Machine Learning and Deep Learning in Health Informatics: Advancements, Applications, and Challenges

CHAIMAE TAOUSSI, IMAD HAFIDI, and ABDELMOUTALIB METRANE

Abstract—**Biomedicine's information storage and consumption are being optimized by the emerging area of medical informatics, which makes use of public health information, computation, and technology. Simultaneously, the demand for individualized care, rising healthcare costs, and the quantity of data available have led to the prevalent integration of big data into healthcare. Within the field of health informatics, machine learning and deep learning are essential because they provide researchers with improved decision-making skills in areas like clinical data mining, diagnosis, and prediction. Regarding a recent thorough systematic review, our paper explores four major machine and deep learning areas in medical informatics. These include advancements in medical diagnosis, disease prediction, enhanced medical care, and online health management, with a specific emphasis on the precise and timely prediction of medical conditions. The large-scale AI models, characterized by large parameter and data scales, perform exceptionally well after pre-training. This convergence not only enriches medical informatics but also opens up new horizons for the continuous improvement of healthcare and medical research. The article also discusses the limits and constraints associated with using these advanced techniques.**

Index Terms—**Health Informatics, Machine Learning, Deep Learning, Big Data, Technology, Pathologies Prediction.**

I. Introduction

INFORMATION science, computer science, and health care are all intersected by the quickly developing interdisciplinary field of health informatics. To capture the information-related difficulties involved in gathering, evaluating, and sharing healthcare information, the term "medical informatics" was created. Medical informatics is a broad term that encompasses many different fields and refers to the use of information and communication technologies in the creation, processing, and dissemination of medical knowledge and information. Many different fields, such as clinical medicine, preventive medicine, public health, nursing, and education management, are affected by its broad influence [1].

Medical or health informatics, a dynamic and transformative research frontier, stands as the cornerstone of a paradigm shift in traditional healthcare, leveraging the power of cutting-edge technologies to redefine the delivery and management of health services. At its essence, health informatics orchestrates a sophisticated fusion of informatics, technology, and information within the public health domain, weaving a comprehensive tapestry that encompasses prevention, vigilant surveillance, and strategic health promotion. Its overarching objective is to meticulously optimize the storage and seamless utilization of information within the intricate landscape of biomedicine [2].

In the current epoch, the meteoric rise of Machine Learning (ML) within the realm of healthcare serves as a beacon of innovation, capturing the imagination of scholars and medical practitioners alike. The pervasive influence of Machine Learning techniques extends deep into the core of medical information systems, playing a pivotal role in critical decision-making processes such as precise diagnosis, proactive prediction, and the nuanced exploration of vast clinical datasets [3].

The systematic review conducted between 2018 and 2024 synthesizes studies that investigate the use of ML and DL in healthcare. It focuses on key algorithms, such as Support Vector Machines (SVM), Convolutional Neural Networks (CNN), and Multi-Layer Perceptrons (MLP), and explores their applications in diagnosing diseases like cancer, cardiovascular conditions, and mental health disorders. The review also highlights gaps in the literature and proposes future research directions, emphasizing the importance of addressing current limitations.

Embedded within this dynamic evolution is the burgeoning field of Health Informatics (HI), an expansive arena that transcends traditional research boundaries. HI emerges as an indispensable catalyst for groundbreaking research, fueled by the exponential surge in digital health data meticulously curated by biomedical and health research institutions. This vast reservoir includes but is not limited to Electronic Health Records (EHR), the intricate visual landscape of biomedical imaging, the intricacies of bio signals, the granularity of sensor data, the complexity of genomic data, the invaluable tapestry of medical his- tory, and the rich narrative woven by social media data. The orchestration and synthesis of this diverse spectrum of health-related datasets underscore the paramount importance and far-reaching impact of Health Informatics [4].

Big data has been more prevalent in healthcare in recent years, due to three primary factors: the vast amount of data available, rising healthcare expenses, and a need for individualized care. In healthcare, the big data processing refers to the development, collecting, analysis, and curation of clinical data that is too enormous or complicated to be inferred using traditional data processing methods [6].

Meanwhile, the widespread use of Electronic Health Records (EHR) has resulted in massive amounts of digital

Manuscript received July 16, 2024; revised January 22, 2025.

CHAIMAE TAOUSSI is a PhD student at University Sultan Moulay Slimane, Beni-Mellal 23000, Morocco. (corresponding author: email: chaimae.taoussi@usms.ac.ma, phone: +212660328381)

IMAD HAFIDI is a professor at University Sultan Moulay Slimane, Beni-Mellal 23000, Morocco. (email : i.hafidi@usms.ma)

ABDELMOUTALIB METRANE is a professor at the Cadi Ayyad University, Marrakech 40000, Morocco. (email: a.metrane@uca.ac.ma)

text, which contain highly multidimensional, heterogeneous, multimodal, irregular, chronological data such as test results, lab data, doctor's notes, drug prescriptions, demographic information, diagnostics, epidemiology, and behavioral data ranging from critical care to long-term planning. Which can aid in therapy selection, finding patient similarities, integrating genomic data for tailored treatment, predicting hospital stay length, and predicting readmission risk [5].

Science has made human life much more accessible through technology. In this twentieth era, the whole world is in the age of big data. Machine learning can predict future data or come up with a preferable decision from a large data set. In previous work, we have seen that machine learning algorithms can quickly predict disease. But many papers pointed out some specific diseases and after implementing machine learning algorithms, they showed that machine learning algorithms can predict diseases [7].

Our article explores four key areas related to ML and DL in medical informatics: the role of ML in enhancing medical diagnostics, the potential of DL to improve patient care, the various applications of ML and DL in healthcare, and the limitations of these techniques, such as data and model complexity. By examining these areas, we can better understand how advanced technologies are shaping the future of healthcare and address the challenges that lie ahead.

This article is organized to provide a comprehensive analysis of the advancements and challenges in health informatics driven by ML and DL. The following section, "Methodology," outlines the systematic approach undertaken, including the study selection and research questions that guided the analysis. Next, the "Machine Learning for Health Informatics" section delves into various machine learning algorithms, their applications, and their contributions to enhancing medical diagnostics and decision-making. Following this, the "Deep Learning for Health Informatics" section discusses deep learning models, including Convolutional Neural Networks and Recurrent Neural Networks, and their impact on predicting and managing complex health conditions. The article then explores "Applications of Machine Learning and Deep Learning in Healthcare," highlighting specific case studies in fields such as cancer diagnosis, mental health monitoring, and COVID-19 response. Finally, the article addresses the "Challenges and Limitations" associated with implementing these advanced technologies in healthcare, concluding with a discussion on future directions and potential solutions to overcome these challenges and limitations.

II. Methodology

This systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines to ensure transparency and rigor throughout the study selection process. The PRISMA flow diagram (Fig. 1) illustrates the selection process.

A. Identification of Studies

We conducted a detailed literature search, focusing on articles published between 2018 and 2024 to base our analysis on recent studies. A review protocol is designed to minimize researcher bias. This review protocol includes the formulation of the research questions, the defined search strategy, and the inclusion and exclusion criteria.

B. Research Questions (RQ)

To clarify the scope of the research and be more focused, RQs have been introduced, which focus on eight main points of view such as on the one hand the discussion of machine learning for health informatics as well as its different types and different algorithms used in this context, on the other hand the discussion of deep learning for health informatics as well as the different algorithms used in this context. In addition, the presentation of machine learning and deep learning applications for health informatics as well as the challenges and limits of these for health informatics. Table I represents the RQs considered in this study.

TABLE I
SPECIFIC RESEARCH QUESTIONS

ID	Research Question
RQ1	How does Machine Learning serve health informatics?
RQ2	What are the types of Machine Learning?
RQ3	What are the different Machine Learning algorithms used in health informatics?
RQ4	How does Deep Learning serve health informatics?
RQ5	What are the different Deep Learning algorithms used for health informatics?
RQ6	What are the applications of Machine Learning and Deep Learning for health informatics?
RQ7	What are the challenges of Machine Learning for health informatics?
RQ8	What are the challenges of Deep Learning for health informatics?

C. Search Strategy

The search strategy followed a semi-automated systematic review process, structured in four distinct stages. As a first step, meticulous attention was paid to the selection of digital libraries, given the essential role they play in identifying relevant primary studies that answer the research questions. Our automated search was conducted across five major academic electronic databases: Scopus, Web of Science (WOS), PubMed, IEEE Xplore, and Google Scholar as detailed in Table 2. These databases were selected for their comprehensive coverage of healthcare and technology-related publications, ensuring that a diverse range of studies could be accessed.

TABLE II
THE LIST OF DIGITAL LIBRARIES USED TO FIND RECENT ARTICLES ON OUR TOPIC

ID	Digital Library	Online Search Interface
1	Scopus	https://www.scopus.com/
2	Web of Science (WOS)	https://www.webofscience.com/
3	PubMed	https://pubmed.ncbi.nlm.nih.gov/
4	IEEE Xplore	https://ieeexplore.ieee.org/
5	Google Scholar	https://scholar.google.com/

Then, as part of this semi-automated approach, the selection of additional research sources was carried out rigorously, considering pre-established criteria. This step was guided by a semi-automated sorting and filtering process to our specific research questions, aimed at maximizing the relevance of the data included.

An exhaustive search strategy was followed to find all related articles. The search terms that were used in this search strategy were:

- "Technology" AND "Machine Learning" OR "Deep Learning" AND "Health Informatics"

- “Big Data” AND “Machine Learning” OR “Deep Learning” AND “Health Informatics”
- “Machine Learning” AND “Health Informatics”
- “Deep Learning” AND “Health Informatics”

D. Inclusion and Exclusion Criteria

Several articles were then produced as a result of the earlier processes. Thus, we have defined inclusion and exclusion criteria to sort out irrelevant articles:

1) Inclusion Criteria:

Studies were included if the article described new technologies and machine learning or deep learning, or if the article presented big data and machine learning or deep learning. Finally, the article is eligible if it proves the application of machine learning or deep learning in health informatics. To ensure relevance and quality, the following inclusion criteria were applied:

- Original research published between 2018 and 2024, focusing on the use of ML/DL algorithms in healthcare applications.
- Studies presenting quantitative results based on performance metrics such as precision, recall, F1-score, or Area Under the Curve (AUC).
- Research applying ML or DL models to real-world medical data (e.g., imaging, patient records, genomic data) for purposes such as disease diagnosis, prediction, or treatment planning.
- Articles published in peer-reviewed journals and available in English.

2) Exclusion Criteria:

Studies were excluded if they were not relevant to the research questions, were not recently published, lacked full-text access, or did not address the application of machine learning (ML) and deep learning (DL) in health informatics. The exclusion criteria were as follows:

- Articles that did not focus on the healthcare domain or that discussed ML/DL in a theoretical context without practical application.
- Studies lacking quantitative performance metrics or methodological clarity.
- Review articles, opinion pieces, conference abstracts, or papers without full-text access.
- Studies published before 2018, to ensure the inclusion of the most recent advancements in ML/DL technology.

E. Results

Our search identified 219 articles published between 2018 and 2024, when analyzing the importance of the research questions, 153 articles were excluded based on title, abstract, keyword, full text (Fig. 1.), leaving 66 eligible items. However, based on the exclusion criteria, 153 articles were excluded, of which 38 contain irrelevant research questions, 35 duplicate articles, 33 are not recent publications, 16 their full text is not available, 13 pay no attention to presenting machine learning and deep learning for health informatics, 10 methodological articles presenting a new method, 5 short or corrigendum communications, 3 validation studies. Thus, 66 articles were finally selected for this systematic review, of which some articles present the application of machine learning and deep learning for physical health informatics (radiation therapy-cancer-cardiovascular diseases, etc.) and the rest of the articles present the application of machine learning and deep learning for mental health informatics (depression-parkinson. . .) as shown in Fig. 1 In the sections below, we discuss in more depth each of the research questions for the analysis Datas.

F. Report review

1) Machine Learning for Health Informatics:

RQ1: How does Machine Learning serve health informatics?

Over time, human interaction was needed to guide machines in solving a specific problem. However, with the emergence of Machine Learning (ML), machines have improved their ability to solve problems autonomously by identifying patterns present in the data associated with a given problem [8]. Machine Learning (ML) is a field of Artificial Intelligence (AI) that involves developing computer systems capable of performing tasks that typically require human intelligence [9]. It is based on the use of statistical models and algorithms to allow computer systems to perform a task without explicit instruction, based on inference and patterns present in the data [10].

Machine Learning (ML) plays a crucial role in smart health, improving the quality of healthcare by providing accurate medical diagnoses, predicting diseases at an early stage, and analyzing illnesses [11]. On the one hand, machine learning enables computer systems to recognize patterns based on existing algorithms and datasets and to develop suitable solution concepts [6]. On the other hand, since there is a great need for computer techniques to analyze the large amounts of biological data generated every day in the world, machine learning algorithms are particularly useful because they analyze all the data, regulate their internal structure according to the data and generate hidden layers which give estimated models to study the results [12].

RQ2: What are the types of Machine Learning?

Machine learning is an ever-evolving field that is playing an increasingly important role in health informatics. There are various types of machine learning, each with its own set of characteristics and applications. The following are the primary types of machine learning:

- **Supervised learning:** is a type of machine learning that involves using a model to learn how to match input examples to a target variable [6]. It requires labeled data, which has inputs and desired outputs, to construct or map this correspondence [13]. The major objective of supervised learning is to accurately predict the future output and it is called regression when the target output is a continuous variable and classification when it is a group of discrete values [12].
- **Unsupervised learning:** is a form of machine learning that does not require labeled training data. Instead of labeled data, unlabeled data is used to group similar data and analyze the data further [12]. Common unsupervised learning techniques include similarity-based data clustering and dimension reduction to project high-dimensional data into lower-dimensional spaces [13].

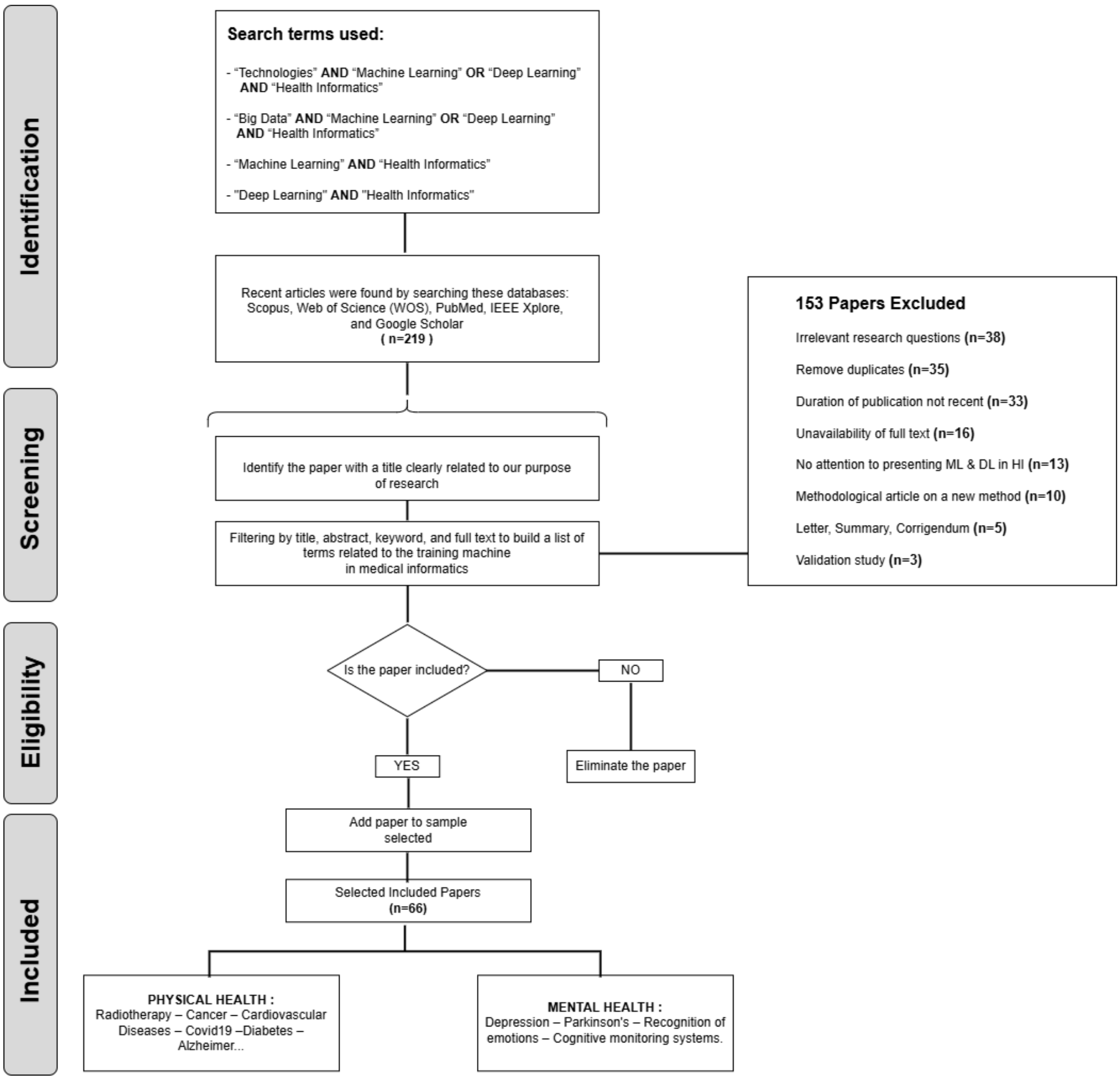

Fig. 1. The PRISMA procedures diagram used for the meta-review

- **Semi-supervised learning:** is a type of machine learning that uses a combination of labeled and unlabeled data to train a model. It is used when labeled data is available but in limited quantity, and unlabeled data is available in large quantity [12]. Semi-supervised learning techniques can be used for a variety of applications, especially in healthcare, where it is often difficult to obtain enough labeled data to train models [13].
- **Reinforcement learning:** is a form of machine learning where an agent learns to interact with an environment using rewards and punishments. It does not use labeled or unlabeled training data. Instead, the agent learns by taking actions in the environment and receiving rewards or punishments in response to those actions [12]. Reinforcement learning methods aim to improve online performance and include techniques such as policy learning and Q-learning [13].

RQ3: What are the different Machine Learning algorithms used for health informatics?

Machine learning is a powerful tool in the healthcare field. It can quickly process large amounts of clinical data, emulating the decision-making processes of clinicians [14]. Machine learning algorithms can solve complex problems by carefully examining evidence to make informed decisions. Recent advances in data analytics have supported a transformative model of healthcare based on machine learning, leading the way to better patient care. In this section, we will explore the different machine learning algorithms most commonly used in health informatics that aim to promote health care and predict medical pathologies accurately, as illustrated in Fig.2:

- **Dimensionality Reduction Algorithms (DRA):** are tools that simplify large datasets by identifying correlations and patterns within them without losing important information. There are different types of DRA algorithms, such as Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA), that reduce the number of dimensions of the input data to output a smaller, more manageable set, while eliminating inconsistencies, redundant data, and highly correlated features [15].
- **Transfer Learning (TL):** is an algorithm used in medical informatics, inspired by human competence in related tasks. It is based on the cognitive principle

of transferring knowledge from one medical task to another, to improve the efficiency and accuracy of algorithms for new medical challenges through the use of pre-trained models. The application of transfer learning refines model performance, helping to improve diagnostic accuracy, predictive modeling, and overall healthcare data analysis. It therefore appears to be a promising algorithm for optimizing model capabilities in the complex field of medical informatics [16].

- **The Random Forest Algorithm (RFs):** a non-parametric ensemble algorithm, has gained popularity in addressing regression and classification challenges, extending its application to medical informatics. Multiple decision trees are constructed by RFs using a randomly selected set of variables, acquired independently and with replacement from the original dataset. An important feature of RFs is their integrated functionality for assessing feature importance, facilitating the ranking of independent variables based on their impact on the outcome variable. This intrinsic capability enhances the value of RFs in data analysis, with promising results showcased in both empirical and theoretical studies, particularly within the realm of medical informatics. Despite their effectiveness in data analysis, the utilization of RFs in the context of patient satisfaction remains constrained [17].
- **Linear Regression (RL):**is a way to establish a relation between a dependent variable and one or more independent variables using a linear approach, it is particularly suited to situations where the independent variables are continuous. There are different methods to build a linear regression model, such as the ordinary least squares method and the gradient descent method. The former involves directly minimizing the sum of squared errors to find the coefficients, while the latter uses an iterative approach to minimize the sum of squared residuals [15].
- **Logistic Regression (LR):** is a statistical tool that predicts the probability of an individual belonging to a certain class, based on explanatory variables. This algorithm uses a sigmoid function to evaluate this probability, which takes a value between 0 and 1. There are two types of logistic regression: binary logistic regression for cases where there are two possible classes, and multinomial logistic regression for cases where there are more than two classes [15].
- **Support Vector Machine (SVM):** is a classification algorithm that relies on the definition of a hyperplane separating the data into two groups with a maximum distance between the points of the groups. The points to the left and right of this hyperplane are assigned to different groups. SVMs can be used for linear classifications, but can also operate with non-linear functions using kernels to classify complex data [15].
- **k-Nearest Neighbors (KNN):** is a supervised classification algorithm where objects are classified based on their similarity to the features of k previously labeled similar objects. The distances between the objects are calculated and the category of the object is determined by voting for the majority category of the k closest objects. The value of k is chosen based on the number of objects and is often an odd number. KNN is useful for classifying data with few training examples and is widely used in many applications [15].
- **Naive Bayes (NB):** is a classification algorithm based on Bayes' theorem. It assumes that all features in the data are independent of each other. The data are separated into a feature matrix and a response vector. Each row of the feature matrix represents an example of data in vector form [15].
- **Gradient Increase and Adaboost:** This method assigns weights to each algorithm based on their accuracy in classifying or estimating the data, then combines these decision boundaries to create a final model [15].
- **Artificial Neural Networks (ANN):** are machine learning models that imitate the decision- making process of the human brain. They have input, hidden and output layers, and work by assigning random weights to inputs to calculate weighted sums, which are then activated to deter- mine the output. When the output is incorrect, the weights are modified through back-propagation according to a cost function until the output is sufficiently accurate [15].

2) Deep Learning for Health Informatics:

RQ4: How does Deep Learning serve health informatics?

Deep learning is an approach to machine learning that focuses on creating meaningful representations from raw data. This method enables valuable information to be extracted from complex systems and is finding increasing application thanks to increasing computing power and data availability. With the advent of Big Data, deep learning has gained in popularity and is seen as an effective solution for processing and analyzing these vast quantities of data [18].

Deep learning in healthcare helps doctors accurately diagnose diseases and treat them more effectively, leading to better medical decisions [6]. DL has surpassed human levels of recognition capabilities and has been very successful in areas such as computer vision, natural language, and speech processing. This is because deep learning is a multilayered nonlinear neural network capable of learning from raw data, extracting features at each layer, and producing results via regression, classification, or ranking. The key factors supporting the success of deep learning are Big Data, Big Modeling and Big Computing [8].

Introduced in 2000, deep learning techniques that use computer-designed models to combine different layers of processing to understand and capture interpretations of data at multiple levels of abstraction have seen a recent surge in popularity due to technology upgrades [12]. They have improved significantly and expanded their application scope, including image recognition [19] speech recognition [20], machine translation [21], and integration into industrial systems like Google DeepMind's AlphaGo.

RQ5: What are the different Deep Learning algorithms used for health informatics?

Deep learning utilizes experience to automatically establish the complex parameters of a network. The utilization of multiple hidden layers within neural networks, which have long been acknowledged, has reached new levels of examination [9]. Deep learning can operate both with and without guidance. In machine learning, patterns are emphasized by simplifying data complexity. Deep learning can achieve this, and it has the ability to independently learn the

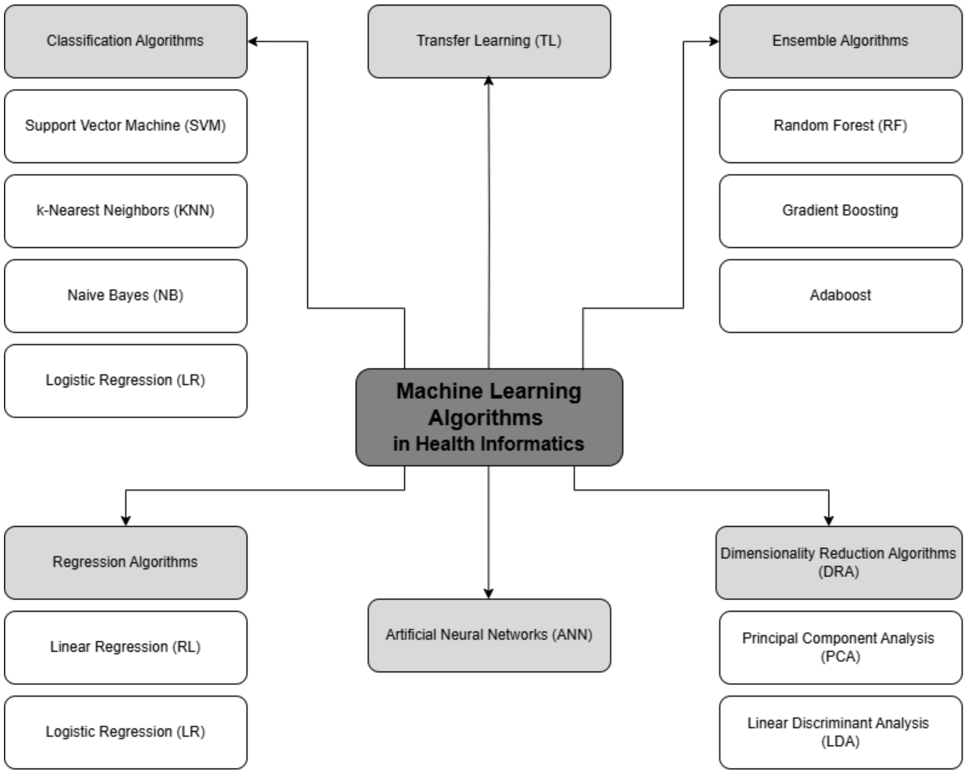

Fig. 2. Overview of Machine Learning Algorithms in Health Informatics

order representation of the input data and typically requires high volume [22].

In this section, we will examine the most widely used deep learning algorithms in health informatics, as illustrated in Fig.3, that aim to improve healthcare and accurately predict medical pathologies:

- **Automatic Representative Encoder:** It is a type of unsupervised representation learning that primarily serves the purpose of feature engineering. It is a preferred alternative to other traditional dimensionality reduction techniques, such as Principal Component Analysis (PCA) [23] and Singular Value Decomposition (SVD) [24], for nonlinear dimensionality reduction.
- **Convolutional Neural Network (CNN):** [25] CNN are specific algorithms that excel in image classification problems. They have a unique architecture where nodes in one layer are not connected to every node in the next layer. Instead, each layer contains one or multiple filters applied to the input to generate intermediate values.
- **Recurrent Neural Network (RNN):** With some types of data in the EHR such as clinical notes, the input data is not of the same length to be used with the basic ANN. RNN can process large amounts of text data, like clinical notes or web-based medical queries, to identify relevant keywords for standard clinical entities like ICD codes and CPT codes [5].
- **Restricted Boltzmann Machine (RBM):** is a form of neural network that has two layers: visible and hidden. It's called "restricted" because there is no communication between neurons within each layer. RBMs are simple to incorporate into deeper neural networks [5].
- **Neural Autoregressive Distribution Estimation (NADE):** NADE is a type of unsupervised neural network based on the autoregressive model and constructed using feedforward neural networks. It is used to estimate the distribution and density of data and has shown to be a reliable and manageable estimator [5].
- **Adversarial Neural Network (AN):** include both discriminators and generators, making them generative neural networks. The two neural networks can compete against each other in a minimax game when they are trained simultaneously [5].
- **An Attentional Model (AM):** is considered to be a differential neural architecture. Their operational implementation is based on addressing the content of input sequences. Due to their ubiquitous nature, most areas of computer vision and natural language processing accept such mechanisms. The attentional mechanism has even been observed in research on deep recommender systems [5].
- **Deep Reinforcement Learning (DRL):** DRL is entirely a trial-and-error operation. The basic components that fall under this paradigm are the agent, actions, environment, states and rewards. DRL is the only learning technique that has attained human-level performance in games, self- driving cars, and a variety of other disciplines. It is a combined result of deep neural networks and reinforcement learning. Agents can acquire knowledge from raw data and provide efficient representations without hand-crafted and heuristic domain features via networks of many advanced models [5].
- **Generative Pre-Trained Transformers (GPT):** constitute a class of algorithmic frameworks within Large Language Models, leveraging deep learning to generate natural language text based on input. GPT has a lot of

potential for use in medical applications because of its distinct architecture and use of unsupervised learning. Its potential for use in the healthcare industry highlights the need for cautious thought, regular updates from dependable sources, accurate data, and strong privacy protections especially considering the sensitive nature of the medical data involved [26].

3) RQ6: What are the applications of Machine Learning and Deep Learning for health informatics:

Cancer Diagnosis and Prediction:

A research project was conducted to create and assess two machine learning algorithms for ranking abstracts related to cancer proliferation in individuals with germline mutations or normal germline gene mutations. The initial model is based on a Support Vector Machine (SVM) that derives a linear decision rule from the n-gram representation of the title and abstract. On the other hand, the second model is a Convolutional Neural Network (CNN) that learns a set of parameters from the unprocessed title and abstract [27].

Alkhawaldeh et al. [28] presented a Multilayer Perceptron Model (MLP) for predicting breast cancer. The MLP model was fine-tuned and split into two categories: one to evaluate the reduction of extracted features and the other to assess the improvement of classification power. The dataset used in the study was obtained from WDBC, consisting of 569 instances and 32 attributes.

Several researchers have utilized machine learning techniques to enhance the prediction of lung cancer. Seenivasagam et al. [29] developed a machine learning approach for staging lung cancer based on mega data. Tabitha Peter et al. proposed a machine learning method for classifying diseases using lung cancer screening image data [30]. Wazir et al. [31] proposed a neural network model for predicting ovarian cancer risk stratification utilizing personal health data. Bansal et al. [32] conducted a recent study where they proposed a ResNet-based framework for the classification and 3D segmentation of lung cancer, achieving high accuracy in detecting early lung cancer using the LUNA16 dataset.

Additionally, a recent study by Shaik et al. [33] evaluated the performance of various machine learning algorithms on the HAM10000 skin cancer dataset. The study applied Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Logistic Regression (LR), and Naïve Bayes (NB) classifiers with dimensionality reduction techniques like Principal Component Analysis (PCA) and Gabor filters. The results showed that RF with PCA achieved an accuracy of 89%, while SVM with Gabor filters reached an accuracy of 84%. This study highlights the potential of machine learning to improve the diagnostic accuracy for skin cancer, showcasing the effectiveness of these algorithms in medical applications.

Moreover, a study by Hashim et al. [34] proposed a soft voting classifier for breast cancer prediction. This classifier integrates logistic regression, decision tree, and support vector machine models, achieving an accuracy of 99.3%, precision of 100%, recall of 98.46%, and an F1 score of 99.2% on the Wisconsin Diagnostic Breast Cancer dataset. Their methodology, which includes the use of SMOTE for dataset balancing, demonstrated superior performance compared to individual models, indicating the robust potential of ensemble learning in cancer diagnosis.

By integrating these advanced ML and DL techniques, healthcare practitioners can significantly improve the accuracy and efficiency of cancer diagnosis and prediction, leading to better patient outcomes and personalized treatment plans. The continuous advancements in these technologies hold great promise for revolutionizing cancer care and management.

Radiology and Radiotherapy:

Machine learning can prove to be beneficial for radiologists as it can help them to detect and diagnose diseases in the early stages by identifying even minor changes in the scan reports [35]. In radiation therapy planning quality assurance, machine learning has the capability to forecast dose-volume parameters, including the dose distribution index [36].

Deep learning is widely used in many radiological specialties. It has demonstrated effectiveness in converting free-form, unstructured conclusions into highly accurate, structured reports. Without noticeably lowering productivity, this technology has helped radiologists and clinicians communicate better. Additionally, it provides improved structured data that is beneficial for data mining and research [37].

Bellini et al. [38] used semantics-aware autoencoders to investigate the effects of knowledge graphs in recommendation systems. According to their research, combining knowledge graphs and deep learning models can enhance the variety and accuracy of recommendations, which is critical for individualized patient care and treatment regimens. This integration demonstrates how structured knowledge can be utilized by deep learning techniques to produce recommendations that are more accurate and contextually relevant.

Mental Health Monitoring:

Within the field of medical computer science, automatic learning has been applied to various predictive models to enhance patient care and streamline healthcare operations. For example, Rodrigues et al. [39] have developed a semi-supervised, whole-person learning approach to predict work-related stress using physiological indicators such as heart rate variability (VRC) and facial expressions. Their model achieved an accuracy of 86.7% and an F1 score of 87%, demonstrating the effectiveness of automatic learning techniques in the non-intrusive detection of stress in work environments. This approach highlights the potential of automatic learning in the creation of real-time health surveillance systems that could result in prompt and effective interventions.

COVID-19 Applications:

The ongoing novel coronavirus disease (COVID-19) pandemic has highlighted the critical importance of effective health informatics systems. Machine learning and deep learning have been pivotal in developing models to predict and manage COVID-19 spread. For instance, Zhang et al. [40] modified SEIR model accounts for local healthcare capacities and intervention techniques, providing valuable insights for decision-making during outbreaks. These models analyze large datasets to forecast trends and evaluate intervention effectiveness, helping optimize pandemic responses and prepare for future outbreaks.

In addition to these models, enhancing semantic prediction methods for COVID-19 using Big Data technologies has proven to be highly effective. ElDahshan et al. [41] developed an Onto-NoSQL framework that integrates Big Data with ontological models to predict COVID-19 prevalence and ana-

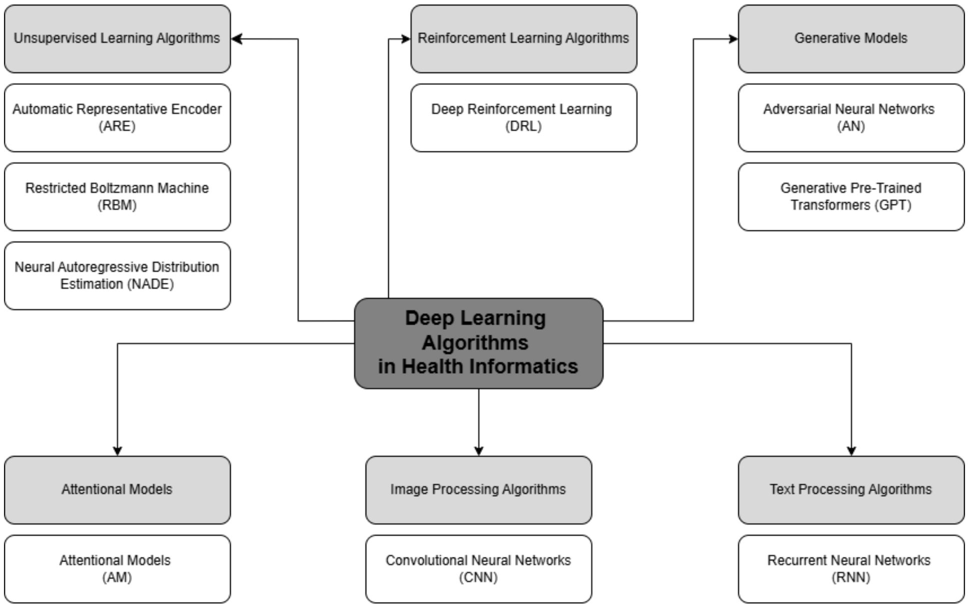

Fig. 3. Overview of Deep Learning Algorithms in Health Informatics

lyze the relationship between COVID-19 and weather factors. This approach demonstrated high accuracy and efficiency in handling large datasets, supporting decision-making and public health interventions.

Furthermore, in the context of the COVID-19 pandemic, Artificial Intelligence (AI) and Machine Learning (ML) models have been widely used to improve disease prediction, diagnosis, and response. Many COVID-19 symptom checkers and online chatbots have been developed, but studies have found significant variations in their results. A study was conducted to evaluate their performance and found that different symptom checkers had varying strengths in terms of sensitivity and specificity. Only two verifiers managed to achieve a balance between specificity and sensitivity [42].

Moreover, due to the rapid spread of COVID-19, it was crucial for governments and healthcare providers to quickly obtain accurate forecasts of the vulnerability of geographic regions and countries at risk of spreading the virus. Artificial intelligence was used to develop a three-step model that utilizes the XGBoost machine learning algorithm to estimate potential occurrences in unaffected countries and assess the likelihood of COVID-19 [43].

Additionally, Saha et al. [44] developed an automated detection system called EMCNet, which uses chest X-ray images to identify patients with COVID-19. The system relies on a specially designed convolutional neural network (CNN) to extract high-level features from patient X-ray images. This enables accurate and rapid detection of COVID-19 cases.

Heart Disease Prediction:

Machine Learning (ML) and Deep Learning (DL) models have revolutionized the medical field by providing new opportunities for the prediction and early detection of heart disease. These models use sophisticated algorithms to analyze large data sets and extract useful information, thereby improving the accuracy of diagnoses and predictions.

In this context, the study conducted by Li et al. [45] provides an accurate and efficient organization for heart disease diagnostics using ML methods. Their approach is based on a classification algorithm composed of SVM, LR, ANN, KNN, NB and DT. By using techniques such as the typical FS algorithm and least absolute removal electoral operators, they were able to maximize relevance and reduce feature redundancy, thereby improving prediction performance.

Similarly, Elhoseny et al. [46] proposed an Automatic Heart Disease Diagnostic (AHDD) scheme that integrates a Convolutional Neural Network (CNN) with an innovative MAFW mockup. The MAFW model includes several software agents working with algorithms such as GA, SVM and NB. This approach enables better exploitation of data from wearable sensors, such as glucose, body temperature, chest and heart rate sensors, providing more accurate diagnostic results.

In another study, Awotunde et al. [47] presented an Internet of Things (IoT WBN) based architecture for a wireless sensor network for heart disease. They used a machine learning model to collect and analyze data from different wearable sensors such as glucose, body temperature, chest and heart rate sensors to detect and predict heart abnormalities.

By exploiting the advantages of decision trees, Xie et al. [48] developed a heartbeat classification algorithm to diagnose arrhythmias, particularly Premature Ventricular Contraction (PVC). In addition, recent advances in cardiovascular disease risk prediction have revealed the effectiveness of the Auto-Prognosis system. This groundbreaking research demonstrated that using this system can significantly improve the accuracy of cardiovascular risk predictions [49].

At the same time, the use of gradient boosting has also shown promising results in the field of heart disease. This model consists of the sequential creation of several learners, where each learner corrects the errors made by the previous one. Extreme gradient amplification is a specific technique used in this setting, which has been able to predict irregular cycles in cardiac patients with remarkable accuracy, reaching a high accuracy rate of 92.1% [50].

Furthermore, the study by Nurmaini et al. [51] highlighted the use of a Deep Neural Network (DNN) structure combined with discrete wavelet transform (DWT) and principal component analysis (PCA) for the classification of cardiac arrhythmia. This method achieved 99.76% accuracy and demonstrated the robustness and generalization performance across multiple datasets, proving its efficacy in improving the generalizability of cardiac arrhythmia detection.

Other Medical Applications:

The advancement of technologies such as Machine Learning (ML), Deep Learning (DL) and the Internet of Things (IoT) has opened new perspectives in the field of health and well-being. These advances make it possible to exploit available data, such as speech, facial expressions, movements, electrical brain signals, and more, to improve diagnostics, patient care, and disease management.

In the field of speech analysis, local hidden patterns have been used to develop systems capable of classifying emotions. These models use convolutional layers that sample entity maps, providing accurate results in emotion recognition [52]. Similarly, the use of bidirectional recurrent neural networks with layers of attention allows the development of chatbots capable of understanding and generating more appropriate conversations [53].

The integration of IoT and NLP enables real-time assessment of various parameters, such as speech, facial expressions, and movements, to provide emergency medical assistance, assess patient status, and monitor their well-being [54]. NLP is also used in the field of mental health, where it is applied to data from social media and IoT devices for various applications [55].

In the field of elderly health, systems based on IoT and deep learning are being developed for emotion recognition. These systems can assess emotions such as calm, joy, sadness, anger, fear, surprise and disgust, which can help improve the quality of life of nursing home residents [56].

Finally, the use of raw Electroencephalogram (EEG) in combination with Convolutional Neural Net- work (CNN) models offers a novel approach to assess depression. This method analyzes the electrical signals of the brain and provides an objective assessment of the depressive state of individuals [57].

The modeling of influenza-based epidemics is an important research area in the field of public health. In this article, several Machine Learning (ML) methods were studied to build predictive models. Tapak et al. [58] examined the performance of Artificial Neural Networks (ANN), Support Vector Machines (SVM), and random series forestry. They used metrics such as Root Mean Square Error (RMSE), Intraclass Correlation Coefficient (ICC), and Mean Absolute Errors (MAE) to assess these models.

Another study by Ginantra et al. [59] proposed an SVM-based classification model to identify people with Influenza-Like Illnesses (ILI), which are acute respiratory infections. Their SVM model was found to perform better than the other classifiers studied.

Additionally, Al Hossain et al. [60] explored the application of a random forest model to predict the number of people infected with influenza in public places. Their random forest model outperformed other models with 95% accuracy. This can be attributed to the model's ability to combine outputs from multiple decision trees.

The application of Machine Learning (ML) techniques in the medical field has become a promising approach for modeling and predicting clinical outcomes. These advances provide valuable insights to support medical decisions, improve patient care and optimize treatment protocols. In this article, we explore two separate studies that used ML methods to develop predictive models in specific medical settings.

In the first study, Lucas et al. [61] focused on predicting recovery from hemorrhagic shock after resuscitation, using rats as test subjects. They applied logistic regression using the Python scikit-learn library to model this prediction system. The performance of this model was evaluated using measures such as mean cross-validation precision, Youden's J statistic, and Cohen's kappa coefficient.

The second study, conducted by Chen et al. [62], proposed an intelligent Internet of Things (IoT) based application for cerebral hemorrhage diagnostics in humans. They used ML algorithms such as Support Vector Machines (SVM) and Forward Propagation Neural Networks (FFNN) for the classification of intracranial datasets based on CT scan images.

The application of Machine Learning (ML) and Deep Learning (DL) in healthcare is rapidly expanding, offering new perspectives for the diagnosis, prediction and treatment of various medical conditions. Among these applications, we look at several studies that use ML and DL to predict bleeding emergencies, diagnose Parkinson's disease and assess diabetes risk.

In the field of medical emergencies, ML models have been developed to predict bleeding and allow early intervention. For instance, Kumar et al. [63] assessed the critical importance of diagnosing diabetes. They developed an Android app that allows customers to input their data, which is then analyzed in real time using pre-trained ML models. Logistic Regression (LR) was used to perform prediction calculations, thereby improving the detection of diabetes disease. The diagnostic results are then displayed to patients in the Android app.

Additionally, research has been conducted to diagnose Parkinson's disease by combining the Internet of Things (IoT) and Linear Discriminant Analysis (LDA) [64]. Lamba et al. [65] proposed a hybrid diagnostic scheme for Parkinson's disease based on the vocal signal. They used a combination of classification algorithms and feature selection approaches to develop an integrated and optimal method.

Furthermore, Salman et al. [66] performed an in-depth review of ML models in the field of electronic Emergency triage (E-triage) for organizing patients for prompt medical services in telemedicine applications. Their study highlighted the effectiveness of ML models in remote telemedicine systems.

These examples underline the diversity of applications of ML and DL in the field of health. Significant advances are being made to improve diagnoses, predict emergency situations and optimize treatments, allowing for more accurate, personalized and effective medical care. However, it is essential to continue validating and refining these models to ensure their reliability and safety in real clinical settings.

4) Challenges and limitations of Machine Learning and Deep Learning for Health Informatics:

RQ7: What are the challenges of Machine Learning for health informatics?

The field of machine learning continues to evolve and holds great potential for enhancing healthcare through data analysis and improved outcome prediction. However, there are various challenges and limitations that must be overcome in order to effectively integrate machine learning into medical informatics. These include issues with acquiring sufficient and reliable data for model training, and concerns over preserving patient confidentiality. To fully leverage the advantages and minimize any risks associated with the use of machine learning in healthcare, it is essential to be aware of the potential challenges and limitations. Some of the drawbacks and challenges that are commonly encountered in the Machine Learning process in health informatics can be listed as follows:

- If the data used to train a machine learning model does not align with the data it will encounter during deployment, its performance may not meet expectations. As machine learning algorithms take the path of least resistance during training, they may learn features from the data that are incorrectly correlated with the target outputs instead of the correct features. This can limit the model's ability to effectively generalize the learned information [65].
- In the process of machine learning, a significant amount of data is required for the training and learning phase. Therefore, the data used must be of high quality and free from bias. As part of the machine learning process with the assistance of software development services, there may be periods of waiting, during which new data can be generated and used for further processing [8].
- Time and resources are important considerations in the machine learning process, as the algorithms manage all functions for data handling and the use of specific data for error correction. This requires sufficient time and reliable resources to ensure the system operates effectively [8].
- Interpretation of the output generated by the algorithms is essential to ensure accuracy. The output must be thoroughly checked for any errors, and corrective measures should be taken accordingly. Choosing the right algorithm is crucial to achieve the desired level of accuracy in the output [8].
- The machine learning process involves using large amounts of data and testing various algorithms, which increases the likelihood of encountering errors. During the training of the dataset, many algorithms are used simultaneously, and any errors in the algorithms can lead to irrelevant results. These errors are a common problem that can occur repeatedly. However, identifying the source of the problem and fixing it is often challenging and time-consuming [8].

RQ8: What are the challenges of Deep Learning for health informatics?

Deep learning, a technique that utilizes neural networks to address complex problems, has seen tremendous success in areas like image recognition, machine translation, and speech recognition. However, its implementation in the realm of computational medicine faces various difficulties and limitations that must be considered. To evaluate the viability and efficiency of deep learning in medical informatics, it is crucial to examine the drawbacks and challenges associated with this field. Some of the drawbacks and challenges that are commonly encountered in the Deep Learning process in health informatics can be listed as follows:

- In the field of healthcare, it can be difficult to establish causality because the learning process is typically based solely on observational data. Asking causal questions based on observational data requires the construction of causal models, which can be challenging. Deep Learning (DL) models are essentially black boxes that lack a fundamental underlying theory, and they operate by exploiting patterns and correlations without considering any causal links [13]. Although there has been some work on representing advanced features in DL models, such as using weighting filters in a CNN, these models are still often difficult to understand. Consequently, many researchers use DL approaches as a black box without needing to explain the excellent results that they can produce [2].
- Deep learning has limitations when it comes to providing explainable predictions as it often relies on a black-box approach. The hidden weights and activations of deep learning models are generally uninterpretable, which makes their predictions difficult to explain. However, the development of neural attention models has addressed this issue and led to the creation of more interpretable deep neural models. Although it is still challenging for neural models to interpret a single neuron, some interpretability has been achieved in current state-of-the-art models, which allows for more explainable recommendations. This problem will be further discussed in the open problems section [8].
- Learning approaches require a large amount of training data to produce accurate predictions. While many medical organizations have made efforts to digitize paper medical records, a dataset specific to a particular disease is still necessary [18]. To ensure that a deep neural network does not require an excessive amount of data, it is crucial to provide sufficient data to support its rich parameterization. While labeled data can be scarce in certain domains, such as language or vision, it is often easier to collect a significant amount of data in the context of recommender systems research. Industry and academic data sets can be found at the scale of millions or billions. However, deep learning algorithms typically require a large amount of training data to achieve high accuracy, which may not be readily available [8]. Insufficient training data can result in overfitting, where the test error is high despite a low training error. Furthermore, deep learning may not be able to provide solutions to certain important questions. High-level visualizations obtained using deep learning may be difficult to interpret, and there may be no provision for making changes in case of classification problems. Additionally, deep learning may not be suitable for all types of diseases, particularly rare diseases [12].
- The arrangement of different hyperparameters that determine how a Deep Neural Network (DNN) operates - such as the size and number of filters in a Convolutional Neural Network (CNN), or its depth - is a trial-and-error process that usually needs to be validated. It can be difficult to find the best combination of hyperparameters and data preprocessing techniques, and this

can make training the model even more challenging [4]. The process requires extensive human expertise and resources, which makes it difficult to create an effective classification model. Setting hyperparameters is a general challenge in machine learning, and deep learning can sometimes introduce even more hyperparameters. For instance, a recent study introduced a new hyperparameter by expanding on the traditional metric learning algorithm [8].

- In many applications, the raw data cannot be used directly as input for deep learning models. As a result, preprocessing, normalization, or modification of the input is often necessary prior to training. It's also important to note that for most applications, the raw data cannot be used directly as input for neural networks [4]. Therefore, preprocessing and normalization are frequently required before training. The process of configuring hyperparameters that affect the structure of a DNN, such as the size and number of filters in a CNN, is always a new exploration process that requires careful validation [2].

III. General Discussion

The integration of Machine Learning (ML) and Deep Learning (DL) into health informatics has transformed the way healthcare systems operate, offering a more data-driven approach to diagnosis, disease prediction, and personalized treatment. This section provides a synthesis of the key achievements in applying these technologies, the challenges faced in their implementation, and potential future directions to improve their efficacy and adoption in clinical settings.

A. Overview of the Impact of ML and DL in Health Informatics:

Machine learning and deep learning have become essential tools in health informatics, revolutionizing the analysis and utilization of healthcare data. These technologies have demonstrated significant potential in enhancing diagnostic accuracy, predicting diseases, and personalizing treatment plans. Through the integration of large clinical databases and advanced algorithms, such as Support Vector Machines (SVM), Convolutional Neural Networks (CNN), and Multi-Layer Perceptrons (MLP), healthcare providers can now make more informed and timely decisions, marking a significant shift in healthcare delivery toward data-driven insights.

B. Key Achievements in the Application of ML and DL:

Several studies have highlighted the success of ML and DL algorithms in various medical applications. In oncology, hybrid models like SVM and CNN have achieved high accuracy in cancer diagnosis. CNN models, in particular, have shown superior performance in image-based cancer detection, including lung and breast cancer. Additionally, optimized models such as MLP have been effective in predicting breast cancer outcomes. These examples demonstrate the potential of ML and DL to significantly improve diagnostic capabilities in healthcare.

In medical imaging, deep learning models, particularly CNNs and ResNets, excel in cancer classification and 3D segmentation, with ResNet achieving high accuracy in cancer segmentation tasks. As summarized in Fig.4, CNN achieved 92% accuracy in lung cancer detection, MLP 89% in breast cancer prediction, and ResNet 91% in cancer 3D segmentation. Fig.4 provides a visual comparison of these accuracy rates, underscoring the superior performance of CNN in lung cancer detection and ResNet in 3D segmentation.

The integration of semantic autoencoders and knowledge graphs has significantly enhanced the accuracy and diversity of recommendations in health systems. Machine learning techniques, such as semi-supervised learning and ensemble algorithms, have demonstrated effectiveness in real-time monitoring of mental and physical health and in the development of non-invasive diagnostic tools. Furthermore, hybrid models that combine AI capabilities with clinical expertise present a promising approach to improving the accuracy and reliability of medical predictions.

As detailed in Table 3, various studies summarize the algorithms used and specific applications in the domain of health informatics. Both machine learning and deep learning algorithms demonstrate high performance across diverse medical applications, highlighting their potential to improve diagnosis and treatment in real-world clinical settings.

C. Challenges in the Adoption of ML and DL in Healthcare:

Despite these remarkable achievements, several challenges hinder the widespread adoption of ML and DL in healthcare:

- **Data Availability and Quality:** The success of these models depends heavily on access to high-quality, annotated datasets. However, limited availability of such data, especially in underrepresented populations, poses a challenge for the scalability and generalizability of ML models.
- **Model Generalization:** Models trained on specific datasets often struggle to generalize across different patient groups or clinical environments, limiting their application in real-world scenarios.
- **Regulatory and Ethical Issues:** The integration of AI models in healthcare requires navigating complex ethical and regulatory challenges, particularly in relation to patient privacy, data security, and the explainability of AI-driven decisions.

D. Future Directions:

To address these challenges, future research should focus on:

- **Data Standardization and Sharing:** Encouraging the secure sharing of anonymized medical data across institutions can help build larger, more diverse datasets, which are essential for training robust models.
- **Improving Interpretability:** Increasing the transparency of ML and DL models is critical for ensuring that clinicians can trust and effectively use AI in decision-making.
- **Hybrid Models:** Integrating clinical expertise with AI models can enhance their accuracy and reliability. Hybrid approaches, such as combining CNNs with traditional clinical assessments, have the potential to deliver more comprehensive and accurate predictions in medical practice.

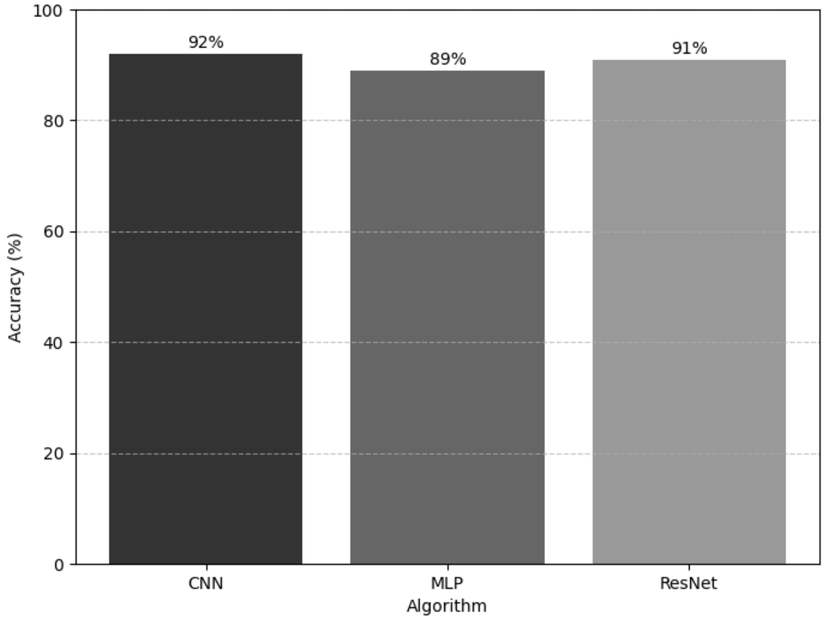

Fig. 4. Algorithm Accuracy Comparison

IV. Conclusion

In this article, we have explored the key applications of machine learning and deep learning in the field of health informatics. These techniques have the potential to revolutionize medicine by enabling more accurate diagnoses, more reliable disease predictions, and improved medical care. The results of predictive models are particularly relevant in healthcare, where fast, accurate decisions are crucial for patients.

By providing reliable information on potential outcomes, these models can help clinicians develop personalized treatment plans and improve patient outcomes. However, it is important to stress that rigorous validation of these models is essential before their widespread use in clinical practice. Further studies are needed to assess the performance and generalizability of these models across large populations and diverse clinical settings.

While the prospects offered by machine learning and deep learning in healthcare are promising, it is important to recognize the challenges and limitations associated with their use. Large datasets, model complexity, and ethical issues relating to data confidentiality are all important considerations.

In conclusion, the application of ML and DL techniques in the medical field opens up exciting new prospects for predicting clinical outcomes and improving patient care. However, it is essential to continue studying, developing, and validating these models to ensure their reliability, safety, and clinical relevance. The future of health informatics lies in continued collaboration between medical experts, informatics researchers, and ML and DL specialists to harness the full potential of these technologies and improve people's health and well-being.

References

[1] El Morr, C., "Introduction to health informatics: a Canadian perspective," Canadian Scholars, 2023.

[2] Tyagi, A. K., and Nair, M. M., "Deep learning for clinical and health informatics," Computational Analysis and Deep Learning for Medical Care: Principles, Methods, and Applications, pp. 107–129, 2021.

[3] Hasan, N., and Bao, Y., "Understanding current states of machine learning approaches in medical informatics: a systematic literature review," Health and Technology, vol. 11, no. 3, pp. 471–482, 2021.

[4] Aloyuni, S. A., "A Systematic Review on Machine Learning and Deep Learning Based Predictive Models for Health Informatics," Journal of Pharmaceutical Research International, vol. 33, no. 47B, pp. 183–194, 2021.

[5] Harerimana, G., et al., "Deep learning for electronic health records analytics," IEEE Access, vol. 7, pp. 101245–101259, 2019.

[6] Supriya, M., and Deepa, A. J., "Machine learning approach on healthcare big data: a review," Big Data and Information Analytics, vol. 5, no. 1, pp. 58–75, 2020.

[7] Ferdous, M., Debnath, J., and Chakraborty, N. R., "Machine learning algorithms in healthcare: A literature survey," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), IEEE, 2020.

[8] Pramod, A., Naicker, H. S., and Tyagi, A. K., "Machine learning and deep learning: Open issues and future research directions for the next 10 years," Computational Analysis and Deep Learning for Medical Care: Principles, Methods, and Applications, pp. 463–490, 2021.

[9] Siddique, S., and Chow, J. C. L., "Machine learning in healthcare communication," Encyclopedia, vol. 1, no. 1, pp. 220–239, 2021.

[10] Dalal, K. R., "Analysing the implementation of machine learning in healthcare," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), IEEE, 2020.

[11] Rayan, Z., Alfonse, M., and Salem, A. B. M., "Machine learning approaches in smart health," Procedia Computer Science, vol. 154, pp. 361–368, 2019.

[12] Mittal, S., and Hasija, Y., "Applications of deep learning in healthcare and biomedicine," Deep Learning Techniques for Biomedical and Health Informatics, pp. 57–77, 2020.

[13] Qayyum, A., et al., "Secure and robust machine learning for healthcare: A survey," IEEE Reviews in Biomedical Engineering, vol. 14, pp. 156–180, 2020.

TABLE III
COMPARATIVE SUMMARY OF MACHINE LEARNING AND DEEP LEARNING ALGORITHMS IN HEALTH INFORMATICS

Category	Article	Algorithm	Application
Machine Learning	Ferdous et al. [7]	SVM, CNN	Cancer Diagnosis
	Alkhawaldeh et al. [28]	MLP	Breast Cancer Prediction
	Hart et al. [31]	Neural Networks	Ovarian Cancer Risk Stratification
	Bansal et al. [32]	ResNet	Cancer 3D Classification and Segmentation
	Ng et al. [36]	Linear Regression (LR), Tree Regression (TR), SVM Gaussian Process Regression	Radiotherapy Planning
	Rodrigues et al. [39]	SVM, Random Forests, Gaussian naive Bayes	Work-Related Stress Prediction
	Lucas et al. [61]	Logistic Regression models (LRMs)	Hemorrhagic Shock Recovery Prediction
	Chen et al. [62]	SVM, Feedforward Neural Network (FNN)	Brain Hemorrhage Diagnosis
	Kumar et al. [63]	Logistic Regression (LR)	Diabetes Detection
	Ginantra et al. [59]	SVM	Acute Respiratory Infections (ISPA) Prediction
	Al Hossain et al. [60]	Random Forest (RF)	Influenza Infection Prediction
	Tapak et al. [58]	ANN, SVM, RF	Influenza Epidemic Modeling
Deep Learning	Bellini et al. [38]	Semantic Autoencoders	Recommender Systems (RSs)
	Spandorfer et al. [37]	Self-supervised convolutional neural network	Medical Report Structuring
	Saha et al. [44]	CNN	COVID-19 Detection
	Li et al. [45]	SVM, Logistic Regression (LR), ANN, KNN, NB, Decision Tree (DT)	Heart Disease Diagnosis
	Elhoseny et al. [46]	CNN, Genetic Algorithm (GA), SVM, NB	Automatic Heart Disease Diagnosis
	Awotunde et al. [47]	Extra Tree Algorithm	Heart Disease Monitoring
	Xie et al. [48]	Decision Tree	Heartbeat Classification
	Tariq et al. [56]	CNN	Speech Emotion Detection (SED)
	Ke et al. [57]	CNN	Depression Assessment via EEG
	Seenivasagam et al. [29]	Machine Learning	Lung Cancer Staging
	Tabitha Peter et al. [30]	Machine Learning	Lung Cancer Classification
	Wazir et al. [31]	Neural Network	Ovarian Cancer Risk Stratification
	Shaik et al. [33]	RF, SVM, KNN, LR, NB	Skin Cancer Diagnosis
	Hashim et al. [34]	Soft Voting Classifier	Breast Cancer Prediction
	ElDahshan et al. [41]	Onto-NoSQL Framework	COVID-19 Prevalence Prediction
	Salman et al. [66]	ML Models	Electronic Emergency Triage (E-triage)
	Lamba et al. [65]	Hybrid Diagnostic Scheme	Parkinson's Disease Diagnosis
	Nurmaini et al. [51]	DNN with DWT and PCA	Cardiac Arrhythmia Classification
	Zhang et al. [40]	Modified SEIR Model	COVID-19 Spread Prediction

[14] Mishra, R., Chaudhary, K., and Mishra, I., ”AI in health science: A perspective,” Current Pharmaceutical Biotechnology, vol. 24, no. 9, pp. 1149–1163, 2023.

[15] Bharadwaj, H. K., et al., ”A review on the role of machine learning in enabling IoT based healthcare applications,” IEEE Access, vol. 9, pp. 38859–38890, 2021.

[16] Kim, H. E., et al., ”Transfer learning for medical image classification: a literature review,” BMC Medical Imaging, vol. 22, no. 1, pp. 69, 2022.

[17] Simsekler, M. C. E., et al., ”Exploring drivers of patient satisfaction using a random forest algorithm,” BMC Medical Informatics and Decision Making, vol. 21, no. 1, pp. 157, 2021.

[18] Navamani, T. M., ”Efficient deep learning approaches for health informatics,” Deep Learning and Parallel Computing Environment for Bioengineering Systems, Academic Press, pp. 123–137, 2019.

[19] Wu, R., et al., ”Deep image: Scaling up image recognition,” arXiv preprint arXiv:1501.02876, 2015.

[20] Noda, K., et al., ”Audio-visual speech recognition using deep learning,” Applied Intelligence, vol. 42, pp. 722–737, 2015.

[21] Tyagi, A. K., ”Machine learning with big data,” Proceedings of International Conference on Sustainable Computing in Science, Technology and Management (SUSCOM), Amity University Rajasthan, Jaipur-India, 2019.

[22] Siddique, S., and Chow, J. C. L., ”Artificial intelligence in radiotherapy,” Reports of Practical Oncology and Radiotherapy, vol. 25, no. 4, pp. 656–666, 2020.

[23] Lovric, M., ”International encyclopedia of statistical science,” Springer, 2011.

[24] Wilkinson, J. H., Bauer, F. L., and Reinsch, C., ”Linear algebra,” vol. 2, Springer, 2013.

[25] Krizhevsky, A., Sutskever, I., and Hinton, G. E., ”ImageNet classification with deep convolutional neural networks,” Communications of the ACM, vol. 60, no. 6, pp. 84–90, 2017.

[26] Javaid, M., Haleem, A., and Singh, R. P., ”ChatGPT for healthcare services: An emerging stage for an innovative perspective,” BenchCouncil Transactions on Benchmarks, Standards and Evaluations, vol. 3, no. 1, pp. 100105, 2023.

[27] Bao, Y., et al., ”Using machine learning and natural language processing to review and classify the medical literature on cancer susceptibility genes,” JCO Clinical Cancer Informatics, vol. 1, pp. 1–9, 2019.

[28] Al-Shargabi, B., Alshami, F., and Alkhawaldeh, R., ”Enhancing multilayer perception for breast cancer prediction,” International Journal of Advanced Science and Technology, 2019.

[29] Sujitha, R., and Seenivasagam, V., ”Retraction Note to: Classification of lung cancer stages with machine learning over big data healthcare framework,” pp. 541–541, 2023.

[30] Delzell, D. A. P., et al., ”Machine learning and feature selection methods for disease classification with application to lung cancer screening image data,” Frontiers in Oncology, vol. 9, pp. 1393, 2019.

[31] Hart, G. R., et al., ”Stratifying ovarian cancer risk using personal health data,” Frontiers in Big Data, vol. 2, pp. 24, 2019.

[32] Bansal, G., et al., ”Deep3DSCan: Deep residual network and morphological descriptor based framework for lung cancer classification and 3D segmentation,” IET Image Processing, vol. 14, no. 7, pp. 1240–1247, 2020.

[33] Shaik, A. R., and Kumar, P. R., ”Performance Evaluation of Machine Learning Algorithms on Skin Cancer Data Set Using Principal Component Analysis and Gabor Filters,” IAENG International Journal of Computer Science, vol. 51, no. 7, pp. 831-841, 2024.

[34] Hashim, M. S., and Yassin, A. A., ”Breast Cancer Prediction Using

Soft Voting Classifier Based on Machine Learning Models," IAENG International Journal of Computer Science, vol. 50, no. 2, pp. 705-714, 2023.

[35] Watson, D. S., et al., "Clinical applications of machine learning algorithms: beyond the black box," BMJ, vol. 364, 2019.

[36] Ng, F., Jiang, R., and Chow, J. C. L., "Predicting radiation treatment planning evaluation parameter using artificial intelligence and machine learning," IOP SciNotes, vol. 1, no. 1, pp. 014003, 2020.

[37] Spandorfer, A., et al., "Deep learning to convert unstructured CT pulmonary angiography reports into structured reports," European Radiology Experimental, vol. 3, pp. 1–8, 2019.

[38] Bellini, V., et al., "A qualitative analysis of knowledge graphs in recommendation scenarios through semantics-aware autoencoders," Journal of Intelligent Information Systems, 2024.

[39] Rodrigues, F., and Correia, H., "Semi-supervised and ensemble learning to predict work-related stress," Journal of Intelligent Information Systems, vol. 62, no. 1, pp. 77–90, 2024.

[40] Zhang, C., "Supporting decision-making for COVID-19 outbreaks with the modified SEIR model," IAENG International Journal of Computer Science, vol. 48, no. 1, pp. 86–95, 2021.

[41] ElDahshan, K., Elsayed, E. K., and Mancy, H., "Enhancement semantic prediction big data method for COVID-19: Onto-NoSQL," IAENG International Journal of Computer Science, vol. 47, no. 4, pp. 613–622, 2020.

[42] Munsch, N., et al., "Diagnostic accuracy of web-based COVID-19 symptom checkers: comparison study," Journal of Medical Internet Research, vol. 22, no. 10, pp. e21299, 2020.

[43] Mehta, M., et al., "Early stage machine learning–based prediction of US county vulnerability to the COVID-19 pandemic: machine learning approach," JMIR Public Health and Surveillance, vol. 6, no. 3, pp. e19446, 2020.

[44] Saha, P., Sadi, M. S., and Islam, M. M., "EMCNet: Automated COVID-19 diagnosis from X-ray images using convolutional neural network and ensemble of machine learning classifiers," Informatics in Medicine Unlocked, vol. 22, pp. 100505, 2021.

[45] Li, J. P., et al., "Heart disease identification method using machine learning classification in e-healthcare," IEEE Access, vol. 8, pp. 107562–107582, 2020.

[46] Elhoseny, M., et al., "A new multi-agent feature wrapper machine learning approach for heart disease diagnosis," 2021.

[47] Awotunde, J. B., et al., "Disease diagnosis system for IoT-based wearable body sensors with machine learning algorithm," Hybrid Artificial Intelligence and IoT in Healthcare, pp. 201–222, 2021.

[48] Xie, T., et al., "Research on heartbeat classification algorithm based on CART decision tree," 2019 8th International Symposium on Next Generation Electronics (ISNE), IEEE, 2019.

[49] Alaa, A. M., et al., "Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK Biobank participants," PLoS One, vol. 14, no. 5, pp. e0213653, 2019.

[50] Shi, H., et al., "A hierarchical method based on weighted extreme gradient boosting in ECG heartbeat classification," Computer Methods and Programs in Biomedicine, vol. 171, pp. 1–10, 2019.

[51] Nurmaini, S., Tutuko, B., Rachmatullah, M. N., Darmawahyuni, A., and Masdung, F., "Machine learning techniques with low-dimensional feature extraction for improving the generalizability of cardiac arrhythmia," IAENG International Journal of Computer Science, vol. 48, no. 2, pp. 369–378, 2021.

[52] Mustaqeem, and Kwon, S., "A CNN-assisted enhanced audio signal processing for speech emotion recognition," Sensors, vol. 20, no. 1, pp. 183, 2019.

[53] Khandia, R., et al., "Emergence of SARS-CoV-2 Omicron (B.1.1.529) variant, salient features, high global health concerns and strategies to counter it amid ongoing COVID-19 pandemic," Environmental Research, vol. 209, pp. 112816, 2022.

[54] Amin, S. U., et al., "Cognitive smart healthcare for pathology detection and monitoring," IEEE Access, vol. 7, pp. 10745–10753, 2019.

[55] Dalal, S., Jain, S., and Dave, M., "A systematic review of smart mental healthcare," Proceedings of the 5th International Conference on Cyber Security Privacy in Communication Networks (ICCS), 2019.

[56] Tariq, Z., Shah, S. K., and Lee, Y., "Speech emotion detection using IoT-based deep learning for healthcare," 2019 IEEE International Conference on Big Data (Big Data), IEEE, 2019.

[57] Ke, H., et al., "Cloud-aided online EEG classification system for brain healthcare: A case study of depression evaluation with a lightweight CNN," Software: Practice and Experience, vol. 50, no. 5, pp. 596–610, 2020.

[58] Tapak, L., et al., "Comparative evaluation of time series models for predicting influenza outbreaks: Application of influenza-like illness data from sentinel sites of healthcare centers in Iran," BMC Research Notes, vol. 12, pp. 1–6, 2019.

[59] Ginantra, N. L. W. S. R., Indradewi, I. G. A. D., and Hartono, E., "Machine learning approach for acute respiratory infections (ISPA) prediction: Case study Indonesia," Journal of Physics: Conference Series, vol. 1469, no. 1, IOP Publishing, 2020.

[60] Al Hossain, F., et al., "FluSense: a contactless syndromic surveillance platform for influenza-like illness in hospital waiting areas," Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 4, no. 1, pp. 1–28, 2020.

[61] Lucas, A., Williams, A. T., and Cabrales, P., "Prediction of recovery from severe hemorrhagic shock using logistic regression," IEEE Journal of Translational Engineering in Health and Medicine, vol. 7, pp. 1–9, 2019.

[62] Chen, H., et al., "A smart machine learning model for the detection of brain hemorrhage diagnosis based on Internet of Things in smart cities," Complexity, vol. 2020, no. 1, pp. 3047869, 2020.

[63] Kumar, N., et al., "Efficient automated disease diagnosis using machine learning models," Journal of Healthcare Engineering, vol. 2021, no. 1, pp. 9983652, 2021.

[64] Lee, S. J., et al., "A dimension-reduction based multilayer perception method for supporting the medical decision making," Pattern Recognition Letters, vol. 131, pp. 15–22, 2020.

[65] Lamba, R., et al., "A hybrid system for Parkinson's disease diagnosis using machine learning techniques," International Journal of Speech Technology, 2022.

[66] Salman, O. H., et al., "A review on utilizing machine learning technology in the fields of electronic emergency triage and patient priority systems in telemedicine: Coherent taxonomy, motivations, open research challenges and recommendations for intelligent future work," Computer Methods and Programs in Biomedicine, vol. 209, pp. 106357, 2021.