*International Journal on Recent and Innovation Trends in Computing and Communication*
***ISSN: 2321-8169 Volume: 11 Issue: 9***
***Article Received:*** *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

# A Comprehensive Review on Machine Learning Based Models for Healthcare Applications

**Sabari Vasan S[1]**
[1]Research Scholar
School of Computer Science Engineering and Information Systems
Vellore Institute of Technology
Vellore, India
sabarivasan.s2022@vitstudent.ac.in

**Jayalakshmi P[2]**
[2]Assistant Professor
School of Computer Science Engineering and Information Systems
Vellore Institute of Technology
Vellore, India
pjayalakshmi@vit.ac.in

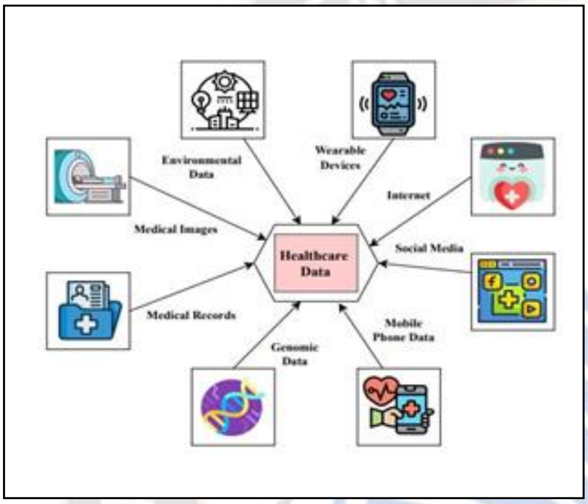
**Abstract**— At present, there has been significant progress concerning AI and machine learning, specifically in medical sector. Artificial intelligence refers to computing programmes that replicate and simulate human intelligence, such as an individual's problem-solving capabilities or their capacity for learning. Moreover, machine learning can be considered as a subfield within the broader domain of artificial intelligence. The process automatically identifies and analyses patterns within unprocessed data. The objective of this work is to facilitate researchers in acquiring an extensive knowledge of machine learning and its utilisation within the healthcare domain. This research commences by providing a categorization of machine learning-based methodologies concerning healthcare. In accordance with the taxonomy, we have put forth, machine learning approaches in the healthcare domain are classified according to various factors. These factors include the methods employed for the process of preparing data for analysis, which includes activities such as data cleansing and data compression techniques. Additionally, the strategies for learning are utilised, such as reinforcement learning, semi-supervised learning, supervised learning, and unsupervised learning. are considered. Also, the evaluation approaches employed encompass simulation-based evaluation as well as evaluation of actual use in everyday situations. Lastly, the applications of these ML-based methods in medicine pertain towards diagnosis and treatment. Based on the classification we have put forward; we proceed to examine a selection of research that have been presented in the framework of machine learning applications within the healthcare domain. This review paper serves as a valuable resource for researchers seeking to gain familiarity with the latest research on ML applications concerning medicine. It aids towards the recognition for obstacles and limitations associated with ML in this domain, while also facilitating the identification of potential future research directions.

**Keywords-** Artificial Intelligence; Machine Learning; Diagnosis; Treatment; Healthcare

## I. INTRODUCTION

From the start of the modern era, there was a substantial rise in the significance attributed to technological advances in regards of both productivity and expansion [1–3]. This tendency is anticipated to persist. Kaplan and Haenlein's work [4] claims that improvements in machine technology have replaced manual labor-intensive tasks, which has aided in human development. Artificial intelligence has become a pivotal technological innovation that enables individuals in different sectors to substitute physical labour with enhanced mental capabilities and cognitive abilities [5,6].The application of AI in the medical industry employs computational methods to extract meaningful insights from unprocessed data enabling accurate and precise decision-making in the field of medicines [7, 8]. The field of machine learning represents a subfield within the broader discipline of artificial intelligence. The system can autonomously identify and uncover trends inside datasets. Machine learning models possess the ability to acquire knowledge and refine their performance through automated learning processes without requiring detailed programming instructions [9, 10]. Basically, a learning framework acquires knowledge by analysing tests, but specific programming conforms to established regulations or an exclusive premise [11, 12]. Machine learning [13, 14] has been shown to enhance productivity and dependability while also decreasing costs in computer usage. In addition, it possesses the capability to generate models efficiently and precisely by means of analysing the data. Machine learning encompasses a variety of technologies allowing the processing of vast quantities of data, surpassing the capacity of human comprehension. An instance of health data may encompass several types of information, such as population statistics, pictures, test outcomes, genetic information, health information and sensor-collected information. Several systems serve in the creation or collection of these information segments. Examples of such platforms include networking servers, health information systems, and genetic sources of information, laptops, handsets, mobile programmes, gauges, and smart watches [15, 16]. Figure 1 illustrates multiple data collection methods in medical fields.

Figure 1. Multiple data collection methods in medical fields.




The field of healthcare is widely recognised as the foremost domain for the use of machine learning as well as AI techniques [17]. During the second half of the twentieth century, numerous researchers introduced a multitude of clinical findings. The topic under discussion pertains to decision-making systems. In the year 1970, rule-based techniques gained significant popularity [18, 19]. They have been effectively utilised for the interpretation of cardiac recordings, the identification of illnesses, and the selection of suitable treatment approaches. Nevertheless, systems based on rules were characterised by their high cost and sensitivity to weaknesses. It is imperative for individuals to effectively comprehend and understand the principles of decision-making with precision. Continuous updates are also recommended. These devices are recognised as the inaugural cohort of artificial intelligence-based platforms [20, 21]. In the context of these systems, it is imperative that health information be applied with precision by knowledgeable professionals in order to develop efficient selection criteria. On the other hand, contemporary AI models employ machine learning methodologies to discern data patterns within intricate contexts [22, 23]. Machine learning offers a wide range of uses within the field of healthcare. These tasks encompass illness recognition and categorization, the assessment of health risks, and the determination of suitable treatment methodologies. Figure 2 illustrates numerous machine learning applications within the healthcare sector. In the last few years, an abundance of investigation studies has been done by experts. With a specific emphasis on various facets of healthcare [24, 25]. In their research, the authors have employed a range of methodologies for machine learning including Support Vector Machine, Naive Bayes, Artificial Neural Networks, Evolutionary Computations, Fuzzy Systems. Additionally, they have explored hybrid approaches such as neuro-fuzzy systems and neuro-genetic systems [26].

Every day, numerous researchers dedicate their efforts to the study of machine learning and AI within the healthcare domain. Hence, it is imperative for carrying out an in-depth analysis of additional scientific inquiries in this field, given the significant progressions observed in machine learning methodologies and their utilisation within the field of medicine. Table 1 displays a compilation of review articles that provide a survey of the medicinal benefits of ML. These papers frequently concentrated on the utilisation of ML approaches within a certain medical domain, such as medical imaging or the utilisation of ML approaches at the background for identifying or treating a specific condition. There is an urge to give fewer attention towards the structural aspects of machine learning models employed in various methodologies. It is vital for AI specialists have an understanding concerning the composition of learning models employed in various methodologies and discern their respective merits and drawbacks, with the aim of enhancing these models within the healthcare domain. There is a limited number of review articles in the healthcare sector which particularly examine the structural aspects of machine learning-based models, as represented by reference [27]. Hence, it is imperative to allocate greater focus to this topic matter. Consequently, the purpose of this study is to investigate the concepts associated with the development of frameworks based on ML in medicine as well as how they could be utilised in the healthcare industry. This study offers a comprehensive perspective for academics in the field of artificial intelligence to address the inquiry, "In what ways can machine learning techniques be leveraged to enhance various healthcare methodologies?" Table 2 presents a comparative analysis of our investigation article with other extant studies within the same field. This study begins by providing a categorization of ML methodologies in medicine. The categorization method separates ML-based methodologies in the medical sector on the basis of their data pre-processing techniques including data cleansing and data lessening, Learning methodologies which include supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning , Evaluation methodologies such as Simulation-based evaluation and Practical implementation based evaluation in everyday life situations and their Applications such as detection and therapy.

TABLE I. A FEW REVIEWS OF MEDICAL APPLICATIONS OF MACHINE LEARNING.

| Review Articles | Description |
|---|---|
| The Application of Machine and Deep Learning in the Diagnosis and Treatment of COVID-19: A Comprehensive Review of Existing Literature, Identified Challenges, and Prospective Avenues for Further Research. | Alafif et al. [27] explored machine learning applications for COVID-19 diagnosis and treatment. They introduced new ML techniques, showcased freely available tools and datasets, and identified future research opportunities. Highlighting ML's diverse potential, they envisioned its use in health |

| | |
|---|---|
| | evaluation, treatment suggestions, prevention, drug development, and vaccine creation. |
| A literature review on the applications of artificial intelligence in the fight against covid-19. | Tayarani [28] explores uses of AI in battling COVID-19, from diagnosis and therapy to virus tracking. Examining ML applications in clinical settings, the paper covers patient evaluation, treatment, and monitoring. Chest image analysis, virus characterization, and outbreak modeling are also discussed, along with available datasets for further research. |
| The current state and prospective challenges of machine learning in the medical field. | Smiti [29] dived into the foundation of ML in medicine. The paper starts by breaking down healthcare's core aspects: early detection, identification, diagnosis, and therapy. Machine learning is defined, then its broad categories explored: Supervised, Unsupervised, Semi-supervised, and Reinforcement Learning. The paper then delves into specific applications like disease identification, drug discovery, surgical robotics, and medical data analysis, highlighting the unique challenges in each. Ultimately, it focuses on healthcare data analysis, examining the specific intricacies this area presents. |
| The application of machine learning and artificial intelligence in the field of haematology. | Shouval et al. [30] offer resources for medical professionals to grasp machine learning in haematology. They detail standards, applications, and learning types. A six-step framework for building ML models is outlined, along with potential barriers and limitations, especially in implementation within haematology. |
| The diagnostic, classification and prognostic applications of machine learning in heart failure. | Olsen et al. [31] explored AI methods in heart failure, reviewing its healthcare applications and highlighting key considerations for building ML models. They categorized ML by training framework (supervised, unsupervised, deep learning) and application (evaluation, categorization, prognosis of coronary artery disease). The authors also discussed challenges facing AI in medicine. |

This Review article is expected to provide significant value to investigators in the area of AI seeking to gain a comprehensive understanding of the most recent advancements in machine learning-based methodologies within the healthcare domain. It facilitates the process of acquainting oneself with the current state of study, identifying the obstacles and constraints encountered in this field, and gaining insight into potential avenues for future investigation. we will analyse and evaluate the many aspects of the subject matter. In this study, our primary focus is on a collection of review articles pertaining to the integration of ML in area of medicine, which have been published in reputable academic journals. In addition, we conducted a comprehensive analysis and examination of multiple review papers, book summaries, and research materials. academic papers and conference papers sourced from reputable publishers such as Springer, Elsevier, and IEEE. Due to the increasing volume of intellectual articles being available and the prevalence of diseases in the healthcare sector is substantial, necessitating a selective approach to studying them due to the constraints of limited resources.

TABLE II. COMPARING OUR REVIEW ARTICLE WITH VARIOUS SURVEY ARTICLES.

| Survey articles | Initial preparation techniques | Framework for learning | Criteria for evaluation | Integration |
|---|---|---|---|---|
| [27] | No | Yes | No | Yes |
| [28] | No | No | No | Yes |
| [29] | No | No | No | Yes |
| [30] | No | Yes | No | Yes |
| [31] | No | Yes | No | No |
| [32] | No | Yes | No | Yes |
| Our survey article | Yes | Yes | Yes | Yes |

Regarding this review paper. Consequently, the selection of articles that were just recently accepted. From the realm of medicine, it is imperative to conduct a comprehensive analysis of the published material, so offering a more intricate evaluation. Additionally, it is essential to incorporate a broader range of sources to enhance the scholarly discourse. The dataset is shared among articles that pertain to the same concept. Subsequently, the elimination of more scholarly articles is undertaken. One can utilise Google Scholar to locate the aforementioned documents and conduct searches using other terms include "Supervised learning in healthcare," "Unsupervised learning in healthcare," "Semi-supervised learning in healthcare," "Reinforcement learning in healthcare," "Deep learning," and "Future hospitals" are all terms that refer to the use of ML and AI in healthcare.

The main contributions of this research are as follows:

- We first introduce a categorization of healthcare plans that utilise machine learning techniques.
- Our proposed taxonomy classifies machine learning-based plans in medical care according to data pre-processing approaches (including data cleaning and data reduction), learning approaches (such as unsupervised, supervised, semi-supervised, and reinforcement learning), evaluation approaches (including simulation-based and practical implementation-based evaluation in real environments), and applications (such as diagnosis and treatment).
- Based on our suggested classification, we analyse various research that explore the use of machine learning in medical applications.

Subsequent sections shows the work of this survey article are structured as follows: Section 2 of the document discusses the area of ML and its various applications within the healthcare industry. In the following section, a comprehensive framework is presented for the development of a learning model within the medical domain. The introduction to our suggested classification is presented in Section 4. In this paper, Section 5 examines ML-based approaches in healthcare, using the classification outlined in the study. In the sixth section, a summary is provided for the debates pertaining to the machine learning-based methods that have been reviewed in this work. In the next section, an overview is provided about the hurdles and limitations related to the utilisation of ML techniques in medicine. Finally, the conclusion of the paper is presented in Section 8.

## II. MACHINE LEARNING

A concept of endowing computers with the capacity to understand like humans is frequently seen as a visionary search, mostly due to the fact machines have fundamental intelligence [16, 18]. There exist observable distinctions among human beings and machines in the execution of their tasks, with intelligence being an important differentiating factor. It also suggests that people possess the capacity to acquire knowledge from past encounters, whereas machines lack this capability. Indeed, it is imperative that they are programmed to adhere to specific information [25, 33]. Today, the field of ML has enabled PCs to acquire knowledge and improve their performance through the process of learning from past events. Historically, conventional computing algorithms were comprised of clearly designed instructions, sometimes referred to as "hard coded". PCs employ this information to address the difficulty, but in contemporary times, machine learning facilitates computers in acquiring decision-making rules autonomously, hence obviating the necessity for programmers to manually formulate such rules [34, 35]. This trend is often referred to as "soft coded". Machine learning is a constituent part of the wider discipline of computational intelligence. Machine learning systems exhibit higher levels of intelligence and has the capability to operate autonomously without requiring human involvement. The word "smart machine" is indeed regarded as a symbol [36]. The term smart machine pertains to the field of artificial intelligence and its associated objectives. In the year 1995, Allan Turing raised the inquiry for the initial occasion: "Is it possible for a machine to exhibit cognitive abilities?" The one in question developed a test known as the "Turing Test". This assessment assesses the cognitive abilities of a computer [37, 38]. At present, there are multiple interpretations and conceptualizations of the field of machine learning. Arthur Samuel provides a definition of machine learning as the study of enabling computers to acquire knowledge and improve performance without the need for specific code [39]. According to Alpaydin, artificial intelligence can be defined as the process of coding devices to enhance their efficiency depending on samples of data as well as previous experiences [40]. The term "machine learning" is used to describe a method of exploring within a representation field to develop the most optimal representation using the data that is available [41, 42]. Moreover, the term "machine" denotes an algorithmic entity that executes search operations. The algorithm presented in this study is a fusion of mathematical principles and logical reasoning [41, 42]. The primary objective of artificial intelligence is to address the inquiry: "How can historical data be utilised to create a computer programme capable of solving a problem and subsequently enhancing its performance through experiential learning?" [43,44]. Machine learning is a technological field concerned with the development of computer programmes that possess the ability to replicate human thinking and gain knowledge from the world around them. Regarding AI, an algorithm is established then instructed utilising a substantial volume of information, usually made up of thousands of data samples, in order to effectively handle complex tasks. The objective of this approach is to make decisions, predictions, or execute activities without the need for explicit programming. For this model to effectively provide the necessary output, it is imperative that it can accommodate the given inputs. At times, individuals possess a propensity for comprehending this model with relative ease. Nevertheless, in certain instances, it bears resemblance to a conventional black box. This implies that the comprehension of this structure for human beings is not readily achievable. The framework provided in this study serves as an approximation of the process that needs to be replicated by a machine, as indicated by previous studies [20.45].

### *A. Applications of ML in Healthcare Industry*

A branch related to medicine encompasses an extensive choice of applications for machine learning. It has the potential to expedite tasks that require a significant amount of time and intricate responsibilities within this field. In recent times, there has been notable and quick innovation in ML and developing more efficient machines and the availability of electronic medical data. These developments have opened up avenues for enhancing the healthcare process. The implementation of these emerging technologies has been shown to effectively lower expenses, speed up identifying suitable pharmaceutical compounds, and enhance the overall efficacy of therapeutic interventions.

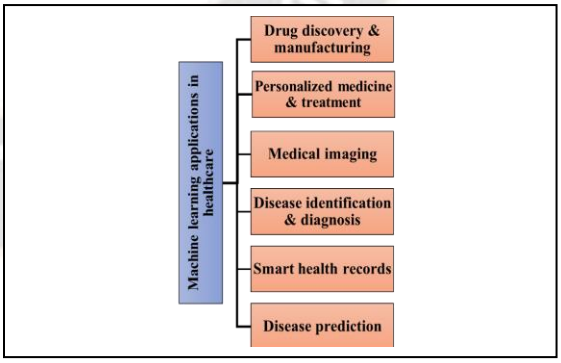



Figure 2. Numerous Machine learning applications in healthcare.

Currently, machine learning has attracted the attention of entrepreneurs and other stakeholders within the healthcare industry [46]. Typically, applications for machine learning within the medical field can be categorised into three distinct groups:

- Initial group [47, 48] enhancing the existing clinical infrastructure. These applications represent most basic machine learning techniques within the medical field. The utilisation of these techniques has been shown to enhance the efficiency of pre-existing shapes. ML methodologies are utilised to establish standardised, rules-driven methods in typical tasks, including emulation and replication of data. The classification of digital pictures used in hospitals is an example of machine learning applications. The utilisation of this method enhances the precision of conventional image processing methodologies. ML may be utilised to analyse the radiological images to provide predictions regarding the presence or absence of specific diseases. Additionally, machine learning can be employed to assess retinal pictures to ascertain the presence or absence of optical hazards in patients. One instance is Aindra, a medical startup that utilises machine learning and AI as its foundation. The classification of medical images is facilitated through the utilisation of a machine learning-based platform and the main goal of the device serves to augment precision as well as effectiveness for cancer diagnosis.
- The Second group pertains to the enhancement of healthcare infrastructure. ML applications under this specific category offer novel capabilities through the implementation of various architectures. There is a trend towards the implementation of personalization. Personalised medicine is a notable example of machine learning applications [8, 49]. This medical intervention is designed to address the individualised requirements of an individual, taking into consideration their unique characteristics, such as their genetic composition. One notable example of a company that is actively pursuing personalised healthcare services is iCarbonx. This approach leverages extensive datasets, biological sciences, and computational intelligence.
- The third group, known as autonomous healthcare frameworks, has experienced rapid growth in the field of machine learning applications. The ML-based models are developed to autonomously execute their tasks according to predetermined objectives [11]. An instance of an eventual application in the healthcare industry involves the construction of a hospital that operates without the presence of physicians [37, 38]. Consequently, it is imperative that we equip ourselves for an impending era characterised by the proliferation of machine learning and computational intelligence technologies. Hence, it is important to strategize the incorporation of robots into forthcoming healthcare facilities. In the foreseeable future, it is anticipated that robots would assume the responsibility of executing all healthcare procedures, encompassing tasks ranging from diagnostic assessments to surgical interventions. Currently, in advanced nations such as Korea, China, and the US, the use of robotic equipment has become prevalent in the field of an operation, enabling surgeons to do intricate procedures inside the confines of the surgical suite [50, 51]. Nevertheless, this emerging technology exhibits certain limitations and flaws; yet, its progress is swift and warrants further development. As an illustration, the Mayo Clinic, for example, is transitioning towards a healthcare facility that operates without the presence of medical practitioners. At present, the design of its components is being undertaken. Nevertheless, it is imperative that these components undergo thorough testing to ensure compliance with a range of requirements. Currently, surgeons employ robotic technology to enhance the efficacy of operations [52, 53].

## III. An Overview of the General Structure for Developing a Healthcare Learning Model

The following part provides an inclusive description of the many steps encompassed by the design process of an instructional approach from the medical field. It's essential that emphasis as the intent of this element is to provide researchers with extensive knowledge of every step entailed during the development of a framework for learning specifically within the subject area of medicine and it is recommended that researchers carefully review and undertake. Further investigation in this field is necessary to attain a comprehensive comprehension and expertise regarding. The learning models have been discussed in previous studies [18, 21]. To develop a learning model within the healthcare domain, it is imperative that there are five primary stages that can be considered in relation to this discussion. These stages include problem formulation, dataset acquisition, data preparation, and advancement of an ML framework. There are three primary stages involved in the procedure are development, implementation, and assessment. The steps are depicted in figure 3. During the subsequent paragraphs, extensive clarification can be found for all the above stages.

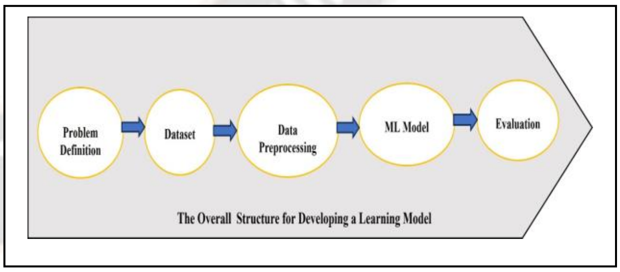



Figure 3. Multiple stages required to construct an instructional framework.

*1) Defining the issue:* The purpose of this section is to clearly define the problem that will be addressed in this study. In the framework of health care, it is imperative to address the fundamental query of the underlying objective behind the development of a learning model. To develop a valuable model, the initial phase involves the identification of issues and obstacles within the healthcare domain. It is imperative for researchers to conduct a thorough analysis on the precise methods through which machine learning might be utilised to enhance medical services. Furthermore, it is imperative to thoroughly analyse the current solutions that have been proposed in this field [31]. During the initial stage, it is crucial to assess the accessibility of data. This implies that researchers must possess knowledge of pre-existing data sources, as the availability of data is crucial for the development and assessment of the instruction framework. From the arena of medicine, lack of information can be caused by many variables like an inadequate supply of electronic information, concerns over persistent confidentiality, marketable considerations or else the prevalence of uncommon medical conditions.

*2) Dataset:* In the framework of healthcare, utilisation of datasets is necessary for the creation of an instructional framework for the purposes of training, verification, and validation. Healthcare datasets commonly encompass a range of information, such as demographic data, pictures, laboratory findings, genetic data, and information collected from devices [54, 55]. Different platforms are utilised for the production or collection of this data. Examples include communicating systems, electronic medicine information, genetic information, portable laptops, cell phones, app stores, and connected gadgets [56, 57]. Currently, taking advantage of online and service clouds. has the potential to enhance worldwide connectivity [58,59]. Consequently, the accessibility of data has been enhanced. Prior to constructing a model for learning in the healthcare domain, it is imperative to establish a suitable framework for assessing the efficacy of the learning model. Merely asserting that a machine learning model exhibits good performance and desirability is insufficient. Machine learning models prioritise the centrality of data in their functioning. Consequently, individuals may encounter a challenge known as excessive or insufficient fit [60, 61]. A successful instructional framework needs to strike a balance among excessive and insufficient fit. This signifies sufficient variance and discrimination are required. Underfitting arises when the level of detail of the learning model is insufficient in relation to the intricacy of the issue at hand as well as the magnitude of the information. The instructional framework's precision is suboptimal in both the training and assessment sets. It suggests there may be a substantial amount of partiality. Conversely, overfitting arises when the learning model exhibits excessive complexity, characterised by a multitude of variables that exceed the scope of the issue at hand and the magnitude of the information. For instance, the framework displays acceptable results on the original dataset, while demonstrating suboptimal performance on the testing set. In this scenario, there is a significant degree of variability. Typically, an effective learning framework should have both minimal bias as well as maximal variance.

*3) Data Pre-processing:* In the healthcare domain, the development of a system for learning represents a significant challenge in the form of data preprocessing. This is because the training phase of a model based on machine learning necessitates the utilisation of high-quality information to attain superior outcomes and enhance precision. The preparation of information is a fundamental procedure that involves analysing and taking care of various problems with data quality, such as noise, values that are missing, duplication, and inconsistencies. The primary aim of the method aims to increase the reliability of the information stored in it prior to the development of the framework for learning. Hence, during the procedure of information initial preparation, it becomes necessary to implement outlier filtering techniques and estimate missing values. When dealing with data that has high dimensions, it is possible to employ data reduction techniques such as selection of features [62, 63] or extracting features [64]. Selection of features is a process that aims to identify the most optimal collection of features. In contrast, feature extraction involves the identification and extraction of a reduced-dimensional dataset derived from the original dataset.

*4) Machine learning model development:* While building a learning model within the healthcare domain, when evaluating the database, it's essential to take into account many elements including the scale of the information, the specific training method employed and the duration required for model inference. The level of difficulty of a framework for learning is determined by considering the dimension of the database in order to prevent. The subject of concern is the phenomenon of overfitting or underfitting in statistical modelling. The training duration of an instructional approach is quite significant. Still, learning models that possess a greater number of features have the potential to generate more precise outcomes. But in this scenario, both models exhibit a higher number of computing processes and it require an extended duration for the training period. Consequently, these systems are unsuitable for real-time applications. Hence, it is more suitable to employ lightweight designs while developing a lean structure. model. The consideration of the learning scheme type holds significant importance in the development of machine learning models [65, 66]. Broadly stating there exist four primary modes of learning, which encompass supervised or unsupervised training, semi-supervised training, and reinforcement modelling [67, 68] are distinct paradigms within the field of machine learning. The strategies will be expanded in a more extensive manner in the fourth part.

*5) Evaluation:* Inspection about an organisation that utilises machine learning requires carrying out a series of operations to identify anomalies among the system's current behaviour and the anticipated behaviour [69]. Once a learning model has been developed in the field of healthcare, it is imperative to conduct the requisite evaluations in order to ascertain whether the model possesses the requisite conditions for implementation in real-world settings. During the evaluation phase, designers employ a range of scales to assess the efficacy of the instructional approach. The resulting evaluation identifies the positive and negative aspects of the subject under consideration. Furthermore, it is imperative to assess the efficacy of the instructional framework in real-world settings and analyse its behaviour during interactions with actual users [70, 71]. Various evaluation features of a system that uses ML

*International Journal on Recent and Innovation Trends in Computing and Communication*
***ISSN: 2321-8169 Volume: 11 Issue: 9***
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

encompass an assessment of the information employed to create the final training framework, the evaluation of the algorithms for learning employed in designing said model and the assessment of the effectiveness demonstrated through the finalised framework.

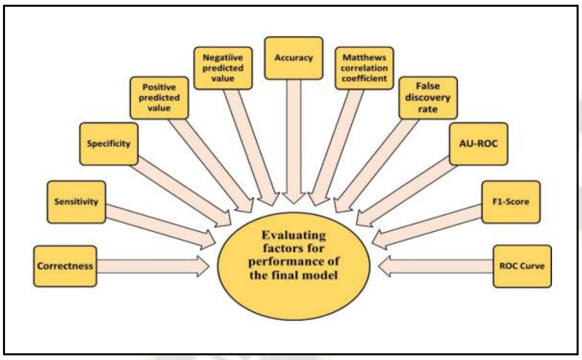



Figure 4. Evaluation factors for the effectiveness of the ultimate framework.

## IV. CATEGORIZATION OF TECHNIQUES THAT USE ML IN A MEDICAL FIELD

This specific domain will provide an extensive listing ML methodology employed in medical sector and the classification as depicted in Figure 5, has four distinct groups. This paper discusses various academic subjects related to initial preparation techniques, instructional approaches, assessment techniques as well as their applications in the field of diagnosis and treatment.

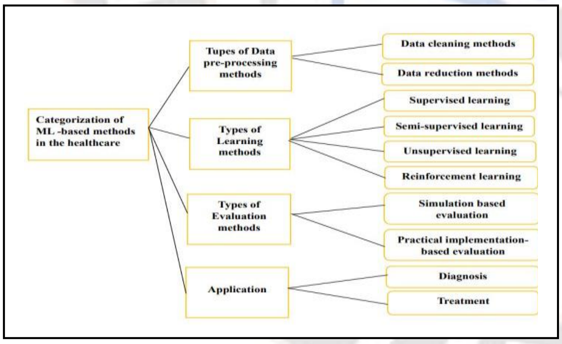



Figure 5. Proposed Categorization of ML founded methods in medical sector.

- The first section explores different initial preparation techniques, including techniques like cleansing information and information minimization.
- The second section focuses on various learning methods, such as unsupervised training, semi-supervised training, supervised training and reinforcement training.
- The third section there are two primary evaluation methodologies that can be employed replication founded assessment and applied execution founded assessment in actual situations. In the subsequent discussion, an extensive representation of each of these portions is provided.

### A. *Various kinds of initial data preparation techniques*

For the outlined categorization framework, machine learning approaches employed in the medical field are categorised into two primary groups, namely initial data preparation techniques, which encompass cleansing information and information minimization techniques. In the subsequent section, a comprehensive explanation of each of these strategies is provided. In addition, Figure 6 showcases various techniques employed in the pre-processing of data.

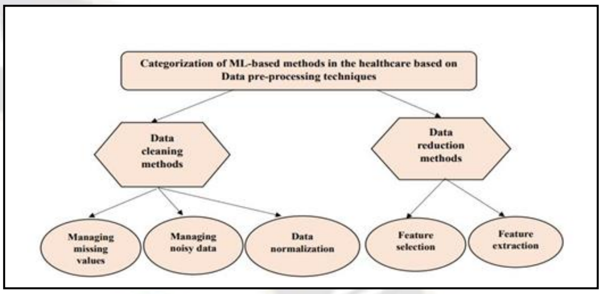



Figure 6. Categorization of ML founded methods via initial data preparation techniques.

1) *Cleansing information techniques:* Several machine learning-based approaches employed in the field of medicine utilise data cleaning techniques to address prevalent issues seen in medical records, such as lost or noisy information. These issues have many causes: The precision of data gathering systems in the medical area is questionable. Thus, certain data may be absent because of the physical constraints built into these devices. There is a possibility of inadvertent recording errors in the data. Additionally, certain data samples are generated manually. By medical professionals or healthcare providers. Consequently, there is a possibility of inaccurate recording of the data. Human errors can manifest in various ways, including instances where patients either unintentionally or intentionally fail to adequately communicate their needs or concerns and detailly regarding their medical condition. This incident leads to inaccuracies in the process of data recording. it might be said that there exist various techniques for data cleaning, such as the management of missing values as well as handling of noisy data management and data standardisation [18,20].

*a) Missing value management:* In the healthcare domain, there are two primary methodologies for addressing the issue of missing values. (1) Eliminating the data points that contain missing values. It should be noted that if there is an excessive amount of information components through value gaps in information gathering, this strategy becomes impractical. (2) Estimation of values that are missing. It should be noted that if the approach employed to estimate the missing data is not precise, it will diminish the precision of the framework for learning.

*b) Missing value management*: In the healthcare domain, there are two primary methodologies for addressing the issue of missing values. (1) Eliminating the data points that contain

*International Journal on Recent and Innovation Trends in Computing and Communication*
***ISSN: 2321-8169 Volume: 11 Issue: 9***
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

missing values. It should be noted that if there is an excessive amount of information components through value gaps in information gathering, this strategy becomes impractical. (2) Estimation of values that are missing. It should be noted that if the approach employed to estimate the missing data is not precise, it will diminish the precision of the framework for learning.

*c) Noisy data management:* Sorting techniques are commonly employed in the field of health data analysis to effectively eliminate noise present in datasets. The change enhances the precision of a learning framework. Still, the identification of noisy data poses a considerable challenge. One potential approach is the evaluation of the collection of data by individuals who possess expertise in the relevant field, such as professionals and physicians, with the aim of enhancing its overall quality. This process leads to enhanced precision in modelling and a decrease in associated errors. Nevertheless, this endeavour incurs significant expenses and demands a substantial investment of time.

*d) Data normalization:* Medical information are typically represented using various scales, such as age and gender, among others. It is not possible to make comparisons between these data samples. In order to address this issue, it is recommended to employ data normalisation techniques, such as the Min-Max approach, which facilitates the transformation of data into the interval [0, 1].

*2) Information minimization techniques:* Frequently, medical data exhibits a great number of features. The previous factor has an adverse effect on the efficacy of machine learning methods, as it diminishes the overall effectiveness of the training phase and compromises the correctness of the resulting learning model. The reduction of dimension refers to the process of representing health data in a condensed format. Consequently, this procedure leads to the loss of certain information. A suitable approach for reducing dimensionality in the medical domain should preserve valuable features. Data reduction approaches can be classified into two primary types: selection of features and extraction of features.

*a) Feature selection:* During this phase, a specific collection of characteristics is chosen from the medical databases for utilisation in the method of learning. The procedure of feature selection can be conducted either automatically or semi-automatically [62,63]. The choice of whether to eliminate or retain a feature is dependent upon the intended application. In a broad information, feature selection approaches can be classified into three distinct classes.

- Wrapper techniques: In the current study, the machine learning-based framework is regarded as a black box in the employed methodologies. Next, the model is provided with various subsets of characteristics and the performance of the system is assessed for each individual subgroup to identify its level of effectiveness. At last, the optimal group of traits is chosen. There exist two possible wrapper techniques, namely selecting forward and reverse selection. In the forward method of choice, the initial step involves considering a subset that is devoid of any elements. Next, a specific feature from the health database is chosen and subsequently included into the designated subset. Then, the effectiveness for ML founded framework is assessed and then inclusion feature in the final subset results in a reduction of system error in comparison to other characteristics, it is incorporated. The process mentioned earlier repeats until a reduction in the error rate is observed. The reverse selection techniques exhibit similarities to the forward choice methodologies. Still, a distinction exists. In the setting of these systems, our initial step involves the consideration of a subset that encompasses all features. Next, an element from the subset is chosen in each phase and later removed from the subset. The process mentioned above repeats until the failure rate of the learning framework diminishes.
- Embedded techniques: The selection of features procedure is an integral part of the learning framework in these methods.
- Filtering techniques: These strategies are regarded as a distinct component of the instructional framework. The previously mentioned approaches include doing a priority check on every unique characteristic inside the database, resulting in an order of these features according to an established standard. Then, the client selects the more advantageous attributes.

*b) Feature extraction:* These techniques are being used for the purpose of reducing health-related data that possess a significant number of dimensions [18,64]. This process preserves the fundamental characteristics of the database while eliminating extraneous data and interdependencies. The execution of this method is expected to expedite the acquisition of knowledge and yield outcomes that are more precise in nature. For instance, we present a selection of the most significant feature extraction approaches.

- Principal Component Analysis [PCA]: It represents an empirical technique that falls within a category of multivariate and unsupervised techniques. PCA is employed to evaluate the data to extract valuable information. Next, the data is represented as a collection of distinct axial values. The elements in discussion are commonly referred to as the primary components.
- Linear Discriminant Analysis [LDA]: The employed methodology constitutes a variant called supervised training. The goal of this approach is to identify an ordered set of attributes that may be categorised into multiple distinct groups. This approach aims to optimise the differentiation between groups and effectively construct linear discriminant values.
- Singular Value Decomposition [SVD]: The type of technology in problem is classified as an unsupervised

*International Journal on Recent and Innovation Trends in Computing and Communication*
***ISSN: 2321-8169 Volume: 11 Issue: 9***
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

learning method. The current approach bears a strong resemblance to PCA. It can be considered a generalised form of principal component analysis. The method in question is classified as a matrix-based factorization technique, known for its effectiveness in lowering the dimensions of data. The singular value decomposition algorithm provides the utilisation of any low-rank matrices and allows for an effective approach and description regarding the actual matrices.

## *B. Varieties of Learning techniques*

A suggested categorization organises ML methods employed by the medical industry according to 4 primary categories according to their instructional approaches: Supervised learning, Semi-Supervised learning, Unsupervised learning & Reinforcement learning. Table 3 presents a comparative analysis of several instructional strategies.

TABLE III. VARIETIES OF LEARNING TECHNIQUES.

| Technique | Goal | Schema |
|---|---|---|
| Supervised learning | The task involves making predictions on the testing set by establishing the correlation between the inputs and the outcomes. | labeled schema |
| Semi-Supervised learning | Predicting the categorization of the testing dataset. | Either labeled schema & unlabeled schema |
| Unsupervised learning | The process of identifying similarities within data and categorising data items. | unlabeled schema |
| Reinforcement learning | The identification of optimal actions by ways of engaging with an environment. | No schema |

*1) Supervised Learning:* The learning approach encompasses the integration of various elements & outcomes, which are referred to as labeled records [33]. Its goal of carrying out this method of instruction aims to ascertain a correlation among each of these inputs & outcomes during instruction procedure [72,73]. The algorithm generates a mathematical function that establishes a correspondence between input data and corresponding labels. Next, it is employed for the purpose of forecasting the classification of data that lacks assigned labels. Supervised learning is employed in cases where the training set possesses corresponding outputs, often known as labels. In the subsequent section, we provide the fundamental supervised learning algorithms.

*a) Decision Tree (DT):* It's a type of ML procedure which falls under certain category of supervised learning. The learning model in question is developed using DT, which utilises a collection of IF-THEN rules derived from the set of training data to make predictions regarding the class of outcomes [74,75]. The creation of the hierarchy tree depends on the features present in the dataset. The decision tree is comprised of three distinct types of nodes: the root node, which is the topmost node in a decision tree; the internal node, which represents an experiment or comparison on each attribute and the leaf node, which denotes the class label or final outcome

*b) Naive bayes (NB):* The classifier in discussion defines as the statistical framework which establishes the conditional probability among the parameters (features) and the target variable (class) [76,77]. The Naive Bayes algorithm is a straightforward approach that relies on the principles of Bayes' theorem. This approach operates on the assumption that the distribution of data points within a single class follows a specified probability distribution. In the field of machine learning, there exists a prominent hypothesis known as the independence of features in the setting of NB models. Still, the previous theory lacks practicality in real-world scenarios due to the prevalent presence of high correlations within most genuine datasets. Undoubtedly, in contemporary times, the use of NB has proven instrumental in addressing a multitude of practical challenges, exhibiting commendable efficacy.

*c) Artificial Neural Network (ANN):* The ANN has inlay variables, output variables, and weights. The behaviour of the network is contingent upon the correlation among the input and output variables [78,79]. Artificial neural networks are composed of three layers. It is important to acknowledge that within each of these layers, there exists a multitude of processing units referred to as neurons. The initial layer is referred to as the input layer, which is responsible for receiving unprocessed data. The second layer, commonly referred to as the hidden layer, is responsible for executing the learning task. It should be noted that certain artificial neural networks possess several hidden layers. The third layer is commonly referred to as the output layer. The activation of the output layer is influenced by the learning process occurring in the hidden layer, as well as the weights associated with the units of input and hidden units. The determination of the number of layers that are hidden and the allocation of neurons inside each layer is a task assigned to the designer. This study is undertaken using a method of experimentation and refinement. It should be noted that there exist numerous methodologies for training artificial neural networks and adjusting the weights to minimise error. The back-propagation algorithm is widely recognised as the most used approach.

*d) Random forest (RF):* This classifier has high speed, accuracy, and resilience to noise. Random Forest (RF) is a machine learning approach that use an ensemble of decision trees to classify data [80]. In this framework, a substantial quantity of autonomous trees is generated by utilising a first training set, such as an NF matrix, where N represents the total number of samples and F denotes the number of features. Once the random forest model has been constructed, it is subsequently employed for the purpose of predicting labels. The determination of the final label for the samples is ultimately achieved through the utilisation of a majority vote approach.

*e) Support vector machine (SVM):* The learning technique in discussion is a supervised binary categorization method. The Support Vector Machine algorithm utilises a labeled training set to acquire knowledge about the distinction among two classes.

This is achieved by transforming the input data into a feature space that is nonlinear in nature [75,81]. The Support Vector Machine algorithm operates on the assumption that a hyperplane exists within the feature space. This hypothesis implies that the data exhibits linear separability. During the training process, Support Vector Machines aim to identify a hyperplane that effectively separates two distinct classes. The hyperplane in question should possess two distinct properties. There are two requirements for the separation of a dataset into two classes: (1) The dataset must be precisely divided into two distinct classes, and (2) The hyperplane used for separation should be positioned equidistantly between the two classes to maximise the margin between them. But this hypothesis lacks practicality. Hence, Support Vector Machines aim to identify the ideal centre is capable of splitting both categories exhibiting uncertainty.

*f) K-Nearest Neighbor (KNN):* This approach utilises the most fundamental form of supervised learning. This learning plan is commonly referred to as a lazy learning method [74,77]. In this methodology, the classification of the fresh sample is determined in the following manner: Initially, a comparison is made between the given sample and learning schema to identify most similar K samples inside learning plan. The individuals in discussion are commonly referred to as "neighbors." The subsequent stage involves determining the grouping of this data samples by a majority voting process among its neighbours. The key parameter, denoted as K is utilised in this approach to represent the number of training samples that are closest in proximity inside the feature space.

*2) Semi-Supervised Training:* The instructional approach includes combined collections of data with and without labels during training procedure. Hence, the utilisation of this method necessitates the implementation of a supervised learning algorithm that undergoes training using a collection of labelled data. Furthermore, it is recommended to employ an unsupervised approach to learning to generate data samples with novel labels [82,83]. The previously mentioned data samples are incorporated into the labeled set of training data for the supervised learning technique.

*3) Unsupervised learning:* This strategy involves utilising a dataset that comprises data samples for which the corresponding output is uncertain [16,33]. This suggests that the data lacks labels or identifying information. This learning technique endeavours to uncover the models and connections within information. Regarding Unsupervised training, the process involves comparing data points using a similarity metric in order to assign them to certain groups or categories. In the subsequent section, we provide many unsupervised learning techniques.

*a) K-means clustering:* The clustering algorithm adopted is easy. The main aim for K-means approach has to divide an array of n variables into subsets of k distinct clusters with each cluster being characterised by its centroid. The strategy employed via the investigation follows a iterative strategy. [84]. In early stages, a set of k random cluster centres is selected and each data point is assigned to the cluster centre that is closest to it. Once clusters are formed, ensuring that every point of data in the database are assigned to particular groups with a trend centroid can computed for every single collection and that implies the collection centres undergo updates during each iteration. The procedure is iteratively executed until there is no change in any of the cluster centres.

*b) Fuzzy-c-means (FCM):* The clustering method applied in this study depends upon the principles of fuzzy logic. In this particular methodology, it is possible for each individual sample to be assigned to several clusters. The Fuzzy-c-means algorithm identifies clusters by considering several similarity measures, such as distance. It should be noted that the method for clustering may utilise one or more similarity scales, and the choice of scale relies on the specific application or dataset being considered. The clustering procedure is iteratively performed to identify the optimal cluster centres. Like the K-means clustering technique, the FCM approach necessitates a previous understanding the overall quantity of segments [84].

*c) Hierarchical clustering:* HC seeks to categorise information points into clusters founded on their resemblance so that each member of a cluster exhibits a higher degree of resemblance with respect to one another when compared to data elements in different clusters [84]. The procedure is executed via a pair of methods: top-to-down, also known as Bottom-to-Up, as well as assortment grouping is the opposite of contentious grouping. Regarding contentious grouping the initial step involves the placement of all data points into a single group. Next, the previous group is subdivided by smaller parts. This process persists until every individual sample has been allocated to a distinct category. In the context of agglomerative clustering, the initial step involves assigning each individual sample to an individual cluster. Subsequently, like groupings are consolidated to form larger groups. The previous method persists until every point of data has been allocated to a single category. The hierarchical clustering approach does not require any prior knowledge on the number of groups. The implementation of this technique is easy.

*4) Reinforcement learning:* The previously discussed learning framework facilitates the acquisition of optimal behaviour by computers or agents in certain circumstances through the utilisation of past experiences [12,24]. A framework derived from reinforcement learning exhibits constant development by engaging in interactions with the environment and gathering data to carry out its tasks [85]. Throughout this period, a multitude of approaches have been proposed to address the challenge of reinforcement learning. An illustration of computational techniques. For instance, adaptive designing and advanced reinforcement training. Next represents an analysis of various significant reinforcement training techniques.

*a) Dynamic programming (DP):* The approach contains a collection of techniques for determining the best option of action within an extensive depiction relating to our surroundings, like an evolutionary selection procedure.

*b) Q-learning:* The algorithm in discussion is widely recognised and extensively utilised in the field of reinforcement learning. Q-Learning facilitates the discovery of optimal actions by an agent. Within this approach, there exists a designated

*International Journal on Recent and Innovation Trends in Computing and Communication*
***ISSN: 2321-8169 Volume: 11 Issue: 9***
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

table referred to as the Q-Table. The table in discussion is responsible for storing action-state pairings and their respective values. For the surroundings of such display, action-state pairings are alluded called sources. While Q-Value represents corresponding outcome. The primary objective of Q-learning is to optimise Q-Value [12,24].

*c) Advanced Reinforcement Training:* The approach involves the integration of both reinforcement training and advanced training approaches. This methodology can be employed to address a multitude of intricate problems [12,24]. The utilisation of this technology enhances the cognitive capabilities of the agents. This enhances their capacity to optimise the policy. Reinforcement learning refers to a type of machine learning methodology that is capable of functioning autonomously without the need for a pre-existing database. Hence, within the realm of Deep Reinforcement Learning agents have the capability to generate a dataset by engaging in interactions with the environment. Subsequently, the database mentioned above is employed for the purpose of training deep networks in the context of Deep Reinforcement Learning.

### *C. Varieties of Assessment techniques*

From the suggested category, machine learning approaches in the medical domain are categorised obsessed by dual primary groups, namely Simulation founded assessment and realistic operation founded assessment and table 4 presents a comparison of the evaluation techniques.

*1) Simulation-based assessment:* Many machine learning founded frameworks developed in medical sector utilise simulation tools for the purpose of assessing their effectiveness, as these technologies are more readily accessible compared to practical deployment. Additionally, they possess a greater degree of adaptability and contribute to cost reduction. To assess the performance of models based on machine learning, it is important to employ appropriate replication devices like Matlab, Weka and R can replicate the learning process and ascertain its effectiveness. The learning models are assessed using a range of evaluation metrics. In broad terms, evaluation criteria can be classified towards dual primary groups:

*a) Bias measurements*: The metrics assess efficiency of machine learning model in terms of its capacity to rank or differentiate between two groups. The discrimination scales of utmost significance includes receiver operating characteristic, area under the roc curve, f1-score, sensitivity and specificity.

*b) Calibration scales:* The measures are utilised to assess the degree of concordance between projected outcomes and real results. In real-life scenarios, the significance of these scales lies in their ability to assess projected revenues or expenses. In instances where the mortality rate associated with a surgical procedure exceeds the mortality rate in the absence of an operation, the doctor may opt to skip the operation and quit treatment.

*2) Practical implementation-based assessment:* The evaluation of machine learning-based models in the medical field holds significant importance due to its practical implementation, as it enables the assessment and analysis of learning models within real-life scenarios. However, the expense associated with building learning models is high due to the inherent hardware difficulties involved. Reiterating settings and doing diverse trials pose significant challenges. In the framework of actual application, it is necessary to assess the model of learning in a dynamic fashion, consistently revising and validating it. Several crucial factors to consider through realistic application of ML techniques in medical sector encompass the capacity of ability to be generalised to novel data, user feedback, the level of trust that healthcare professional's places in the designed framework, assessing the effectiveness of the model against domain experts, and contrasting framework routine against further previous frameworks.

TABLE IV. ANALYSIS OF VARIOUS ASSESSMENT TECHNIQUES.

| **Method** | **Price** | **Accessibility** | **Assessment Outcome** | **Execution** |
|---|---|---|---|---|
| Simulation-founded assessment | L | H | Outcomes can appear implausible | simple |
| Practical implementation-founded assessment | H | H | Outcomes appear more feasible | complicated |

*L - LOW, H - HIGH*

### *D. Utilisations*

The suggested categorization, machine learning approaches utilised in the healthcare domain are categorised into two primary groups, namely diagnosis and treatment depending on their respective applications.

*1) Diagnosis:* This area holds significant importance within the medical industry. It possesses the capacity to gain insights from ML to assist clinicians in the early diagnosis of diseases, hence minimising the time required for detection. Machine learning has the potential to enhance medical imaging, analyse laboratory findings, split and recognise features inside pictures, diagnose illnesses and their extent, as well as analyse impulses via equipment including cardiograms over cardiac abnormalities and EEG to evaluate cognitive function.

*2) Treatment:* Machine learning algorithms have demonstrated potential in aiding the management and treatment of many diseases. It possesses the capacity to gain insights from ML utilising several applications such as determining appropriate dosages, tailoring therapy to individual patients, tracking the effects of medication, and forecasting the advancement of the disease. These technologies have the potential to decrease treatment expenses, lower costs associated with drug manufacturing, enhance the treatment process, expedite the identification of suitable medications, and address issues arising from the scarcity of specialised medical practitioners. ML can also encompass the application of surgical procedures to assist in the execution of highly complex surgeries that are challenging for human surgeons to do.

*International Journal on Recent and Innovation Trends in Computing and Communication*
***ISSN: 2321-8169 Volume: 11 Issue: 9***
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

## V. EXPLORING VARIOUS MACHINE LEARNING-BASED TECHNIQUES IN HEALTHCARE

In this part, we provide some techniques using machine learning in medicine that are based on the model described in this research, as well as their drawbacks and merits.

*1) FCMIM-SVM*

Li et al. [86] developed a machine learning-based system designed to identify heart failure conditions. The authors put out a feature selection technique referred to as FCMIM. Furthermore, the authors conducted an analysis of various learning methodologies, including Artificial Neural Networks, Support Vector Machines, Decision Tree, Naive Bayes, K-Nearest Neighbour and Logistic Regression to construct ultimate training framework. At last, researchers developed the ultimate educational framework known as FCMIM-SVM. It utilises a Cleveland-related heart disease dataset.

**Drawbacks:** When contemplating a limited database, which lacks generalizability and the ability to define the degree of disease, it is important to acknowledge the potential for prolonged runtime.

**Merits:** When analysing the process of data initial processing, it is important to employ a suitable feature selection strategy. It is also necessary to compare the anticipated assortment of attributes approach with further existing designs. Additionally, estimating various categories is crucial to construct a final learning framework that achieves high precision.

*2) Nested Ensemble Method (NE)*

The nested ensemble method was presented by Abdar et al. [87] as an automated approach to identifying breast cancer. The NE method consists of two layers containing classifiers and meta-classifiers. The nested ensemble utilised the breast cancer Wisconsin diagnostic database.

**Drawbacks:** The lack of an appropriate feature selection approach, failure to provide a clear methodology for selecting ten characteristics that impact breast cancer, use of a limited database, absence of generalizability and failure to validate the suggested method using other accessible databases are notable shortcomings.

**Merits:** The final model was designed through the execution of complete tests and ensuring a high level of precision.

*3) A LOG and RF-Based Integrated Model*

Qin et al. [88] proposed an ML approach to prompt the revealing of chronic kidney disease. Initially, the researchers employed the K-nearest neighbours imputation method to approximate the absent values inside the database. Additionally, optimal subset regression and random forest were employed to effectively reduce dimensionality & identify the major appropriate items within the schema. Subsequently, a framework for training was developed by employing a range of classifiers. ML source utilised via CKD record provided at the University of California Irvine.

**Drawbacks:** When analysing a limited database, one must acknowledge its lack of adaptability and its difficulty to accurately define the extent of the condition.

**Merits:** When evaluating a suitable approach for predicting data that is missing, it is important to do thorough testing in order to develop the final model effectively.

*4) HMANN*

In this paper Ma et al. [89] indicated a refined Neural Network architecture known as HMANN. This methodology is employed for the purpose of discovering, categorising & characterising severe kidney damage. The system deployed on Internet of medical things boards and the approach integrates the Support Vector Machines, Multi-Layer Perceptron & back propagation system and utilisation of images from the UCI chronic renal disease dataset is employed by the authors for the purpose of training and evaluating the performance of HMANN.

**Drawbacks:** The tool for simulation was not adequately explained, and the dataset was not described. Additionally, there were not enough tests conducted to thoroughly analyse the final model.

**Merits:** The use of several pre-processing procedures significantly enhances the overall accuracy of the system.

*5) CWV-BANN-SVM*

Abdar and Makarenkov [90] explored the development of an expert system designed to identify breast cancer. The approach employed in this study utilises an ensemble learning methodology that combines Support Vector Machines & Artificial Neural Network techniques. The appropriate parameters of Support Vector Machines are identified through several experimental procedures in this approach. The ensemble system comprises of two Support Vector Machines, Multi-Layer Perceptron & Radial Basis Function Neural Network. Boosting technique is employed to improve the efficiency of Neural Networks and the writers employed Wisconsin Breast Cancer Dataset for their study.

**Drawbacks:** When examining a limited database, the lack of adaptability is apparent. In addition, the suggested approach has not been tested using other databases that are readily available. Furthermore, the evaluation of Support Vector Machines with distinct kernel purposes hasn't been conducted. The reasoning behind selecting SVM, MLP & RBF to build collective approach hasn't been provided. Moreover, an appropriate preprocessing of data approach has not been outlined. Lastly, the time required for execution of the final system for learning has not been reported.

**Merits:** This study aims to investigate the high accuracy in ranking the helpful aspects associated with breast cancer

TABLE V. THE MOST SIGNIFICANT FEATURES OF MODELS BASED ON SUPERVISED TRAINING.

| Plan | Goal | Initial data preparation techniques | Learning model | Evaluation measures |
|---|---|---|---|---|
| [86] | Detection of cardiovascular illness | Fuzzy C-Means Image Segmentation algorithm | support vector machine | 92.37% of accuracy 89% of sensitivity 98% of specificity 90% of mcc |

| | | | | |
|---|---|---|---|---|
| [87] | Breast carcinoma detection | Distinct choices technique | Multimodal learning framework | 98.07% of accuracy 98.10% of precision 98.10% of recall 98.10% of f1 score 97.60% of roc |
| [88] | CKD disease diagnosis | KNN estimation for inaccurate value estimation, ideal split failure & RF for valuable feature selection | Combining logistic regression & random forest | 99.83% of accuracy 99.84% of sensitivity 99.80% of specificity 99.86% of f1 score |
| [89] | Detection, segmentation & identification of kidney stones | Eliminating noise, providing an approach for feature extraction termed GLCM, and proposing a technique for distinct choices | multi-layer perceptron & support vector machine | 97.5% of accuracy 99.7% of prediction rate 0.922 of area under the curve |
| [90] | Breast Cancer Detection | Eliminating missing values from data points and determining the significance of every characteristic using CPG-support vector machine | Multimodal learning framework | 100% of accuracy 100% of sensitivity 100% of specificity 100% of precision 1.0 of area under the curve |

*6) SS-BLSTM*

Gupta et al. [91] introduced an approach known as SS-BLSTM, which utilises a recurrent neural network. The objective of the Semi-Supervised methodology to isolate references pertaining to drug-interrelated side effects reported on social media. During the supervised learning stage, the researcher's utilised Adverse Drug Reaction schema compiled from social media.

**Drawbacks:** This study did not consider the runtime aspect did not employ multiple DL techniques to assess the training framework and didn't examine an appropriate initial data preparation technique.

**Merits:** Evaluating the training framework's performance is conducted by analysing its efficacy across various datasets and under various scenarios.

*7) A Deep Learning Model for Image Segmentation of the Retina Cornea*

The Authors [92] proposed a DL framework to perform visual disc segmentation in the visual cortex. The approach employed in this study incorporates two distinct learning strategies, namely Semi-Supervised training and transferral training and they utilise a range of databases. The databases under investigation include 1) The diabetic eye disease resource provided by kaggle & used the already labelled dataset to train the auto-encoder network. 2) DRISHTI GS1 database is a significant resource in the field. The dataset is employed by the authors for the purpose of training the separation network. 3) RIM-ONE database is a significant resource in the field. Professionals analyse and partition these pictures, afterwards assessing the optical density inside them. The data set is used for the segmentation setup.

**Drawbacks:** The omission of several deep learning methods for evaluating the learning model, the failure to incorporate an appropriate pre-processing plan, and the requirement of a substantial amount of effort for the training process are notable limitations.

**Merits:** When assessing a learning framework over several datasets, it's vital to consider multiple criteria in order to assess its performance. These criteria include achieving a high level of accuracy while maintaining an acceptable duration.

*8) Deep learning architecture that combines temporal encoding with long short-term memory*

They [93] introduced a Semi-Supervised training approach for the purpose of detecting bodily functions through the utilisation of inertial sensors. The approach employed in this study involves the utilisation of deep long short-term memory for the purpose of extracting key characteristics. In this study, the researchers employed the UCI database.

**Drawbacks:** The evaluation of the final model is hindered by a lack of adequate experiments. Additionally, the impact of several variables on ultimate training framework hasn't been considered. Furthermore, an evaluation of learning model has not been conducted using multiple deep learning methods.

**Merits:** The present study aims to design a semi-supervised deep learning methodology that achieves a notable level of accuracy while maintaining a reasonable duration.

*9) Generative Adversarial Networks Founded Semi-Supervised Training Technique*

Using generative adversarial networks the experts [94] developed a Semi-Supervised training technique. This approach aims to enhance medical findings in the healthcare sector based on IoT. This method can address two issues, namely the lack of labeled diagnostic data and the disparity of groups, ten UCI symmetrical schemas & ten UCI asymmetrical schemas are utilised by authors.

**Drawbacks**: The neglect of many fundamental techniques for evaluating the learning model, the absence of any rationale for utilising SVM and KNN as primary classifiers, the failure to develop an appropriate pre-processing method, and the lack of description about the procedure for choosing features.

**Merits:** When assessing the efficacy of the learning approach across several datasets, it is observed that the model consistently achieves a high level of accuracy.

*10) Semi-Supervised Learning Based ECG Categorization Framework*

The experts [95] recommended a Semi-Supervised training framework to facilitate the categorization of electrocardiogram signals and the primary intent of the categorization process is to expose and diagnose the occurrence of arrhythmia & the present learning challenge pertains to the classification of time series signals characterised by imbalanced classes. It can be divided into three categories: There are Normal heartbeats, Supraventricular abnormal heartbeats and abnormal atrial heartbeats this strategy's objective aims to accurately determine the existence of Supraventricular abnormal heartbeats and abnormal atrial heartbeats through an analysis of electrocardiogram data without the need for explicit labelling. It should be noted that experts employ a Two-Dimensional CNN within framework. Two datasets are utilised for the purpose of modelling the system: (1) Repository of arrhythmias developed by MIT and the BIH a well-known collection of heart rate data, and (2) an unlabeled database, which frequently contains data samples representing normal heartbeats.

**Drawbacks:** The study fails to consider the execution dimension, neglects the use of diverse artificial neural networks for assessing the learning model, overlooks the consideration of an appropriate pre-processing scheme, and lacks an adequate number of tests to assess the final model.

**Merits:** Achieving high precision in classifier design entails the use of many classes and employing a Semi-Supervised training approach to iteratively renovate the anticipated labels.

TABLE VI. THE MOST SIGNIFICANT FEATURES OF MODELS BASED ON SEMI-SUPERVISED TRAINING.

| Plan | Goal | Initial data preparation methods | Learning model | Evaluation measures |
|---|---|---|---|---|
| [91] | ADR statement extraction from Twitter | Method of data normalisation | DNN and semi-supervised learning technique | 75.1% of f1-score 73.1% of precision 77.4% of recall |
| [92] | Image separation of the retinal fundus | Method of data normalisation & increasing data sample size | Deep neural network, semi-supervised learning method and transfer learning | Drishti schema 0.99% of accuracy 0.9539 of sensitivity Rmi-one schema 0.99% of accuracy 0.873 of sensitivity |
| [93] | Utilising a Semi-Supervised training technique to extract high-level features | Technique of feature extraction | Deep neural network and Semi-Supervised learning technique | 97.21% of accuracy 2.118 secs of runtime |
| [94] | Developing a decision support system for medical use | Increasing sample sizes of data & the procedure of feature selection | Method of Semi-Supervised learning | Symmetric schema 90.9% of accuracy Asymmetric schema 87.2% of accuracy |
| [95] | Identifying normal beats, Supraventricular abnormal heartbeats, and Abnormal atrial heartbeats from an Unlabeled schema | Method of information normalisation | CNN and the Semi-Supervised learning technique | Supraventricular abnormal heartbeats 97.4% of accuracy 93.38% of sensitivity 97.2% of specificity 72.5% of f1-score Abnormal atrial heartbeats 98.6% of accuracy 87.5% of sensitivity 99.4% of specificity 89.2% of f1-score |

*11) A medically founded detection approach for societal anxiety syndrome*

The Author's [96] developed a healthcare assistance approach for societal anxiety syndrome detection (SAD) and they identified chaotic datasets via the self-organizing map. Utilising a system that adapts to neuro-fuzzy inference and SAD is detected. Through an online platform, the researchers obtain original unmodified data in this approach.

**Drawbacks**: Never assessing the efficacy of the training framework utilising multiple clustering strategies not analysing the efficacy of the training framework founded various categorizers and not determining the execution time, conducting inadequate tests to assess the ultimate model and doesn't verify the training framework using further accessible records.

**Merits:** High precision, development of an appropriate data preprocessing technique.

*12) UDR-RC*

The Unsupervised DL Aided Reconstructed Coder was introduced by Janarthanan et al. [97]. This method's objective is to provide a data preprocessing technique designed to optimise the dataset. UDR-RC makes use of the WISDM database.

**Drawbacks:** Lack of trials to assess the final framework, and a failure to verify the framework for learning with other accessible datasets.

**Merits:** Developing an appropriate data preprocessing method, minimizing time for computation, decreasing error rate, and achieving high precision.

*13) Hybrid fuzzy clustering technique*

On brain MRI images, Kanniappan et al. [98] separated aberrant regions. They modelled a semi-automatic method to identify standard and aberrant regions in every brain MRI scan using fuzzy clustering. The researchers utilised two magnetic resonance imaging schemas: (1) an actual health record and (2) a BRAT'S record.

**Drawbacks:** The assessment of the model's efficacy is conducted regarding alternative clustering approaches. The amount of time required for implementing the suggested approach is calculated. The presence of a high error rate is observed, along with the absence of a noise removal method specifically designed for MRI pictures. The rapidity of clustering is not evaluated. The analysis of the model is performed utilising a large dataset.

**Merits:** The efficacy of the model was assessed by various experiments, employing the Symmetry score to figure out the

*International Journal on Recent and Innovation Trends in Computing and Communication*
***ISSN: 2321-8169 Volume: 11 Issue: 9***
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

correct group dimensions & assessment of model's efficacy was conducted alongside its following post-processing procedures.

*14) Computational algorithms used in data analysis and machine*

The experts [99] introduced a bunching founded accusation method known as Clustimp and they made use of dual records, namely aggregate radiographic sample and the human breast carcinoma sample.

**Drawbacks:** A learning strategy wasn't evaluated using massive datasets, and runtime calculations weren't performed.

**Merits:** The goal of the research seems to create a appropriate data pre-processing methodology for the estimation of missing values and the reduction of error rates.

*15) AFGC*

Huang [100] proposed a flexible rapidly generalised Fuzzy C-means Clustering Algorithm and goal of the approach is effectively partition nodules of the thyroid pictures within a context of high noise levels with the aim of precisely identifying dangerous thyroid tumors. They utilised the Jinshan Hospital database which encompasses a collection of nodules in the thyroid pictures.

**Drawbacks:** Failure to implement an appropriate pre-processing technique and neglecting to evaluate the learning framework using additional accessible sources.

**Merits:** This study aims to propose a noise removal strategy that achieves high precision in segmentation while considering computational efficiency. Additionally, the efficacy of the learning model will be evaluated under various noise conditions.

TABLE VII. THE MOST SIGNIFICANT FEATURES OF MODELS BASED ON UNSUPERVISED TRAINING.

| **Plan** | **Goal** | **Initial data preparation methods** | **Learning model** | **Evaluation measures** |
|---|---|---|---|---|
| [96] | The identification of Social nerves syndrome | min-max, version 18.0 of IBM modeller & SOM technique to recognize data with noise | Supervised learning & Unsupervised learning | 98.67% of accuracy 97.14% of sensitivity 10% of specificity |
| [97] | Developing a Human Activity Recognition system | FT aimed at establishing information models and eliminating noisy data, developing a framework for extracting features founded the Coder design & z-layer technique | Unsupervised learning & Supervised learning | 97.5% of accuracy 0.52% of mse 11.255 n of runtime |
| [98] | Brain MRI segmentation and detection of brain tumour | Data standardisation approach | Unsupervised learning (fuzzy-clustering) | 28.802 of psnr 0.717 of ncc 0.455 of nae 0.814 of ssim 0.88 of disc |
| [99] | Developing a decay technique for missing value estimation | Utilising Clustimp for managing absent values | Unsupervised learning & Supervised learning | Aggregate radiographic sample Error rate: 9–11% Human breast carcinoma sample Error rate: 4–5% |
| [100] | Classifying pictures of thyroid nodules to detect Papillary Thyroid Cancers. | No data preparation methods | Unsupervised learning | 99.87% of sa 99.78% of cs |

*16) A Healthcare Processing System Using Closed-Loop*

Dai et al. [101] employed deep neural networks to replicate the body of a person and applied DRL techniques for identifying appropriate therapy strategies to replicated physique. The approach involves utilising a replicated body as a representation of a patient, while employing deep reinforcement learning as a surrogate for a doctor. The researchers fail to give precise explanations for the database; rather they have used tongue images as a way for implementation.

**Drawbacks:** The evaluation of the final model is limited due to a lack of appropriate tests. Additionally, a suitable pre-processing methodology was not taken into consideration, and the proposed method was not tested with other existing databases.

**Merits:** The integration of reinforcement learning and deep learning techniques is employed in the development of a comprehensive learning model that combines disease diagnostic and therapy suggestion processes.

*17) HQLA*

Khalilpourazari and Hashemi [102] proposed a novel algorithm known as HQLA, which is based on reinforcement learning principles. The method employed in this study utilises the Quebec database to forecast the prevalence of the Coronavirus. The approach employed in this study includes two methodologies, namely reinforcement learning and evolutionary algorithms.

**Drawbacks:** The evaluation of the final model is hindered by a lack of adequate tests. Additionally, there is a failure to consider an appropriate preliminary processing method, along with neglecting to account for various aspects including age, gender, and other relevant factors in the construction of the final learning framework.

**Merits:** The final framework for learning is designed by including both reinforcement learning and metaheuristic techniques. The selection of the optimal operator is achieved through the utilisation of Q-learning.

*18) SRL-RNN*

The authors [103] introduced an ML framework called SRL-RNN. This approach employs reinforcement learning in conjunction with a Recurrent Neural Network and the main goal of Semantic Role Labelling Recurrent Neural Network framework is for solving the inherent difficulty associated with the flexible therapy paradigm issue & primary objective of this approach is to concurrently integrate two signals, namely an indication and an evaluation. The researchers use the MIMIC-3 v1.4 database, which is extensive and easily accessible to evaluate its efficacy for SRL-RNN.

**Drawbacks:** This modelling software was not adequately explained, and there were not enough experiments conducted to thoroughly analyse ultimate framework. Also proper initial preparation technique wasn't taken into consideration.

**Merits:** By employing a combination of replication founded assessment and applied execution founded assessment, this study aims to assess the performance of the learning model. The evaluation process will involve utilising a substantial database for testing and training purposes.

*19) tVAE*

Transitional Variational Auto-Encoders were introduced by Baucum et al. [104]. The objective is to acquire knowledge about the development of an illness to establish a correlation between someone's current condition and their anticipated state in the subsequent time interval. The researchers exploited the MIMIC database.

**Drawbacks:** The evaluation of the resultant model is limited due to a lack of adequate experiments. Additionally, there is a failure to consider an appropriate pre-processing scheme, the impact of multiple variables in ultimate training framework & influence of schema dimension on model routine and utilisation of several RL procedures to assess training mode's effectiveness.

**Merits:** The utilisation of ongoing samples is considered when developing the model, employing an arbitrary path within the framework of on-policy reinforcement learning technique.

*20) GAN + RAE + DQN*

The experts [105] suggested a comprehensive approach founded Deep Reinforcement training to facilitate the process of determining choices regarding treatment. This methodology comprises three constituent elements: 1) This study focuses on the utilisation of generative adversarial networks to the purpose for producing synthetic data from a limited dataset. 2) The use of a transition deep neural network is aimed at developing a virtual radiation environment. 3) The Deep Q-Network algorithm is utilised to determine the appropriate dose of radiation for the radiotherapy procedure. The present study used a database of persons identified with Non-Small Cell Lung cancer.

**Drawbacks:** The number of experiments conducted to assess the end product is inadequate, resulting in an insufficient amount of data available for evaluation. Thus, the final product has a significant error frequency.

**Merits:** The completed learning framework is designed by including both reinforcement learning and deep learning techniques. Additionally, the dataset size is augmented using Generative Adversarial Networks. Furthermore, an appropriate feature selection strategy is employed.

TABLE VIII. THE MOST SIGNIFICANT FEATURES OF MODELS BASED ON REINFORCEMENT TRAINING.

| **Plan** | **Goal** | **Initial data preparation methods** | **Learning model** | **Evaluation measures** |
|---|---|---|---|---|
| [101] | Creating a virtual organism and virtual physician | No data preparation methods | RNN & RL | No evaluation measures |
| [102] | Predicting the development of a covid-19 pandemic | No data preparation methods | RL | 6.29 × 10−6 of mse |
| [103] | Developing treatment suggestions for DTR | Eliminating certain points of data with many values that are missing and approximating others with a restricted amount of entries are absent | RNN & advanced RL | 0.409 of jaccard coefficient 0.157 of mortality rate |
| [104] | Simulating fake patients and the virtual treatment strategy | Using hold and sample iteration & ANN to estimate values that are missing | ANN & DRL | 12.15 of mae |
| [105] | Developing a virtual radiotherapy setting and figuring out the optimal dose of radiation for pulmonary carcinoma treatment | BN grid idea for identifying valuable characteristics | RNN & advanced RL | 100% of accuracy 0.76 of rmse |

## VI. DISCUSSION

According to this part, we bring you a discussion on the ML founded approaches in the medical domain, focusing on the learning models that were analysed in Section 5. It is important to acknowledge that datasets in the healthcare domain frequently encounter a range of challenges, including missing values, noisy information, high data complexity (i.e., an excessive number of features), and other related issues. These issues diminish the general standard of datasets. This issue negatively impacts the efficacy of machine learning-based models. From the details highlighted throughout this investigation, this may be deduced as many ML methodologies employed in the area of medicine highlight the utilisation of data pre-processing approaches. The prevalence of missing values is a prominent issue encountered in medical schemas. According to machine learning founded methodologies examined in the research, it is observed that there exist two primary approaches for addressing this issue: There are two primary approaches to handling missing data in research studies: (1) the deletion of data points that contain absent standards and (2) the estimation of absent standards. qin et al. [88], wang et al. [103], baucum et al. [104] and savarimuthu & shobha [130] provided a variety of estimation strategies for values that are missing. li et al. [86], abdar and makarenkov [90] and wang et al. [103] eliminate information from database. It is an easy strategy to resolve this issue, but it may give rise to a new issue known as unbalanced groups. This issue is detrimental to the efficacy of learning models. Thus,

*International Journal on Recent and Innovation Trends in Computing and Communication*
***ISSN: 2321-8169 Volume: 11 Issue: 9***
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

methodologies that fill in missing values offer an improved solution to this problem. When developing a technique for approximating values that are missing, however, it is crucial to figure out missing values precisely. Otherwise, the effectiveness of the training framework wasn't acceptable. Wang et al. [103] proposed a combination solution to this problem. This implies that certain information models through numerous absent standards were eliminated from the records, while others with fewer absent standards were implicated. Also, data normalisation is considered by the majority of ML-based methodologies. The goal of data normalisation is to standardize variables with various scales within a certain range, such as [0, 1], so that they have the same impact on the approach to learning. Several studies have employed data normalisation approaches, including li et al. [86], baucum et al. [104], gupta et al. [91], zhai et al. [95], bengani et al. [92], kanniappan et al. [98], fathi et al. [96] & janarrhanan et al. [97]. Noise is an additional factor. The issue pertaining to medical datasets. The introduction of noise into algorithms for learning diminishes their precision while simultaneously augmenting the mistake made by the individual. Hence, it is of extreme importance to develop methodologies for the elimination of data that contains noise. To enhance the efficacy of machine learning models. Data encompasses various types, such as the types of data commonly utilised in research include digital photographs, numerical information, and qualitative information. The method of noise reduction exhibits variability. Based on the classification of data types inside datasets. This paper explores various methodologies for eliminating several forms of stochasticity present towards multiple schemas. For instance, ma et al. [89], fathi et al. [96], huang [100] and janarrhanan et al. [97] provided a variety of techniques for removing noise from datasets. In Section 5, we investigated these methodologies. Furthermore, the scopes associated with medical datasets are frequently high. It suggests the information segments has numerous characteristics and that can raise framework complication, learning time & the likelihood of overfitting. Utilising methods for eliminating dimensionality, like feature selection and extraction of features, is the most effective method for resolving this issue. Some studies are centred on the selection and extraction of characteristics. qin et al. [88], li et al. [86], abdar et al. [87], ma et al. [89], tseng et al. [105], zhu et al. [93], yang et al. [94], fathi et al. [96] & jannrthanan et al. [97] presented techniques correspond to diminishing complexity. But a few of the approaches examined in the review doesn't demonstrate a technique employed for the diminution of measurements. Thus, significant drawback of these approaches, as we are unable to verify the outcomes provided by these models to evaluate the impact of the approach to selecting features according to effectiveness. For instance, abdar et al. [87] & yang et al. [94] failed to clarify the feature selection procedure. Table 9 classifies ML-based methods according to data preprocessing techniques.

TABLE IX. CATEGORIZATION OF MACHINE LEARNING TECHNIQUES ON INITIAL DATA PREPARATION METHODS.

| | Initial data preparation techniques | | | | |
|---|---|---|---|---|---|
| | *Cleansing information techniques* | | | *Information minimization techniques* | |
| ***Scheme*** | ***Lacking value analysis*** | ***Chaotic information organization*** | ***Information stabilization*** | ***Selecting features*** | ***Extracting features*** |
| [86] | yes | no | yes | yes | no |
| [87] | no | no | no | yes | no |
| [88] | yes | no | no | yes | no |
| [89] | no | yes | no | yes | yes |
| [90] | yes | no | no | no | no |
| [91] | no | no | yes | no | no |
| [92] | no | no | yes | no | no |
| [93] | no | no | no | no | yes |
| [94] | no | no | no | yes | no |
| [95] | no | no | yes | no | no |
| [96] | no | yes | yes | yes | no |
| [97] | no | yes | yes | yes | yes |
| [98] | no | no | yes | no | no |
| [99] | yes | no | no | no | no |
| [100] | no | yes | no | no | no |
| [101] | no | no | no | no | no |
| [102] | no | no | no | no | no |
| [103] | yes | no | no | no | no |
| [104] | yes | no | yes | no | no |
| [105] | no | no | no | yes | no |

The sort of learning procedure utilised in the development of ML-based models is a second crucial element. According to our research in this article, unsupervised learning-based approaches are frequently employed in data preprocessing tasks. For instance, Fathi et al. [96] utilised the autonomous sound recognition map. To decrease processing time, Janarrhanan et al. [97] introduce an Unsupervised DL methodology towards the needs of extracting features, choosing features, & reducing sound. In [99], Savarimuthu and Shoha presented an Unsupervised NN towards cost prediction that is missing in a dataset. whereas Supervised training techniques is frequently managed to identify & categorise the illness, it is also possible to learn on your own. For instance, qin et al. [88], li et al. [86], abdar & makarenkov [90], abdar et al. [87], and ma et al. [89] provide training approaches. Currently, treatment suggestion engines are also designed using deep learning techniques. However, a significant issue with these techniques is that their efficacy is dependent on the labeled database. When sufficient labelled data are accessible for testing and training an algorithm for supervised learning, its outcome is satisfactory. But in the medical sector, large labeled datasets are rarely accessible. This can result in an overfitting issue. This decreases the training frameworks generalizability & raises the fault. Additionally, few experts gave resolutions to this problem. Reinforcement learning is a potential cure for such problems. wang et al. [103], dia et al. 101], tseng et al. [105], khalilpourazari & hashemi [102] and baucum et al. [104] used reinforcement learning to develop frameworks for training, for instance, to figure out the best treatment plans, a reinforcement learning approach need to constantly monitor the condition of the patient. This is the most

*International Journal on Recent and Innovation Trends in Computing and Communication*
***ISSN: 2321-8169 Volume: 11 Issue: 9***
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

significant challenge associated with the application of this technique in healthcare. Based on the content shown previously, monitoring the condition of a patient seems a particularly difficult task. Second, researchers are prohibited from conducting unauthorised experiments on a patient's body. The solution to these issues is to construct a virtual setting for frameworks based on reinforcement learning. By applying deep learning approaches, dia et al. [101], tseng et al. [105], and baucum et al. [104] developed an artificial setting for interacting with reinforcement learning-based models. Producing fake samples of data is a further remedy for data scarcity. tseng et al. [105] & yang et al. [94] applied a DNN known as Generative Adversarial Networks to generate fake tests of data & extend the preliminary set of data, respectively. Using semi-supervised learning methods is an additional solution for data scarcity. Using a mix of Labeled and Unlabeled records, these techniques construct the training framework. In addition, these approaches employ both Supervised training & Unsupervised training methods. For instance, zhu et al. [93], gupta et al. [91], zhai et al. [95], bengani et al. [92] & yang et al. in [94] designed the learning framework via semi-supervised techniques. Table 10 classifies the machine learning-based methods utilised in the medical sector based on different strategies for learning.

TABLE X. CATEGORIZATION OF MACHINE LEARNING-BASED TECHNIQUES BASED ON NUMEROUS LEARNING METHODS.

| Several learning techniques | | | | |
|---|---|---|---|---|
| ***Scheme*** | ***Supervised learning*** | ***Semi-supervised learning*** | ***Unsupervised learning*** | ***Reinforcement learning*** |
| [86] | yes | no | no | no |
| [87] | yes | no | no | no |
| [88] | yes | no | no | no |
| [89] | yes | no | no | no |
| [90] | yes | no | no | no |
| [91] | yes | yes | no | no |
| [92] | yes | yes | no | no |
| [93] | yes | yes | no | no |
| [94] | yes | yes | no | no |
| [95] | yes | yes | no | no |
| [96] | yes | no | yes | no |
| [97] | yes | no | yes | no |
| [98] | no | no | yes | no |
| [99] | yes | no | yes | no |
| [100] | no | no | yes | no |
| [101] | yes | no | no | yes |
| [102] | no | no | no | yes |
| [103] | no | no | no | yes |
| [104] | yes | no | no | yes |
| [105] | yes | no | no | yes |

In the field of healthcare, it is worth noting that study authors frequently assess the efficacy of their machine learning models by employing simulation systems. But while recognising the significance of this evaluation technique, we suggest that the results may be insufficient. The analysis of machine learning techniques in healthcare needs their review in real-life scenarios and evaluation by medical professionals and experts to detect any inherent limitations. During the present investigation merely wang et al. [103] & kanniappan et al. [98] conducted empirical evaluations of their methodologies in a real-life scenario. However, it is important to note that these investigations were subject to some limitations. It should be noted that the application of learning approaches in healthcare carries a significant financial burden. They address the challenges associated with hardware difficulties in the execution of frameworks founded on ML. Furthermore, the replication of diverse events poses a significant challenge. These challenges are frequently regarded as significant hurdles for experts in the discipline of ML as they necessitate the ongoing evaluation and updating of their own models. The ML founded techniques employed in medical domain is classified depending on the evaluation techniques employed as shown in Table 11.

TABLE XI. CATEGORIZATION OF MACHINE LEARNING-BASED TECHNIQUES BASED ON EVALUATION METHODS.

| Evaluation methods | | |
|---|---|---|
| ***Scheme*** | ***Simulation-based evaluation*** | ***Practical implementation-based evaluation*** |
| [86] | yes | no |
| [87] | yes | no |
| [88] | yes | no |
| [89] | yes | no |
| [90] | yes | no |
| [91] | yes | no |
| [92] | yes | no |
| [93] | yes | no |
| [94] | yes | no |
| [95] | yes | no |
| [96] | yes | no |
| [97] | yes | no |
| [98] | yes | yes |
| [99] | yes | no |
| [100] | yes | no |
| [101] | yes | no |
| [102] | yes | no |
| [103] | yes | yes |
| [104] | yes | no |
| [105] | yes | no |

The last discussion regarding ML-based frameworks in the medical sector is that most of the techniques using ML are used for disease diagnosis. Very few papers have been provided in the treatment area that employ machine learning approaches. Thus, experts are required to address these issues. Example, wang et al. [103], dai et al. [101], tseng et al. [105] & baucum et al. [104] have conducted studies on this topic. Table 12 presents a comparative analysis of machine learning-based approaches in the medical sector area, focusing on their application-specific aspects.

TABLE XII. CATEGORIZATION OF MACHINE LEARNING-BASED TECHNIQUES BASED ON NUMEROUS APPLICATIONS.

| Applications | | |
|---|---|---|
| ***Scheme*** | ***Diagnosis*** | ***Treatment*** |
| [86] | yes | no |
| [87] | yes | no |
| [88] | yes | no |
| [89] | yes | no |
| [90] | yes | no |
| [91] | yes | no |
| [92] | yes | no |
| [93] | yes | no |

| | | |
|---|---|---|
| [94] | yes | no |
| [95] | yes | no |
| [96] | yes | no |
| [97] | yes | no |
| [98] | yes | no |
| [99] | yes | no |
| [100] | yes | no |
| [101] | yes | yes |
| [102] | yes | no |
| [103] | no | yes |
| [104] | no | yes |
| [105] | no | yes |

The above-mentioned survey provided an in-depth review of the various uses of ML within the area of medical research. Numerous strategies were put forth explored and significant approaches have been proposed, spanning from the study of coronary illness [106] to research on pandemics [107]. The field of machine learning has experienced a significant growth in COVID-19 research, with a notable rise in the development of novel methodologies [108-115]. Ensemble, deep learning, and hybrid approaches have gained significant traction in recent years, as evidenced by earlier surveys [115–121]. The advancements regarding the utilisation of evolutionary approaches, such as those discussed in references [122-125] for algorithm training in machine learning have not seen the same level of progress as other domains.

## VII. Barriers and Unresolved Problems

In this part, we will outline several obstacles & limits that arise during the creation of strategies founded on ML.

- Accessibility of information: Machine learning models frequently necessitate extensive datasets for the purpose of training. When datasets are of a substantial size, the efficacy of these frameworks is commendable and their error rate is minimal. To tackle this issue, it is imperative to develop novel approaches for electronically capturing health-related information.
- The quality of data: An additional crucial aspect to consider that the presence of either accidental or purposeful errors through the process of gathering information has the potential to elevate the overall error rate. Hence, ensuring the accuracy of data holds significant importance. The occurrence of such issues may arise due to insufficient caution exercised by doctors and experts in the process of assigning labels to collected samples. The implementation of data preparation techniques has the potential to effectively mitigate these issues and enhance the overall integrity and reliability of the records.
- High dimensions: The medical records in real-time exhibit a high degree of dimensionality. This issue exacerbates the intricate design of the framework, accelerates the duration of the learning process, and results in excessive fitting. Hence, it is imperative for machine learning-based approaches to consistently consider this particular concern. There exist several viable strategies for the reduction of dimensionality. For instance, the utilisation of selecting features and methods to extract features has proven to be effective in addressing this challenge. Nevertheless, further investigation is warranted in this domain to develop better strategies for decreasing the complexity.
- Effectiveness: Models based on machine learning have proven to be advantageous in the medical sector when they effectively address significant challenges within this field. In certain scenarios, the utilisation of machine learning approaches may not be imperative for problem-solving, as traditional approaches can effectively address the issue at hand. Machine learning procedures are indispensable in situations where datasets exhibit high dimensionality or if every element are not readily predicted. Additionally, ML methods are beneficial when inferring accurate findings requires a significant amount of time or when conventional approaches prove to be expensive in addressing the problem at hand. Hence, it is imperative for investigators to employ contemporary and authentic machine learning methodologies.
- Security: Being aware of security concerns is crucial when constructing frameworks for machine learning, as the potential identification of individuals via unidentified information needs to be considered. The preservation of patient confidentiality is a critical and indispensable issue that warrants the attention of investigators to conduct further investigations aimed at solving this issue.

## VIII. Conclusion

In this work, we investigated ML-based healthcare techniques. To begin, we provided a concise the significance of ML to the medical sector is examined. Subsequently, an in-depth structure was presented for the purpose of constructing machine learning-based models in the field of medicine. ML-based techniques in healthcare can be categorised according to various factors. One such factor is the pre-processing of data techniques, which include cleaning of data procedures as well as data elimination approaches. Another factor is the techniques of learning employed, which encompass Unsupervised training, Semi-Supervised training, Supervised training & Reinforcement training. Additionally, assessment techniques play a crucial role, with replication founded assessment and applied execution founded assessment in actual environments being commonly utilised. Lastly, the applications of ML techniques in healthcare span across areas such as treatment and diagnosis. In conclusion, we examined many ML founded approaches in area of medicine and delineated their respective advantages and limitations. This study aims to offer investigators a complete understanding via application of ML medicinal field and acquaint them regarding the latest advancements in ML applications for medicine. The objective is to enable researchers to develop new approaches to address the current challenges in this domain. In the forthcoming era, there is a concerted effort to prioritize the utilisation of methods such as DL & RL approaches due to their formidable efficacy in addressing healthcare challenges.

## REFERENCES

[1] Aerts, A., & Bogdan-Martin, D. (2021). Leveraging data and AI to deliver on the promise of digital health. *International Journal of Medical Informatics*, *150*, 104456.

[2] Ali, O., Abdelbaki, W., Shrestha, A., Elbasi, E., Alryalat, M. A. A., & Dwivedi, Y. K. (2023). A systematic literature review of artificial intelligence in the healthcare sector: Benefits, challenges, methodologies, and functionalities. *Journal of Innovation & Knowledge*, *8*(1), 100333.

[3] Li, B. H., Hou, B. C., Yu, W. T., Lu, X. B., & Yang, C. W. (2017). Applications of artificial intelligence in intelligent manufacturing: a review. *Frontiers of Information Technology & Electronic Engineering*, *18*, 86-96.

[4] Kaplan, A., & Haenlein, M. (2020). Rulers of the world, unite! The challenges and opportunities of artificial intelligence. *Business Horizons*, *63*(1), 37-50.

[5] Chien, C. F., Dauzère-Pérès, S., Huh, W. T., Jang, Y. J., & Morrison, J. R. (2020). Artificial intelligence in manufacturing and logistics systems: algorithms, applications, and case studies. *International Journal of Production Research*, *58*(9), 2730-2731.

[6] Kumar, P., Sharma, S. K., & Dutot, V. (2023). Artificial intelligence (AI)-enabled CRM capability in healthcare: The impact on service innovation. *International Journal of Information Management*, *69*, 102598.

[7] Demirhan, A. (2017). Neuroimage‐based clinical prediction using machine learning tools. *International Journal of Imaging Systems and Technology*, *27*(1), 89-97.

[8] Datta, S., Barua, R., & Das, J. (2019). Application of artificial intelligence in modern healthcare system. *Alginates-recent uses of this natural polymer*.

[9] Elsebakhi, E., Lee, F., Schendel, E., Haque, A., Kathireason, N., Pathare, T., ... & Al-Ali, R. (2015). Large-scale machine learning based on functional networks for biomedical big data with high performance computing platforms. *Journal of Computational Science*, *11*, 69-81.

[10] Bashir, S., Qamar, U., Khan, F. H., & Naseem, L. (2016). HMV: A medical decision support framework using multi-layer classifiers for disease prediction. *Journal of Computational Science*, *13*, 10-25.

[11] Waring, J., Lindvall, C., & Umeton, R. (2020). Automated machine learning: Review of the state-of-the-art and opportunities for healthcare. *Artificial intelligence in medicine*, *104*, 101822.

[12] Coronato, A., Naeem, M., De Pietro, G., & Paragliola, G. (2020). Reinforcement learning for intelligent healthcare applications: A survey. *Artificial Intelligence in Medicine*, *109*, 101964.

[13] Yousefpoor, E., Barati, H., & Barati, A. (2021). A hierarchical secure data aggregation method using the dragonfly algorithm in wireless sensor networks. *Peer-to-Peer Networking and Applications*, *14*(4), 1917-1942.

[14] Yousefpoor, M. S., Yousefpoor, E., Barati, H., Barati, A., Movaghar, A., & Hosseinzadeh, M. (2021). Secure data aggregation methods and countermeasures against various attacks in wireless sensor networks: A comprehensive review. *Journal of Network and Computer Applications*, *190*, 103118.

[15] Rong, G., Mendez, A., Assi, E. B., Zhao, B., & Sawan, M. (2020). Artificial intelligence in healthcare: review and prediction case studies. *Engineering*, *6*(3), 291-301.

[16] Seaton.H. (2021). The Construction Technology Handbook; *John Wiley & Sons*: Hoboken, NJ, USA, 2021; ISBN 978-1-119-71995-3.

[17] Chen, Y., Liu, Q., & Guo, D. (2020). Emerging coronaviruses: genome structure, replication, and pathogenesis. *Journal of medical virology*, *92*(4), 418-423.

[18] Mohammed, M., Khan, M. B., & Bashier, E. B. M. (2016). *Machine learning: algorithms and applications*. Crc Press.

[19] Seo, H., Badiei Khuzani, M., Vasudevan, V., Huang, C., Ren, H., Xiao, R., ... & Xing, L. (2020). Machine learning techniques for biomedical image segmentation: an overview of technical aspects and introduction to state‐of‐art applications. *Medical physics*, *47*(5), e148-e167.

[20] Zhang, X. D. (2020). A matrix algebra approach to artificial intelligence.

[21] Chen, P. H. C., Liu, Y., & Peng, L. (2019). How to develop machine learning models for healthcare. *Nature materials*, *18*(5), 410-414.

[22] He, J., Baxter, S. L., Xu, J., Xu, J., Zhou, X., & Zhang, K. (2019). The practical implementation of artificial intelligence technologies in medicine. *Nature medicine*, *25*(1), 30-36.

[23] Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, *18*(8), 2674.

[24] Uprety, A., & Rawat, D. B. (2020). Reinforcement learning for iot security: A comprehensive survey. *IEEE Internet of Things Journal*, *8*(11), 8693-8706.

[25] Yu, K. H., Beam, A. L., & Kohane, I. S. (2018). Artificial intelligence in healthcare. *Nature biomedical engineering*, *2*(10), 719-731.

[26] Yousefpoor, M. S., & Barati, H. (2020). DSKMS: A dynamic smart key management system based on fuzzy logic in wireless sensor networks. *Wireless Networks*, *26*(4), 2515-2535.

[27] Alafif, T., Tehame, A. M., Bajaba, S., Barnawi, A., & Zia, S. (2021). Machine and deep learning towards COVID-19 diagnosis and treatment: survey, challenges, and future directions. *International journal of environmental research and public health*, *18*(3), 1117.

[28] Tayarani, M. (2020). Applications of artificial intelligence in battling against covid-19: A literature review. *Chaos, Solitons & Fractals*.

[29] Smiti, A. (2020). When machine learning meets medical world: Current status and future challenges. *Computer Science Review*, *37*, 100280.

[30] Shouval, R., Fein, J. A., Savani, B., Mohty, M., & Nagler, A. (2021). Machine learning and artificial intelligence in haematology. *British journal of haematology*, *192*(2), 239-250.

[31] Olsen, C. R., Mentz, R. J., Anstrom, K. J., Page, D., & Patel, P. A. (2020). Clinical applications of machine learning in the diagnosis, classification, and prediction of heart failure. *American Heart Journal*, *229*, 1-17.

[32] Berry, M. W., Mohamed, A., & Yap, B. W. (Eds.). (2019). *Supervised and unsupervised learning for data science*. Springer Nature.

[33] Mabrouk, E., Ayman, A., Raslan, Y., & Hedar, A. R. (2019). Immune system programming for medical image segmentation. *Journal of Computational Science*, *31*, 111-125.

[34] Forsch, N., Govil, S., Perry, J. C., Hegde, S., Young, A. A., Omens, J. H., & McCulloch, A. D. (2021). Computational analysis of cardiac structure and function in congenital heart disease: Translating discoveries to clinical strategies. *Journal of computational science*, *52*, 101211.

[35] Surendar, P. (2021). Diagnosis of lung cancer using hybrid deep neural network with adaptive sine cosine crow search algorithm. *Journal of Computational Science*, *53*, 101374.

[36] Saxena, A. (2021). *Artificial intelligence and machine learning in healthcare*. S. Chandra (Ed.). Singapore: Springer.

[37] Pucchio, A., Eisenhauer, E. A., & Moraes, F. Y. (2021). Medical students need artificial intelligence and machine learning training. *Nature biotechnology*, *39*(3), 388-389.

[38] Samuel, A. L. (2000). Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, *44*(1.2), 206-226.

[39] Alpaydin, E. (2014). Introduction to machine learning, 3rd editio. ed.

[40] Kubat, M., & Kubat, J. A. (2017). *An introduction to machine learning* (Vol. 2, pp. 321-329). Cham, Switzerland: Springer International Publishing.

[41] Belciug, S., Gorunescu, F., Belciug, S., & Gorunescu, F. (2020). Era of intelligent systems in healthcare. *Intelligent Decision Support Systems—A Journey to Smarter Healthcare*, 1-55.

[42] El Naqa, I., & Murphy, M. J. (2015). *What is machine learning?* (pp. 3-11). Springer International Publishing.

[43] Dulhare, U. N., Ahmad, K., & Ahmad, K. A. B. (Eds.). (2020). *Machine learning and big data: concepts, algorithms, tools and applications*. John Wiley & sons.

[44] Shobha, G.; Rangaswamy, S. (2018). *Machine learning. In Handbook of Statistics*; Elsevier: Amsterdam, The Netherlands, 2018; Volume 38, pp. 197–228.

[45] Alsuliman, T., Humaidan, D., & Sliman, L. (2020). Machine learning and artificial intelligence in the service of medicine: Necessity or potentiality?. *Current research in translational medicine*, *68*(4), 245-251.

[46] Kandhway, P., Bhandari, A. K., & Singh, A. (2020). A novel reformed histogram equalization based medical image contrast enhancement using krill herd optimization. *Biomedical Signal Processing and Control*, *56*, 101677.

[47] Zerouaoui, H., & Idri, A. (2021). Reviewing machine learning and image processing based decision-making systems for breast cancer imaging. *Journal of Medical Systems*, *45*(1), 8.

[48] Handelman, G. S., Kok, H. K., Chandra, R. V., Razavi, A. H., Lee, M. J., & Asadi, H. (2018). eD octor: machine learning and the future of medicine. *Journal of internal medicine*, *284*(6), 603-619.

[49] Vatandoost, M., & Litkouhi, S. (2019). The future of healthcare facilities: how technology and medical advances may shape hospitals of the future. *Hospital Practices and Research*, *4*(1), 1-11.

[50] Himidan, S., & Kim, P. (2015, June). The evolving identity, capacity, and capability of the future surgeon. In *Seminars in pediatric surgery* (Vol. 24, No. 3, pp. 145-149). WB Saunders.

[51] Assaf, D., Rayman, S., Segev, L., Neuman, Y., Zippel, D., & Goitein, D. (2022). Improving pre-bariatric surgery diagnosis of hiatal hernia using machine learning models. *Minimally Invasive Therapy & Allied Technologies*, *31*(5), 760-767.

[52] De Bruyne, S., Speeckaert, M. M., Van Biesen, W., & Delanghe, J. R. (2021). Recent evolutions of machine learning applications in clinical laboratory medicine. *Critical reviews in clinical laboratory sciences*, *58*(2), 131-152.

[53] Rahmani, A. M., Ali, S., Yousefpoor, M. S., Yousefpoor, E., Naqvi, R. A., Siddique, K., & Hosseinzadeh, M. (2021). An area coverage scheme based on fuzzy logic and shuffled frog-leaping algorithm (sfla) in heterogeneous wireless sensor networks. *Mathematics*, *9*(18), 2251.

[54] Lee, S. W., Ali, S., Yousefpoor, M. S., Yousefpoor, E., Lalbakhsh, P., Javaheri, D., ... & Hosseinzadeh, M. (2021). An energy-aware and predictive fuzzy logic-based routing scheme in flying ad hoc networks (fanets). *IEEE Access*, *9*, 129977-130005.

[55] Tao, W., Concepcion, A. N., Vianen, M., Marijnissen, A. C., Lafeber, F. P., Radstake, T. R., & Pandit, A. (2021). Multiomics and machine learning accurately predict clinical response to adalimumab and etanercept therapy in patients with rheumatoid arthritis. *Arthritis & Rheumatology*, *73*(2), 212-222.

[56] Alizadehsani, R., Roshanzamir, M., Abdar, M., Beykikhoshk, A., Khosravi, A., Panahiazar, M., ... & Sarrafzadegan, N. (2019). A database for using machine learning and data mining techniques for coronary artery disease diagnosis. *Scientific data*, *6*(1), 227.

[57] Ben-Israel, D., Jacobs, W. B., Casha, S., Lang, S., Ryu, W. H. A., de Lotbiniere-Bassett, M., & Cadotte, D. W. (2020). The impact of machine learning on patient care: a systematic review. *Artificial intelligence in medicine*, *103*, 101785.

[58] Yousefpoor, M. S., & Barati, H. (2019). Dynamic key management algorithms in wireless sensor networks: A survey. *Computer Communications*, *134*, 52-69.

[59] Golsorkhtabar, M., Nia, F. K., Hosseinzadeh, M., & Vejdanparast, Y. (2010, July). The novel energy adaptive protocol for heterogeneous wireless sensor networks. In *2010 3rd International Conference on Computer Science and Information Technology* (Vol. 2, pp. 178-182). IEEE.

[60] Nikravan, M., Movaghar, A., & Hosseinzadeh, M. (2018). A lightweight defense approach to mitigate version number and rank attacks in low-power and lossy networks. *Wireless Personal Communications*, *99*, 1035-1059.

[61] Dhal, P., & Azad, C. (2022). A comprehensive survey on feature selection in the various fields of machine learning. *Applied Intelligence*, 1-39.

[62] Tiwari, S. R., & Rana, K. K. (2021). Feature selection in big data: Trends and challenges. *Data Science and Intelligent Applications: Proceedings of ICDSIA 2020*, 83-98.

[63] Guyon, I.; Elisseeff, A. (2006). An introduction to feature extraction. In Feature Extraction; *Springer*: Berlin/Heidelberg, Germany, 2006; pp. 1–25.

[64] Xiong, Z., Cui, Y., Liu, Z., Zhao, Y., Hu, M., & Hu, J. (2020). Evaluating explorative prediction power of machine learning algorithms for materials discovery using k-fold forward cross-validation. *Computational Materials Science*, *171*, 109203.

[65] Xu, Z., Qin, W., Tang, Q., & Jiang, D. (2014). Energy-efficient cognitive access approach to convergence communications. *Science China Information Sciences*, *57*, 1-12.

[66] Mandal, I. (2017). Machine learning algorithms for the creation of clinical healthcare enterprise systems. *Enterprise Information Systems*, *11*(9), 1374-1400.

[67] Feldman, K., Faust, L., Wu, X., Huang, C., & Chawla, N. V. (2017). Beyond volume: The impact of complex healthcare data on the machine learning pipeline. In *Towards Integrative Machine Learning and Knowledge Extraction: BIRS Workshop, Banff, AB, Canada, July 24-26, 2015, Revised Selected Papers* (pp. 150-169). Springer International Publishing.

[68] Zhang, J. M., Harman, M., Ma, L., & Liu, Y. (2020). Machine learning testing: Survey, landscapes and horizons. *IEEE Transactions on Software Engineering*, *48*(1), 1-36.

[69] Javaheri, D., Hosseinzadeh, M., & Rahmani, A. M. (2018). Detection and elimination of spyware and ransomware by intercepting kernel-level system routines. *IEEE Access*, *6*, 78321-78332.

[70] Mesbahi, M. R., Rahmani, A. M., & Hosseinzadeh, M. (2017). Highly reliable architecture using the 80/20 rule in cloud computing datacenters. *Future Generation Computer Systems*, *77*, 77-86.

[71] Caruana, R., & Niculescu-Mizil, A. (2006, June). An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd international conference on Machine learning* (pp. 161-168).

[72] alias Balamurugan, A., Rajaram, R., Pramala, S., Rajalakshmi, S., Jeyendran, C., & Prakash, J. D. S. (2011). NB+: an improved naive Bayesian algorithm. *Knowledge-Based Systems*, *24*(5), 563-569.

[73] Miorelli, R., Kulakovskyi, A., Chapuis, B., D'almeida, O., & Mesnil, O. (2021). Supervised learning strategy for classification and regression tasks applied to aeronautical structural health monitoring problems. *Ultrasonics*, *113*, 106372.

[74] Dhasaradhan, K., Jaichandran, R., Shunmuganathan, K. L., Usha Kiruthika, S., & Rajaprakash, S. (2021). Hybrid machine learning model using decision tree and support vector machine for diabetes identification. In *Data*

*Engineering and Intelligent Computing: Proceedings of ICICC 2020* (pp. 293-305). Springer Singapore.

[75] alias Balamurugan, A., Rajaram, R., Pramala, S., Rajalakshmi, S., Jeyendran, C., & Prakash, J. D. S. (2011). NB+: an improved naive Bayesian algorithm. *Knowledge-Based Systems*, *24*(5), 563-569.

[76] Ballard, Z., Brown, C., Madni, A. M., & Ozcan, A. (2021). Machine learning and computation-enabled intelligent sensor design. *Nature Machine Intelligence*, *3*(7), 556-565.

[77] Shrestha, Y. R., Krishna, V., & von Krogh, G. (2021). Augmenting organizational decision-making with deep learning algorithms: Principles, promises, and challenges. *Journal of Business Research*, *123*, 588-603.

[78] Villarrubia, G., De Paz, J. F., Chamoso, P., & De la Prieta, F. (2018). Artificial neural networks used in optimization problems. *Neurocomputing*, *272*, 10-16.

[79] Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5-32.

[80] Sharma, S., Singh, G., & Sharma, M. (2021). A comprehensive review and analysis of supervised-learning and soft computing techniques for stress diagnosis in humans. *Computers in Biology and Medicine*, *134*, 104450.

[81] Zhang, L., Liu, P., Zhao, L., Wang, G., Zhang, W., & Liu, J. (2021). Air quality predictions with a semi-supervised bidirectional LSTM neural network. *Atmospheric Pollution Research*, *12*(1), 328-339.

[82] Bull, L. A., Worden, K., & Dervilis, N. (2020). Towards semi-supervised and probabilistic classification in structural health monitoring. *Mechanical Systems and Signal Processing*, *140*, 106653.

[83] Celebi, M.E.; Aydin, K. (2016). Unsupervised Learning Algorithms; *Springer*: Berlin/Heidelberg, Germany, 2006; pp. 1-25.

[84] Xu, X., Zuo, L., & Huang, Z. (2014). Reinforcement learning algorithms with function approximation: Recent advances and applications. *Information Sciences*, *261*, 1-31.

[85] Li, J. P., Haq, A. U., Din, S. U., Khan, J., Khan, A., & Saboor, A. (2020). Heart disease identification method using machine learning classification in e-healthcare. *IEEE access*, *8*, 107562-107582.

[86] Abdar, M., Zomorodi-Moghadam, M., Zhou, X., Gururajan, R., Tao, X., Barua, P. D., & Gururajan, R. (2020). A new nested ensemble technique for automated diagnosis of breast cancer. *Pattern Recognition Letters*, *132*, 123-131.

[87] Qin, J., Chen, L., Liu, Y., Liu, C., Feng, C., & Chen, B. (2019). A machine learning methodology for diagnosing chronic kidney disease. *IEEE Access*, *8*, 20991-21002.

[88] Ma, F., Sun, T., Liu, L., & Jing, H. (2020). Detection and diagnosis of chronic kidney disease using deep learning-based heterogeneous modified artificial neural network. *Future Generation Computer Systems*, *111*, 17-26.

[89] Abdar, M., & Makarenkov, V. (2019). CWV-BANN-SVM ensemble learning classifier for an accurate diagnosis of breast cancer. *Measurement*, *146*, 557-570.

[90] Gupta, S., Pawar, S., Ramrakhiyani, N., Palshikar, G. K., & Varma, V. (2018). Semi-supervised recurrent neural network for adverse drug reaction mention extraction. *BMC bioinformatics*, *19*, 1-7.

[91] Bengani, S. (2021). Automatic segmentation of optic disc in retinal fundus images using semi-supervised deep learning. *Multimedia Tools and Applications*, *80*, 3443-3468.

[92] Zhu, Q., Chen, Z., & Soh, Y. C. (2018). A novel semisupervised deep learning method for human activity recognition. *IEEE Transactions on Industrial Informatics*, *15*(7), 3821-3830.

[93] Yang, Y., Nan, F., Yang, P., Meng, Q., Xie, Y., Zhang, D., & Muhammad, K. (2019). GAN-based semi-supervised learning approach for clinical decision support in health-IoT platform. *Ieee Access*, *7*, 8048-8057.

[94] Zhai, X., Zhou, Z., & Tin, C. (2020). Semi-supervised learning for ECG classification without patient-specific labeled data. *Expert Systems with Applications*, *158*, 113411.

[95] Fathi, S., Ahmadi, M., Birashk, B., & Dehnad, A. (2020). Development and use of a clinical decision support system for the diagnosis of social anxiety disorder. *Computer Methods and Programs in Biomedicine*, *190*, 105354.

[96] Janarthanan, R., Doss, S., & Baskar, S. (2020). Optimized unsupervised deep learning assisted reconstructed coder in the on-nodule wearable sensor for human activity recognition. *Measurement*, *164*, 108050.

[97] Kanniappan, S., Samiayya, D., Vincent PM, D. R., Srinivasan, K., Jayakody, D. N. K., Reina, D. G., & Inoue, A. (2020). An efficient hybrid fuzzy-clustering driven 3D-modeling of magnetic resonance imagery for enhanced brain tumor diagnosis. *Electronics*, *9*(3), 475.

[98] Shobha, K., & Savarimuthu, N. (2021). Clustering based imputation algorithm using unsupervised neural network for enhancing the quality of healthcare data. *Journal of Ambient Intelligence and Humanized Computing*, *12*, 1771-1781.

[99] Huang, W. (2020). Segmentation and diagnosis of papillary thyroid carcinomas based on generalized clustering algorithm in ultrasound elastography. *Journal of Medical Systems*, *44*(1), 13.

[100] Dai, Y., Wang, G., Muhammad, K., & Liu, S. (2022). A closed-loop healthcare processing approach based on deep reinforcement learning. *Multimedia Tools and Applications*, *81*(3), 3107-3129.

[101] Khalilpourazari, S., & Hashemi Doulabi, H. (2021). Designing a hybrid reinforcement learning based algorithm with application in prediction of the COVID-19 pandemic in Quebec. *Annals of Operations Research*, 1-45.

[102] Wang, L., Zhang, W., He, X., & Zha, H. (2018, July). Supervised reinforcement learning with recurrent neural network for dynamic treatment recommendation. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 2447-2456).

[103] Baucum, M., Khojandi, A., & Vasudevan, R. (2020). Improving deep reinforcement learning with transitional variational autoencoders: A healthcare application. *IEEE Journal of Biomedical and Health Informatics*, *25*(6), 2273-2280.

[104] Tseng, H. H., Luo, Y., Cui, S., Chien, J. T., Ten Haken, R. K., & Naqa, I. E. (2017). Deep reinforcement learning for automated radiation adaptation in lung cancer. *Medical physics*, *44*(12), 6690-6705.

[105] Joloudari, J. H., Hassannataj Joloudari, E., Saadatfar, H., Ghasemigol, M., Razavi, S. M., Mosavi, A., ... & Nadai, L. (2020). Coronary artery disease diagnosis; ranking the significant features using a random trees model. *International journal of environmental research and public health*, *17*(3), 731.

[106] Ardabili, S. F., Mosavi, A., Ghamisi, P., Ferdinand, F., Varkonyi-Koczy, A. R., Reuter, U., ... & Atkinson, P. M. (2020). Covid-19 outbreak prediction with machine learning. *Algorithms*, *13*(10), 249.

[107] Pinter, G., Felde, I., Mosavi, A., Ghamisi, P., & Gloaguen, R. (2020). COVID-19 pandemic prediction for Hungary; a hybrid machine learning approach. *Mathematics*, *8*(6), 890.

[108] Mahmoudi, M. R., Baleanu, D., Band, S. S., & Mosavi, A. (2021). Factor analysis approach to classify COVID-19 datasets in several regions. *Results in Physics*, *25*, 104071.

[109] Ayoobi, N., Sharifrazi, D., Alizadehsani, R., Shoeibi, A., Gorriz, J. M., Moosaei, H., ... & Mosavi, A. (2021). Time series forecasting of new cases and new deaths rate for COVID-19 using deep learning methods. *Results in physics*, *27*, 104495.

[110] Mahmoudi, M. R., Heydari, M. H., Qasem, S. N., Mosavi, A., & Band, S. S. (2021). Principal component analysis to study the relations between the spread rates of COVID-19 in high risks countries. *Alexandria Engineering Journal*, *60*(1), 457-464.

[111] Mahmoudi, M. R., Baleanu, D., Qasem, S. N., Mosavi, A., & Band, S. S. (2021). Fuzzy clustering to classify several time series models with fractional Brownian motion errors. *Alexandria Engineering Journal*, *60*(1), 1137-1145.

[112] Ardabili, S., Mosavi, A., Band, S. S., & Varkonyi-Koczy, A. R. (2020, November). Coronavirus disease (COVID-19) global prediction using hybrid artificial intelligence method

of ANN trained with Grey Wolf optimizer. In *2020 IEEE 3rd International Conference and Workshop in Óbuda on Electrical and Power Engineering (CANDO-EPE)* (pp. 000251-000254). IEEE.

[113] Kumar, R. L., Khan, F., Din, S., Band, S. S., Mosavi, A., & Ibeke, E. (2021). Recurrent neural network and reinforcement learning model for COVID-19 prediction. *Frontiers in public health*, *9*, 744100.

[114] Yang, F., Moayedi, H., & Mosavi, A. (2021). Predicting the degree of dissolved oxygen using three types of multi-layer perceptron-based artificial neural networks. *Sustainability*, *13*(17), 9898.

[115] Anwar, F., Ejaz, M. Y., & Mosavi, A. (2020). A comparative analysis on diagnosis of diabetes mellitus using different approaches–A survey. *Informatics in Medicine Unlocked*, *21*, 100482.

[116] Nosratabadi, S., Mosavi, A., Duan, P., Ghamisi, P., Filip, F., Band, S. S., ... & Gandomi, A. H. (2020). Data science in economics: comprehensive review of advanced machine learning and deep learning methods. *Mathematics*, *8*(10), 1799.

[117] Mosavi, A., Faghan, Y., Ghamisi, P., Duan, P., Ardabili, S. F., Salwana, E., & Band, S. S. (2020). Comprehensive review of deep reinforcement learning methods and applications in economics. *Mathematics*, *8*(10), 1640.

[118] Chen, H., Heidari, A. A., Chen, H., Wang, M., Pan, Z., & Gandomi, A. H. (2020). Multi-population differential evolution-assisted Harris hawks optimization: Framework and case studies. *Future Generation Computer Systems*, *111*, 175-198.

[119] Hu, J., Chen, H., Heidari, A. A., Wang, M., Zhang, X., Chen, Y., & Pan, Z. (2021). Orthogonal learning covariance matrix for defects of grey wolf optimizer: Insights, balance, diversity, and feature selection. *Knowledge-Based Systems*, *213*, 106684.

[120] Zhang, Y., Liu, R., Heidari, A. A., Wang, X., Chen, Y., Wang, M., & Chen, H. (2021). Towards augmented kernel extreme learning models for bankruptcy prediction: algorithmic behavior and comprehensive analysis. *Neurocomputing*, *430*, 185-212.

[121] Zhao, D., Liu, L., Yu, F., Heidari, A. A., Wang, M., Liang, G., ... & Chen, H. (2021). Chaotic random spare ant colony optimization for multi-threshold image segmentation of 2D Kapur entropy. *Knowledge-Based Systems*, *216*, 106510.

[122] Tu, J., Chen, H., Liu, J., Heidari, A. A., Zhang, X., Wang, M., ... & Pham, Q. V. (2021). Evolutionary biogeography-based whale optimization methods with communication structure: Towards measuring the balance. *Knowledge-Based Systems*, *212*, 106642.

[123] Dehghani, E., Ranjbar, S. H., Atashafrooz, M., Negarestani, H., Mosavi, A., & Kovacs, L. (2021). Introducing Copula as a Novel Statistical Method in Psychological Analysis. *International Journal of Environmental Research and Public Health*, *18*(15), 7972.

[124] Shan, W., Qiao, Z., Heidari, A. A., Chen, H., Turabieh, H., & Teng, Y. (2021). Double adaptive weights for stabilization of moth flame optimizer: Balance analysis, engineering cases, and medical diagnosis. *Knowledge-Based Systems*, *214*, 106728.