

A Survey of Machine Learning Techniques for Disease Detection, Prognosis, and Drug Response Analysis

Hossein Eyvazi

Department of Computer Science, Tarbiat Modares University, Tehran

Ali Rajaei

Department of Computer Science, Tarbiat Modares University, Tehran

Abstract

The integration of artificial intelligence (AI) and machine learning (ML) into healthcare has revolutionized disease detection, prediction, and drug response analysis. This survey provides a comprehensive overview of the transformative role of AI and ML in addressing critical challenges in modern medicine. By leveraging advanced techniques such as convolutional neural networks (CNNs), graph neural networks (GNNs), and explainable AI (XAI), these technologies enable early and accurate disease detection, precise prognosis, and personalized treatment strategies. This paper explores the application of AI and ML in various domains, including leukemia detection, cancer staging, chronic disease prediction, and drug response analysis. We highlight key methodologies, datasets, and performance metrics, while addressing challenges such as data dependency, computational costs, and model interpretability. Furthermore, we discuss future directions, including the integration of multi-omics data, transfer learning, and the development of robust, generalizable models. By bridging the gap between computational science and clinical practice, AI and ML hold immense potential to enhance healthcare delivery, improve patient outcomes, and pave the way for precision medicine.

Keywords

AI, Machine Learning, Helath, Disease Detection

1. Introduction

The synergy between artificial intelligence (AI) and machine learning (ML) with medical science is catalysing a paradigm shift in disease detection, prediction, and drug response analysis. Conventional diagnostic and prognostic approaches, which often rely on expensive and invasive techniques, are being increasingly augmented—and in some instances, superseded—by AI-driven solutions that promise enhanced accuracy, efficiency, and accessibility.

1.1 The Imperative for Innovation

Early and precise disease detection remains a pivotal challenge in healthcare, significantly influencing treatment efficacy and patient survival rates. Traditional diagnostic modalities can be arduous, time-intensive, and contingent on the proficiency of medical practitioners. The deployment of AI and ML seeks to mitigate these constraints by automating and refining the analysis of intricate datasets.

1.2 Scope of AI and ML Applications

This survey elucidates the multifaceted applications of AI and ML techniques across several critical domains:

- **Disease Detection:** Identifying pathologies such as cancer, including leukaemia, at nascent stages via advanced image analysis and data-driven methodologies. For example, one study used a combination of segmented and original microscopic blood images to train a model to detect B-ALL [1,2].
- **Disease Prediction:** Forecasting the propensity for disease manifestation or progression based on an array of patient-specific attributes, genetic predispositions, and administrative claims data. For instance, one study used age, gender and smoking history as node features in a GNN to predict chronic diseases [3].
- **Drug Response Prediction:** Ascertaining the potential effectiveness of specific pharmacological agents for individual patients or cancer subtypes through the interrogation of gene expression profiles and other omics data[4].

1.3 Technological Advancements

The exponential growth of deep learning (DL), particularly convolutional neural networks (CNNs) and graph neural networks (GNNs), has markedly amplified the capabilities of AI in medical contexts. These sophisticated models are adept at extracting intricate features from medical imagery, genomic sequences, and patient records, thereby enabling more accurate and personalised predictive outcomes [5].

This introduction lays the groundwork for a comprehensive exploration of how AI and ML are transforming healthcare, establishing a context for the detailed sections that will follow.

2. Background and Related Works

This section provides an overview of the foundational concepts and prior research that underpin the application of AI and ML in disease detection, prediction, and drug response. It will cover essential machine learning techniques, the importance of data quality and preprocessing, and a review of relevant literature in each application area.

2.1 Machine Learning Fundamentals

At its core, machine learning involves the development of algorithms that allow computers to learn from data without explicit programming. Several ML techniques are pivotal in biomedical applications:

- **Supervised Learning:** Algorithms trained on labelled data to predict outcomes. Common methods include:
 - **Convolutional Neural Networks (CNNs):** Widely used for image analysis, CNNs can automatically learn spatial hierarchies of features, making them ideal for detecting patterns in medical images. For instance, CNNs are used to classify leukocytes in blood samples [6].
 - **Graph Neural Networks (GNNs):** Designed to operate on graph-structured data, GNNs are employed to model relationships between genes, proteins, and other biological entities. GNNs learn node embeddings automatically from patient networks constructed from administrative healthcare claim data.
 - **Support Vector Machines (SVMs):** Effective for classification tasks, SVMs find an optimal hyperplane to separate data into different classes.
 - **Decision Trees and Ensemble Methods:** Decision trees partition data based on feature values, while ensemble methods like Random Forests and Gradient Boosting combine multiple trees to improve predictive accuracy[7].
- **Unsupervised Learning:** Algorithms that identify patterns in unlabelled data. Key techniques include:
 - **Clustering:** Methods like K-means and hierarchical clustering group similar data points together, helping to identify disease subtypes or patient subgroups[8].
 - **Dimensionality Reduction:** Techniques such as Principal Component Analysis (PCA) reduce the number of variables while retaining essential information.
- **Semi-Supervised Learning:** Approaches that use both labelled and unlabelled data to improve model performance, useful when labelled data is scarce[9].

2.2 Data Quality and Preprocessing

The performance of AI/ML models heavily relies on the quality of input data. Preprocessing steps are crucial for enhancing data quality and preparing it for analysis:

- **Image Preprocessing:** Normalization of bone marrow aspirates can be done to account for hemodilution.
- **Data Cleaning:** Handling missing values, removing outliers, and correcting inconsistencies[10].

- Feature Extraction: Identifying relevant features from raw data, which can be done automatically using deep learning or manually through domain expertise.
- Data Augmentation: Increasing the size of the training dataset by applying transformations such as rotations, mirroring, and noise addition to images.

2.3 Review of Relevant Literature

Prior research has demonstrated the potential of AI and ML in various biomedical applications:

- Leukaemia Detection and Classification: Studies have employed CNNs to classify blood microscopic images and distinguish between normal and malignant cells. The C-NMC 2019 dataset is often used.
- Cancer Gene Module Dissection: Explainable GNN frameworks have been developed to dissect cancer gene modules and identify high-order relationships between genes[11].
- Chronic Disease Prediction: GNNs have been applied to predict chronic diseases using administrative claims data, modelling patient relationships and aggregating information from patient networks.
- Drug Response Prediction: Machine learning models, including graph convolutional networks, have been used to predict drug response in cancer cells based on genomic signatures and chemical structures.

This background provides a solid foundation for understanding the more detailed discussions in the subsequent sections, which will delve into specific applications and methodologies.

3. AI and Machine Learning in Disease Detection

Early and accurate disease detection is crucial for improving treatment outcomes and patient survival. This section explores how AI and ML techniques are applied to enhance the detection of various diseases, with a focus on leukaemia and other cancers.

3.1 Leukaemia Detection

Leukaemia, a cancer of the blood-forming tissues, benefits significantly from early detection. Traditional methods, such as flow cytometry[12], can be laborious and require specialised equipment. AI and ML offer alternative, efficient approaches:

- Image Analysis of Blood Smears:
 - Optimised Convolutional Neural Networks (OCNN): An optimised CNN (OCNN) can classify blood microscopic images to distinguish between leukaemia-free and leukaemia-affected samples. Fuzzy logic optimises the CNN hyperparameters. The OCNN has three steps: (i) Image Preprocessing, (ii) Feature Extraction, and (iii) Classification.
 - Datasets: The C-NMC 2019 dataset, containing lymphocyte images, is commonly used for training and validation. This dataset includes malignant and healthy cells.
 - Performance Metrics: Accuracy, precision, sensitivity, and F1-score are used to evaluate the performance of these models. OCNN models have achieved high accuracy (99.99%) with the C-NMC_Leukemia dataset.
- Deep Learning-Based Approaches:
 - Deep learning models can automatically extract features from blood smear images, reducing the need for manual feature engineering.
 - CNNs can detect crucial characteristics without human interaction, improving performance by optimising hyperparameters.
- Data Augmentation:
 - Techniques such as mirroring, rotation, blurring, and noise addition are employed to balance training and validation datasets.
 - Data augmentation helps improve the generalisability and robustness of the models.

3.2 Cancer Detection Using Multi-Omics and Graph Neural Networks

Beyond leukaemia, AI and ML are also transforming the detection of other cancers:

- Multi-Omics Data Integration:
 - CGMega, a framework using explainable graph attention[13], dissects cancer gene modules by constructing a multi-omics representation graph.
 - Nodes represent genes, and edges represent protein-protein interactions (PPIs).
 - Node features combine Hi-C features, promoter densities, histone modifications, and frequencies of single nucleotide variants (SNVs) and copy number variants (CNVs).
- Graph Neural Networks (GNNs):
 - GNNs are used to predict cancer genes in a semi-supervised manner, achieving high performance in cancer gene prediction.
 - GNNExplainer is used to interpret the contribution factors to cancer genes, detecting cancer gene modules with representative features.
- Sample-Specific Networks:
 - Graph Convolutional Networks (GCNs) are employed to understand cancer development by classifying cancer development stages.

- These networks help identify significant subgraphs and genes associated with different cancer stages.
- Intratumour Heterogeneity:
 - Intratumour Graph Neural Networks (IGNN) can evaluate prognostic risk by minimising the negative log partial likelihood of Cox proportional hazards regression loss.
 - IGNNs use GRU-based attention blocks for node information propagation, extracting prognostic graph representations.

3.3 General Machine Learning Approaches

Various machine learning algorithms are applied for cancer detection:

- Neural Networks: Effective for both large and small datasets and capable of detecting complex non-linear relationships.
- Support Vector Machines (SVM): Offer high accuracy and work well with unstructured and semi-structured data.
- Deep Learning: Applied to various data types, automatically deducing features without requiring manual feature engineering.

3.4 Challenges and Future Directions

Despite the advancements, challenges remain:

- Data Dependency: The performance of AI/ML models heavily relies on the quality and size of datasets.
- Computational Costs: Training deep learning models can be computationally intensive and require expensive GPUs.
- Interpretability: Understanding the decision-making process of complex models remains a challenge.

Future research directions include:

- Improving model interpretability to gain insights into disease mechanisms.
- Developing more robust methods for handling noisy and heterogeneous data.
- Exploring transfer learning techniques to leverage knowledge from well-studied cancers to less common ones.
- Integrating multi-omics data to provide a more comprehensive view of cancer biology.

4. AI and Machine Learning in Disease Prediction

The ability to predict disease onset or progression is a key goal of modern medicine, enabling proactive interventions and personalised treatment strategies. This section examines how AI and ML techniques are being employed to forecast disease risk and outcomes, focusing on cancer prognosis and chronic disease prediction.

4.1 Cancer Prognosis

Predicting cancer prognosis involves assessing the likelihood of disease recurrence, progression, and patient survival. AI and ML models are increasingly used to integrate diverse data types and improve the accuracy of prognostic predictions:

- Intratumour Heterogeneity Analysis:
 - Intratumour Graph Neural Networks (IGNN) can evaluate prognostic risk in breast cancer by analysing spatial heterogeneity. These models minimise the negative log partial likelihood of Cox proportional hazards regression loss.
 - GRU-based attention blocks are used for node information propagation and aggregation along the graph structure.
- Gene Expression Analysis:
 - Machine learning models can be used to predict survival outcomes based on gene expression profiles.
 - Identifying gene expression signatures associated with prolonged overall survival in cancers such as lung adenocarcinoma (LUAD) can improve prognostic accuracy.
- Integration of Clinical and Pathological Data:
 - AI models can integrate clinical data (e.g., patient age, tumour size, stage) with pathological images to predict cancer outcomes.
 - Combining computational histopathology with genomic data enhances the accuracy of outcome predictions.

4.2 Chronic Disease Prediction

Chronic diseases, such as cardiovascular and pulmonary conditions, are major public health challenges. AI and ML offer opportunities to identify individuals at high risk and implement preventive measures:

- Graph Neural Networks (GNNs):
 - GNNs are used to model patient relationships from administrative healthcare claim data.
 - These models learn node embeddings automatically from patient networks, using this information to predict chronic diseases like cardiovascular disease (CVD) and chronic pulmonary disease (CPD)[15].
- Weighted Patient Networks (WPN):

- A bipartite graph is created and the bipartite projection technique to create a WPN.
- An edge list is created from this WPN, and patient features are used to train and test the GNN-based models to predict the chronic disease.
- Feature Selection and Engineering:
 - Identifying relevant features from administrative data, such as medical codes, demographics, and healthcare utilisation patterns, is crucial for accurate prediction.
 - Chi-square is a non-parametric statistical method used for feature selection, specifically selecting the top n features. It assesses the independence between a given phrase and the presence of a specific class.
 -

4.3 Predictive Models and Techniques

Various ML techniques are employed for disease prediction:

- Ensemble Methods: Combine multiple models to improve prediction accuracy and robustness.
- Survival Analysis Models: Such as Cox proportional hazards regression, are used to model time-to-event data and predict survival probabilities.
- Deep Learning Models: Including CNNs and recurrent neural networks (RNNs), can capture complex temporal dependencies in longitudinal data.

4.4 Challenges and Future Directions

Several challenges need to be addressed to enhance the effectiveness of AI and ML in disease prediction:

- Data Quality and Availability: Ensuring access to high-quality, well-curated datasets is essential.
- Model Generalisability: Developing models that can generalise across different populations and healthcare settings.
- Ethical Considerations: Addressing issues related to data privacy, algorithmic bias, and the responsible use of predictive models.

Future research directions include:

- Incorporating longitudinal data to capture disease progression over time.
- Developing personalised prediction models that account for individual patient characteristics.
- Evaluating the clinical utility of AI/ML-based prediction tools in real-world settings.
- Using transfer learning to improve prediction accuracy in data-scarce environments.

5. AI and Machine Learning in Drug Response Prediction

Predicting how patients will respond to specific drugs is vital for personalised medicine, enabling clinicians to select the most effective treatments while minimising adverse effects. This section explores the application of AI and ML techniques in forecasting drug responses, with a focus on anti-cancer drug response prediction.

5.1. Molecular Representation and Feature Engineering

- Graph Neural Networks (GNNs) Enhanced predictive power through meticulous molecular depiction.
- Morgan Algorithm Adaptation: Utilising the Morgan algorithm, traditionally employed for computing Extended-Connectivity Fingerprints (ECFPs), to formulate node features [16].
 - Assigning the feature vector of each atom as the membership of the identified structures.
- Incorporating Chemical Bond Types: Including the type of chemical bonds as edge features to enhance molecular representation.

5.2. GNN-Based Models for Drug Response Prediction

- Relational Graph Convolutional Networks (RGCN): Describing edges as distinct relations and developing unique message-passing patterns for each.
 - Showed improved performance over GCNs without edge features.
- Dual Graph Convolutional Network (DualGCN): A model to predict cancer drug response [17].
- Graph Attention Networks: Pushing the boundaries of molecular representation for drug discovery.

5.3 Explainable AI (XAI) for Mechanistic Insights

- XGDP Framework: A novel framework to predict drug response levels and uncover the underlying mechanisms of action.
- GNNExplainer: An interpretable method to provide explanations for predictions of any GNN-based model on any graph-based machine learning task.
 - A model-agnostic approach for providing interpretable explanations for predictions of any GNN-based model on any graph-based machine learning task, and could avoid issues related to gradient-based methods.
- Application of CGMega: Utilization of GNNExplainer help provide a comprehensive evaluation for all gene neighbors (genes that connect with predicted gene via PPIs) and molecular features (including SNVs, CNVs [18], histone modifications, chromatin accessibility, and 3D genome architectures) in a masking-based manner.

5.4. Challenges and Future Directions

- Limited GPU Resources: Hindering the training of complex models like RGAT with optimal batch sizes.
- Data Heterogeneity: Handling the diverse nature of drug response data, including genomic, proteomic, and clinical information.
- Model Interpretability: Enhancing the transparency and interpretability of AI/ML models to facilitate clinical adoption.

Future research directions include:

- Exploring knowledge transfer between different cancer types to predict cancer-specific driver genes.
- Further investigation of relational GNNs with optimized training configurations.
- Combining diverse data sources to build more comprehensive and accurate predictive models.
- Validating AI/ML-based predictions in prospective clinical trials to assess their real-world utility.

6. AI and Machine Learning in Acute Lymphoblastic Leukemia (ALL) Detection and Classification

Acute Lymphoblastic Leukemia (ALL) is a type of cancer that originates in the blood and stems from white blood cells (WBCs) in the bone marrow [19]. Early and accurate detection of ALL is crucial for effective treatment and improved survival rates. This section explores the application of AI and ML techniques in the detection and classification of ALL, focusing on image analysis of blood smear images and microarray data analysis.

6.1. Image Analysis of Blood Smear Images

- Machine Learning (ML) and Deep Learning (DL) Techniques: Utilised to examine blood smear images to identify, classify, and count cells in different types of leukaemia. Numerous ML and DL techniques have been employed, each with different features and levels of accuracy.
- Optimised CNN (OCNN): An optimised Convolutional Neural Network (CNN) is used for classification, detecting and classifying images as "normal" or "abnormal". Fuzzy optimisation is used to optimise the CNN hyperparameters, enhancing performance.
 - An OCNN classifier, after hyperparameter optimisation, achieved 99.99% accuracy with the C-NMC_Leukemia dataset.
- 3SNet Model: A model using Grey and Histogram of Oriented Gradients (HOG) features that achieved high performance measures in leukaemia detection. The average accuracy was 96.13%, precision 86.29%, recall 84.61%, and F1-Score 85.42%.
- Deep Learning-Based Feature Extraction and Ensemble Learning: Utilised for lung and colon cancer detection.
- Stacked Convolutional Neural Network: Used for skin lesion classification in dermoscopic images.
- Modified Post Transfer Learning with MobileNetV2 Architecture: Applied to classify acute lymphoblastic leukaemia.
- UNET Based Segmentation and Statistical Analysis of Fused Deep Features: Used for automatic detection of acute lymphoblastic leukaemia[20].
- Convolutional Neural Networks: Used for automatic detection and classification of leukocytes.

6.2. Microarray Data Analysis

- Microarray techniques offer detailed examination of methodologies for gene expression analysis.
- Feature Selection Methods:
 - Chi-square: A non-parametric statistical method used for feature selection, specifically selecting the top n features. It assesses the independence between a given phrase and the presence of a specific class.
- Ensemble Convolutional Networks (ECN):
 - ECN-2: A sequential convolution model with a 1-D convolutional layer and additional layers to prevent overfitting.
 - ECN-3: A sequential network with a 1-D Convolutional Layer and a softmax classifier. Adam is utilised as the optimizer for all models in the ensemble.

6.3. Challenges and Future Directions

- Time-Consuming and Challenging Diagnosis: Diagnosing leukaemia can be time-consuming and challenging, prompting the increased adoption of ML approaches to expedite and streamline the diagnostic process.
- Data Quality and Availability: Ensuring access to high-quality, well-curated datasets for training and validation of AI/ML models.
 - The C-NMC 2019 dataset is one such resource, containing lymphocyte images collected from subjects and divided into training and testing sets.
- Model Generalisability: Developing models that can generalise across different patient populations and laboratory settings.
- Interpretability: Improving the interpretability of AI/ML models to provide insights into the factors driving predictions.

Future research directions include:

- Integrating multi-omics data to improve the accuracy and robustness of ALL detection and classification.
- Using advanced GNN techniques to model the complex relationships between genes and other molecular features in ALL.
- Developing personalised prediction models that can tailor treatment strategies based on individual patient characteristics.
- Evaluating the clinical utility of AI/ML-based diagnostic tools in real-world settings to assess their impact on patient outcomes.
- Combining CNNs with other machine learning algorithms for enhanced classification accuracy.
- Applying transfer learning to improve diagnostic accuracy in resource-limited settings.

7. AI and Machine Learning in Cancer Staging and Prognosis

Cancer staging, such as the Tumour/Node/Metastasis (TNM) system, is widely used to classify cancer development, guiding further investigation and understanding at molecular and clinical levels. This section explores how AI and ML techniques contribute to cancer staging and prognosis, enhancing the precision and personalisation of cancer treatment strategies[21].

7.1. Graph-based Approaches

- Sample-Specific Networks: Utilising graph-based methods to construct sample-specific networks for personalised characterisation of diseases.
- Graph Neural Networks (GNNs): Employing GNNs to model complex relationships within biological data.
 - Enable the identification of significant subgraphs and genes associated with cancer development and progression.
- Graph Pooling Methods: Utilising graph pooling methods like SAGpool, ASAP, DGCNN, Graph U-Net, GMN and Diffpool to refine graph representations[22].

7.2. Identification of Significant Genes and Pathways

- RHOA, CTNNA2, and SIX2: Identifying genes such as RHOA, CTNNA2, and SIX2, which are associated with cancer development and progression, particularly in late stages.
 - RHOA: Overexpression is common in breast cancer and is involved in cytoskeletal organisation and tumour cell motility.
 - CTNNA2: Identified as a tumour suppressor in laryngeal cancer, with mutations related to migration and invasion advantages[23].
 - SIX2: Expression is significantly increased in non-small cell lung cancer (NSCLC) tissues and is correlated with advanced tumour stages.
- KEGG Pathway Analysis: Using KEGG pathway analysis to identify stage-specific processes in different cancers.
 - EGFR tyrosine kinase inhibitor resistance: Identified in early-stage BRCA.
 - Parathyroid hormone synthesis, secretion, and action: Identified in late-stage BRCA.

7.3. Survival Analysis

- Cox Proportional Hazards Regression Loss: Minimising the negative log partial likelihood of Cox proportional hazards regression loss to evaluate prognostic risk.
- Intratumour Heterogeneity: Analysing intratumour heterogeneity as a prognostic determinant of survival across different cancer types.

7.4. Challenges and Future Directions

- Complexity of Cancer: Addressing the complex, dynamic, and progressive nature of cancer involving diverse gene-environment interactions.
- Limited Survival Rates: Improving the comprehension of cancer development, especially in early stages, to increase survival rates.
- Data Integration: Combining multi-omics data to provide a comprehensive view of cancer biology and improve prognostic accuracy.
- Explainability: Enhancing the interpretability of AI/ML models to identify key factors influencing cancer progression and treatment response.

Future research directions include:

- Developing more sophisticated GNN architectures to capture complex dependencies in cancer networks.
- Integrating spatial information from imaging data with genomic and proteomic data to refine cancer staging.
- Creating personalised models that incorporate individual patient characteristics and tumour profiles to predict prognosis accurately.
- Validating AI/ML-based prognostic tools in prospective clinical trials to evaluate their impact on patient outcomes.
- Refining methods to identify personalised driver genes in cancer.

8. AI and Machine Learning in Chronic Disease Prediction

Chronic diseases, such as Cardiovascular Disease (CVD) and Chronic Pulmonary Disease (CPD), are significant global health challenges. Predicting the onset and progression of these diseases is crucial for early intervention and management. This section explores the application of AI and ML techniques in predicting chronic diseases using various data sources, including administrative claims data and electronic health records.

8.1. GNN-Based Frameworks

- Weighted Graph-Based Frameworks:
 - Utilise Graph Neural Networks (GNNs) to predict chronic diseases using administrative claim data.
 - Automatically learn node embeddings from patient networks constructed from healthcare claim data.
 - Model patients' latent relationships from comprehensive weighted patient networks.
- Bipartite Graph Creation:
 - Create bipartite graphs and use bipartite projection techniques to construct weighted patient networks (WPN).
 - Use edge lists created from WPNs and patient features to train and test GNN-based models.

8.2. Feature Selection and Analysis

- Network Features:
 - Employ network features on GNN-based models for node classification and graph classification tasks.
 - Features include degree, DeepWalk, and PageRank.
- Association of Node Characteristics and Labels:
 - GNNs perform well when node characteristics and node labels have a strong association.

8.3. Data Sources and Integration

- Administrative Data:
 - Use administrative data provided by private health funds to create study cohorts.
- Electronic Health Records (EHR):
 - Mine electronic health records to improve research applications and clinical care.
- Comprehensive Weighted Patient Network:
 - Model patients' latent relationships from a comprehensive weighted patient network.

8.4. Challenges and Future Directions

- Loss of Latent Information:
 - Some latent information may be lost if it is impossible to use all network features.
- Generalisability:
 - Developing models that can generalise across different patient populations and healthcare systems.
- Interpretability:
 - Enhancing the interpretability of AI/ML models to provide insights into the factors driving predictions.

Future research directions include:

- Combining network features with other patient characteristics to enhance prediction accuracy.
- Using more sophisticated GNN architectures to capture complex relationships in patient networks.
- Developing personalised prediction models that can tailor intervention strategies based on individual patient characteristics.
- Evaluating the clinical utility of AI/ML-based predictive tools in real-world settings to assess their impact on patient outcomes.
- Exploring the integration of multi-modal data sources, such as genomic data and lifestyle factors, to improve prediction accuracy.

8. AI and Machine Learning in Chronic Disease Prediction

Chronic diseases, such as Cardiovascular Disease (CVD) and Chronic Pulmonary Disease (CPD), are significant global health challenges. Predicting the onset and progression of these diseases is crucial for early intervention and management. This section explores the application of AI and ML techniques in predicting chronic diseases using various data sources, including administrative claims data and electronic health records.

8.1. GNN-Based Frameworks

- Weighted Graph-Based Frameworks:
 - Utilise Graph Neural Networks (GNNs) to predict chronic diseases using administrative claim data.
 - Automatically learn node embeddings from patient networks constructed from healthcare claim data.
 - Model patients' latent relationships from comprehensive weighted patient networks.
- Bipartite Graph Creation:

- Create bipartite graphs and use bipartite projection techniques to construct weighted patient networks (WPN).
- Use edge lists created from WPNs and patient features to train and test GNN-based models.

8.2. Feature Selection and Analysis

- Network Features:
 - Employ network features on GNN-based models for node classification and graph classification tasks.
 - Features include degree, DeepWalk, and PageRank.
- Association of Node Characteristics and Labels:
 - GNNs perform well when node characteristics and node labels have a strong association.

8.3. Data Sources and Integration

- Administrative Data:
 - Use administrative data provided by private health funds to create study cohorts.
- Electronic Health Records (EHR):
 - Mine electronic health records to improve research applications and clinical care.
- Comprehensive Weighted Patient Network:
 - Model patients' latent relationships from a comprehensive weighted patient network.

8.4. Challenges and Future Directions

- Loss of Latent Information:
 - Some latent information may be lost if it is impossible to use all network features.
- Generalisability:
 - Developing models that can generalise across different patient populations and healthcare systems.
- Interpretability:
 - Enhancing the interpretability of AI/ML models to provide insights into the factors driving predictions.

Future research directions include:

- Combining network features with other patient characteristics to enhance prediction accuracy.
- Using more sophisticated GNN architectures to capture complex relationships in patient networks.
- Developing personalised prediction models that can tailor intervention strategies based on individual patient characteristics.
- Evaluating the clinical utility of AI/ML-based predictive tools in real-world settings to assess their impact on patient outcomes.
- Exploring the integration of multi-modal data sources, such as genomic data and lifestyle factors, to improve prediction accuracy.

10. AI and Machine Learning in Leukaemia Detection and Classification

Leukaemia, a cancer of the blood-forming tissues, necessitates early and accurate detection for effective treatment. Traditional diagnostic methods can be laborious and resource-intensive, making the application of AI and ML particularly valuable. This section reviews the use of AI and ML in leukaemia detection and classification, highlighting key techniques and datasets.

10.1. Image-Based Detection and Classification

- Optimised Convolutional Neural Networks (OCNN):
 - Utilising OCNNs for classifying blood microscopic images to distinguish between leukaemia-free and leukaemia-affected samples.
 - Employing fuzzy logic to optimise CNN hyperparameters, enhancing performance.
 - OCNNs detect and classify images as "normal" or "abnormal" with high accuracy.
- Convolutional Neural Networks (CNNs):
 - Applying CNNs to automatically detect and classify leukocytes in blood smear images.
 - Using CNNs to detect crucial characteristics without human interaction, improving performance by optimising hyperparameters.
- Data Augmentation:
 - Using techniques such as mirroring, rotation, blurring, shearing, and noise addition to balance training and validation sets.

10.2. Feature Extraction and Selection

- Feature Extraction Techniques:
 - Combining image pre-processing and feature extraction as essential steps in the classification process.
 - Using deep learning-based feature extractors like LeuFeatx for white blood cell image classification.
- Chi-Square Method: Employing the Chi-square method for feature selection, specifically selecting the top n features for data analysis tasks.

10.3. Datasets and Evaluation

- C-NMC_Leukemia Dataset:
 - Using the C-NMC 2019 dataset, which contains lymphocyte images collected from multiple subjects, to train and evaluate models[24].
 - Achieving high accuracy (e.g., 99.99% with OCNN) on the C-NMC_Leukemia dataset.
- Munich AML Morphology Dataset: Using the Munich AML Morphology Dataset, containing expert-labelled single-cell images.

10.4. Machine Learning Models

- Support Vector Machines (SVM):
 - Employing SVM classifiers for Acute Myeloid Leukaemia (AML) subtype classification.
 - Using SVM in conjunction with Particle Swarm Optimisation (PSO) to enhance diagnostic processes.
- Ensemble Methods:
 - Developing ensemble models, such as Enhanced Predictive Heterogeneous Ensemble Models, for leukaemia prediction.
 - Using ensemble convolutional networks (ECN) to get class probability values.
- Graph Neural Networks (GNN):
 - Using graph-based methods to construct patient networks from administrative healthcare claim data and GNNs to learn node embeddings automatically.
- Other Classifiers:
 - Comparing the performance of OCNN with common classifiers like Decision Tree (DT), Linear Discriminant (LD), Support Vector Machine (SVM), and K-Nearest Neighbour (KNN).
 - Using Naïve Bayes classifiers, which require minimal training data and offer quick convergence.

10.5 Challenges and Future Directions

- High Computational Costs: Addressing the high computational demands associated with complex models like neural networks.
- Overfitting: Implementing strategies such as dropout layers and adaptive stopping schedulers to prevent overfitting, especially with limited datasets.
- Data Dependency: Reducing the reliance on extensive feature engineering by using models that automatically deduce features.

Future research directions include:

- Enhancing the integration of diverse data types, such as genetic and proteomic data, with image-based analysis to improve diagnostic accuracy.
- Developing more robust and generalisable models that can perform effectively across different patient populations and datasets.
- Improving the interpretability of deep learning models to provide insights into the morphological features and biomarkers driving predictions.
- Creating more efficient and cost-effective diagnostic tools that can be deployed in resource-limited settings.
- Validating AI/ML-based diagnostic tools in prospective clinical trials to evaluate their impact on patient outcomes.
- Leveraging transfer learning to improve performance on cancers with less training data.

11. AI and Machine Learning in Cancer Staging

Cancer staging is a critical process in oncology that determines the extent of cancer progression, informing treatment decisions and predicting patient outcomes. Traditional cancer staging relies on the TNM (tumour, node, metastasis) system, which classifies cancer development into several stages. However, integrating molecular and genetic data with clinical information can enhance the precision and personalisation of cancer staging. This section reviews the application of AI and ML techniques in cancer staging, focusing on methodologies that leverage diverse data types to improve staging accuracy and prognostic predictions.

11.1. Graph-Based Approaches

- Sample-Specific Networks: Constructing sample-specific networks to capture individual variations in cancer. These networks are used to characterise diseases at a personalised level.
- Graph Neural Networks (GNNs):
 - Utilising GNNs to integrate complex interactions and predict cancer stages.
 - Employing GNNs to learn node embeddings from patient networks constructed from administrative healthcare claim data.
- Graph Pooling Methods: Applying graph pooling methods to extract significant subgraphs and key features relevant to cancer staging.
- Multi-Structure Graph Classification: Using multi-structure graph classification methods with attention-based pooling to improve the accuracy of cancer staging.

11.2. Integration of Multi-Omics Data

- Multi-Omics Data Analysis: Combining genomic, transcriptomic, and proteomic data to provide a comprehensive view of cancer progression.
- Feature Selection: Using methods like Chi-square to select the most relevant features from high-dimensional omics data for improved model performance.
- Cancer Gene Module Dissection:
 - Employing explainable graph attention to dissect cancer gene modules, uncovering high-order relationships between genes.
 - Using frameworks like CGMega to integrate Hi-C features, promoter densities, histone modifications, SNVs and CNVs for cancer gene prediction.
- KEGG Pathway Enrichment: Performing KEGG pathway enrichment analysis to identify stage-specific processes and pathways involved in cancer progression.

11.3. Machine Learning Models

- Survival Analysis Models:
 - Using models like Cox proportional hazards regression within GNN frameworks to predict survival outcomes.
 - Applying Graph CNN for survival analysis on whole slide pathological images.
- Deep Learning Models: Developing deep learning models based on chest X-ray images for cancer detection and staging.
- Ensemble Methods: Combining multiple machine learning models to improve prediction accuracy and robustness.

11.4. Challenges and Future Directions

- Data Heterogeneity: Addressing the challenges of integrating heterogeneous data types and formats.
- Interpretability: Enhancing the interpretability of complex AI/ML models to provide biological insights into cancer staging.
- Personalisation: Developing personalised models that account for individual patient characteristics and tumour heterogeneity.
- Validation: Validating AI/ML-based staging tools in clinical trials to ensure their reliability and effectiveness.
- Computational Complexity: Managing the computational demands of graph-based and deep learning approaches.

Future research directions include:

- Developing more sophisticated GNN architectures that can effectively capture the complex interactions within cancer networks.
- Integrating spatial transcriptomics data to account for tumour microenvironment heterogeneity.
- Improving the generalisability of AI/ML models across diverse patient populations and cancer subtypes.
- Creating user-friendly tools that enable clinicians to easily access and interpret AI/ML-based staging predictions.
- Combining AI/ML-based staging with liquid biopsies for non-invasive monitoring of cancer progression.

12. AI and Machine Learning in Drug Response Prediction

Predicting how patients will respond to specific drugs is a significant challenge in cancer treatment. AI and ML offer powerful tools to analyse complex molecular and clinical data to forecast drug response, enabling more personalised and effective treatment strategies. This section explores the application of AI and ML in drug response prediction, highlighting methodologies, data integration techniques, and future directions.

12.1. Molecular Representation and Feature Engineering

- Genomic Signatures:
 - Using cancer genomic signatures to predict drug response.
 - Analysing mutations, gene expression levels, and copy number variations to identify predictive biomarkers.
- Chemical Structure Encoding:
 - Employing methods like SMILES and ECFP (Extended Connectivity Fingerprints) to represent drug structures.
 - Adapting the Morgan algorithm to compute ECFPs for node features, enhancing the predictive power of GNN models.
- Multi-Omics Integration: Combining genomic, transcriptomic, and proteomic data to provide a comprehensive view of drug response determinants.

12.2. Graph-Based Approaches

- Graph Neural Networks (GNNs):
 - Utilising GNNs to model the interactions between drugs and biological systems.

- Employing GNNs for drug repositioning over heterogeneous information networks.
- Exploring relational GNNs to describe edges as different relations and develop distinct message-passing patterns.
- Drug-Target Interaction Prediction: Predicting drug-target interactions using GNNs and other machine-learning techniques.
- Knowledge Graph Construction: Building knowledge graphs to represent the relationships between drugs, targets, and diseases.

12.3. Machine Learning Models

- Regression Models: Using regression models to predict drug response levels based on molecular features.
- Deep Learning Models:
 - Developing deep learning models for drug response prediction from cancer genomic signatures and compound chemical structures.
 - Applying convolutional neural networks and other deep-learning architectures to analyse molecular data.
- Ensemble Methods: Combining multiple machine learning models to improve prediction accuracy and robustness.
- Support Vector Machines (SVM): Using SVM in conjunction with Particle Swarm Optimisation (PSO) to enhance diagnostic processes.

12.4. Explainability and Interpretability

- Explainable AI (XAI):
 - Using techniques like GNNExplainer to interpret contributing factors in drug response prediction.
 - Applying model-agnostic interpretation approaches to dissect cancer gene modules[26].
- Saliency Maps: Generating saliency maps to visualise the important features influencing drug response predictions.

12.5. Challenges and Future Directions

- Data Availability and Quality: Addressing the limited availability of high-quality, well-annotated drug response data.
- Model Generalisability: Improving the generalisability of drug response prediction models across diverse patient populations and cancer subtypes.
- Integration of Clinical Data: Incorporating clinical data, such as patient demographics and treatment history, to enhance prediction accuracy.
- Computational Complexity: Managing the computational demands of complex models like GNNs and deep learning architectures.

Future research directions include:

- Developing more sophisticated GNN architectures that can effectively capture the complex interactions within biological systems.
- Integrating 3D genome architecture to improve cancer gene module dissection.
- Creating user-friendly tools that enable clinicians to easily access and interpret AI/ML-based drug response predictions.
- Validating AI/ML-based predictive tools in prospective clinical trials to evaluate their impact on patient outcomes.
- Leveraging transfer learning to improve performance on cancers with less training data.

13. AI and Machine Learning in Other Diseases

While AI and ML have shown promise in cancer research, their applications extend to various other diseases. This section reviews the use of AI and ML in detecting, predicting, and managing conditions beyond cancer, demonstrating the versatility and broad applicability of these technologies.

13.1. Cardiovascular Diseases (CVD)

- Prediction and Risk Assessment:
 - Utilising ML models to predict the risk of cardiovascular diseases.
 - Employing GNN-based techniques to aggregate information directly from patient networks constructed from administrative healthcare claim data.
- Analysis of Electrocardiogram (ECG) Data: Applying deep learning models to analyse ECG data for detecting cardiac abnormalities.
- Integration of Clinical and Genetic Data: Combining clinical data with genetic information to improve the accuracy of CVD risk prediction.

13.2. Chronic Pulmonary Diseases (CPD)

- Prediction and Diagnosis:
 - Using ML models to predict and diagnose chronic pulmonary diseases.
 - Applying GNN-based techniques to model patient relationships and predict disease outcomes.

- Analysis of Medical Images: Employing convolutional neural networks to analyse chest X-rays and CT scans for detecting lung abnormalities.
- Personalised Treatment Strategies: Developing personalised treatment strategies based on AI-driven analysis of patient data.

13.3. Neurological Disorders

- Brain Tumour Segmentation:
 - Using deep learning models for brain tumour segmentation in multimodal MRI images.
 - Applying novel LSIS operators and deep learning techniques for enhanced segmentation accuracy.
- Alzheimer's Disease Prediction: Predicting the onset and progression of Alzheimer's disease using ML models and neuroimaging data.
- Epilepsy Detection: Detecting epileptic seizures using AI algorithms applied to EEG data.

13.4. Infectious Diseases

- Pneumonia Detection:
 - Using deep learning models based on chest X-ray images for pneumonia detection.
 - Developing deep-pneumonia frameworks for improved diagnostic accuracy.
- COVID-19 Diagnosis:
 - Employing customised efficient neural networks for COVID-19 infected region identification in CT images.
- Tuberculosis Screening: Utilising AI algorithms to screen for tuberculosis from chest X-rays in resource-limited settings.

13.5. Other Diseases

- Diabetes Prediction:
 - Utilizing fog computing and explainable deep learning techniques for gestational diabetes prediction.
- Leukemia Detection:
 - Applying machine learning in the detection and classification of leukemia using the C-NMC_Leukemia dataset.
 - Using optimized CNNs (OCNN) for classifying microscopic blood images to distinguish between leukemia-free and leukemia-affected.
 - Employing transfer learning for white blood cell image classification.
- Skin Lesion Classification:
 - Using stacked convolutional neural network for skin lesion classification in dermoscopic images.

13.6. Challenges and Future Directions

- Data Standardisation: Addressing the lack of standardisation in medical data to facilitate the development of robust AI/ML models.
- Model Validation: Ensuring the reliability and generalisability of AI/ML models through rigorous validation on diverse datasets.
- Ethical Considerations: Addressing ethical considerations related to data privacy, algorithmic bias, and the responsible use of AI in healthcare.
- Integration with Clinical Workflows: Integrating AI/ML-based tools into clinical workflows to support clinicians in decision-making.
- Explainability and Transparency: Enhancing the explainability and transparency of AI/ML models to build trust among clinicians and patients.

Future research directions include:

- Developing AI/ML models that can integrate diverse data types, including clinical, imaging, and omics data, to provide a holistic view of disease.
- Creating user-friendly tools that enable clinicians to easily access and interpret AI/ML-based predictions.
- Validating AI/ML-based tools in prospective clinical trials to evaluate their impact on patient outcomes.
- Leveraging federated learning to train AI/ML models on distributed datasets while preserving patient privacy.
- Exploring the use of AI/ML in drug discovery and development for diseases beyond cancer.

14. AI and Machine Learning Methodologies

To facilitate a comprehensive understanding of the applications of AI and ML in disease research, it is essential to outline the primary methodologies employed. This section provides an overview of the machine learning algorithms, deep learning architectures, and data processing techniques commonly used in the studies discussed.

14.1. Machine Learning Algorithms

- Supervised Learning:

- Support Vector Machines (SVM): Effective for classification and regression tasks, particularly in scenarios with high-dimensional data. SVM is also used in conjunction with Particle Swarm Optimisation (PSO) to enhance diagnostic processes.
- Decision Trees: Used for classification and regression, offering interpretability and ease of implementation.
- K-Nearest Neighbors (KNN): A simple yet effective algorithm for classification and regression based on proximity to training data.
- Naïve Bayes: A probabilistic classifier based on Bayes' theorem, often used for its speed and simplicity.
- Linear Discriminant Analysis (LDA): A dimensionality reduction technique used for classification problems.
- Unsupervised Learning:
 - Clustering Algorithms (e.g., K-Means): Used to identify patterns and group similar data points, aiding in the discovery of disease subtypes and patient stratification.
- Ensemble Methods:
 - Random Forests: An ensemble of decision trees, providing improved accuracy and robustness compared to individual trees.
 - Boosting Algorithms (e.g., XGBoost): Combining multiple weak learners to create a strong learner, often used for high-performance prediction.

14.2. Deep Learning Architectures

- Convolutional Neural Networks (CNN):
 - Effective for image analysis, particularly in medical imaging applications such as detecting abnormalities in X-rays and microscopic images.
 - OCNN (Optimized CNN) is used for classifying microscopic blood images to distinguish between leukemia-free and leukemia-affected images.
 - Can automatically detect crucial characteristics without human interaction.
- Recurrent Neural Networks (RNN):
 - Useful for sequential data analysis, such as time-series gene expression data or patient history records.
- Graph Neural Networks (GNN):
 - Designed for graph-structured data, enabling the modelling of relationships between genes, proteins, and other biological entities.
 - Used to model interactions between drugs and biological systems.
 - scGNN (single-cell graph neural network) is a novel GNN framework for single-cell RNA-Seq analyses.
 - CGMega is an explainable and graph attention-based deep learning framework to perform cancer gene module dissection.
 - Employed for drug repositioning over heterogeneous information networks.
- Autoencoders:
 - Used for dimensionality reduction and feature learning, enabling the extraction of relevant information from high-dimensional data.

14.3. Data Processing Techniques

- Feature Selection:
 - Chi-square: A non-parametric statistical method used for feature selection, specifically selecting the top n features.
 - Selecting relevant features from high-dimensional datasets to improve model performance and interpretability.
- Dimensionality Reduction:
 - Principal Component Analysis (PCA): Reducing the number of variables in a dataset while retaining essential information.
- Data Augmentation:
 - Increasing the size of the training dataset by applying transformations to existing data, particularly useful in image analysis tasks.
- Normalisation and Standardisation:
 - Scaling data to a standard range to prevent variables with larger values from dominating the analysis.
- Handling Missing Data:
 - Employing imputation techniques to fill in missing values in datasets, ensuring completeness and accuracy.

14.4. Explainable AI (XAI) Techniques

- GNNExplainer: A model-agnostic approach for providing interpretable explanations for predictions of any GNN-based model on any graph-based machine learning task.
- Attention Mechanisms:
 - Used to identify the most important features or connections in a neural network.
- Saliency Maps:
 - Visualising the parts of an input (e.g., an image) that most influence the model's prediction.

14.5. Performance Evaluation Metrics

- Accuracy, Precision, Recall, and F1-Score:
 - Common metrics for evaluating the performance of classification models.
- AUC-ROC (Area Under the Receiver Operating Characteristic Curve):
 - A measure of the ability of a classifier to distinguish between classes.
- RMSE (Root Mean Squared Error):
 - A metric for evaluating the performance of regression models.
- PCC (Pearson Correlation Coefficient):
 - Measures the linear correlation between predicted and actual values.

14.6. Challenges and Future Directions

- Algorithm Selection: Determining the most appropriate AI/ML algorithm for a specific disease and dataset.
- Hyperparameter Optimization: Tuning the parameters of AI/ML models to achieve optimal performance.
- Model Interpretability: Improving the interpretability of complex AI/ML models to facilitate clinical adoption.
- Scalability: Developing AI/ML models that can scale to large datasets and handle complex computational tasks.
- Reproducibility: Ensuring the reproducibility of AI/ML research through standardised data processing and model validation techniques.

Future research directions include:

- Developing new AI/ML algorithms that are specifically tailored to the unique characteristics of medical data.
- Creating automated machine learning (AutoML) platforms that can streamline the process of model selection and hyperparameter optimization.
- Integrating AI/ML with other advanced technologies, such as cloud computing and edge computing, to enable real-time disease detection and prediction.
- Establishing best practices for the development and validation of AI/ML-based diagnostic and prognostic tools in medicine.

15. Conclusion

This survey has provided an overview of the current state of AI and ML in disease detection, prediction, and drug response. AI and ML are emerging as powerful tools with the potential to revolutionise healthcare. By leveraging large datasets, advanced algorithms, and explainable AI techniques, these approaches offer new insights into disease mechanisms, personalised treatment strategies, and improved healthcare delivery. To unlock the full potential of AI and ML in healthcare, future efforts should focus on addressing the challenges related to data scarcity, interpretability, generalisability, and ethical considerations. With continued innovation and collaboration, AI and ML can transform medical practice, improve patient outcomes, and create a healthier future for all.

References

- [1] Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection—harnessing the Web for public health surveillance. *The New England journal of medicine*. 2009 May 7;360(21):2153.
- [2] Duetz C, Bachas C, Westers TM, van de Loosdrecht AA. Computational analysis of flow cytometry data in hematological malignancies: future clinical practice?. *Current Opinion in Oncology*. 2020 Mar 1;32(2):162-9.
- [3] Lu H, Uddin S. A weighted patient network-based framework for predicting chronic diseases using graph neural networks. *Scientific reports*. 2021 Nov 19;11(1):22607.
- [4] Adam G, Rampášek L, Safikhani Z, Smirnov P, Haibe-Kains B, Goldenberg A. Machine learning approaches to drug response prediction: challenges and recent progress. *NPJ precision oncology*. 2020 Jun 15;4(1):19.
- [6] Hosseini M, Bani-Hani D, Lam SS. Leukocytes image classification using optimized convolutional neural networks. *Expert Systems with Applications*. 2022 Nov 1;205:117672.

- [7]Fratello M, Tagliaferri R. Decision trees and random forests. Encyclopedia of bioinformatics and computational biology: ABC of bioinformatics. 2018 Aug 21;1(S3):374.
- [8]Luong DT, Chandola V. A k-means approach to clustering disease progressions. In2017 IEEE International conference on healthcare informatics (ICHI) 2017 Aug 23 (pp. 268-274). IEEE.
- [9]Yu G, Sun K, Xu C, Shi XH, Wu C, Xie T, Meng RQ, Meng XH, Wang KS, Xiao HM, Deng HW. Accurate recognition of colorectal cancer with semi-supervised deep learning on pathological images. Nature communications. 2021 Nov 2;12(1):6311.
- [10]Ramadhan NG, Maharani W, Gozali AA. Chronic diseases prediction using machine learning with data preprocessing handling: a critical review. IEEE Access. 2024 May 29.
- [11]Ahmed IA, Senan EM, Shatnawi HS, Alkhraisha ZM, Al-Azzam MM. Hybrid techniques for the diagnosis of acute lymphoblastic leukemia based on fusion of CNN features. Diagnostics. 2023 Jan;13(6):1026.
- [12]Ratley A, Minj J, Patre P. Leukemia disease detection and classification using machine learning approaches: a review. In2020 First International Conference on Power, Control and Computing Technologies (ICPC2T) 2020 Jan 3 (pp. 161-165). IEEE.
- [13]Xiao S, Lin H, Wang C, Wang S, Rajapakse JC. Graph neural networks with multiple prior knowledge for multi-omics data analysis. IEEE Journal of Biomedical and Health Informatics. 2023 Jun 12;27(9):4591-600.
- [14]Liu, Q., Hu, Z., Jiang, R. and Zhou, M., 2020. DeepCDR: a hybrid graph convolutional network for predicting cancer drug response. Bioinformatics, 36(Supplement_2), pp.i911-i918.
- [15]Oh, S.M., Stefani, K.M. and Kim, H.C., 2014. Development and application of chronic disease risk prediction models. Yonsei medical journal, 55(4), pp.853-860.
- [16]Zhang B, Zhang J, Xia Y, Chen P, Wang B. Prediction of electron ionization mass spectra based on graph convolutional networks. International Journal of Mass Spectrometry. 2022 May 1;475:116817.
- [17]Ma T, Liu Q, Li H, Zhou M, Jiang R, Zhang X. DualGCN: a dual graph convolutional network model to predict cancer drug response. BMC bioinformatics. 2022 Apr 15;23(Suppl 4):129.
- [18]Li, H., Han, Z., Sun, Y., Wang, F., Hu, P., Gao, Y., Bai, X., Peng, S., Ren, C., Xu, X. and Liu, Z., 2024. CGMega: explainable graph neural network framework with attention mechanisms for cancer gene module dissection. Nature Communications, 15(1), p.5997.
- [19]Das PK, Diya VA, Meher S, Panda R, Abraham A. A systematic review on recent advancements in deep and machine learning based detection and classification of acute lymphoblastic leukemia. IEEE access. 2022 Aug 3;10:81741-63.
- [20]Elsayed, Basel, Mohamed Elhadary, Raghad Mohamed Elshoeibi, Amgad Mohamed Elshoeibi, Ahmed Badr, Omar Metwally, Raghad Alaa ElSherif et al. "Deep learning enhances acute lymphoblastic leukemia diagnosis and classification using bone marrow images." Frontiers in Oncology 13 (2023): 1330977.
- [21]Hunter, B., Hindocha, S. and Lee, R.W., 2022. The role of artificial intelligence in early cancer diagnosis. Cancers, 14(6), p.1524.
- [22]Hou, W., Wang, Y., Zhao, Z., Cong, Y., Pang, W. and Tian, Y., 2024. Hierarchical graph neural network with subgraph perturbations for key gene cluster discovery in cancer staging. Complex & Intelligent Systems, 10(1), pp.111-128.
- [23]Fanjul-Fernández, M., Quesada, V., Cabanillas, R., Cadiñanos, J., Fontanil, T., Obaya, Á., Ramsay, A.J., Llorente, J.L., Astudillo, A., Cal, S. and López-Otín, C., 2013. Cell–cell adhesion genes CTNNA2 and CTNNA3 are tumour suppressors frequently mutated in laryngeal carcinomas. Nature communications, 4(1), p.2531.
- [24]Gupta, R., Gehlot, S. and Gupta, A., 2022. C-NMC: B-lineage acute lymphoblastic leukaemia: A blood cancer dataset. Medical Engineering & Physics, 103, p.103793.



[26]Dwivedi R, Dave D, Naik H, Singhal S, Omer R, Patel P, Qian B, Wen Z, Shah T, Morgan G, Ranjan R. Explainable AI (XAI): Core ideas, techniques, and solutions. ACM Computing Surveys. 2023 Jan 13;55(9):1-33

• ,