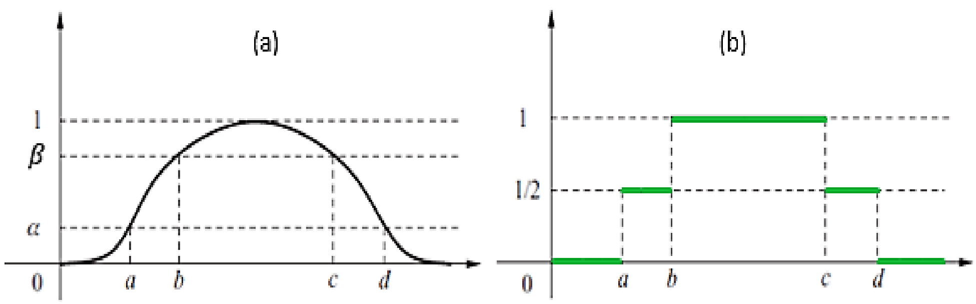


Fig. 2. Inducing a shadow set (b) from a fuzzy set (a) by Cattaneo and Ciucci [14]

Deng and **Yao** in [11] suggested a generalized definition where they replaced the interval [0,1] with a constant value δ where $0 < \delta < 1$ giving rise to the following definition:

$$S_A(x) = \begin{Bmatrix} 0 & \mu_A(x) < \alpha \\ 1 & \mu_A(x) > \beta \\ \delta & \alpha \leq \mu_A(x) \leq \beta \end{Bmatrix} \tag{22}$$

As per **Deng** and **Yao** "*The value 0.5 may be a good approximation of some objects, but may not be a good approximation of other objects*", they calculated δ as the mean value of the membership grade:

$$\delta = \frac{\sum_{x \in S_A} \mu_A(x)}{card\ (S_A)} \tag{23}$$

The membership value of each element in the shadow region can be either elevated or reduced to δ as can be shown in Fig. 3. The δ value is taking different values based on the application at hand.

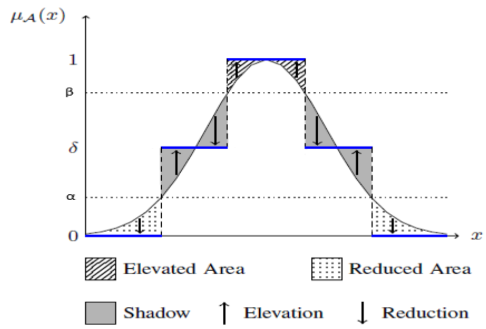


Fig. 3. Shadow set defined by the mean value δ proposed by Deng and Yao [11]

Zhang et al. in [4] replaced the interval [0,1] with the interval [α, β] in their proposed new model which they called interval shadow sets, that model aimed to reduce the fuzzy entropy loss that they found to be large between the original fuzzy set and the induced shadow set. The interval shadow set has the following definition:

$$S_A(x) = \begin{Bmatrix} 0 & \mu_A(x) < \alpha \\ 1 & \mu_A(x) > \beta \\ [\alpha, \beta] & \alpha \le \mu_A(x) \le \beta \end{Bmatrix} \tag{24}$$

Dynamic fuzzy sets are those sets varying with time where the membership $\mu_A(x, t)$ is a function of two elements $x \in U$ and $t \in T$, unlike the normal static fuzzy sets where the membership $\mu_A(x)$ is a function of only one element $x \in U$. **Cai et al.** in [6] produced a definition of a dynamic shadow set as follows:

$$S_A(x,t) = \begin{Bmatrix} 0 & \mu_A(x,t) < \alpha(t) \\ 1 & \mu_A(xx,t) > \beta(t) \\ [0,1] & \alpha(t) \le \mu_A(x,t) \le \beta(t) \end{Bmatrix} \tag{25}$$

They followed the steps of Deng and Yao in calculating the threshold values $\alpha(t)$ and $\beta(t)$ based on the loss values having δ = 0.5.

III. OPERATIONS ON SHADOW SETS [15, 16]

Some of the operations carried on shadow sets are represented in this section which are union, intersection, and complement, they are found to be consistent with logic, especially the three-valued logic.

Union: $A \cup B$

Shadow set

	0	1	[0,1]
0	0	1	[0,1]
1	1	1	1
[0,1]	[0,1]	1	[0,1]

Fuzzy set

	0	1
0	0	1
1	1	1

Intersection: $A \cap B$

Shadow set

	0	1	[0,1]
0	0	0	0
1	0	1	[0,1]
[0,1]	0	[0,1]	[0,1]

Fuzzy set

	0	1
0	0	1
1	1	1

Complement: $\overline{A}$

Shadow set

0	1
1	0
[0,1]	[0,1]

Fuzzy set

0	1
1	0

Notes:

$0 \cup [0,1] = [0,1]$

$1 \cap [0,1] = [0,1]$

For $a \in [0,1]$ there is:

$\min\ ([0,1], a) = [0, a]$

$\max\ ([0,1], a) = [a, 1]$

A. Operation properties [16]

Commutativity:

$$A \cup B = B \cup A$$
$$A \cap B = B \cap A$$

Distributivity:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

I

Associativity:

$$A \cup (B \cup C) = (A \cup B) \cup C = A \cup B \cup C$$
$$A \cap (B \cap C) = (A \cap B) \cap C = A \cap B \cap C$$

Boundary conditions:

$$A \cup \emptyset = A \qquad A \cup X = X$$
$$A \cap \emptyset = \emptyset \qquad A \cap X = A$$

Idempotency:

$$A \cup A = A$$
$$A \cap A = A$$

Involution:

$$\bar{\bar{A}} = A$$

B. T-norm and S-norm (T-conorm) operators [1, 10, 14]

The function $T(x, y)$ mapping $[0,1] \times [0,1] \rightarrow [0,1]$, is a T-norm if it is commutative, associative, monotone (non-decreasing), and satisfies a boundary condition where 1 is the neutral (identity) element, then for any $x \in [0,1]$, $T(0,0) = 0, T(1,1) = 1, and\ T(x, 1) = T(1, x) = x$. The dual operator of T-norm is T-conorm denoted as S-norm, a function $S(x, y)$ mapping $[0,1] \times [0,1] \rightarrow [0,1]$, is a S-norm if it satisfies the previous four conditions with the only difference in the neutral element which will be 0 instead of 1, then for any $x \in [0,1]$, $S(0,0) = 0, S(1,1) = 1, and\ S(x, 0) = S(0, x) = x$.

There are many T-norm and S-norm operators, two of the most commonly used of them are the minimum T-norm and the drastic T-norm operators, and their corresponding duals, maximum S-norm and drastic S-norm operators respectively. The **Gödel** minimum T-norm and maximum S-norm operators are given by the following functions: $T(x, y) = min\ \{x, y\}$, and $S(x, y) = max\ \{x, y\}$, they are represented by the following tabular form:

Min T-norm

	0	1	[0,1]
0	0	0	0
1	0	1	[0,1]
[0,1]	0	[0,1]	[0,1]

Max S-norm

	0	1	[0,1]
0	0	1	[0,1]
1	1	1	1
[0,1]	[0,1]	1	[0,1]

While the drastic T-norm and S-norm are given by the following functions:

$$T(x,y) = \begin{Bmatrix} x & if\ y = 1 \\ y & if\ x = 1 \\ 0 & otherwise \end{Bmatrix}, \qquad S(x,y) = \begin{Bmatrix} x & if\ y = 0 \\ y & if\ x = 0 \\ 1 & otherwise \end{Bmatrix}$$

And are represented by the following tabular form:

Drastic T-norm

	0	1	[0,1]
0	0	0	0
1	0	1	[0,1]
[0,1]	0	[0,1]	0

Drastic S-norm

	0	1	[0,1]
0	0	1	[0,1]
1	1	1	1
[0,1]	[0,1]	1	1

Also, **Łukasiewicz** T-norm operator is applied on shadow sets which is given by the function: $T(x,y) = max\,\{0, x+y-1\}$.

C. Shadowed relation [13]

The shadow set relation R between elements in two different finite universes of discourse X and Y is a mapping of their cartesian product:

$$R: X \times Y \rightarrow \big[0,1,[0,1]\big] \tag{26}$$

Given the shadow set X, then a max-min composition with R ($X \circ R$) is possible giving rise to a new shadow set Y as in the following equation where j = 1.2…., m, then card (X) = n and card (Y) = m

$$Y(y_i) = \max_{i=1,2,\ldots,n}\Big[\min\Big(X(x_i), R\big(x_i, y_j\big)\Big)\Big] \tag{27}$$

The implication operator ($A \rightarrow B$), is used to solve many problems related to the shadowed relational systems, it represents the implication relation between two shadowed fuzzy propositions: $R(x,y)\colon if\ A(x)\ then\ B(y)$, or simply written as $R(x,y)\colon A(x)\ \rightarrow B(y)$, The implication operator for a shadow set is defined as follows:

A \ B	0	1	[0,1]
0	1	1	1
1	0	1	[0,1]
[0,1]	0	1	[0,1]

IV. APPLICATIONS

A. Image processing [16]

Image processing methodology involves a feature extraction algorithm aiming to extract important features from the image like the edges or the boundaries of an object inside the image. The extraction process is based mainly upon the difference in the color intensities between the pixels of the image. Thinking of the boundary as separating two different regions of the image while constructing a third region, which is the boundary itself, then shadow sets can help determining such a boundary. The shadow region of the shadow set will represent the boundary while the other two regions surrounding the boundary are represented by the core and the exclusion regions of the shadow set.

B. Image segmentation using shadow C-means (SCM) clustering

Image segmentation is the process of dividing the image into segments (clusters) having a semantic meaning. Fuzzy C-means clustering (FCM), is one of the most important techniques used in fuzzy clustering (fuzzy clustering is also known as soft clustering compared to the hard clustering technique which is k-means clustering) and is normally used for image segmentation. The number of clusters is initially determined, then the center (centroid) of each cluster is calculated. The distance between each pixel and the center of each cluster is calculated to determine the membership degree of that pixel to each cluster. Each pixel belongs then to each one of the available clusters with a certain membership degree and is assigned to the cluster associated with the highest membership degree [7,16].

A matrix is constructed for each cluster representing the fuzzy relation between the pixels in the image and that cluster, each element in the matrix represents the membership degree μ_{ik} of the specified pixel to the cluster where i is the pixel number and k is the cluster number, the membership value is calculated as follows:

$$\mu_{ik} = \frac{1}{\sum_{j=1}^{c}\left(\frac{\|x_i - v_k\|^2}{\|x_i - v_j\|^2}\right)^{\frac{2}{m-1}}} \tag{28}$$

Shadow C-means (SCM) is used as an enhancement of FCM, in which each cluster is represented by a shadow set, therefore each cluster is typically divided into three regions. The membership values in the matrix are hence distributed among the three regions of the shadow set based on the threshold value $\alpha \in [0,1]$ which is determined by the following equation [2, 16]:

$$V = \left| \sum_{i,k:\mu_{ik}<\alpha} \mu_{ik} + \sum_{i,k:\mu_{ik}>1-\alpha} (1-\mu_{ik}) - card\,\{(i,k)|\alpha \le \mu_{ik} \le 1-\alpha\} \right| \tag{29}$$

The pixels are given weights based on their assigned region and the centroid of the cluster is calculated again to start the new iteration until the maximum number of iterations is reached or the improvement between the last two iterations is smaller than a specified value. The pixels in the core region are not assigned any minimizing weight since they are supposed to have the main impact on the centroid calculations, the pixels in the shadow region are assigned a minimizing weight that minimizes their impact on the centroid calculations, and the pixels in the exclusion region are assigned a higher minimizing weight than that assigned to the shadow region, as they are supposed to have the least impact on the centroid calculations, therefore the noisy effect of the outliers, which are the noise pixels with abnormal values, is being minimized [2, 7].

In his interpretation of the fuzzy clustering using shadow sets, **Pedrycz** described the data structures created by the clustering technique as a hierarchy of structures which is illustrated in Fig. 4. The most important data is in the core structure then moving outwards to the less important ones in the shadow structure and ending with the uncertain (exclusion) structure which contains the least important data [13].

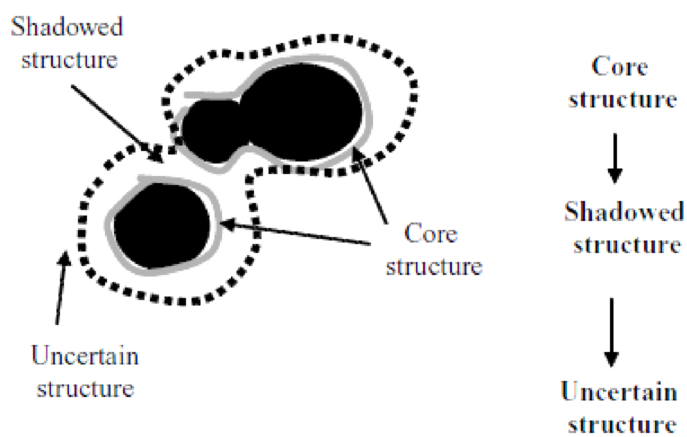


Fig. 4. Illustration of the hierarchy of clustered data structures [13]

C. Decision-Making [16]

The decision, in a decision-making problem, is normally obtained by treating each one of the decision objectives as a fuzzy set and taking the intersection of all of these fuzzy sets forming the decision fuzzy set. Considering the decision-making problem with two decision objectives represented by the two shadow sets A and B, then the decision is made by taking the intersection D where $D = A \cap B$, then we have, as **Pedrycz** explained in [16], three possible solutions:

1- No decision can be made, where $D = \emptyset$.

2- The risky or critical decision, where although there is an intersection, but D does not contain a core region.
3- The decision is clear, where D contains a core region.

D. Granular computing

Granular computing (GrC) is an approach aiming to divide complicated problems into simpler and easier-to-be-analyzed subproblems. This requires dividing information into information granules of similar characteristics, which may be considered as the information building blocks leading to a clear interpretation of that information, removing therefore the computation complications from complex systems [9, 14].

The shadow set is among several techniques used in granular computing like the rough set and the random set, as it can represent the uncertainty present in any information granule and preserve it until the problem is solved and the decision is made, and it can play this role regardless of the nature of the information granular structure used. Shadow sets can be used in various applications in this regard including fuzzy modeling and data quantization processes since it gives the ability to create new regions to deal with the uncertainty associated with such processes [5, 15].

V. SHADOW SET WITH HIGHER NUMBER OF APPROXIMATION REGIONS [2]

A proposal of inducing a shadow set by a fuzzy set through a five-way approximation was introduced by **Ibrahim et al.** in [2], increasing the number of regions in the shadow set (S) is supposed to enhance the uncertainty amount. To do that, four threshold values have to be determined which are α, β, γ, and ρ, these four thresholds will divide the shadow set into five regions by adding two regions to the original three ones, these two regions are the semi-core region and the semi-exclusion region. Fig. 5, shows the distribution of the five regions where:

$$\text{Core (S)} = \{x \in A: \mu_A(x) \in [\rho, 1]\} \quad (30)$$
$$\text{Semi-core (S)} = \{x \in A: \mu_A(x) \in [\gamma, \rho]\} \quad (31)$$
$$\text{Shadow (S)} = \{x \in A: \mu_A(x) \in [\beta, \gamma]\} \quad (32)$$
$$\text{Semi-exclusion (S)} = \{x \in A: \mu_A(x) \in [\alpha, \beta]\} \quad (33)$$
$$\text{Exclusion (S)} = \{x \in A: \mu_A(x) \in [0, \alpha]\} \quad (34)$$

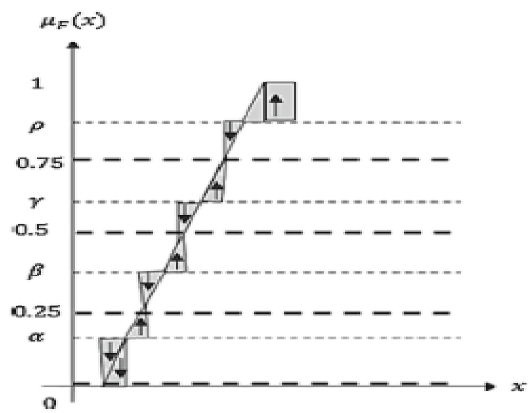


Fig. 5. The five-approximation proposal [2]

The four thresholds are determined such that $0 \leq \alpha \leq \beta$, $\alpha + \beta = 0.5$, $\alpha + \rho = 1$, $\beta + \gamma = 1$, and $\gamma \leq \rho$. The two additional regions formed, the semi-core and semi-exclusion regions, are supposed to reduce the probability of not making optimal decisions since the semi-core region contains the likely included elements and the semi-exclusion region contains the likely excluded elements. These elements in both regions are very close to being included or excluded, but they are not because there is not enough information to do so, viz. Their presence in a separate area leads to the filtering of possible decisions, thus focusing on the best ones.

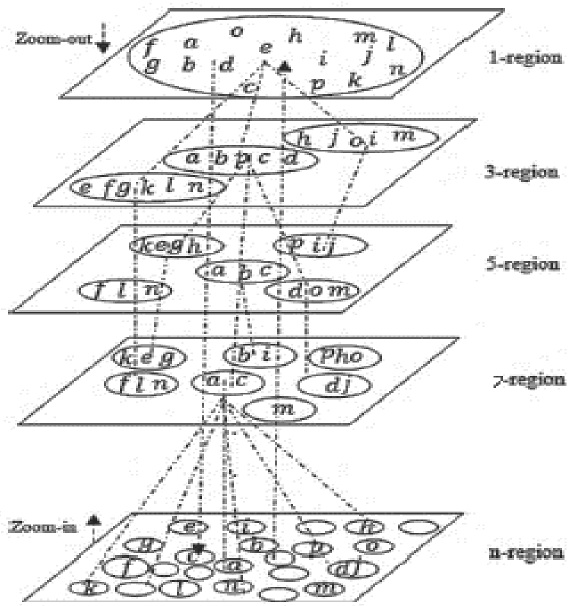


Fig. 6. The granularity levels of higher approximation degrees [2]

In addition to the four threshold values, five membership values are determined, one for each region, these values are 0, 0.25, 0.5, 0.75, 1 as shown in Fig. 5. The threshold and the membership values are interchangeably arranged such that; if the membership value of an element is greater than a specific threshold, it is being elevated to the next higher specified membership value and if it is smaller than the threshold value, it will be reduced to the next smaller specified membership value as explained in Fig. 5. The amount of uncertainty is being distributed then upon three regions, which are the shadow, the semi-core, and the semi-exclusion regions, instead of only the shadow region as in the three-way approximation approach.

The three-way approximation approach (three regions), the normal shadow set induced by a fuzzy set, can then be considered as an intermediate between the fuzzy set (one region) and the five-way approximation (five regions), it is like a zoom-out process is carried out on the fuzzy set taking it to a more detailed level of granularity. The zoom-in process is also possible, which changes the granularity level from five regions to three and back to the original one region fuzzy set. Fig. 6 shows the change in the granular level from the fuzzy set (F) to the shadow set (S_n) and vice versa, where *n* can be more than five indicating n-way approximation and *n* regions.

VI. CHALLENGES

The main challenge in inducing a shadow set by a fuzzy set is the retention of the total amount of the uncertainty of the original fuzzy. Associated with this challenge is the localization (distribution) of the uncertainty throughout the three regions of the induced shadow set, while the uncertainty is being spread among the whole space of the fuzzy set, it is being localized in the shadow region of the shadow set. In fact, the uncertainty of the elements presented in both the core and the exclusion regions are totally eliminated since the membership values are either elevated to one or reduced to zero as previously explained, the shadow region in **Pedrycz**'s model had to account for that amount of the eliminated uncertainty, but it does not succeed in capturing the actual total amount of uncertainty received from the fuzzy set. Tackling this challenge has much to do with the determination of the threshold α-cut values to cover the limitation found in the original α-cut proposed by **Pedrycz**. In fact, there is no closed formula that can be used to determine the threshold values in a way that preserves the total amount of certainty received by the fuzzy set [12, 13].

VII. CONCLUSION

The shadow set is induced by the fuzzy set as a three-way approximation giving rise to three disjoint regions, these three regions provide the shadow set with the ability to decompose the information received from the fuzzy set into information granules distributed among the three regions, reducing, therefore, the computation complexity associated with the fuzzy set, and being consequently capable of dealing with real-world problems and interpreting them with more flexibility. In addition, the shadow set is characterized by its ability to retain the uncertainty preserved in the original fuzzy set through the localization of the amount of uncertainty received among its three regions.

The first step to induce a shadow set by a fuzzy set is to determine two threshold values which will divide the space of the shadow set into three regions namely the core, the exclusion, and the shadow regions. The first two regions totally eliminate the uncertainty by performing either an elevation or a reduction operation on the fuzzy set membership function giving rise to 1 or 0 values respectively, it is therefore the shadow region that is responsible for restoring the balance and preserving the uncertainty. Therefore, the choice of the threshold values is the most important task in inducing the shadow set since they control the size of the shadow region and consequently the uncertainty balance. Many researches, some of them mentioned in this paper, have been therefore directed toward the determination of an optimum threshold values that can retain the maximum amount of uncertainty as one of the challenges facing the shadow set.

In this paper, a brief overview of the shadow set is carried out where the definition and induction by the fuzzy set is discussed covering various attempts to refine the threshold value determination proposed by **Pedrycz**. Some of the operations and relations have been discussed together with their properties. The shadow set applications like image processing and segmentation, decision making and granular computing are briefly discussed.

There are considerably few researches on shadow sets, that are not proportional to the importance and value the shadow set can add to many domains, some of them were briefly discussed in this paper. Future work may pay more attention to a higher number of approximations of the shadow set, as introduced in section five where more than three regions can be obtained, hence filtering the possible decisions and focusing only on the main ones. Also, the search for a closed formula of the threshold value that can pass all the uncertainty measures will continue.

REFERENCES

[1] Boffa, Stefania, Andrea Campagner, Davide Ciucci, and Yiyu Yao. "Aggregation operators on shadowed sets." *Information Sciences* 595 (2022): 313-333.

[2] Ibrahim, M. A., Tamunokuro Opubo William-West, Armand Florentin Donfack Kana, and D. Singh. "Shadowed sets with higher approximation regions." *Soft Computing* 24, no. 22 (2020): 17009-17033.

[3] Ibrahim, Adeku Musa, and Tamunokuro Opubo William-West. "Induction of shadowed sets from fuzzy sets." *Granular Computing* 4, no. 1 (2019): 27-38.

[4] Zhang, Qinghua, Yuhong Chen, Jie Yang, and Guoyin Wang. "Fuzzy entropy: A more comprehensible perspective for interval shadowed sets of fuzzy sets." *IEEE Transactions on Fuzzy Systems* 28, no. 11 (2019): 3008-3022.

[5] Zhou, Jie, Can Gao, Witold Pedrycz, Zhihui Lai, and Xiaodong Yue. "Constrained shadowed sets and fast optimization algorithm." *International Journal of Intelligent Systems* 34, no. 10 (2019): 2655-2675.

[6] Cai, Mingjie, Qingguo Li, and Guangming Lang. "Shadowed sets of dynamic fuzzy sets." *Granular Computing* 2, no. 2 (2017): 85-94.

[7] Chen, Long, and Shihong Chen. "New shadowed fuzzy C-means algorithm for image segmentation." In *2016 3rd International Conference on Informative and Cybernetics for Computational Social Systems (ICCSS)*, pp. 43-46. IEEE, 2016.

[8] Kacprzyk, Janusz, and Witold Pedrycz, eds. *Springer handbook of computational intelligence*. Springer, 2015.

[9] El_Hawy, Mohamed AH, Hesham A. Hassan, Hesham A. Hefny, and Khaled T. Wassif. "An improved fuzzy number approximation using shadowed sets." International Journal of Computer Applications 118, no. 25 (2015).

[10] Balas, Valentina Emilia, and János Fodor. *New Concepts and Applications in Soft Computing*. Edited by Annamária R. Várkonyi-Kóczy. Springer, 2013.

[11] Deng, Xiaofei, and Yiyu Yao. "Mean-value-based decision-theoretic shadowed sets." In *2013 Joint IFSA world congress and NAFIPS annual meeting (IFSA/NAFIPS)*, pp. 1382-1387. IEEE, 2013.

[12] Tahayori, Hooman, Alireza Sadeghian, and Witold Pedrycz. "Induction of shadowed sets based on the gradual grade of fuzziness." *IEEE Transactions on Fuzzy Systems* 21, no. 5 (2013): 937-949.

[13] Pedrycz, Witold. "From fuzzy sets to shadowed sets: interpretation and computing." *international journal of intelligent systems* 24, no. 1 (2009): 48-61.

[14] Cattaneo, Gianpiero, and Davide Ciucci. "Theoretical aspects of shadowed sets." Handbook of granular computing (2008): 603-627.

[15] Pedrycz, Witold, and George Vukovich. "Granular computing with shadowed sets." International Journal of Intelligent Systems 17, no. 2 (2002): 173-197.

[16] Pedrycz, Witold. "Shadowed sets: representing and processing fuzzy sets." *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 28, no. 1 (1998): 103-109.

[17] Klir, George J. "Principles of uncertainty: What are they? Why do we need them?" *Fuzzy sets and systems* 74, no. 1 (1995): 15-31.

Soft Set Theory and its Applications in Multi-Criteria Decision-Making Problems

Souad Abu Jabal

Department of Computer Science, Faculty of Graduate Studies for Statistical Research,

Cairo University

Egypt, Cairo

Sabujabal2@gmail.com

Hesham Hefny

Department of Computer Science, Faculty of Graduate Studies for Statistical Research,

Cairo University

Egypt, Cairo

hehefny@cu.edu.eg

Abstract: Although the existence of many well-known and significant tools to handle uncertainty of data such as fuzzy set theory, rough set theory, and intuitionistic fuzzy set theory, each of these theories has its own difficulty. For example, even though the fuzzy set theory of Zadeh is powerful and has a great attention, it does not offer a parameterized tool to accurately set the membership degree of an element. Consequently, in 1999, the soft set theory of Molodtsov was presented as a different technique for modeling uncertainty using parameters. Simply, in soft sets, the problem of setting the membership function is not existing. Many extensions of the theory were proposed later to improve its performance and for making it more convenient to be applied in different areas. Through this paper, we shed the light on the soft set theory supported by an application in decision making problems.

Keywords: *Soft sets, Soft-set operations, Fuzzy soft sets, Weighted-soft sets, Decision making.*

1. Introduction

Recently, vague and uncertainty concepts become heavily used in different areas. Classical techniques are unable to successfully handle these problems because they rely on crisp approaches that make a certain phenomenon either true or false. Although, the existence of many theories to handle uncertainty problems, each of them has its own difficulties.

The most commonly used theory in handling vagueness and uncertainty is Zadeh's fuzzy set theory (Zadeh 1965) [1]. However, it suffers from the difficulty of identifying the membership degree of an element accurately. The reason for this difficulty due to the lack of a parametrization tool of the theory. Consequently, (Molodtsov 1999) [2] presented the soft set theory as a different technique that successfully handle uncertainty using parameters. Soft set theory is considered a general form of fuzzy sets. It was described as "soft" because the boundaries of the set depend on the number of parameters given. And also, because it doesn't impose any restrictions on the types of the parameters used (i.e., parameters may be words, sentences, numbers, and so on). This flexibility about parameters makes the theory more convenient and easier to be applied in practice. One of the important notes about the soft set is that it is not a set as the known sense of sets, it is simply a parameterized family of subsets of a universal set.

Although the great success of the soft set theory, few of its power were presented in the pioneer work of Molodtsov. The different mindset of the theory encourages many mathematical researchers to extend it by proposing several hybrid models. Most of these models combine the soft set theory with one or more other theories of uncertainty to get all of their benefits. In this section, we present some of the researches introduced in this topic, for example (Maji et al. 2001) [3] presented a model that combines the soft set theory with the fuzzy set theory. This model in decision making was introduced as the first practical application in this area and is referred to as a fuzzy soft model. In addition, (Maji et al. 2002) [4] added some concepts of the rough set theory of (Pawlak 1991) [5] to the soft sets to minimize the number of parameters used in the soft set. Unfortunately, this model produces impractical optimal results. (Maji et al. 2003) [6] introduced some new algebraic operations on soft sets. (Keles et al. 2007) [7] applied the soft set theory concepts in designing a classifier called a neuro−fuzzy classifier. It can be used to diagnose the benign prostatic hyperplasia (BPH) diseases and the prostate cancer. (Yang et al. 2009) [8] presented the concept of the interval−valued fuzzy soft sets by combining the interval−valued fuzzy set and soft set together. (Ali et al. 2010) [9] presented some new operations on soft sets. (Simsekler and Yuksel 2012)[10] introduced what is called a Fuzzy soft topological space. (Hakim, Sari, and Herawan 2014) [11] showed that the soft set theory is capable to be used as a general tool in many mathematical purposes and also can describe objects in form of information systems. They also proposed a multi−dimensional scaling technique to find the optimal soft solution in decision making.

(Zakri, Hossen, and Shari 2016) [12] suggests an application that applies the soft sets in diagnosing the students' educational problems. (Malathy and Meenakshi 2019) [13] presented the concept of fuzzy soft bi-partite graph supported by an application that shows how it can be applied in solving decision making problems in organizations. (Akram et al. 2021)[14] introduced a Parameter reduction model of N- soft sets supported with applications in decision- making. (Palanikumar and Iampan 2022) [15] provided a new approach in decision making depends on Type-II bipolar fuzzy soft sets.

In this work, we shed the light on the soft set theory in solving decision-making problems. The next sections of the paper are organized as follows. Section 2, recalls some basic notations and operations of the theory. Section 3, presents the main differences between the soft set theory and the fuzzy set theory. Section 4, shows how to obtain an accurate fuzzy membership degree based on the soft sets. In section 5, we present one extension of the soft set theory supported by an application. Section 6 views the challenges and future work. Finally, the conclusion is reached in section 7.

2. Preliminaries

This section presents some basic notations and operations of the soft set theory of (Molodtsov 1999) [2].

Definition 2.1

Soft sets are defined by Molodtsov as following:

Let U be the initiate universe set, E is the parameters set, and P(U) is the power set of U and B⊂E.

A pair (Q, B) is called a soft set over U if and only if Q is a mapping as Q: B → P(U), and for e ∈ B, Q(e) is the set of e-approximate elements of the soft set (Q, B). It's obvious that the soft set is not a set. It is a parameterized family of subsets of the universe U. To understand the idea, Molodtsov presented a set of examples. We choose the following one to present.

Example 2.1

Let U is the set of available chalets that Mr. Smith is intended to select one of them to buy.

Assume that, we have 7 chalets in the universe U as following:

U = {ch_l, ch_2, ch_3, ch_4, ch_5, $ch_{6,}$ ch_7),

E is the parameters set,

E = {nice; modern; in good location; cheap; in good state; in bad state; luxury; sunny}, (Note that, the parameter e_1 means '*nice*', e_2 means '*modern*', and so on.), and

B is the parameters' set of Mr. Smith such that $B \subset E$, and

B= {e_1, e_2, e_3, e_4, e_5}.

Let the soft set (Q, B) shows the "description of the chalets" that may be suitable to Mr. Smith.

Q(e_1) = {ch_1, ch_3}, Q(e_2) = {ch_2, ch_4}, Q(e_3) = {ch_1, ch_3, ch_5}, Q(e_4) = {ch_2, ch_4, ch_6}, Q(e_5) = {ch_2}.

The mapping Q(e_1) means the "nice chalets" whose functional value is the set {h_1, h_3}, and the mapping Q(e_2) means modern chalets whose functional value is the set {h_2, h_4}, and so on.

Thus, the soft set (Q, B) can be viewed as a collection of approximations as following:

(Q, B) = {nice chalets = {ch_1, ch_3}, modern chalets = {ch_2, ch_4}, in good location chalets = {ch_l, ch_3, ch_5}, cheap chalets = {ch_2, ch_4, ch_6}, in good state chalets = {ch_2}}.

Notes

- It is not necessary that each Q(e) be nonempty. It may be empty, if no value matches the parameter.
- Any approximation consists of two parts:

(i) Predicate name, referred to as (p), and

(ii) Approximate value−set (it is also called value−set), and referred to as (ν).

For example, the approximation (nice chalet = {ch_1, ch_3}) includes two parts:

"nice chalet" is the predicate name, and {h_1, h_3}is the approximate value set.

- A soft set can be represented in a tabular form, i.e., the soft set (Q, B) of the previous example is represented as shown in table 1 below.

TABLE 1 TABULAR REPRESENTATION OF THE SOFT SET (Q, B)

U	'nice' e_1	'modern' e_2	'in good location' e_3	'cheap' e_4	'In good state' e_5
ch_1	1	0	1	0	0
ch_2	0	1	0	1	1
ch_3	1	0	1	0	0
ch_4	0	1	0	1	0
ch_5	0	0	1	0	0
ch_6	0	0	0	1	0

It may also be shown in another form as a collection of approximations as following:

(Q, B) = {$p_l = v_l$, $p_2 = v_2$,……,$p_m = v_m$}.Such that p is the predicate name, and v is the value set.

Definition 2.2

Let (Q, B) and (F, D) are 2 soft sets over U, then (Q, B) is a soft subset of (F, D) if B ⊂ D.and ∀e ∈ B, both of the approximations Q(e) and F(e) are identical. Then, we denote it as (Q, B) $\tilde{\subset}$ (F, D)

Definition 2.3

Assume that B is the set of parameters, such that B= {b_l, b_2, b_3,…………, b_n}.

We refer to the NOT set of B as ꓶB, and ꓶB is defined as:

ꓶB = {ꓶb_l, ꓶb_2, ꓶb_3,…. , ꓶb_n} where, ꓶb_i = not b_i $\forall_i$ (Note that, the two operators ꓶ and ꓶ are different).

Definition 2.4

The complement of' a soft set (Q, D) is referred to as soft complement $(Q, D)^C$ and is defined by:

$(Q, D)^C = (Q^C, ꓶ D)$ where Q^C : ꓶ D → P(U) is a mapping given by $Q^C(\alpha) = U – Q(ꓶ\alpha)$ ∀ α ∈ ꓶ D.

And $((Q, D)^C)^C = (Q, D)$

Definition 2.5

A soft set (Q, B) over U is a NULL soft set and denoted by ϕ if ∀e ∈ B, all the approximations Q(e) = ϕ.

Definition 2.6

A soft set (Q, B) over U is said to be absolute soft set and denoted by Ã, if ∀ e ∈ A, Q(e) = U.

Therefore, $\tilde{A}^C = \phi$, and $\phi^C = \tilde{A}$.

Definition 2.7

The union of two soft sets (Q, B) and (F, C) over the common universe U is (H, Z) where Z = B ∪ C, and

∀ a ∈ Z, [H (a) = Q(a), if a ∈ B – C], [H(a) =F(a), if a ∈ C – B], or [H(a) = Q(a) ∪ F(a) if a ∈ B ∩ C].

And, we denote it as (H, Z) = (Q, B) $\tilde{\cup}$ (F, C).

Definition 2.8

The intersection of two sets (Q, B) and (F, C) over a common universe U is the soft set (H, Z) were,

Z = B ∩ C, and ∀ a ∈ Z, [H(a) = Q(a)] or [H(a) = F(a)], (Since they are the same).

And we denote it as (H, Z) = (Q, B) $\tilde{\cap}$ (F, C).

3. Soft set theory VS. Fuzzy set theory

Through this section, we will introduce the main differences between the soft set theory and fuzzy set theory through an illustrative example. The example shows the use of soft sets of (Molodstov 1999) [2] in solving decision making problems. It uses the concept of the choice value as a means to compare the alternatives. The choice value of each alternative is calculated by adding the value sets of the existing parameters together. Each parameter's value may be one, if it is satisfied, or zero otherwise. After this example, we will be able to understand the differences between the two theories.

Example 3.1

Assume that, the problem of Mr. Smith is how to choose the most suitable chalet among a set of chalets according to his parameters set. To start solving the problem:

Let U = {ch_1, ch_2, ch_3, ch_4, ch_5, ch_6} be a set of chalets, and

E = {sunny, expensive, in good location, cheap, modern, with good state, surrounded by gardens, with bad state} be a set of parameters.

Let the soft set (F, E) = {sunny chalets (e_0) ={ ch_1, ch_3, ch_4, ch_5, ch_6}, expensive chalets (e_1) = φ, in good location chalets (e_2) = {ch_1, ch_2, ch_6}, cheap chalets (e_3) = {ch_1, ch_2, ch_3, ch_4, ch_5}, modern chalets (e_4) = {ch_1, ch_2, ch_4, ch_5}, with good state chalets (e_5) = {ch_1, ch_5}, surrounded by gardens chalets (e_6) = {ch_1, ch_2, ch_6}, with bad state chalets (e_7} = φ}.

The first step is to represent the soft set (F, E) that represents all the parameters set as shown in Table 2 below. Each row represents an object (chalet) and each column represents a parameter.

TABLE 2. SHOWS THE REPRESENTATION OF SOFT SET (F, E) IN TABULAR FORMAT

U	e_0	e_1	e_2	e_3	e_4	e_5	e_6	e_7
ch_1	1	0	1	1	1	1	1	0
ch_2	0	0	1	1	1	0	1	0
ch_3	1	0	0	1	0	0	0	0
ch_4	1	0	0	1	1	0	0	0
ch_5	1	0	0	1	1	1	0	0
ch_6	1	0	1	0	0	0	1	0

The second step is to construct the soft set (F, B) which contains only the parameters set of Mr. Smith (Suppose that, Mr. Smith interested in purchasing a chalet according to his parameter set B, such that: $B \subset E$, and B = {sunny, in good location, cheap, modern, with good state} = {e_0, e_2 e_3, $e_{4,}$ e_5}.

Then, the soft set (F, B) that describes the available chalets which matches Mr. Smith' parameters set B will be shown in table 3 below. A new column is added to the right to calculate the choice value d_i for each alternative. The choice value is the sum of the values of the parameters in each row as we mentioned.

TABLE 3. SHOWS THE CHOICE VALUES OF THE SOFT SET (F, B)

U	e_0	e_2	e_3	e_4	e_5	choice value (di)
ch_1	1	1	1	1	1	5
ch_2	0	1	1	1	0	3
ch_3	1	0	1	0	0	2
ch_4	1	0	1	1	0	3
ch_5	1	0	1	1	1	4
ch_6	1	1	0	0	0	2

The last step is to compare the choice values obtained in table 3 together. Thus, will find that ch_1 has the highest choice value $d_{i=5}$, so, it is the optimal choice for Mr. Smith. Whereas, ch_5 has the second highest value which is 4 and it can be considered the sub-optimal choice. Finally, each of ch_3 and ch_6 is considered the worst choice.

Now, we can summarize the main differences between soft set theory and fuzzy set theory as following:

- A soft set solution gives a detailed description in form of parameters. It uses the concept of choice values that are calculated depending upon the parameters set. The choice values are compared together to find the optimal decision. On the other hand, in the fuzzy set theory. The membership degree of an element is identified as a restricted value without detailed information.
- In soft sets, the presence or absence of each parameter is described in crisp form by giving one to the existing parameter and zero to express its absence, while in the fuzzy set, no parameters are used.
- In soft sets, the choice value of an object may be exceeding one, while in the fuzzy sets, the membership function values should be range between zero and one,
- Soft set is considered as a general form of the fuzzy set in which the membership degree of an element not necessarily to be ranged between zero and one. It can exceed one.

4. Constructing accurate fuzzy membership degrees depending on the soft sets

This section presents an improvement solution for the problem of Mr. Smith mentioned in example 3.1. To start the solution, let's consider table 4 below. It is the same as table 3 except that we add a new column to the right to calculate the fuzzy membership degree in each case. In our case, the fuzzy membership degree is calculated depending on the choice value of soft sets by dividing each choice value in each row on the total number of parameters determined by Mr. Smith (which are five parameters). For example, Ch_1 satisfies five parameters out of the total number of parameters which is also five, then the fuzzy membership degree of ch_1 is calculated as $f\ (ch_1) = 5\ /\ 5 = 1$, the fuzzy membership degree of ch_2 which is $f(ch_2) = 3\ /5 = 0.6$, and so on.

TABLE 4. FUZZY MEMBERSHIP DEGREES DEPENDING ON SOFT SETS

U	Soft Set Parameters					Choice value "di" Using the soft set	Fuzzy membership degree "f"
	e_0	e_2	e_3	e_4	e_5		
ch_1	1	1	1	1	1	5	1
ch_2	0	1	1	1	0	3	0.6
ch_3	1	0	1	0	0	2	0.4
ch_4	1	0	1	1	0	3	0.6
ch_5	1	0	1	1	1	4	0.8
ch_6	1	1	0	0	0	2	0.4

We summarize the similarities and the differences between the soft set solution and the fuzzy solution obtained as following:

- Both solutions depend on the idea of a parameters and both lead to the same decisions.
- The main difference between them is that the choice value of the soft set solution can exceeds one, while the fuzzy solution should range between zero and one.

5. Soft set theory in decision making

This section presents a decision-making application that is built upon one a significant extension of the soft set theory (Maji et al 2002) [4]. This model combines some concepts of the rough set theory (i.e., the concept of reduction of parameters) to the soft set theory to achieve parameter reduction. The solution obtained is improved by applying the W-soft set theory of (T. Y. Lin 1996) [16].

Application 5.1

Assume that, Mr. Smith is interested to buy a chalet according to his selection parameters set A. The problem is how to help him selecting the best chalet that matches all (or most) of his parameters set A. To solve the problem:

Let U be a set of 8 chalet, U = {ch_1, ch_2, ch_3, ch_4, ch_5, ch_6; ch_7; ch_8}, and

E is the set of all parameters, E = {nice; in good location; cheap; sunny; in good state; luxury; modern; expensive}.

A is the parameter set of Mr. Smith, A= “nice, in good location, cheap, sunny, in good state”, such that $A \subset E$.

In other form, A = {e_1 = nice; e_2 = in good location; e_3 = cheap; e_4 = sunny; e_5 = in good state}.

Let (G, E) is the soft set that shows the” features of chalets” such that:

(G, E) = {nice chalet = ϕ, in good location chalet = {ch_1, ch_2, ch_3, ch_4, ch_5, ch_6}, cheap chalet = {ch_1, ch_2, ch_6}, sunny chalet = {ch_1, ch_2, ch_6}, in good state chalet = {ch_2, ch_4, ch_5}, luxury chalet = {ch_1, ch_2, ch_3, ch_4, ch_5, ch_6}, modern state chalet = {ch_1, ch_3, ch_6}, expensive chalet = {ch_1, ch_2, ch_3, ch_4, ch_6}}.

Let (G, A) be the soft set that is constructed depending on Mr. smith’s preferred parameters set only.

a) Tabular Representation of a Soft Set (G, A)

(G, A) represents the soft set that depends on the set A of choice parameters of Mr. Smith.

The tabular representation of (G, A) is shown in table (5) below.

TABLE 5. SHOWS A TABULAR REPRESENTATION FOR THE SOFT SET (G, A)

u	“nice” e_1	“in good location” e_2	“cheap” e_3	“sunny” e_4	“in good state” e_5
ch_1	1	1	1	1	1
ch_2	1	1	1	1	0
ch_3	0	1	1	1	0
ch_4	0	1	1	1	0
ch_5	0	1	0	1	0
ch_6	1	1	1	1	1

b) Reduct−Table of a soft set

The reduct−table is a table that only includes the necessary part of a soft set which is sufficient to describe all of the essential approximations of the soft set. Accordingly, to the tabular form of the soft set (G, A), if B is a reduction of A, then the soft set (G, B) is referred to as the reduct soft set of the soft set (G, A). Before we construct the reduct table, let's present the selection algorithm to reach an optimal decision.

c) Selection Algorithm

Mr. Smith can use the following algorithm to help him in choosing the most suitable chalet for him:

1. Identify the soft set (G, E).
2. Identify the soft sub-set (G, A) such that A is the parameter set of Mr. Smith, and A $\subset$E.
3. Find all possible reduct-soft sets of (G, A).
4. Select randomly one reduct-soft set of (G, A); say (G, B).
5. Search about k that satisfies ν_k = max ν_i. And ν_k is the most suitable selection for Mr. Smith.

Note that, *k* may have more than one value; in this case Mr. Smith can select any of them.

Now, let's apply the previous algorithm to our problem.

We noted from table 5 that: There are two reductions of the set A = {e_1, e_2, e_3, e_4, e_5}.

They are B = {e_1, e_3, e_4, e_5}, and C= {e_1, e_2, e_3, e_5}. So, we can select any of them randomly, say B which is equal to {e_1, e_3, e_4, e_5}. Then the reduct-table of the soft set (G, B) is shown in table 6 below:

TABLE 6. SHOWS THE REDUCT TABLE OF THE SOFT SET (G, B)

U	In good location e_1	cheap e_3	sunny e_4	In good state e_5
ch_1	1	1	1	1
ch_2	1	1	1	0
ch_3	0	1	1	0
ch_4	0	1	1	0
ch_5	0	0	1	0
ch_6	1	1	1	1

Now, we will calculate the choice values as shown in table 7 below.

TABLE 7. SHOWS THE REDUCT-TABLE AFTER ADDING THE CHOICE VALUES

U	'in good location' e_1	'cheap' e_3	'sunny' e_4	'in good state' e_5	Choice value (ν_i)
ch1	1	1	1	1	4
ch2	1	1	1	0	3
ch3	0	1	1	0	2
ch4	0	1	1	0	2
ch5	0	0	1	0	1
ch6	1	1	1	1	4

Then, the decision of Mr. Smith is to choose either the chalet ch_1 or ch_6 according to his desire because they have the highest choice values (4). However, this solution can be improved as following:

d) Improved solution supported by weights:

Due to the fact that, not all the parameters of Mr. Smith have the same importance to him. Accordingly, he may prefer to set weights to his parameters. So, a weight $w_i \in (0, 1]$ is given to each parameter as follows: 0.7, 0.8, 0.9, and 0.6 are given to e_1, e_3, e_4, and e_5 respectively. Then, table 8 below is produced. Here, the choice value v_1= 0.7+ 0.8 +0.9+ 0/6 = 3.0, and v_2= 0.7+ 0.8 +0.9 = 2.4 and so on.

TABLE 8. THE WEIGHTED REDUCT–TABLE

U	"In good location" e_1 w1=0.7	"cheap" e_3 w3=0.8	"sunny" e_4 w4=0.9	"In good state" e_5 w5=0.6	Choice value (v_i) According to weights
ch_1	1	1	1	1	3.0
ch_2	1	1	1	0	2.4
ch_3	0	1	1	0	1.7
ch_4	0	1	1	0	1.7
ch_5	0	0	1	0	0.9
ch_6	1	1	1	1	3.0

Accordingly, ch_1 and ch_6 are the best choices for Mr. Smith. He can select one of them.

6. Challenges and future work

The main difficulty of the soft set theory is that, in case of a problem with a large number of parameters, the theory may not work properly. some researchers tried to handle this Parameter reduction issue such as (Maji et al. 2002) [4] and (Akram et al. 2021) [14]. However, the obtained results are not relatively acceptable because some of them are impractical (although they are optimal) and the other some needs further efforts to be significant. Another difficulty of the theory is that it doesn't consider the concept of non-membership degree of an element. It still considers the concept of membership degree only.

7. Conclusion

The soft set theory could successfully handle the problem of other theories related to the inadequacy of a parametrization tool. One advantage of the theory is that it doesn't impose any restrictions on the parameters used. This makes it more convenient. Due to the different mindset of the theory, many researches were presented to show its use lonely or in a combination with other theories of uncertainty. Through this paper we introduced the soft set theory supported by a decision–making application.

REFERENCES

[1] Zadeh.L. A: Fuzzy sets. Inf. Control 1965, 8: 338–353. 10.1016/S0019-9958(65)90241-X
[2] Molodtsov.D, Soft set theory-first results, Computers Math. Applic., 37 (4/5), (1999), 19-31.
[3] Maji.P.K, Roy.A.R, and Biswas.R: Fuzzy soft sets. J. Fuzzy Math. 2001, 9(3):589–602.
[4] Maji.P.K, Biswas.R, Roy.A.R, an application of soft sets in a decision-making problem, Computers and Mathematics with Applications 44 (2002), 1077-1083.
[5] Pawlak Z, Rough sets, Theoretical aspects of reasoing about data, Kluwer Academic, Boston, MA, (1991).
[6] Maji PK, Biswas R, Roy.A.R, Soft set theory, Comput. Math. Appl. 45 (4–5) (2003) 555–562.
[7] Keles.A, Hasiloglu.A.S, Aksoy.Y: Neuro-fuzzy classification of prostate cancer using NEFCLASS-J. Comput. Biol. Med. 2007, 37: 1617–1628. 10.1016/j.compbiomed.2007.03.006
[8] Yang.X.B, Lin.T.Y, Yang.J.Y, Li.Y, Yu.D, Combination of interval-valued fuzzy set and soft set, Comput. Math. Appl. 58 (3) (2009) 521–527.
[9] Ali.M.I, Feng.F, Liu.X, Min.W.K, Shabir.M, On some new operations in soft set theory, Comput. Math. Appl. 57 (9) (2009) 1547–1553.
[10] Simsekler.T.H, Yuksel.S: Fuzzy soft topological spaces. Ann. Fuzzy Math. Inf. 2012, 5(1):87–96
[11] Hakim.R.BF, Sari.E.N, and Herawan.T,Soft Solution of Soft Set Theory for Recommendation in Decision Making, Recent Advances on Soft Computing and Data Mining SCDM , Pp-313-319,2014.
[12] Zakri.A.H, Hossen.H.M, Erwi.L.Y and Shari. E.A, Application of Soft Sets to Diagnose the Educational Obstacles for Students, Journal of Innovative Technology and Education, Vol. 3, no. 1, pp.61 - 70, 2016.
[13] Malathy.K and Meenakshi.S, Fuzzy soft bi-partite graph and its Application in Employee Selection for An organization, International Journal of Recent Technology and Engineering (IJRTE),Vol-7, Issue-5C, 2019.
[14] Akram, M., Ali, G., Alcantud, J.C. and Fatimah, F., 2021. Parameter reductions in N-soft sets and their applications in decision-making. *Expert Systems*, *38*(1), p.e12601.
[15] Palanikumar, M. and Iampan, A., 2022. A Novel Approach to Decision Making Based on Type-II Generalized Fermatean Bipolar Fuzzy Soft Sets. *International Journal of Innovative Computing, Information and Control*, *18*(3), pp.769-783.
[16] Lin.T.Y, A set theory for soft computing, a unified view of fuzzy sets via neighbourhoods, In proceedings of 1996, IEEE International Conference on Fuzzy Systems, New Orleans, LA, September, (1996), 8-11.

Breast Cancer Prediction Using Transfer Learning

Ghada Mohammed Mansour Eissa, Shahira Shaaban Azab, Hesham Ahmed Hefny
Department of Computer Science, Faculty of Graduate Studies for Statistical Research, Cairo University, Giza, Egypt
ghadaeissa@pg.cu.edu.eg

***Abstract*—**
Breast cancer is one of the fatal diseases in the world. The early prediction of breast cancer is a global challenge. Deep Learning models have become efficient tools to classify and predict Breast Cancer Tumors. AlexNet and GoogleNet are powerful deep learning models used to predict Breast Cancer. Early detection of breast cancer leads to can make the treatment more effective. in this comparison, AlexNet and GoogleNet are employed to predict breast cancer in 50 female Cases. The dataset of these cases has been taken from the oncology center of Nasser Institute for Research and Treatment in Egypt. A comparison between the results of AlexNet and GoogleNet is held. GoogleNet is more accurate as it achieves 82% accuracy while AlexNet achieves 78% Accuracy.

Keywords—Breast Cancer Prediction; Deep Learning; Transfer Learning; AlexNet; Google Net.

I. INTRODUCTION

In the last few decades, breast cancer has become a life-threatening disease. According to the World Health Organization (WHO) reports, Breast cancer treatment is very essential as it can achieve more than 90 % survival probability when the disease is detected early. At the beginning, the tumor is restricted to the duct or lobule with no symptoms. The tumor may be developed and start invading the breast and the lymph nodes or any other organs in the body. Survival percentage of breast cancer after early detection and treatment in wealthy countries. is higher than that in the developing countries. Breast cancer treatment protocols have been developed during the previous few years. Several studies have proposed new treatments protocols to improve the survival rate of breast cancer. World Health Organization (WHO) explain the prediction of breast cancer mortality in the period between 2020 to 2040 with comparison of worldwide and Egypt cases. (See Table 1) [1], the comparison between Estimated number of new cases of Breast Cancer from 2020 to 2025, Both sexes, age [0-85+] to some countries with Egypt and Estimated number of new cases of Breast Cancer from 2020 to 2040, Both sexes, age [0-85+] . (See Figure 1)

Table 1 Breast Cancer Worldwide vs. Egypt[2]

Cancer label	Population	Sex	Type	Cases base in 2020	Year	Prediction	Risk change	Risk percentage (%)
Breast	World	0	1	684996	2020	684996	0	0
Breast	World	0	1	684996	2025	768761	83765	**12.228**
Breast	World	0	1	684996	2030	857319	172323	**25.156**
Breast	World	0	1	684996	2035	949530	264534	**38.618**
Breast	World	0	1	684996	2040	1040813	**355817**	**51.944**
Breast	Egypt	0	1	9148	2020	9148	0	0
Breast	Egypt	0	1	9148	2025	10592	1444	**15.784**
Breast	Egypt	0	1	9148	2030	12342	3194	**34.914**
Breast	Egypt	0	1	9148	2035	14381	5233	**57.203**
Breast	Egypt	0	1	9148	2040	16651	**7503**	**82.017**

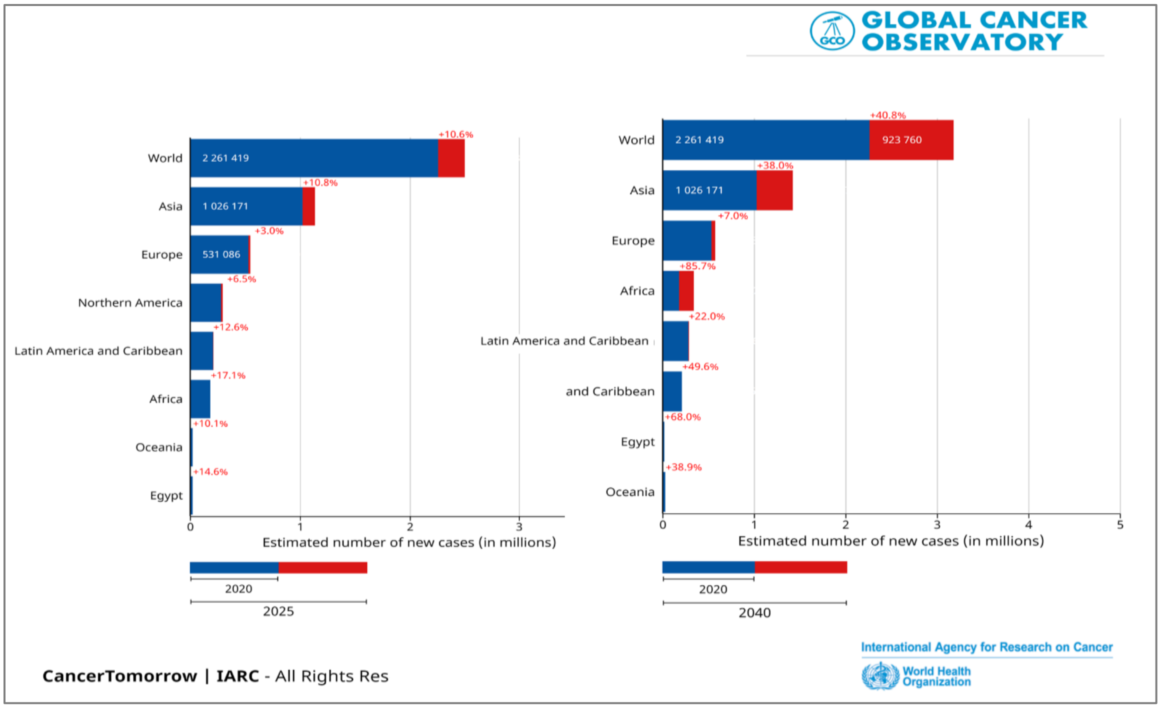


Figure 1 the comparison between Estimated number of new cases of Breast Cancer from 2025 to 2040

Our study was conducted in 50 female Cases coming to the mammogram imaging unit of oncology center of Nasser Institute for Research and Treatment in Egypt between the period of February **2021** to September **2022**, these Cases underwent screening and diagnostic full field digital mammography unit with two main views:
1-Craniocaudal View (RCC and LCC)
2-Mediolateral Oblique (RMLO and LMLO)
to gives us **180** Images and we increased it to 780 images after using the data Augmentation Method, the following Results has been obtained by using sample of dataset of 180 Images have been converted from Mammogram DICOM files to X-Ray PNG Images, these X-Ray Images were divided into 9 classes based on 6 categories of Breast Imaging Reporting & Data System (BI-RADS) classification and we classified the category 4 into 3 classes (4A, 4B, 4C) to become 9 classes from 0 to 8 as following:
Category (0) = class0, Category (1) = class1, Category (2) = class2, Category (3) = class3, Category (4A) = class4, Category (4B) = class5, Category (4C) = class6, Category (5) = class7, Category (6) = class8. The rest of the paper is organized as follows, in section two, a literature review about Breast Cancer Prediction using deep learning models while section three is devoted to experiments and results of AlexNet[3] and Google Net[4] Models. In section four the conclusion and future work of this study is presented.

II. Literature Review

Several research papers have been proposed to develop Breast Cancer prediction techniques.
Al-Antari et al. [(2018) presented a model using INbreast database [5] to assess the accuracy, segmentation, and classification of the proposed integrated computer-aided detection (CAD) system [6]. The INbreast database accomplished an accuracy of 98.96%, a Mathews Correlation Coefficient (MCC) of 97.62 %, and an F1 score of 99.24 %, based on CAD evaluation system. In addition, the cluster segmentation archived accuracy of 92.97% by YOLO, MCC 85.93%, and Dice() (F1 score) while Jaccard Similarity coefficient measured as 86.37%, respectively [7].

Mohani Viglu et al. (2020) proposed a new CAD system for mass classification. This proposed system accomplished a high accuracy using a few numbers of features. The authors classified mammograms into two classes, normal class, and abnormal class. Their system gives tumors mass classification (benign or malignant). They used the Database for Digital Screening Mammography (DDSM) [8], the Mammography Analysis Society

(MIAS) [9] and the Breast Cancer Digital Repository (BCDR) datasets [10] to validate their proposed system. Such technique accomplished 99.62% accuracy for MIAS and 99.92% accuracy for DDSM, for both of normal and abnormal classification. The benign/malignant classification archives accuracy of 99.28%, 99.63%, 99.60% by using MIAS, DDSM, BCDR respectively [11].

Shin et al. (2020) based on antagonistic learning attempted to facilitate the annotation of masses in mammograms by introducing a new automated process. The two grids of the frame are proposed as follows. The first one is a Full Convolutional Network (FCN) which is proposed for spatial density prediction, while the second one is a domain. This approach gives 0.9083 AUC score for the private data set and 0.8522 for the INBreast. [12].

Ansar et al. (2020) presented MobileNet which is based on an architectural model Such approach can successfully classify masses in mammograms into malignant and benign tumors with high accuracy and lower computational cost. In addition, the proposed approach can successfully classify mammograms into two categories, cancerous and non-cancerous by using a Convolution Neural Network (CNN) approach. The performance of this model is better compared to some other models like VGG-16 and VGG-19 [13].

Cao et al. (2021) presented A new model for the detection of breast masses in mammograms used INbreast and DDSM dataset. with a new data augmentation technique to overcome overfitting problem. The proposed technique improves the model's performance of the model by using Feature Selective Anchor-Free (FSAF) [14], but with slow computational speed Proposed truncation normalization method for image preprocessing and combination with adaptive histogram equalization algorithm to improve image contrast [15].

Naveed Chouhan et al. (2021) presented an automated breast cancer detection system. The system which is Diverse Features based Breast Cancer Detection (DFeBCD) is presented to classify mammograms. The results show that, the dynamic features performance of which is generated by the proposed highway network based on the CNN highway network is better than all the other individual groups of features. In addition, by using the hybrid feature space, the ELiEC classifier gives a better performance than SVM (see Table 3, Table 4) [16].

Soham Chattopadhyay et al (2022) proposed a new deep learning model, called Multi-scale Dual Residual Recurrent Network (MTRRE-Net) [17] depended on four different magnification levels of BreakHis dataset. [18] on the challenging BreaKHis dataset, our model achieves accuracy, the images accuracy values as folowing, 40x= 97.12%, 100x=95.22%, 200x= 96.85%, 400x=97.81%, to classify benign and malignant cells. We have compared the obtained results with some pretrained models that include DenseNet161, DenseNet169, GoogLeNet, ResNet18, VGG16, and we have found that it outperforms all these models, thereby confirming the superiority of the model. Moreover, we have compared our model with some state-of-the-art models which reported the benchmark results, and the comparison validates the effectiveness of the proposed model, this dataset is Breast Cancer Histopathological Database (BreakHis) publicly available standard dataset.

Table 2 Pros and Limitation Breast Cancer Deep Learning Dataset

Ref.	Task	Dataset	Pros	Limitation
Al-antari et al. [(2018)[7]	Detection segmentation Classification	INBreast	Classification accuracy has increased with the hash that was Performed with perfect precision Convolutional Network (FrCN). Their approach reveals the masses with a single shot detector (YOLO) which means less information is lost.	detection of small blocks. They removed the wrong compiler manually. Masses before segmentation, which is impractical for an automated CAD system.
Mohanty Figlu et al. (2020) [11]	Mammogram Classification	DDSM MIAS BCDR	Their proposed scheme classifies mammograms in real time.	ROIs are manually cropped.
Shen et al. (2020) [12]	Detection	INBreast Private Dataset	Their training strategy addressed: -The problem of Oscillation.	Their methodology should be investigated in more medical terms shooting scenarios.
Ansar et al. (2020) [13]	Classification	DDSM CBIS-DDSM	Their approach is reduced computational cost compared to current models.	Low classification accuracy compared to other methods.
Cao et al. (2021) [15]	Detection	INbreast DDSM	It turns out that the proposed BMassD-Net could be a competitor. Exposing current top-ranking methods.	Their enlargement technology is based on localized elastic deformation, this technology. boosted their model performance; However, its calculation speed is slower compared to traditional increment techniques.
Naveed Chouhan et al. (2021)[16]	Xception SqueezeNe	IRMA Mammogram Dataset	DFeBCD uses a Emotional learning based band Classifier (ELiEC) for Category Mammograms normal or abnormal.	the performance The comparison indicates that proposed technique It excels from end-to-end CNN models.
Soham Chattopadhyay et al (2022)[17]	MTRRE-Net	BreakHis	The model achieves accuracy of images to classify benign and malignant cells	MTRRE-Net more accurate of DenseNet161, DenseNet169, GoogleNet, ResNet18, VGG16

Table 3 Results of Mass Detection

Ref.	Method	Results				
		Detection / Segmentation			Classification	
		AUC	ACC. Detection	ACC. Segmentation	AUC	ACC.
Al-antari et al. [(2018)[7]	-Using (FrCN) to increase classification Accuracy. - A deep model based on YOLO to accurately detect masses from mammograms is fully adopted - CNN model, ResNet-50 and Inception ResNet-V2 are used to identify the mass as either benign or malignant.		98.96%	92.97%	94.78%	95.64%
Mohanty Figlu et al. (2020) [11]	Principal Component Analysis (PCA) An enhanced envelope-based extreme learning machine (KELM) using the weighted salp swarm algorithm (WC-SSA) has been proposed as a classifier for digital mammogram classification.					**Normal & Abnormal** DDSM: Acc=99.92% MIAS: Acc=99.62% **Benign & malignant** MIAS: Acc=99.28% DDSM: Acc=99.63% BCDR: Acc=99.60% **MIAS Normal & Abnormal** Acc=99.9%
Shen et al. (2020) [12]	- fully convolutional network (FCN) - Distinguish the domain to move the domain	**Private dataset** AUC=0.9083 **INBreast** AUC=0.8522				

Ansar et al. (2020) [13]	- CNN Models like AlexNet 2012 VGG-16 2014 VGG-19 2014 GoogLeNet 2014 ResNet-50 2015 MobileNet-v1 2017 MobileNet-v2 2018	DDSM CBIS-DDSM				**DDSM** MobileNet-v1 **Acc**=86.8% **CBIS-DDSM** MobileNet-v1 **Acc**=74.5%
Cao et al. (2021) [15]	BMass detection method based on FSAF network architecture to improve the robustness of the detection system.	INbreast DDSM CBIS-DDSM	FSAF for mass detection. Each image has an average **0.4950 false positive rate** for INBreast DDSM dataset for each image **0.5990 false positives**			
Naveed Chouhan et al. (2021)[16]	Miscellaneous features Existing breast cancer detection (DFeBCD) proposed system to classify a Mammography normal or abnormal. ELiEC file using Techniques: exception SqueezeNet	IRMA mammogram dataset.			**Xception** ROC-AUC=0.75% PR-AUC=0.75% **SqueezeNe** ROC-AUC=0.78% PR-AUC=0.85%	**Xception ACC**= 80.38 % **SqueezeNe ACC**= 81.07 %

Soham Chattopadhyay et al (2022)[17]	used the deep learning model, called Multi-scale Dual Residual Recurrent Network (MTRRE-Net) depended on four different magnification levels of BreakHis dataset.	BreakHis dataset.				40x=97.12% 100x=95.22% 200x=96.85%, 400x=97.81% to classify benign and malignant cells.

III. Experiments and Results

We collected 180 x-ray images from 50 Cases, and we used the mammogram dataset of Nasser Institute for Research and Treatment in Egypt. we increased this dataset by using the Augmentation Method till 780 images, and we divided our dataset based on Breast Imaging Reporting & Data System (BI-RADS) Classification as following:

class0 = 98 Images, class1 = 80 Images, class2 = 92 Images, class3 = 76 Images, class4A = 97 Images, class4B = 91 Images, class4C = 73 Images, class5 = 88 Images, class6 = 85 Images.

BI-RADS classification of Breast lesion according to the American college of radiology used for Breast assessment combined with recommendation			
Category & Recommendations			**Suspected Cancer Risk**
Category (0)	**Non-informative study.**		
Category (1)	**N**ormal study, routing screening.		**0%**
Category (2)	**B**enign findings, routing screening.		**0%**
Category (3)	**P**robably benign, for close follow up.		**2%**
Category (4)	**P**robably malignant, for biopsy		**50 %**
	A	mild suspicion.	
	B	moderate suspicion.	
	C	high suspicion.	
Category (5)	**M**alignant, for biopsy.		**98-100 %**
Category (6)	**P**athological proven malignant lesion		

Table 4 BI-RADS classification[19]

This figure shows the balanced dataset between the 9 classes Breast Cancer dataset after using the augmentation method and shows the number of images in each class depend on 6 categories of BI-RADS classification, given that the class 4 divided into 4A,4B,4C so the total is 9 classes.

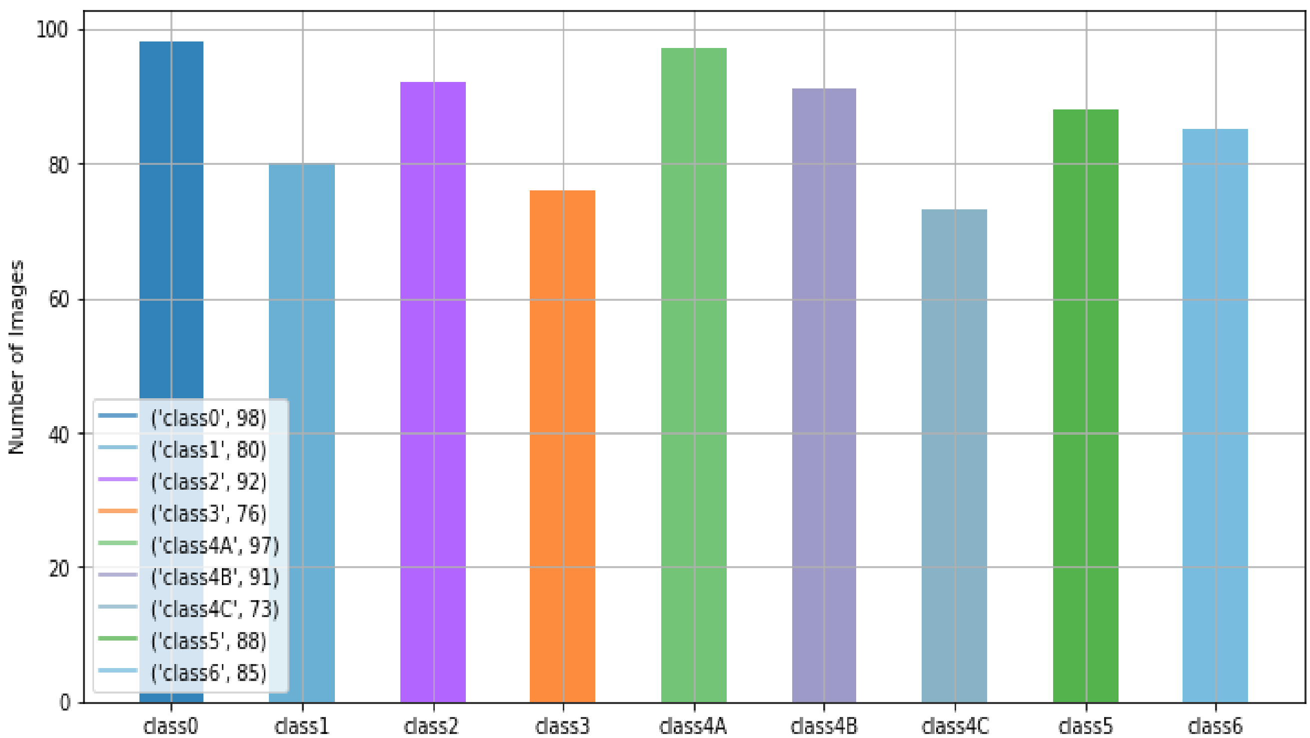


Figure 2 visualization of the balanced dataset

1-AlexNet Model Confusion Matrix

AlexNet uses 227 image size with Epochs = 20, Validation size=0.1, Random State=41.
(702 Training Images, 9 Classes) (78 Validation Images, 9 Classes)
and the Confusion Matrix results of validation images were the following:

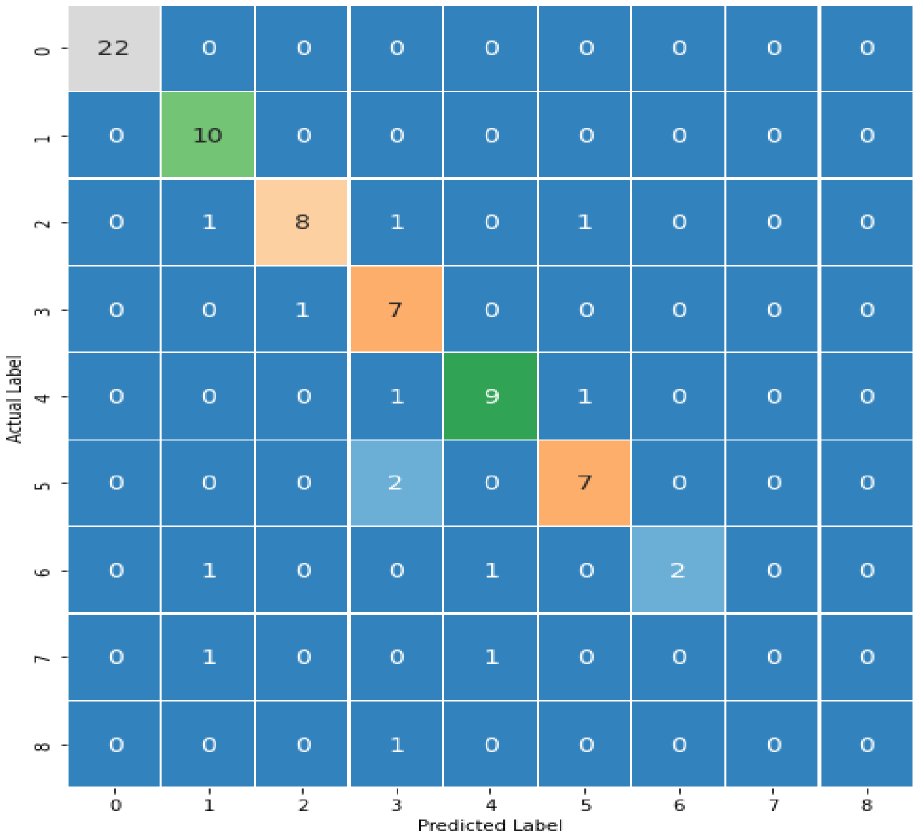


Figure 3 AlexNet Model Confusion Matrix of Validation Images

Table 5 AlexNet Metrics Results

AlexNet Metrics	Result
AlexNet Accuracy	83.33 %
AlexNet Precision	64.86%
AlexNet Recall	63.31 %
AlexNet F1 score	64.07%
AlexNet AUC	91.64%

Table 6 The Equations of these Metrics [20]

Accuracy =	$\frac{\mathbf{TP+TN}}{\mathbf{TP+TN+FP+FN}}$
Precision =	$\frac{\mathbf{TP}}{\mathbf{TP+FP}}$
Recall =	$\frac{\mathbf{TP}}{\mathbf{TP+FN}}$
F1 Score =	$\frac{\mathbf{2*precision*recall}}{\mathbf{precision+recall}}$
AUC=	$\int \boldsymbol{TPR\ d(FPR)}$

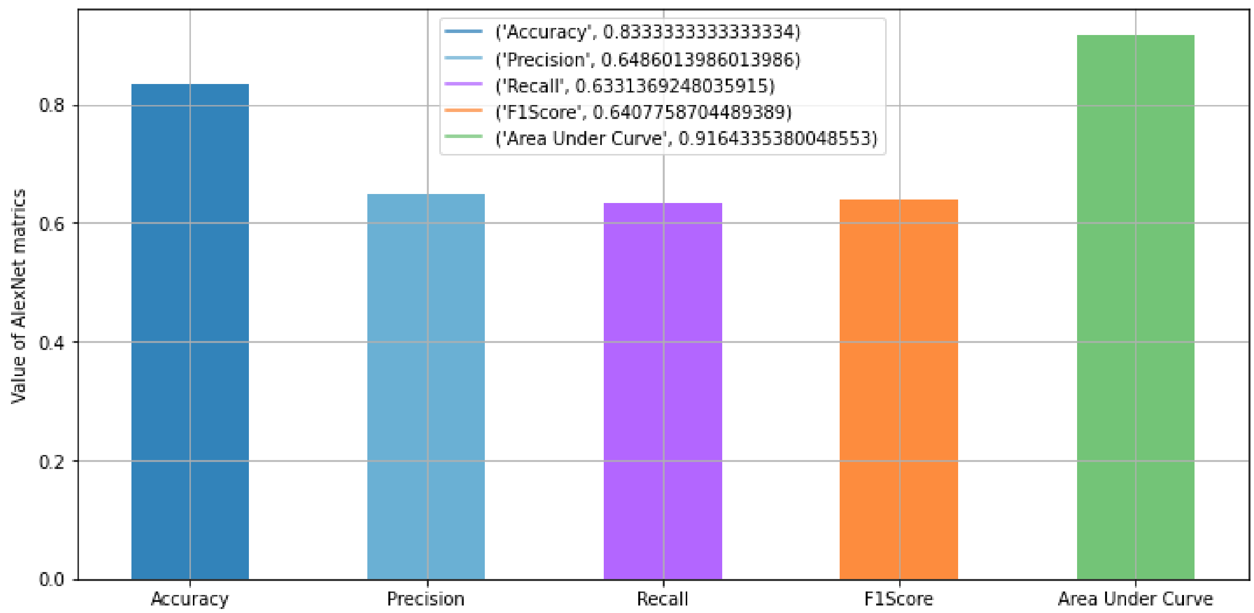


Figure 4 AlexNet Matrices Summary

Area Under a Receiver Operating Characteristic Curve (ROC-AUC)

An area under a receiver operating characteristic curve (ROC-AUC) describes the sensitivity of a binary classification model to changes in its threshold. In other words, ROC-AUC determines how well a model will classify a given set of test data as positive if "true" positives are changed in the model's decision function. It is also sometimes called an area under the ROC curve (AUROC) or just an AUC for short. The ROC-AUC is often measured using Hanley's approach, but there are several other methods as well. It also can be visualized graphically as a plot of sensitivity versus 1-specificity at various thresholds in a continuous scale from -1 to +1. The graph is then called a receiver operating characteristic (ROC) curve, which may also be termed a "receiver operating characteristic [20].

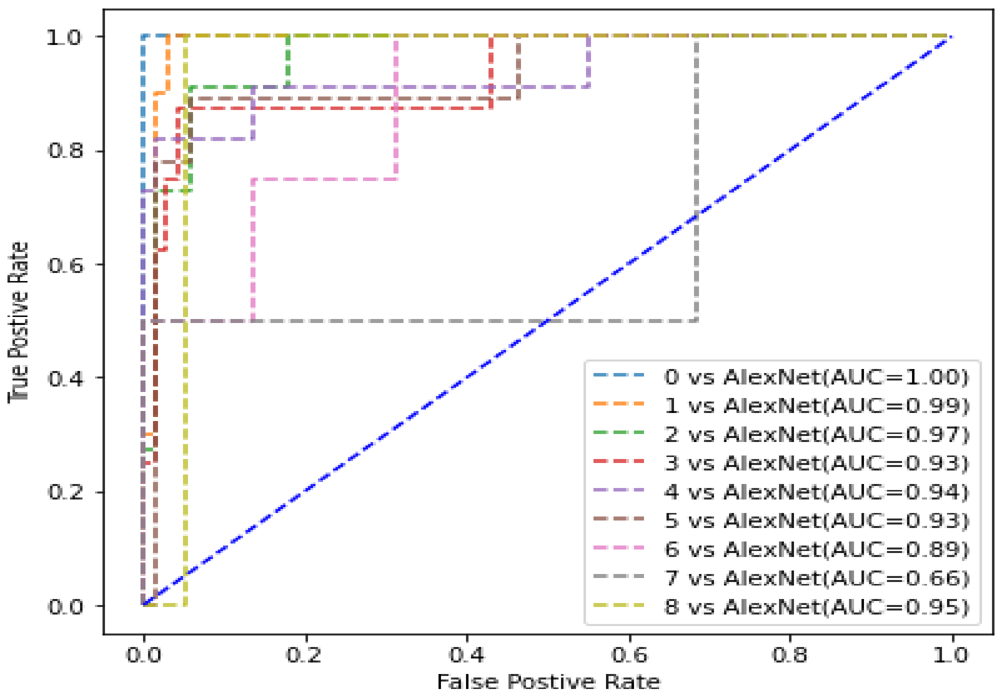


Figure 5 AlexNet ROC Curve

2-GoogleNet Model Confusion Matrix

GoogleNet uses 299 image size with Epochs = 20, Validation size=0.1, Random State=41.
(702 Training Images, 9 Classes) (78 Validation Images, 9 Classes)
and the Confusion Matrix results of validation images were the following:

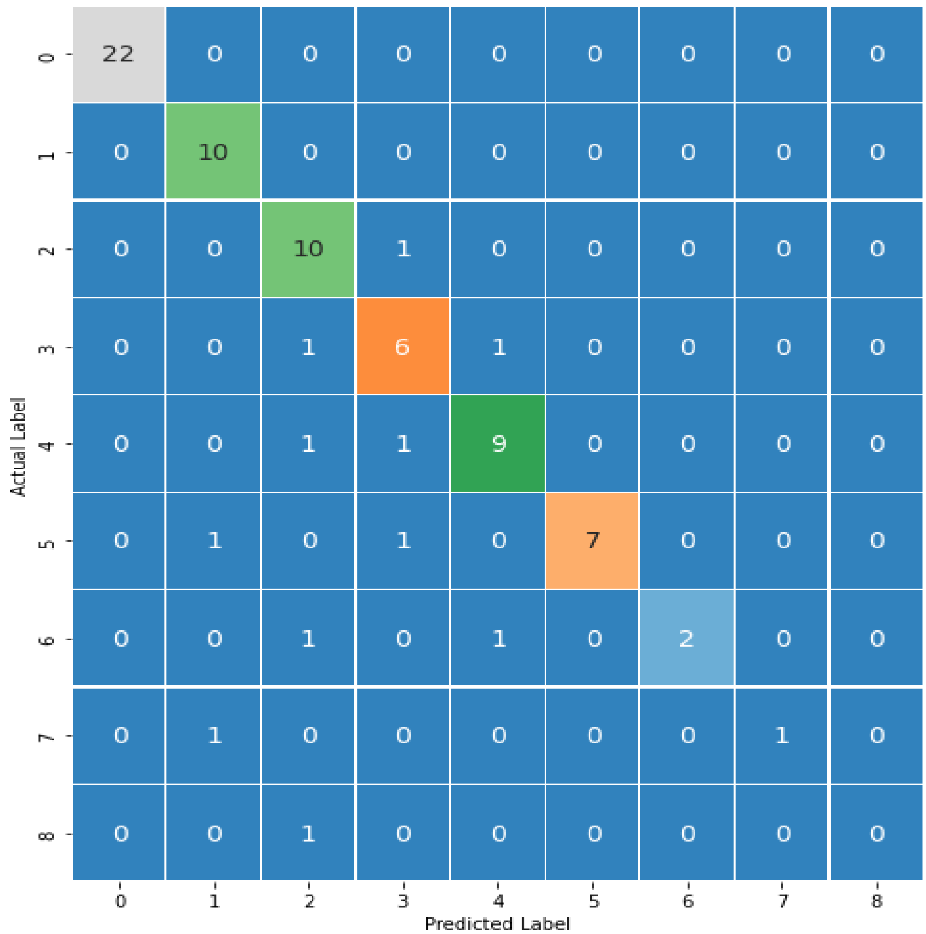


Figure 6 GoogleNet Model Confusion Matrix of Validation Image

Table 6 GoogleNet Matrices Results

GoogleNet Metrics	Result
GoogleNet Accuracy	85.89 %
GoogleNet Precision	87.13 %
GoogleNet Recall	69.50 %
GoogleNet F1 score	73.56%
GoogleNet AUC	96.08%

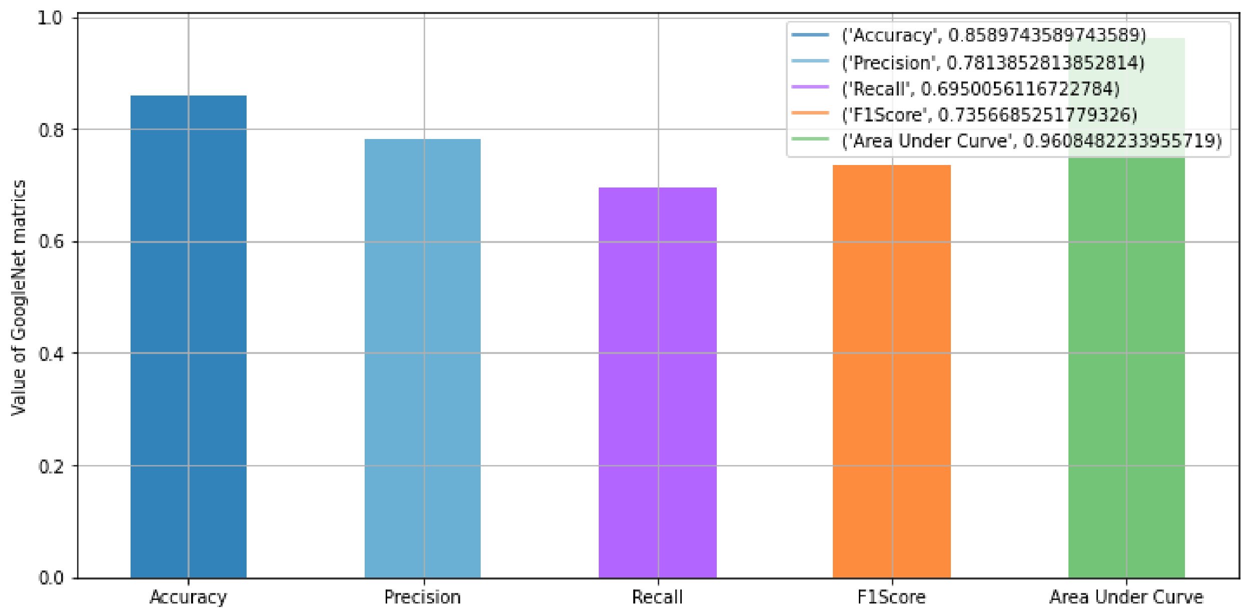


Figure 7 GoogleNet Matrices Summary

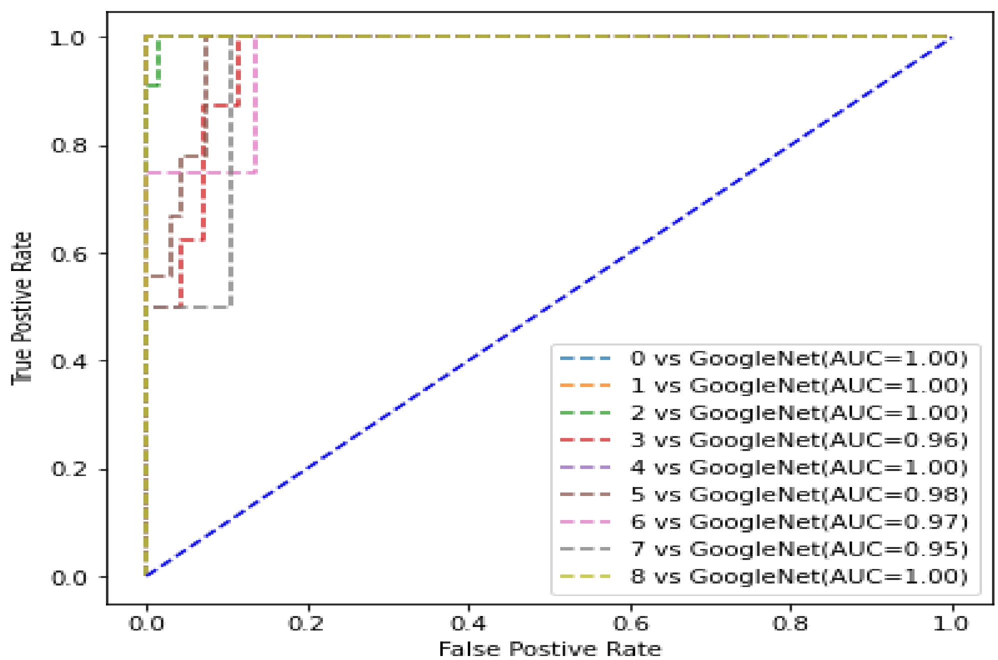
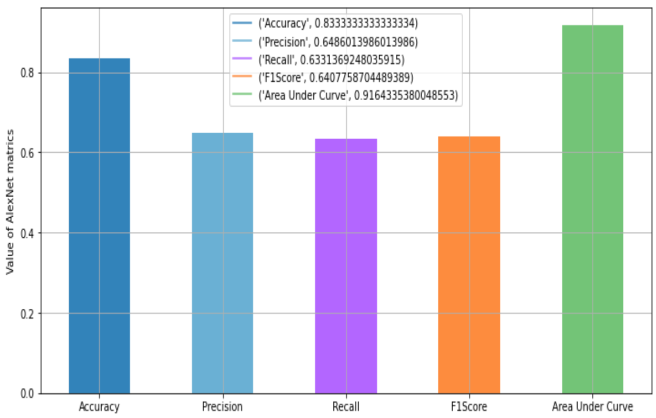
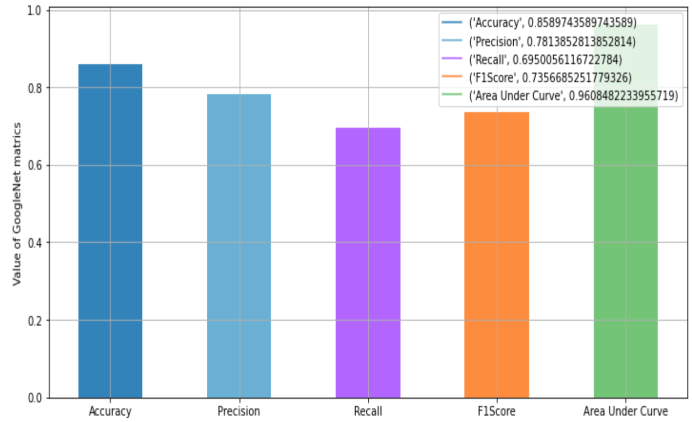


Figure 8 GoogleNet ROC Curve

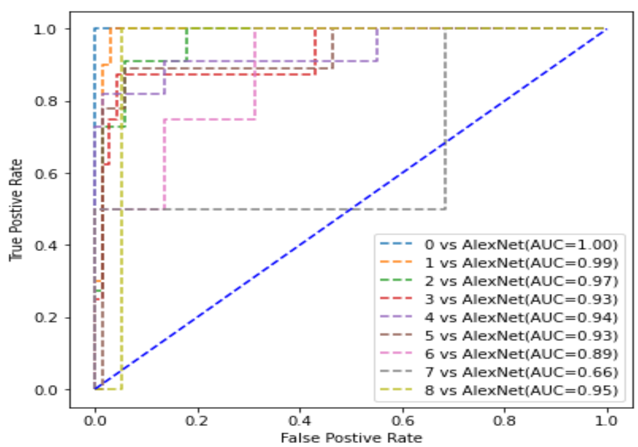
Table 7 comparison between AlexNet and GoogleNet in terms of the same performance metrics and ROC.



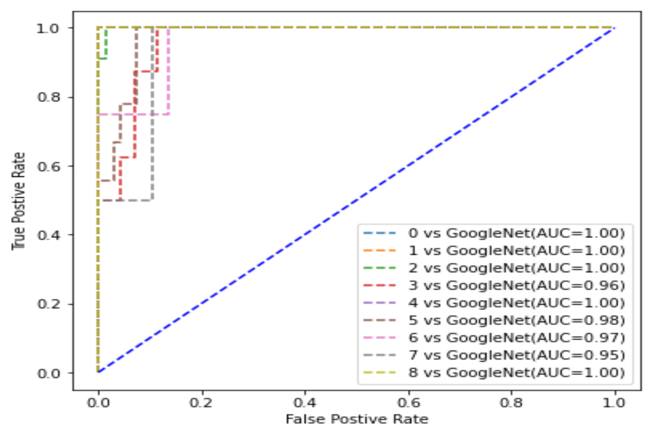
AlexNet Matrices Summary



GoogleNet Matrices Summary



AlexNet ROC



GoogleNet ROC

IV. CONCLUSION

In this paper we used deep learning to predict Breast Cancer. AlexNet and Google Net Models have been used the Egyptian X-Ray dataset collected from the mammography unit of oncology center of Nasser Institute for Research and Treatment, we are collected 180 images then increased this dataset by using the Augmentation Method till 780 images collected from 50 Cases. The result of comparison between AlexNet and GoogleNet Models, it's clear that Accuracy of Google Net is better than Accuracy of AlexNet.

V. REFERENCES

[1] World Health Orgnization, "Estimated number of deaths from 2020 to 2040, Both sexes, age [0-85+] Breast Cancer-Table." https://gco.iarc.fr/tomorrow/en/dataviz/bars?mode=population&cancers=20&populations=900_818&types=1.

[2] World Health Orgnization, "Estimated number of new cases from 2020 to 2025, Both sexes, age [0-85+]." https://gco.iarc.fr/tomorrow/en/dataviz/bars?mode=population&cancers=20&populations=903_904_905_908_909_935_818&years=2025.

[3] I. Singh, G. Goyal, and A. Chandel, "AlexNet architecture based convolutional neural network for toxic comments classification," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 9, pp. 7547–7558, 2022, doi: 10.1016/j.jksuci.2022.06.007.

[4] C. Szegedy *et al.*, "Going deeper with convolutions," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June, pp. 1–9, 2015, doi: 10.1109/CVPR.2015.7298594.

[5] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, "INbreast: Toward a Full-field Digital Mammographic Database.," *Acad. Radiol.*, vol. 19, no. 2, pp. 236–248, 2012, doi: 10.1016/j.acra.2011.09.014.

[6] R. A. Castellino, "Computer aided detection (CAD): An overview," *Cancer Imaging*, vol. 5, no. 1, pp. 17–19, 2005, doi: 10.1102/1470-7330.2005.0018.

[7] M. A. Al-antari, M. A. Al-masni, M. T. Choi, S. M. Han, and T. S. Kim, "A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification," *Int. J. Med. Inform.*, vol. 117, no. April, pp. 44–54, 2018, doi: 10.1016/j.ijmedinf.2018.06.003.

[8] E. A. Scuccimarra, A. D. S. at N. R. Center, and S. Lausanne, Vaud, "DDSM Mammography." https://www.kaggle.com/datasets/skooch/ddsm-mammography.

[9] J. Parker, I. Ricketts, J. Savage, and E. Stamatakis, "MAMMOGRAPHIC IMAGE ANALYSIS SOCIETY Mammographic Database," no. August, 2015.

[10] D. C. Moura and M. A. Guevara López, "An evaluation of image descriptors combined with clinical data for breast cancer diagnosis," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 8, no. 4, pp. 561–574, 2013, doi: 10.1007/s11548-013-0838-2.

[11] F. Mohanty, S. Rup, B. Dash, B. Majhi, and M. N. S. Swamy, "An improved scheme for digital mammogram classification using weighted chaotic salp swarm algorithm-based kernel extreme learning machine," *Appl. Soft Comput. J.*, vol. 91, p. 106266, 2020, doi: 10.1016/j.asoc.2020.106266.

[12] R. Shen, J. Yao, K. Yan, K. Tian, C. Jiang, and K. Zhou, "Unsupervised domain adaptation with adversarial learning for mass detection in mammogram," *Neurocomputing*, vol. 393, pp. 27–37, 2020, doi: 10.1016/j.neucom.2020.01.099.

[13] N. Baker Al Barghuthi and H. E. Said, *Sentiment analysis on predicting presidential election: Twitter used case*, vol. 1187 CCIS. 2020.

[14] L. J. Biffi *et al.*, "Article atss deep learning-based approach to detect apple fruits," *Remote Sens.*, vol. 13, no. 1, pp. 1–23, 2021, doi: 10.3390/rs13010054.

[15] H. Cao, S. Pu, W. Tan, and J. Tong, "Breast mass detection in digital mammography based on anchor-free architecture," *Comput. Methods Programs Biomed.*, vol. 205, 2021, doi: 10.1016/j.cmpb.2021.106033.

[16] N. Chouhan, A. Khan, J. Z. Shah, M. Hussnain, and M. W. Khan, "Deep convolutional neural network and emotional learning based breast cancer detection using digital mammography," *Comput. Biol. Med.*, vol. 132, no. March, p. 104318, 2021, doi: 10.1016/j.compbiomed.2021.104318.

[17] S. Chattopadhyay, A. Dey, P. K. Singh, D. Oliva, E. Cuevas, and R. Sarkar, "MTRRE-Net: A deep learning model for detection of breast cancer from histopathological images," *Comput. Biol. Med.*, vol. 150, no. August, 2022, doi: 10.1016/j.compbiomed.2022.106155.

[18] B. P&D Laboratory – Pathological Anatomy and Cytopathology, Parana, "Breast Cancer Histopathological Database (BreakHis)," 2019. https://www.kaggle.com/datasets/ambarish/breakhis.

[19] Harmien Zonderland and Robin Smithuis and the N. Radiology department of the Academical Medical Centre in Amsterdam and the Rijnland hospital in Leiderdorp, "Bi-RADS for Mammography and Ultrasound 2013 Updated version," *Radiol. Assist.*, doi: https://radiologyassistant.nl/breast/bi-rads/bi-rads-for-mammography-and-ultrasound-2013.

[20] Turing, "Understanding AUC-ROC Curves and Their Usage for Classification in Python." https://www.turing.com/kb/auc-roc-curves-and-their-usage-for-classification-in-python

Using Blockchain Artificial Intelligence in Healthcare Cybersecurity - Case Study on Organ Trading Using Blockchain –

Khalid Litim[1], Rania Boubidi[2]
[1] k.litim@univ-jijel.dz
[2] rania.boubidi@univ-jijel.dz
Economics of organizations and sustainable development laboratory, University of Jijel (Algeria)

Abstract:

The purpose of this paper is to investigate how blockchain technology can be used to transform cyberattacks on the healthcare industry by applying it to organ trading. This is accomplished through a review of the literature on cyberattacks and artificial intelligence applications in healthcare.

Electronic healthcare records contain sensitive information that must be effectively protected from cyber-attacks. To address this issue, a security system based on smart contract technology embedded in the blockchain network is required. The study found that using Blockchain technology for organ donation creates reliable, time-stamped, and cryptographically linked records, and that the transparency and immutability of blockchain chains for organ donation improve traceability and deter organ trafficking. According to studies, this technology helps to improve the efficiency of health-care operations by shortening the transplantation process compared to the traditional transplantation process.

Keywords: Digital Health, Artificial Intelligence, Blockchain, Cybersecurity. Organ trading.

INTRODUCTION

As a result of the Covid-19 pandemic, there has been a push toward the use of digital health care platforms like mobile phone apps to contact doctors and order medicines via the Internet, as well as a general trend towards the establishment of remote treatment traditions and health prevention guidelines. According to the pandemic scenario, the digital health care sector faces numerous opportunities and challenges, and in 2020, the health care sector is hit hard by the Covid-19 pandemic. Digitization is widely acknowledged to be important in the healthcare system due to its contributions in facilitating patient-doctor communication remotely, lowering costs through the use of information, communication, and technology, facilitating the storage and electronic access of medical records, and assisting digital healthcare platforms and mobile applications in monitoring health conditions.

Conversely, the expansion of digital technologies into the medical field has spawned novel difficulties. Consumers' use of digital platforms to improve their health presents a number of challenges, one of the most significant being the security of their data. Privacy concerns, data breaches, and identity theft are all examples of this expanding set of problems. Beginning in 2020, cases involving the healthcare sector are expected to be among the five largest cybercrime cases, according to statistics and evidence. Chinese hackers, for instance, broke into the Spanish Research Center's files on the Covid-19 vaccine while taking advantage of the current epidemiological crisis. It is estimated that by 2026, the cybersecurity industry will lose $27.1 billion due to the difficulty of combating cybercriminals' efforts to gain access to sensitive healthcare information. Cybercrime poses a significant threat to digital healthcare due to the high financial rewards for breaching patient data. In this context, hackers misuse the benefits and obstacles presented by the foundational system of electronic medical records in order to extort money from victims through means like ransomware and phishing.

One cannot fathom the urgency and desperation a person feels when a loved one needs an organ to continue living and they cannot locate a suitable donor, which is why organ donation is the noblest act that requires a revolution. Because there aren't enough donors to go around, desperate patients are forced to look for alternative ways to get their hands on transplant organs, which only serves to strengthen the underground market for organs. Organ transplants are complex; organs deteriorate rapidly; and the donated organs must be from a person with a compatible blood type. On the other hand, people who want to donate, if not trade, their organs face the problem of maintaining identity confidentiality. Mortgaging the future of life-saving organs, the demand for transplantation has skyrocketed in the past decade as organ failure has become more common and transplantation has become more successful around the world. More than 120,000 people are on the transplant waiting list, and 22 people die every day; this has prompted an increase in the trafficking of human organs, as evidenced by an increase in cases of electronic data piracy of them.

Study problem:

Most people agree that the healthcare industry is vulnerable to cyberattacks because it has a lot of personal and financial information that hackers want in exchange for their high-tech skills. Because of this, it is very important to act quickly to stop the effects of these crimes. When making healthcare software, companies need to take extra steps to protect their customers' personal information, follow confidentiality rules, and deal with privacy and cyber security issues. On the other hand, the digital world led to the creation of new products that helped businesses get close to their customers. Recently, Blockchain technology changed the digital world by giving us a new way to think about security, reliability, and how well systems work. On the other hand, managing organ donation groups is very important when it comes to making sure that organs are available and that the list of donors and recipients is managed correctly. It's also important to talk about how well the two people's health and personal choices, like religion and hobbies, fit together. In this case, giving this management to blockchain platforms could be helpful.

The problem of this study can be formulated in the following main question:

What role does blockchain technology play in establishing rules and standards in the organ trade and healthcare sectors?

Study objectives:

The goal of this paper is to look at digital health systems that can't be hacked by using blockchain technology and that can be trusted and checked by using a decentralized network. This is done by taking a look at the digital healthcare sector as a whole and the level of cyber security in the healthcare sector in particular. In addition to identifying cybercrime problems in the health care sector, the next step is to understand the scope of cybercrime in health care and figure out how important the factors are that affect security measures or precautions in digital health. Also, the goal of the research is to figure out how cybersecurity BLOCKCHAIN platforms work in health care and to figure out how much BLOCKCHAIN platforms help with health electronic crimes, especially crimes involving human organ trafficking.

study importance:

In medical practice, digital health technologies like telemedicine, ambulatory health, and telemonitoring are very important. Safely managing medical information improves both digital health and public health. Digital health can also help keep costs down by making it easier for people to get care and connecting them with healthcare providers. Digital platforms and mobile apps give patients and healthcare providers the tools they need to monitor and treat medical conditions in real time, no matter where the patient or healthcare provider is located. Also, mobile health data is stored on servers, so it is important for both medical practice and clinical trials to develop software that manages data in a way that is similar to blockchain technology and helps stop all kinds of tampering.

Study Approach:

The study used a descriptive-analytical method based on a review of the most important literature about cyber security in health care and how blockchain technology affects this issue. The study was approved after a careful look at some of the research, which included suggestions for how blockchain could be used in healthcare. The literature that was chosen was found in international information databases.

Literary review:

Many studies have looked at how blockchain can be used in healthcare, and most of them have focused on electronic medical records (EMR) and telemedicine monitoring procedures (RPM), as well as clinical research and supply chains. D. V. Dimitrov [1], talked about some of the ways blockchain technology could be used, such as data management for electronic medical records (EMR), data management platforms for personal health records (PHR), and a healthcare genomics blockchain. Oderkirk and Slawomirski [2] looked at how important it is to use Blockchain to manage medical supply chains. This includes identifying products and using a distributed ledger that automatically checks the authenticity of products and related information. In addition to warning stakeholders about products that don't meet standards or pose a high risk. The study by Mettler and Hsg [3], which focused on clinical research, showed that research and development (R&D) and drug production in the pharmaceutical industry are good places to use Blockchain technology. This technology can be used to keep an eye on how drugs are made and deal with the problem of fake medicines. Alandjan [4] says that the intelligent communication tools and hashes of Blockchain technology help make it safe for the organ donor, the patient, and the whole healthcare unit to talk to each other and share information. In references [5] [6], cybersecurity in Blockchain health care was limited to the security of the Internet of Medical Things, decentralization of medium storage, remote consultations, and mobile health. To answer these questions, we decided to split this research paper along the following axes:

I. The issue of digital security in health care

1. Concept of digital healthcare

In the past few years, there has been a big rise in the use of electronic health care practices, which are now called "digital health." The World Health Organization defines "digital health" as "the safe and cost-effective use of information and communication technologies to support health and health-related areas, such as health care services, health surveillance, health documentation, health education, knowledge, and research" [7]. The World Health Organization has written about the idea of "digital health" in recent documents. This idea includes e-health, including the use of mobile wireless communications (mHealth), as well as other active new areas like the collection and processing of "big data," computer technologies used in genomics, and artificial intelligence. The Global Strategy for Digital Health (2020-2025) says that digital health is "the area of knowledge and practice related to the development and use of new digital technologies to improve health" [8]. So, Internet of Things (IoT) applications can help solve the problems of protecting the health of citizens.

The World Health Organization puts a lot of emphasis on the development of digital health. Its goals and principles show what it stands for. On May 26, 2018, the World Health Assembly passed a resolution about digital health. The resolution asked Member States to "consider how digital technologies can be integrated into existing health systems, infrastructures, and regulatory frameworks." To move forward with national and global health priorities by improving existing platforms and services to strengthen the social orientation of health care and disease prevention measures and to reduce the burden on health systems. People were also told that digital health should "become an important part of health priorities and help people from an ethical and safety point of view." ICTs must be used in a way that is "safe, reliable, fair, and sustainable," and digital health innovations must be built on principles like openness, accessibility, interoperability, privacy, and security [9].

2. The concept of healthcare cyber security

Today, anyone can send or receive any kind of information by email or video with the click of a button. Have you ever thought about how to send that information to someone else in a safe way? This is what cyber security is about [10]. Even when we do everything right, we can't keep our information safe because of new technologies. As a result, cybercrime is on the rise. At the moment, more than 60% of financial transactions happen over the Internet. Because of this, the industry needs the best security for a lot of transactions. Because of this, cybercrime in these areas is a very serious issue for society [11].

A cyber-attack is the intentional use of computer systems and businesses that depend on technology and networks. This can cause problems with data and lead to cybercrimes like identity theft [6]. "Cybercrime" is any illegal act where a computer is used to gather evidence, says the US Department of Justice [10]. Cybercrime is the illegal use of information and communication technologies, such as the Internet, malicious software, unauthorized access to websites, copyright violations, and the theft of credit card numbers.

Cyber security is a big problem because bad things that happen online are getting worse and more people, businesses, and governments are at risk. In the Global Risks Report 2021 from the World Economic Forum, cyber security failures were listed as clear and present risks. Also, as the value of some assets has gone up, people who want to hack systems for financial, psychological, or reputational gains are more likely to do so. In 2011, Mulligan and Schneider suggested that cybersecurity be seen as a public good and managed as such. The idea of cybersecurity as a public good is meant to remind everyone that they are all responsible for making cybersecurity better. To reach the SDGs in societies that depend on digital infrastructure and are highly connected, it is important to make sure cybersecurity through more awareness and strong partnerships between different groups.

3. Healthcare cybercrime issues

According to the 2022 Medicare Network Safety Statistics, the medical services industry alone has lost $25 billion in the last two years. Cyber Risk Management's (CyRiM) report says that Medicare companies are likely to be among the companies that cyberattacks will affect [7]. A report by the online security group Critical Insights says that cyberattacks on healthcare have reached a level that has never been seen before, putting a record number of patients' Protected Health Information (PHI) at risk. In 2021, attacks on medical services hurt 45 million people, which is more than the 34 million people who were hurt in 2020. This number has gone up a lot in just three years, from 14 million in 2018 to 42 million in 2022 [7], because India has seen a lot of cybercrimes [12]. In September 2020, hackers from China stole information about the Covid-19 vaccine from a Spanish research center. Here are the different kinds of cybercrime that happen in the healthcare sector [7]:

- **Ransomware:** ransomware is a type of malware that infects systems and files and makes them useless until the amount demanded is paid. Other industries can wait a while if their system is broken, but not the healthcare industry. It's a big problem if a clinic can't get to their records. Digital tasks that have to do with healthcare slow down or stop working altogether. The Maryland Department of Health said that emergency clinics were having trouble because of a ransomware attack in the middle of the COVID-19 crisis.

- **Data breaches:** data breaches happen when data is taken or changed without the owner's permission. Between 2009 and 2021, there were 4,419 Medicare data breaches involving at least 500 records. These breaches led to the theft of 314,063,186 Medicare records. Also, very important information like social security numbers, dates of birth, and addresses were among the records that were lost.
- **Phishing attacks:** phishing attacks are when attackers send malicious emails to try to trick people into giving them financial information, access to their company's email, or other important information by making it look like the message came from their company.

II. Healthcare blockchain artificial intelligence applications

Blockchain is the technology behind Bitcoin, Ethereum, Dogecoin, and every other cryptocurrency. It has changed the way the digital economy works. It is a technology that uses the Internet to exchange, distribute, and store digital information. This is done by connecting and distributing blocks of data across the network of this technology.

1. Blockchain technology concept

A blockchain network is a decentralized peer-to-peer (P2P) network that sends all the information needed to support a cryptocurrency system. The Bitcoin system is one of the most popular cryptocurrencies. It works on a network [13], which is made up of many blocks linked together to form a blockchain. Each block contains many transactions that are put together using "cryptographic" encryption techniques to keep them safe [14]. As can be seen in figure 1 [13].

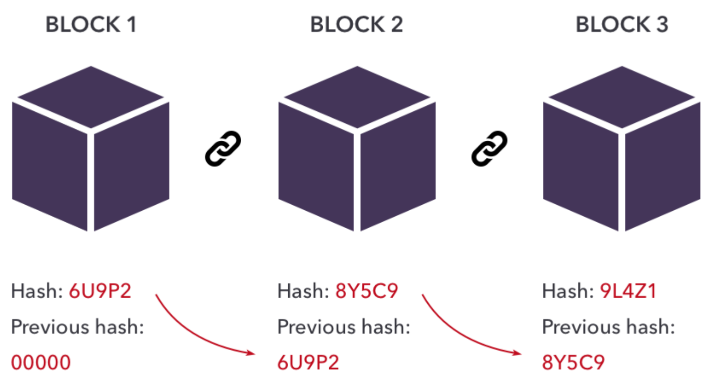


Figure 1: The blockchain network

A blockchain has three main parts: a block, which is a list of all the transactions recorded in a ledger during a certain time period, a chain, which is a hash that links one block to the next, and a node, which is a computer that is part of the blockchain. A network of "full nodes" and the information in the previous block are used to create the hash in the blockchain. Each node has a full record of every transaction that has ever been added to the blockchain [15]. A process called "mining" links the blocks together in the chain. This process turns pending transactions into a math puzzle that is solved by computer systems. The result is a string of letters and numbers in the block that are all different from each other. Each block has the transaction data, a timestamp, and a cryptographic hash of the block before it [16]. Each blockchain has its own set of rules for how its network can agree on new entries. Some blockchains can be used to make money, others can be used to store data, and others can be used to keep systems and contracts safe. There are many different ways to get people to agree on something [15]. The following figure shows one of the models [17]:

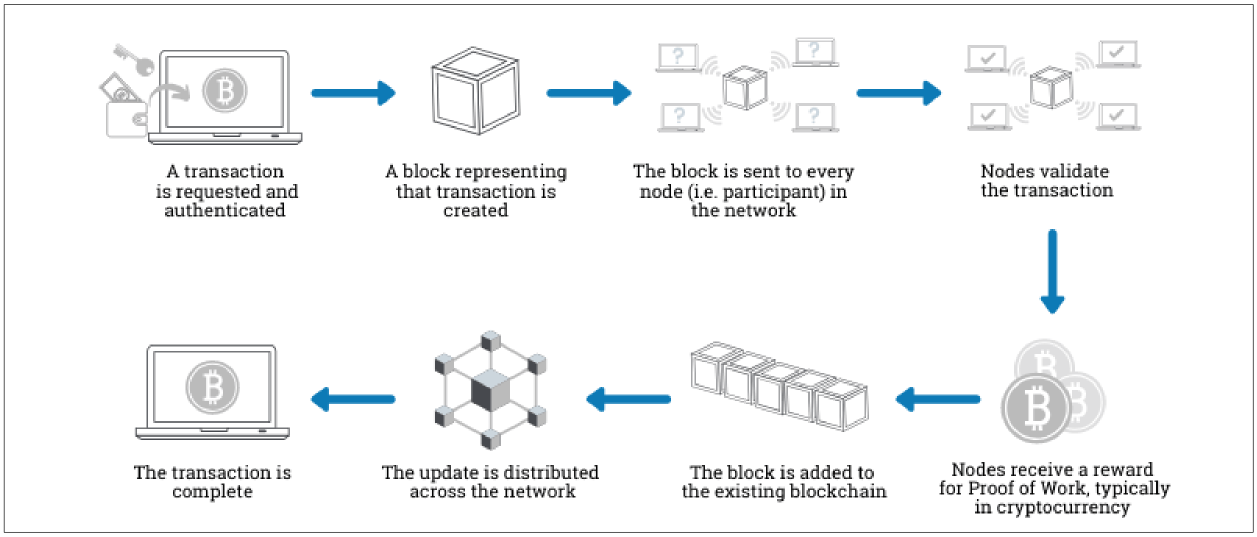


Figure 2: How does the blockchain work in the network

A blockchain is a special kind of database that can only be read. This is how it was made and how it is defined. This means that blockchain databases are only meant to be added to. They can't be changed or taken away [1].

2. Concept of artificial intelligence

The Oxford American Dictionary says that "artificial intelligence" is the study of "getting computers to act like smart humans." [17] It was also defined as: "The ability of computers to do things that only humans can do, such as pattern recognition, inference, and speech. It also refers to the ability of machines to learn and make decisions based on what they have learned." [18] It also means "a branch of computer science that focuses on making smart machines that act and talk like people." Artificial intelligence lets machines learn from their own mistakes, adapt to new information, and do things that humans do [19]. So, artificial intelligence means machines that are smart enough to make decisions that look like human intelligence. Because of how important they are in the financial world, smart contracts and cryptocurrencies are becoming more and more popular in business as financial technology is used and improved.

A smart contract is a codec that reflects computer protocols that check the performance of a contract, negotiate or simplify it, or enforce or cancel unexpected clauses in the contract [20]. There are many transactions between authorized parties that are carried out by computer codes. The size and complexity of these transactions vary greatly. As one of the most important parts of blockchain technology, a smart contract is a piece of software on a network (blockchain) that makes it possible for agreements between parties to be carried out automatically [21]. From this, we can say that a smart contract is a piece of code in the blockchain network that does a lot of things to make sure transactions between the right people happen.

Blockchain technology is a new technology that can change how e-commerce works. Blockchain technology can be used to make improvements in many areas, such as the financial sector, e-government, privacy, and security [6]. Both the network (blockchain) and artificial intelligence can help the pharmaceutical industry. This includes making clinical trials more likely to be successful and looking into how the supply chain for drugs works. When you put together smart data analysis and a decentralized framework for clinical trials, you have a chance to integrate and share data. And the growing trend of using historical data about supply puts a lot of pressure on businesses because of their receivable account balances, especially for small and medium-sized businesses. Because billing is the main form of the supply chain and accounts receivable are used for financial transactions, billing plays a very important role in the supply chain. Business billing procedures are an important part of the supply chain, and blockchain-based smart contract technology has a lot to offer in this area. This includes the division and transfer of creditors' rights certificates, factoring financing for upstream suppliers, payment due to core institutions, and the development of their implementation processes [22].

3. The Health Sector and Blockchain: Potential Applications and Use Cases

Most people agree that any information about a person's health, including demographic and genetic information, that can be used to identify that person is protected health information. Even though the use of the cloud to create a better database to protect medical information is growing quickly, this won't solve the problem of privacy violations by itself. Researchers think that using Blockchain technology to get around the problem mentioned above would lead to better care for patients. This is mostly because it can protect shared medical data in a clear way. Because of this, the health care industry has become more efficient and spent less money. Many studies have looked at how blockchain can be used in healthcare, and most of them have focused on electronic medical records (EMR) and remote patient monitoring (RPM) procedures, as well as biomedical and pharmaceutical research, supply chain and data analytics, insurance ordering, and medical reviews. Most articles talk about the different ways blockchains can be used in healthcare and their use cases:

- **Blockchain Medical Data Management:** This includes electronic medical record (EMR) data management and personal health record (PHR) data management platforms. As personal health records, wearable sensors or medical IoT devices have recently started collecting personal life history data (PHR). AI-powered healthcare analytics are sent in real time to the people who need them, such as patients, doctors, and the pharmaceutical industry. When Blockchain is used for healthcare genomics, a blockchain platform company called Timi Inc says that a single patient's data is worth up to $7,000 per year. For example, a company called Nebula Genomics offers free complete gene sequencing. Once a user has their genome sequenced, they can charge tokens to anyone who wants to see it. Then, you can use those tokens to buy more tests and products that will help you figure out what the DNA means [1].
- **Blockchain Medical Supply Chain Management:** Blockchains are now being used to manage the supply chains for medicines, clinical supplies, blood products, and medical devices in the health sector. In this area, blockchain can be used to identify products, use a distributed ledger that automatically checks relevant information, and verify the authenticity of products. The blockchain system also helps regulators find out about products that don't meet standards, are fake, or are high-risk [2].
- **Blockchain Clinical Research:** Blockchain technology can also be used in research and development and drug production in the pharmaceutical industry. This is because Blockchain technology can be used to keep track of how drugs are made. In this situation, the problem of fake medicines is an important and urgent one. The World Health

Organization says that 10 percent of medicines around the world are fake, and that number jumps to 30 percent in developing countries [3]. Because fake medicines bring in a lot of money, drug crimes are growing very quickly. The World Health Organization says that sales of fake drugs could bring in as much as $200 billion in the global pharmaceutical market [23]. The pharmaceutical industry needs a strategy that is clear, reliable, and can't be changed so that medical products can be tracked from the manufacturer to the consumer [24].

This is in addition to the use of blockchain technology in the field of organ transplantation, which we will discuss in detail in the next chapter.

III. The impact of Blockchain technology on healthcare cybersecurity

1. Blockchain Cybersecurity (BCT)

BCT is popular because it has security features that make it safer, like the fact that it is distributed. Bitcoin is the most important use of Blockchain. Bitcoin is a digital currency that is worth more than real gold right now. Big companies worry more about how to handle a particular cyberattack than about getting ahead of the crime [25]. Studies have found that cybersecurity problems have been looked into the most in healthcare and smart cities. But there aren't many studies in the literature that focus on cybersecurity problems in BCT that are specific to certain industries, as shown in Figure 3 [25].

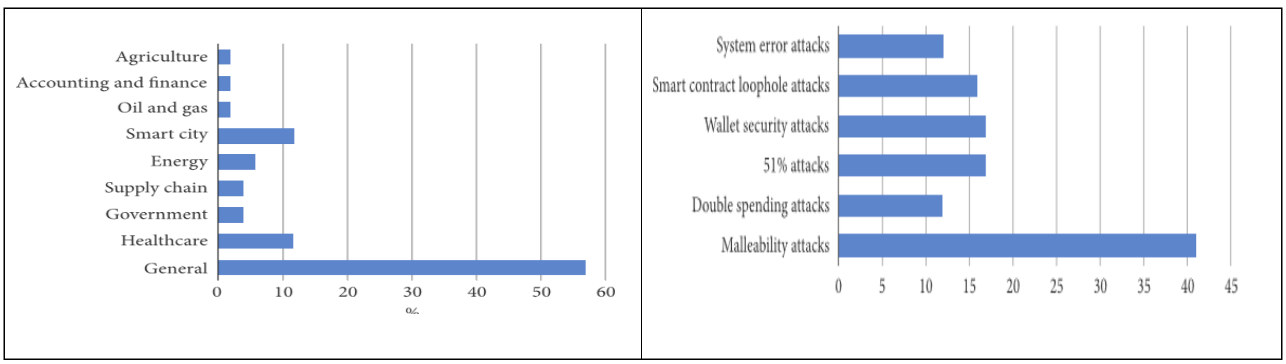


Figure 3: The scope and nature of blockchain cyber security issues in healthcare, broken down by industry

Studies show that wallet security attacks and 51% attacks are the next most common types of cybersecurity problems in BCT. As shown in Figure 3, there have also been smart contract vulnerabilities attacks, double spending attacks, and system bug attacks that were reported in BCT [25].

2. Cyber-Security Dimensions in Blockchain Healthcare

A lot of research has been done on how blockchain can be used in healthcare. Most of these studies have focused on electronic medical records (EMR) and remote patient monitoring (RPM) systems. Biomedical and pharmaceutical research, as well as supply chain and data analytics, medical reviews, and ordering insurance[5]. When it comes to measures to protect against electronic attacks and cyber security.

- **Teleconsultation and mobile health:** Many of the attacks have happened on social networking sites like Twitter and Facebook, which are used by doctors to plan how to talk to their patients. Too many records were lost or stolen, and customer information got into the hands of people who shouldn't have had it. If blockchain innovations were used everywhere in these media systems, they might stop these kinds of cyberattacks from happening in the future. Mobile phones are used in every home and by everyone every day in the world we live in now. Many groups are looking for ways to protect mobile apps from cyberattacks by using blockchain. Blockchain can be used to make a traceable, clear, and decentralized agreement, which can help make a mobile health infrastructure that can't be changed. In a recent study, researchers showed how to use Blockchain to spread messages about cybersecurity. They made a message confirmation model based on (SM2) to stop spoofing and replay attacks. Next, they made a cryptographic hash based on (SM3) to make sure the integrity of messages. Finally, they made a message encryption model based on (SM4) to protect customer security [7].

- **Medical IoT Security:** Because AI is always in the news, it's getting easier for programmers to work with big systems like hospital automation by using smart tools like "smart" switches. A large number of IoT devices have security points that could be better. Putting IoT information on the blockchain makes it safer. Because Blockchain uses high-level encryption, it is hard to get to personal records. In 2018, an article described a new method that combines the benefits of private key, public key, blockchain, and a few other native cryptographic tokens to make it easier for patients to agree to use electronic clinical records and to keep them safe [7]. There are three types of blockchain-based health care applications: data management, smart contract medical applications, and the medical internet of things. Data governance includes global scientific data sharing for research and development (R&D), data warehousing, data management, electronic health care records, and medical applications of smart contracts, such as clinical trials and medicines [6].
- **Decentralizing Medium Storage:** Hacking of healthcare information and ransomware give the healthcare industry important reasons to invest in cybersecurity. When this kind of attack happens, sensitive and secret information gets out, like financial records for healthcare and private patient information. With blockchain, "data can be kept safe by making sure that information is stored in different places. With this strategy for moderation, it will be harder for programmers to get into information capabilities frameworks. Several health care organizations are looking into how blockchain can be used to keep information safe from hackers. Blockchain technology is used to keep track of patient data and manage the supply chain of medicines, among other things [7]. One of the most important problems in developing countries is that fake medicines are sold that don't have the right amount of ingredients or ingredients that don't work. If medicines are checked through the blockchain network, fake medicines can no longer be made. RFID tags that are part of the block chain technology are used to prove ownership of a product from the manufacturer to the end user [5].

IV. Blockchain Transplantation: Mechanism and Added Value

Organ donation is the process of taking an organ from a donor, who may be alive or dead, and putting it into a person who needs it for their health [26]. Organ transplants are hard because organs break down quickly and must come from a person with the same blood type as the recipient. The University of Michigan Transplant Center says that a heart or lung transplant is usually done in less than 10 hours. This is done without a working system, which means that organs that could save someone's life are put up as collateral. More than 120,000 people are waiting for organ transplants, and 22 people die every day on average [16]. In order to learn more about how Blockchain can be used in health care, a study made an app that makes it easier to match organ donors and recipients. With the help of Blockchain technology, the app connects donors and patients and makes hospitals look like the middlemen between them [7].

1. **Blockchain mechanism in organ transplantation**

Blockchain technology's intelligent communication tools and hashes make it safe for the organ donor, the patient, and the healthcare unit to talk to each other and share information. Figure 4 shows all the steps of the organ donation process in a set order.

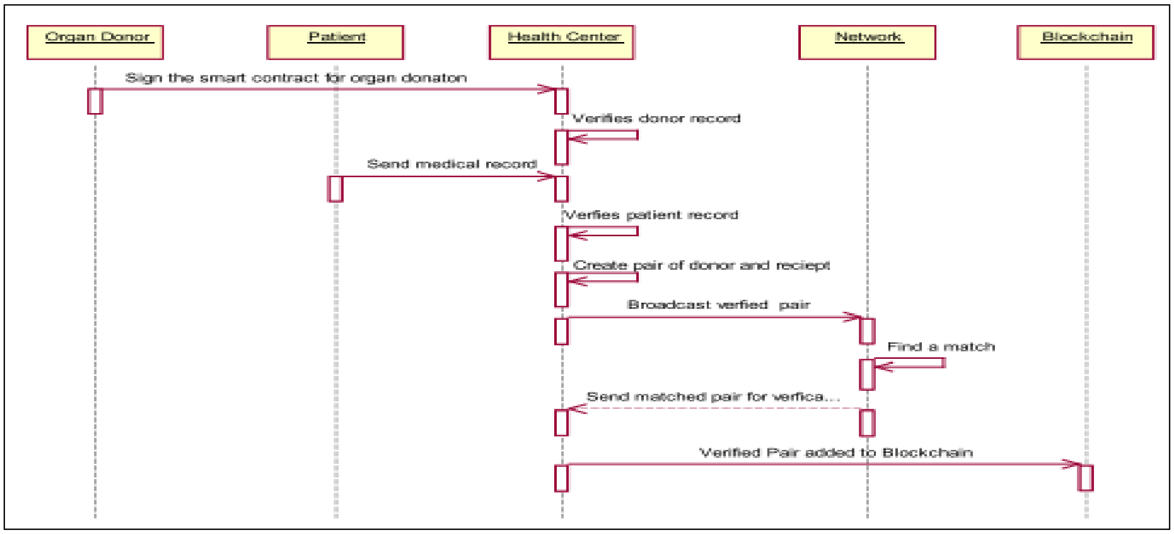


Figure 4: Sequence diagram of the organ donation system

The process of organ donation begins with an agreement between the organ donor and the patient. One of the two parties signs the document, which is then sent to the healthcare unit for verification of medical records and sample matching. After that, the system runs into two situations:

- If the samples don't match, the application will be turned down and a note from medical experts will explain why.
- If the samples are compatible, the patient's request will be sent to the network using digital signatures. If the request is approved, it will be added to the Blockchain, which will end this process and make it impossible to change anything else.

A blockchain-based organ transplantation system lets different entities share data in a safe, encrypted, and trusted way. Also, it gives access to the whole database, where the data is updated regularly by the entities and can be checked. No participant can control and process the data on their own. All communication happens between peer nodes, which store information and send it to all their neighbors since there is no central node to handle communication other than exchanging digital patient data directly from Electronic Health Record (EHR) transactions. Because of all of this, the Blockchain-based organ transplantation system can cut the time it takes to do a transplant by five times compared to the old way of doing things. Studies have shown that a Blockchain-based implant lasts for 14 hours, while a traditional implant lasts for around 60 hours [27].

2. **Blockchain transplant applications**

Blockchain Organtree is the first company in the world to have a decentralized organ donation database. It uses blockchain technology to connect donors, patients, and healthcare facilities

[16]. It also uses AI technologies to make the process of matching organ donors and patients more efficient [7]. It wants to set up a better and safer way for people to donate organs, check and match organs with patients, and use AI to improve the transplantation process [16]. Since data is kept in one place, switching to a distributed system will make it safer and stop hackers [28]. The first blockchain application for organ donation was made in the United Arab Emirates. The Ministry of Health worked with Dhonor Health Tech, a leading national company that provides blockchain technology for global healthcare, to use the "Hayat" app, in which the donor records his promise to donate a specific organ after his death. A list of people who need an organ is also part of the application. The blockchain will make sure that organs are checked by storing DNA swaps from donors. This will let hospitals check where the organ came from. Not only will this process make it more likely that transplants will work, but the audit trail will also stop people from trading organs [26].

The World Health Organization (WHO) says that there are about 70,000 people on waiting lists for kidney transplants. About 20,000 kidneys come from neighborhoods, and at least half of those are sold on the black market. Table 1 shows how much different body parts are worth on the black market [29].

Organ	Value ($)	Organ	Value ($)
Corneas	19,800	Liver	137,000
Skeleton	6,600	Blood	$630/$297/pint
Lungs	272,000	Skin	$1.24/cm³ ($8/in²)
Kidney	138,700	Bones + ligaments	4,800

Table.1: Black market value of various body parts($)

The kidney is the most popular organ to sell. This is because humans have two kidneys, so giving up one of them can bring in a lot of money, especially when a person is having trouble paying their bills. But it helps close the gap between the number of kidneys available and the number of people who need them.

People with kidney failure have to deal with a lot of problems, like being put on a waiting list and having to wait 3 to 5 years on average for a kidney transplant in the United States. During this time, patients have to have heavy dialysis at least three times a week, which is very expensive compared to a transplant [30]. Every year, 4,000 people on waiting lists die in the United States [30]. Kidner has created a "kidney exchange" or "double kidney donation" system based on the blockchain. For example, if a person wants to donate a kidney to a family member but the kidney is not compatible with the family member, the system matches the donor kidney with another patient who also has a non-compatible donor kidney. This makes it possible to extend the bilateral exchange to all pairs of unmatched donors and recipients in a safe and quick way. Encryption tools are used to keep the information safe. When a match is found, Kidner is notified, and the patient's doctor and other healthcare professionals get all the information they need to set up the logistics of the operation [31].

CONCLUSION

Cyberattacks are a worldwide problem that threatens the growth of online services. The healthcare industry is one of the biggest ones that has been hit hard by cyberattacks, which have led to identity theft and data manipulation by people who shouldn't have access to the information. Electronic healthcare records contain sensitive information about patients, such as their medical history, lab reports, payment information, and contact and payment information.

These records must be shared between healthcare providers and kept safe from cyberattacks. But the use of blockchain technology has made cyber security stronger. This study gave an overview of how blockchain can be used in healthcare to help prevent cyberattacks. For example, it showed how blockchain can be used to take care of organ transplants.

Decentralized blockchain technology is based on the interaction of stakeholders (hospitals, patients, taxpayers, etc.) without the control of a central intermediary. It is good for storing health data and rules for recording important information, and it provides audit trails that can't be changed (insurance records). It can also be used to manage digital assets and find out where they came from. Lastly, it improves data security and privacy because the data is encrypted in the block chains and can only be decrypted with the patient's private key. Even if a bad person gets into the network, they won't be able to read the patient's information.

Blockchain technology helps manage organ transplants and solves some of the problems that come with them. It can make the process better by making it easy to check if an organ donor and a patient are a good match and by making it quick to match a donor organ to a patient's needs. This technology also makes it possible to store personal choices in a safe way. This makes it possible to cut the time it takes to do a transplant by five times compared to the old way of doing things. It also makes the whole process of donating clearer and easier to track, which helps stop organ trafficking.

Bibliography

[1] Dimitrov D. V., "Blockchain applications for healthcare data management," *Healthc. Inform,* vol. 25, no. 1, pp. 51-56, 2019.

[2] J. ilian Oderkirk and L. Slawomirski, *Opportunities and Challenges of Blockchain Technologies in Health Care,* 2020, p. 12.

[3] M. Mettler and M. A. Hsg, in *18th International Conference on e-Health Networking, Applications and Services*, 2016.

[4] G. Alandjani, "Blockchain Based Auditable Medical Transaction Scheme For Organ Transplant Services," *3C Tecnol. Glosas innovación Apl,* no. Special Issue, p. 41–63, 2019.

[5] F. Alam Khan, M. Asif, A. Ahmad, M. Alharbi and H. Aljuaid, "Blockchain technology, improvement suggestions, security challenges on smart grid and its application in healthcare for sustainable development," *Sustain. Cities Soc,* vol. 55, no. December 2019, 2020.

[6] D. Mandal, R. Kar, S. Das, B. Ketan and P. Editors, "Advances in Intelligent Systems and Computing 343 Intelligent Computing and Applications," 2014. [Online]. Available: http://www.springer.com/series/11156.

[7] A. S. Jat and T. M. Grønli, "Blockchain for Cybersecure Healthcare," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics),* vol. 13475, p. 106–117, 2022.

[8] WHO, "Global strategy on digital health 2020–2025," *Geneva,* 2021.

[9] WHO, *Global action plan on physical activity 2018–2030: more active people for a healthier world,* World Health Organization, 2018.

[10] N. Z. Pazilovich, "The Concept Of Cybercrime , Its Peculiarities And Cybersecurity Issues," vol. 7, no. 9, p. 67–74, 2021.

[11] M. Kianpour, S. J. Kowalski and H. Øverby, "Advancing the concept of cybersecurity as a public good," *Simulation Modelling Practice and Theory,* vol. 116, pp. 1-16, 04 2021.

[12] Avani Rachh, "A Study of Future Opportunities and Challenges in Digital Healthcare Sector: Cyber security vs. crimes in digital healthcare sector," *Asia Pacific Journal of Health Management,* pp. 7-15, 2021.

[13] W. Bi, H. Yang and M. Zheng, "An Accelerated Method for Message Propagation in Blockchain Networks," ARXIV, 2018. [Online]. Available: http://arxiv.org/abs/1809.00455.

[14] Mubashir Husain Rehmani, "Blockchain Systems and Communication Networks: From Concepts to Implementation," *Textbook in Telecommunication Engineerin,* 2021.

[15] Shubham Hapse, "Blockchain 101: The Simplest Guide You Will Ever Read," VELOTIO, 02 11 2022. [Online]. Available: https://www.velotio.com/engineering-blog/introduction-to-blockchain-and-how-bitcoin-works. [Accessed 13 12 2022].

[16] tiana laurence, Blockchain for Dummies, Canada: John Wiley & Sons, 2019.

[17] M. Attaran, "Blockchain technology in healthcare: Challenges and opportunities," *Healthc. Manag,* p. 70–83, 2022.

[18] euromoney, "How does a transaction get into the blockchain?," EUROMONEY, [Online]. Available: https://www.euromoney.com/learning/blockchain-explained/how-transactions-get-into-the-blockchain. [Accessed 13 12 2022].

[19] D. Oxford Advanced Learner's, *artificial-intelligence noun - Definition, pictures, pronunciation and usage notes,* Oxford Advanced Learner's Dictionary, 2022.

[20] D. Mohan, "The Financial Services Guide to Fintech: Driving Banking Innovation Through Effective Partnerships," 2020. [Online]. Available: https://books.google.com/books?hl=en&lr=&id=e7zEDwAAQBAJ&oi=fnd&pg=PP1&dq=%22fintech+adoption%22+and+%22india%22&ots=17VvFNRKRe&sig=dXsT1tURXqRDl-suR6YDQFsKtc8.

[21] H. Mohamed and H. Ali, Blockchain, Fintech, and Islamic Finance: Building the Future in the New Islamic Digital Economy, Berlin: Walter de Gruyter, 2019.

[22] A. Shojaei, I. Flood, H. I. Moud, M. Hatami and X. Zhang, "Implementation of Smart Contracts by Integrating BIM and Blockchain," *Adv. Intell. Syst. Comput,* vol. 1070, p. 519–527, 2020.

[23] M. Grazia and V. Haydn, The Executive Guide to Blockchain: Using Smart Contracts and Digital Currencies in Your Business, Switzerland: Springer , 2021.

[24] K. Zheng, Z. Zhang and J. Gauthier, "Blockchain-based intelligent contract for factoring business in supply chains," *nn. Oper. Res,* vol. 308, no. 1, p. 777–797, 2022.

[25] H. L. Pham, T. H. Tran and Y. Nakashima, "Practical Anti-Counterfeit Medicine Management System Based on Blockchain Technology," in *4th Technol. Innov. Manag. Eng. Sci. Int. Conf.,* 2019.

[26] S. Kalarani, K. Raghu and S. K. Aakash, to Combat Counterfeit Drug Transactions, Springer Nature Singapore.

[27] S. Mahmood, M. Chadhar and S. Firmin, "Cybersecurity Challenges in Blockchain Technology: A Scoping Review," *Hum. Behav. Emerg. Technol,* pp. 1-11, 2022.

[28] V. Srivastava, T. Mahara and P. Yadav, "An analysis of the ethical challenges of

blockchain-enabled E-healthcare applications in 6G networks," *nt. J. Cogn. Comput. Eng,* vol. 2, p. 171–179, 2021.

[29] B. J. Chalissery and V. Asha, *Blockchain Based System for Human Organ Transplantation Management,* Springer, 2020, p. 829–838.

[30] S. Srivastava, V. Gupta, S. Tapaswi and N. Kumar, "Decentralised and distributed system for organ/tissue donation and transplantation," *Conf. Inf. Commun. Technol. CICT,* 2019.

[31] S. V. Shinde, P. N. Mahalle , V. Bendre and O. Castillo, *Disruptive Developments in Biomedical Applications,* 2022.

[32] Sajida Zouhari, "Kidner – a Worldwide Decentralised Matching System for Kidney," *J. Int. Soc. Telemed. Ehealth,* pp. 1-4, 2017.

[33] A. Zouarhi and Sajida Maroussia, "The smartest way to manage transplants; Kidner Solution," Kidner, 2017. [Online]. Available: https://www.kidner-project.com. [Accessed 08 12 2022].

جامعة القاهرة
كلية الدراسات العليا للبحوث الإحصائية

المؤتمرالسنوي الدولي الخامس والخمسون
لعلوم البيانات

علوم الحاسب والمعلومات

5-7 ديسمبر 2022