Human-Centric Intelligent Systems (2023) 3:588–615
https://doi.org/10.1007/s44230-023-00041-3

REVIEW ARTICLE

Applications and Techniques of Machine Learning in Cancer Classification: A Systematic Review

Abrar Yaqoob[1] · Rabia Musheer Aziz[1] · Navneet Kumar verma[1]

Received: 10 April 2023 / Accepted: 10 August 2023 / Published online: 11 September 2023

Abstract
The domain of Machine learning has experienced Substantial advancement and development. Recently, showcasing a Broad spectrum of uses like Computational linguistics, image identification, and autonomous systems. With the increasing demand for intelligent systems, it has become crucial to comprehend the different categories of machine acquiring knowledge systems along with their applications in the present world. This paper presents actual use cases of machine learning, including cancer classification, and how machine learning algorithms have been implemented on medical data to categorize diverse forms of cancer and anticipate their outcomes. The paper also discusses supervised, unsupervised, and reinforcement learning, highlighting the benefits and disadvantages of each category of Computational intelligence system. The conclusions of this systematic study on machine learning methods and applications in cancer classification have numerous implications. The main lesson is that through accurate classification of cancer kinds, patient outcome prediction, and identification of possible therapeutic targets, machine learning holds enormous potential for improving cancer diagnosis and therapy. This review offers readers with a broad understanding as of the present advancements in machine learning applied to cancer classification today, empowering them to decide for themselves whether to use these methods in clinical settings. Lastly, the paper wraps up by engaging in a discussion on the future of machine learning, including the potential for new types of systems to be developed as the field advances. Overall, the information included in this survey article is useful for scholars, practitioners, and individuals interested in gaining knowledge about the fundamentals of machine learning and its various applications in different areas of activities.

Keywords Dimensionality reduction · Deep learning · Machine learning · Reinforcement learning · Supervised learning · Unsupervised learning

Abbreviations

KNN	K-nearest neighbors
LR	Logistic regression
LR	Linear regression
RF	Random forest
DT	Decision tree
NN	Neural network
SVM	Support vector machine

1 Introduction

The inception of "machine learning" dates back to 1959, when Arthur Samuel, a pioneering figure in the domains of artificial intelligence and computer gaming hailing from the United States, coined this term. Samuel devised a program that played checkers and utilized self-play to enhance its gameplay over time, paving the way for contemporary machine-learning algorithms. Presently, machine learning has become ubiquitous in various domains like natural language processing, image recognition, and other recommendation systems. For instance, image recognition algorithms can learn from vast collections of labelled images to recognize objects in new pictures, while natural language processing algorithms can learn from massive datasets of text to detect speech or translate languages [1]. Machine learning, which falls under the umbrella of artificial intelligence and computer science, involves the development of

✉ Abrar Yaqoob
abrar.yaqoob2022@vitbhopal.ac.in

Rabia Musheer Aziz
rabia.musheer@vitbhopal.ac.in

Navneet Kumar verma
navneet.verma@vitbhopal.ac.in

[1] School of Advanced Sciences and Languages, VIT Bhopal University, Kothrikalan, Sehore 466114, India

algorithms and models that enable computers to learn and make predictions or decisions autonomously, without the need for explicit programming instructions. This involves feeding extensive data to an algorithm, enabling it to identify patterns and connections within the data. Machine learning has emerged as a pivotal technology in numerous sectors, such as healthcare, finance, and e-commerce, fundamentally transforming how we analyze data and make informed decisions. However, it also poses several challenges, such as the necessity for enormous amount of data and the comprehensibility of the decision-making procedure of the algorithm. The development of more robust and ethical machine learning systems is an ongoing research area in the field [2, 3].

Machine learning roots can be detected to the mid-twentieth century and it has since been applied in a broad spectrum of practical applications. One of the earliest examples of machine learning was Cybertron, an experimental "learning machine" created by Raytheon Company in the 1960s. This device used punched tape memory to analyze sonar data, electrocardiograms and speech patterns. The Cybertron was repeatedly underwent training by a human operator to recognize patterns and was equipped with a "goof" button to reconsider bad choices. During this period, machine learning research primarily focused on pattern categorization, as demonstrated by the book Learning Machines by Nilsson. Pattern recognition continued to be an area of interest as highlighted by Duda and Hart in 1973, In 1981, researchers presented. Studies have been conducted to train a neural network in recognizing a set of 40 characters that are commonly found in computer terminals. These characters consist of 26 letters, 10 numbers, and 4 special symbols. This marked a significant development in machine learning and paved the way for future advancements in the field [4].

Two primary objectives of modern machine learning are: one is to group data into categories using pre-existing patterns, and the other is to predict future outcomes based on those patterns. For example, by using computer vision and supervised learning, we can train an algorithm to recognize malignant moles based on their appearance. A stock trading machine learning algorithm could alert the trader to potential future predictions [5].

In summation, Machine Learning is ideal for:

- Machine learning can reduce the need for extensive manual adjusting or lengthy lists of rules for solving complex issues that existing solutions struggle to handle effectively [6].
- Machine learning can solve complicated issues that traditional methods cannot realistically address by utilizing the best techniques and adapting to new data in different circumstances [7].
- Machine learning can gain valuable insights and knowledge from massive amounts of data and complex situations, simplifying the information into understandable sentences [8] (Fig. 1).

Artificial learning techniques have a pivotal role to play in cancer classification by analyzing complex and high-dimensional datasets. These algorithms automatically select relevant features or biomarkers from large-scale datasets, improving the accuracy and efficiency of cancer classification models. By identifying hidden patterns and relationships in the data, machine learning algorithms can discover subtle associations between genetic or molecular markers and different cancer types, leading to improved classification accuracy. The algorithms learn from labeled examples to build predictive models, iteratively adjusting their parameters to optimize performance. Ensemble methods, including random forests and gradient boosting, amalgamate multiple models to improve accuracy of predictions, enhancing the robustness of categorization cancer models [9]. Machine learning algorithms can be deployed in real-time systems to provide rapid analysis and decision support for cancer diagnosis and treatment, improving patient care and treatment outcomes. Their application in cancer research and clinical practice enables advancements in precision medicine and

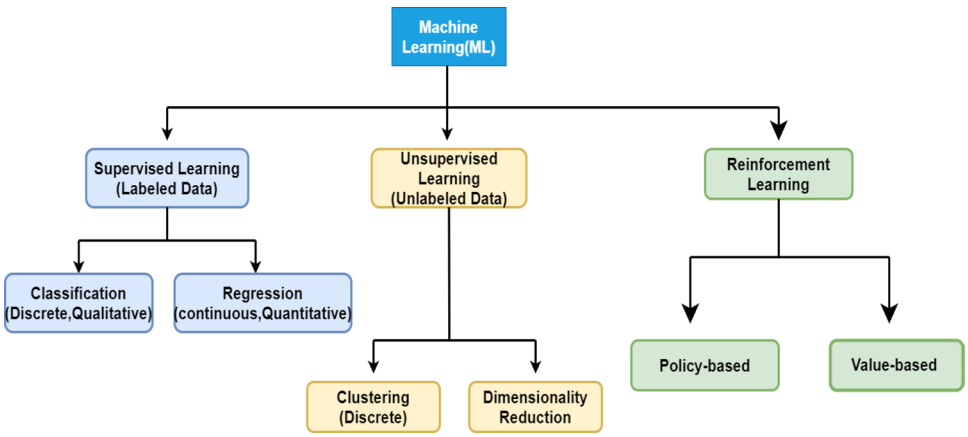

Fig. 1 Overview diagram of machine learning

personalized treatment strategies. Moreover, machine learning algorithms excel at handling the inherent complexity of cancer datasets, which often contain an overwhelming number of variables and intricate relationships. By employing advanced mathematical and statistical techniques, these algorithms can effectively navigate through the vast data landscape, uncovering nuanced patterns and connections that may elude human observation.

The capacity of machine learning algorithms to understand labeled examples is a cornerstone of their success in cancer classification [10]. By leveraging large annotated datasets, these algorithms acquire the ability to recognize subtle variations and distinguish between different cancer types with remarkable accuracy. The iterative process of adjusting parameters and fine-tuning models enables continuous refinement and optimization, leading to ever-improving performance and enhanced diagnostic precision.

Ensemble methods, such as random forests and gradient boosting, provide an additional layer of strength and reliability to cancer classification models [11]. By combining multiple individual models, each with its own strengths and weaknesses, these ensemble methods harness the collective intelligence of diverse algorithms, resulting in more robust and resilient predictions. The collaborative nature of ensemble learning mitigates the risks of overfitting and enhances generalization capabilities, ultimately improving the overall accuracy and stability of cancer classification systems. When integrated into real-time systems, machine learning algorithms become invaluable tools for rapid analysis and decision support in cancer diagnosis and treatment. These algorithms has the capability to handle massive volumes of data in real-time, quickly extracting pertinent information and providing clinicians with evidence-based insights. The timely delivery of accurate and actionable information empowers medical practitioners can utilize this technology to make well-informed decisions, customize treatment approaches, and ultimately enhance patient outcomes [12, 13].

To conclude, the utilization of machine learning algorithms in categorizing cancer has shown significant progress represents a remarkable developments in the oncology field. The capacity to analyze complex databases, identify hidden patterns, optimize models, leverage ensemble methods, and provide real-time decision support has revolutionized the way we approach cancer diagnosis and treatment. As we continue to refine and expand these algorithms, their impact on precision medicine and personalized treatment strategies will undoubtedly continue to grow, offering hope and improved outcomes for cancer patients worldwide.

Machine learning techniques continue to evolve, and researchers are exploring new approaches for cancer prediction beyond traditional methods. Here are some novel approaches in clustering, feature selection, and other areas of machine learning in predicting cancer:

1. Deep Learning and Neural Networks: Deep learning techniques, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have gained popularity in cancer prediction. These models have the ability to autonomously acquire pertinent features from intricate medical data, including images and genomic sequences [13, 14].
2. Transfer Learning: Transfer learning involves leveraging pre-trained models on large datasets from related domains and adapting them for cancer prediction tasks. By utilizing knowledge from other domains, transfer learning can improve forecasting accuracy, especially when the amount of cancer-specific data is limited [15].
3. Unsupervised Clustering: Unsupervised clustering algorithms help identify distinct subtypes by analyzing the patterns of gene expression or other molecular data. This allows for personalized treatment strategies and a better understanding of tumor heterogeneity [16].
4. Ensemble Methods: Ensemble methods bring together various machine learning approaches to create a unified models to generate more precise forecasts. It can be strengthened using methods such as bagging, boosting, and stacking and expansion of cancer prediction models [17].
5. Feature Selection with Genetic Algorithms: Genetic algorithms can optimize the feature selection process by iteratively selecting subsets of relevant features that maximize the performance of cancer prediction models. This approach helps reduce dimensionality and improve model interpretability [18].
6. Multi-Omics Integration: Cancer prediction often involves integrating data from multiple sources, such as genomics, transcriptomics, proteomics, and clinical data. Machine learning techniques that can effectively integrate multi-omics data offer a comprehensive view of cancer biology and improve prediction accuracy [19].
7. Explainable AI: Interpretability is crucial in healthcare applications. Researchers are developing machine learning models with built-in explainability, allowing clinicians to understand the reasoning behind predictions. This helps build trust and facilitates the adoption of machine learning models in clinical practice [20].
8. Longitudinal Data Analysis: Cancer progression involves temporal changes, and analyzing longitudinal data can provide valuable insights. Machine learning approaches that model the temporal dynamics of cancer, such as recurrent neural networks and hidden Markov models, enable accurate prediction of disease progression and treatment response [21].

These are just a few examples of the evolving approaches in machine learning for cancer prediction. Ongoing research

and advancements in the field continue to expand the repertoire of techniques and improve our understanding of cancer biology and treatment.

1.1 Various Types of Cancer Data Analysis

When conducting analysis various forms of data are available in the field of cancer that are commonly used to gain insights and make sensible decisions. The type of data utilized depends on the specific research objectives and the available resources. Here are some commonly used data types for cancer analysis used in this survey:

1. Clinical Data: This includes patient-related information such as demographic data, medical history, symptoms, treatment records, laboratory results, pathology reports, and clinical outcomes. Clinical data provides essential insights into patient characteristics and disease progression [22].
2. Genomic Data: Genomic data involves studying the genetic makeup of cancer cells, including DNA sequencing data, gene expression profiles, and genetic variations. This data helps identify genetic mutations, gene expression patterns, and potential biomarkers for cancer diagnosis, prognosis, and treatment selection [23].
3. Imaging Data: Imaging data is produced by medical imaging modalities like X-rays, CT scans, MRI scans, and PET scans. These images provide detailed information about tumor location, size, shape, and characteristics, aiding in cancer diagnosis, staging, and treatment planning [24].
4. Omics Data: Omics data refers to large-scale molecular data, including transcriptomics, proteomics, metabolomics, and epigenomics. These data provide insights into the molecular changes occurring in cancer cells and can help identify novel therapeutic targets and biomarkers [25].
5. Electronic Health Records (EHR): EHRs contain comprehensive patient information, including medical history, diagnoses, treatments, and outcomes. Mining EHR data allows researchers to study large patient populations and identify patterns and trends related to cancer incidence, treatment response, and patient outcomes [26].
6. Publicly Available Datasets: Various public databases and repositories provide researchers with access to curated cancer datasets, such as The Cancer Genome Atlas (TCGA), Gene Expression Omnibus (GEO), and International Cancer Genome Consortium (ICGC). These datasets enable comparative analyses, validation of findings, and collaborative research [27] (Table 1).

1.2 Implication and Uniqueness of Comprehensive Review of the Literature with Existing SLR in Cancer Prediction

The implication and uniqueness of comprehensive review of the literature (SLR) in cancer prediction lie in its unique contribution to the understanding of applications and techniques of machine learning in cancer classification. Compared to existing SLRs, this review offers several distinctive features:

Table 1 Cancer data sets and short Description [28–34]

Data set	Short description
Colon Cancer	Colorectal cancer refers to the formation of cancerous cells in either the colon or rectum. This type of cancer can be recognized by a range of symptoms including the presence of blood in the stool, alterations in bowel movements, weight loss, and feelings of fatigue [28]
Acute Leukema	Acute leukemia is a group of severe symptoms linked to a leukemia diagnosis and can be categorized based on the myeloid or lymphoid lineage of the malignant cells [29]
Prostate Tumor	Prostate cancer is a male gland cancer that can develop slowly or aggressively and requires monitoring or other therapies [30]
High Grade Glioma	High-grade gliomas are fast-growing brain tumours affecting people of all ages and include glioblastomas and anaplastic oligodendrogliomas [31]
Lung Cancer II	The lymph nodes close to the affected lung contain cancer cells, and the tumour measures up to 5 cm. In addition, the lung may be completely or partially collapsed or inflamed, or the cancer may have spread to the primary airway and/or into the layer of the membrane covering the lung Samples of malignant pleural mesothelioma and lung adenocarcinoma tissue acquired from stage II lung cancer [32]
Leukemia 2	Leukemia is a type of cancer that affects the blood and is marked by the rapid growth of abnormal blood cells. This abnormal growth primarily occurs in the bone marrow, which is responsible for producing most of the blood in the body. Leukemia cells are usually immature or underdeveloped white blood cells [33]
Bladder Cancer	Description: Bladder cancer affects the bladder, which is part of the urinary system. It often starts in the lining of the bladder and can cause symptoms like blood in the urine, frequent urination, and pelvic pain [34]

- Comprehensive scope: This SLR encompasses a diverse array of machine learning applications and methodologies specifically focused on cancer classification. It provides a holistic view of the field, covering various data sources, such as imaging, genetic markers, and clinical data, and their utilization in accurately classifying different types of cancer.
- Actual use cases: The review presents real-world use cases where machine learning algorithms have been implemented on medical data to classify cancer and predict outcomes. By highlighting these practical examples, it demonstrates the efficacy and potential of machine learning in clinical settings.
- Comparative analysis: A comparative analysis table is provided, enabling readers to compare this SLR with existing ones. This analysis emphasizes the unique aspects and contributions of the current review, such as its focus on specific applications, comprehensive scope, or novel insights, setting it apart from previous studies.
- Implications for readers: The review's key takeaways and implications provide valuable insights for readers. This statement emphasizes the capacity of machine learning to enhance cancer detection, tailor treatment approaches, and forecast patient outcomes. It enables individuals to make well-informed choices regarding the implementation of machine learning methods in medical settings.

Overall, the implication and originality of this SLR in cancer prediction stem from its comprehensive scope, real-world use cases, comparative analysis, and actionable insights, making it a valuable resource for researchers, practitioners, and anyone interested in machine learning applications for cancer classification (Fig. 2).

1.3 Search Criteria, Inclusion/Exclusion Criteria for Conducting a Review Work

In our review of distinct types of machine learning systems, in the context of systematic reviews and meta-analyses, we used the Recommended Reporting Items (PRISMA) methodology to ensure a comprehensive and systematic approach. The PRISMA method consists of four key steps:

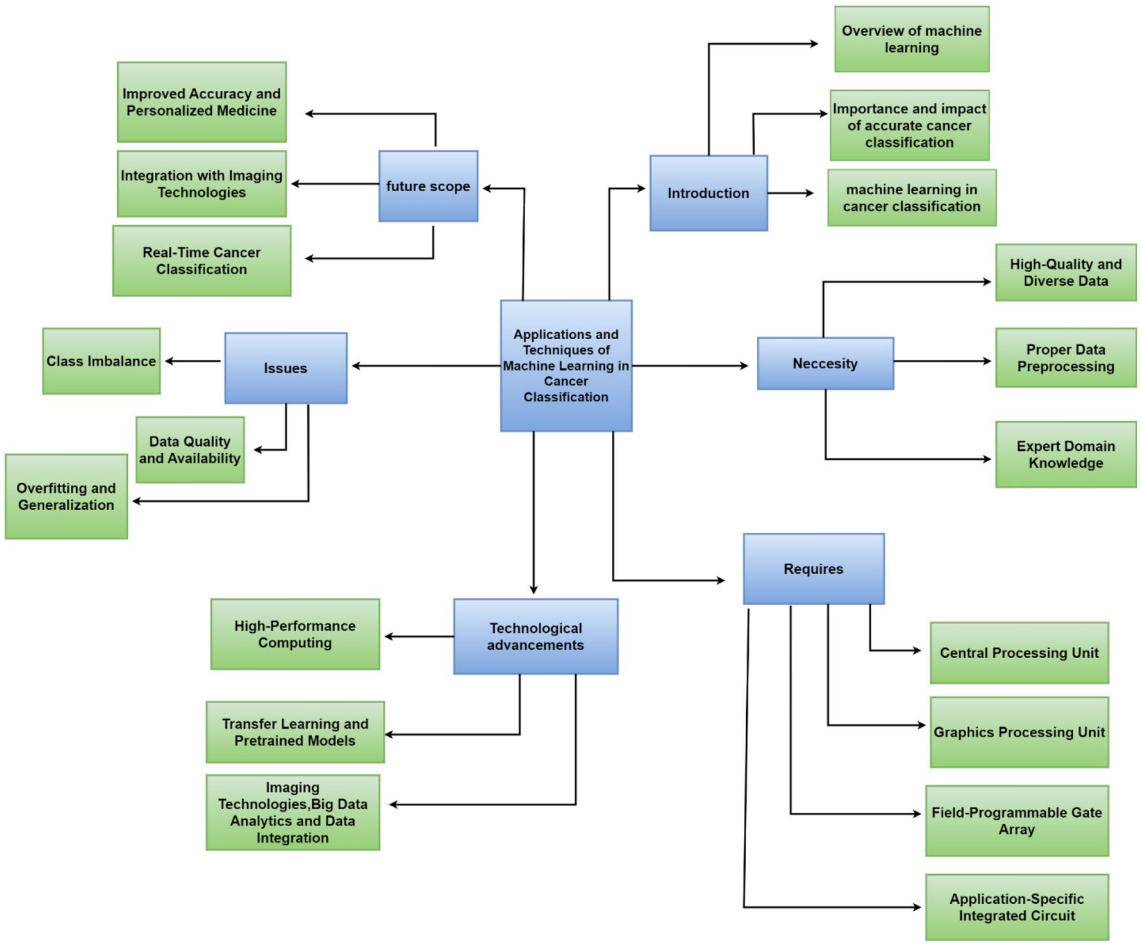

Fig. 2 Concept map of machine learning for cancer classification

identification, screening, eligibility, and inclusion. During the identification step, we utilized relevant and important keywords to search for eligible articles, limiting our study to those published in the English language. We also employed a snowballing technique to identify additional relevant articles by examining the references of the selected articles. Next, we conducted a screening process to evaluate the relevance of each article based on our inclusion criteria. We excluded articles that did not meet our criteria or were duplicates, resulting in a final set of eligible articles. In the eligibility step, we evaluate the quality of each eligible article based on its scientific rigor, methodology, and relevance to our review topic. Articles that did not meet our quality criteria were excluded from our analysis. Finally, in the inclusion step, we selected the final set of articles that met all of our standard and involved them in our review.

By utilizing the PRISMA method, we ensured a comprehensive and systematic approach to our review of different types of machine learning systems. This method enabled us to identify and evaluate a high-quality set of articles that provided relevant and informative insights into the field of machine learning (Fig. 3) (Table 2).

1.4 The Objective of the Paper is as Follows

- To systematically review the applications and techniques of machine learning in cancer classification.
- To assess the efficacy and precision of machine learning algorithms in differentiating and classifying various types of cancer.
- To identify the potential of machine learning in predicting patient outcomes and personalizing cancer treatment approaches.
- To provide insights and implications for the application of machine learning methods in clinical environments to enhance cancer diagnosis and treatment.

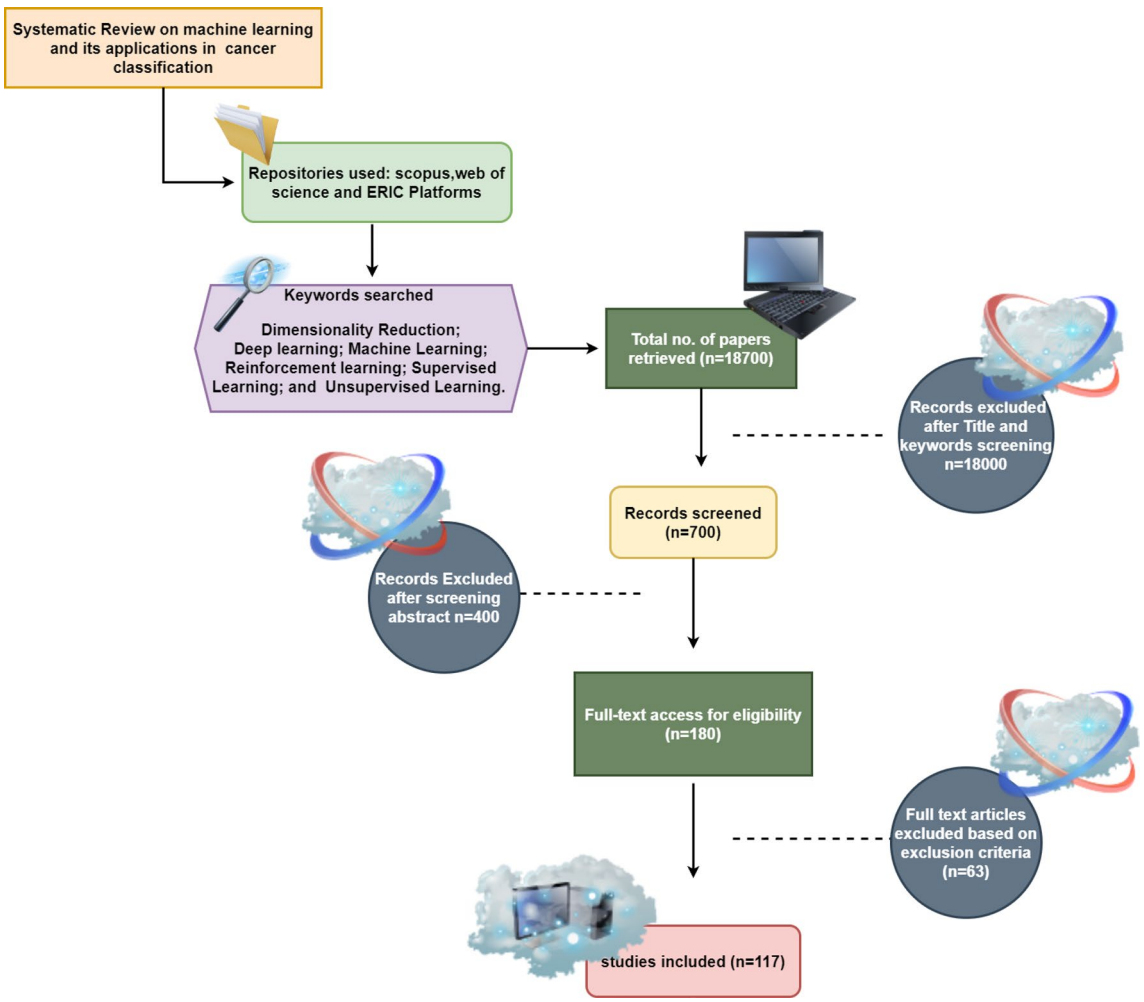

Fig. 3 Systematic approach for conducting a literature review

Table 2 Features of Machine Learning in Medical Field [2, 3, 5, 7, 9].

Feature	Description	New technologies
Medical Imaging	ML algorithms assist in analyzing medical images	Deep learning, convolutional neural networks (CNNs)
Diagnostics	ML aids in diagnosing diseases and conditions	Support Vector Machines (SVMs), decision trees, random forests
Drug Discovery	ML helps identify potential drug candidates	Generative adversarial networks (GANs), reinforcement learning
Personalized Medicine	ML enables tailored treatment plans for individuals	Predictive analytics, natural language processing (NLP)
Electronic Health Records (EHRs)	ML improves data management and analysis of patient records	Natural language processing (NLP), recurrent neural networks (RNNs)
Patient Monitoring	ML assists in real-time patient monitoring and early detection	Internet of Things (IoT), wearable devices, sensor technology
Precision Medicine	ML facilitates genetic analysis for personalized treatments	Genomic sequencing, gene expression profiling, bioinformatics
Risk Prediction	ML predicts patient outcomes and identifies high-risk cases	Bayesian networks, logistic regression, gradient boosting
Virtual Assistants	ML powers intelligent virtual assistants for healthcare tasks	Natural language processing (NLP), chatbots, voice recognition

1.4.1 Outline of the Paper

This research paper is divided into six sections that make it easy to follow and understand. In the beginning Sect. 1, the paper talks about why it's important to use machine learning in diagnosing and classifying cancer. It also explains the different ways researchers analyze cancer data and how they conducted their review. The Sect. 2 discusses the different types of machine learning systems, explaining how they work. Then, the Sect. 3 of the paper looks at the advantages and disadvantages of using machine learning in cancer research in the third section. The Sect. 4 is all about how machine learning is applied to classify different types of cancer. Section 5 goes into more detail about machine learning, discussing its ins and outs. Finally, the Sect. 6 concludes the paper and suggests ideas for future research and how to apply the findings. Overall, this paper is a fascinating read that shows how machine learning can help with cancer diagnosis and opens up possibilities for future advancements. Section eight discusses the future of machine learning, including its potential and challenges, and the future direction of research in this field. Finally, the conclusion summarizes the paper, explores the implications of the findings, and provides suggestions for future research.

2 Different Categories of Machine Learning Systems

Purpose: Discussing different machine learning categories helps to understand their applications and choose the most suitable approach for specific problems. It also drives research, promotes interdisciplinary learning, and aids in addressing ethical concerns.

2.1 Learning with Guidance

Learning with Guidance (Supervised Learning).

When it comes to supervised learning, we furnish a model that has access to train dataset with labelled instances. The algorithm then leverages this labelled data will be used to acquire knowledge about the relationship between input variables (feature) and output (labels) using a specific algorithm for learning. During training, the algorithm attempts to identify the underlying pattern or the relationship between the variables used as input and output, by minimizing the error or loss function. The ultimate objective is to generalize this connection to previously unknown data, i.e., to generate accurate predictions on previously unseen input data. Supervised learning finds utility in diverse domains, such as image classification, audio identification, and natural language processing, and predictive modeling. In image classification, for example, the input variables may be the pixel values of a photograph, while the output variable could be the associated label or class of the object in the image. The beauty of supervised learning is its ability to learn from labeled data and generalize to new, unseen examples. This is achieved through the application of different learning algorithms, such as decision trees, neural networks, support vector machines, and linear regression, to name a few. These algorithms are capable of handling large and complex datasets, and can be trained on a variety of input and output data types, including numerical, categorical, and textual data [35].

In summary, supervised Training is a strong technique that enables robots to understand from labelled data and

make accurate predictions on fresh, previously unseen samples. It serves as a fundamental technique in machine learning and finds extensive usage in a wide range of practical scenarios applications (Fig. 4).

There are two types of supervised learning:

2.1.1 Classification

Classification constitutes a fundamental undertaking in supervised learning, where the objective is to train a model to forecast the class designation of a given input by considering its distinctive attributes. A common example of this is a spam filter, which is designed to classify incoming emails as either spam or legitimate. To train a spam filter, the model is fed a large dataset of example emails along with their corresponding labels, which indicate whether each email is spam or not. The model then uses these examples to learn patterns and features that are indicative of spam, such as specific keywords, phrases, or formatting. After trained, the model may be used to categorize data new incoming emails by analyzing their features and predicting their label. A well-trained spam filter can significantly improve the user experience by reducing the amount of unwanted or malicious emails that are received, while also ensuring that legitimate emails are not mistakenly flagged as spam [37].

Cancer classification is the process of categorizing different types of cancers based on their characteristics, such as the site of origin, histological features, genetic mutations, and clinical behavior. Accurate classification of cancer plays a significant part in ensuring precise diagnosis, treatment planning, and predicting patient outcomes [38].

Here are some common methods used for cancer classification:

Histopathology: This involves examining cancerous tissue samples under a microscope to analyze their cellular and tissue characteristics. Pathologists classify tumors based on their morphology, including cell type, degree of differentiation, and tissue architecture [39].

Immunohistochemistry (IHC): IHC uses specific antibodies to identify and classify cancer cells by considering the presence or absence of particular proteins, cancer can be classified effectively or markers. It helps determine the origin of the tumor and can provide information about potential therapeutic targets [40].

Molecular profiling: This approach involves analyzing the genetic and molecular alterations within cancer cells. Techniques such as DNA sequencing, gene expression profiling, and proteomics can identify specific mutations, gene amplifications, or changes in gene expression patterns. Molecular profiling can help classify tumors into different subtypes and guide targeted therapies [41].

Imaging techniques: Imaging modalities like Computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET) scans offer valuable insights into the position, dimensions, and scope of tumors, furnishing crucial information in the process. Radiologists use imaging findings to classify cancers and determine the stage of the disease [42].

Classification models: Machine learning and artificial intelligence algorithms can be trained using large datasets to develop predictive models for cancer classification. These models can incorporate various data types, such as clinical information, imaging data, and molecular profiles, to classify tumors and assist in diagnosis and treatment decisions [43].

To further illustrate the concept of classification, let's consider a few more examples:

Image Classification: In image classification, a model undergoes training to anticipate or forecast the category of a given image based on its visual features. For instance, a model can be trained to classify images of animals, such as cats and dogs. The model is fed an extensive collection of labeled images of cats and dogs, which it uses to learn features such as fur texture, shape, and size. Once trained, the model may be used to categorize fresh dog and cat photographs [44].

Sentiment analysis: Sentiment analysis involves classifying text data based on the sentiment expressed in it. For example, a model can be trained to predict whether a given movie review is positive or negative. The model is fed a large dataset of labeled movie reviews, which it uses to learn patterns in the language that are associated with positive or negative sentiment. The model, once trained, may be used to categorize fresh reviews based on their sentiment [45].

Fraud Detection: In fraud detection, a model is trained to identify fraudulent transactions in a given dataset. For instance, a bank can train a model to detect fraudulent credit card transactions by analyzing patterns in the data, such as unusual spending behavior or geographical location. Once trained, the model can be used to flag potentially fraudulent transactions in real-time [46].

In all of these examples, the key to successful classification is the model's capacity to extract significant patterns and features from labeled data is enhanced as it undergoes

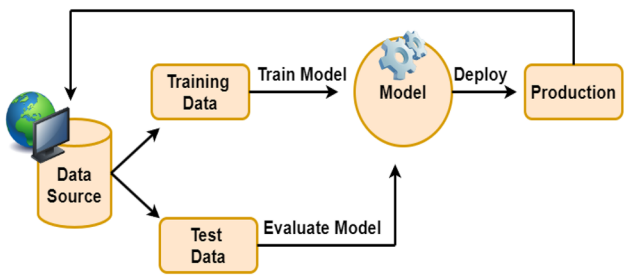

Fig. 4 Supervised Learning workflow [36]

the training phase. These patterns and traits may then be utilized to produce precise predictions based on fresh, previously unknown data.

Regression: Regression is a popular approach for predicting the significance of a numerical attribute using a collection of input characteristics or predictors. For example, estimating the price of an automobile based on factors such as mileage, age, brand, and model is a common use of regression. To train a regression model for this task, we need a set of examples that include both the labels (i.e., the prices of the cars) and their predictors (i.e., the mileage, age, brand, and model). Once the model has been trained, it can accurately predict the cost of a vehicle determined by its characteristics [47].

Regression can be applied to various domains, such as healthcare, finance, transportation, marketing, and education. In healthcare, for example, based on their medical history and other criteria, regression can be used to forecast the chance of a patient getting an illness. In finance, regression analysis can be employed to make predictions about stock prices by considering various market indicators. In transportation, regression can be used to predict the duration required for a vehicle to cover a specific distance based on various factors such as weather conditions, traffic, and road type [48].

Overall, regression is a valuable tool for predicting numerical values based on input features, and it has numerous applications in different domains, including the car industry, where it can help car dealerships estimate the value of trade-ins or buyers compare the prices of different cars based on their features [49].

2.1.2 Some of the Most Important Supervised Algorithms

KNN: Machine learning techniques such as K-Nearest Neighbours (KNN) are used for both classification and regression problems. Its operation is based on selecting the "k" closest neighbors to an object in the training dataset, where "k" is a positive integer of the user's choosing. The algorithm determines the distance between the object and the rest of the dataset's objects, and then selects the "k" nearest neighbors based on their proximity. After identifying the "k" nearest neighbors, the object is classified by assigning it to the class that is most prevalent among its neighbors. In regression tasks, the algorithm predicts the average value of the "k" nearest neighbors as the anticipated value for the object [50].

Overall, KNN is a straightforward algorithm that can be easily implemented and utilized in a diverse range of applications in machine learning.

LR: Logistic Regression is a widely adopted statistical technique employed to analyze datasets that encompass one or more independent variables, which have the potential to influence the outcome. It is widely used for binary classification tasks in which one of two probable outcomes must be predicted. For example, determining whether a patient is healthy or sick, or whether a candidate will pass or fail an exam. The relationship between the independent variables and a logistic function is used to describe the dependent variable, which converts the projected output to a number between 0 and 1. This value represents the apparently that the dependent variable is a positive outcome. Maximum likelihood estimation is used to train the logistic regression model, The procedure entails finding the parameter values that maximize the likelihood of observing the training data. Once trained, the model can be employed to predict outcomes for new data points. This is done by inputting the independent variable values into the logistic function and calculating the estimated probability of obtaining a positive outcome [51].

LR: Linear Regression is a statistical method for building a model that represents the relationship between one or more independent variables and one or more dependent variables. The main objective is to predict the value of the dependent variable based on the values of the independent variables. Linear regression is widely employed across various disciplines to comprehend the association between variables and make predictions. The model is built by finding a linear equation that best fits the data, enabling estimation of the dependent variable when provided with the values of the independent variables [52].

Finding the best line to depict the relationship between the independent and dependent variables is the goal of linear regression. The equation $Y = ax + b$, where Y stands for the dependent variable, X for the independent variable, a for the line's slope, and b for the intercept, represents the line of best fit. It is possible to use linear regression for both simple linear regression, which involves a single independent variable, and multiple linear regression, which encompasses multiple independent variables. The linear regression model's parameters are estimated using a number of strategies, including the ordinary least squares method and gradient descent. Once the parameters have been computed, the linear regression model can be utilized to predict new data points by inputting the values of the independent variables into the equation and computing the anticipated or expected value of the dependent variable [53].

SVM: A supervised machine learning method used for classification and regression analysis is the support vector machine (SVM). SVM is a powerful and widely used method because it can handle data that cannot be linearly divided by translating it into a higher-dimensional space with a linear boundary to separate the classes. In SVM, the purpose is to select the optimal hyperplane for classifying the data. The hyperplane is defined as the line or plane that is closest to the data each class's points. The minimum distance between

data points, known as the margin, is a crucial concept in SVM (Support Vector Machines). By identifying the hyperplane that is farthest from the nearest data points of each class, called support vectors, the margin is expanded. This technique allows SVM to handle both linear and nonlinear data by utilizing the kernel method, which transforms the data into a higher-dimensional space. A linear border can be created in this modified space to separate the classes. After locating the hyperplane, SVM can be applied to categorize new data points by assessing which side of the hyperplane they lie on. If a new data point is positioned on one side of the hyperplane, it will be assigned to a specific class. Conversely, if it falls on the other side, it will be assigned to a different class [54].

DT: The Decision Tree is a widely used algorithm utilized in machine learning and artificial intelligence to address classification and regression tasks. It takes the shape of a tree, describing a series of decisions and their accompanying results. Within a decision tree, an internal node signifies an attribute test, a branch signifies the result of the test, and a leaf node represents a prediction or class label. The construction of the tree involves recursively dividing the data into smaller groups using attribute values and selecting the split that generates the most consistent subsets (i.e., subsets with the highest proportion of instances belonging to the same class) [55].

The process of constructing a decision tree is iterated until certain stopping criteria are met, such as exceeding a certain level of occurrences in a leaf or a maximum tree depth. The entire tree can be used to produce predictions for new data points by following the path from the root to a leaf node and basing predictions on the outcomes of the tests at each internal node. Because it can handle both numerical and categorical data and is easily interpretable, decision trees are frequently employed. They are also relatively fast to and can handle large datasets. However, they are prone to overfitting, particularly when the tree becomes too deep, and can benefit from pruning or ensemble methods, such as random forests [56].

RF: A classification and regression ensemble learning system is called Random Forest. It is a kind of decision tree method that generates numerous decision trees and combines their prediction to obtain a more reliable and accurate outcome. Each tree in a random forest is produced using a random subset of the data and a random subset of the attributes. This process is repeated multiple times to build multiple trees, and the estimates made by each tree individually are combined through a majority vote (for classification) or average (for regression). Random forests offer several advantages over an individual decision tree. By combining the forecasts from many trees, they typically lessen overfitting and increase model accuracy, and increase its stability. They can also manage complex, non-linear connections between the attributes and the desired outcome, and are capable of handling a combination of numerical and category features. One of the key benefits of random forests is that they are simple to understand, as the feature importance can be estimated from the trees, and decision trees themselves are relatively easy to understand. Despite these advantages, random forests computation costs may be high to train, particularly for significant databases and large numbers of trees, and they It might not be the optimal option for datasets that possess a high number of dimensional features. However, for many problems, they provide a good balance between accuracy and interpretability, making them a popular choice for many machine learning practitioners [57].

NN: A machine learning system called a neural network is modelled after the way the human brain is structured and functions. It resembles a kind of synthetic neural network made up of several linked nodes, or artificial neurons, organized into layers.

The information that is transferred through by means of numerous layers in a neural network, where each layer performs a mathematical operation on the data and the result of each a layer is connected to the subsequent layer in the sequence. The model's predictions are produced by the output layer called the final layer. During training, the model's parameters, referred to as weights and biases, are adjusted to minimize the disparity between the predicted and observed outputs. Neural networks excel at handling intricate and non-linear connections between input and output variables, making them applicable to various tasks like image classification, natural language processing, and time series forecasting. Furthermore, they can be combined with other machine learning techniques, like decision trees, to create hybrid models that leverage the strengths of multiple algorithms. Despite their capabilities, neural networks can be challenging to design, train, and interpret, particularly for large and complex models, and can require significant computational resources. Additionally, they can be prone to overfitting, and may require regularization techniques, to avoid this, employ measures such as dropout or early termination. However, breakthroughs in processing power and novel designs, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have resulted in considerable gains in neural network performance and interpretability in recent years [58].

2.1.2.1 Unsupervised Learning Unsupervised learning is a type of machine learning where algorithms are trained using datasets that lack explicit labels or annotations allowing models to learn from patterns and correlations in the data without the requirement for prior supervision or predetermined classifications. Unlike supervised learning, where the algorithms require both input and output data for training, unsupervised learning focuses only on input data, making

it a practical and economical method for analyzing large and complex data sets. Clustering, anomaly detection, and dimensionality reduction are three prominent unsupervised learning approaches. Clustering algorithms bring together similar data points to find underlying relationships and patterns in the data. Anomaly detection is the process of discovering data points that vary considerably from the norm, whereas dimensionality reduction is the process of simplifying data by lowering its complexity and the number of variables [59].

Unsupervised learning finds extensive application across diverse domains, such as computer vision, natural language processing, and data mining. It has many practical applications, such as image and speech recognition, fraud detection, customer segmentation, and recommendation systems. In addition, unsupervised learning has proven useful in scientific research, such as clustering and identifying patterns in genetic data and analyzing social network structures. One of the main advantages of unsupervised learning is that it can be used to discover previously unknown patterns and relationships in data, without relying on human-defined categories or labels. This makes it ideal for applications where the data is too complex or too large to be manually labeled. Additionally, unsupervised learning can unsupervised learning offers valuable insights into the underlying structure and behavior of the data, enabling researchers to better understand the underlying linkages and trends in order to make more educated decisions. In conclusion, unsupervised learning is a powerful technique that has numerous applications across a broad range of fields. By leveraging unsupervised learning, researchers and practitioners can gain insights into large and complex data sets, allowing them to identify previously unknown patterns and relationships that can inform decision-making and drive innovation [60] (Fig. 5).

Unsupervised learning is subdivided into two groups.

2.2 Learning Without Guidance

2.2.1 Clustering

Clustering is a statistical technique that is widely employed in many different domains, including as bioinformatics, pattern recognition, machine learning, and many more. Its major purpose is to organise a collection of items so that objects in the same group have comparable properties to those in other groups. Clustering includes the use of different algorithms, each with its own idea of what constitutes a cluster and how to discover them quickly. Clusters are commonly defined as dense regions of data space, groupings with minimal distances between cluster members, intervals, or unique statistical distributions. Clustering might be considered a multi-objective optimisation problem, with the proper method and parameter settings determined by the dataset and intended application of the results. It is not an automatic process, and instead requires an interactive multi-objective optimization process of knowledge discovery that involves trials and errors. To achieve the desired outcomes, it is often necessary to modify the data preparation and model parameters. In summary, clustering is a valuable exploratory data analysis technique that enables efficient grouping of data for various applications, but it requires careful consideration of the data and a thorough understanding of the algorithms involved to produce useful results [62, 63].

2.2.1.1 Widely used Clustering Algorithms

I. K-means: K-means clustering technique separates a given collection of data points into K groups based on their similarity. Based on the input data, this unsupervised learning technique determines the best cluster centroids. The objective of the process is to reduce the total sum of squared distances between individual data points and the centroid of the cluster [64]. Here are the steps involved in the K-means algorithm:

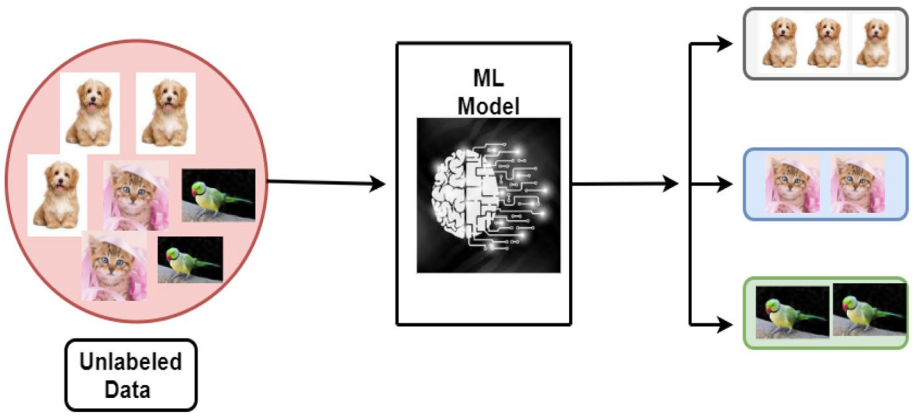

Fig. 5 Unsupervised learning [61]

1. Choose K initial cluster centroids at random from the data points.
2. Assign each data point to the cluster that has the closest centroid to it.
3. Recalculate each cluster's centroids by calculating the average of the data points within that cluster.
4. Continue with steps 2 and 3 iteratively until either the cluster assignments remain unchanged or the maximum number of iterations is reached.

The K-means algorithm is designed to reach a minimum of the goal function on a local scale through a series of iterations. It is crucial to consider that the initial selection of centroids can impact the final clustering outcome, leading to different cluster assignments and local optima. To address this issue, running the K-means algorithm with multiple initial centroid selections can yield the best possible clustering result. K-means has found wide application in data science, machine learning, and computer vision for clustering and image segmentation due to its simplicity, scalability, and efficiency. However, it comes with limitations as well as the requirement to indicate the number of clusters beforehand, sensitivity to initial centroid selection, and the assumption of a spherical cluster shape and equal cluster size. Numerous extensions and variants have been introduced, including hierarchical K-means, fuzzy K-means, and spectral clustering, to improve K-means performance in diverse scenarios and address some of its limitations [65].

II. Hierarchical clustering.

Hierarchical clustering is an approach used to arrange data points into hierarchical and tree-like structures. Initially, each data point is considered as an individual cluster, and subsequently, clusters that are closest to each other are merged iteratively until a single cluster remains. There are two kinds of hierarchical clustering techniques: agglomerative and divisive. Each data point is the starting point for agglomerative clustering at each stage, the algorithm treats each cluster individually and combines the two nearest clusters. On the other hand, divisive clustering initiates with all data points in a single cluster and progressively divides them into two clusters at each stage. The distance between two clusters is established that utilize distance metrics, such as Euclidean distance, Manhattan distance, or correlation distance. Which is chosen based on the specific data and problem domain. A dendrogram is produced by hierarchical clustering, which is a tree-like diagram that depicts the order of cluster mergers. Each leaf node in the dendrogram corresponds to a data point, while each internal node represents a merged cluster. The distance between the merged clusters is represented by the height of each internal node. The figure of clusters to choose can be determined by examining the dendrogram. The cut-off point, the number of clusters corresponds to where the dendrogram is terminated. The cut-off point is determined by the desired level of granularity and the problem domain [66].

Hierarchical clustering offers several benefits, including its ability to handle non-convex clusters, the freedom of not having to predefine the desired number of clusters in advance, and the ease of interpretation provided by the dendrogram output. However, the agglomerative method can be resource-intensive in terms of computation, particularly when dealing with large datasets, and the choice of distance metric and linkage method can impact the clustering result. There are several linkage methods available, such as single, complete, average, and Ward's method, each with its strengths and weaknesses, making it important to select the suitable linkage method based on the characteristics of the data and the specific problem domain. Hierarchical clustering finds extensive application in diverse fields, including biology, social science, and computer science, for clustering and classification tasks [67].

III. Density-based clustering.

Density-based clustering is a method for grouping data points in a particular data space based on their density. The technique identifies high-density regions as clusters and distinguishes them from areas of low density. The clusters produced can have diverse shapes and sizes and do not need to be predetermined [68]. DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is widely recognized as the most frequently used density-based clustering method.

The process of the algorithm is as follows:

1. Locate all the neighboring data points within a radius, epsilon, for each data point.
2. Designate a data point as a core point if it has a minimum number of neighboring points; otherwise, mark it as a noise point.
3. For each core point and its neighbors, recursively identify all the connected points within epsilon and categorize them as part of the same cluster.
4. Repeat steps 1 to 3 until all the data points have been visited.

DBSCAN produces a collection of clusters and noise points. The shapes and sizes of clusters are determined by the connected components of core points and their neighbors, and the term "noise points" refers to Data points that are not assigned to any cluster. DBSCAN has several advantages, including the ability to handle non-convex clusters, the lack of a requirement to define the number of clusters earlier, and the capability to tolerate outliers and noise. However, it also has some shortcomings, such as its sensitivity to the choice of distance metric and epsilon parameter and its difficulty in handling clusters with varying densities.

To overcome these limitations, various modifications and alternatives to DBSCAN have been developed, such as Clustering Structure Ordering Points (OPTICS), Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) and Density-Based Clustering of Applications with Noise are two clustering techniques that focus on identifying clusters in data while also considering noisy or outlier points. Using a Localized Gaussian Mixture Model (DENCLUE), each designed to improve performance under different circumstances. Density-based clustering is a prominent approach for clustering and classification in a variety of fields, comprised of computer vision, image segmentation, and anomaly detection [69].

2.2.2 Dimensionality Reduction

Dimensionality reduction is a method of converting complex, high-dimensional data into a simpler form that preserves the most relevant characteristics. This approach attempts to lower the amount of variables while preserving the data's underlying structure and connections. The resulting reduced representation should ideally have the same dimensionality as the data, which refers to the fewest parameters necessary to explain its attributes. Dimensionality reduction is an important technique in many industries since high-dimensional data can be difficult to store, handle, and analyze [70]. Data visualization can benefit from dimensionality reduction, which minimises the number of dimensions, categorization, and compression, among other things. This technique enables researchers and practitioners to gain insights from complex data sets that would otherwise be difficult or impossible to analyse [71]. However, it is important to use dimensionality reduction methods responsibly and to avoid any potential issues related to plagiarism by giving credit to the original authors of any related works [72, 73].

A technique called "dimensionality reduction" (DR) reduces the number of input variables in a dataset before employing machine learning models. It can be performed through either feature extraction or feature selection. Feature extraction reduces the size of the original dataset by deleting redundant and unnecessary characteristics, while preserving the maximum amount of information. Alternatively, the feature selection algorithm finds the most pertinent subset of characteristics from the input data that are relevant to the given problem. Employing the right DR approach will help you save time and effort when choosing and extracting important features for analysis [74]. There are several Dimensionality Reduction Techniques (DRTs) that may be used to shorten calculation time and make better use of computer resources. These strategies can be used during the preprocessing stage, prior to data analysis and machine learning model creation. Nevertheless, choosing the best DRT might be difficult because each approach was designed to preserve specific elements of the original data. Thus, a specific DRT may be suitable for some types of data or applications, but not appropriate for others. Additionally, some DRTs may be created with limitations that limit their scope and use [75]. In conclusion model, DRTs offer an efficient way to decrease the number of input variables prior to employing machine learning models, feature selection or dimensionality reduction techniques are commonly employed. However, it is crucial to carefully choose the appropriate dimensionality reduction technique based on the data type and the specific application in order to achieve optimal results (Fig. 6).

2.2.2.1 Feature Selection Approach for Dimensionality Reduction Feature selection plays a vital role in the context of various disciplines such as pattern recognition, data mining, and statistical analysis. It involves identifying and selecting the most relevant features from a dataset while eliminating unnecessary or redundant information that can lead to biases or inaccurate models. Feature selection is particularly important when creating models for classification, regression, or clustering tasks. One of the key benefits of feature selection is that it makes it easier to visualize and analyse complex datasets, leading to a more accurate understanding of the underlying patterns and relationships. Additionally, feature selection can result in more compact models that are easier to interpret and have superior generalization capabilities. As a result, feature selection has become an increasingly popular area of research, with numerous methods and techniques developed throughout the last many decades to address the various dispute involved. Overall, effective feature selection is essential for achieving accurate and reliable results in many different fields of study [77].

Feature selection strategies can be classified into three types depending on the availability of labeled data in the dataset: supervised, semi-supervised, and unsupervised. In

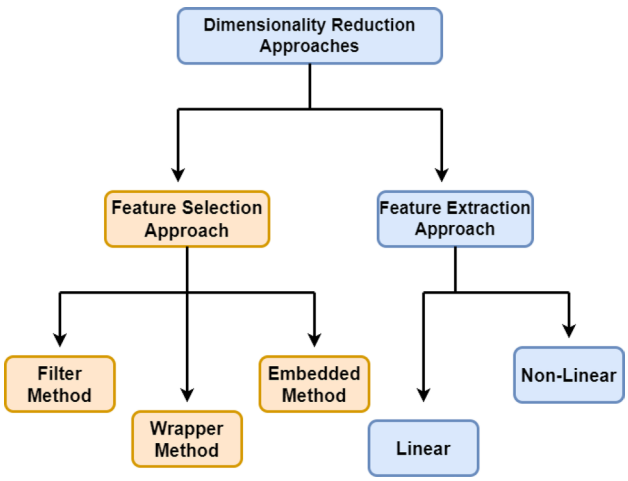

Fig. 6 Overview of Dimensionality Reduction [76]

supervised feature selection, labeled data is necessary to identify and choose relevant features. Labels can be in the form of categories, ordered values, or real values, and must be assigned to each object in the dataset. Semi-supervised approaches may only require labels for some objects. However, to find essential characteristics, unsupervised feature selection algorithms do not rely on labeled data. Instead, they use statistical and mathematical methods to analyze the dataset and identify relevant features based on their characteristics and correlations with other features [78]. Numerous feature selection in recent years, numerous techniques have been devised primarily for supervised classification problems. However, due to recent technological advances and an abundance of unlabelled data in unsupervised feature selection (UFS) techniques have garnered considerable attention in the scientific community, particularly in various applications like text mining, bioinformatics, image retrieval, analysis of social media and intrusion detection. Moreover, UFS techniques offer two significant advantages over supervised techniques. First, they are objective and do not rely on prior knowledge, making them suitable for handling new classes of data. Second, they can aid in lowering the danger of data overfitting, which is a common problem with supervised feature selection techniques [79].

Three major methods of feature selection can be distinguished.

- Filter methods are a category of feature selection technique that selects based on the most significant aspects solely on their characteristics within the dataset. These techniques do not utilize clustering algorithms to guide the search for important features. Instead, filter techniques evaluate each feature's inherent features to determine its relevance to the target variable. One of the primary advantages one notable advantage of filter methods is their speed and scalability. Since they do not rely on complex algorithms or iterative processes, filter methods can analyse large datasets quickly and efficiently. Additionally, may be used on a variety of datasets, as well as those with high-dimensional features or large numbers of observations. Filter methods employ diverse statistical or mathematical techniques to assess the significance or importance of each feature. Some common approaches include correlation analysis, mutual information, chi-square test, and information gain. To evaluate relevance, these approaches examine the connection between the target variable and the characteristic. Correlation analysis, for example, checks linear connection refers to the correlation or relationship between a characteristic and the goal variable, while mutual information measures the extent of dependency between the two variables. Filter methods are widely used in various applications, including bioinformatics, text mining, and image analysis. They provide a simple and effective means of selecting relevant features, it might amplify the efficacy and accuracy of machine learning algorithms. However, filter methods do have limitations, such as the potential for irrelevant or redundant features to be retained, which can negatively impact model performance [80].
- Wrapper approaches are a popular method used in machine learning to analyse feature subsets by utilizing the findings of a specific clustering algorithm. This method's main goal is to find feature subsets that can enhance the calibre of the outcomes produced by the grouping method employed in the collection phase. One of the main advantages of wrapper approaches is that they are designed to be highly targeted and specific, resulting in improved accuracy and precision when compared to other approaches. However, this precision often comes at a cost, as wrapper approaches can be computationally expensive and may only be compatible with certain clustering algorithms. Despite their limitations, wrapper approaches are extensively used in many fields, including bioinformatics, image recognition, and natural language processing, where the need for precise feature selection is critical. By combining the findings of clustering algorithms with wrapper approaches, researchers and practitioners can uncover critical insights and improve the overall quality of their data analysis [81].

In conclusion, wrapper approaches are a powerful tool for analysing feature subsets and improving the accuracy and precision of clustering algorithms. While they do have their limitations, their effectiveness in targeted scenarios makes them a valuable asset to any machine learning practitioner's toolbox**.**

- Embedded methods in machine learning goal in order to balance effectiveness and efficiency when selecting relevant features for a given objective task. To accomplish their objectives, these strategies take advantage of the benefits of both filter and wrapper approaches are utilized. Filter methods use statistical techniques to identify relevant features by measuring their correlation with the output variable. These methods are efficient in terms of computation and can quickly process although they might not always be able to capture the complex connections between the goal variable and the characteristics. In contrast, wrapper approaches employ a trial-and-error methodology to evaluate the performance of a subset of qualities. While wrapper methods can capture complex feature interactions, they are computationally expensive and may overfit the data. Embedded methods attempt to strike a balance between these two approaches by integrating feature selection into the model-building process. These methods aim

to identify relevant features during the model-building process, As a result, the number of characteristics that must be examined is reduced, as is the danger of overfitting. One popular example of an embedded method is the Lasso algorithm, which uses regularization to shrink the coefficients of irrelevant features to zero. This approach not only identifies relevant features but also performs feature selection during the model-building process, resulting in a more efficient and effective model [82].

In conclusion, embedded methods in machine learning offer a compromise between filter and wrapper methods by integrating feature selection into the model-building process. These techniques seek to balance efficacy and efficiency, resulting in more accurate models that are less prone to overfitting.

2.2.2.2 Feature Extraction Approach for Dimensionality Reduction The method of feature extraction involves obtaining discriminatory data from a collection of samples. For the extraction of medically useful information from the textures, features must be computed. The traits, which may not be visually visible but are relevant to the diagnostic issue, can be thought of as supplements to the researchers' visual abilities. Effective and distinctive features are extracted using a variety of feature extraction techniques. Below is an explanation of a few feature extraction techniques [83]. In order to extract valuable information from images, many feature extraction approaches are used in the processing of medical images.

(a) Gray-Level Co-occurrence Matrix (GLCM): In medical image processing, GLCM is a popular texture analysis approach. It entails calculating the likelihood of pixel values co-occurring at particular pixel distances and directions in an image. The co-occurrence matrix is then used to compute various statistical measures such as contrast, correlation, energy, and homogeneity. These metrics can be utilised for classification or segmentation tasks as well as features to characterise the texture of a picture [84].
(b) Local Binary Patterns (LBP): LBP is a basic yet strong texture analysis tool. It entails comparing each pixel in a picture to its neighbours and assigning a binary value based on whether the neighbours have greater or lower values than the centre pixel. Each pixel in the picture goes through this procedure once again to produce a binary pattern. These patterns can then be used as features to describe the texture of an image [85].
(c) Gabor Wavelets: Gabor wavelets are a type of filter that is used to analyse the frequency and orientation content of an image. They are particularly useful for analysing texture because they can capture both the fine and coarse details of an image. Gabor wavelets can be used to extract features such as mean amplitude, mean frequency, and classification or segmentation task-specific orientation [86]
(d) Histogram of Oriented Gradients (HOG): The HOG feature extraction technique that involves computing the gradient magnitude and direction of an image and then grouping these gradients into histograms based on their orientation. These histograms can then be used as features to describe the texture of an image. HOG has been shown to be particularly effective for object detection and recognition tasks [87].
(e) Convolutional Neural Networks (CNN): A popular deep learning algorithm is CNNs, which are used for feature extraction from medical images. These networks are created to automatically recognise and extract characteristics from images by utilizing convolutional layers. Filters are applied using convolutional layers to an image, allowing the extraction of specific information and patterns. This enables CNNs to effectively capture complex spatial relationships and distinctive features in medical images. The output from the convolutional layers can then be used as features for classification or segmentation tasks. CNNs have been demonstrated to be extremely successful for a variety of medical image processing applications, including tumour identification and segmentation [88].
(f) Recurrent Neural Network (RNN): RNN is a kind of neural network that is designed to process consecutive data, where the output at each step is influenced by previous steps. It has feedback connections that allow information to persist across different time steps, making it appropriate for problems like speech recognition, language modelling, and time series analysis. RNNs have a recurrent hidden state that captures and updates information as new input is fed into the network. The vanishing gradient problem, which affects traditional RNNs, limits their ability to detect long-term dependencies [89].
(g) Long Short-Term Memory (LSTM): The LSTM represents a recurrent neural network developed to replace the vanishing gradient problem in typical RNNs. It includes a memory cell that enables the network to selectively store data, read, and write information over long sequences. LSTMs have gated mechanisms, comprising input, forget, and output gates that regulate information flow and enable the network to store pertinent data over time. For modelling long-term dependencies in sequential data, LSTMs are highly helpful. They are frequently employed in time-series prediction issues including machine translation, audio recognition, and natural language processing [90].

2.2.2.3 Linear and Non-Linear Approaches of Feature Extraction The practise of lowering the number of variables in a dataset while maintaining important data and linkages is known as dimensionality reduction. It is possible to reduce dimension in two ways: linearly and non-linearly. In order to create a new set of variables that retains the majority of the crucial information inherent in the original data, linear dimensionality reduction techniques combine the original variables in linear combinations. The method of linear dimensionality reduction known as principal component analysis (PCA) determines the directions of the data that account for the most variance. On the other hand, non-linear dimensionality reduction techniques employ non-linear transformations of the original variables to generate a new set of variables that effectively capture the significant details within the data. Examples of non-linear dimensionality reduction techniques include Iso-map, t-SNE, and UMAP. These techniques are useful when the relationships between the variables in the data are non-linear [91].

Both linear and nonlinear dimensionality reduction solutions have advantages and disadvantages, and the methodology utilized is determined by the specific scenario and data type. Dimensionality reduction plays a critical role in various machine learning and data processing analysis applications. It offers several benefits, including noise reduction, enhanced visualization capabilities, and improved performance of other machine learning algorithms.

2.2.2.4 Feature Classification Algorithms The process of grouping features based on criteria that divide data into multiple classes is known as feature classification, and it uses a variety of techniques, including:

A. Support Vector Machine (SVM): Support Vector Machines (SVMs) are utilized as supervised learning algorithms in both classification and regression tasks. They are members of the generalised linear classification family. The capacity of SVM to maximise the geometric margin while simultaneously minimising the empirical classification error distinguishes it. SVM uses Maximum Margin Classifiers as a result. SVMs construct a maximum separation hyperplane and create two parallel hyperplanes on either side of the data-separating hyperplane. By mapping input vectors to a higher-dimensional space, SVM aims to find the hyperplane with the largest separation between the parallel hyperplanes, known as the separating hyperplane. According to the theory, increasing the distance or margin between these hyperplanes leads to a reduction in the classifier's generalization error [92].
B. Radial Basis Function (RBF): RBF is a classification method that uses nonlinear activation functions like sigmoidal and Gaussian Kernel for functional approximation and classification. As a result, the Gaussian function's response is positive for all values of x and it approaches zero at |x tends 0|. As its name implies, RBF is proven to be radically symmetric because it yields the same results for any input values coming from the kernel's centre [93].
C. K-Nearest Neighbour (KNN): The most fundamental approach for classification and regression on k-nearest neighbour data sets is known as KNN. In order to forecast how a fresh data set will be classified, the KNN model divides the provided data into a number of classes. Acting as a "clustering model". The majority decision of its k nearest neighbors is used as the basis for classification, it forecasts the membership of a class. Regression analysis uses the mean (average) of its k closest neighbours to represent the class. The Euclidean distance is used to calculate it [94].
D. Linear Discriminant Analysis (LDA): Linear discriminant analysis is a widely used technique for reducing linear dimensionality (LDA) for classification and pattern recognition applications. LDA seeks a linear combination of the variables in the data that best divides the data's multiple classes. Unlike PCA, which seeks to capture the maximum variance in the data, the objective of linear discriminant analysis (LDA) is to identify the directions that maximize the distinction between classes. LDA is particularly valuable when the classes in the data are well-separated and exhibit similar covariance matrices. In such cases, LDA can provide a better representation of the data for classification compared to other linear dimensionality reduction techniques like PCA [95].

In addition to its use for classification and pattern recognition, LDA is also used for data visualization, feature extraction, and for reducing the dimensionality of data for other machine learning algorithms. LDA is a rapid and computationally efficient algorithm that is frequently used in image and audio recognition, text classification, and bioinformatics applications.

2.2.2.5 Reinforcement Learning Reinforcement Learning (RL) is a method within the field of machine learning that empowers an entity, known as an agent, to acquire knowledge and improve its performance from its interactions with its environment in order to maximise a reward signalIn the context of reinforcement learning, an agent interacts within an environment to achieve a specific goal. The agent receives feedback in the form of rewards or penalties based on its actions and decisions. In response to this feedback, the agent adjusts its behavior with the objective of incrementally maximizing the cumulative reward over time [96].

Reinforcement learning has been utilised in a variety of areas such as robotics, games, banking, and transportation. RL has been used in robotics to train robots to perform tasks such as grasping and manipulation, whilst in gaming, it has been utilised to construct AI agents capable of playing games at a human-like level. In finance, RL has been used to optimize portfolio selection and algorithmic trading, and in transportation, it has been used to develop autonomous driving systems [97].

Reinforcement learning is a highly effective tool for tackling intricate problems where a clear mathematical formulation of the problem is not available. It is especially beneficial in situations where an agent needs to learn from experience and make decisions based on its current state and environment. However, reinforcement learning can be challenging to implement and requires a lot of data to train the agent effectively. Additionally, it can be difficult to design effective reward functions, and the agent's behavior may not always align with the desired outcome [98].

There are two primary types or approaches of reinforcement learning algorithms: value-based techniques and policy-based methods (Fig. 7).

2.3 Learning Through Interaction

Value-based approaches estimate the value of doing a certain action in a specific condition. The value of an action measures how advantageous it is for the agent to do that action in the present condition. The agent determines which actions to take by assessing the estimated values. Value-based methods include Q-Learning and SARSA [100].

Policy-based approaches, on the other hand, are concerned with estimating a policy directly, which is linking states to actions. The policy defines the optimal action to take in each state, and the agent uses the estimated policy to make decisions. Policy-based methods include REINFORCE and Proximal Policy Optimization (PPO) [101].

In addition to value-based and policy-based strategies, there are actor-critic methods, which incorporate the qualities of both. The evaluation of the value of an action in a given state and the selection of the optimal action to take are crucial tasks in decision-making processes, actor-critic approaches combine a value function and a policy function.

The selection of the reinforcement learning algorithm is dependent on the characteristics or nature of the problem at hand and complexity of the problem at hand and the data. In some cases, value-based methods may be more appropriate, while in others, policy-based methods may be more suitable. The method used is also determined by the computer capabilities available and the amount of data available for training.

2.3.1 Widely used Reinforcement learning algorithms

2.3.1.1 Q-learning Q-learning is a reinforcement learning technique employed to determine the optimal strategy for selecting actions within a Markov Decision Process (MDP) framework. In an MDP setting, an agent interacts with the environment, doing actions that result in new states and being rewarded appropriately. To maximise the overall accrued reward over time is the agent's goal. A Q-value function, which calculates the predicted utility of carrying out a certain action in a specific condition, is updated through the Q-learning process. The Q-value function, denoted as Q(s, a), represents the expected cumulative reward that an agent can achieve by taking action "a" in state "s" and then following the optimal course of action thereafter. The policy that maximizes the expected cumulative reward is considered the best action to take. The Q-value function is updated iteratively using the Bellman equation, which takes into account

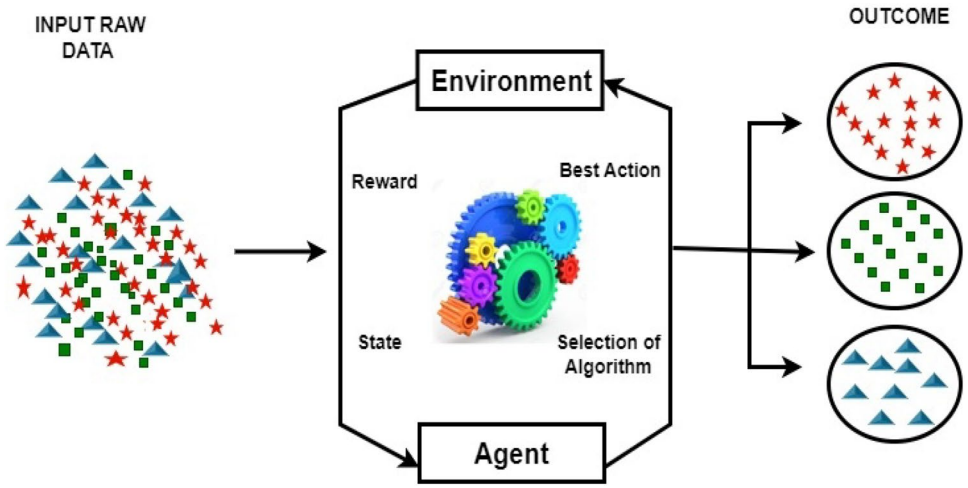

Fig. 7 Reinforcement learning workflow [99]

the immediate reward, the discounted future rewards, and the Q-values of the next state-action pairs.

$$Q(\mathrm{s}, \mathrm{a}) \;=\; \mathrm{r} \;+\; \gamma \max\left(Q\left(\mathrm{s}', \mathrm{a}'\right)\right)$$

In the Q-learning process, the Q-value function is updated using the Bellman equation, which takes into account the reward 'r' obtained for performing action 'an' in state 's', The discount factor 'γ' is a parameter that determines the significance of future rewards, and the resulting state 's'. The term max (Q(s', a')) reflects the projected cumulative reward that the agent may expect to get in the next state "s" by choosing the action that maximises the Q-value function. Q-learning begins by arbitrarily initialising the Q-value function, and then iteratively updates the Q-values for every observed (s, a, r, s') transitions using the Bellman equation. Through iterative updates, the Q-value function progressively converges towards the optimal Q-values, which correspond to the actions that maximize the expected cumulative reward. By selecting the action that maximizes the Q-value function for a given state, the agent can determine the optimal policy to follow. Q-learning is a model-free technique, which means that prior knowledge of the MDP transition and reward functions is not required. It is extensively utilised in numerous applications, such as game playing, robotics, and control systems. However, Q-learning may suffer from slow convergence and high variance due to the stochastic nature of the environment, and various improvements have been proposed, such as SARSA and Deep Q-Network (DQN) [102].

2.3.1.2 SARSA (State-Action-Reward-State-Action) A reinforcement learning method called SARSA (State-Action-Reward-State-Action) was created for sequential environmental decision-making. It is an on-policy algorithm, which implies that when learning, it maintains the same policy while updating it. The goal of SARSA is to discover the best Q-values for each state-action pair that exists in the world. The predicted cumulative return that the agent will obtain by performing a certain action in a particular state is represented by the Q-value. Based on the observed reward, the agent's subsequent state, and the action they decide to take, SARSA modifies the Q-values. The Q-values are modified using the next update rule:

$$\mathrm{Q}\ (\mathrm{s}, \mathrm{a}) \;<\; -Q\,(\mathrm{s}, \mathrm{a}) \;+\; \mathrm{alpha} * \left(\mathrm{r} + \mathrm{gamma} * \mathrm{Q}\left(\mathrm{s}', \;\mathrm{a}'\right) - \mathrm{Q}\ (\mathrm{s}, \mathrm{a})\right)$$

The learning rate (alpha), the immediate reward (r) received by the agent, the discount factor (gamma) for future rewards, the next state (s'), and the subsequent action (a') decided upon by the agent are all factors in updating the Q-value for a state-action pair (s, a) in SARSA. The SARSA uses an epsilon-greedy approach to strike a balance between exploration and exploitation. The agent prioritises exploitation and chooses the action with the highest Q-value with a probability of 1-epsilon. However, with a probability of epsilon, the agent chooses a random action, promoting exploration. The value of epsilon is typically decreased over time to encourage the agent to rely more on its learned policy and explore less frequently. SARSA is appropriate for environments in which the agent's actions have a direct impact on the subsequent states, and When dealing with environments with continuous states, either a probability distribution across the action space is used, or the Q-values for all viable actions in that state are taken into account as inputs and outputs. Backpropagation and stochastic gradient descent techniques are used to train the network by reducing the mean squared error between the actual and expected Q-values or policies as determined by the Bellman equation. Value-based and policy-based deep reinforcement learning algorithms can be distinguished from one another. The goal of value-based methods like Deep Q-Networks (DQNs), which assess the predicted cumulative reward for each state-action pair, is to learn the best Q-value function. Actor-Critic algorithms, on the other hand, are policy-based techniques that concentrate on directly learning the best policy [103].

Actor-critic methods, which use deep neural networks to represent both the value function and the policy function, have made considerable strides recently in merging both approaches. Deep Deterministic Policy Gradients (DDPG) and Trust Region Policy Optimisation (TRPO), among others, are notable examples of these techniques. These methods make use of deep neural networks to simultaneously learn and optimise the policy and value functions, which improves task performance and flexibility for reinforcement learning tasks. The unstable nature of learning is one of the main issues with deep reinforcement learning, which can result from the non-stationary targets and the correlation between samples. To tackle this problem, numerous techniques have been developed such as experience replay, target networks, and batch normalization. Over all, deep Reinforcement learning has demonstrated significant promise in addressing difficult issues that were previously assumed to be beyond the reach of classic reinforcement learning approaches. However, it still requires careful tuning of hyper parameters and significant amounts of training data to achieve good results.

3 Pros and Cons of Machine Learning Systems

See Table 3.

Table 3 Summarizes the advantages and disadvantages of each type of machine learning system [12]

Type of machine learning	Pros	Cons
Supervised learning	High accuracy, clear objectives, well-understood evaluation metrics	Requires labeled data, limited generalization ability, susceptible to bias
Unsupervised learning	Can spot patterns and links that people would find difficult or impossible to discover, and can deal with unlabeled data	Difficulty in evaluating performance, may produce uninterpretable results
Reinforcement learning	Can learn from experience and improve over time, can handle complex tasks in dynamic environments	Requires significant computational resources and time, susceptible to unexpected or undesirable behaviors, difficulty in defining appropriate rewards

4 Application of Machine Learning in Cancer Classification

In a study, Sung-Bae Cho and Hong-Hee used three benchmark datasets to evaluate various traits and classifiers. Their goal was to objectively assess feature selection techniques and machine learning classifiers. The Leukaemia cancer dataset, the Colon cancer dataset, and the Lymphoma cancer dataset served as the study's benchmark datasets. To choose the features, they considered a number of factors, including the signal-to-noise ratio, information gain, mutual information, Euclidean distance, Pearson's and Spearman's correlation coefficients, and cosine coefficient. They used support vector machines, multi-layer perceptrons, k-nearest neighbours, and structure-adaptive self-organizing maps for the classification process. They also integrated classifiers to boost classification performance. According to the experimental findings, the ensemble of numerous basic classifiers offered the benchmark dataset's greatest recognition rate [104].

Using the Wisconsin Diagnostic Breast Cancer (WDBC) dataset, Nikita Rane and her fellow researchers conducted a study in which they examined six different machine learning methods. Naive Bayes (NB), Random Forest (RF), Artificial Neural Networks (ANN), Nearest Neighbour (KNN), Support Vector Machine (SVM), and Decision Tree (DT) were the algorithms they took into consideration in their study. This dataset was being produced from a digitized picture of an MRI scan. The training phase and testing phase of the dataset were being used to implement the machine learning algorithms. The website's backend was going to be built using the algorithm that was performing the best, and the model was going to categorize cancers as benign or malignant [105].

Epimack Michael and his coworkers have presented a computer-aided diagnostic (CAD) system that has the ability to automatically design an optimal algorithm. Out of the 185 possible qualities, they selected 13 features to train the machine learning models. To differentiate between cancerous and benign tumours, they applied five different machine learning classifiers. The results of the experiment showed that employing a tree-structured Parzen estimator with a machine learning classifier in Bayesian optimization for tenfold cross-validation yielded promising outcomes. Light GBM emerged as the top performer among the five classifiers, achieving an accuracy of 99.86%, precision of 100.0%, recall of 99.60%, and FI score of 99.80% [70].

A novel approach for classifying and segmenting skin lesions using image processing and machine learning was put forth by Javaid et al. For image segmentation, they employed contrast stretching and OTSU thresholding, and they retrieved features including GLCM, HOG, and colour identification. They used SMOTE sampling to address class imbalance and PCA to reduce dimensionality. They carried out feature selection and employed Random Forest, SVM, and Quadratic Discriminant classifiers for classification. For the ISIC-ISBI 2016 dataset, their suggested approach had an accuracy of 93.89% [106].

David A. Omondiagbe and colleagues developed a combination technique to identify breast cancer by lowering the complexity. The proposed method involved utilizing linear discriminant analysis (LDA) to reduce the feature set, followed by implementing Support Vector Machine using the reduced features. In terms of performance, the approach achieved an accuracy of 98.82%, sensitivity of 98.41%, specificity of 99.07%, and an area under the receiver operating characteristic curve of 0.9994 [107].

A novel technique for choosing genes from gene expression data was created by Guyon et al. utilizing Support Vector Machine techniques with Recursive Feature Elimination. In contrast to other approaches, their research demonstrated that the chosen genes produced higher classification performance and more condensed gene subsets [10].

Dr. Shahin Ali and associates developed a deep convolutional neural network (DCNN) to reliably differentiating normal from cancerous tumours skin lesions. During the preprocessing stage, the input images are filtered to remove noise and artefacts, normalised, and feature extraction is carried out to aid in correct classification. To enhance the quantity of images and improve categorization accuracy, data augmentation techniques are employed. The performance of the DCNN model is compared to several transfer learning

models such as AlexNet, ResNet, VGG-16, DenseNet, and MobileNet. The evaluation of the model's effectiveness is conducted using the HAM10000 dataset. The training accuracy of the model was determined to be 93.16%, while the testing accuracy reached 91.93% [89].

The goal of the study conducted by Nurul Amirah Mashudi et al. was to assess how well various machine learning algorithms performed for classifying benign and malignant breast cancers, including Support Vector Machine (SVM), Random Forest, and k-Nearest Neighbours (k-NN). The scientists also used ensemble methods and tenfold cross-validation to forecast breast cancer survival. The recommended methodologies were further tried utilising twofold, threefold, and fivefold cross-validation in order to get the highest accuracy rate practical was 91.93%. An accuracy rate of 70% was observed by the research [108].

The technique proposed by Khalil Maalmi et al. consists of two parts. Firstly, Association Rules are employed to eliminate unnecessary parts (AR). Second, a number of classifiers are used to differentiate entering tumours with AR, the feature space is divided into eight and four qualities instead of nine. Using the Wisconsin Breast Cancer Diagnostic (WBCD) dataset from the University of California Irvine machine learning repository, the performance of the proposed system is assessed during the test phase using a threefold cross-validation technique. The greatest classification accuracy for the Support Vector Machine (SVM) model with AR was 98.00% for eight features and 96.14% for four attributes [109].

An innovative method for classifying breast cancer dubbed DLXGB, developed by Xin Yu Liew, analyses histopathology pictures of breast cancer from the BreaKH is dataset using Deep Learning and extreme Gradient Boosting algorithms. Using pre-processing methods including data augmentation and stain normalization, a pre-trained DenseNet201 is used to learn image characteristics, which are then combined with a potent gradient boosting classifier. Adenosis, Fibroadenoma, Phyllodes Tumor, Tubular Adenoma, Ductal Carcinoma, Lobular Carcinoma, Mucinous Carcinoma, and Papillary Carcinoma are among the eight non-overlapping/overlapping categories that will be used to classify breast cancer histology images in addition to binary benign and malignant categories. Using the BreaKHis dataset, the suggested DLXGB approach outperformed previous studies with an accuracy of 97% for binary and multi-classification [110].

Support Vector Machine (SVM), Logistic Regression (LR), and Neural Network (NN) were three machine learning approaches used in a study by Kristoffersen et al. To differentiate between benign and malignant breast cancer, a research study was conducted using the breast cancer Wisconsin diagnostic (BCWD) dataset. The study aimed to evaluate various machine learning techniques by testing multiple models, each with its own unique parameter values. The performance of these models was assessed using a confusion matrix and k-fold cross-validation. The results obtained through the k-fold cross-validation method revealed that Support Vector Machine (SVM) outperformed Logistic Regression (LR) and Neural Networks (NN) in terms of classification accuracy, precision, recall, and specificity. However, when using train-test split validation, the Neural Network model achieved the highest accuracy at 99.4%, surpassing both SVM and LR [111].

The output layer of two novel hybrid CNN models developed by Duggani Keerthana and her associates uses an SVM classifier to categorise dermoscopy pictures as benign or malignant lesions. The suggested model utilizes two Convolutional Neural Network (CNN) models to extract characteristics of the data. The extracted features are then concatenated and fed into an SVM classifier for classification. To assess this model's performance, the predicted outcomes are compared to the labels assigned by a dermatological expert. This allows for an assessment of how well the model performs in classifying the data accurately [112].

Kosmia Loizidou et al. conducted a comprehensive study that reviewed recent research on the automated detection and/or classification of breast cancer in mammograms. Their study covered both traditional feature-based machine learning methods as well as deep learning approaches. The paper contrasts algorithms designed to identify and/or categorise microcalcifications and masses, two different forms of breast abnormalities, and utilizing sequential mammograms has been explored as a means to enhance the performance. The authors also discuss open access mammography datasets and show various FDA-approved CAD systems for the triage and detection of breast cancer in mammograms. The study closes by pointing up potential avenues for further research in this area. This comprehensive overview might serve as a field introduction and provide direction for upcoming research applications [113].

The classification of lung CT scans has undergone a new method of development by Ebtasam Ahmad Siddiqui et al. Their approach combines the use of Gabor filters in conjunction with an enhanced Deep Belief Network (E-DBN) that incorporates numerous categorization techniques. The Gaussian-Bernoulli (GB) and Bernoulli-Bernoulli (BB) RBMs make up the E-two DBN's cascaded RBMs. The authors were able to acquire the best performance parameters out of all the applicable approaches by using a support vector machine (SVM). The suggested model combines an SVM with an E-DBN to improve lung CT image classification's precision, sensitivity, specificity, F-1 score, false positive rate (FPR), false negative rate (FNR), and ROC curve. Three publicly accessible datasets, including the LUNA-16 and LIDC-IDRI datasets, were used to test and assess the suggested technique [114].

Table 4 Comparative Study Table for the Comparison of Different Machine-Learning Techniques [118–122]

Algorithms used	Accuracy/Precision	F1—score	References
Support vector machine (SVM), decision tree naïve bayes (NB) and k nearest neighbors (k-NN)	(97.13%)	–	[118]
ANN, KNN, and RF	87.23	–	[119]
support vector machine (SVM), K-nearest neighbors, random forests, artificial neural networks (ANNs) and logistic regression	–	98.57%,	[120]
Support vector machine (SVM), Naive Bayes (NB), Boosted Decision Tree (BDT), and Decision Forest (DF)	83.1	78%	[121]
k-nearest neighbor, support vector machines, random forest, logistic regression, linear regression, Naive Bayes, linear discrimination analysis, linear classification, multi-layer perceptron and deep neural network	97%	–	[122]

In order to diagnose breast cancer, Md. Mehedi Hassan et al. compare multiple machine learning models using various categorization techniques. They use techniques like correlation matrices, histograms, and data distribution for systematic data collection, preparation, transformation, and exploratory analysis. To determine the most crucial characteristics, they also use the Least Absolute Shrinkage and Selection Operator (LASSO) technique. For evaluation and analysis, the study uses the techniques Logistic Regression, K-Nearest Neighbours, Extreme Gradient Boosting, Gradient Boosting, Random Forest, Multilayer Perceptron, and Support Vector Machine. Notably, their findings demonstrate that Random Forest outperforms the LASSO method with a maximum accuracy of 90.68%. Furthermore, K-Nearest Neighbors achieves a recall of 98.80%, Multilayer Perceptron exhibits a precision of 92.50%, and Random Forest attains an F1 score of 94.60% [115].

For categorization objectives, Abdullah-Al Nahid et al. apply novel deep neural network (DNN) approaches led by structural and statistical data from biological breast cancer images (BreakHis dataset). To categorise the breast cancer photos, they suggest using a Convolutional Neural Network (CNN), a Long-Short-Term Memory (LSTM), or a mix of CNN and LSTM. Following the feature extraction step utilising the suggested DNN models, the decision-making stage makes use of Softmax and Support Vector Machine (SVM) layers. The experimental findings demonstrate the best precision of 96.00% on the 40 × dataset, the best F-Measure on both the 40 × and 100 × datasets, and the greatest accuracy of 91.00% on the 200 × dataset [116].

In a study, Mehedi Masud and associates analysed histopathological images to create a classification system that could discriminate between two benign and three malignant forms of lung and colon tissues, making a total of five different types. Their suggested framework produced encouraging outcomes, with the ability to identify malignant tissues with a maximum accuracy of 96.33%. The results indicate that this model may be used as an automated and trustworthy method for medical professionals to correctly diagnose various forms of colon and lung cancer [117].

From the literature survey we conclude that Determining the "best" machine learning algorithm for cancer classification depends on several factors like the specific characteristics of the dataset, the size of the dataset, the nature of the cancer classification problem, and the desired performance metrics. There is no one-size-fits-all answer as different algorithms may perform differently in various scenarios. However, some commonly regarded powerful algorithms for cancer classification tasks include:

1. Support Vector Machines (SVM): Capable of handling both linear and non-linear classification tasks, and effective for processing high-dimensional data.
2. Random Forests: Robust and versatile, multiple decision trees are combined in the ensemble learning method for improved accuracy.
3. Gradient Boosting methods (e.g., XGBoost, AdaBoost): Can effectively handle imbalanced datasets and often achieve high predictive performance.
4. Deep Learning models (e.g., Convolutional Neural Networks): Particularly useful for image-based cancer classification tasks, leveraging complex patterns and features (Table 4).

5 Discussion on Machine Learning

Overview of the Findings: The comprehensive study found that machine learning algorithms had promising results for classifying cancer. Numerous machine learning techniques, such as Support Vector Machines (SVM), Artificial Neural Networks (ANN), Random Forest, Decision Trees, and Deep Learning, have been used to categorise various types of cancer [14, 123–126].

The results of the study indicate that these techniques have high accuracy and can effectively differentiate between cancerous and non-cancerous tissues. Additionally, these techniques can also identify different subtypes

of cancer and predict the likelihood of cancer recurrence. The potential clinical applications of these techniques are also noteworthy. Machine learning algorithms can help in early cancer diagnosis, patient stratification, personalized treatment planning, and monitoring of cancer progression. These techniques can also aid in drug discovery and development. However, there are some limitations to the use of these techniques in cancer classification, such is the requirement for extensive and varied datasets, potential bias in data selection, and the interpretability of the models.

Overall, the systematic review suggests that machine learning techniques have great potential in cancer classification and can significantly enhance the detection and management of cancer.

5.1 Significance of the Findings

The findings of the systematic review have significant implications for cancer diagnosis, prognosis, and treatment. The application of machine learning techniques to the classification of cancer can lead to more precise and effective diagnosis as well as better patient outcomes. The ability of machine learning to analyze massive volumes of data rapidly and accurately is a key advantage it has over conventional approaches [127]. This can help in identifying subtle differences in cancer types, subtypes, and stages, which can be missed by human experts. Additionally, machine learning can also help in identifying potential biomarkers and drug targets [128], which can lead to the development of new and more effective cancer therapies. The use of machine learning in cancer classification can also improve patient stratification, allowing clinicians to provide personalized treatment plans based on individual patient characteristics. This can help in reducing treatment-related adverse effects and improving treatment efficacy. However, there are some limitations and challenges to implementing these techniques in clinical practice. One significant limitation is the need for large and diverse datasets to train the models accurately [19]. Additionally, there is a potential risk of bias in data selection, which can affect the performance of the models. There is also a need to address the interpretability of the models to ensure that clinicians can understand the reasoning behind the model's predictions.

Despite these difficulties, there are substantial potential advantages to employing machine learning for cancer categorization. The systematic review's conclusions emphasize the need for more study and advancement in this field in order to fully realize machine learning's promise for cancer diagnosis, prognosis, and therapy.

5.2 Comparison with Previous Studies

Previous studies have also investigated the use of machine learning techniques in cancer classification, and the results of the present investigation agree with those from these studies. The excellent accuracy of machine learning approaches in the categorization of cancer has been documented in a number of research. For instance, According to a study by Esteva et al. a deep-learning system can accurately identify skin cancer in photographs with performance that is on par with or superior to that of board-certified dermatologists [129]. Similarly, another study by S Cui et al. reported that a deep learning algorithms could effectively classify lung nodules on CT images [130]. The results of the current study, which also show the great accuracy of machine learning techniques in cancer classification, are consistent with these findings. However, there are also some discrepancies between the current findings and previous studies. For example, some studies have reported lower accuracy rates for machine learning algorithms in cancer classification. These discrepancies could be due to differences in the datasets used, the machine learning techniques employed, or the specific cancer types being classified.

Overall, the current study's findings are consistent with previous studies, which have shown the use of machine learning methods in the categorization of cancer. To address some of the restrictions and difficulties related to the application of these approaches in clinical practise, more study is necessary.

Some of the limitations which we addressed from the literature review are:

1. Availability and quality of labeled data: The limited availability of large-scale, well-annotated datasets for training machine learning models in cancer classification poses a challenge. We will highlight the importance of access to improve the models' precision and generalizability, use a range of different and representative datasets.
2. Class imbalance: Imbalanced class distributions in cancer datasets can affect the performance of machine learning algorithms since they frequently favour the dominant class. To solve this issue and enhance the categorization of minority classes, we will talk about strategies like oversampling, under sampling, and cost-sensitive learning.
3. Feature selection and dimensionality: Cancer datasets often contain many features, some of which can be unnecessary or superfluous. To increase efficiency, we will investigate feature selection strategies and dimensionality reduction techniques and effectiveness of cancer classification models.

4. Interpretability and transparency: Machine learning models used for cancer classification often exhibit a black-box nature, making it challenging to interpret the underlying reasoning behind their predictions. We will discuss the importance of model interpretability and potential methods, such as feature importance analysis and model-agnostic interpretability techniques, to improve transparency and trust in the decision-making process.

In addition to discussing these issues, some of the objectives for improvising cancer classification, which may include:

- Developing novel machine learning algorithms specifically tailored for cancer classification to improve accuracy and interpretability.
- Integrating multimodal data sources, such as genomics, imaging, and clinical data, to enhance the predictive power of machine learning models.
- Exploring ensemble learning techniques and model combination approaches to leverage Synthesize the advantages of multiple approaches and augment overall effectiveness.
- Investigating the potential of Utilize deep learning models like convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to effectively capture intricate patterns and temporal relationships, resulting in improved precision when classifying cancer.
- Conducting further research on transfer learning and domain adaptation to address the challenges of limited labeled data in specific cancer types or subtypes.

By addressing these issues and stating the objectives for improvement, the discussion section will provide insights into the potential future directions and advancements in Utilize machine learning techniques to classify cancer accurately.

5.2.1 Contribution of the paper

1. Comprehensive review of the applications and techniques of machine learning in cancer classification.
2. Presentation of actual use cases of machine learning in cancer classification, demonstrating their implementation on medical data.
3. Discussion on supervised, unsupervised, and reinforcement learning algorithms, highlighting their advantages and disadvantages in the context of cancer classification.
4. Exploration of the implications and future potential of machine learning in improving cancer diagnosis, patient outcome prediction, and identification of therapeutic targets.

By organizing the contributions in bullet points, we aim to provide a clear and concise overview of the key contributions of the paper.

6 Conclusion

In conclusion, the systematic review has highlighted the promise of machine learning methods for identifying cancer. The findings suggest that these techniques have high accuracy rates and can be used to classify various cancer types, potentially aiding in diagnosis, prognosis, and treatment planning. The significance of the findings lies in the potential impact on improving cancer care, with the ability to provide faster and more accurate diagnoses, improved treatment planning, and personalized treatment strategies. The review has also identified limitations and challenges associated with the implementation of these techniques in clinical practice, including data standardization, interpretability, and ethical considerations. The study's contribution to the field of machine learning in cancer classification is in identifying key areas of focus for future research, examining, for instance, how well machine learning models function in actual healthcare contexts, improving interpretability, and evaluating the potential of machine learning in combination with other diagnostic and treatment modalities.

Overall, the systematic review demonstrates the tremendous promise of machine learning methods for cancer classification, and more study in this field may lead to better cancer treatment and patient outcomes.

Implications for Future Research: The findings of the systematic review have several implications for future research in the field of machine learning and cancer classification.

Firstly, there is a need for more standardized approaches to data collection and analysis. This can help in ensuring that the datasets used in machine learning studies are diverse, representative, and unbiased. Additionally, there is a need to address the issue of data privacy and security to ensure that patient data is protected while still being accessible for research purposes.

Secondly, there is a need to investigate the interpretability of machine learning models in cancer classification. This can help in improving the transparency of the models and ensuring that clinicians can understand the reasoning behind the model's predictions.

Thirdly, the effectiveness of machine learning models in actual healthcare situations has to be examined. This can help in evaluating the feasibility and clinical applicability of these techniques and identifying any potential barriers to their implementation in clinical practice.

Fourthly, there is a need to investigate the potential of machine learning techniques in combination with other diagnostic and treatment modalities. For example, combining

machine learning with imaging or genomic data can provide more accurate and comprehensive cancer diagnosis and treatment planning.

Finally, there is a need to investigate the ethical implications of using machine learning in cancer classification. This can help in ensuring that the use of these techniques is consistent with ethical principles and patient rights.

In conclusion, the findings of the systematic review highlight the potential of machine learning methods in cancer classification and the need for further research to address the limitations and challenges associated with their implementation in clinical practice.

Author Contributions AY and RM were responsible for collecting all the data for the project, including figures and tables. NKV was responsible for correcting the English language and formatting references.

Funding No funding available.

Data Availability NA.

Declarations

Conflict of interest The authors declare that they have no conflicts of interest to disclose.

References

1. Samuel AL, Gabel F. Artificial Intelligence for games: seminar some studies in machine learning using the game of checkers. IBM J Res Dev. 1959;3(1959):210–29.
2. Batta M. Machine learning algorithms—a review. Int J Sci Res. 2018;18(8):381–6. https://doi.org/10.21275/ART20203995.
3. Aziz RM, Sharma P, Hussain A. Machine learning algorithms for crime prediction under Indian penal code. Berlin: Springer; 2022. https://doi.org/10.1007/s40745-022-00424-6.
4. Nilashi M, Minaei-Bidgoli B, Alghamdi A, Alrizq M, Alghamdi O, Nayer FK, Aljehane NO, Khosravi A, Mohd S. Knowledge discovery for course choice decision in Massive Open Online Courses using machine learning approaches. Exp Syst Appl. 2022;199:117092. https://doi.org/10.1016/j.eswa.2022.117092.
5. Mahesh B. Machine learning algorithms-A review self flowing generator view project machine learning algorithms-A review view project Batta Mahesh independent researcher machine learning algorithms-A review. Int J Sci Res. 2018. https://doi.org/10.21275/ART20203995.
6. Géron A (Ed.) Book review: hands-on machine learning with Scikit-Learn, Keras, and Tensorflow, 2nd Edn. https://doi.org/10.1007/s13246-020-00913-z.
7. Ray S. A quick review of machine learning algorithms. In: 2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon). IEEE; 2019. p. 35–9. https://doi.org/10.1109/COMITCon.2019.8862451.
8. Kelleher JD, Mac Namee B, D'Arcy A. Fundamentals of Machine Learning for Predictive Data Analytics. Igarss 2014. 2015;(1):1–691
9. Alharbi F, Vakanski A. Machine learning methods for cancer classification using gene expression data: a review. Bioengineering. 2023. https://doi.org/10.3390/bioengineering10020173.
10. Aziz RM, Joshi AA, Kumar K, Gaani AH. Hybrid feature selection techniques utilizing soft computing methods for cancer data. In: Computational and analytic methods in biological sciences. River Publishers; 2023. p. 23–39.
11. Aziz RM. Application of nature inspired soft computing techniques for gene selection: a novel frame work for classification of cancer. Soft Comput. 2022;26(22):12179–96. https://doi.org/10.1007/s00500-022-07032-9.
12. Yaqoob A, Aziz RM, Verma NK, Lalwani P, Makrariya A, Kumar P. A review on nature-inspired algorithms for cancer disease prediction and classification. Mathematics. 2023;11(5):1081. https://doi.org/10.3390/math11051081.
13. Aziz RM. Nature-inspired metaheuristics model for gene selection and classification of biomedical microarray data. Med Biol Eng Comput. 2022;60(6):1627–46. https://doi.org/10.1007/s11517-022-02555-7.
14. Qasim Gilani S, Syed T, Umair M, Marques O. Skin cancer classification using deep spiking neural network. J Digit Imaging. 2023. https://doi.org/10.1007/s10278-023-00776-2.
15. Balaha HM, Hassan AES. Skin cancer diagnosis based on deep transfer learning and sparrow search algorithm, vol. 35. London: Springer; 2023. https://doi.org/10.1007/s00521-022-07762-9.
16. Ke J, Shen Y, Lu Y, Guo Y, Shen D. Mine local homogeneous representation by interaction information clustering with unsupervised learning in histopathology images. Comput Methods Programs Biomed. 2023;235:107520. https://doi.org/10.1016/j.cmpb.2023.107520.
17. Li T, et al. Ensemble learning-based gene signature and risk model for predicting prognosis of triple-negative breast cancer. Funct Integr Genom. 2023;23(2):1–16. https://doi.org/10.1007/s10142-023-01009-z.
18. Wang Z, Zhou Y, Takagi T, Song J, Tian YS, Shibuya T. Genetic algorithm-based feature selection with manifold learning for cancer classification using microarray data. BMC Bioinf. 2023;24(1):139. https://doi.org/10.1186/s12859-023-05267-3.
19. Dhillon A, Singh A, Kumar V. A systematic review on biomarker identification for cancer diagnosis and prognosis in multi-omics : from computational needs to machine learning and deep learning, vol. 30. Netherlands: Springer; 2023. https://doi.org/10.1007/s11831-022-09821-9.
20. Massafra R, et al. Analyzing breast cancer invasive disease event classification through explainable artificial intelligence. Front Med. 2023. https://doi.org/10.3389/fmed.2023.1116354.
21. Suthahar N, et al. Association of initial and longitudinal changes in C-reactive protein with the risk of cardiovascular disease, cancer, and mortality. Mayo Clin Proc. 2023;98(4):549–58. https://doi.org/10.1016/j.mayocp.2022.10.013.
22. Botlagunta M, et al. Classification and diagnostic prediction of breast cancer metastasis on clinical data using machine learning algorithms. Sci Rep. 2023;13(1):1–17. https://doi.org/10.1038/s41598-023-27548-w.

23. Zhao M, Lau MC, Haruki K, Väyrynen JP, Gurjao C, Väyrynen SA, Dias Costa A, Borowsky J, Fujiyoshi K, Arima K, Hamada T. Bayesian risk prediction model for colorectal cancer mortality through integration of clinicopathologic and genomic data. NPJ Prec Oncol. 2023;7(1):57. https://doi.org/10.1038/s41698-023-00406-8.
24. Srikantamurthy MM, Rallabandi VPS, Dudekula DB, Natarajan S, Park J. Classification of benign and malignant subtypes of breast cancer histopathology imaging using hybrid CNN-LSTM based transfer learning. BMC Med Imaging. 2023;23(1):1–15. https://doi.org/10.1186/s12880-023-00964-0.
25. Mohammed MA, Lakhan A, Abdulkareem KH, Garcia-Zapirain B. A hybrid cancer prediction based on multi-omics data and reinforcement learning state action reward state action (SARSA). Comput Biol Med. 2023;154:106617. https://doi.org/10.1016/j.compbiomed.2023.106617.
26. Kotevski DP, Smee RI, Vajdic CM, Field M. Empirical comparison of routinely collected electronic health record data for head and neck cancer-specific survival in machine-learnt prognostic models. Head Neck. 2023;45(2):365–79. https://doi.org/10.1002/hed.27241.
27. Zhang S, Xie W, Li W, Wang L, Feng C. GAMB-GNN: graph neural Networks learning from gene structure relations and Markov Blanket ranking for cancer classification in microarray data. Chemom Intell Lab Syst. 2023;232:104713. https://doi.org/10.1016/j.chemolab.2022.104713.
28. Otchy D, et al. Practice parameters for colon cancer. Dis Colon Rectum. 2004;47(8):1269–84. https://doi.org/10.1007/s10350-004-0598-8.
29. Mrózek K, Heerema NA, Bloomfield CD. Cytogenetics in acute leukemia. Blood Rev. 2004;18(2):115–36. https://doi.org/10.1016/S0268-960X(03)00040-7.
30. Bastian PJ, Mangold LA, Epstein JI, Partin AW. Characteristics of insignificant clinical T1c prostate tumors: a contemporary analysis. Cancer. 2004;101(9):2001–5. https://doi.org/10.1002/cncr.20586.
31. Wang Y, Jiang T. Understanding high grade glioma: Molecular mechanism, therapy and comprehensive management. Cancer Lett. 2013;331(2):139–46. https://doi.org/10.1016/j.canlet.2012.12.024.
32. Rowell NP, Williams C. Radical radiotherapy for stage I/II non-small cell lung cancer in patients not sufficiently fit for or declining surgery (medically inoperable). Cochrane Database Syst Rev. 2001. https://doi.org/10.1002/14651858.cd002935.
33. Ready N, et al. Chemoradiotherapy and gefitinib in stage III non-small cell lung cancer with epidermal growth factor receptor and KRAS mutation analysis: cancer and leukemia group B (CALEB) 30106, a CALGB-stratified phase II trial. J Thorac Oncol. 2010;5(9):1382–90. https://doi.org/10.1097/JTO.0b013e3181eba657.
34. Kamat AM, Hahn NM, Efstathiou JA. 5 Yalcin AD, Celik B, Yalcin AN. Omalizumab (anti-IgE) therapy in the asthma-COPD overlap syndrome (ACOS) and its effects on circulating cytokine levels. Immunopharmacol Immunotoxicol. 2016;38:253–56. https://doi.org/10.1016/S0140-6736(16)32112-2.
35. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. Comput Struct Biotechnol J. 2015;13:8–17. https://doi.org/10.1016/j.csbj.2014.11.005.
36. Basavaraju A, Du J, Zhou F, Ji J. A machine learning approach to road surface anomaly assessment using smartphone sensors. IEEE Sens J. 2020;20(5):2635–47. https://doi.org/10.1109/JSEN.2019.2952857.
37. Bonnot K, Benoit P, Mamy L, Patureau D. Transformation of PPCPs in the environment: review of knowledge and classification of pathways according to parent molecule structures. Crit Rev Environ Sci Technol. 2023;53(1):47–69. https://doi.org/10.1080/10643389.2022.2045159.
38. Aziz RM, Desai NP, Baluch MF. Computer vision model with novel cuckoo search based deep learning approach for classification of fish image. 2023; 3677–3696
39. Mandair D, Reis-Filho JS, Ashworth A. Biological insights and novel biomarker discovery through deep learning approaches in breast cancer histopathology. npj Breast Cancer. 2023;9(1):1–11. https://doi.org/10.1038/s41523-023-00518-1.
40. Vermij L, et al. Prognostic refinement of NSMP high-risk endometrial cancers using oestrogen receptor immunohistochemistry. Br J Cancer. 2023. https://doi.org/10.1038/s41416-023-02141-0.
41. Chitalia R, Miliotis M, Jahani N, Tastsoglou S, McDonald ES, Belenky V, Cohen EA, Newitt D, van't Veer LJ, Esserman L, Hylton N. Radiomic tumor phenotypes augment molecular profiling in predicting recurrence free survival after breast neoadjuvant chemotherapy. Commun Med. 2023;3(1):46. https://doi.org/10.1038/s43856-023-00273-1.
42. Pandit BR, et al. Deep learning neural network for lung cancer classification: enhanced optimization function. Multimed Tools Appl. 2023;82(5):6605–24. https://doi.org/10.1007/s11042-022-13566-9.
43. Jakhar AK, Gupta A, Singh M. SELF: a stacked-based ensemble learning framework for breast cancer classification. Evol Intell. 2023. https://doi.org/10.1007/s12065-023-00824-4.
44. Dimitrovski I, Kitanovski I, Kocev D, Simidjievski N. Current trends in deep learning for earth observation: an open-source benchmark arena for image classification. ISPRS J Photogramm Remote Sens. 2022;197(February):18–35. https://doi.org/10.1016/j.isprsjprs.2023.01.014.
45. Le Chan JY, Bea KT, Leow SMH, Phoong SW, Cheng WK. State of the art: a review of sentiment analysis based on sequential transfer learning, vol. 56. Netherlands: Springer; 2023. https://doi.org/10.1007/s10462-022-10183-8.
46. Abdallah A, Maarof MA, Zainal A. Fraud detection system: a survey. J Netw Comput Appl. 2016;68:90–113. https://doi.org/10.1016/j.jnca.2016.04.007.
47. Liang KY, Zeger SL. Regression analysis for correlated data. Annu Rev Public Health. 1993;14:43–68. https://doi.org/10.1146/annurev.pu.14.050193.000355.
48. Meer P, Mintz D, Rosenfeld A, Kim DY. Robust regression methods for computer vision: a review. Int J Comput Vis. 1991;6(1):59–70. https://doi.org/10.1007/BF00127126.
49. Nasteski V. An overview of the supervised machine learning methods. Horizons B. 2017;4(December 2017):51–62. https://doi.org/10.20544/horizons.b.04.1.17.p05.
50. Dhanabal S, Chandramathi S. A review of various k-nearest neighbor query processing techniques. Int J Comput Appl. 2011;31(7):14–22, [Online]. Available: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:A+Review+of+various+k-Nearest+Neighbor+Query+Processing+Techniques#0
51. Berg-Kirkpatrick T, Bouchard-Côté A, DeNero J, Klein D. Painless unsupervised learning with features. In: Human language technologies: the 2010 annual conference of the North American chapter of the association for computational linguistics; 2010. p. 582–90.
52. Cunningham JP, Ghahramani Z. Linear dimensionality reduction: survey, insights, and generalizations. J Mach Learn Res. 2015;16:2859–900.
53. Rustam F, et al. COVID-19 future forecasting using supervised machine learning models. IEEE Access. 2020;8:101489–99. https://doi.org/10.1109/ACCESS.2020.2997311.
54. Noble WS. What is a support vector machine? Nat Biotechnol. 2006;24(12):1565–7. https://doi.org/10.1038/nbt1206-1565.

55. Rana A, Pandey R. A review of popular decision tree algorithms in data mining. Asian J Multidimens Res. 2021;10(10):230–7. https://doi.org/10.5958/2278-4853.2021.00837.5.
56. Fletcher S, Islam MZ. Decision tree classification with differential privacy: a survey. ACM Comput Surv. 2019. https://doi.org/10.1145/3337064.
57. Biau G, Scornet E. A random forest guided tour. Test. 2016;25(2):197–227. https://doi.org/10.1007/s11749-016-0481-7.
58. Bishop CM. Neural networks and their applications. Rev Sci Instrum. 1994;65(6):1803.
59. Yuwono M, Moulton BD, Su SW, Celler BG, Nguyen HT. Unsupervised machine-learning method for improving the performance of ambulatory fall-detection systems. Biomed Eng Online. 2012. https://doi.org/10.1186/1475-925X-11-9.
60. García-Díaz P, Sánchez-Berriel I, Martínez-Rojas JA, Diez-Pascual AM. Unsupervised feature selection algorithm for multiclass cancer classification of gene expression RNA-Seq data. Genomics. 2020;112(2):1916–25. https://doi.org/10.1016/j.ygeno.2019.11.004.
61. Sun W. Sports performance prediction based on chaos theory and machine learning. Wirel Commun Mob Comput. 2022. https://doi.org/10.1155/2022/3916383.
62. Khurma RA, Aljarah I, Sharieh A, Elaziz MA, Damaševičius R, Krilavičius T. A review of the modification strategies of the nature inspired algorithms for feature selection problem. Mathematics. 2022;10(3):1–45. https://doi.org/10.3390/math10030464.
63. Jawad K, Mahto R, Das A, Ahmed SU, Aziz RM, Kumar P. Applied sciences novel cuckoo search-based metaheuristic approach for deep learning prediction of depression. Appl Sci. 2023. https://doi.org/10.3390/app13095322.
64. Ahmed M, Seraj R, Mohammed S, Islam S. The k-means algorithm: a comprehensive survey and performance evaluation. Electronics. 2020. https://doi.org/10.3390/electronics9081295.
65. Jain AK. Data clustering: 50 years beyond K-means. Pattern Recogn Lett. 2010;31(8):651–66. https://doi.org/10.1016/j.patrec.2009.09.011.
66. Murtagh F, Contreras P. Algorithms for hierarchical clustering: an overview. Wiley Interdiscip Rev Data Min Knowl Discov. 2012;2(1):86–97. https://doi.org/10.1002/widm.53.
67. Dabhi DP, Patel MR, Dipak MRP, Dabhi P. Extensive survey on hierarchical clustering methods in data mining. Int Res J Eng Technol, 2016; 03(11):659–665, [Online]. Available: https://www.irjet.net/archives/V3/i11/IRJET-V3I11115.pdf
68. Kriegel HP, Kröger P, Sander J, Zimek A. Density-based clustering. Wiley Interdiscip Rev Data Min Knowl Discov. 2011;1(3):231–40. https://doi.org/10.1002/widm.30.
69. Moulavi D, Jaskowiak PA, Campello RJGB, Zimek A, Sander J. Density-based clustering validation. SIAM Int Conf Data Min 2014, SDM 2014. 2014; 2(i):839–847. https://doi.org/10.1137/1.9781611973440.96.
70. Aziz R, Verma CK, Srivastava N. A novel approach for dimension reduction of microarray. Comput Biol Chem. 2017;71:161–9. https://doi.org/10.1016/j.compbiolchem.2017.10.009.
71. Aziz R, Verma CK, Srivastava N. Dimension reduction methods for microarray data: a review. AIMS Bioeng. 2017;4(1):179–97. https://doi.org/10.3934/bioeng.2017.1.179.
72. Van Der Maaten LJP, Postma EO, Van Den Herik HJ. Dimensionality reduction: a comparative review. J Mach Learn Res. 2009;10:1–41. https://doi.org/10.1080/13506280444000102.
73. Musheer RA, Verma CK, Srivastava N. Novel machine learning approach for classification of high-dimensional microarray data. Soft Comput. 2019;23(24):13409–21. https://doi.org/10.1007/s00500-019-03879-7.
74. Aziz R, Verma CK, Jha M, Srivastava N. Artificial neural network classification of microarray data using new hybrid gene selection method. Int J Data Min Bioinf. 2017;17(1):42–65. https://doi.org/10.1504/IJDMB.2017.084026.
75. Box PO, Van Der Maaten L, Postma E, Van Den Herik J. Tilburg centre for creative computing dimensionality reduction: a comparative review dimensionality reduction: a comparative review 2009. [Online]. Available: http://www.uvt.nl/ticc
76. Washington P, Paskov KM, Kalantarian H, Stockham N, Voss C, Kline A, Patnaik R, Chrisman B, Varma M, Tariq Q, Dunlap K. Feature selection and dimension reduction of social autism data. In: Pacific symposium on biocomputing 2020. 2019. p. 707–18. https://doi.org/10.1142/9789811215636_0062.
77. Khalid S, Khalil T, Nasreen S. A survey of feature selection and feature extraction techniques in machine learning. Proc 2014 Sci Inf Conf SAI 2014, 2014; (October 2016):372–378. https://doi.org/10.1109/SAI.2014.6918213.
78. Solorio-Fernández S, Carrasco-Ochoa JA, Martínez-Trinidad JF. A review of unsupervised feature selection methods. Artif Intell Rev. 2020;53(2):907–48. https://doi.org/10.1007/s10462-019-09682-y.
79. Cai J, Luo J, Wang S, Yang S. Feature selection in machine learning: a new perspective. Neurocomputing. 2018;300:70–9. https://doi.org/10.1016/j.neucom.2017.11.077.
80. Bommert A, Sun X, Bischl B, Rahnenführer J, Lang M. Benchmark for filter methods for feature selection in high-dimensional classification data. Comput Stat Data Anal. 2020;143:106839. https://doi.org/10.1016/j.csda.2019.106839.
81. Shukla AK, Singh P, Vardhan M. A two-stage gene selection method for biomarker discovery from microarray data for cancer classification. Chemom Intell Lab Syst. 2018;183:47–58. https://doi.org/10.1016/j.chemolab.2018.10.009.
82. Wang X, Bo D, Shi C, Fan S, Ye Y, Yu PS. A survey on heterogeneous graph embedding: methods, techniques, applications and sources. IEEE Trans Big Data. 2022. https://doi.org/10.1109/TBDATA.2022.3177455.
83. Lamba P, Rawal K. A Survey of Algorithms for Feature Extraction and Feature Classification Methods
84. Santoni MM, Sensuse DI, Arymurthy AM, Fanany MI. Cattle race classification using gray level co-occurrence matrix convolutional neural networks. Procedia Comput Sci. 2015;59(Iccsci):493–502. https://doi.org/10.1016/j.procs.2015.07.525.
85. Huang D, Shan C, Ardabilian M, Wang Y, Chen L. Local binary patterns and its application to facial image analysis: a survey. IEEE Trans Syst Man Cybern Part C Appl Rev. 2011;41(6):765–81. https://doi.org/10.1109/TSMCC.2011.2118750.
86. Yudistiro K, Suharto G, Fatah A, Ari L, Wibawa N. Detection of aflatoxin contamination in corn using the simplified Gabor Wavelet algorithm. Internet Things Artif Intell J. 2023. https://doi.org/10.31763/iota.v3i1.576.
87. Yu R, Huang Y, Peng Y, Wang K. Monitoring of butt weld penetration based on infrared sensing and improved histograms of oriented gradients. J Mater Res Technol. 2023;22:3280–93. https://doi.org/10.1016/j.jmrt.2022.12.139.
88. Li Z, Liu F, Yang W, Peng S, Zhou J. A survey of convolutional neural networks: analysis, applications, and prospects. IEEE Trans Neural Netw Learn Syst. 2021. https://doi.org/10.1109/TNNLS.2021.3084827.
89. Ali S, Miah S, Haque J, Rahman M. Machine Learning with Applications An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models. Mach Learn Appl. 2021;5(February):100036. https://doi.org/10.1016/j.mlwa.2021.100036.

90. Saqib M, Şentürk E, Sahu SA, Adil MA. Comparisons of autoregressive integrated moving average (ARIMA) and long short term memory (LSTM) network models for ionospheric anomalies detection: a study on Haiti (Mw = 7.0) earthquake. Acta Geod Geophys. 2022;57(1):195–213. https://doi.org/10.1007/s40328-021-00371-3.
91. Khodr J, Younes R. Dimensionality reduction on hyperspectral images: A comparative review based on artificial datas. Proc - 4th Int Congr Image Signal Process. CISP 2011. 2011; 4(October):1875–1883. https://doi.org/10.1109/CISP.2011.6100531.
92. Pirra M, Diana M. A study of tour-based mode choice based on a support vector machine classifier. Transp Plan Technol. 2019;42(1):23–36. https://doi.org/10.1080/03081060.2018.1541280.
93. Zhou S, Yang C, Su Z, Yu P, Jiao J. An aeromagnetic compensation algorithm based on radial basis function artificial neural network. Appl Sci. 2023. https://doi.org/10.3390/app13010136.
94. Nethra Betgeri S, Reddy Vadyala S, Matthews JC, Madadi M, Vladeanu G. Wastewater pipe condition rating model using K-nearest neighbors. Tunn Undergr Sp Technol. 2023;132(December 2022):104921. https://doi.org/10.1016/j.tust.2022.104921.
95. Saha KK, et al. Classification of starfruit maturity using smartphone-image and multivariate analysis. J Agric Food Res. 2023;11(June 2022):100473. https://doi.org/10.1016/j.jafr.2022.100473.
96. Plaat A, Kosters W, Preuss M. High-accuracy model-based reinforcement learning, a survey. Netherlands: Springer; 2023. https://doi.org/10.1007/s10462-022-10335-w.
97. Mousavi SS, Schukat M, Howley E. Deep reinforcement learning: an overview. Lect Notes Networks Syst. 2018;16:426–40. https://doi.org/10.1007/978-3-319-56991-8_32.
98. Spanò S, et al. An efficient hardware implementation of reinforcement learning: the q-learning algorithm. IEEE Access. 2019;7:186340–51. https://doi.org/10.1109/ACCESS.2019.2961174.
99. Leong CP, Liew CS, Chan CS, Rehman MHU. Optimizing workflow task clustering using reinforcement learning. IEEE Access. 2021;9(July):110614–26. https://doi.org/10.1109/ACCESS.2021.3101454.
100. Arulkumaran K, Deisenroth MP, Brundage M, Bharath AA. Deep reinforcement learning: a brief survey. IEEE Signal Process Mag. 2017;34(6):26–38. https://doi.org/10.1109/MSP.2017.2743240.
101. Qiu D, Wang Y, Hua W, Strbac G. Reinforcement learning for electric vehicle applications in power systems: a critical review. Renew Sustain Energy Rev. 2023;173(June 2023):113052. https://doi.org/10.1016/j.rser.2022.113052.
102. Hamad Q, Samma H, Suandi SA. Q-Learning based metaheuristic optimization algorithms: a short review and perspectives. https://doi.org/10.21203/rs.3.rs-1950095/v1.
103. Chen X, Yao L, McAuley J, Zhou G, Wang X. Deep reinforcement learning in recommender systems: a survey and new perspectives. Knowledge Based Syst. 2023;264:110335. https://doi.org/10.1016/j.knosys.2023.110335.
104. Cho S, Won H. Machine learning in DNA microarray analysis for cancer classification. 2018; (May 2014)
105. Priyanka KS. A review paper on breast cancer detection using deep learning. In: IOP conference series: materials science and engineering, Vol. 1022, No. 1. IOP Publishing; 2021. p. 012071. https://doi.org/10.1088/1757-899X/1022/1/012071.
106. Javaid A, Sadiq M, Akram F. Skin cancer classification using image processing and machine learning. In: 2021 international Bhurban conference on applied sciences and technologies (IBCAST). IEEE; 2021. p. 439–44. https://doi.org/10.1109/IBCAST51254.2021.9393198.
107. Omondiagbe DA, Veeramani S, Sidhu AS. Machine learning classification techniques for breast cancer diagnosis. In: IOP conference series: materials science and engineering, Vol. 495. IOP Publishing; 2019. p. 012033. https://doi.org/10.1088/1757-899X/495/1/012033.
108. Mashudi NA, Rossli SA, Ahmad N, Noor NM. Comparison on some machine learning techniques in breast cancer classification. In: 2020 IEEE-EMBS conference on biomedical engineering and sciences (IECBES). IEEE; 2021. p. 499–504. https://doi.org/10.1109/IECBES48179.2021.9398837.
109. Ed A. Breast cancer classification with reduced feature set using association rules and support vector machine. Netw Model Anal Heal Informatics Bioinf. 2020;9(1):1–10. https://doi.org/10.1007/s13721-020-00237-8.
110. Javed Mehedi Shamrat FM, Raihan MA, Rahman AKMS, Mahmud I, Akter R. An analysis on breast disease prediction using machine learning approaches. Int J Sci Technol Res. 2020;9(2):2450–5.
111. Lomboy KEMR, Hernandez RM. A comparative performance of breast cancer classification using hyper-parameterized machine learning models. Int J Adv Technol Eng Explor. 2021;8(82):1080–101. https://doi.org/10.19101/IJATEE.2021.874380.
112. Keerthana D, Venugopal V, Nath MK, Mishra M. Hybrid convolutional neural networks with SVM classifier for classification of skin cancer. Biomed Eng Adv. 2023;5(December 2022):100069. https://doi.org/10.1016/j.bea.2022.100069.
113. Loizidou K, Elia R, Pitris C. Computer-aided breast cancer detection and classification in mammography: a comprehensive review. Comput Biol Med. 2023;153(December 2022):106554. https://doi.org/10.1016/j.compbiomed.2023.106554.
114. Siddiqui EA, Chaurasia V, Shandilya M. Detection and classification of lung cancer computed tomography images using a novel improved deep belief network with Gabor filters. Chemom Intell Lab Syst. 2023;235(January):104763. https://doi.org/10.1016/j.chemolab.2023.104763.
115. Hassan MM, et al. A comparative assessment of machine learning algorithms with the least absolute shrinkage and selection operator for breast cancer detection and prediction. Decis Anal J. 2023;7(April):100245. https://doi.org/10.1016/j.dajour.2023.100245.
116. Al Nahid A, Mehrabi MA, Kong Y. Histopathological breast cancer image classification by deep neural network techniques guided by local clustering. Biomed Res Int. 2018. https://doi.org/10.1155/2018/2362108.
117. Masud M, Sikder N, Al Nahid A, Bairagi AK, Alzain MA. A machine learning approach to diagnosing lung and colon cancer using a deep learning-based classification framework. Sensors (Switzerland). 2021;21(3):1–21. https://doi.org/10.3390/s21030748.
118. Asri H, Mousannif H, Al Moatassime H, Noel T. Using machine learning algorithms for breast cancer risk prediction and diagnosis. Procedia Comput Sci. 2016;83(Fams):1064–9. https://doi.org/10.1016/j.procs.2016.04.224.
119. Nageswaran S, et al. Lung cancer classification and prediction using machine learning and image processing. Biomed Res Int. 2022. https://doi.org/10.1155/2022/1755460.
120. Zhou X, Liu KY, Wong STC. Cancer classification and prediction using logistic regression with Bayesian gene selection. J Biomed Inform. 2004;37(4):249–59. https://doi.org/10.1016/j.jbi.2004.07.009.
121. Alabi RO, et al. Comparison of supervised machine learning classification techniques in prediction of locoregional recurrences in early oral tongue cancer. Int J Med Inform. 2020;136(December 2019):104068. https://doi.org/10.1016/j.ijmedinf.2019.104068.

122. Erdem E, Bozkurt F. Prostat kanseri tahmini için çeşitli denetimli makine öğrenimi tekniklerinin karşılaştırılması. Eur J Sci Technol. 2021;21:610–20. https://doi.org/10.31590/ejosat.802810.
123. Abunasser BS, Al-hiealy MRJ, Zaqout IS. Convolution neural network for breast cancer detection and classification using deep learning. Asian Pac J Cancer Prev. 2023;24:531–44. https://doi.org/10.31557/APJCP.2023.24.2.531.
124. Minnoor M, Baths V. Science direct sciencedirect diagnosis of breast cancer using random forests diagnosis of breast cancer using random forests. Procedia Comput Sci. 2023;218(2022):429–37. https://doi.org/10.1016/j.procs.2023.01.025.
125. Gupta V, Gaur H, Vashishtha S, Das U, Singh VK, Hemanth DJ. A fuzzy rule-based system with decision tree for breast cancer detection. IET Image Process. 2023;17(7):2083–96. https://doi.org/10.1049/ipr2.12774.
126. Asif S, Zhao M, Tang F, Zhu Y. An enhanced deep learning method for multi-class brain tumor classification using deep transfer learning. Multimed Tools Appl. 2023;82:31709.
127. Kavitha R, Jothi DK, Saravanan K, Swain MP, Gonzáles JL, Bhardwaj RJ, Adomako E. Ant colony optimization-enabled CNN deep learning technique for accurate detection of cervical cancer. BioMed Res Int. 2023. https://doi.org/10.1155/2023/1742891.
128. Wang Y, et al. Bioinformatics analysis of ferroptosis-related gene AKR1C3 as a potential biomarker of asthma and its identification in BEAS-2B cells. Comput Biol Med. 2023;158(March):106740. https://doi.org/10.1016/j.compbiomed.2023.106740.
129. Esteva A, et al. Corrigendum: dermatologist-level deep neural networks. Nat Publ Gr. 2017;546(7660):686. https://doi.org/10.1038/nature22985.
130. Cui S, et al. Development and clinical application of deep learning model for lung nodules screening on CT images. Sci Rep. 2020. https://doi.org/10.1038/s41598-020-70629-3.