



Review of Machine Learning Algorithm in Medical Health

Zahra Ghorbani¹, Sahar Behrouzi-Moghaddam¹, Shahram Zandiyan², Babak Nouri-Moghaddam¹, Nasser Mikaeilvand³, Sajjad Jahanbakhsh⁴, Ailin Khosravani¹, Fatemeh Tahmasebizadeh¹, Abbas Mirzaei^{1*}

1 Department of Computer Engineering, Ard.C., Islamic Azad University, Ardabil, Iran

2 Department of Computer Science and Mathematics, Central Tehran branch, CT.C., Islamic Azad University, Tehran, Iran

3 Department of Computer Engineering, ST.C., Islamic Azad University, Tehran, Iran

4 Department of Computer Engineering, Germi.C., Islamic Azad University, Germi, Iran

Article Info

Article History:

Received: 2025/04/13

Revised: 2025/06/10

Accepted: 2025/06/20

DOI:

Keywords:

Machine Learning; Medical;

Supervised Learning;

Classification.

*Corresponding Author's Email
Address: a.mirzaei@iau.ac.ir

Abstract

Recently, health-related data has been analyzed using a variety of cutting-edge methods, including artificial intelligence and machine learning. The application of machine learning technologies in the healthcare industry is enhancing medical professionals' proficiency in diagnosis and treatment. Researchers have extensively used medical data to identify patterns and diagnose illnesses. Nevertheless, little research has been done on using machine learning algorithms to enhance the precision and usefulness of medical data. An extensive analysis of the many machine learning methods applied to healthcare applications is given in this work. We first examine supervised and unsupervised machine learning techniques, and then we investigate the applicability of time series tasks on historical data, evaluating their appropriateness for datasets of varying sizes.

1. Introduction

The healthcare service system plays a crucial role in the medical domain, addressing significant demands on human life. To advance, healthcare providers in developing countries are increasingly adopting intelligent technologies such as artificial intelligence (AI) and machine learning. The integration of AI has spurred advancements in human-centered healthcare systems. AI technologies have notably influenced the development of intensive care and supervisory activities in hospitals and clinics [1-3].

Extensive research by Jafar Abdollahi since 2019 has highlighted the successful application of AI, including machine learning and deep learning, in medical image and healthcare analysis. His research covers a range of conditions such as bupropion, diphenhydramine, breast cancer, medicinal plants, epidemics, stroke, lung cancer, social networks, diabetes, suicides, coronary artery

disease, and more, demonstrating promising results [2-5].

Machine learning, an automated process that enables computers to learn and improve performance without explicit programming, is central to these advancements. Unlike systems reliant on preset rules, machine learning utilizes complex algorithms and statistical techniques to analyze data and make accurate predictions. The dataset's quality is critical for machine learning accuracy, leading to more precise forecasts as the data improves [3,18].

Machine learning has found applications across various industries, including banking, retail, and healthcare. In healthcare, it offers significant opportunities for disease detection and treatment. One of its key benefits is enhancing the accuracy of data forecasting and classification, which is particularly valuable in medical analysis. As more data is collected, the prediction model's ability to make precise decisions improves.

Overall, the healthcare service system is vital in addressing human needs, and advanced technologies like AI and machine learning are instrumental in advancing and refining healthcare services. The integration of AI has led to significant developments in human-centered healthcare systems. Since 2019, Jafar Abdollahi's research has demonstrated the successful application of AI in diagnosing various diseases and analyzing medical images, yielding encouraging results across multiple conditions.

2. OVERVIEW OF MACHINE-LEARNING IN HEALTHCARE

Machine learning, a branch of artificial intelligence, focuses on using data to train algorithms so they can act or anticipate without explicit programming. Machine learning has the ability to completely change how the healthcare sector recognizes, treats, and prevents diseases, as seen in Figure 1. The following are a few possible uses of machine learning in the medical field [4,19]:

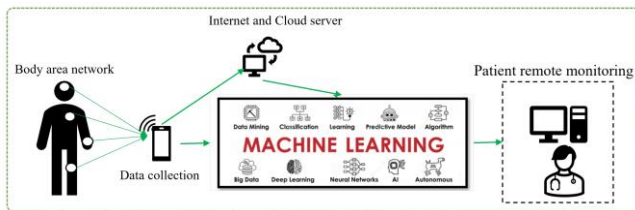


Fig 1 illustrates, machine learning has the power to radically alter how we identify, manage, and prevent diseases in the healthcare industry [19].

- A. *Predictive analytics:* Utilizing information from claims data, electronic health records, and other sources, machine learning algorithms can forecast the probability of certain health outcomes, like hospital readmissions or the onset of chronic illnesses. This capacity permits early preventative treatments and enables medical practitioners to identify individuals who are more likely to have negative effects [6,20].
- B. *Diagnosis and treatment:* Diagnoses and the best course of treatment for a patient can be made with the assistance of machine learning algorithms that have been trained on medical images, such as CT scans and X-rays [5,21].
- C. *Personalized medicine:* Using unique patient variables like genetics and medical history, machine learning may be used to forecast which drugs a patient is most likely to react to [7, 22].
- D. *Clinical decision support:* Clinical decision support systems may incorporate machine

learning algorithms to help medical personnel make better decisions about patient care [8, 23].

- E. *Population health management:* Data from huge populations may be analyzed using machine learning to find trends and patterns that can guide the creation of public health programs. All things considered, using machine learning to healthcare might lead to better patient outcomes, lower costs, and increased system efficiency [9, 24].

3. REVIEW OF MACHINE LEARNING

The two main subcategories of machine learning are supervised learning and unsupervised learning, as seen in Figure 2. In order to forecast future results, supervised learning algorithms are trained with input and output data from previous occurrences. Unsupervised learning algorithms, on the other hand, find underlying structures or hidden patterns in the given data without the need for pre-existing labels. Unsupervised learning mostly concentrates on clustering tasks, whereas supervised learning is appropriate for both classification and regression problems [10- 13].

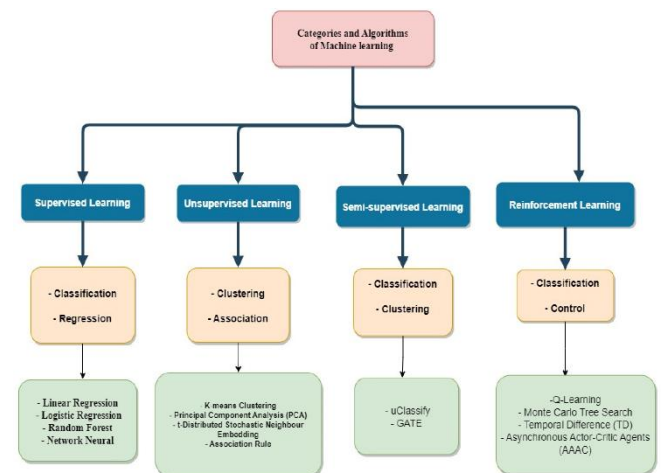


Fig 2. Supervised learning and unsupervised learning are the two primary subcategories of machine learning [13].

Classification algorithms, which predict categorical outcomes, are a subset of supervised machine learning approaches. Unlike unsupervised learning, supervised learning relies on known and labeled training data. The data is divided into training and testing sets [14-17]. Classification algorithms sort incoming data into distinct categories to make predictions. Supervised machine learning is commonly applied in fields such as speech recognition, medical image interpretation, and heart attack prediction [7, 18].

Using the supplied training data, categorization models are created in supervised learning. Then, more unlabeled data can be classified by these models. One output variable from the training dataset needs to be categorized. In order to categorize the test data, classification algorithms first recognize unique patterns in the training data [13]. Neural networks, decision trees, naïve Bayes, K-nearest neighbors, and support vector machines are examples of common classification techniques.

A. Supervised Machine Learning

• Decision Trees(DT)

A decision tree classifier uses a tree-like diagram to illustrate possible outcomes, final values, and options. This method involves computing the probabilities of selecting different actions through a computer algorithm. The process begins with samples of training data and their associated category labels. The decision tree method recursively partitions the training data into subsets based on feature values, resulting in subgroups with more homogeneous data compared to the parent set [19, 25].

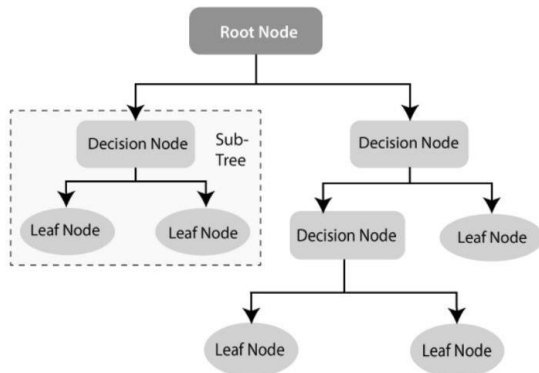


Fig 3. Visual illustration of the DT algorithm [13]

In a decision tree, every internal node denotes a test feature, every branch node shows the test's outcome, and every leaf node shows the class label. The decision tree classifier classifies an unknown sample using the route from a root node to a leaf node, and it utilizes this path to derive the category label [15–17].

• Support Vector Machine (SVM)

The Support Vector Machine (SVM) is a classical machine learning method used for addressing classification problems. SVM plays a vital role in supporting a wide range of applications in extensive data mining environments [30]. It leverages specific characteristics of a model to train data and generate accurate predictions from a given dataset [20–22].

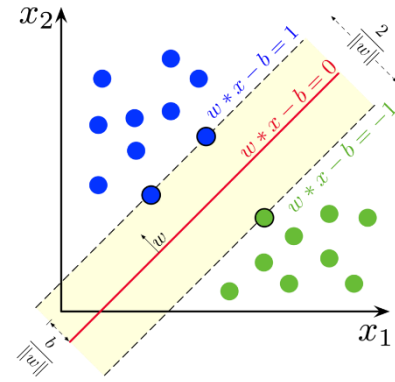


Fig 4. Visual illustration of the SVM algorithm [22].

Support Vector Machine's mathematical intuition: Think of a binary classification task where there are two classes, denoted by the labels +1 and -1. The input feature vectors (X) and the matching class labels (Y) comprise our training dataset. The equation for the linear hyperplane can be written as:

$$w^T x + b = 0 \quad (1)$$

The direction perpendicular to the hyperplane, or the normal vector, is represented by the vector W . The offset, or distance, of the hyperplane from the origin along the normal vector w is represented by the parameter b in the equation. The distance between a data point x_i and the decision boundary can be calculated as:

$$d_i = \frac{w^T x_i + b}{\|w\|} \quad (2)$$

where $\|w\|$ represents the Euclidean norm of the weight vector w . Euclidean norm of the normal vector W .

• Naïve Bayes (NB)

An approach that is frequently used for classification jobs is the Naïve Bayes algorithm. It is one of the most basic types of Bayesian networks since it is predicated on the idea that there is a single parent node with a finite number of independent child nodes. As shown in Figure 6, the Naïve Bayes technique multiplies the individual probabilities of each attribute-value combination to determine the probability of a classification. This approach works incredibly well in cases where the qualities are independent. The Naïve Bayes method's effective computational training time is one of its main benefits. The classification performance of the algorithm can also be improved by eliminating unnecessary characteristics [23-26].

$$P(y = 1 | x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{n-1} x_{n-1})}} \quad (8)$$

- $P(c | x)$ = Posterior probability the probability of class C given the features X .
- $P(X | C)$ = Likelihood the probability of the features X given the class C .

- $P(c)$ = Class Prior Probability the probability of the class C occurring.
- $P(x)$ = Predictor Prior Probability the probability of the features X occurring.

- **K-Nearest Neighbours (K-NN)**

In data mining classification technology, the K-nearest neighbors (K-NN) classification technique is a straightforward and intuitive method. The K-NN algorithm operates on the principle that an unknown pattern can be classified by considering the K nearest neighbors. By specifying a value for K, the algorithm identifies the category based on the majority class of the K training samples most similar to the unknown pattern. Factors such as the chosen K-value and the distance metric play crucial roles in the performance of the classifier [27].

Euclidean=

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2} \quad (4)$$

$$\text{Manhattan} = \sum_{i=1}^k |X_i - y_i| \quad (5)$$

$$\text{Minkowski} = \left\{ \sum_{i=1}^k (|X_i - y_i|) \right\}^{1/q} \quad (6)$$

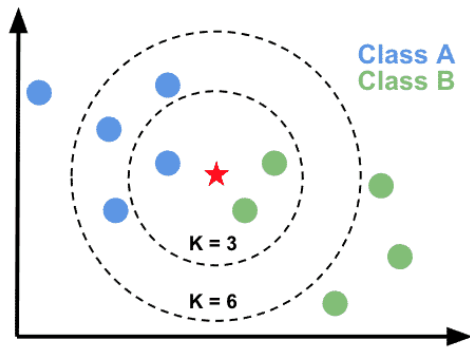


Fig 5. Visual illustration of the KNN algorithm [27].

One advantage of the K-NN method is its relatively low training time compared to other machine learning algorithms. However, it may require more computational time during classification. Despite this, K-NN is favored for its simplicity and ease of use in classification tasks. It is particularly effective when dealing with datasets that have multiple class labels. Additionally, the data training phase of K-NN tends to be faster than that of other machine learning algorithms [27, 28].

- **Linear Regression (LR)**

Linear regression is a straightforward and commonly used method for quantifying the relationship

between response variables and continuous predictors. Its simplicity makes it an optimal choice for analyzing small datasets with high accuracy, as it is relatively easy to understand and interpret. However, if there is an excessive number of predictor variables, the model may struggle to produce reliable results and might not provide the desired outcome [29-31].

$$y = \beta_0 + \beta_1 x + \epsilon \quad (7)$$

Where:

- The dependent variable, often known as the target or outcome variable, is Y .
- The independent variable, often known as the predictor or feature, is x .
- The value of y when $x=0$ is the regression line's intercept, or β_0 .
- The regression line's slope, or the change in y for every unit change in x , is β_1 .
- The error term, denoted by ϵ , is the discrepancy between the observed and model-predicted values.

- **Logistic Regression (LR)**

Unlike linear regression, which predicts continuous data, logistic regression is primarily used for predicting discrete class labels. In classification problems, logistic regression estimates the probability of a sample belonging to one of two possible categories. This is achieved by applying a logistic function, which maps the predicted values to a binary outcome of either 0 or 1. Consequently, logistic regression can indicate the category to which a sample belongs based on the output variable. Researchers have utilized logistic regression to predict health-related behaviors [32-35].

$$P(y = 1 | x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}} \quad (8)$$

Where:

- $P(y=1|x)$ is the probability that the dependent variable y is 1 given the independent variables x_1, x_2, \dots, x_n .
- β_0 is the intercept (the bias term).
- $\beta_1, \beta_2, \dots, \beta_n$ are the coefficients for the independent variables x_1, x_2, \dots, x_n .
- e is the base of the natural logarithm (approximately equal to 2.718).

- **Ensemble Methods**

Ensemble methods leverage the strengths of multiple machine learning algorithms rather than relying on a single algorithm. By combining and integrating various models, ensemble approaches enhance the overall learning process. One key advantage of ensemble methods is their ability to achieve high

predictive accuracy, which can be superior to that of individual models. However, this increased accuracy often comes at the cost of a more complex training process, which can impact efficiency [36, 37].

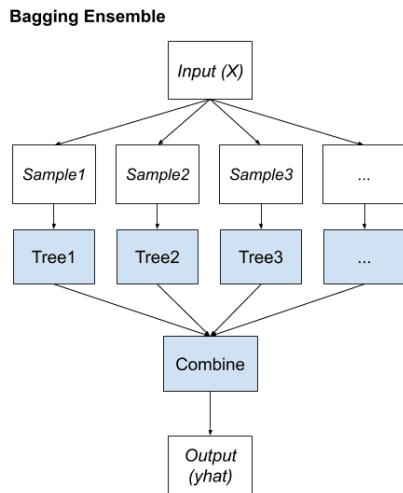


Fig 6. Visual illustration of the Ensemble algorithm [36].

Currently, two common types of ensemble learning techniques are bagging-based methods and boosting-based methods. For instance, Random Forest is a representative algorithm of bagging, while Adaboost, Gradient Boosting Decision Trees (GBDT), and XGBoost are examples of boosting-based algorithms [38, 39].

• Support Vector Regression (SVR)

An examination of the connection between one or more independent variables and a continuous dependent variable is done using the supervised regression approach known as support vector regression (SVR).

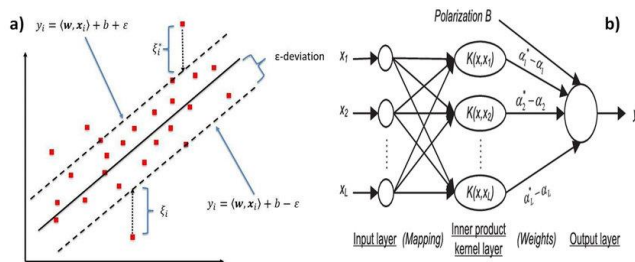


Fig 7. Visual illustration of the SVR algorithm.

While linear regression techniques depend on certain model assumptions, support vector regression (SVR) focusses on identifying the significance of variables in order to describe the connection between inputs and outputs. By keeping the inaccuracy within a certain tolerance margin, this method improves the modelling and prediction of continuous data [33].

B. Unsupervised Machine Learning

Unsupervised machine learning techniques use sophisticated models with millions of parameters to

analyze vast quantities of unlabeled data in a highly non-linear manner. These methods are popular tools for clustering and exploratory data analysis, allowing for the discovery of hidden patterns within the data. Unlike supervised learning, which relies on labeled data, unsupervised learning draws inferences from datasets that lack explicit output labels. Key applications of unsupervised learning include market research, item recognition, and DNA sequence analysis [34].

Putting incoming data into meaningful categories based on similarities and features rather than predetermined labels is the fundamental idea behind unsupervised learning. This is grouping data according to innate patterns instead of precise categorizations. Hard clustering and soft clustering are the two primary categories of clustering techniques. While soft clustering permits data points to belong to numerous clusters with differing degrees of membership, hard clustering allocates each data point to a single cluster. Popular techniques for unsupervised machine learning are covered in the section that follows [40, 41].

A. K-Means

K-means is a well-liked unsupervised learning method that is effective and straightforward for handling clustering issues. By minimizing the total squared distances between each point and the centroid of its designated cluster, the K-means method divides data points into k clusters. This technique is popular for a variety of clustering applications because it effectively divides data into clusters with low intra-cluster variance [42, 43].

B. K-Medoids

Unlike K-Means, which uses the mean value of data points in a cluster as a reference point, K-Medoids employs actual data points as the central objects, or medoids, to determine cluster centers. K-Medoids assigns each data point to the nearest medoid and builds clusters around these central objects. Although K-Medoids can produce conflicting results depending on the initial medoids, it is less sensitive to outliers and can adapt cluster memberships more effectively than K-Means [44, 45].

C. Using Hierarchical Grouping

One popular technique in data mining for cluster analysis is hierarchical cluster analysis (HCA), also referred to as hierarchical clustering. By comparing the traits inside each cluster, it seeks to establish a hierarchical structure of clusters. This methodology creates tiered sets of clusters repeatedly, resulting in a diagram that resembles a tree called a dendrogram. The relationships between data points and clusters are visually represented by the dendrogram, where each level denotes a distinct stage of cluster development [46-49].

Table 1. Gives a quick overview of the pros and cons of each algorithm in a clear and concise manner [46-49].

SUPERVISED			UNSUPERVISED		
Algorithms	Pros	Cons	Algorithms	Pros	Cons
DT	Easy to interpret, handles both categorical and numerical data, works well with non-linear data.	Prone to overfitting, especially with noisy data.	K-Means	Fast, scalable, and works well with large datasets. Effective for spherical clusters.	Sensitive to initial centroids and outliers, struggles with clusters of varying sizes or densities, and assumes spherical clusters.
SVM	Effective in high-dimensional spaces, robust to outliers, and works well with clear margin separation.	Computationally intensive, less effective with large datasets, and difficult to interpret.	K-Medoids	More robust to outliers and noise compared to K-Means, as it uses medoids instead of means.	Slower and more computationally intensive than K-Means, especially with large datasets.
KNN	Simple to implement, no training phase, effective with small datasets.	Computationally expensive with large datasets, sensitive to irrelevant features, and storage-intensive.	Using Hierarchical Grouping	Does not require the number of clusters to be specified, provides a hierarchy of clusters, and can capture complex cluster structures.	Computationally expensive, especially for large datasets, and sensitive to noise and outliers.
Linear Regression	Simple and interpretable, works well with linear relationships, and easy to implement.	Assumes linearity, sensitive to outliers, and may underperform with non-linear data.			
Logistic Regression	Interpretable, works well with binary classification, and can handle linear decision boundaries.	Assumes linearity, struggles with complex relationships, and sensitive to outliers.			
Ensemble Methods	Combines multiple models to improve performance, reduces overfitting, and increases accuracy and robustness.	More complex and computationally expensive, less interpretable, and requires careful tuning.			
SVR	Effective for regression with high-dimensional data, robust against overfitting in high-dimensional spaces.	Similar challenges as SVM, including computational complexity, and less intuitive to interpret.			

The table compares various algorithms, highlighting that **Decision Trees** are easy to interpret and handle different data types but are prone to overfitting. **SVM** is effective in high-dimensional spaces but is computationally intensive. **KNN** is simple and effective with small datasets but struggles with large datasets and irrelevant features. **Linear and Logistic Regression** are interpretable and handle linear relationships well but are limited by their assumption of linearity and sensitivity to outliers. **Ensemble Methods** improve accuracy and reduce overfitting by combining models but are more complex and less interpretable. **SVR** shares SVM's strengths in high-dimensional regression but also its computational challenges. **K-Means** is fast and scalable but sensitive to outliers and initial centroids, while **K-Medoids** is more robust to outliers but slower. **Hierarchical Grouping** captures complex structures without needing a preset number of clusters but is computationally expensive and sensitive to noise.

4. EVALUATION MATRIX OF SUPERVISED CLASSIFICATION ALGORITHMS

Three standard measures are used to assess the performance of supervised classification algorithms: specificity, sensitivity, and accuracy. Specificity is the amount of true negative data points identified in actual negative data points (TP = true positive, TN = true negative, FN = false negative, and FP = false positive); accuracy is the percentage of prediction rate in the model; and sensitivity is the amount of true positive data points correctly identified in actual positive data points [3- 5].

Accuracy: A category's accuracy is calculated by dividing its "correct predictions made" total by the number of "total predictions made" by a category that is similar.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (9)$$

Sensitivity: Real positive rate: If the individual has a positive result, the model will be positive in a tiny fraction of situations, according to the formula below.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (10)$$

Specificity: If the person gets a poor result, it will only happen in a tiny portion of situations. This is calculated with the following formula [50-53].

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (11)$$

5. DISSCUSSION

Healthcare has shown considerable promise for both supervised and unsupervised machine learning

technologies. The applications of these approaches vary based on the type of data and the specific tasks at hand, each with its own advantages and limitations.

With supervised learning, a model is trained using labeled data in order to forecast outcomes based on input features [65-67]. It has been widely applied in the medical field to diagnose, classify, and predict prognoses [68-70]. To predict cardiovascular risk, identify malignant cells, and classify medical images, for instance, supervised learning algorithms such as decision trees, logistic regression, and support vector machines have been used [71-75]. Supervised learning is useful, but it needs a lot of labeled data, and it can be biased if the training set isn't typical of the general population.

On the other hand, unsupervised learning makes use of unlabeled data to train a model that, in the absence of explicit guidance, finds patterns and correlations on its own [14, 15, 16]. This method works well in the medical field for tasks like clustering, anomaly detection, and feature extraction [54-57]. For example, K-means clustering techniques have been used to identify uncommon conditions, extract pertinent information from medical images, and classify patients based on shared features [58-60]. Unsupervised learning, however, occasionally yields outcomes that are difficult to interpret and are not clinically relevant.

Thus, there are clear advantages and disadvantages to both supervised and unsupervised learning in the healthcare industry. The particular task at hand, the type of data, and the resources at hand all influence which of these approaches is best. Machine learning will be essential to enhancing patient outcomes and advancing medical research as long as healthcare data is available.

Machine learning algorithms each have distinct strengths and weaknesses in medical health applications. Decision Trees are easy to understand but can overfit and be biased. Random Forests reduce overfitting but can become complex and resource-intensive. Support Vector Machines (SVM) perform well in high-dimensional spaces but are expensive and hard to interpret. Neural Networks capture complex patterns across various data types but require significant resources and are often seen as a "black box." K-Nearest Neighbors (KNN) is simple and flexible but computationally expensive and sensitive to irrelevant features. Logistic Regression is efficient and interpretable but limited to linear relationships. Gradient Boosting Machines (GBMs) offer high accuracy but are prone to overfitting and are complex to implement. Principal Component Analysis (PCA) reduces dimensionality effectively but may lose important information, while Naive Bayes is efficient but struggles with correlated features due to its assumption of independence [61-64].

Table 2. summarizing the strengths and weaknesses of machine learning algorithms in medical health, along with references for each aspect [61-64].

ASPECT	STRENGTHS	WEAKNESSES	REF
Decision Trees	<ul style="list-style-type: none"> - Simple to understand and interpret. - Useful for classification tasks with clear decision rules. 	<ul style="list-style-type: none"> - Prone to overfitting, especially with complex data. - Can be biased towards features with more levels. 	[25]
Random Forest	<ul style="list-style-type: none"> - Reduces overfitting by averaging multiple decision trees. - Handles large datasets with higher accuracy. 	<ul style="list-style-type: none"> - Computationally intensive. - Model can become complex and less interpretable. 	[26]
Support Vector Machines (SVM)	<ul style="list-style-type: none"> - Effective in high-dimensional spaces. - Works well for classification problems with clear margins. 	<ul style="list-style-type: none"> - Memory and computationally expensive. - Difficult to interpret results and tune hyperparameters. 	[27]
Neural Networks (ANNs)	<ul style="list-style-type: none"> - Capable of capturing complex patterns in data. - Suitable for various types of data including images and texts. 	<ul style="list-style-type: none"> - Requires large amounts of data and computational resources. - Can be seen as a "black box" with poor interpretability. 	[28]
K-Nearest Neighbors (KNN)	<ul style="list-style-type: none"> - Simple to implement and understand. - Flexible and adaptable to different types of data. 	<ul style="list-style-type: none"> - Computationally expensive with large datasets. - Sensitive to irrelevant features and the choice of 'k'. 	[29]
Logistic Regression	<ul style="list-style-type: none"> - Provides probabilistic outputs and is easy to interpret. - Computationally efficient for binary classification. 	<ul style="list-style-type: none"> - Assumes a linear relationship between features and the log odds of the outcome. - Limited to binary classification without extensions. 	[30]
Gradient Boosting Machines (GBMs)	<ul style="list-style-type: none"> - Provides high prediction accuracy. - Can handle various types of data and feature interactions. 	<ul style="list-style-type: none"> - Prone to overfitting if not tuned properly. - Computationally expensive and complex to implement. 	[31]
Principal Component Analysis (PCA)	<ul style="list-style-type: none"> - Reduces dimensionality while retaining most variance in data. - Simplifies models and speeds up computation. 	<ul style="list-style-type: none"> - May discard useful information. - Difficult to interpret principal components. 	[32]
Naive Bayes	<ul style="list-style-type: none"> - Simple and efficient for large datasets. - Good performance with small to moderate-sized datasets. 	<ul style="list-style-type: none"> - Assumes feature independence, which may not hold true. - Less accurate with highly correlated features. 	[33]

6. CONCLUSION

In conclusion, the application of machine learning algorithms in healthcare offers significant potential to enhance diagnostic accuracy and treatment effectiveness. As highlighted in this review, different algorithms come with their own strengths and limitations. For instance, while decision trees and ensemble methods provide interpretability and improved accuracy through model combinations, they can suffer from overfitting and complexity. On the other hand, algorithms like SVM and SVR excel in handling high-dimensional data but require substantial computational resources. Simpler algorithms such as KNN and logistic regression, though effective in specific contexts, face challenges with scalability and handling non-linear relationships.

Moreover, unsupervised techniques like K-Means and hierarchical grouping offer valuable insights into data patterns without requiring labeled datasets, but they can be sensitive to initial conditions and computationally intensive. The review underscores the importance of selecting the appropriate algorithm based on the specific characteristics of the healthcare data at hand, whether it involves small or large datasets, linear or non-linear relationships, or the need for scalability and robustness. Ultimately, the integration of these machine learning models into healthcare systems must consider these trade-offs to optimize patient outcomes and improve the efficiency of medical practices.

REFERENCES

- [1] Mehrpour, O., Saeedi, F., Vohra, V., Abdollahi, J., Shirazi, F. M., & Goss, F. (2023). The role of decision tree and machine learning models for outcome prediction of bupropion exposure: A nationwide analysis of more than 14 000 patients in the United States. *Basic & Clinical Pharmacology & Toxicology*, 133(1), 98-110.
- [2] Duan, H., & Mirzaei, A. (2023). Adaptive Rate Maximization and Hierarchical Resource Management for Underlay Spectrum Sharing NOMA HetNets with Hybrid Power Supplies. *Mobile Networks and Applications*, 1-17.
- [3] Abdollahi, J. (2022, February). Identification of medicinal plants in ardabil using deep learning: identification of medicinal plants using deep learning. In *2022 27th International Computer Conference, Computer Society of Iran (CSICC)* (pp. 1-6). IEEE.
- [4] Mirzaei, A., & Najafi Souha, A. (2021). Towards optimal configuration in MEC Neural networks: deep learning-based optimal resource allocation. *Wireless Personal Communications*, 121(1), 221-243.
- [5] Abdollahi, J., Davari, N., Panahi, Y., & Gardaneh, M. (2022). Detection of Metastatic Breast Cancer from Whole-Slide Pathology Images Using an Ensemble Deep-Learning Method: Detection of Breast Cancer using Deep-Learning. *Archives of Breast Cancer*, 364-376.
- [6] Shahriyar, O., Moghaddam, B. N., Yousefi, D., Mirzaei, A., & Hoseini, F. (2025). An analysis of the combination of feature selection and machine learning methods for an accurate and timely detection of lung cancer. *arXiv preprint arXiv:2501.10980*.
- [7] Mikaeilvand, N., Ojaroudi, M., & Ghadimi, N. (2015). Band-Notched Small Slot Antenna Based on Time-Domain Reflectometry Modeling for UWB Applications. *The Applied Computational Electromagnetics Society Journal (ACES)*, 682-687.
- [8] Li, X., Lan, X., Mirzaei, A., & Bonab, M. J. A. (2022). Reliability and robust resource allocation for Cache-enabled HetNets: QoS-aware mobile edge computing. *Reliability Engineering & System Safety*, 220, 108272.
- [9] Abdollahi, J., & Mahmoudi, L. (2022, February). An Artificial Intelligence System for Detecting the Types of the Epidemic from X-rays: Artificial Intelligence System for Detecting the Types of the Epidemic from X-rays. In *2022 27th International Computer Conference, Computer Society of Iran (CSICC)* (pp. 1-6). IEEE.
- [10] Somarin, A. M., Barari, M., & Zarrabi, H. (2018). Big data based self-optimization networking in next generation mobile networks. *Wireless Personal Communications*, 101(3), 1499-1518.
- [11] Amani, F., & Abdollahi, J. (2022). Using Stacking methods based Genetic Algorithm to predict the time between symptom onset and hospital arrival in stroke patients and its related factors. *Journal of Biostatistics and Epidemiology*, 8(1), 8-23.
- [12] Jahanbakhsh Gudakahriz, S., Momtaz, V., Nouri-Moghadam, B., Mirzaei, A., & Vajed Khiavi, M. (2025). Link life time and energy-aware stable routing for MANETs. *International Journal of Nonlinear Analysis and Applications*.
- [13] Abdollahi, J., Keshandehghan, A., Gardaneh, M., Panahi, Y., & Gardaneh, M. (2020). Accurate detection of breast cancer metastasis using a hybrid model of artificial intelligence algorithm. *Archives of Breast Cancer*, 22-28.
- [14] PARVAR, M. E., SOMARIN, A. M., TAHERNEZHAD, M. R., & ALAEI, Y. (2015). Proposing a new method for routing improvement in wireless ad hoc networks (optional). *Fen Bilimleri Dergisi (CFD)*, 36(4).
- [15] Barzaki, M. A. J. Z., Abdollahi, J., Negaresh, M., Salimi, M., Zolfaghari, H., Mohammadi, M., ... & Amani, F. (2023, November). Using Deep Learning for Classification of Lung Cancer on CT Images in Ardabil Province: Classification of Lung Cancer using Xception. In *2023 13th International*

- Conference on Computer and Knowledge Engineering (ICCKE) (pp. 375-382). IEEE.
- [16] Narimani, Y., Zeinali, E., & Mirzaei, A. (2022). QoS-aware resource allocation and fault tolerant operation in hybrid SDN using stochastic network calculus. *Physical Communication*, 53, 101709.
- [17] Abdollahi, J. (2023). Evaluating LeNet Algorithms in Classification Lung Cancer from Iraq-Oncology Teaching Hospital/National Center for Cancer Diseases. *arXiv preprint arXiv:2305.13333*.
- [18] Mirzaei, A. (2022). A novel approach to QoS-aware resource allocation in NOMA cellular HetNets using multi-layer optimization. *Concurrency and Computation: Practice and Experience*, 34(21), e7068.
- [19] Khavandi, H., Moghadam, B. N., Abdollahi, J., & Branch, A. (2023). Maximizing the Impact on Social Networks using the Combination of PSO and GA Algorithms. *Future Generation in Distributed Systems*, 5, 1-13.
- [20] Jahandideh, Y., & Mirzaei, A. (2021). Allocating Duplicate Copies for IoT Data in Cloud Computing based on Harmony Search Algorithm. *IETE Journal of Research*, 1-14.
- [21] Abdollahi, J., NouriMoghaddam, B., & MIRZAEI, A. (2023). Diabetes Data Classification using Deep Learning Approach and Feature Selection based on Genetic.
- [22] Mirzaei, A., Barari, M., & Zarrabi, H. (2019). Efficient resource management for non-orthogonal multiple access: A novel approach towards green hetnets. *Intelligent Data Analysis*, 23(2), 425-447.
- [23] Javadzadeh Barzaki, M. A., Negaresh, M., Abdollahi, J., Mohammadi, M., Ghobadi, H., Mohammadzadeh, B., & Amani, F. (2022, July). USING DEEP LEARNING NETWORKS FOR CLASSIFICATION OF LUNG CANCER NODULES IN CT IMAGES. In *Iranian Congress of Radiology* (Vol. 37, No. 2, pp. 34-34). Iranian Society of Radiology.
- [24] Mirzaei, A., Barari, M., & Zarrabi, H. (2021). An Optimal Load Balanced Resource Allocation Scheme for Heterogeneous Wireless Networks based on Big Data Technology. *arXiv preprint arXiv:2101.02666*.
- [25] Abdollahi, J., Aref, S. Early Prediction of Diabetes Using Feature Selection and Machine Learning Algorithms. *SN COMPUT. SCI.* 5, 217 (2024). <https://doi.org/10.1007/s42979-023-02545-y>.
- [26] Mirzaei, A., & Rahimi, A. (2019). A Novel Approach for Cluster Self-Optimization Using Big Data Analytics. *Information Systems & Telecommunication*, 50.
- [27] Narimani, Y., Zeinali, E., & Mirzaei, A. (2025). A new approach in fault tolerance in control level of SDN. *International Journal of Nonlinear Analysis and Applications*, 16(5), 69-76.
- [28] Rad, K. J., & Mirzaei, A. (2022). Hierarchical capacity management and load balancing for HetNets using multi-layer optimisation methods. *International Journal of Ad Hoc and Ubiquitous Computing*, 41(1), 44-57.
- [29] Amani, F., Abdollahi, J., & Amani, P. (2024, February). Identify the Factors Influencing Suicide among Ardabil city People Using Feature Selection: Identify the Factors Influencing Suicide among Ardabil using machine learning. In *2024 10th International Conference on Artificial Intelligence and Robotics (QICAR)* (pp. 17-23). IEEE.
- [30] Barari, M., Zarrabi, H., & Somarin, A. M. (2016). A New Scheme for Resource Allocation in Heterogeneous Wireless Networks based on Big Data. *Bulletin de la Société Royale des Sciences de Liège*, 85, 340-347.
- [31] Abdollahi, J., & Mehrpour, O. (2024, February). Using Machine Learning Algorithms for Coronary Artery Disease (CAD) Prediction Prediction of Coronary Artery Disease (CAD) Using Machine Learning Algorithms. In *2024 10th International Conference on Artificial Intelligence and Robotics (QICAR)* (pp. 164-172). IEEE.
- [32] Mirzaei, A. (2021). QoS-aware Resource Allocation for Live Streaming in Edge-Clouds Aided HetNets Using Stochastic Network Calculus.
- [33] Javaid, M., Haleem, A., Singh, R. P., Suman, R., & Rab, S. (2022). Significance of machine learning in healthcare: Features, pillars and applications. *International Journal of Intelligent Networks*, 3, 58-73.
- [34] NOSRATIP, M., HOSEINIP, M., SHIRMARZP, A., SOMARINP, A. M., HOSEININIAP, N., BARARIP, M., & Ardebil, I. (2016). Application of MLP and RBF Methods in Prediction of Travelling within the city. *Bulletin de la Société Royale des Sciences de Liège*, 85, 1392-1396.
- [35] Habehh, H., & Gohel, S. (2021). Machine learning in healthcare. *Current genomics*, 22(4), 291.
- [36] Somarin, A. M., Nosrati, M., Barari, M., & Zarrabi, H. (2016). A new Joint Radio Resource Management scheme in heterogeneous wireless networks based on handover. *Bulletin de la Société Royale des Sciences de Liège*
- [37] [19] Swain, S., Bhushan, B., Dhiman, G., & Viriyasitavat, W. (2022). Appositeness of optimized and reliable machine learning for healthcare: a survey. *Archives of Computational Methods in Engineering*, 29(6), 3981-4003.
- [38] Nemati, Z., Mohammadi, A., Bayat, A., & Mirzaei, A. (2024). Metaheuristic and Data Mining Algorithms-based Feature Selection Approach for Anomaly Detection. *IETE Journal of Research*, 1-15.
- [39] Eckerson, W. W. (2007). Predictive analytics. Extending the Value of Your Data Warehousing Investment. *TDWI Best Practices Report*, 1, 1-36.
- [40] Nemati, Z., Mohammadi, A., Bayat, A., & Mirzaei, A. (2025). Fraud Prediction in Financial Statements through Comparative Analysis of Data Mining

- Methods. *International Journal of Finance & Managerial Accounting*, 10(38), 151-166.
- [41] Khanna, N. N., Maindarkar, M. A., Viswanathan, V., Fernandes, J. F. E., Paul, S., Bhagawati, M., ... & Suri, J. S. (2022, December). Economics of artificial intelligence in healthcare: diagnosis vs. treatment. In *Healthcare* (Vol. 10, No. 12, p. 2493). MDPI.
- [42] Nematia, Z., Mohammadia, A., Bayata, A., & Mirzaeib, A. (2024). Predicting fraud in financial statements using supervised methods: An analytical comparison. *International Journal of Nonlinear Analysis and Applications*, 15(8), 259-272.
- [43] Mehrpour, O., Saeedi, F., Abdollahi, J., Amirabadizadeh, A., & Goss, F. (2023). The value of machine learning for prognosis prediction of diphenhydramine exposure: National analysis of 50,000 patients in the United States. *Journal of Research in Medical Sciences*, 28(1), 49.
- [44] Nemati, Z., Mohammadi, A., Bayat, A., & Mirzaei, A. (2024). The impact of financial ratio reduction on supervised methods' ability to detect financial statement fraud. *Karafan Quarterly Scientific Journal*.
- [45] Mathur, S., & Sutton, J. (2017). Personalized medicine could transform healthcare. *Biomedical reports*, 7(1), 3-5.
- [46] Nemati, Z., Mohammadi, A., Bayat, A., & Mirzaei, A. (2024). Fraud Risk Prediction in Financial Statements through Comparative Analysis of Genetic Algorithm, Grey Wolf Optimization, and Particle Swarm Optimization. *Iranian Journal of Finance*, 8(1), 98-130
- [47] Berner, E. S. (2007). *Clinical decision support systems* (Vol. 233). New York: Springer Science+ Business Media, LLC.
- [48] Nemati, Z., Mohammadi, A., Bayat, A., & Mirzaei, A. (2023). Financial Ratios and Efficient Classification Algorithms for Fraud Risk Detection in Financial Statements. *International Journal of Industrial Mathematics*.
- [49] Swarthout, M., & Bishop, M. A. (2017). Population health management: review of concepts and definitions. *American Journal of Health-System Pharmacy*, 74(18), 1405-1411.
- [50] Nematollahi, M., Ghaffari, A., & Mirzaei, A. (2024). Task offloading in Internet of Things based on the improved multi-objective aquila optimizer. *Signal, Image and Video Processing*, 18(1), 545-552.
- [51] De Ville, B. (2013). Decision trees. *Wiley Interdisciplinary Reviews: Computational Statistics*, 5(6), 448-455.
- [52] Hozouri, A., EffatParvar, M., Yousefi, D., & Mirzaei, A. Scheduling algorithm for bidirectional LPT.
- [53] Webb, G. I., Keogh, E., & Miikkulainen, R. (2010). Naïve Bayes. *Encyclopedia of machine learning*, 15(1), 713-714.
- [54] Babazadeh, Z., & Mirzaei, A. A review on methods, ways of decrease delay duties at calculations cloudy.
- [55] Peterson, L. E. (2009). K-nearest neighbor. *Scholarpedia*, 4(2), 1883.
- [56] Movahhedi, R., Effatparvar, M., & Somarin, A. M. A study on load balancing methods in cloud computing environment.
- [57] Cunningham, P., & Delany, S. J. (2021). K-nearest neighbour classifiers-a tutorial. *ACM computing surveys (CSUR)*, 54(6), 1-25.
- [58] Ziaeddini, A., Mohajer, A., Yousefi, D., Mirzaei, A., & Gonglee, S. (2022). An optimized multi-layer resource management in mobile edge computing networks: a joint computation offloading and caching solution. *arXiv preprint arXiv:2211.15487*
- [59] Su, X., Yan, X., & Tsai, C. L. (2012). Linear regression. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(3), 275-294.
- [60] Nokhostin, P., Mirzaei, A., & Jahanbakhsh, S. Proposed Methods for Establishing Load Balancing in Fog Computing: A Survey.
- [61] LaValley, M. P. (2008). Logistic regression. *Circulation*, 117(18), 2395-2399.
- [62] Mehri, R., & Somarin, A. M. (2020). Designing an Energy-Efficient Mechanism to Regulate the Transmission Power Rate in Wireless Sensor Networks. *World*, 9(S1), 189-194.
- [63] Dietterich, T. G. (2000, June). Ensemble methods in machine learning. In *International workshop on multiple classifier systems* (pp. 1-15). Berlin, Heidelberg: Springer Berlin Heidelberg.
- [64] Mohammad Zadeh, M., & Mirzaei Somarin, A. (2017). Attack Detection in Mobile Ad Hoc.
- [65] Opitz, D., & Maclin, R. (1999). Popular ensemble methods: An empirical study. *Journal of artificial intelligence research*, 11, 169-198.
- [66] Zhang, S., Madadkhani, M., Shafieezadeh, M., & Mirzaei, A. (2019). A novel approach to optimize power consumption in orchard WSN: Efficient opportunistic routing. *Wireless Personal Communications*, 108(3), 1611-1634.
- [67] Derakhshandeh, S., & Mikaeilvand, N. (2011). New framework for comparing information security risk assessment methodologies. *Australian Journal of Basic and Applied Sciences*, 5(9), 160-166.
- [68] Mirzaei, A., & Zandiyan, S. (2023). A Novel Approach for Establishing Connectivity in Partitioned Mobile Sensor Networks using Beamforming Techniques. *arXiv preprint arXiv:2308.04797*.
- [69] Allahviranloo, T., & Mikaeilvand, N. (2011). Non zero solutions of the fully fuzzy linear systems. *Appl. Comput. Math*, 10(2), 271-282.
- [70] Awad, M., Khanna, R., Awad, M., & Khanna, R. (2015). Support vector regression. *Efficient learning machines: Theories, concepts, and*

- applications for engineers and system designers, 67-80.
- [71] Javid, S., & Mirzaei, A. (2021). Highly Reliable Routing In Healthcare Systems Based on Internet of Things
 - [72] Hahne, F., Huber, W., Gentleman, R., Falcon, S., Gentleman, R., & Carey, V. J. (2008). Unsupervised machine learning. Bioconductor case studies, 137-157.
 - [73] Nematollahi, M., Ghaffari, A., & Mirzaei, A. (2024). Task and resource allocation in the internet of things based on an improved version of the moth-flame optimization algorithm. Cluster Computing, 27(2), 1775-1797.
 - [74] Kaur, N. K., Kaur, U., & Singh, D. (2014). K-Medoid clustering algorithm-a review. Int. J. Comput. Appl. Technol, 1(1), 42-45.\
 - [75] Navarro-Cerdán, J. R., Sánchez-Gomis, M., Pons, P., Gálvez-Settier, S., Valverde, F., Ferrer-Albero, A., ... & Redon, J. (2023). Towards a personalized health care using a divisive hierarchical clustering approach for comorbidity and the prediction of conditioned group risks. Health Informatics Journal, 29(4), 14604582231212494.