A Review of Machine Learning Techniques in the Medical Domain

Enas M.F. El Houby
Department of Systems and Information, National Research Centre, Giza, Egypt
E-mail: em.fahmy@nrc.sci.eg, enas_mfahmy@yahoo.com

Keywords: active learning, curriculum learning, deep learning, federated learning, medical, transfer learning

Received: August 19, 2024

We have witnessed a rapid exponential growth of all types of data in all domains specifically in the medical domain. The utilization of machine learning techniques has made significant strides across various domains, with deep learning achieving notable success in recent years. Lately, deep learning has gained increasing attention in the medical field. While deep learning excels at automatically learning discriminative features from raw data, it is still challenging to achieve high performance without a huge amount of data and some handcrafted steps. To address these challenges, deep learning has been incorporated with other new trends and domain knowledge to enhance deep learning's capabilities and improve performance covering the ever-growing needs. Transfer learning utilizes knowledge from natural images, curriculum learning integrates domain-specific knowledge, active learning selects the most informative samples to reduce reliance on labeled data, and federated learning enables collaborative training across organizations while ensuring data privacy. In this review paper, these new trends incorporated with deep learning have been investigated and presented as applications in the medical domain by investigating articles that have applied these trends and published in highly reputable journals in the Science Direct database in recent years.

Povzetek: V pregledni študiji so predstavljeni sodobni trendi strojnega učenja v medicini, kot so transferno, aktivno in federativno učenje na podlagi učnih načrtov, ki v kombinaciji z globokim učenjem izboljšujejo diagnostiko, personalizacijo zdravljenja in varnost podatkov.

1 Introduction

Recently, we have witnessed the growth of all types of data in all domains. Medical data specifically has grown dramatically in the last few years due to the exponential increase of knowledge in the medical domain. Medical data can be found in various forms such as clinical and biomedical data. Biomedical data contains data related to genomics, drug discovery, and biomedicine. Clinical data contains patient records such as medical patients' history, laboratory investigation, and image data from magnetic resonance imaging (MRI), ultrasound (US), X-rays, and computerized tomography (CT) scans. Clinical data exists in 2 forms, structured and unstructured. The structured format includes the disease history and living habits of the patients. While unstructured clinical data such as doctor's investigation records and the conversation between the doctor and patients [1-3]. Therefore, this rapidly growing volume of medical data requires advanced methods for analysis.

Applying artificial intelligence (AI) in the medical domain comprises a promising technology for different healthcare providers. These technologies, particularly data mining, help extract hidden patterns and insights from large datasets using machine learning techniques (MLTs). Traditional MLT includes Artificial Neural Networks (ANN), Decision Trees (DT), Support Vector Machines (SVM), and many other techniques. Machine learning techniques are usually categorized into supervised, unsupervised, semi-supervised, and reinforcement learning. In supervised learning, the labeled data is available, therefore, the model can be trained using this manually tagged data to extract patterns. When there is no labeled training data, unsupervised techniques are employed. It groups similar entities in the same cluster. Each cluster demonstrates a relation between these grouped entities. Semi-supervised depends on a set of hand-crafted extraction patterns and a few tagged instances as initial seeds of the target relation to start the training. The training output is used as the training input for the following generation. The process of learning is repeated for many generations. Reinforcement learning is based on evaluative feedback, so, it can automatically perform goal-oriented learning and process decision-making problems [4, 5].

Deep learning is an advanced form of artificial neural network (ANN), with a larger number of layers than a conventional ANN model to automatically learn the features from the data which makes more refined predictions possible. In numerous recent medical image classification tasks, convolutional neural networks (CNNs), which are a kind of deep learning network particularized in image analysis, were utilized and achieved high performance. The success of CNN in the classification of medical images has motivated researchers to utilize pre-trained models in building new ones. These high-performing CNN pre-trained models have been utilized for different image classification tasks by employing the transfer learning (TL) approach. Pre-trained CNN models utilize features that were learned from a specific domain to fine-tune any other data. They can be utilized as-is to classify new images or to extract features using the output from the layer previous to the output layer and introduce it to another classifier [6].

However, many challenges face the application of machine learning techniques generally and in the medical domain specifically such as I) **the limitations of available datasets** for training the models, that is because collecting and labeling the data is a labor-intensive and expensive task, especially in the case of medical images data such as Ultrasonic imaging (US), CT, MRI. The annotation of data includes the segmentation annotations of abnormality regions and classification labels such as (normal, benign, and malignant). Also, that limitation may result from the scarcity of some diseases with which it is difficult to obtain enough positive cases. II) **The low quality of some data** is another major challenge, where some of the data can be found unlabeled, inconsistent, inaccurate, or in an unstructured format—such as handwritten notes, radiology reports, and conversations between doctors and patients—which are difficult for machine learning algorithms to process effectively. In the case of medical image modalities, there may be variations in image resolution and quality. III) **The shortage of explanations** of pathological basis such as the diagnosis reasons, where the techniques depend only on the differences between the normal and patient cases. For healthcare professionals to trust and act on ML-generated results, it is essential to understand how these models arrive at their predictions. IV) **Ethical and regulatory concerns** play a crucial role, where the healthcare industry is tightly regulated, and machine learning models must comply with stringent standards to ensure patient privacy, data security, and model safety. Furthermore, any biases in the data could lead to unequal or unfair treatment recommendations, making fairness an ongoing concern in the application of ML in healthcare [7-10].

Despite these challenges, machine learning presents a wealth of opportunities that can significantly improve healthcare outcomes such as I) **diagnostic accuracy and speed** where ML algorithms, particularly deep learning models, have demonstrated remarkable success in automating and enhancing diagnostic processes, especially in medical imaging. For instance, ML models can analyze radiographs, MRI scans, and other images to identify abnormalities such as tumors or lesions with a level of precision that often rivals or exceeds that of human experts. This capability can lead to earlier detection, which is critical for improving patient outcomes, particularly in cancer and cardiovascular diseases. II) **personalized medicine** by analyzing large datasets, including patient demographics, genetic information, and medical history, machine learning can help tailor treatments to individual patients, optimizing therapeutic interventions based on their unique characteristics. III) **Predictive analytics** is another powerful opportunity that ML offers. By analyzing trends in patient data, machine learning models can predict disease progression, forecast complications in chronic conditions, and identify high-risk patients who may benefit from earlier interventions. IV) **Automation** is another key opportunity in healthcare, with ML models capable of automating routine tasks such as image analysis, patient triaging, and administrative work. This allows healthcare providers to focus more on direct patient care, improving overall efficiency. V) **Drug discovery** by identifying promising drug candidates and predicting their behavior in the human body, which can reduce the time and cost associated with bringing new medications to market [7-10].

In response to the mentioned challenges, recent research has shifted towards using advanced techniques such as deep learning with some incorporated techniques and domain knowledge like transfer learning which provides deep learning with information from natural images. Curriculum learning integrates domain knowledge through training patterns of the processed task. Active learning explores the most informative samples and retrieves them from an unlabeled pool to fulfill better performance with less labeled data. Federated learning allows many organizations to collaborate on deep learning without sharing clients' data or devices which provides efficient data access and security and an improvement of the learning model utilizing a large decentralizing dataset. The purpose of this research is to illustrate the new trends of machine learning in the medical domain. The selected articles that are reviewed show these new trends in the medical domain using different medical dataset types including medical images, tabular datasets, genes, etc. in different tasks. The remainder of this research is organized as follows: Section 2 illustrates the different types of medical data. Section 3 presents some data preprocessing steps. Section 4 presents the new trends of MLTs. Section 5 describes the search methodology for articles that apply the mentioned new trends of MLTs in the medical domain. Section 6 presents some of the applications of new trends of MLTs in the medical domain. Section 7 presents the conclusion and some of the recommended points for future work.

2 Types of medical data

Medical data can be found in different forms such as arrays of numerical data, images, sequences of DNA, amino acids, ...etc. For developing any ML model, the data is split into three parts which are training, validation, and testing. The training part is used to learn to tune the parameters of the model. The validation part is used to stop overfitting, and the test part is used to assess the performance of the model. In the next subsections, a brief overview of different medical data forms will be presented.

2.1 Numerical data

Different diseases' related data are found as an array of lab tests which is numerical data. These numerical datasets can be used to manage the related diseases such as the datasets available on the UCI machine learning repository [11]. Most numerical data are available in table form such as Excel sheets or database tables where rows represent samples from patients and columns represent different features that describe the intended diseases or vice versa. A huge number of numerical datasets are available such as the patient demographics of some diseases like COVID-19, and the lab results for different diseases such as thyroid, heart disease, dermatology, cancer, etc.

2.2 Microarray gene expression data

Microarray techniques provide a platform for measuring the expression levels of thousands of genes in various conditions. It is composed of a small glass slide or membrane that contains samples of many genes arranged in a regular pattern. It is used to find genes associated with specific diseases by analyzing and finding the differences between two mRNA sets, one set is from normal cells and the other set includes cells from pathological tissues such as cancer cells. Microarray data contains a lot of redundant genes, and many genes include inappropriate information for the accurate classification of diseases. Thus, the analysis of the large amount of data generated by this technology is not an easy task for biologists [12]. Figure 1 shows a cDNA microarray spotted on a glass surface. While, in Figure 2 the general structure of the microarray is illustrated, which is represented as an array of numerical values. Cancer gene expression datasets for leukemia, lung, prostate, etc. can be found in [13].

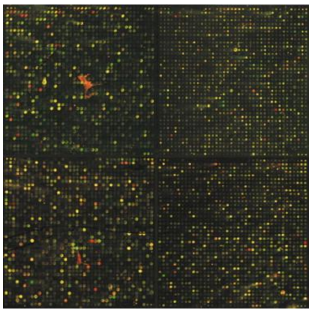


Figure 1: cDNA microarray spotted on a glass surface. https://www.cell.com/fulltext/S0960-9822%2898%2970103-4

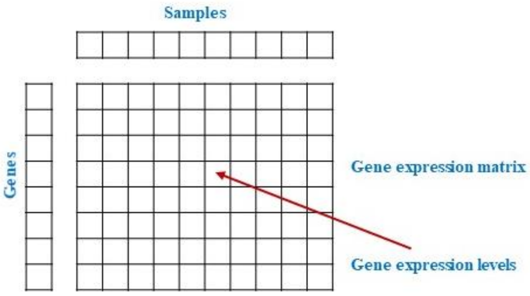


Figure 2: General structure of microarray.

2.3 Image modalities

Information obtained from medical imaging modalities is clinically beneficial in many applications like computer-aided detection, diagnosis, and treatment planning. Many imaging modalities can be used to check abnormalities in different body organs. They include radiation such as CT, X-rays, US, and MRI. They are categorized according to the method of producing images. They help radiologists to recognize abnormal regions. The interpretation of different image modalities needs expertise, where it is operator dependent. Therefore, the process of reading image modalities is exhausting, costly, and prone to error.

Ultrasound (US) is a suitable modality for tumor detection. It can estimate the size of the tumor and distinguish abnormalities. Its capability of detecting contra-lateral malignant lesions is limited [14].

Magnetic resonance imaging (MRI) produces images relevant to the displaying of hydrogen atoms to radio waves and magnetic fields. MRI images are valuable as they present physiology and anatomy. It images the target organ and prepares it as thin slices; moreover, it provides information about the vascularity of the tissue [15].

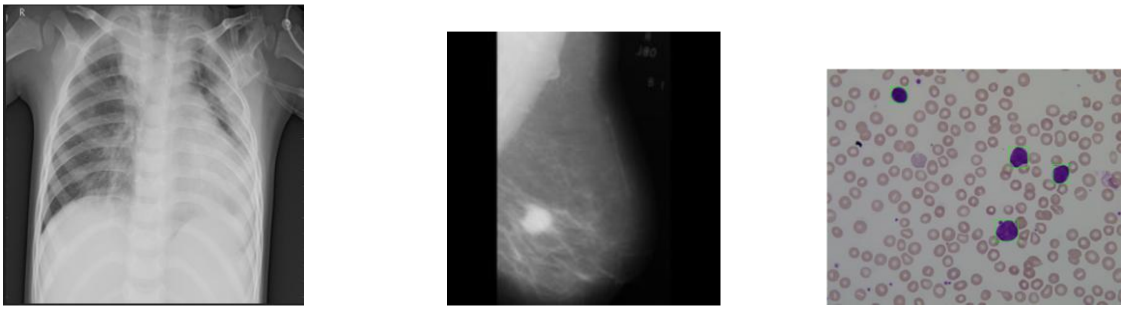
Computed tomography (CT) scanners display better image clarity using multiple X-ray sources and detectors [15]. Radiation X-ray generated images are 2-dimensional images. Fluoroscopy units show real-time moving images produced by X-ray exposure. Angiography is a widespread usage of fluoroscopy, imaging blood flow in vessels [15].

Digital Mammography (DM) is an X-ray imaging that is specialized for breast tissue. DM is the most common and most important screening method in clinical practice. It can detect tumors before they develop further and become easily detected and felt by the physician [16].

Microscopic images are the images that are captured by the microscope to enlarge small scanned objects and extract fine details that cannot be obtained otherwise [17]. Figure 3 shows samples of different image modalities for different body organs.

2.4 DNA and protein sequences

The fast growth of sequencing resulted in huge numbers of DNA and protein sequences. Sequences can be used to predict diseases associated with a given either DNA or protein sequences. DNA is a long polymer chain of units named nucleotides; it exists in a double helical shape as shown in Figure 4. There are 4 types of nucleotides which are A (adenine), C (cytosine), G (guanine), and T (thymine), they are considered the alphabet of DNA. They are arranged into sequences of 3-letter called codons. A double-stranded helical structure of DNA would be complementary, where “G” is chemically combined with “C”, and "A" with "T" within the replication of DNA [18].



(a) X-ray of Lung [19] (b) DM of breast [20] (c) Microscopic blood image [21]

Figure 3: Samples of different image modalities for different body organs.

Amino acids are linked into linear chains to produce proteins. The properties of proteins are defined by the composition of their amino acids. The triplets of consecutive DNA nucleotides which are called codons are responsible for the forming amino acid sequence in a protein. There are $4^3 = 64$ various codons formed from the 4-letters [22], which is more than 3 times larger than the number of amino acids which is 20 amino acids, 3 of which represent stop codons and one is a start codon. While the remaining codons are responsible for generating the 20 amino acids. So, it is possible that more than one codon maps the same amino acid [18]. Figure 5 shows the transcription of DNA sequence into molecules of mRNA; then, the translation of the transcribed mRNA into the associated chain of amino acid sequence, which later fold into fully functional proteins.

Single nucleotide polymorphisms (SNPs) are the most common human genetic variations as mutations or insertions/deletions (indels). If SNPs have changed the codon triplets without changing the encoded amino acid, it is synonymous (sSNPs) while the gene is not mutated. Otherwise, it is non-synonymous (nsSNPs), as it changes the codon while the encoded amino acid is changed into various amino acids which are called missense mutations which are the reason for many diseases [23, 24]. Figure 6 shows single nucleotide polymorphisms (SNPs).

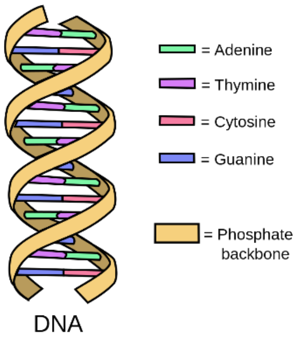


Figure 4: Chain of DNA sequence.
http://acer.disl.org/news/2016/08/17/tool-talk-gene-sequencing/

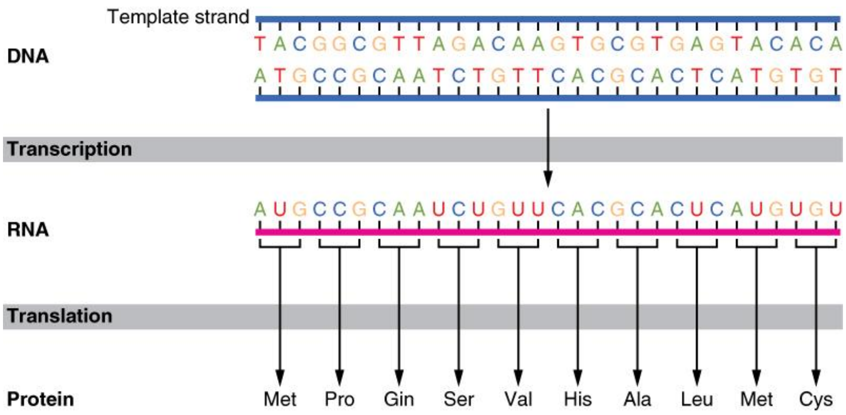


Figure 5: The process of translation from DNA sequence to the associated amino acid sequence.
https://courses.lumenlearning.com/suny-ap1/chapter/3-4-protein-synthesis/

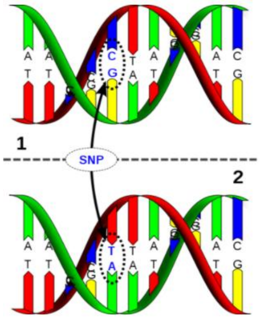


Figure 6: Single nucleotide polymorphisms (SNPs).
https://isogg.org/wiki/Single-nucleotide_polymorphism

3 Data preprocessing

The knowledge discovery includes 3 main phases which are the preprocessing phase, the data mining phase, and the post-processing phase. Data pre-processing is a crucial phase of knowledge discovery to build an accurate machine learning model. In the preprocessing phase, a set of data preprocessing steps are performed (cleaning the data from the noise, handling missed values, merging appropriate data from different databases, normalizing the data, extracting features, and selecting the most informative features) to prepare the data for data mining phase. Datasets can also be small therefore the relevant features have not been captured and thus data augmentation is performed by applying different data augmentation techniques. Data mining, which is the core phase in knowledge discovery, is performed by applying MLTs. The preprocessing facilitates the application of the MLTs to extract important patterns or correlations. In the post-processing phase, the discovered knowledge is refined and improved then interpreted into meaningful knowledge for the user’s presentation [25].

Feature selection is a key difference preprocessing step that should be highlighted when comparing deep learning with the traditional MLTs so it will be tackled in more detail in the next subsection.

3.1 Features selection

Feature selection is the process of finding the optimal feature subset that is strongly distinguishing among different classes. The purpose of this process is the reduction of the dataset and the elimination of redundant and irrelevant features that impact the classification process negatively. Feature selection is a combinatorial optimization problem its aim is to select the feature subset with the least number of features that achieves the highest possible classification accuracy. It is one of the data preprocessing for pattern recognition and data mining specifically when working on high-dimensional datasets [26, 27].

Feature selection has 2 main approaches: the filter and the wrapper. In the filter approach, the feature selection is based on statistical individual feature ranking. It is easily implemented but eliminates the interaction among features and does not rely on the applied ML algorithm to the selected features. Whereas, for the wrapper approach the feature selection depends on the outcome of the ML algorithm to decide how favorable the features subset is. Candidate solutions of feature subsets are iteratively generated and their characteristics are assessed by the applied ML algorithm [28].

The wrapper-based feature selection approach evaluates the quality of feature subsets using the learning algorithm. Thus, it can determine and discard irrelevant and redundant features effectively. As the learning algorithm is frequently used in the search process, high computational time is required, especially when the datasets are large. On the other hand, the hybrid methods aim to utilize the advantages of both approaches; the efficiency of computation of the filter approach and the high performance of the wrapper approach [29].

Feature selection algorithms based on heuristic search methods are needed as the computation of a huge number of features is not feasible. Many meta-heuristics approaches have been used for feature selection. From these algorithms are the nature inspired algorithms such as genetic algorithm (GA), firefly [30, 31] and ant colony optimization (ACO) [32, 33].

4 New trends of machine learning techniques

After data preprocessing, various Machine Learning Techniques (MLTs) are applied to uncover hidden patterns and correlations in the data. As mentioned earlier, disease-related data, often represented through numerical lab tests, microarray data, medical imaging, and genetic sequences, can be processed to predict disease presence or other related tasks. Traditional MLTs like Support Vector Machines (SVM), Decision Trees (DT), and K-Nearest Neighbors (KNN) are effective across these data types for decision-making. However, Artificial Neural Networks (ANNs), which mimic the brain’s neural structure, are increasingly used for more complex tasks and in various domains, especially the medical domain. ANNs consist of input, hidden, and output layers, and are trained through techniques like backpropagation. Their success across domains, especially healthcare, has led to the development of Deep Learning (DL), a more advanced form of ANN. This section explores the role of deep learning in modern machine learning.

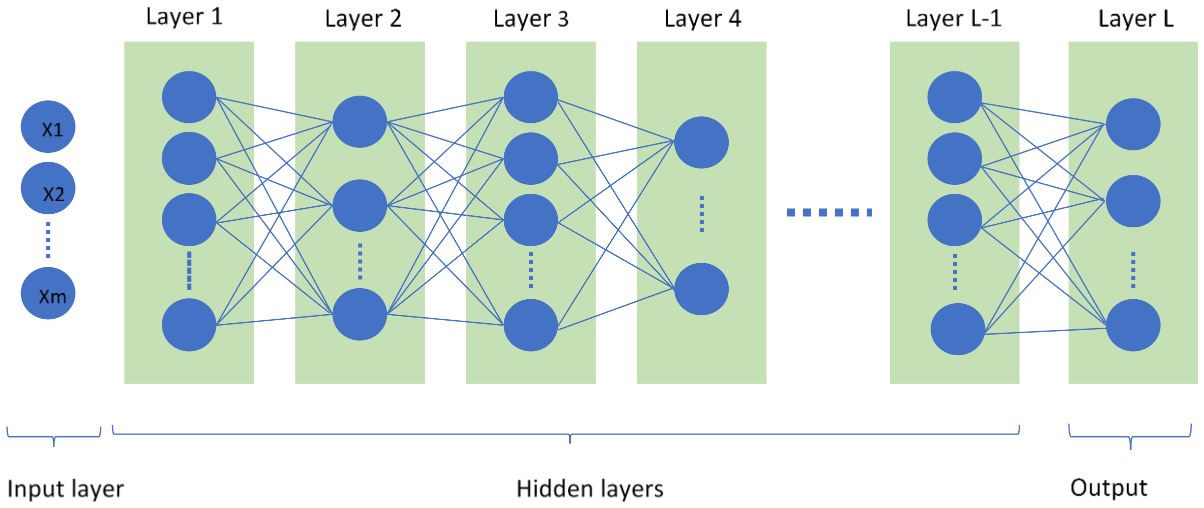


Figure 7: The general structure of Deep Neural Network.

4.1 Deep learning (DL)

Deep Learning (DL) models have gained prominence due to their ability to automatically extract complex patterns from data, eliminating the need for manual feature engineering. However, DL models require large datasets, making them particularly suited for high-dimensional data, such as in medical fields, where they can uncover intricate structures through multiple intermediate layers. The depth of a DL model—referring to the number of hidden layers—enables it to learn complex mappings between input and output. Unlike shallow networks, which struggle with intricate data patterns, deeper networks excel at learning these relationships [34, 35]. Figure 7 shows the general structure of Deep Neural Networks (DNNs).

There are several deep learning algorithms such as Convolution Neural Network (CNN), radial basis function networks, deep belief networks, autoencoders, and Recurrent Neural Network (RNN) [35, 36].

Deep learning depends on hyperparameters such as activation function, learning rate, batch size, number of epochs, optimizer, dropout rate, etc. Different deep learning algorithms, like RNNs and CNNs, also have additional specific hyperparameters. Adjusting these hyperparameters is critical, as their values significantly affect the model's behavior. Finding the optimal combination of hyperparameters can be an exhaustive task, requiring substantial computational resources and time [37, 38].

The performance of a DL model heavily depends on the selection of these hyperparameters, particularly in complex domains like medical data analysis. Medical data often have high dimensionality, noise, and imbalanced class distributions, making hyperparameter optimization crucial to enhancing model performance. Careful selection improves robustness and generalizability, ensuring reliability in real-world clinical settings. While methods like grid search and random search are widely used, more advanced techniques, such as Bayesian optimization, offer significant advantages [37, 38].

Common techniques for hyperparameter optimization:

1. **Grid search**: Exhaustively searches across all possible hyperparameter combinations. While it guarantees to find the best parameter set within the grid, it can be computationally expensive, especially for models with many hyperparameters.

2. **Random search**: A more computationally efficient approach, randomly selecting combinations of hyperparameters from specified ranges. It often achieves comparable or better results than grid search in fewer trials.

3. **Bayesian optimization:** An advanced method that builds a probabilistic model of the objective function. It predicts the best hyperparameters based on past performance, guiding the search toward the most promising regions with fewer trials. Libraries like Optuna and Hyperopt can implement Bayesian optimization efficiently.

For example, CNN can be used to classify medical images like X-rays or MRI scans. Random search can explore different values for hyperparameters (e.g., learning rate, batch size, number of layers). Alternatively, Bayesian optimization can be used for a more efficient search, predicting the most promising hyperparameters configurations based on prior evaluations. By optimizing the model’s parameters using these methods, we can improve classification accuracy, reduce overfitting, and ensure the model performs well on unseen medical data.

Advantages and Limitations of DNNs for Medical Data

Advantages:

- **Versatility:** DNNs can be adapted to work with various data types, including structured clinical data (e.g., patient demographics, lab results), unstructured data (e.g., free-text medical records), and image data.

- **Feature Learning:** DNNs can automatically learn relevant features from the data, making them more flexible than traditional machine learning algorithms that rely on feature engineering.

Limitations:

- **Training complexity:** Training deep neural networks can be computationally expensive and time-consuming. Additionally, DNNs require large datasets to avoid overfitting.

- **Overfitting:** If not carefully tuned, DNNs can overfit to small or imbalanced datasets, a common issue in medical data where datasets may not be as large or diverse as needed for training.

Use Cases in Medicine: DNNs have been applied to a variety of tasks in medicine, including predicting patient outcomes, disease progression modeling, and disease classification from images and clinical data.

CNNs and RNNs are two of the most common and promising deep learning algorithms used in medical applications. These algorithms have demonstrated success in a variety of tasks, such as medical image classification and time-series data analysis. Further details on these algorithms will be discussed in the next subsections.

4.1.1 Recurrent neural networks (RNN)

Recurrent neural networks (RNNs) are neural networks that contain memories that can capture the stored information in the prior element of the given sequence. Therefore, RNN is suitable for processing sequential data types such as the diagnostic history of patients, DNA and protein sequences, etc., where the information is remembered through the network. RNN is called recurrent because it executes the same task for each element of the input sequence while its output is based on the prior computations (memory). Thus, the decision of recurrent net at time t-1 affects the decision that will be taken later at time t. Therefore, RNN has 2 sources of input, the recent past and the present, which are combined to define the response to new data. Figure 8 shows the architecture of the RNN in which a set of input x values are mapped into a sequence of output o values. A loss L measures the difference between the expected output o and the actual output y [35].

Where x, h, o, L, y symbolizes input, hidden state, output, loss, and target value. A weight matrix U defines the input to hidden connection, a weight matrix W defines the hidden-to-hidden connection, and a weight matrix V defines the hidden-to-output connection. Then, from time step t = 1 through time step t = n, the following equations are used:

$$a_t = b + Wh_{t-1} + Ux_t \quad (1)$$

$$h_t = \tanh(a_t) \quad (2)$$

$$o_t = \mathrm{c} + Vh_t \quad (3)$$

$$\widehat{y_t} = \mathrm{SoftMax}(O_t) \quad (4)$$

The forwarding propagation of RNN is defined by the preceding equations, where b and c are the bias vectors, while tanh and SoftMax are the activation functions. To update the weight matrices U, V, and W, we compute the gradient of the loss function for each weight matrix. Gradient computation requires both forward and backward propagation of the network. Any loss function can be used depending on the goal. At each time step, the sum of all losses is the total loss for a particular sequence of x values.

However, traditional RNNs suffer from gradient exploding and gradient vanishing issues, making them unsuitable for long-term dependencies. On the other hand, long short-term memory (LSTM) is effective in capturing long-term time dependence. LSTM networks address this by introducing gating mechanisms that control the memory flow, allowing for better long-term sequence learning. Gated Recurrent Units (GRUs) offer a simplified version of LSTMs with similar benefits but fewer parameters.

Advantages and Limitations of RNNs for medical data

Advantages:

- **Sequential data processing:** RNNs can handle different types of sequential data, including time series and text, where the past medical history or a series of clinical events influence future outcomes.

- **Memory of past inputs**: RNNs can remember information from previous time steps in the sequence, allowing them to capture temporal dependencies in data. This is particularly useful for tracking disease progression over time or analyzing patient histories.

Limitations:

- **Training difficulties**: RNNs are prone to the vanishing gradient problem, especially in long sequences, making them harder to train effectively.

- **Data complexity:** RNNs are best suited for data where the relationship between input and output is sequential. For static data like images or tabular data, CNNs or DNNs might be more appropriate.

- **Resource intensive**: Training RNNs, especially on long sequences, can be computationally expensive.

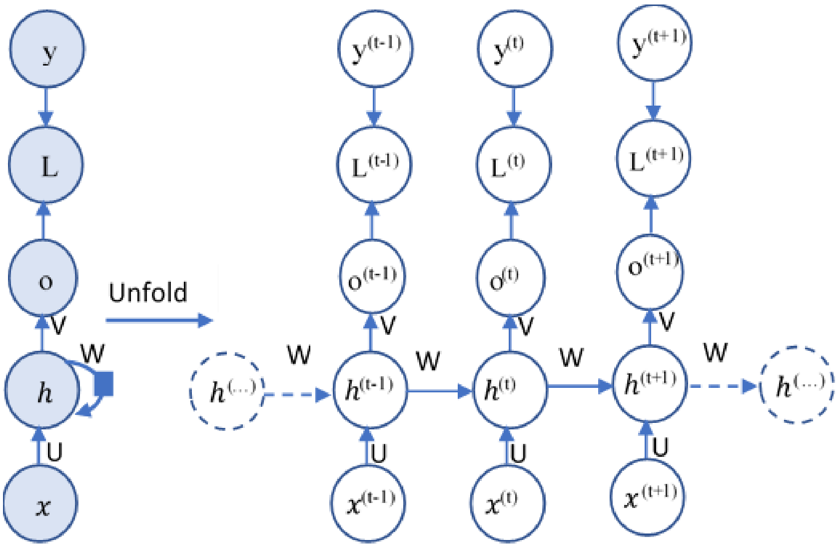


Figure 8: The architecture of recurrent neural network.

Use cases in medicine: RNNs (and their variants like LSTMs) are commonly used in medical applications such as gene sequence classification, predicting disease progression over time, and analyzing time-series medical signals (e.g., ECG readings) [35].

4.1.2 Convolution Neural Network

CNNs are a type of deep learning network specialized for image analysis. Unlike traditional MLTs that rely on manual feature extraction, CNNs can automatically learn hierarchical features from raw image data. This is especially useful in the medical field, where CNNs are applied to analyze medical images for tasks like disease detection and classification [36, 39].

It contains an input, an output and many hidden layers which represent convolutional networks. Convolutional network includes three types of layers: convolutional, activation, and pooling. The convolutional layers apply filters to detect features (edges, textures, etc.). As the image proceeds through layers, the filters can detect more sophisticated features. The activation function like Rectified Linear Unit (ReLU) follows the convolution layer to control the output, it introduces non-linearity. Pooling layers reduce the dimensionality of the data, making the model more computationally efficient and less sensitive to minor positional changes in the features. The final layer is fully connected, producing predictions for classification tasks. The overall number of network parameters is defined by the number of layers, the number of neurons in each layer, and the connection between neurons. The weights should be tuned through the training phase to achieve good performance [40].

convnet processes the image (I) using a matrix of weights called filters which can recognize certain features at specific positions. At a specific layer l, the feature map at position (i, j) is defined as h_{ij}^{l}, the bias as b^{l}, and the weight as W^{l}. The feature map can be expressed as follows:

$$h_{ij}^{l} = \mathrm{ReLU}((W^{l} * I)_{ij} + b^{l}) \qquad (5)$$

Where ReLU is activation function which controls the output. The basic structure of the CNN is shown in Figure 9.

Advantages and limitations of CNNs for medical data

Advantages:

- **Feature extraction**: CNNs automatically learn hierarchical features from raw image data, eliminating the need for manual feature extraction, which is time-consuming in traditional methods.

- **Spatial relationships**: The convolutional layers can detect local patterns (e.g., edges, textures) in images, which are crucial for tasks like tumor detection or organ segmentation.

- **Efficiency**: CNNs are computationally efficient due to shared weights in convolution layers, allowing them to process large datasets more effectively.

Limitations:

- **Data requirements**: CNNs require large labeled datasets to perform well, which may not always be available in medical settings.
- **Limited to spatial data**: While CNNs excel in image-based data, they are not as effective for

non-spatial data like time-series or sequential data.

Use cases in medicine: CNNs have been widely applied in diagnostic tasks such as detecting cancers, classifying lesions, and analyzing radiological images (e.g., X-rays, MRIs, CT scans) [35, 36, 41].

Recent advances in CNN, like AlexNet [42], VGGNet [43], GoogLeNet [44], and ResNet [45], have significantly improved image classification accuracy, with models now outperforming human experts in some cases. These networks have been trained on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) using millions of annotated images [46, 47]. And their success has spurred the rise of transfer learning, where pre-trained models are fine-tuned for specific tasks [48]. In the next subsection, a brief description of some of the high-performance pre-trained models using ImageNet.

4.1.2.1 Visual geometry group network VGG 16–19

VGG16 network is the winning model architecture of the 2014 ImageNet competition. VGG consists of 16–19 layers. The size of the input image to VGG is (224×224). VGG has a set of convolutional filters with small sizes (3×3) to capture the information of the up/down and left/right center. The size of the pre-trained weights is 528 MB. The overall number of parameters of VGG16 is 138 357 544 parameters [43].

4.1.2.2 InceptionV3 model architecture

InceptionV3 network is the winning model architecture of the 2015 ImageNet competition. The Inception V3 model has a total of 48 layers. The size of the input image to InceptionV3 is (299×299). It is deeper than VGG16 but with fewer parameters. The size of the pre-trained weights is 92 MB. It has 23 851 784 parameters [44].

4.1.2.3 Residual neural network (ResNet)

ResNet network is the winning model architecture of the 2016 ImageNet competition. ResNet-50 contains a 50-layer architecture. The size of the input image to ResNet is (224×244). The size of the pre-trained weights is 99 MB. It has 25 636 712 parameters [45].

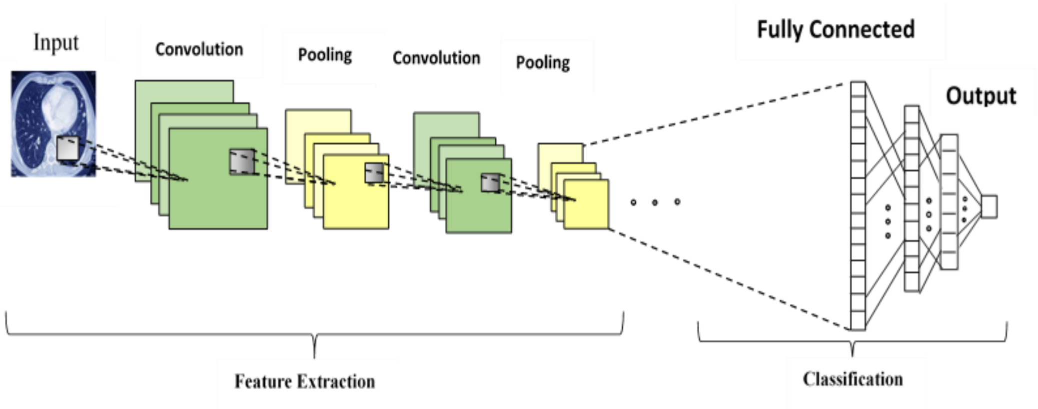


Figure 9: The basic structure of CNN.

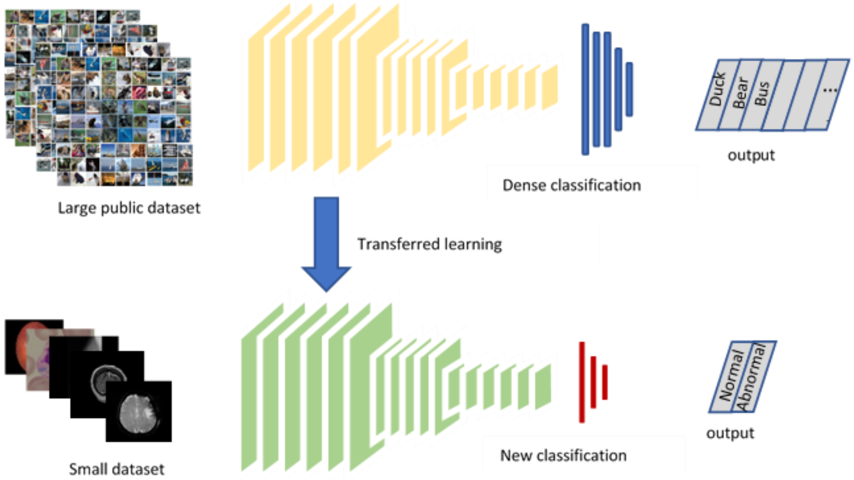


Figure 10: Transfer learning architecture.

4.2 Transfer learning

Transfer learning is a more appropriate approach when the available data for training is limited. In transfer learning, an intricate model can be trained using available large-scale annotated images such as natural images. Therefore, the TL process drives knowledge from source domain (e.g. natural images) to a target domain or network where the domain images are limited. Only the small amount of available annotated data of the target domain is used to tune the model. Where the fundamental features used for classification are similar between domains, retraining the entire model is unnecessary. In such cases, TL allows for the transfer of learned features, with only the classification layer(s) being retrained on the small new dataset [48]. TL leverages pre-trained models such as VGG [43], NasNetLarge [22], Inception GoogLeNet [44], ResNet [45], etc—that have been developed for image classification and they have been presented at the annual ILSVRC [46, 47]. TL saves a great amount of time lost in developing and training CNN models. The pre-trained model or the required part of the model can be incorporated directly into the new model and used as a classifier, standalone feature extractor, integrated feature extractor, or weight initializer [48, 49]. Figure 10 shows transfer learning architecture.

4.3 Curriculum learning

In the standard educational method, learning depends on a curriculum that presents new concepts based on previously collected ones. The rationale beyond this is that people pick up better if the information is introduced in a meaningful method instead of randomly. By using the same ideas to train neural networks starting with simple cases, it was noticed that the networks perform better, which indicates the significance of gradual and systematic learning [50].

The curriculum learning (CL) approach is motivated by the capacity of humans to pick up new tasks fast with finite "training sets". Similarly, the training procedure of medical students called teacher-student curriculum learning is based on training by tasks with gradually growing difficulty. While each task uses smaller datasets than those utilized in machine learning. Like, students can start with a simple task, such as deciding if an image includes lesions, and later are asked to determine if the lesions are malignant or benign which is a more complicated task. With time, they will progress to a more complex task, like recognizing the subtypes of lesions [8].

In machine learning, CL works with a series of training samples sorted in increasing order according to learning difficulty. The order in which the samples are introduced to the model is critical, as it can significantly impact the model's performance. Curriculum learning is an active area of research, particularly in applications such as medical image diagnosis [8].

A key point in CL is the design of data schedulers that control the sequence in which training samples are fed into the model. These schedulers can use a variety of methods to determine sample difficulty, such as expert input, heuristics, or natural language processing (NLP) applied to radiology reports.

Given a sample xi which should be assigned to a class label Ci ϵ {C1, C2, …., Cm}. Suppose the training set consisted of pairs {X, C}, and the training is processed in batches of size B for a total of E epochs. To train CNN with CL, prefer to start the training with simpler samples. Practically, CL is performed by assigning a probability to every training pair, where the simpler samples are given higher probabilities to be chosen first. Initially, every sample xi is assigned a probability pi(0). At the beginning of each epoch e, the training set {X, C} is permuted to $\{X, C\}^k$ by the reordering function F(e). Where this mapping is produced by sampling the training set based on the probabilities at the present epoch pi(e). After executing many iterations, these probabilities are updated using a scheduler, aiming to achieve a regular distribution by the end of the training process [50].

4.4 Active learning (optimal experimental design)

Supervised learning techniques rely heavily on annotated data. Although more datasets are becoming available, the effort, cost, and time required to annotate them remain significant. On the other side, any error especially in some important applications such as those in the medical domain can have severe consequences. Achieving reliable outcomes often requires an interactive process where predictions are reviewed or modified by an oracle or user. This means users must be able to override and adjust automated predictions to meet specific criteria. Techniques such as Active Learning (AL) or what is called Human-in-the-Loop computing have witnessed progress in overcoming these challenges [51].

Active learning is a semi-supervised learning approach that begins with a small set of labeled samples (seed samples) and iteratively selects the most informative samples from a pool of unlabeled data for annotation. By focusing training on the most informative subset of samples, AL improves model performance and reduces the annotation burden, particularly for image data. In AL, an MLT scans unlabeled data and recognizes the most informative samples. These samples are then presented to a human annotator (oracle) for labeling. This makes AL a part of the Human-in-the-Loop paradigm, where only selected samples are used for training, often far fewer than in traditional supervised learning [51].

Formally, suppose that U is available big pool of unannotated data and that there are oracles to request annotations for any unannotated sample x_U to be added to annotated set L. The goal is to train a model $f(x \mid L*)$ using the annotated set $L* \subseteq L$. A brute-force solution would involve requesting the oracle(s) to annotate each sample x_U, resulting in $L* = L$. However, this is a costly and not practical solution. Theoretically, there is an optimal subset $L*$ of data that can achieve performance equivalent to that obtained using the whole annotated dataset L, i.e. $(f(x \mid L*) \approx f(x \mid L))$. AL is a trend of ML that tries to explore this optimal subset $L*$, where the current model is $f´(x \mid L´)$, $L´$ is an intermediate annotated data. AL intends to iteratively explore the most informative data samples x_i^* to train the model, assuming that the unannotated data samples and the model will evolve through time, rather than choosing a constant subset of samples once for training.

The selection of samples to be annotated is based on the informativeness of these requested samples. The evaluation of the informativeness of each un-annotated data sample x_U is done given $f'(x_U \mid L')$, then all selected samples are demanded to be annotated. After the annotations, the new annotated data has been used to improve the model. This is done by retraining the whole model using all available annotated data L', or by using the most recently annotated sample x_i^* to fine-tune the network [51].

Active Learning typically employs three methods to select samples for annotation:

Stream-based selective sampling supposes the existence of a continuous flow of unannotated data samples x_U. In this method, the present model and informativeness $I(x_U)$ measure are the criteria used to specify, for each incoming sample whether or not to require an annotation from the oracle(s). Thus, while the model is being trained, it is offered a data sample and instantly decides if it needs to query for the label. Although this type of query is inexpensive, its performance is limited because it does not consider the broader context of the underlying distribution, but it depends on the separation nature of each decision, therefore the balance between the exploration and exploitation is less than in other query kinds.

Membership query synthesis generates the sample x_G^* that the model believes to be most informative, rather than selecting from real-world data. Therefore, it is annotated by the oracle(s). This method may be very effective in bounded domains, but it may struggle when the model has no knowledge of unrepresented areas of the data distribution, similar to stream-based methods.

Pool-based sampling selects N data samples $x_0^*, \ldots, x_N^*$ from a large unlabeled dataset U to pull samples from. Pool-based approaches use the present model to do a prediction on un-annotated data samples to get a ranked measure of informativeness for each data sample in the un-annotated data. The highest N informative samples are selected for annotation by the oracle(s). Therefore, the model is initially trained on labeled samples which are then used to find which data samples would be most informative to be inserted into the training set for the next AL loop. This approach has proved to be the most promising, which depends on batch-based training. Figure 11 shows the full process of active learning.

AL uses some informativeness measures of unlabeled samples to select the most informative samples. They depend on probabilities, these approaches are least confidence sampling, margin sampling, and entropy sampling [51].

Least confidence sampling the model selects the highest uncertainty sample or least confidence for annotation and therefore is given to the oracle to be labeled.

Margin sampling can be utilized in a multi-class, it uses the first and second most likely labels and computes the difference between them, then considers the sample that has the smallest difference between the first and second most likely labels to be annotated.

Entropy sampling uses entropy as it is a measure of uncertainty to select a sample to be annotated. Entropy measures the amount of information gained by considering a sample and so selects the sample that maximizes the information that has the largest entropy value [51].

4.5 Federated learning

Federated learning (FL), developed by Google in 2017 is a collective distributed decentralized learning method that allows many organizations to collaborate on machine learning or deep learning models without sharing clients' or devices' data. It allows the training data to be on the decentralized edge devices rather than keeping it in a data center. These individual nodes or devices jointly train a machine learning or deep learning model from their local data and then aggregate the devices' training outputs on the server to update the global model without sharing edge data. The resulting model can be shared among all participating devices or clients. Therefore, it provides secure models that fulfill an efficient solution while providing data access and security [52-54].

One major issue with centralized models is that medical organizations do not allow to break doctor-patient confidentiality by providing medical images such as CT and X-ray images for training purposes because of privacy, legal, and data-ownership issues. To develop deep learning models for the medical domain, large medical data is needed to develop these models. Therefore, many medical researchers illustrated that federated learning is a good technique to connect different medical organizations and let them share their experiences while keeping privacy. Furthermore, the performance of the learning model will be improved using a large medical dataset. However, the resulting models may be biased toward organizations that have larger training datasets [53].

In federated learning, the process begins by sending a global model with unified initial weights to each client. At each client side, there is a local dataset, where the model is trained in each separately. After completing local training, the client sends its model updates back to the server, which aggregates these updates to refine the global model, while the data at the clients remains local in each client. The server has the authority to manage the whole process where it sends the model to the client, collects the updates, and synchronizes them to build the updated model with the new parameters. This method enables medical organizations to collaborate on training models while maintaining data privacy. There are different federated learning algorithms according to the computation method of gradients such as federated stochastic gradient descent, federated averaging, and federated learning with dynamic regularization. [53, 54]. Figure 12 shows the architecture of federated learning.

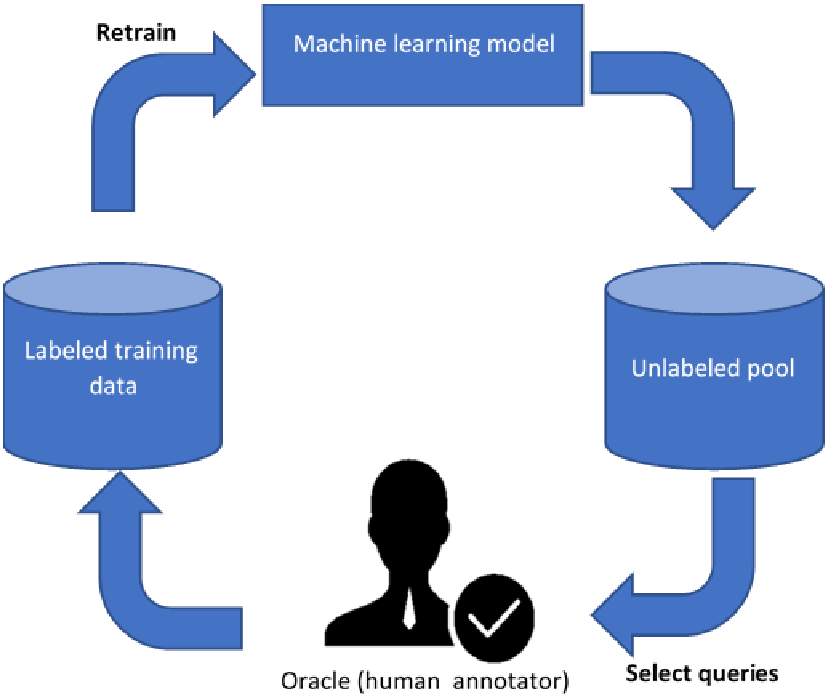


Figure 11: The process of active learning.

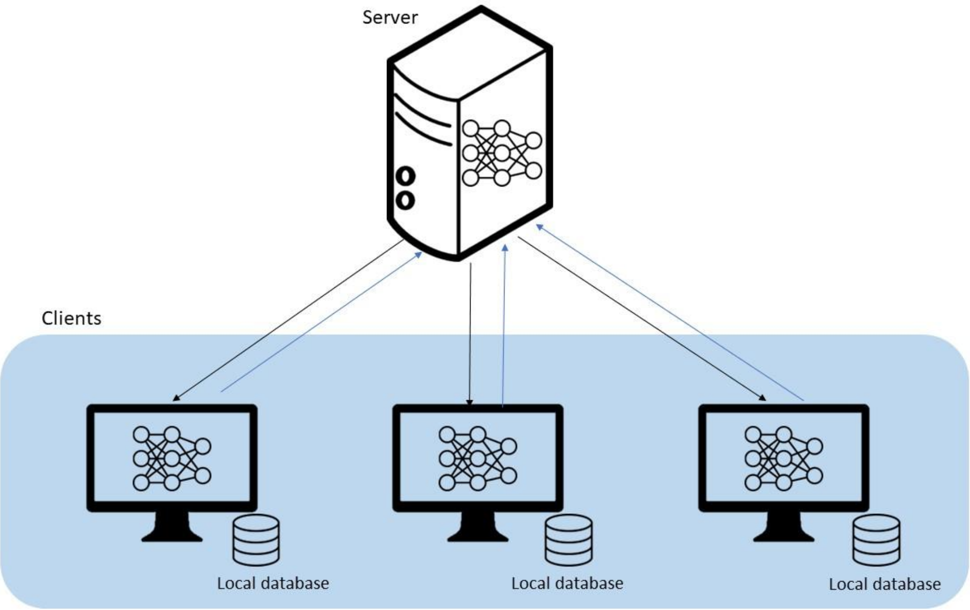


Figure 12: Federated Learning architecture.

5 Search methodology

5.1. Search criteria

This research investigates recent trends in machine learning (ML) within the medical domain. To achieve this, we explored a ScienceDirect (Elsevier) (http://www.sciencedirect.com). The following keywords were used in the search: “active learning”, “curriculum learning”, “deep learning”, “transfer learning” and “federated learning” to investigate the different research that utilizes these recent trends. Additional keywords—"medical", "disease", "cancer" and "gene" were included to focus the search on medical applications that used these new trends. Although the search intended to retrieve the articles related to any disease, "cancer" was added to retrieve more relevant results, given that much of the recent research in ML is focused on cancer. Publications from 2016 to 2024 were considered. The composition of the used terms to form the search query used for deep learning-based techniques in the medical domain was:

"Deep learning" AND ("medical" OR "Disease" OR "Cancer" OR "Gene").

Where the aim of this research is to find the new trends in machine learning techniques which after accurate investigation were found to be mostly based on "deep learning" either alone or combined with other new techniques such as "transfer learning", "active learning", "curriculum learning" and "federated learning", so the same query was used as for deep learning-based techniques with adding the other techniques' keyword as follow:

("Deep learning" AND "*") AND ("medical" OR "Disease" OR "Cancer" OR "Gene")

Where "*" can be replaced by any of the other techniques' keywords ("transfer learning", "active learning", "curriculum learning", and "federated learning").

The following criteria were applied to select the publications: (1) Articles related to human diseases (other organisms' related diseases are excluded); (2) Inclusion of at least one of the new ML techniques; (3) Only complete research articles were included (excluding letters, surveys, book chapters, and non-English articles); (4) publications published from 2016 to 2024.

5.2. Data extraction

As the search retrieved a large number of articles, therefore only a subset of the retrieved articles was selected for analysis. Figures 13-15 illustrate the number of publications per year for the various techniques between 2016- 2024, based on "Elsevier" database to show the growth rates of these new trends.

- Figure 13 shows the steady increase in deep learning publications, from 17 articles in 2016 to 2958 in 2024, indicating a growing interest in applying deep learning in the medical domain.
- Figure 14 shows that transfer learning started to be applied in the medical domain in 2017 with 2 articles only and reached 218 in 2024.
- Figures 15 shows that the number of publications of active learning, curriculum learning, and federated learning is limited and scattered across the years as they are newly emerged trends.

The selected articles were drawn from top journals in ScienceDirect, adhering to the criteria mentioned above. The references provide a sample of the applications of these new ML techniques in the medical domain, rather than an exhaustive list. For each reference, key details such as the task, disease, technique(s) used, evaluation results, and data type are presented.

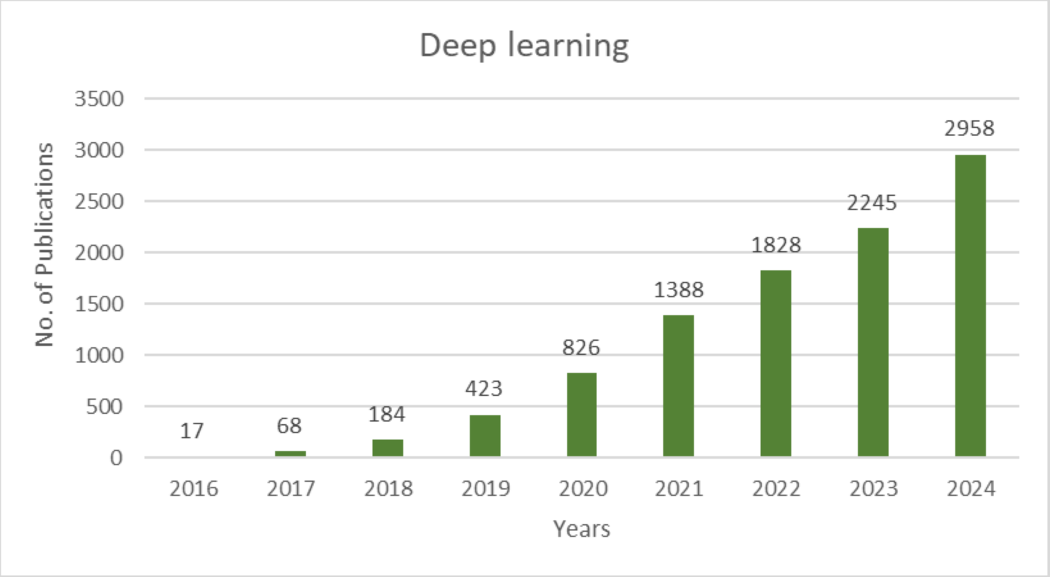


Figure 13: The number of articles published on deep learning from 2016- 2024 in Elsevier database.

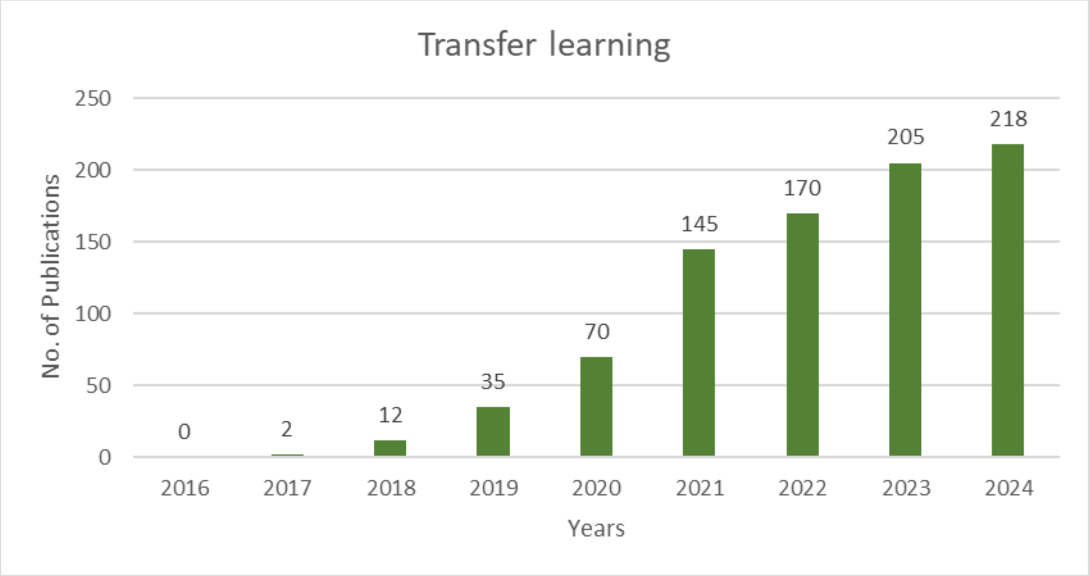


Figure 14: The number of articles published on transfer learning from 2016- 2024 in Elsevier database.

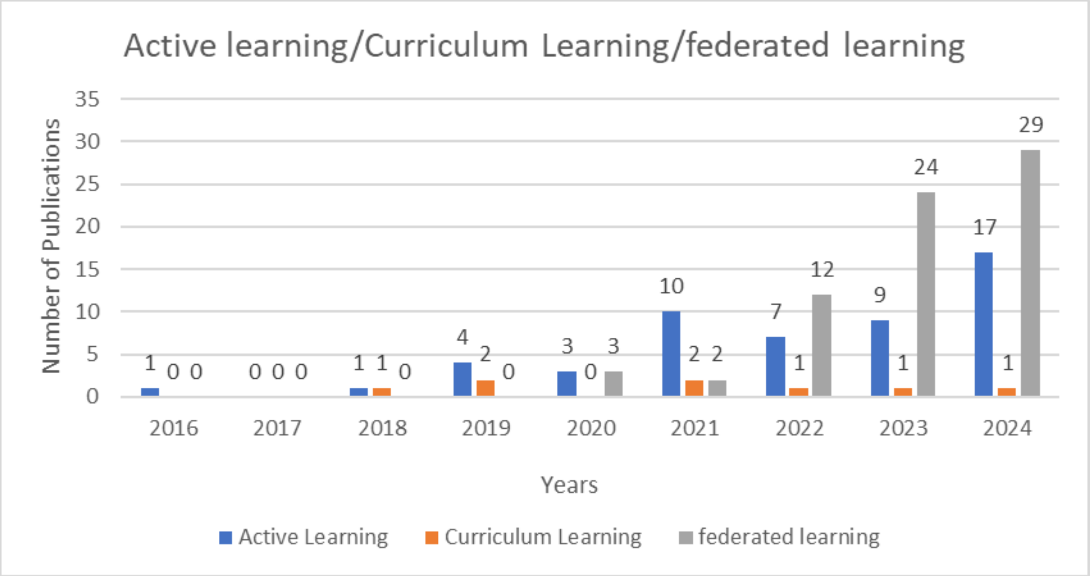


Figure 15: The number of articles published on active/curriculum/Federated learning from 2016- 2024 in Elsevier database.

6 Some applications of new trends of MLTs in the medical domain

This section illustrates the selected articles from the retrieved ones from searching the databases which represent the applications of previously discussed emerging ML trends in the medical domain.

Li, X., et al., [55] proposed a DL model to detect lung nodules. First, segmentation and rib suppression were applied to extract the region of interest and enhance the nodules' visibility. Then, the histogram was applied to enhance the images. After that, patch-based multi-resolution CNN was used for feature extraction, and 4 fusion methods were employed for classification, the best performance method to detect lung nodules achieved an accuracy of more than 99% and FAUC of 0.982 when applied to the Chest X-ray radiographs dataset [56].

El Houby & Yassin [57] developed a CNN model to classify the breast mammographic images' into nonmalignant or malignant. They used 2 methods, the first is based on patches of region of interest (ROI) in the mammogram and the second is based on the whole breast. The accuracy, specificity, sensitivity, and AUC were 95.3%, 92.6%, 98%, and 0.974 respectively using MIAS [20] dataset, while they were 96.52%, 96.49%, 96.55%, and 0.98 using INbreast [58] dataset.

Dai, Y., et al., [59] developed a deep learning CNN model for detecting coronary artery disease utilizing raw heart sound signals. It extracts 206 multidomain features and 126 medical multidomain features. The heart sound signal datasets have been collected from 400 patients from the hospital of Xinjiang Medical University. The model achieved an accuracy of 87.86, sensitivity of 90.67, specificity of 82.38, and AUC of 94.70 using multidomain features. It achieved an accuracy of 85.6, sensitivity of 88.04, specificity of 80.83, and AUC of 92.74 using medical multidomain features.

Alassafi et al., [60] proposed a model that predict the distribution of the COVID-19 outbreak in Saudi Arabia, Malaysia, and Morocco. A DL RNN and LSTM network were developed to predict the number of possible cases of COVID-19. The LSTM achieved an accuracy of 98.58%, while the RNN achieved an accuracy of 93.45%. A comparison was conducted between the number of resulting deaths and the number of coronavirus cases in each of the 3 countries. The model predicted the number of certain COVID-19 cases and deaths for the following 7 days. The model was tested using a public dataset from the European Centre for Disease Prevention and Control [61].

Maiti et al., [62] developed a deep learning (DL)-based framework to automatically detect and segment the optic disc from fundus images for the diagnosis of diabetic retinopathy. The framework utilized an adjusted CNN, experimenting with seven different encoder networks: DenseNet121, InceptionV3, ResNet34, GG11, VGG19, VGG13, and VGG16. VGG16 was selected as the adopted encoder, while the decoder was designed with a harmonic structure based on that of the encoder to improve segmentation performance. The framework was applied to several fundus image datasets, including DIARETDB1, MESSIDOR, IDRiD, DIARETDB0, CHASE-DB1, DRIVE, and STARE. It achieved an impressive accuracy of 99.44%.

Zareen, et al. [63] developed a skin cancer classification deep learning CNN-RNN model with a ResNet-50 for spatial features extraction and LSTM for temporal dependencies. The model has been applied to a dataset of 9000 images of skin lesions representing 9 cancer types. The model achieved an accuracy of 94.48, a sensitivity of 94.38, and a specificity of 93.

Ge, R., et al., [64] Proposed a Dual-Enhanced Convolutional Ensemble Neural Network (DECENN) to detect the presence or absence of metastasis in the whole slide imaging patches of breast cancer. It utilizes VGG16 and DenseNet121 in the network. It was applied to the updated version of a benchmark dataset of microscopic images and histopathologic scans of lymph node sections for the breast [65]. It achieved an accuracy of about 98.92%, an AUC of 99.70%, and a F-score of 98.93%.

Liu, Q., X. She & Q. Xia [66] proposed a model to classify osteosarcoma cells and other cell types using an updated version of CA-MobileNet V3 based on the transfer learning model. It was applied to osteosarcoma cells microscopy imaging of bone cancer dataset [67]. It achieved an accuracy of 98.69 % and f1-score of 94.11.

Oommen & Arunnehru [68] proposed a model to diagnose Alzheimer’s disease in its early stages. The proposed model contains 3 phases: preprocessing the images, extracting features using TL with ResNet-18, which are then compressed by cascaded autoencoders (AE), and finally classifying the disease to one of its 5 stages using DNN. The model was applied to the MRI Neuroimaging dataset [69] and achieved an accuracy of 98.54% recalls of 98.9%, a precision of 98.98%, and an F1 score of 98.82%.

Kumar et al. [70] developed a CNN model using the Resnet152 TL approach with feature extractors to classify the brain tumor images into normal, benign, and malignant. The model has been applied to the Brats MRI image dataset. The proposed transfer Learning model achieved high accuracy reaching 99.57%.

Manickam, et al., [71] proposed a deep TL model for pneumonia detection. The chest X-ray images were preprocessed to recognize the existence of pneumonia based on the U-Net segmentation network, then classify the cases as normal or abnormal (Bacteria, viral) using pre-trained models such as ResNet50, InceptionV3, and Inception ResNetV2. It was evaluated using a publicly available database which includes 5,232 chest x-ray images. ResNet50 model achieved an accuracy of 93.06%, precision of 88.97%, Recall of 96.78%, and F1-score of 92.71%.

Veknugopal, et al., [72] developed a DNN using modified EfficientNetV2-M based on transfer learning to detect skin cancer on dermoscopic images. The model was applied to 58,032 dermoscopic images collected from [73-77]. The model was tested for binary classification tasks and the multiclass classification tasks. It achieved an accuracy reached 97.62 for the multiclass classification of the ISIC 2020 dataset, while it achieved an accuracy of 99.23 for the binary classification of the same dataset.

Mehmood, et al., [78] developed a model to diagnose Alzheimer’s disease (AD) in its early stage based on TL using VGG-19 pre-trained model. The model distinguishes among 4 classes which are AD, late mild cognitive impairment (LMCI), early mild cognitive impairment (EMCI), and normal control (NC). The used dataset was collected from the AD Neuroimaging Initiative (ADNI) [69] database. In the pre-processing phase, the gray matter (GM) tissue was segmented from brain MRI, and then VGG-19 was used to classify the segmented parts. The model achieved an accuracy of 98.73% to distinguish between AD and NC, 83.72% to distinguish between LMCI and EMCI cases, and more than 80% to distinguish between the other combinations of classes.

Al-Shabi, Shak, and Tan, [79] developed a Progressive Growing Channel Attentive Non-Local (ProCAN) deep learning model to classify lung nodules as benign or malignant. Curriculum Learning (CL) was used to train easy samples before hard samples. The model has gradually grown to improve the possibility of classifying the samples based on CL. The model has been applied to samples from 2 publicly available CT scan datasets LIDC-IDRI [80] and LUNGx [81]. It achieved an accuracy of 95.28% and AUC of 98.05%, precision of 95.75, sensitivity of 94.33 and F1-Score of 95.04.

Cho, Y., et al., [82] proposed CL model using a DL CNN to classify chest radiograph (CXR) images into normal and five types of pulmonary abnormalities. The model used ResNet-50 for training on patches of CXR images based on the various patch ratios according to pre-trained weights, with fine-tuning using transfer learning

(TL). The model was applied to CXR from hospitals, including Seoul National University Bundang Hospital (SNUBH) and Asan Medical Center (AMC). It achieved the following accuracies: 90.97% for 20% of the dataset at SNUBH, 91.92% for 50%, and 93.00% for 100%. At AMC, the accuracies were 93.90%, 94.54%, and 95.39%, respectively.

Wong et al., [83] developed a CL-based method for classifying medical images, using features from segmentation networks. The model first learns simpler shapes and features through a segmentation network pre-trained on similar data, then applies this knowledge for more complex classification tasks. The M-Net, a CNN modified from U-Net for working with fewer training samples, was used for segmentation. Then the CNN classifier receives the features from a segmentation network as inputs. The model achieved an accuracy of 82% in a 3D 3-classes brain tumor classification and 86% on a 2D nine-class cardiac semantic level classification problem.

Wu, et al. [84] developed a weakly-supervised deep AL framework to diagnose COVID-19 using CT scans. The framework contains a 2D U-Net for segmentation of the lung region and a hybrid active learning approach, which keeps sample diversity and predicted loss diagnosis of COVID-19. The framework classifies the CT scans into one of three classes which are pneumonia, coronavirus pneumonia caused by SARS-CoV-2, and normal cases. The framework was validated on a CT scan dataset from the China Consortium of Chest CT Image Investigation (CC-CCII) [85]. With only 30% of the labeled data, the accuracy of the framework reached 0.867, while AUC was 0.968.

Wu, X., et al., [86] proposed a hybrid active learning (HAL) framework that combines AL with deep TL using ResNet18. The framework applies data augmentation to the unlabeled data pool and uses a hybrid sampling approach that maintains sample variety and classification loss (data uncertainty). The diversity sampling is based on data augmentation, while the generated data noise is discarded with an outlier detection process. The HAL was validated on 3 medical image datasets which are the Hyper-Kvasir for gastrointestinal disease [87], Messidor for eye fundus images [88], and breast cancer datasets [89]. By applying the proposed framework to the Hyper-Kvasir dataset it achieves an accuracy of 0.871, precision of 0.602, recall of 0.587, and F1-score of 0.594.

Meirelles, et al., [90] used Pool-based AL to train DL models for classifying Tumor Infiltrating Lymphocytes. The proposed approach selects image patches based on feature grouping and prediction uncertainty. They introduced a Diversity-Aware Data Acquisition (DADA) method, which ensures diverse batch selection by clustering images based on features and then choosing uncertain patches from each cluster. The most uncertain patches from each cluster are prioritized for selection, the clusters with the most uncertain patches must participate with more patches, the pool is updated by removing the selected patches. By applying the proposed model to the cancer tissue image dataset [91], it achieved an AUC of 0.78 with fewer tissue patches and execution time.

Zhang, et al., [92] developed a semi-supervised framework for brain segmentation that incorporates quality-driven active learning (QDAL). In the AL module, deep supervision loss and attention mechanism improve the accuracy of segmentation and return quality information for the unlabeled slices. The AL module chooses the most informative slices to be annotated, and the segmentation process is trained iteratively using the updated labeled data. The framework was tested on two brain MRI datasets [93, 94]. The experiment results showed that the segmentation utilizing the QDAL only wants 15–20% annotated slices for the brain extraction task, and 30–40% for tissue segmentation, achieving competitive results with full supervision and an accuracy of 90.7.

Lu, Q., et al.,[95] presented a blood cell classification method called MAE4AL, which combines the self-supervised Masked Autoencoder (MAE) and active learning (AL). It chooses the most remarkable samples for labeling based on self-supervised loss of MAE and sample uncertainty. Tested on blood smear samples obtained from [96], MAE4AL needed labeling only 20% of the data to perform the same as ResNeXt, which was trained on the full dataset. When it trained using half of the labeled data, MAE4AL achieved an accuracy of 96.36%, outperforming ResNeXt which trained on all the data.

Kumbhare et al. [97] developed a FL method for breast cancer diagnosis using mammogram images from the “Curated Breast Imaging Subset of DDSM (CBIS-DDSM)” dataset [98]. The DenseNet pre-trained model was used for feature extraction and the extracted features were classified using Enhanced Recurrent Neural Networks (E-RNN). FL was employed to reduce processing time and improve model performance. The method achieved an accuracy of 95%.

Feki, et al. [53] proposed a decentralized FL framework that permits different medical organizations to screen COVID-19 using Chest X-ray images based on deep learning while keeping the privacy of patient data. Two pre-trained models which are VGG16 and ResNet50 were used for classification. The framework was tested using four clients, where each client has his private dataset and the same CNN models. The proposed FL framework achieved competitive results compared to those models trained by sharing data. The best achieved accuracy was 97% using the ResNet50 model with data augmentation.

Zhang, et al. [99] proposed a FL based on DL framework for the diagnosing brain disorders. The proposed framework was tested on Autism Brain Imaging Data Exchange (ABIDE) [100] dataset. The proposed framework achieved an average accuracy of 79% and reduced the communication burden of FL.

Shaikh, et al. [101] developed an FL-based DL method to classify respiratory diseases by listening to lung sounds. Generative Adversarial Networks created new lung sounds to train a neural network that classifies 4 lung diseases, heart attack and normal breathing patterns. Using two datasets [102, 103], the proposed method achieved an accuracy of 92% for the classification of different respiratory diseases and heart failure.

Table 1 provides a summary of 25 selected articles from top journals on Science Direct, published between 2016 and 2024, based on a database search. These articles showcase applications of recent trends of MLTs in the medical domain and are intended to illustrate these trends, not to present a comprehensive list. For each reference, the table includes the task, disease, techniques used, evaluation results, and data type.

Table 1: Summary of the selected articles from search results for applications of the new ML in medical domain.

Ref.	Task	Disease	Used Technique(s)	Evaluation results	Data Type
[55]	Detection	Chest Lung cancer	DL-CNN	Acc.= 99% FAUC = 0.982	x-ray radiographs
[57]	Classification	Breast cancer	DL-CNN	Acc.=96.52% spec.=96.4% Sen.=96.5% AUC =0.98	mammogram
[59]	Detecting	Coronary artery disease	DL-CNN	Acc.=87.86, % Sen.=90.67% Spec.=82.38 % AUC = 94.70	heart sound signals
[60]	Prediction	COVID-19	DL-RNN LSTM	Acc. = 93.45% Acc.= 98.58%	Numerical
[62]	Segmentation diagnosis	Diabetic retinopathy	DL-CNN	Acc.= 99.44%	fundus images
[63]	Classification	Skin cancer	DL-CNN-RNN	Acc.=94.48, Sen.= 94.38 spec. = 93	skin lesion images
[64]	Detection	Breast cancer	DL-TL-VGG16-DenseNet121	Acc.=98.92%, AUC =99.70%, F-score = 98.93%	histopathologic image of lymph node
[66]	Classification	Bone cancer	TL CA-MobileNetV3	Acc.=98.69 % f1-score= 94.11%	microscopic images of bone cancer
[68]	Classification	Alzheimer's disease	TL- ResNet-18- AE-DL	Acc. = 98.54% recalls = 98.9% prec.=98.98% F1-score=98.82%	MRI Neuroimaging dataset
[70]	Classification	Brain tumor	TL-Resnet152-CNN	Acc. =99.57%	MRI
[71]	Segmentation Detection	Pneumonia	U-Net TL- ResNet50	Acc.=93.06% prec.=88.97% Rec.=96.78% F1score=92.7	chest X-ray
[72]	Classification	Skin cancer	TL-EfficientNetV2-M	Acc. = 99.23	dermoscopic images
[78]	Classification	Alzheimer	TL-VGG19	Acc.=98.73%	MRI
[79]	Classification	lung nodules	DL-CNN-CL	Acc. = 95.28% AUC = 98.05% Precision = 95.75 Sen = 94.33 F1-Score = 95.04	CT scans
[82]	Classification	pulmonary abnormalities	TL-ResNet-50 -CL	Acc. =93.90, 94.54, 95.39 For 20%, 50%, 100% of dataset	CXR
[83]	Segmentation Classification	brain tumor cardiac	TL-M-Net DL-CNN-CL	Acc.= 82% Acc.= 86%	MR

[84]	Segmentation Classification	COVID-19	TL-U-Net DL-AL	Acc=0.866 ROC= 0.968	CT scans
[86]	Classification	gastrointestinal disease	TL-ResNet18 AL	Acc. =0.871 Prec. =0.602 Recall=0.587 F1score=0.594	Images
[90]	Classification	Tumor Infiltrating Lymphocytes	DL_CNN_AL	AUC = 0.78	histology image
[92]	Segmentation	Brain	DL_CNN_AL	Acc.= 90.7	MRI
[95]	Classification	blood diseases (Leukemia)	MaskedAutoencoder (MAE4AL)	Acc.= 96.36%	blood smear samples
[97]	Classification	Breast Cancer	FL-TL-DenseNet-RNN	Acc.=95%	mammogram
[53]	Classification	COVID-19	FL-TL-VGG16/ ResNet50	Acc. = 97%	X-ray images
[99]	Classification	Brain disorders	FL-CNN	Acc. = 79%	Autism Brain Imaging
[101]	Classification	respiratory diseases& heart failure	FL-DL	Acc. = 92%	breathing sounds

7 Conclusion and future work

This research explored the emerging trends in machine learning techniques (MLTs) within the medical domain. Through a comprehensive literature review, we found that deep learning has become the dominant trend, holding significant promise for developing intelligent medical applications. A key advantage of deep learning is its ability to perform automatic feature engineering, simplifying the model-building process and reducing reliance on manual input. Current research predominantly addresses diagnostic tasks, with disease classification being the most common approach. Other tasks, such as segmentation, are also explored. Cancer, in its various forms, is the most frequently studied condition, while the COVID-19 pandemic has notably led to a surge in research on lung diseases.

In the realm of medical imaging, traditional machine learning approaches require extensive pre-processing, including feature extraction and selection. Deep learning, particularly Convolutional Neural Networks (CNNs), has advanced the field by automating feature engineering, reducing the need for manual intervention. However, this comes with an increased demand for large datasets and significant computational resources. To address these challenges, recent trends like transfer learning, curriculum learning, active learning, and federated learning have been introduced to enhance model performance, expedite the training process, and improve data security. In summary, the overarching goal in this field is to automate processes, reduce human intervention, and maximize the value derived from limited labeled data, thereby enhancing medical decision-making and patient outcomes.

Looking ahead, there are several key areas where further work is needed. While the number of publications on deep learning in the medical domain has steadily increased since its initial applications in 2016, and although these applications have yielded promising results, further research is essential to address several key challenges. Areas such as active learning, curriculum learning, and federated learning have shown promise but remain under-explored and require more attention in future research. A critical direction for future is to focus on reducing the time and computational costs associated with deep learning models and other trends. These processes often consume substantial energy, indirectly contributing to environmental and climate concerns. Therefore, developing more energy-efficient techniques will be crucial. Additionally, data augmentation, a significant pre-processing step in deep learning, could be integrated more effectively into the model-building process itself, thereby enhancing sample diversity and improving class representation with less manual effort.

Another important aspect for future research is the development of standardized, public databases that include diverse patient data, such as DNA sequences. These databases would enable more comprehensive studies and improve the accuracy of predictive models by providing a richer set of input data. Additionally, integrating knowledge from multiple domains could further enhance the performance of deep learning models in different medical applications. Despite the progress made, the real challenge lies in translating these advancements into practical, real-world applications that can be implemented in clinical settings. Bridging the gap between theoretical research and clinical deployment will be vital to realizing the full potential of deep learning in medicine.

Conflicts of interest

The author has no competing interests to declare.

References

[1] Chen, M., et al., *Disease prediction by machine learning over big data from healthcare communities.* Ieee Access, 2017. **5**: p. 8869-8879.

[2] Grossman, R.L., et al., *Toward a shared vision for cancer genomic data.* New England Journal of Medicine, 2016. **375**(12): p. 1109-1112.

[3] Schaekermann, M., et al., *Understanding expert disagreement in medical data analysis through structured adjudication.* Proceedings of the ACM on Human-Computer Interaction, 2019. **3**(CSCW): p. 1-23.

[4] Garg, A. and V. Mago, *Role of machine learning in medical research: A survey*. Computer Science Review, 2021. **40**: p. 100370.

[5] Dallora, A.L., et al., *Machine learning and microsimulation techniques on the prognosis of dementia: A systematic literature review.* PloS one, 2017. **12**(6): p. e0179804.

[6] Kharazmi, P., et al., *A computer-aided decision support system for detection and localization of cutaneous vasculature in dermoscopy images via deep feature learning.* Journal of medical systems, 2018. **42**(2): p. 1-11.

[7] Xu, J., K. Xue, and K. Zhang, *Current status and future trends of clinical diagnoses via image-based deep learning.* Theranostics, 2019. **9**(25): p. 7556.

[8] Xie, X., et al., *A survey on incorporating domain knowledge into deep learning for medical image analysis.* Medical Image Analysis, 2021. **69**: p. 101985.

[9] Lee, C.H. and H.-J. Yoon, *Medical big data: promise and challenges.* Kidney research and clinical practice, 2017. **36**(1): p. 3.

[10] Dinov, I.D., *Methodological challenges and analytic opportunities for modeling and interpreting Big Healthcare Data.* Gigascience, 2016. **5**(1): p. s13742-016-0117-6.

[11] http://archive.ics.uci.edu/ml/datasets/.

[12] Available from: http://image-net.org/challenges/LSVRC/.

[13] https://file.biolab.si/biolab/supp/bicancer/projections/.

[14] Corsetti, V., et al., *Evidence of the effect of adjunct ultrasound screening in women with mammography-negative dense breasts: interval breast cancers at 1 year follow-up.* European journal of cancer, 2011. **47**(7): p. 1021-1026.

[15] Zhang, Z. and E. Sejdić, *Radiological images and machine learning: trends, perspectives, and prospects.* Computers in biology and medicine, 2019. **108**: p. 354-370.

[16] Saslow, D., et al., *American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography.* CA: a cancer journal for clinicians, 2007. **57**(2): p. 75-89.

[17] http://medicaldictionary.thefreedictionary.com/operating+microscope.

[18] Jones, N.C. and P.A. Pevzner, *An introduction to bioinformatics algorithms*. 2004: MIT press.

[19] Rahman, T., et al., *COVID-19 radiography database.* https://www.kaggle.com/tawsifurrahman/covid19-radiography-database.

[20] Suckling, J., et al., *Mammographic image analysis society (mias) database v1. 21.* 2015.

[21] Labati, R.D., V. Piuri, and F. Scotti. *All-IDB: The acute lymphoblastic leukemia image database for image processing*. in *2011 18th IEEE international conference on image processing*. 2011. IEEE.

[22] Zoph, B., et al. *Learning transferable architectures for scalable image recognition*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.

[23] Thusberg, J. and M. Vihinen, *Pathogenic or not? And if so, then how? Studying the effects of missense mutations using bioinformatics methods.* Human mutation, 2009. **30**(5): p. 703-714.

[24] El Houby, E.M., *Machine learning techniques for pathogenicity prediction of non-synonymous single nucleotide polymorphisms in human body.* Journal of Ambient Intelligence and Humanized Computing, 2023. **14**(7): p. 8099-8113.

[25] Han, J., J. Pei, and M. Kamber, *Data mining: concepts and techniques*. 2011: Elsevier.

[26] Hamla, H. and K. Ghanem, *A hybrid feature selection based on fisher score and SVM-RFE for microarray data.* Informatica, 2024. **48**(1).

[27] Fahrudin, T.M., I. Syarif, and A.R. Barakbah. *Ant colony algorithm for feature selection on microarray datasets*. in *2016 International Electronics Symposium (IES)*. 2016. IEEE.

[28] Talavera, L. *An evaluation of filter and wrapper methods for feature selection in categorical clustering*. in *International Symposium on Intelligent Data Analysis*. 2005. Springer.

[29] Tabakhi, S. and P. Moradi, *Relevance–redundancy feature selection based on ant colony optimization.* Pattern recognition, 2015. **48**(9): p. 2798-2811.

[30] Yang, X.-S., *Firefly algorithm, stochastic test functions and design optimisation.* International journal of bio-inspired computation, 2010. **2**(2): p. 78-84.

[31] Mashhour, E.M., et al., *Feature Selection Approach based on Firefly Algorithm and Chi-square.* International Journal of Electrical & Computer Engineering (2088-8708), 2018. **8**(4).

[32] Neagoe, V.-E. and E.-C. Neghina. *Feature selection with ant colony optimization and its*

applications for pattern recognition in space imagery. in *2016 international conference on communications (COMM)*. 2016. IEEE.

[33] El Houby, E.M., N.I. Yassin, and S. Omran, *A hybrid approach from ant colony optimization and K-nearest neighbor for classifying datasets using selected features.* Informatica, 2017. **41**(4).

[34] LeCun, Y., Y. Bengio, and G. Hinton, *Deep learning.* nature, 2015. **521**(7553): p. 436-444.

[35] Goodfellow, I., Y. Bengio, and A. Courville, *Deep learning*. 2016: MIT press.

[36] Dezaki, F.T., et al., *Cardiac phase detection in echocardiograms with densely gated recurrent neural networks and global extrema loss.* IEEE transactions on medical imaging, 2018. **38**(8): p. 1821-1832.

[37] Raiaan, M.A.K., et al., *A systematic review of hyperparameter optimization techniques in Convolutional Neural Networks.* Decision Analytics Journal, 2024: p. 100470.

[38] Ali, M.J., et al., *A review of AutoML optimization techniques for medical image applications.* Computerized Medical Imaging and Graphics, 2024: p. 102441.

[39] Ambekar, S. and R. Phalnikar. *Disease risk prediction by using convolutional neural network*. in *2018 Fourth international conference on computing communication control and automation (ICCUBEA)*. 2018. IEEE.

[40] Anwar, S.M., et al., *Medical image analysis using convolutional neural networks: a review.* Journal of medical systems, 2018. **42**(11): p. 1-13.

[41] Oraibi, Z.A. and S. Albasri, *A robust end-to-end cnn architecture for efficient covid-19 prediction form x-ray images with imbalanced data.* Informatica, 2023. **47**(7).

[42] Krizhevsky, A., I. Sutskever, and G.E. Hinton, *Imagenet classification with deep convolutional neural networks.* Advances in neural information processing systems, 2012. **25**.

[43] Simonyan, K. and A. Zisserman, *Very deep convolutional networks for large-scale image recognition.* arXiv preprint arXiv:1409.1556, 2014.

[44] Szegedy, C., et al. *Rethinking the inception architecture for computer vision*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

[45] He, K., et al. *Deep residual learning for image recognition*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

[46] Deng, J., et al. *Imagenet: A large-scale hierarchical image database*. in *2009 IEEE conference on computer vision and pattern recognition*. 2009. Ieee.

[47] Russakovsky, O., et al., *Imagenet large scale visual recognition challenge.* International journal of computer vision, 2015. **115**(3): p. 211-252.

[48] Oquab, M., et al. *Learning and transferring mid-level image representations using convolutional neural networks*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014.

[49] Wang, L., et al., *Trends in the application of deep learning networks in medical image analysis: Evolution between 2012 and 2020.* European Journal of Radiology, 2022. **146**: p. 110069.

[50] Jiménez-Sánchez, A., et al. *Medical-based deep curriculum learning for improved fracture classification*. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2019. Springer.

[51] Budd, S., E.C. Robinson, and B. Kainz, *A survey on active learning and human-in-the-loop deep learning for medical image analysis.* Medical Image Analysis, 2021. **71**: p. 102062.

[52] KhoKhar, F.A., et al., *A review on federated learning towards image processing.* Computers and Electrical Engineering, 2022. **99**: p. 107818.

[53] Feki, I., et al., *Federated learning for COVID-19 screening from Chest X-ray images.* Applied Soft Computing, 2021. **106**: p. 107330.

[54] Wu, J.C.-H., et al., *Dynamically Synthetic Images for Federated Learning of Medical Images.* Computer Methods and Programs in Biomedicine, 2023: p. 107845.

[55] Li, X., et al., *Multi-resolution convolutional networks for chest X-ray radiograph based lung nodule detection.* Artificial intelligence in medicine, 2020. **103**: p. 101744.

[56] Li, X., et al. *Rib suppression in chest radiographs for lung nodule enhancement*. in *2015 IEEE International Conference on Information and Automation*. 2015. IEEE.

[57] El Houby, E.M. and N.I. Yassin, *Malignant and nonmalignant classification of breast lesions in mammograms using convolutional neural networks.* Biomedical Signal Processing and Control, 2021. **70**: p. 102954.

[58] I. C. Moreira, I.A., I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, *Inbreast: toward a full-field digital mammographic database.* Academic radiology, 2012. **19**(2): p. 236-248.

[59] Dai, Y., et al., *Deep learning fusion framework for automated coronary artery disease detection using raw heart sound signals.* Heliyon, 2024. **10**(16).

[60] Alassafi, M.O., M. Jarrah, and R. Alotaibi, *Time series predicting of COVID-19 based on deep learning.* Neurocomputing, 2022. **468**: p. 335-344.

[61] https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic- and distribution-covid-19-cases-worldwide.

[62] Maiti, S., et al., *Automatic detection and segmentation of optic disc using a modified convolution network.* Biomedical Signal Processing and Control, 2022. **76**: p. 103633.

[63] Zareen, S.S., et al., *Enhancing Skin Cancer Diagnosis with Deep Learning: A Hybrid CNN-RNN Approach.* Computers, Materials & Continua, 2024. **79**(1).

[64] Ge, R., et al., *Detection of presence or absence of metastasis in WSI patches of breast cancer using the dual-enhanced convolutional ensemble neural network.* Machine Learning with Applications, 2024. **17**: p. 100579.

[65] Cukierski, W., *Histopathologic cancer detection.* Kaggle. https://kaggle.com/competitions/histopathologic-cancer-detection, 2018.

[66] Liu, Q., X. She, and Q. Xia, *AI based diagnostics product design for osteosarcoma cells microscopy imaging of bone cancer patients using CA-MobileNet V3.* Journal of Bone Oncology, 2024: p. 100644.

[67] Charilaou, P. and R. Battat, *Machine learning models and over-fitting considerations.* World Journal of Gastroenterology, 2022. **28**(5): p. 605.

[68] Oommen, D.K. and J. Arunnehru, *Alzheimer's Disease Stage Classification Using a Deep Transfer Learning and Sparse Auto Encoder Method.* Computers, Materials & Continua, 2023. **76**(1).

[69] http://adni.loni.usc.edu.

[70] Kumar, K.A., A. Prasad, and J. Metan, *A hybrid deep CNN-Cov-19-Res-Net Transfer learning architype for an enhanced Brain tumor Detection and Classification scheme in medical image processing.* Biomedical Signal Processing and Control, 2022. **76**: p. 103631.

[71] Manickam, A., et al., *Automated pneumonia detection on chest X-ray images: A deep learning approach with different optimizers and transfer learning architectures.* Measurement, 2021. **184**: p. 109953.

[72] Venugopal, V., et al., *A deep neural network using modified EfficientNet for skin cancer detection in dermoscopic images.* Decision Analytics Journal, 2023. **8**: p. 100278.

[73] Rotemberg, V., et al., *A patient-centric dataset of images and metadata for identifying melanomas using clinical context.* Scientific data, 2021. **8**(1): p. 34.

[74] Tschandl, P., C. Rosendahl, and H. Kittler, *The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions.* Scientific data, 2018. **5**(1): p. 1-9.

[75] Codella, N.C., et al. *Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic).* in *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*. 2018. IEEE.

[76] Combalia, M., et al., *Bcn20000: Dermoscopic lesions in the wild.* arXiv preprint arXiv:1908.02288, 2019.

[77] Codella, N., et al., *Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic).* arXiv preprint arXiv:1902.03368, 2019.

[78] Mehmood, A., et al., *A transfer learning approach for early diagnosis of Alzheimer's disease on MRI images.* Neuroscience, 2021. **460**: p. 43-52.

[79] Al-Shabi, M., K. Shak, and M. Tan, *ProCAN: Progressive growing channel attentive non-local network for lung nodule classification.* Pattern Recognition, 2022. **122**: p. 108309.

[80] Armato III, S.G., et al., *The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans.* Medical physics, 2011. **38**(2): p. 915-931.

[81] Armato III, S.G., et al., *LUNGx Challenge for computerized lung nodule classification.* Journal of Medical Imaging, 2016. **3**(4): p. 044506-044506.

[82] Cho, Y., et al., *Optimal number of strong labels for curriculum learning with convolutional neural network to classify pulmonary abnormalities in chest radiographs.* Computers in Biology and Medicine, 2021. **136**: p. 104750.

[83] Wong, K.C., T. Syeda-Mahmood, and M. Moradi, *Building medical image classifiers with very limited data using segmentation networks.* Medical image analysis, 2018. **49**: p. 105-116.

[84] Wu, X., et al., *COVID-AL: The diagnosis of COVID-19 with deep active learning.* Medical Image Analysis, 2021. **68**: p. 101913.

[85] Zhang, K., et al., *Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography.* Cell, 2020. **181**(6): p. 1423-1433. e11.

[86] Wu, X., et al., *HAL: Hybrid active learning for efficient labeling in medical domain.* Neurocomputing, 2021. **456**: p. 563-572.

[87] Borgli, H., et al., *HyperKvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy.* Scientific data, 2020. **7**(1): p. 1-14.

[88] Decencière, E., et al., *Feedback on a publicly distributed image database: the Messidor database.* Image Analysis & Stereology, 2014. **33**(3): p. 231-234.

[89] Aresta, G., et al., *Bach: Grand challenge on breast cancer histology images.* Medical image analysis, 2019. **56**: p. 122-139.

[90] Meirelles, A.L., et al., *Effective Active Learning in Digital Pathology: A Case Study in Tumor Infiltrating Lymphocytes.* Computer Methods and Programs in Biomedicine, 2022: p. 106828.

[91] Saltz, J., et al., *Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images.* Cell reports, 2018. **23**(1): p. 181-193. e7.

[92] Zhang, Z., et al., *Quality-driven deep active learning method for 3D brain MRI segmentation.* Neurocomputing, 2021. **446**: p. 106-117.

[93] Shattuck, D.W., et al., *Construction of a 3D probabilistic atlas of human cortical structures.* Neuroimage, 2008. **39**(3): p. 1064-1080.

[94] https://www.nitrc.org/projects/ibsr.

[95] Lu, Q., et al., *A blood cell classification method based on MAE and active learning.* Biomedical Signal Processing and Control, 2024. **90**: p. 105813.

[96] Matek, C., et al., *Human-level recognition of blast cells in acute myeloid leukaemia with convolutional neural networks.* Nature Machine Intelligence, 2019. **1**(11): p. 538-544.

[97] Kumbhare, S., A.B. Kathole, and S. Shinde, *Federated learning aided breast cancer detection with intelligent Heuristic-based deep learning framework.* Biomedical Signal Processing and Control, 2023. **86**: p. 105080.

[98] Lee, R.S., et al., *A curated mammography data set for use in computer-aided detection and diagnosis research.* Scientific data, 2017. **4**(1): p. 1-9.

[99] Zhang, C., et al., *FedBrain: A robust multi-site brain network analysis framework based on federated learning for brain disease diagnosis.* Neurocomputing, 2023. **559**: p. 126791.

[100] Heinsfeld, A.S., et al., *Identification of autism spectrum disorder using deep learning and the ABIDE dataset.* NeuroImage: Clinical, 2018. **17**: p. 16-23.

[101] Shaikh, A.A.S. and M. Bhargavi, *Weighted aggregation through probability based ranking: An optimized federated learning architecture to classify respiratory diseases.* Computer Methods and Programs in Biomedicine, 2023. **242**: p. 107821.

[102] Fraiwan, L., et al., *Automatic identification of respiratory diseases from stethoscopic lung sound signals using ensemble classifiers.* Biocybernetics and Biomedical Engineering, 2021. **41**(1): p. 1-14.

[103] Rocha, B., et al. *A respiratory sound database for the development of automated classification*. in *Precision Medicine Powered by pHealth and Connected Health: ICBHI 2017, Thessaloniki, Greece, 18-21 November 2017*. 2018. Springer.