A Survey of Artificial Intelligence Methods for Clinical Trial Outcome Prediction

Long Qian[1], Xin Lu[1], Parvez Haris[2], Jianyong Zhu[3], Shuo Li[1], Yingjie Yang[1*]

[1]Faculty of Computing Engineering Media, De Montfort University, Leicester, UK

[2]Faculty of Health & Life Sciences, De Montfort University, Leicester, UK

[3]Scitops Corporation, Shanghai, China

*Corresponding author: Yingjie Yang (yyang@dmu.ac.uk)

Abstract

Clinical trials are crucial for drug development, but they require significant time and financial resources. Additionally, uncertainties may arise during these trials concerning their results due to concerns surrounding effectiveness, safety, or the enrollment of participants. If robust AI (artificial intelligence) models exist that can accurately forecast clinical trial results, it would effectively prevent potential failures in such trials and also speed up the drug discovery process. Consequently, more resources could be allocated towards potentially successful trials, ultimately enhancing the success rate of new drug development. This article systematically reviews the research works on the three main scenarios of AI affecting clinical trial outcomes. Clinical text embedding, complex trial relations and trial prediction methods. Then, the challenges and opportunities of predicting clinical trial outcomes is discussed in real-world applications.

1. Introduction

Although the development of modern molecular biology disciplines, such as genomics, proteomics, and bioinformatics, has brought great strides to drug R&D theory, new drug development has not escaped empirical nature due to the complexity of the biochemical reactions that drug molecules undergo in humans. Traditional drug research and development is dominated by medicinal chemistry experts, who typically empirically conduct drug screens for every 5,000 to 1,0000 compounds proposed, with only one compound ultimately eligible for clinical testing and eventual marketing. A new drug takes more than 10 years and costs nearly $2.6 billion from development to approval to the market, with a clinical success rate of less than 10% [1]. The long R&D cycle, high R&D cost and low success rate have become three huge barriers to the development of new drugs [2].

The application of AI technology in natural language processing and image recognition is attributed to its exceptional capacity for handling vast quantities of data. In recent years, AI has also been applied at different stages of new drug development including target identification [4], prediction of absorption, distribution, metabolism, excretion and toxicity (ADMET) properties [5-9], virtual screening [10], De novo drug design [11], automated synthesis [12], and precision medicine [13].

There are several milestones for AI for drug discovery. For example, AlphaFold2 accurately predicted protein structure [14], considered one of the most challenging tasks in computational biology. Segler et al. [15] utilized Monte Carlo tree search (MCTS) to combine three distinct neural networks trained on all available published reactions. This approach was employed to forecast the optimal retrosynthetic pathways for a given molecule. Zhavoronkov et al. [16] used a rein
forcement learning model to design new DDR1 kinase inhibitors and test activity in wet lab. LinearDesign [17] is capable of effectively

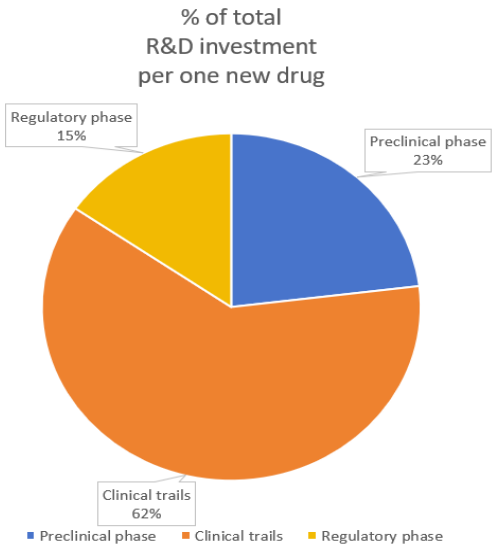

Figure 1: More than half of R&D investment is spent on clinical trials.

improving the stability of mRNA vaccine sequences and protein translation efficiency. It can complete sequence design for the COVID-19 spike protein mRNA vaccine in just 10 minutes.

These breakthroughs demonstrate the potential of AI in exploring chemical and biological data. Therefore, it is estimated that AI-assisted computational approaches can reduce the time required for traditional R&D approaches from 3 to 6 years to 1 to 2 years, spinning from target identification to the clinical candidate drug. This reduction leads to a significant gain in efficiency and cost savings. Currently, AI plus new

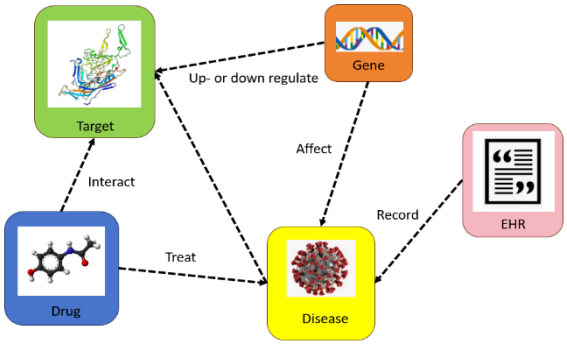

Figure 2: The illustration of different biological units and EHR interact with each other.

drug research and development has emerged as a hotspot in contemporary cutting-edge medical pharmaceutical research and entrepreneurship.

However, current AI applications are severely inadequate in predicting clinical trial outcomes. There are more articles on AI applied to drug development than to clinical trials due to factors such as data availability and complexity. In fact, efficiency gains or cost reductions in the clinical trial phase have a much larger impact on investment in new drug R&D than in the drug discovery phase as illustrated by Fig. 1. There has been a surge in the interest surrounding AI in the prediction of clinical trial outcomes due to two main factors. Firstly, the widespread adoption of electronic health records (EHRs) and electronic data capturing systems (EDC) has resulted in an unprecedented volume of patient data being available. Second, AI has had numerous successful applications in fields such as chatbot, object detection and art design.

Clinical practice heavily relies on EHRs as they provide a comprehensive and diverse range of information formats. These formats encompass various types of data that are crucial for healthcare professionals to make informed decisions and provide optimal patient care.

One important type of data found in EHRs is tabular data, which includes essential demographic information such as age, gender, and contact details. Additionally, it encompasses medical procedures performed on patients, their medical history including past illnesses or surgeries, and any relevant diagnostic test results.

Another significant component of EHRs is image data. This includes photographs capturing physical conditions or injuries, x-rays providing detailed images of bones and internal organs, computerized-tomography scans offering cross-sectional views of the body's structures, magnetic resonance imaging (MRI) scans revealing detailed anatomical information about soft tissues like muscles or organs, as well as pathology slides displaying microscopic tissue samples for analysis.

Time-series data plays a crucial role in EHRs, encompassing intermittent pulse-oximetry readings that measure oxygen saturation levels in the blood over time, blood chemistry results indicating various biochemical parameters such as glucose levels or liver function tests, respiratory analysis findings assessing lung function through spirometry measurements, electrocardiograms (ECG/EKG) recording electrical activity of the heart, ultrasounds providing real-time images during pregnancy or evaluating organ abnormalities, in-vitro test outcomes determining laboratory-based diagnostic results like viral load counts or hormone levels, and wearable sensor measurements tracking vital signs such as heart rate variability or sleep patterns.

Structured sequence data within EHRs comprises genomics information related to an individual's genetic makeup along with proteomics detailing protein expression profiles and metabolomics describing metabolic processes occurring within the body. These molecular-level insights can aid in personalized medicine approaches by identifying potential genetic predispositions to diseases or guiding targeted therapies based on specific protein markers. Lastly, unstructured sequence data adds another layer to the richness of EHRs by incorporating notes, documentation forms, completion reports written by healthcare providers during patient encounters. It also includes voice recordings capturing verbal discussions between clinicians and patients regarding symptoms or treatment plans as well as videos documenting surgical procedures for educational purposes.

Given this, the present article provides a summary of recent applications of AI in predicting clinical trial outcomes and discusses the challenges and opportunities associated with them.

2. Related Work

Clinical text embedding refers to the process of learning representations (i.e., features or embeddings) of clinical trial data that capture its underlying structure and patterns. The goal of TRL is to get an accurate latent space of data from clinical trials that can be used for downstream tasks like clinical trial outcome prediction.

Complex trial relations refer to the relations of different data type from clinical trials, such as electronic health record, imaging data, genomics data, and clinical assessments. By combining multiple data modalities, researchers can leverage complementary information and enhance the accuracy and generalizability of predictive models.

Trial prediction methods is the way of using AI models to predict the out-comes of clinical trials, this is a critical step, as it can inform clinical decision-making and facilitate the selection of the most promising trial design for further investigation.

Clinical text embedding and complex trial relations are key basements of trial prediction methods, as they enable the construction of more informative and robust models that can capture the precise representations and complex interactions between different clinical variables and predict outcomes more accurately.

2.1 Clinical text embedding

Clinical trials play a crucial role in advancing and assessing novel medical therapies and interventions. However, analyzing and interpreting clinical trial data can be challenging due to the complexity and heterogeneity of the data. Electronic health records (EHRs) and eligibility criteria (EC) are two critical sources of information that offer important insights into patient data, such as demographics, medical history, diagnoses, and treatments. EC outlines the inclusion and exclusion criteria for clinical trials, defining the eligible patient population for clinical trials.

Extracting meaningful information from EHRs and EC is essential for predicting clinical trial outcomes and improving patient care. However, the text-based nature of these data sources makes it challenging to analyze and extract useful information. Clinical text embedding, which involves transforming the raw data into a more structured and meaningful format, can help to address this challenge. By learning to recognize patterns and relationships between words and phrases, representation learning algorithms can generate high-quality representations that capture the essential information in the data.

Pre-training models, like BERT [18] have demonstrated great promise in the field of representation learning. BERT is a language model based on the transformer architecture and is pre-trained on a large

corpus of text. Throughout the pre-training phase, BERT learns to recognize patterns and connections among words by engaging in self-supervised tasks, such as masked language modeling and predicting the next sentence. The resulting contextualized word embeddings can be fine-tuned for downstream tasks such as sentiment analysis, question answering, and named entity recognition.

In the biomedical domain, researchers have applied pre-training techniques, such as BERT, to various applications, including the analysis of clinical trials. For example, BioBERT [19] is a domain-specific BERT model pre-trained on biomedical articles, achieving state-of-the-art performance on NLP tasks related to biomedical questions. Other approaches, such as Doctor2Vec [20], DeepEnrol [21], and Compose [22], also utilize pre-trained models like BERT to embed clinical trial information, yielding promising results in various tasks. MiME [23] utilizes the inherent hierarchical organization of EHRs data and the encoded associations among medical codes to address the challenge of large data volume. These approaches demonstrate the potential of pre-training techniques to advance the analysis and understanding of clinical trial data, potentially leading to improved patient outcomes and better-informed clinical decision-making.

2.2 Complex trial relations

In addition to clinical text embedding, complex trial relations involve intricate interactions and dependencies within trial data that can significantly impact the outcomes and interpretations of the trials. Trial data can be highly heterogeneous, encompassing structured data (e.g., lab results), unstructured data (e.g., clinical notes), and semi-structured data (e.g., questionnaires).

Multimodal learning is a specialized field that focuses on developing and training models capable of utilizing various types of data and their relations. The goal is to enable these models to understand the relationships between different modalities and effectively combine them to enhance prediction performance. One prominent example of a multimodal learning framework is CLIP. This framework has been trained on millions of image-text pairs and has demonstrated comparable zero-shot performance to fully supervised models [24]. Based on this idea, MedCLIP introduces a novel strategy that replaces the InfoNCE loss with a medical knowledge-based semantic matching loss. The objective of this adjustment is to tackle the issue of false negatives in contrastive learning [25].

From a biology perspective, various data modalities are critical for predicting the success of clinical trials as illustrated previously because every aspect of human function is achieved through a series of biological units.

Moreover, most biological units execute their functions by interacting with other biological units such as proteins, metabolites, small molecules, genes, and DNA. These elements react and cooperate with each other shown in Figure 2. The formidable nature of the potential intricate in diverse trial components, their intricate interconnections, and their impact on trial outcomes poses a daunting challenge.

In previous research, multimodal learning has been applied to diagnose several diseases [27-45] and address public health [46-50]. It has been proven that multimodal learning can improve the performance of downstream tasks compared to the AI models using a single modality.

Most researcher used a combination of two modalities. Yiwen et al. [50] created a Bidirectional Representation Learning model using EHR and text data to predict depression. Jordan et al. [51] used medical imaging and EHRs to classify skin lesions. Larry et al. [52] used EHR and time series (ECG) data to monitor patient, maintaining the temporal relationship by assembling ECG data into tensors. Jae et al. [53] used imaging (MRI) and genomic data (polygenic risk scores) for ADHD diagnosis. Some researchers used three modalities. Janani et al. [54] used imaging, EHR and genomic data for early detection of Alzheimer's disease. Zeng et al. [55] used imaging, EHR and text data to analyze individuals infected with COVID-19.

These methods process model-specific data using various machine-learning techniques. Subsequently, a fusion module with early, intermediate, or late strategies [56] is used to combine these features for finial prediction. However, these approaches may fail to encode different modalities during training, limiting their ability to fully exploit multimodal information.

To address this challenge, IRENE [57] utilizes bidirectional blocks that incorporate intramodal and intermodal attention, generating a comprehensive representation from both medical images and textual clinical information for the detection of pulmonary disease. It achieves superior performance compared to previous methods that employ data early or late fusion, with an average improvement of 9% and 10%, respectively. To expedite multimodal research, Aliper et al. [58] introduces the HAIM framework which facilitates the development and evaluation of AI systems utilizing multiple types of inputs.

2.3 Trial prediction methods

The process of developing and introducing a novel pharmaceutical product to the market entails a comprehensive and costly endeavor, with a low success rate. Accurately predicting clinical trial outcomes is therefore essential for economic considerations in drug development. Recent advances in artificial intelligence have enabled the use of real-world data (RWD) to predict trial outcomes with increased accuracy. Companies could even use the system to buy and sell pharmaceutical companies in financial markets [59]. This part summarizes the findings of several studies that have focused on clinical trial outcome prediction using artificial intelligence.

In clinical trials, AI has already been applied to personal disease prediction. Rajpurkar et al. [60] utilized gradient boosted decision trees to predict the progression of depressive symptoms in patients receiving antidepressant therapy, incorporating pre-treatment symptom scores and electroencephalographic measurements. Hong et al. [61] utilized an ensemble of classifiers to predict toxicity by considering drug properties and target property features. de Jong et al. [62] constructed a model that integrates genetics data to anticipate drug response in patients with neurological disorders. Wang and Sun [63] developed a transformer-based approach [64] for modeling and predicting the survival rate of breast oncology patients. Additionally, they suggested a transferable transformer model that utilizes information from various oncology trials to improve mortality predictions for individual trials, showcasing its potential adaptability across heterogeneous datasets [65].

However, while most previous works have focused on patient-level, trial-level prediction is more challenging due to the complex relationships and features among the trial components.

Gayvert et al. (2017) used the structures and properties of drugs and targets to predict drug toxicity based on a random-forest model [66]. Lo et al. (2019) explored seven commonly used classifiers and found that kNN gives the highest AUCs (0.81) in predicting drug approvals [67]. Seo et al. proposed an outer product-based convolutional neural network that employs the augmented outer product to combine chemical features of drugs and target-based features to predict the odds of clinical

Table 1: Trial outcome prediction results for three phase trials.

Method	Phase I			Phase II			Phase III		
	PR-AUC	F1	ROC-AUC	PR-AUC	F1	ROC-AUC	PR-AUC	F1	ROC-AUC
LR[67]	0.500 ± 0.005	0.604 ± 0.005	0.520 ± 0.006	0.565 ± 0.005	0.555 ± 0.006	0.587 ± 0.009	0.687 ± 0.005	0.698 ± 0.005	0.650 ± 0.007
RF[67]	0.518 ± 0.005	0.621 ± 0.005	0.525 ± 0.006	0.578 ± 0.008	0.563 ± 0.009	0.588 ± 0.009	0.692 ± 0.004	0.686 ± 0.010	0.663 ± 0.007
XGBoost[67]	0.513 ± 0.06	0.621 ± 0.007	0.518 ± 0.006	0.586 ± 0.006	0.570 ± 0.009	0.600 ± 0.007	0.697 ± 0.007	0.696 ± 0.005	0.667 ± 0.005
Ada-Boost[69]	0.519 ± 0.005	0.622 ± 0.007	0.526 ± 0.006	0.586 ± 0.009	0.583 ± 0.008	0.603 ± 0.007	0.701 ± 0.005	0.695 ± 0.005	0.670 ± 0.004
kNN+RF[67]	0.531 ± 0.006	0.625 ± 0.007	0.538 ± 0.005	0.594 ± 0.008	0.590 ± 0.006	0.597 ± 0.008	0.707 ± 0.007	0.698 ± 0.008	0.678 ± 0.010
FFNN[70]	0.547 ± 0.010	0.634 ± 0.015	0.550 ± 0.010	0.604 ± 0.010	0.599 ± 0.012	0.611 ± 0.011	0.747 ± 0.011	0.748 ± 0.009	0.681 ± 0.008
DeepEn-roll[21]	0.568 ± 0.007	0.648 ± 0.011	0.575 ± 0.013	0.600 ± 0.010	0.598 ± 0.007	0.625 ± 0.008	0.777 ± 0.008	0.786 ± 0.007	0.699 ± 0.008
COM-POSE[22]	0.564 ± 0.007	0.658 ± 0.009	0.571 ± 0.011	0.604 ± 0.007	0.597 ± 0.006	0.628 ± 0.009	0.782 ± 0.008	0.792 ± 0.007	0.700 ± 0.007
HINT[75]	0.567 ± 0.010	0.665 ± 0.010	0.576 ± 0.008	0.629 ± 0.009	0.620 ± 0.008	0.645 ± 0.006	0.811 ± 0.007	0.847 ± 0.009	0.723 ± 0.006
SPOT[77]	0.689 ± 0.009	0.714 ± 0.011	0.660 ± 0.008	0.685 ± 0.010	0.656 ± 0.009	0.630 ± 0.007	0.856 ± 0.008	0.857 ± 0.008	0.711 ± 0.005

trial outcomes. Vidhya et al. employed a combination of biological activities, physicochemical properties, target-related features, and NLP-based compound representation to accurately forecast trial outcomes. This was achieved through the integration of Graph Database and Ensemble Learning techniques [68]. Qi and Tang (2019) built a recurrent neural network model to predict phase III trial outcomes based on patient records from the previous phase II trial [71]. Siah et al. (2020) predicted drug approvals using a statistical machine learning approach that considers both drug characteristics and trial characteristics [67]. Abidi et al. (2020) developed a machine learning model that predicts clinical trial enrollment rates based on historical data and trial characteristics, allowing sponsors to optimize their recruitment strategies.

There is a lack of benchmark data for trial-related tasks, with only a small portion of clinical trial records available for use. To encourage further research and compare different trial prediction method, it would be beneficial to create and release benchmark data. One notable exception is the TOP benchmark developed by Fu et al. [75], which focuses on predicting trial outcomes. This dataset comprises information about drugs, diseases, and eligibility criteria from a total of 17,538 clinical trials. The success rates vary across phases. Phase I (1,787 trials) has a success rate of 56.3%, phase II (6,102 trials) has 49.8%, while phase III (4,576 trials) has the highest success rate at 67.8%. HINT [75] involved the integration of various sources of real-world data (RWD), such as drug compounds, disease ontology, and trial eligibility criteria. This integration was aimed at facilitating outcome predictions for trials across all phases. SPOT [77] employed a meta-learning technique to organize trials with the same subject into a chronological sequence, leveraging insights from related trials and their predictive advancements. Other methods which are not specifically designed for clinical trial outcome prediction also are compared including Logistic regression (LR) [67], Random Forest (RF) [67] , XGBoost [67], Adaptive boosting (AdaBoost) [69], k Nearest Neighbor (kNN) + RF [67] and deep learning models, such as Feedforward Neural Network (FFNN) [70], DeepEnroll [21], COMPOSE [22] . The latest results of the benchmark are shown in Table 1.

3. Challenges

However, the development of accurate and reliable representations faces several challenges. These challenges include data heterogeneity, limited data availability, data quality, interpretability, bias and generalizability, and scalability.

3.1 Comprehensive factor

As discussed in Section 2.1 regarding clinical text embedding, there are some challenges in achieving accurate representation. Firstly, EHR is longitudinal and high-dimensional which poses challenges for AI models to learn statistical properties from complex data. Secondly, the medical field involves concepts of varying granularity, making it challenging to align medical concepts across different data modalities with heterogeneous levels of detail. For instance, a patient exhibiting pleuropericardial adhesion in their EHRs may be eligible for a clinical trial focused on broader cardiovascular conditions [22]. Thirdly, detailed information such as numerical values or units is often overlooked in existing work. Information like ages, values of lab results and medication dosage could significantly impact the results. Therefore, building accurate feature requires not only professional data processing but also a well-designed network for information extraction. Fourthly, some eligibility criteria may have temporal aspects, such as a patient's medical history over a certain time frame. Modeling these temporal dependencies requires specialized techniques.

Apart from EHRs and EC, various other different biomedical components, such as molecules or proteins, also present challenges at the intersection of biology and artificial intelligence.

For molecules, constructing AI models to accurately represent them is a challenging endeavor due to the inherent complexity and vast combinatorial space of molecular structures. The challenge lies in capturing the intricate spatial arrangements of atoms, the diverse chemical bonds, and the nuanced interactions between atoms and functional groups. Additionally, molecules can exist in multiple conformations, making it essential for AI models to encompass the flexibility and dynamics of these structures. MPNN [72] performs iterative message passing between nodes in a graph, allowing information to propagate through the graph's structure to predict molecular property, SphereNet [73] takes into account the 3D position information of the node, EGNN [74] maintains molecular invariance or equivariance under certain transformations such as translation or rotation. In future work, quantum mechanical effects also need to be considered for accurate representation, but it can be computationally intensive and may require specialized techniques.

For proteins, intricate three-dimensional structures, dynamic conformational changes, and diverse functional roles are inherent to these biomolecules. Capturing the complex interplay of amino acid interactions, hydrogen bonding, hydrophobicity, and electrostatic forces demands a nuanced understanding of biophysical principles.

3.2 Data availability

Limited data availability poses a significant challenge to the progress of AI in healthcare, particularly in the context of small sample sizes and privacy concerns [76]. Additionally**,** clinical trial data can be noisy, incomplete, and error-prone, impacting the performance of the AI models.

Artificial Intelligence Generated Content (AIGC) has achieved considerable success in image generation [78-79] and chatbots [80]. It is natural to apply related techniques, such as Generative Adversarial Networks (GANs) [81] and the Diffusion model [82], to synthesize trial data for AI training. Many GAN-based approaches have been applied to the EHR generation, including CONAN [83], CorGAN [84], EHR-M-GAN [85], EMR-WGAN [86], HGAN [87], MedGan [88], MedWGAN [89], SynTEG [90]. However, these GAN-based methods have limitations when generating sparse and high-dimensional data like EHR data. To address this issue, Theodorou et al. [91] proposes a Hierarchical Autoregressive Language Model for generating longitudinal high-dimensional EHR, capturing the hierarchical distribution of EHR records and their temporal relationships without the need for variable selection or aggregation. Another consideration is that GANs are challenging to train and prone to mode collapse. EHRDiff [92] introduces diffusion models for realistic EHR synthesis, achieving better quality of synthetic EHR data for the first time.

However, when using AIGC to generate clinical data, it is crucial to also prioritize the protection of private information in real training EHR data. Researchers need to establish a valid filter for this information.

3.3 Data Imbalance

Clinical trial data can often exhibit a significant imbalance. Certain data modalities, such as imaging or genomic data, may not be available for a given clinical trial, posing a challenge in integrating diverse data types in a meaningful manner. This challenge may result in the development of models that are biased towards specific modalities, affecting their accuracy and generalizability.

To deal with issue of missing data, various imputation techniques can be utilized. Fu et al. [75] designed an imputation module to handle missing molecular data from disease and protocol. Lo et al. [67] experimented with an AI model employing four distinct imputation techniques to handle missing data.

As an increasing number of scientific research institutions disseminate their data, the issue of data imbalance is expected to be partially alleviated.

3.4 Model generalizability

The generalization of AI models is critical for their clinical application. In short, the generality of the model can be expressed in two ways: the first scenario is the performance of the prediction model on data with a similar distribution and the second scenario is predicting how the model behaves on data from different distributions. These differences may include information related to time, treatment regimens, geography, and so on. However, the reality is that most AI models perform well on training data but struggle to maintain consistent performance during internal and external independent validation, indicating poor generalization. The majority of clinical trials focus on common diseases with limited attention given to rare diseases. This imbalance presents significant challenges when applying AI models to rare disease clinical trials. One of the main obstacles faced by AI models in generalizing to rare diseases is the issue of data distribution shifts. Since most clinical trial datasets primarily consist of data from common diseases, there is a lack of diverse and representative data for rare conditions. As a result, AI models trained on such imbalanced datasets struggle to accurately predict outcomes or make informed decisions when applied to new unseen trials involving rare diseases. Furthermore, small sample sizes pose another hurdle for achieving decent out-of-distribution (OOD) performance using AI models in this context. Rare diseases often affect only a small number of individuals within the population, making it challenging to gather sufficient data for robust model training and evaluation. The scarcity of labeled samples limits the ability of AI algorithms to effectively learn patterns specific to these conditions. Addressing limitations and improving OOD performance for new, unseen trials involving rare diseases is challenging. Typically, transfer learning was devised to tackle this problem by initially pretraining certain representations on extensive unannotated datasets and subsequently adjusting them for guiding other tasks [93].

In healthcare applications, recent models for domain generalization (DG) are typically designed in collaborative settings across different institutions to eliminate the distinct covariates of each individual hospital. [94-97]. Relevant techniques have been employed to construct domain generalization methods with a wider scope, including style-based data augmentations [98-100], episodic meta-learning strategies [101-103] and domain-invariant feature learning [104] using heuristic metrics [105-106] or adversarial learning [107]. Most of these methods [108-109] limit the scope to CNN-based models [110] and batch normalization architecture [111] for image classification tasks.

Despite these improvements, achieving decent performance for new, unseen trials using AI models remains a challenge. Designing large models, such as ChatGPT, represents a promising direction for improving the generalizability of AI models in clinical trials.

3.5 Model interpretability

One key challenge is interpretability**.** The relationship between AI model-based prediction results and the occurrence, development, and associated features of diseases is unclear, relying solely on machine learning methods to analyze the data. This approach is insufficient to explain the relationship among the different components of clinical trials.

A common method involves mechanisms incorporated within deep learning models to explore whether the attention region of the model has clinical diagnostic decision-making significance [112-114]. Ma et al. [115] employed attention mechanisms and RNNs to achieve interpretable predictions of medical codes. Kang et al. [117] applied an attention mechanism to multi-omics data [118] to interpret gene expression predictions.

Other post-hoc interpretability techniques such as saliency maps rely on qualitative visual interpretations commonly used in computer vision applications. Chen et al. [119] achieved modality-specific interpretability through Grad-CAM [120] for whole slide image.

However, saliency maps may not fully meet the requirements of biological interpretability, and the associated error is significant. In addition, some relationships of biological potential may be found in medical imaging AI studies. For instance, certain imaging features with high predictive power might be linked to the high expression of specific genes or proteins. Exploring the connection between these genes or proteins and clinical endpoint events can further enhance the biological interpretability of AI models.

4. Conclusion

In this review, papers under three main topics related to clinical trial outcome prediction, clinical text embedding, complex trial relations, and trial prediction methods are systematically reviewed. These studies demonstrate that AI has the potential to extract features from multimodal biomedical data and make valid predictions. However, there is still much room for improvement, especially in terms of representing biomedical data more comprehensively, processing missing or imbalanced data, and enhancing the generalizability of AI models.

As AI techniques continue to improve and more data becomes available, it will be important for researchers, clinicians, and regulators to collaborate in addressing these challenges. This collaboration is essential for harnessing the full potential of AI in clinical trial outcome prediction, enabling us to better manage complex diseases and provide personalized medical treatment to patients.

Reference

[1] Benjamin E Blass. 2015. Basic principles of drug discovery and development. Elsevier.

[2] Martin L, Hutchens M, Hawkins C, Radnov A. How much do clinical trials cost? Nat Rev Drug Discov 2017;16(6):381 – 2.

[3] Moore TJ, Zhang H, Anderson G, Alexander GC. Estimated costs of pivotal trials for novel therapeutic agents approved by the US food and drug administration, 2015-2016. JAMA Intern Med 2018;178(11):1451 – 7.

[4] Jeon J, Nim S, Teyra J, Datti A, Wrana JL, Sidhu SS, et al. A systematic approach to identify novel cancer drug targets using machine learning, inhibitor design and high-throughput screening. Genome Med 2014;6(7):57.

[5] Göller A, Kuhnke L, Montanari F, et al. Bayer's in silico ADMET platform: a journey of machine learning over the past two decades. Drug Discov Today. 2020;25(9):1702–1709.

[6] Winiwarter S, Ahlberg E, Watson E, et al. In silico ADME in drug design – enhancing the impact. ADMET DMPK. 2018;6(1):15– 33.

[7] Beresford AP, Segall M, Tarbit MH. In silico prediction of ADME properties: are Imaking progress? Curr. Opin. Drug Discov. Devel. 2004;7(1):36–42.

[8] Norinder U, Bergström CAS. Prediction of ADMET properties. ChemMedChem. 2006;1(9):920–937.

[9] Beck B, Geppert T. Industrial applications of in silico ADMET. J. Mol. Model. 2014;20(7). DOI:10.1007/s00894-014-2322-5

[10] Xue W, Fu T, Deng S, Yang F, Yang J, Zhu F. Molecular mechanism for the allosteric inhibition of the human serotonin transporter by antidepressant escitalopram. ACS Chem Neurosci 2022;13(3):340-51.

[11] Riniker S, Wang Y, Jenkins JL, Landrum GA. Using information from historical high-throughput screens to predict active compounds. J Chem Inf Model 2014;54(7):1880 – 91.

[12] Corey EJ, Wipke WT. Computer-assisted design of complex organic syntheses. Science 1969;166(3902):178-92.

[13] König IR, Fuchs O, Hansen G, von Mutius E, Kopp MV. What is precision medicine? Eur Respir J 2017;50(4):1700391.

[14] Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Žídek A, et al. Highly accurate protein structure prediction for the human proteome. Nature 2021;596(7873):590 – 6.

[15] Segler MHS, Preuss M, Waller MP. Planning chemical syntheses with deep neural networks and symbolic AI. Nature 2018;555(7698):604-10.

[16] Zhavoronkov A, Ivanenkov YA, Aliper A, Veselov MS, Aladinskiy VA, Aladinskaya AV, et al. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. Nat Biotechnol 2019;37(9):1038–40.

[17] Zhang H, Zhang L, Lin A, et al. Algorithm for optimized mRNA design improves stability and immunogenicity[J]. Nature, 2023: 1-3.

[18] Jacob Devlin et al. Bert: Pre-training of deep bidirectional transformers for language understanding. In NAACL-HLT, 2019.

[19] Lee J, Yoon W, Kim S, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining[J]. Bioinformatics, 2020, 36(4): 1234-1240.

[20] S. Biswal et al. Doctor2vec: Dynamic doctor representation learning for clinical trial recruitment. In AAAI, 2020.

[21] Xingyao Zhang, Cao Xiao, Lucas M Glass, and Jimeng Sun. 2020. DeepEnroll:
patient-trial matching with deep embedding and entailment prediction. In Proceedings of The Web Conference 2020. 1029–1037.

[22] Junyi Gao, Cao Xiao, Lucas M Glass, and Jimeng Sun. Compose: Cross-modal pseudo-siamese network for patient trial matching. In KDD, 2020.

[23] Choi E, Xiao C, Stewart W, et al. Mime: Multilevel medical embedding of electronic health records for predictive healthcare[J]. Advances in neural information processing systems, 2018, 31.

[24] Radford, A. et al. Learning transferable visual models from natural language supervision. In Proc. 38th International Conference on Machine Learning (eds. Meila, M. & Zhang, T.) vol. 139, 8748 – 8763 (PMLR, 18 – 24 July 2021).

[25] Wang Z, Wu Z, Agarwal D, et al. Medclip: Contrastive learning from unpaired medical images and text[J]. arXiv preprint arXiv:2210.10163, 2022.

[26] Huang S C, Pareek A, Seyyedi S, et al. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines[J]. NPJ digital medicine, 2020, 3(1): 136.

[27] Agrawal, S. et al. Selection of 51 predictors from 13,782 candidate multimodal features using machine learning improves coronary artery disease prediction. Patterns 2, 100364 (2021).

[28] Bagheri, A. et al. Multimodal learning for cardiovascular risk prediction using EHR data. arXiv preprint arXiv:2008.11979 (2020).

[29] Li, P., Hu, Y. & Liu, Z.-P. Prediction of cardiovascular diseases by integrating multimodal features with machine learning methods. Biomed. Signal Process. Control 66, 102474 (2021).

[30] Liu, Y. et al. A deep learning system for differential diagnosis of skin diseases. Nat. Med. 26, 900 – 908 (2020).

[31] Stidham, R. W. Artifificial Intelligence for Understanding Imaging, Text, and Data in Gastroenterology. Gastroenterol. Hepatol. 16, 341 (2020).

[32] Paquette, A. G., Hood, L., Price, N. D. & Sadovsky, Y. Deep Phenotyping During Pregnancy for Delivery of Predictive and Preventive Medicine. Sci.Transl. Med. 12, 2 – 4 (2020).

[33] Purwar, S., Tripathi, R. K., Ranjan, R. & Saxena, R. Detection of microcytic hypochromia using cbc and blood fifilm features extracted from convolution neural network by different classififiers. Multimed. Tools Appl. 79, 4573 – 4595 (2020).

[34] Hügle, M., Kalweit, G., Hügle, T. & Boedecker, J. In Explainable AI in Healthcare and Medicine 79 – 92 (Springer, 2021).

[35] Tomašev, N. et al. A clinically applicable approach to continuous prediction of future acute kidney injury. Nature 572, 116 – 119 (2019).

[36] Ieracitano, C., Mammone, N., Hussain, A. & Morabito, F. C. A novel multimodal machine learning based approach for automatic classifification of EEG recordings in dementia. Neural Netw. 123, 176 – 190 (2020).

[37] Prashanth, R., Roy, S. D., Mandal, P. K. & Ghosh, S. High-accuracy detection of early Parkinson’ s disease through multimodal features and machine learning. Int. J. Med. Inform. 90, 13–21 (2016).
[38] Hyun, S. H., Ahn, M. S., Koh, Y. W. & Lee, S. J. A machine-learning approach using PET-based radiomics to predict the histological subtypes of lung cancer. Clin. Nucl. Med. 44, 956 – 960 (2019).
[39] Yala, A., Lehman, C., Schuster, T., Portnoi, T. & Barzilay, R. A deep learning mammography-based model for improved breast cancer risk prediction. Radiology 292, 60–66 (2019).
[40] Reda, I. et al. Deep learning role in early diagnosis of prostate cancer. Technol. Cancer Res. Treat. 17, 1533034618775530 (2018).
[41] An, G. et al. Comparison of machine-learning classifification models for glaucoma management. J. Healthcare Eng. 2018, 2 – 7 (2018).
[42] Patel, M. J. et al. Machine learning approaches for integrating clinical and imaging features in late - life depression classifification and response prediction. Int. J. Geriatr. Psychiatry 30, 1056–1067 (2015).
[43] Huang, S.-C., Pareek, A., Zamanian, R., Banerjee, I. & Lungren, M. P. Multimodal fusion with deep neural networks for leveraging CT imaging and electronic health record: a case-study in pulmonary embolism detection. Sci. Rep. 10, 1 – 9 (2020).
[44] Tiulpin, A. et al. Multimodal machine learning-based knee osteoarthritis progression prediction from plain radiographs and clinical data. Sci. Rep. 9, 1–11 (2019).
[45] Wu, J. et al. Radiological tumour classifification across imaging modality and histology. Nat. Mach. Intell. 3, 787–798 (2021).
[46] Mei, X. et al. Artifificial intelligence - enabled rapid diagnosis of patients with COVID-19. Nat. Med. 26, 1224–1228 (2020).
[47] 27. Bardak, B. & Tan, M. Improving clinical outcome predictions using convolution over medical entities with multimodal learning. Artif. Intell. Med. 117, 102112 (2021).
[48] Jin, M. et al. Improving hospital mortality prediction with medical named entities and multimodal learning. arXiv preprint arXiv:1811.12276 (2018).
[49] Rajkomar, A. et al. Scalable and accurate deep learning with electronic health records. NPJ Digital Med. 1, 1 – 10 (2018).
[50] Yiwen, M., William, S., Michael, K. O. & Corey, W. A. Bidirectional representation learning from transformers using multimodal electronic health record data to predict depression. IEEE J. Biomed. Health Inform. 25, 3121–3129 (2021).
[51] Jordan, Y., William, Y. & Philipp, T. Multimodal skin lesion classification using deep learning. Exp. Dermatol. 27, 1261–1267 (2018).
[52] Larry, H. et al. Multimodal tensor-based method for integrative and continuous patient monitoring during postoperative cardiac care. Artif. Intell. Med. 113, 102032 (2021).
[53] Jae Hyun, Y., Johanna Inhyang, K., Bung Nyun, K. & Bumseok, J. Exploring characteristic features of attention-deficit/hyperactivity disorder: Findings from multi-modal MRI and candidate genetic data. Brain Imaging Behav. 14, 2132–2147 (2020).
[54] Janani, V., Li, T., Hamid Reza, H. & May, D. W. Multimodal deep learning models for early detection of Alzheimer’s disease stage. Sci. Rep. 11, 3254 (2021).
[55] Zeng, Wenhuan, Anupam Gautam, and Daniel H. Huson. "On the application of advanced machine learning methods to analyze enhanced, multimodal data from persons infected with COVID-19." Computation 9.1 (2021): 4.
[56] Kline, A., Wang, H., Li, Y., Dennis, S., Hutch, M., Xu, Z., Wang, F., Cheng, F. and Luo, Y., 2022. Multimodal machine learning in precision health: A scoping review. npj Digital Medicine, 5(1), p.171.
[57] Zhou, Hong-Yu, et al. "A transformer-based representation-learning model with unified processing of multimodal input for clinical diagnostics." Nature Biomedical Engineering (2023): 1-13.
[58] Aliper, Alex, et al. "Prediction of clinical trials outcomes based on target choice and clinical trial design with multi - modal artificial intelligence." Clinical Pharmacology & Therapeutics (2023).
[59] Li, Y. et al. Inferring multimodal latent topics from electronic health records. Nat. Commun. 11, 1–17 (2020).
[60] Pranav Rajpurkar, Jingbo Yang, Nathan Dass, Vinjai Vale, Arielle S Keller, Jeremy Irvin, Zachary Taylor, Sanjay Basu, Andrew Ng, and Leanne M Williams. Evaluation of a machine learning model based on pretreatment symptoms and electroencephalographic features to predict outcomes of an tidepressant treatment in adults with depression: a prespeci- fified secondary analysis of a randomized clinical trial. JAMAnetwork open, 3(6):e206653–e206653, 2020.
[61] Zhen-Yu Hong, Jooyong Shim, Woo Chan Son, and Changha Hwang. Predicting successes and failures of clinical trials with an ensemble ls-svr. medRxiv, 2020.
[62] Johann de Jong, Ioana Cutcutache, Matthew Page, Sami Elmoufti, Cynthia Dilley, Holger Fröhlich, and Martin Armstrong. Towards realizing the vision of precision medicine: Ai based prediction of clinical drug response. Brain, 144(6): 1738–1750, 2021.
[63] Zifeng Wang and Jimeng Sun. Survtrace: Transformers for survival analysis with competing events. arXiv preprint arXiv:2110.00855, 2021.
[64] Ashish Vaswani, Noam Shazeer, Niki Parmar, JakobUszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. Advances in Neural Information Processing Systems, 30, 2017.
[65] Zifeng Wang and Jimeng Sun. Transtab: Learning transferable tabular transformers across tables. arXiv preprint arXiv:2205.09328, 2022.
[66] Kaitlyn M Gayvert, Neel S Madhukar, and Olivier Elemento. A data-driven approach to predicting successes andfailures of clinical trials. Cell chemical biology, 23(10):1294 – 1301, 2016.
[67] Lo A W, Siah K W, Wong C H. Machine learning with statistical imputation for predicting drug approvals[M]. SSRN, 2019.
[68] Murali V, Muralidhar Y P, Königs C, et al. Predicting clinical trial outcomes using drug bioactivities through graph database integration and machine learning[J]. Chemical Biology & Drug Design, 2022, 100(2): 169-184.
[69] Zhao Fan, Fanyu Xu, Cai Li, and Lili Yao. 2020. Application of KPCA and AdaBoost algorithm in classification of functional magnetic resonance imaging of Alzheimer’s disease. Neural Computing and Applications 32, 10 (2020), 5329–5338.
[70] Léon-Charles Tranchevent, Francisco Azuaje, and Jagath C Rajapakse. 2019. A deep neural network approach to predicting clinical outcomes of neuroblastoma patients. BMC medical genomics 12, 8 (2019), 1–11.
[71] Youran Qi and Qi Tang. Predicting phase 3 clinical trial results by modeling phase 2 clinical trial subject level data using deep learning. volume 106 of Proceedings of Machine Learning Research, pages 288 – 303, 2019.
[72] Gilmer, Justin, et al. "Neural message passing for quantum chemistry." International conference on machine learning. PMLR, 2017.
[73] Liu, Yi, et al. "Spherical message passing for 3d graph networks." arXiv preprint arXiv:2102.05013 (2021).
[74] Satorras, Vıctor Garcia, Emiel Hoogeboom, and Max Welling. "E (n) equivariant graph neural networks." International conference on machine learning. PMLR, 2021.
[75] Tianfan Fu, Kexin Huang, Cao Xiao, Lucas M Glass, and Jimeng Sun. Hint: Hierarchical interaction network for clinicaltrial-outcome predictions. Patterns, 3(4):100445, 2022.
[76] Gentil, M.-L. et al. Factors influencing the development of primary care data collection projects from electronic health records: a systematic review of the literature. BMC Med. Inform. Decis. Mak. 17, 139 (2017).
[77] Wang, Zifeng, Cao Xiao, and Jimeng Sun. "SPOT: sequential predictive modeling of clinical trial outcome with meta-learning." Proceedings of the 14th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics. 2023.
[78] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, “High-Resolution Image Synthesis with Latent Diffusion Models,” Apr. 2022. arXiv:2112.10752 [cs].
[79] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen, “Hierarchical Text-Conditional Image Generation with CLIP Latents,” Apr. 2022. arXiv:2204.06125 [cs].
[80] “Chatgpt: Optimizing language models for dialogue,” Nov. 2022.
[81] Goodfellow, I. et al. Generative adversarial nets. Advances in neural information processing systems 27 (2014).
[82] J. Ho, A. Jain, and P. Abbeel, “Denoising diffusion probabilistic models,” Advances in Neural Information Processing Systems, vol. 33, pp. 6840 – 6851, 2020.
[83] Cui, L. et al. Conan: Complementary pattern augmentation for rare disease detection. In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, 614 – 621 (2020).
[84] Torfi, A. & Fox, E. A. Corgan: Correlation-capturing convolutional generative adversarial networks for generating synthetic healthcare records. In The Thirty Third International Flairs Conference (2020).
[85] Li, J., Cairns, B. J., Li, J. & Zhu, T. Generating synthetic mixed-type longitudinal electronic health records for artificial intelligent applications. arXiv preprint arXiv:2112.12047 (2021).
[86] Zhang, Z., Yan, C., Mesa, D. A., Sun, J. & Malin, B. A. Ensuring electronic medical record simulation through better training, modeling, and evaluation. Journal of the American Medical Informatics Association 27, 99 – 108 (2020).
[87] Yan, C., Zhang, Z., Nyemba, S. & Malin, B. A. Generating electronic health records with multiple data types and constraints. In AMIA Annual Symposium Proceedings, vol. 2020, 1335 (American Medical Informatics Association, 2020).

[88] Choi, E. et al. Generating multi-label discrete patient records using generative adversarial networks. In Machine learning for healthcare conference, 286 – 305 (PMLR, 2017).
[89] Baowaly, M. K., Lin, C.-C., Liu, C.-L. & Chen, K.-T. Synthesizing electronic health records using improved generative adversarial networks. Journal of the American Medical Informatics Association 26, 228 – 241 (2019).
[90] Zhang, Z., Yan, C., Lasko, T. A., Sun, J. & Malin, B. A. Synteg: a framework for temporal structured electronic health data simulation. Journal of the American Medical Informatics Association 28, 596 – 604 (2021).
[91] Theodorou, B., Xiao, C. and Sun, J., 2023. Synthesize Extremely High-dimensional Longitudinal Electronic Health Records via Hierarchical Autoregressive Language Model. arXiv preprint arXiv:2304.02169.
[92] Yuan H, Zhou S, Yu S. EHRDiff: Exploring Realistic EHR Synthesis with Diffusion Models[J]. arXiv preprint arXiv:2303.05656, 2023.
[93] Pan, S. J. & Yang, Q. A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 22, 1345 – 1359 (2009).
[94] Hao Wang, Hao He, and Dina Katabi. Continuously indexed domain adaptation. In International Conference on Machine Learning, 2020.
[95] Haoran Zhang, Natalie Dullerud, Laleh Seyyed-Kalantari, Quaid Morris, Shalmali Joshi, and Marzyeh Ghassemi. An empirical framework for domain generalization in clinical settings. In Proceedings of the Conference on Health, Inference, and Learning, pp. 279 – 290, 2021.
[96] Jifan Gao, Philip L Mar, and Guanhua Chen. More generalizable models for sepsis detection under covariate shift. AMIA Summits on Translational Science Proceedings, 2021:220, 2021.
[97] Jenna Marie Reps, Ross D Williams, Martijn J Schuemie, Patrick B Ryan, and Peter R Rijnbeek. Learning patient-level prediction models across multiple healthcare databases: evaluation of ensembles for increasing model transportability. BMC medical informatics and decision making, 22 (1):1 – 14, 2022.
[98] Hyeonseob Nam, HyunJae Lee, Jongchan Park, Wonjun Yoon, and Donggeun Yoo. Reducing domain gap by reducing style bias. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8690 – 8699, 2021.
[99] Juwon Kang, Sohyun Lee, Namyup Kim, and Suha Kwak. Style neophile: Constantly seeking novel styles for domain generalization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.
[100] Riccardo Volpi, Hongseok Namkoong, Ozan Sener, John C Duchi, Vittorio Murino, and Silvio Savarese. Generalizing to unseen domains via adversarial data augmentation. NeurIPS, 31, 2018.
[101] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy Hospedales. Learning to generalize: Meta learning for domain generalization. In Proceedings of the AAAI conference on artifificial intelligence, 2018a.
[102] Yogesh Balaji, Swami Sankaranarayanan, and Rama Chellappa. Metareg: Towards domain generalization using meta-regularization. Advances in neural information processing systems, 2018.
[103] Da Li, Jianshu Zhang, Yongxin Yang, Cong Liu, Yi-Zhe Song, and Timothy M Hospedales. Episodic training for domain generalization. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1446 – 1455, 2019.
[104] Rui Shao, Xiangyuan Lan, Jiawei Li, and Pong C Yuen. Multi-adversarial discriminative deep domain generalization for face presentation attack detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10023 – 10031, 2019.
[105] Krikamol Muandet, David Balduzzi, and Bernhard Scholkopf. Domain generalization via invariant feature representation. In International Conference on Machine Learning. PMLR, 2013.
[106] Muhammad Ghifary, David Balduzzi, W Bastiaan Kleijn, and Mengjie Zhang. Scatter component analysis: A unifified framework for domain adaptation and domain generalization. IEEE transactions on pattern analysis and machine intelligence, 39(7):1414 – 1430, 2016.
[107] Mingmin Zhao, Shichao Yue, Dina Katabi, Tommi S Jaakkola, and Matt T Bianchi. Learning sleep stages from radio signals: A conditional adversarial architecture. In International Conference on Machine Learning, pp. 4100 – 4109. PMLR, 2017.
[108] Woong-Gi Chang, Tackgeun You, Seonguk Seo, Suha Kwak, and Bohyung Han. Domain-specifific batch normalization for unsupervised domain adaptation. In Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition, pp. 7354 – 7362, 2019.
[109] Kaiyang Zhou, Yongxin Yang, Yu Qiao, and Tao Xiang. Domain generalization with mixstyle. In International Conference on Learning Representations (ICLR), 2021.
[110] Nourhan Bayasi, Ghassan Hamarneh, and Rafeef Garbi. Boosternet: Improving domain generalization of deep neural nets using culpability-ranked features. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 538 – 548, June 2022.
[111] Yanghao Li, Naiyan Wang, Jianping Shi, Jiaying Liu, and Xiaodi Hou. Revisiting batch normalization for practical domain adaptation. In International Conference on Learning Representations (ICLR), 2017.
[112] H. Song, D. Rajan, J. J. Thiagarajan, and A. Spanias, "Attend and diagnose: Clinical time series analysis using attention models," in Proc. 32nd AAAI Conf. Artif. Intell., 2018, pp. 4091 – 4098.
[113] E. Choi, M. T. Bahadori, A. Schuetz, W. F. Stewart, and J. Sun, "Doctor AI: Predicting clinical events via recurrent neural networks," in Proc. Mach. Learn. Healthcare Conf., 2015, pp. 301 – 318.
[114] F. Ma, R. Chitta, J. Zhou, Q. You, T. Sun, and J. Gao, "Dipole: Diagnosis prediction in healthcare via attention-based bidirectional recurrent neural networks," inProc. 23rd ACM SIGKDDInt. Conf. Knowl. Discovery Data Mining, 2017, pp. 1903 – 1911.
[115] Fenglong Ma, Radha Chitta, Jing Zhou, Quanzeng You, Tong Sun, and Jing Gao. Dipole: Diagnosis prediction in healthcare via attention-based bidirectional recurrent neural networks. In Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, pages 1903 – 1911, 2017
[116] Chen RJ, Lu MY, Williamson DFK, et al. Pan-Cancer Integrative Histology-Genomic Analysis via interpretable multimodal deep learning 2021;1–46 ArXiv:2108.02278v1. https://arxiv.org/abs/2108.02278 (1 December 2021, date last accessed).
[117] Kang M, Lee S, Lee D, et al. Learning cell-type-specific gene regulation mechanisms by multi-attention based deep learning with regulatory latent space. Front Genet 2020;11:869.
[118] Zadeh A, Vij P, Liang PP, et al. Multi-attention recurrent network for human communication comprehension. In: 32nd AAAI Conference on Artificial Intelligence, Vol. 2018. AAAI, 2018, 5642–9.
[119] Chen RJ, Lu MY, Wang J, Williamson DFK, Rodig SJ, Lindeman NI, et al. Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. 2020. ArXiv:1912.08937. https://arxiv.org/ abs/1912.08937 (1 December 2021, date last accessed).
[120] Selvaraju RR, Cogswell M, Das A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization. Int J Comput Vis 2020;128(2):336–59.