WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

AI-Driven Approaches for Drug Repurposing: Accelerating Drug Discovery with Machine Learning and Deep Learning

Atharva A. Digambar[1], Jasshan S. Bhalgat[2], Aditya N. Adak[3], Vaibhav M. Agade[4]

[1]Department of Artificial Intelligence and Data Science AISSMS IOIT College, Pune, India Email: digambaratharva@gmail.com
[2]Department of Artificial Intelligence and Data Science AISSMS IOIT College, Pune, India Email: jasshanbhalgat24@email.com
[3]Department of Artificial Intelligence and Data Science AISSMS IOIT College, Pune, India Email: adityaadak07@email.com
[4]Department of Artificial Intelligence and Data Science AISSMS IOIT College, Pune, India Email: vagade441801@email.com
DOI : https://doi.org/10.55248/gengpi.6.0425.16144

ABSTRACT—

Traditional medicate revelation is time-consuming, costly, and frequently unsuccessful, taking over a decade and billions of dollars with tall disappointment rates in clinical trials. To address these challenges, sedate repurposing—finding unused employments for approved drugs—offers a quicker, more cost- effective elective by leveraging existing security information and bypassing early improvement stages. Artificial Insights (AI) has changed medicate repurposing by empowering the investigation of tremendous biomedical datasets to identify promising drug- disease intuitive. Machine Learning (ML), deep Learning (DL), Chart Neural Systems (GNNs), and Natural Dialect Handling (NLP) are key instruments in this domain. These strategies have effectively anticipated modern medicate applications, even amid basic times like the COVID-19 widespread. However, challenges stay, counting information heterogeneity, need of labelled data, restricted interpretability, exploratory approval needs, and regulatory obstacles. This paper surveys AI-driven sedate repurposing methods—covering ML, DL, GNNs, and NLP—alongside real-world applications and case considers. We too highlight current confinements and propose future bearings to improve model reliability, straightforwardness, and clinical pertinence. By investigating the synergy between AI and medicate repurposing, this paper points to guide analysts and partners toward more shrewdly and efficient medicate discovery.

Index Terms—Drug Repurposing, Manufactured Insights, Ma- chine Learning, Profound Learning, Chart Neural Systems, NLP, Drug Discovery.

I. INTRODUCTION

The conventional medicate revelation pipeline is a long, expensive, and resource-intensive prepare, characterized by tall attrition rates and broad timelines. A single modern drug typically takes 10 to 15 years to develop and com- mercialize, costing over $2.6 billion, and has a less than 10% success rate from preclinical to final approval. These challenges are exacerbated in confront of rising illnesses, advancing pathogens, and complex incessant conditions, where convenient restorative arrangements are critical. To address these confinements, the logical community has progressively turned its consideration to sedate repurposing, also known as medicate repositioning—the handle of distinguishing new therapeutic employments for existing drugs that are as of now approved or in late-stage clinical trials. Repurposing offers a compelling alternative by altogether diminishing the taken a toll, time, and risk associated with sedate improvement, as repurposed drugs have well-established pharmacokinetics, security profiles, and fabricating pathways. Established results like the use of minoxidil for hair growth, sildenafil for erectile dysfunction and aspiratory hypertension, and more recently, the repurposing of Remdesivir and dexamethasone for COVID-19, show the strategy's practical promise. However, conventional strategies for medicate repurposing are to a great extent driven by hypothesis-based clinical bits of knowledge, fortunate findings, or difficult exploratory screening, which limit their flexibility and thoroughness when exploring the vast natural and chemical universe. Artificial Intelligence (AI) has transformed the biomedical field in recent years by providing powerful tools to accelerate and broaden drug repurposing ef- forts. AI envelops a wide extend of computational techniques, counting machine learning (ML), profound learning (DL), fortification learning, common dialect processing (NLP), and graph-based learning, all of which empower systems to learn from complex and high-dimensional information. Through the coordination of diverse information sources, including genetic data, transcriptome profiles, chemical structures, protein intel- ligence, disease phenotypes, and clinical trial outcomes, these methods can uncover hidden patterns and connections that are difficult to discern through traditional analysis. Machine learning calculations such as Arbitrary Woodland, Support Vector Machines (SVM), and XGBoost have been utilized to predict drug-disease affiliations based on designed features, while profound learning structures like Convolutional Neu- ral Networks (CNNs), Repetitive Neural Systems (RNNs), and Transformer-based models have appeared guarantee in learning intricate representations from unstructured biomedical content and molecular charts. Chart Neural Systems (GNNs) in particular have risen as capable apparatuses for demonstrating organic systems, permitting the investigation of drug-target- disease relationships in all-encompassing and versatile way. Furthermore, natural dialect preparing procedures can extri- cate knowledge from millions of investigate articles, clinical reports, and patents, automating the curation of drug-related experiences and generating repurposing speculations from writing mining. These advancements collectively empower a worldview move from manual, trial and-error repurposing to data-driven, precise exploration of restorative openings. In spite of

the developing adoption of AI in this field, a few challenges stay. Biomedical data is frequently boisterous, imbalanced, and divided over diverse designs and sources. Numerous AI models, particularly deep learning-based ones, work ”black boxes,” making it difficult to decipher their fore- casts and construct believe among clinicians and administrative bodies. The shortage of high-quality labelled information for uncommon maladies, moral concerns surrounding patient information utilization, and require for exploratory or clinical validation of AI-generated theories advance complicate the landscape. Besides, interpreting AI yields into actionable and clinically significant bits of knowledge requires intrigue col- laboration among information researchers, scholars, chemists, and healthcare experts. As such, whereas AI presents trans- formative openings for sedate repurposing, its full potential can as it were be realized through cautious methodological design, robust approval, and dependable arrangement. The purpose of this paper is to provide a thorough and funda- mental overview of the state of AI-driven drug repurposing. It explores the foundational concepts, conventional strategies, and how AI augments these endeavours through progressed learning paradigms. Through case thinks about, methodologi- cal comparisons, and future perspectives, the paper looks for to direct progressing inquire about and foster development at the crossing point of counterfeit intelligence and sedate disclosure.

II. DRUG REPURPOSING – CONCEPT AND TRADITIONAL METHODS

Drug repurposing is an advancing procedure in pharmaceuti- cal research pointed at recognizing unused helpful applications for existing drugs. This strategy stands out as a quicker and more cost-effective elective to conventional medicate discovery, which regularly includes a long time of experi- mentation and clinical trials. Repurposing is especially im- portant when time-sensitive health challenges emerge or when there’s require to address neglected or uncommon maladies. Conventional strategies, including experimental screening and early computational apparatuses, have contributed to eminent repurposing victory stories. However, these approaches to have confinements, making space for the integration of more progressed innovations like artificial intelligence.

A. Definition and Importance of Drug Repurposing

Drug repurposing, too named sedate repositioning, is an evolving and profoundly impactful methodology in the field of biomedical and pharmaceutical sciences that includes recog- nizing new clinical employments for existing drugs that were at first developed to treat distinctive restorative conditions. Not at all like the traditional drug disclosure has pipeline, which is regularly time-consuming, costly, and loaded with tall dis- appointment rates, medicated repurposing offers a essentially quicker, cost-effective, and lower-risk alternative. Ordinary sedate improvement ordinarily takes 10 to 15 a long time, with a normal taken a toll surpassing $2 billion, and faces various obstacles counting broad preclinical testing, multi- phase clinical trials, and administrative approvals. By differen- tiate, repurposing capitalizes on already generated data related to pharmacokinetics, pharmacodynamics, safety, toxicity, and fabricating, empowering analysts to bypass the prior stages of improvement and move straightforwardly into advanced stages of clinical trials, such as Stage II or III. This can radically abbreviate the time required to bring a therapeutic arrangement to patients, which is particularly critical during wellbeing crises or in cases of uncommon illnesses where im- mediate arrangements are missing. The fundamental principle of medicate repurposing is polypharmacology—the concept that most drugs associated with different natural targets rather than a single one. This suggests that medicate planned for one malady may moreover impact pathways included in other diseases, subsequently making it a candidate for reposition- ing. This broadens the scope of helpful applications and maximizes the utility of existing pharmaceutical compounds. Furthermore, the significance of medicate repurposing has developed in response to a few present day challenges, count- ing the expanding global burden of unremitting sicknesses such as cancer, diabetes, cardiovascular and neurodegenerative maladies; the disturbing rise in antimicrobial resistance; and the tirelessness of neglected tropical and uncommon hered- itary maladies that need compelling treatment options. In numerous of these cases, pharmaceutical companies may need the monetary motivations to contribute in modern medicate advancement due to littler advertise sizes or dubious returns, making repurposing an appealing and doable approach. One of the most compelling exhibits of the adequacy of drug repurposing was seen amid the COVID-19 pandemic, where the pressing require for helpful arrangements driven to the rapid assessment of existing drugs. Antivirals like Remdesivir, originally created for Ebola, and anti-inflammatory agents like Dexamethasone, already utilized for other inflammatory con- ditions, were rapidly repurposed to treat COVID-19 patients. These cases emphasize how sedate repurposing can respond to prompt open wellbeing emergencies, giving timely and life-saving mediations when ordinary medicate advancement cannot keep pace with pressing request. In addition, drug repurposing adjusts well with developing innovations such as artificial insights (AI), machine learning, and computational science, which are progressively being utilized to mine large datasets, analyse atomic intelligent, and predict drug-disease affiliations. By identifying the most promising candidates for underutilized indications, these tools improve the accuracy and efficiency of repurposing efforts.

B. Traditional Computational Approaches

Traditional computational strategies in medicate repurpos- ing have historically served as profitable instruments in recog- nizing potential new applications for existing drugs by lever- aging natural, chemical, and clinical data through algorithmic analysis. These strategies point to diminish the time, taken a toll, and labor related with research facility testing by narrow- ing down candidate compounds for advance examination. One of the foundational approaches is atomic docking, which re- enacts the interaction between medicate (ligand) and its target protein (receptor). It predicts how well a sedate can tie to a protein’s dynamic location based on shape complementarity and binding vitality. This approach is instrumental in virtual screening, but its adequacy depends intensely on the accessi- bility of high-resolution auxiliary information and may come up short when the protein structure is obscure or adaptable. Another customary method is ligand-based closeness looking, which expect that fundamentally comparable compounds are likely to have comparative organic impacts. Devices such as Quantitative Structure-Activity Relationship (QSAR) models drop beneath this category. QSAR models utilize factual relationships between chemical features and organic action, but they are restricted in their capacity to generalize past known compound classes. Network-based approaches such as protein-protein interaction (PPI) networks and gene-disease affiliation systems give a systems level understanding of natural forms. By representing proteins, qualities, drugs, and infections as hubs and their intuitive as edges, analysts can recognize

repurposing candidates based on their topological properties in these systems. For instance, drugs that associated with proteins closely connected to disease-associated hubs might be compelling therapeutic agents. Quality expression signature coordinating, especially using resources like the Network Outline (CMap) and LINCS L1000 dataset, includes comparing the transcriptional response of a malady with the quality expression profile initiated by drugs. If a sedate can switch the illness signature (i.e., downregulate upregulated qualities and bad habit versa), it is hypothesized to have therapeutic potential. Whereas effective, this approach requires large, high-quality transcriptomic datasets and may overlook post-transcriptional or metabolic impacts. Content mining and semantic investigation of biomedical writing have too been used to distinguish backhanded or already neglected affiliations between drugs and maladies. These tools use natural lan- guage processing to extract structured data from unstructured texts. Be that as it may, ordinary content mining tools often battle with semantic subtleties, equivalent words, and context sensitivity, which can influence precision. In spite of the guarantee of these conventional computational strategies, they are frequently siloed, meaning that each strategy employments a limit cut of biological or chemical information. This need of integrator capacity restricts the encompassing understanding vital for precisely predicting complex drug-disease intelligent. Additionally, conventional strategies are regularly hypothesis- driven or, rather than data-driven, which can introduce bias and miss out on novel, data-emergent patterns. Their static nature also makes them less adaptive to new biomedical discoveries or large-scale data integration — a gap that modern AI-based approaches aim to fill.

C. Limitations of Traditional Drug Repurposing

Although early-stage sedate repurposing experiences have been made possible by traditional computational method- ologies, they exhibit a few key limitations that make them insufficient to address contemporary biomedical concerns. Essential limitation is the inadequate and divided nature of natural data. Traditional approaches frequently depend on curated databases, many of which are not overhauled in real-time or need coverage for certain illnesses, populaces, or atomic targets. For instance, when target protein high- resolution 3D structures are unavailable or receptor flexibility is ignored, the effectiveness of atomic docking is significantly compromised. This limits the utility of such strategies to well characterized proteins and infections, subsequently bar- ring many unmet therapeutic needs. Besides, these strategies typically analyse a single measurement of information, such as chemical similarity, gene expression, or protein interac- tion. Natural systems, however, are intrinsically multi-layered and energetic, involving intricate crosstalk among qualities, proteins, metabolites, and natural components. The failure of conventional approaches to integrate differing datasets — counting genomics, proteomics, metabolomics, understanding wellbeing records, and clinical results — leads to halfway and regularly wrong forecasts. Another major issue is the rigidity and constrained adaptability of classical models. Most of them are built with settled presumptions and rules, meaning they cannot adjust as modern information or knowledge emerges. This unbending engineering is unacceptable for repurposing in fast-evolving areas like oncology, irresistible illnesses, and neurological clutters, where sedate instruments and disease pathways are continually being re-imagined. The need of interpretability and unthinking knowledge moreover presents a boundary to clinical interpretation. For occurrence, network- based strategies may identify a candidate medicate based on vicinity in a protein interaction outline but come up short to clarify how the sedate influences the specific malady pathway. This need of natural clarity often necessitates broad follow- up approval, subsequently delaying the pipeline. In expansion, conventional strategies battle with rare or complex illnesses, where the test estimate is too small for measurable models, or where pathogenesis involves multifactorial intelligence. In such cases, existing knowledge bases are inadequately, and data-driven investigation becomes essential. Finally, numerous of these instruments require manual curation and space ability for operation and translation. This not only increments the workload and time-to-discovery but also introduces human inclination, diminishing the objectivity of findings. The computational taken a toll and time required to perform thor- ough likeness looks or docking recreations too scale poorly with expansive datasets, making them unreasonable for high throughput screening.

III. AI-BASED APPROACHES FOR DRUG REPURPOSING

Artificial Insights (AI) has revolutionized the landscape of medicate repurposing by empowering the integration and investigation of massive biomedical datasets. These datasets, which run from molecular structures and quality expressions to quiet records and logical writing, are frequently as well tremendous and complex for traditional examination strategies Artificial intelligence (AI) techniques, particularly machine learning, deep learning, graph-based models, and natural lan- guage processing, promote automated information discovery by identifying hidden patterns, connections, and forecasts that may provide unique evidence of contemporary restorative uses for currently available medications. AI-based methods decrease the time, fetched, and vulnerability involved in drug development pipelines by fusing computational efficiency with predictive control.

A. Machine Learning in Drug Repurposing

Machine Learning (ML), a centre subset of counterfeit insights, has revolutionized the pharmaceutical scene by en- abling the extraction of significant bits of knowledge from tremendous and complex biomedical information. In sedate repurposing, ML models facilitate the distinguishing proof of modern restorative employments for existing drugs by reveal- ing non-obvious designs, correlations, and inactive structures inside multi-modal datasets, including chemical structures, genomic profiles, malady phenotypes, and clinical results. This data-driven worldview marks a critical take off from con- ventional hypothesis-driven techniques, advertising a speedier, more cost-effective, and scalable solution to the sedate disclo- sure bottleneck.ML methods in drug repurposing are broadly categorized into directed, unsupervised, and fortification learn- ing strategies, each serving distinct parts depending on the nature of the information and the predictive errands included. In administered learning, algorithms such as Arbitrary Wood- lands, Back Vector Machines (SVM), and Angle Boosting are prepared on named datasets, where drug-disease affiliations are known. These models learn to predict potential helpful impacts of drugs on previously unassociated illnesses based on input highlights like molecular descriptors, organic targets, and qual- ity expression signatures. For case, a administered demonstrate prepared on pharmacological profiles can anticipate whether a medicate initially created for hypertension might moreover appear viability in treating Alzheimer's disease by examining likenesses in target quality interactions. Unsupervised learning, on the other hand, is particularly powerful in finding covered up designs without requiring labelled information. Methods such as k-means clustering, various

levelled clustering, and central component examination (PCA) are utilized to gather drugs and illnesses based on likenesses in their chemical or genomic highlights. For occurrence, clustering algorithms can recognize subsets of drugs that display similar effects over different cell lines, indicating toward potential repurposing openings. Lattice factorization strategies like Non-negative Framework Factorization (NMF) are moreover utilized to re- veal idle drug-disease interaction spaces, allowing researchers to prioritize promising candidates for experimental validation. Support learning (RL), in spite of the fact that still emerging in this space, gives energetic learning system where agents learn ideal procedures for sedate choice or molecule modifi- cation through trial-and-error intelligent with the environment. RL models recreate successive decision-making processes in sedate disclosure pipelines, optimizing ways that lead to remedially successful candidates. These models have appeared guarantee in scenarios where the repurposing task can be surrounded as a multi-step optimization issue, such as select- ing the best combination of repurposed drugs for a specific infection phenotype. A noteworthy commitment of ML to drug repurposing lies in the development of prescient models that outline drugs to potential malady targets. These models integrate heterogeneous information sources—such as quality expression profiles (e.g., from the LINCS L1000 dataset), chemical structure databases (e.g., PubChem), and protein- protein interaction networks—to construct encompassing rep- resentations of sedate and disease mechanisms. Prescient ML models can survey whether a compound's transcriptomic sig- nature conversely connects with that of infection, showing potential helpful efficacy. For occurrence, the Network Outline (CMap) system uses ML calculations to compare disease- related quality expression signatures with drug-induced marks to anticipate unused signs. In expansion to essential drug- disease expectation, ML also assists in auxiliary examination, such as side impact prediction, toxicity profiling, and persis- tent stratification. These functions are basic in repurposing since the security profile of the drug must be reassessed in the setting of the unused indication. Ensemble learning approaches that combine the yields of multiple models have illustrated predominant execution in capturing such multi- faceted sedate properties. The integration of real-world prove (RWE) from electronic wellbeing records (EHRs), clinical trials, and biomedical writing advance enhances ML-driven repurposing methodologies. Naturalistic data, when handled through ML pipelines, can uncover off-label usage patterns and post-marketing security information that recommend viable repurposing ways. Procedures such as include determina- tion and dimensionality lessening are utilized to oversee the high dimensionality and commotion inborn in RWE datasets, ensuring robust and interpretable forecasts. In spite of its transformative potential, ML-based medicate repurposing is not without challenges. Show generalizability, interpretability, and data quality stay noteworthy obstacles. Black-box models often lack the straightforwardness required for administrative endorsement, and biases in preparing information can lead to deluding forecasts. To address these issues, later endeavours have centred on developing explainable AI (XAI) systems and consolidating domain knowledge into show models. Straight- forward ML models not as it were improve believe among clinicians and administrative bodies but moreover give bits of knowledge into the natural credibility of repurposed candi- dates.

B. Deep Learning Techniques

Deep learning, a specialized subfield of machine learning, has risen as a transformative approach in the domain of medicate repurposing due to its capacity to handle endless, un- structured, and high-dimensional biomedical information with negligible highlight designing. It works through manufactured neural systems composed of different layers that mimic the functioning of the human brain, empowering the extraction of hierarchical highlights and complex designs. This capa- bility is particularly profitable in sedate repurposing, where information from diverse sources such as chemical struc- tures, genomics, transcriptomics, proteomics, and biomedical writing require to be integrated to reveal novel drug-disease connections. One of the key qualities of profound learning lies in its flexibility, offering various designs custom fitted for particular sorts of biomedical data and prescient errands. Convolutional Neural Networks (CNNs), initially planned for picture handling, have found applications in dissecting atomic pictures and 3D protein ligand interaction structures. In med- icate repurposing, CNNs can be prepared to translate spatial designs in atomic charts or docking recreations, subsequently foreseeing how well medicate might bind to a target protein related to a distinctive infection. For example, CNNs have been connected to extricate highlights from 2D or 3D representa- tions of chemical compounds, enabling automated expectation of pharmacological properties and target interactions. Repet- itive Neural Systems (RNNs), particularly Long Short-Term Memory (LSTM) systems, are capable at taking care of con- secutive information, making them appropriate for examining time series quality expression profiles and biomedical content. RNNs can be utilized to show the movement of illness states over time in reaction to sedate intercessions, encouraging the identification of drugs that may return malady directions. In addition, RNNs have been utilized to handle Grins (Simplified Molecular Input Line Passage Framework) strings, which are textual representations of chemical compounds, empowering the era and classification of atomic structures for repurposing applications. The approach of Transformer-based models has significantly progressed profound learning in the biomedical domain. Unlike RNNs, Transformers can prepare whole se- quences in parallel and capture long-range conditions more viably. Models like BERT (Bidirectional Encoder Representa- tions from Transformers) and its biomedical variations, such as BioBERT and PubMedBERT, have been instrumental in mining drug-disease connections from large-scale literature. These models are pre-trained on broad biomedical corpora and fine-tuned on particular assignments, such as named substance acknowledgment (NER), connection extraction, and address replying. In the setting of sedate repurposing, they can filter millions of scientific distributions to extricate inactive joins between existing drugs and elective infection signs based on contextual cues and semantic similitudes. Another essential profound learning architecture utilized in sedate repurposing is the autoencoder, an unsupervised neural organize outlined to learn compressed representations (idle vectors) of infor- mation. Autoencoders consist of an encoder that maps input information to a lower-dimensional space and a decoder that remakes the unique input from this compressed representa- tion. In medicate repurposing, autoencoders are utilized to decrease the dimensionality of complex omics datasets (e.g., quality expression or proteomic information) while preserving fundamental organic signals. This dimensionality reduction empowers the comparison of malady and medicate profiles in a common inactive space, where closeness measurements can suggest potential repurposing candidates. By introducing probabilistic information into the inactive space, variation autoencoders (VAEs), an extension of standard autoencoders, enable the creation of underutilized atomic structures with desired useful features, potentially leading to the discovery of new applications for well-known medications. Profound learning too plays a significant part in multi-modal integration, where heterogeneous information sorts such as chemical, natural, clinical, and printed information are combined to make encompassing models of sedate activity. Multi-modal neural networks can learn from distinctive information sources simultaneously, enhancing the vigour and prescient control of repurposing models. For occasion, a profound neural organize might take as input a drug's chemical structure, its quality

expression signature in infected cells, and clinical side-effect profiles to predict potential unused signs. This integrator approach is especially advantageous in overcoming the con- finements of individual data modalities, such as lost values or estimation clamour. The application of Generative Adversarial Networks (GANs) is another frontier in deep learning-driven drug repurposing. The generator and discriminator neural networks, which make up a GAN, compete with one another in a zero-sum game to get better at what they do. In repurposing, GANs can be trained to generate novel drug-like molecules or simulate disease states, which can then be matched with existing drugs to identify therapeutic candidates. For example, a GAN might generate synthetic gene expression profiles of a disease state, which are then screened against a database of drug-induced profiles to find compounds that reverse the synthetic disease signature.

C. Graph-Based Models for Drug-Target Interaction Predic- tion

Graph-based models have picked up noteworthy footing in the field of medicate repurposing, particularly for drug-target interaction (DTI) forecast. The center thought behind these models is to represent organic and chemical entities—such as drugs, proteins, qualities, and diseases—as hubs in a chart, and the relationships or intelligent between them as edges. The rich topological data and pertinent circumstances embedded in intricate biomedical systems can be abused by analysts thanks to this structure. By demonstrating substances as interconnected systems, graph-based approaches can reveal inactive affiliations, empowering the recognizable proof of unused helpful uses for existing drugs. One of the founda- tional components in this domain is the Information Chart (KG). A information chart is organized representation that encodes connections among diverse biomedical substances utilizing a triplet format—typically subject, predicate, and protest (e.g., drug–interacts with–protein or gene–associated with–disease). These charts can integrate data from different heterogeneous sources, counting Drug Bank, ChEMBL, Pub- Chem, UniProt, DisGeNET, and writing mining databases. The quality of KGs lies in their ability to capture both coordinate and backhanded associations between substances, permitting analysts to perform thinking and link forecast for sedate repurposing. Conventional knowledge graph-based approaches frequently utilize measurable or path-based algorithms to recognize potential intuitive. In any case, these techniques have impediments when managing with large-scale or sparsely associated systems. This has driven to the adoption of pro- gressed machine learning models, especially Graph Neural Systems (GNNs), which are well-suited for learning over graph-structured information. GNNs amplify the capabilities of traditional neural systems by consolidating message passing mechanisms, where each hub totals data from its neighbours to upgrade its representation iteratively. This allows demonstrate to learn complex conditions and patterns that reflect both nearby and worldwide chart structures. In the context of DTI forecast, GNNs can be utilized to model both drugs and target proteins as hubs in a bipartite graph. Chemical structure similitudes, atomic fingerprints, and side impact profiles can characterize edges between sedate hubs, while protein-protein intuitive, grouping homology, or functional annotations can interface protein hubs. The drug-target edges represent known intuitive, which the GNN show uses during preparing to learn highlight representations. Once trained, demonstrate can antic- ipate lost or novel edges—i.e., previously unknown drug-target interactions—thus recognizing candidates for repurposing. A few GNN variations have been developed to suit particular repurposing errands. Chart Convolutional Systems (GCNs) apply convolution operations on chart information to propagate data over neighbours. GCNs have been employed to learn embeddings for drugs and proteins by combining topological data with hub highlights, such as chemical properties or amino corrosive arrangements. Another popular variation is the Chart Consideration Arrange (GAT), which introduces consideration components to weigh the significance of neighbouring hubs in an unexpected way amid message passing. This helps the demonstrate center on more important intelligent and improves forecast precision. An excellent application of GNNs in medicate repurposing is the Decagon demonstrate, which predicts polypharmacy side impacts by demonstrating drugs an proteins as a multimodal chart and utilizing GCNs to induce side effect sorts between sedate sets. In spite of the fact that fundamentally designed for unfavourable impact forecast, such models can be adapted for distinguishing po- tential drug-disease joins or discovering off-target intelligent demonstrative of repurposing opportunities. Another impactful approach is heterogeneous chart modelling, which consolidates different sorts of hubs and edges into a single chart. In such models, drugs, infections, qualities, pathways, and clinical phenotypes coexist, permitting for a more comprehensive investigation. In order to collect new affiliations, meta-path- based techniques can extract complicated social groupings (e.g., sedate → quality → sickness). These ways are at that point utilized to train graph-based models that can prioritize sedate candidates for repurposing based on their network designs and semantic similarity to known helpful ways. In expansion to predictive modelling, graph-based approaches empower visual exploration and theory era. Instruments like Cytoscape, Neo4j, and NetworkX permit analysts to visualize interaction networks, identify profoundly associated center points, and investigate communities or modules inside the chart. This makes a difference in defining natural theories around potential medicate components and understanding the broader organic setting of repurposing candidates. Real-world applications of graph-based models have illustrated their es- teem in pressing sedate repurposing scenarios. Amid the COVID-19 widespread, a few studies leveraged graph-based methods to recognize drugs that could interact with SARS- CoV-2 targets or tweak have response pathways. For case, information charts coordination viral host protein intuitive, drug-target information, and clinical trial information were analysed utilizing GNNs to prioritize antiviral agents and safe modulators for clinical assessment. This rapid, data- driven prioritization was instrumental in quickening restorative inquire about amid a worldwide wellbeing emergency.

D. Natural Language Processing (NLP) for Biomedical Liter- ature Analysis

Natural Dialect Preparing (NLP), a department of artificial intelligence centered on the interaction between computers and human dialect, has risen as a transformative apparatus in drug repurposing, especially through the examination of biomedical literature. The biomedical space creates endless sums of unstructured content information daily—from log- ical diaries, clinical trial reports, and sedate databases to electronic wellbeing records (EHRs) and licenses. Extricating important experiences from this literary information physically is overwhelming and time-consuming task. NLP empowers mechanized and brilliantly preparing of this data to distinguish covered up designs, affiliations, and knowledge that can fuel sedate repurposing procedures. At the core of NLP-based sedate repurposing lies the capacity to extract drug-disease, drug-target, and disease-gene connections from large corpora. Conventional databases, in spite of the fact that wealthy in curated knowledge, cannot keep up with the quick pace of scientific publication. NLP bridges this crevice by efficiently scanning the writing, translating biomedical settings, and syn- thesizing modern theories for potential medicate applications. Techniques such as named substance acknowledgment (NER), connection extraction, text classification, and data recovery are broadly used

to reveal affiliations that might something else stay buried in text. In biomedical NLP, Named Substance Ac- knowledgment (NER) is fundamental. It includes distinguishing and categorizing key biomedical substances such as sedate names, diseases, proteins, qualities, and pathways. Not at all like nonexclusive NER tasks, must biomedical NER handle complex phrasings, truncations, equivalent words, and multi- word expressions. Progressed NER systems utilize domain- specific ontologies like Work, UMLS, and SNOMED CT to progress exactness and consistency. Once substances are rec- ognized, connection extraction techniques determine how they are connected—for occasion, identifying whether medicate hinders a particular protein or is effective against infection. In later a long time, the rise of transformer-based language models has essentially boosted the performance of NLP appli- cations in the biomedical field. These models, built on designs such as BERT (Bidirectional Encoder Representations from Transformers), have set modern benchmarks in understanding relevant data in content. However, generic BERT models prepared on general-domain information like Wikipedia or Book Corpus are not optimized for biomedical terminology. To address this, analysts created domain specific models like BioBERT, SciBERT, and PubMedBERT, which are pre trained on biomedical writing from PubMed abstracts and full-text articles. BioBERT, for case, extends BERT's capabilities by preparing on over 18 billion words from biomedical corpora. It exceeds expectations in assignments like biomedical question answering, connection extraction, and record classification. PubMedBERT, on the other hand, is prepared completely from scratch utilizing as it were PubMed writings, empowering it to create a deeper understanding of biomedical seman- tics. These models have illustrated exceptional execution in recognizing drug- disease joins, foreseeing off-label medicate employments, and summarizing writing for repurposing bits of knowledge. Another innovative application of NLP in medicate repurposing is literature-based discovery (LBD). This strategy leverages co-occurrence designs and certain connections in writing to recommend novel hypotheses. For occasion, if medicate A is known to influence quality B, and quality B is related with infection C in diverse articles, LBD frameworks can propose that sedate A may be viable for malady C. This "ABC model" of induction has driven to a few high profile repurposing victories and proceeds to be a powerful paradigm. Besides, semantic look and question-answering sys- tems fuelled by NLP permit analysts to inquiry vast biomed- ical storehouses with common dialect. These systems return relevant and evidence-backed comes about, sparing time and enhancing the accuracy of writing audit forms. For example, inquiry like "Which drugs have been appeared to decrease irritation in COVID-19 patients?" can return targeted insights from thousands of papers, empowering quick identification of candidates for advance examination. In expansion to drug target disclosure, NLP helps in dissecting clinical trial out- comes, extracting unfavourable sedate responses (ADRs) from understanding reports, and understanding real- world prove (RWE) from EHRs. Mining ADRs is especially critical for repurposing, as off target impacts might propose modern signs. NLP techniques can parse account clinical notes and social media information to identify such unintended benefits or side impacts, advertising clue for repurposing pathways. In spite of its guarantee, biomedical NLP faces a few challenges. The uncertainty and variability of biomedical dialect can prevent precise substance recognition. Terms frequently have different implications (e.g., "ACE" can mean angiotensin-converting protein or unfavourable cardiac occasion), and abbreviations are habitually reused. Moreover, the need of labelled preparing information in particular subdomains can constrain the exe- cution of administered models. Biomedical text explanation takes a lot of work and demands spatial skills. To overcome these boundaries, analysts are turning to unsupervised and semi-supervised learning methods, leveraging huge unlabelled corpora to prepare vigorous models. Furthermore, knowledge graphs and ontologies are progressively coordinates with NLP systems to improve understanding and disambiguation. For instance, substance connecting frameworks utilize ontologies like UMLS to outline extricated terms to canonical identifiers, improving consistency and interpretability. An developing slant is the integration of NLP with other AI techniques, such as machine learning and chart analytics. For occurrence, yields from NLP systems—like extricated drug-target relation- ships—can be used to improve information charts, which are at that point dissected using Graph Neural Systems (GNNs). This multimodal approach leverages printed information and organized information to provide more comprehensive sedate repurposing experiences.

E. Reinforcement Learning for Drug Discovery

Reinforcement Learning (RL) has emerged as a powerful approach in the domain of drug discovery and repurposing, complementing traditional methods and other AI techniques like machine learning, deep learning, and natural language processing. As a type of machine learning inspired by behav- ioral psychology, reinforcement learning focuses on training agents to make sequential decisions by interacting with an environment, learning from rewards or penalties received based on their actions. In the context of drug discovery, this translates into learning strategies that can navigate the vast chemical space, optimize molecular structures, and identify new therapeutic opportunities efficiently and accurately. A its core, reinforcement learning operates through an agent environment interaction loop. The agent observes the state of the environment, takes an action, and receives a reward or punishment in return. The learning process is guided by this incentive, which enables the agent to progressively modify its policy in order to optimize cumulative reward. In drug discovery, the "environment" can be the chemical space, where the agent modifies molecular structures; the "actions" may involve adding, deleting, or altering functional groups in a compound; and the "rewards" can be designed based on properties such as drug-likeness, binding affinity, toxicity, or novelty. De novo drug design is one of the most fascinating uses of RL in drug discovery and repurposing.In this scenario, reinforcement learning agents generate novel molecular struc- tures from scratch. These molecules are evaluated in silico for desired pharmacological properties using predictive models or simulations. The feedback received from these evaluations helps the agent improve its molecule-generating policy. This loop continues until a high-potential molecule is identified. Notably, reinforcement learning is highly suitable for this task because it can explore vast, high- dimensional chemical spaces more effectively than random sampling or exhaustive search. An example of this is the REINVENT framework, which utilizes recurrent neural networks (RNNs) combined with reinforcement learning to generate new molecules represented in SMILES (Simplified Molecular Input Line Entry System) format. The model is trained to produce valid chemical struc- tures and is guided by a reward function that incorporates multiple objectives such as synthetic accessibility, potency against a particular target solubility, and ADMET (absorp- tion, distribution, metabolism, excretion, and toxicity) profiles. By iteratively improving the policy through reinforcement learning, the system evolves towards optimal molecules that satisfy the desired criteria. Reinforcement learning also plays a crucial role in lead optimization, which involves improving the properties of existing molecules or drug candidates. For instance, after identifying a repurposable drug with moderate efficacy against a new disease, reinforcement learning can help design analogs with better activity, lower toxicity, or improved pharmacokinetics. The agent explores chemical modifications, simulates their outcomes, and learns which changes yield better results. This is especially useful when adapting drugs for populations with specific genetic

backgrounds or co- morbid conditions. Another impactful application of RL is in multi-objective optimization, where the agent must balance conflicting drug design goals—such as maximizing binding affinity while minimizing toxicity. In conventional approaches, this trade-off is often addressed through heuristic weighting or Pareto front analysis. Reinforcement learning, on the other hand, can dynamically learn these trade-offs during the train- ing process, adapting its strategy to optimize all objectives simultaneously. This is crucial in repurposing scenarios where a drug effective for one condition must be adjusted for a new use-case with different safety or efficacy constraints. Furthermore, RL is being applied in adaptive clinical trial design, a revolutionary approach in drug development. Tradi- tional clinical trials follow rigid protocols, but reinforcement learning enables dynamic decision-making during the trial process. By modeling the trial as a Markov Decision Process (MDP), RL algorithms can adjust parameters such as dosage levels, patient stratification, or treatment paths in real time, based on ongoing results. This adaptive strategy not only accelerates the discovery of effective repurposed drugs but also minimizes patient risk and trial costs. Reinforcement learning's usefulness is further increased by combining it with other AI techniques. For instance, RL combined with gener- ative adversarial networks (GANs) can produce structurally novel and biologically relevant compounds by maintaining a balance between exploration and exploitation. Moreover, reinforcement learning agents can utilize transfer learning to apply knowledge gained from one domain (e.g., oncology) to another (e.g., neurology), thus accelerating the repurposing process across therapeutic areas. The application of reinforce- ment learning to drug repurposing is difficult, despite its potential. The creation of efficient reward systems is one of the main challenges. If the reward signal is too sparse or not aligned with real-world drug discovery goals, the RL agent may converge on suboptimal or trivial solutions. Additionally, computational cost is a significant concern, as simulating molecular interactions or conducting virtual screenings at scale requires substantial resources. Efforts are ongoing to improve the efficiency of RL algorithms, such as through prioritized experience replay, model-based RL, and hierarchi- cal reinforcement learning. Real-world applications of RL in drug discovery are beginning to show promise. For example, In silico Medicine used RL and deep generative models to design a DDR1 kinase inhibitor, which progressed from design to preclinical validation in under 46 days—a process that traditionally takes years. This success demonstrated the speed, flexibility, and precision that RL-based systems can offer in real-world drug discovery pipelines.

IV. CHALLENGES IN AI-DRIVEN DRUG REPURPOSING

Despite the promising advancements, AI-driven drug repur- posing faces several critical challenges. One of the foremost issues is the heterogeneity and incompleteness of biomedical data, which can affect model accuracy and reliability. The lack of standardized datasets and variations across sources often make integration difficult. Additionally, many AI models act as" black boxes," making their predictions hard to interpret, especially for regulatory approval. The lack of high-quality labeled data also hampers the development of robust models. Lastly, experimental validation and ethical considerations re- main essential to ensure the safety, efficacy, and acceptability of AI-suggested drug candidates.

A. Data Heterogeneity

A key challenge in AI-driven drug repurposing is the diversity and inconsistency of biomedical data. Biomedical information originates from a variety of sources, including genomics, proteomics, transcriptomics, metabolomics, elec- tronic health records (EHRs), clinical trial outcomes, drug chemical structure databases, and scientific literature. Each of these data types is generated through different platforms and methodologies, often with their own unique terminologies, units of measurement, and data structures. This inherent diver- sity makes it exceptionally difficult to integrate and analyze the data in a cohesive manner suitable for machine learning (ML) or deep learning (DL) applications. For example, genomic datasets may include gene sequences or expression profiles measured using different sequencing technologies and plat- forms. Proteomics data might use mass spectrometry results formatted differently based on the software used. EHRs may contain structured data such as lab test results, but also unstructured data like doctor's notes or discharge summaries, often written in natural language with varying degrees of specificity and terminology. Clinical trial data can further com- plicate integration due to differences in patient demographics,trial designs, and outcome measures. The inconsistencies and incompatibilities across these datasets pose significant barriers to developing robust AI models. To enable effective data integration, a substantial amount of preprocessing is required. This involves normalizing values, standardizing terminologies through ontologies like SNOMED CT or MeSH, and imputing missing or incomplete data. Inadequate data harmonization can result in flawed or biased predictions, limiting the potential benefits of AI. Moreover, cross-platform variability and batch effects can introduce noise, making it harder for models to generalize. For instance, patient data collected from different hospitals might reflect institutional practices rather than true biological differences. Addressing these concerns requires so- phisticated data engineering pipelines and a deep understand- ing of both biological and computational nuances. Another dimension of data heterogeneity is the constantly evolving nature of biomedical knowledge. As scientific understanding evolves, new relationships between genes, proteins, and dis- eases are discovered, necessitating continuous updates to the underlying datasets. AI models must therefore be adaptable to incorporate newly generated data and insights. Ultimately, overcoming data heterogeneity demands collaborative efforts among biologists, data scientists, clinicians, and software engineers. Interdisciplinary approaches and the adoption of universal data standards are essential to create interoperable and scalable platforms that can support the next generation of AI applications in drug repurposing.

B. Model Interpretability

As AI and machine learning algorithms become increas- ingly complex, especially with the widespread adoption of deep learning, the interpretability and explainability of these models emerge as critical concerns—particularly in healthcare applications like drug repurposing. The concept of a" black box" model—one whose internal workings are opaque to users—presents a serious challenge in clinical environments where trust, transparency, and accountability are essential. In drug repurposing, AI models often analyze complex, high dimensional biomedical data to predict novel uses for existing drugs. While such models can achieve impressive accuracy, their decision-making processes are not always readily un- derstandable, even to domain experts. Clinicians, regulatory bodies, and patients alike require clear and justifiable explana- tions for how a model arrived at a particular

recommendation. Without this transparency, it becomes difficult to validate findings, build trust, or make informed clinical decisions. Inter- pretability is especially important when AI models influence patient treatment options. For instance, if a model suggests repurposing an anti-inflammatory drug for a rare neurode- generative disease, healthcare providers must understand the rationale behind the suggestion—whether it's based on shared gene expression profiles, similar protein interaction networks, or literature-based evidence. A variety of approaches have been developed to address these challenges. Model agnostic tools like SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) can provide post-hoc explanations by identifying which features most influenced a given prediction. In deep learning models, at- tention mechanisms and saliency maps can highlight the most relevant input features, such as specific genes or molecular pathways. However, these methods are not without limita- tions. They often provide only local explanations—i.e., they explain individual predictions rather than the model's overall behavior. Furthermore, explanations generated by such tools may still be difficult to interpret without substantial domain knowledge. To promote model transparency, researchers are exploring the use of inherently interpretable models, such as decision trees or rule-based systems, in combination with more powerful black-box models. Hybrid models work to achieve a balance between precision and explainability. Additionally, incorporating domain-specific knowledge and constraints into the model design can enhance interpretability while maintain- ing predictive performance. Explainability is also crucial for regulatory approval. Agencies like the FDA and EMA require detailed documentation of how AI models work, what data they use, and how decisions are validated. Transparent models are more likely to meet these stringent requirements and gain acceptance in clinical settings.

C. Limited Availability of High-Quality Labelled Data

One of the fundamental challenges in applying artificial intelligence (AI) to drug repurposing lies in the scarcity of high-quality labelled data. AI models, especially those based on supervised learning paradigms, require large amounts of annotated data to learn meaningful patterns. However, in the biomedical domain, obtaining such data is particularly difficult due to several reasons, including privacy regula- tions, the complexity of biological systems, and the lack of standardized labelling protocols. Most biomedical datasets are fragmented, stored in incompatible formats, or confined within proprietary databases, which limits their accessibility and usability for training robust AI models. Moreover, clinical data often contain inconsistencies, missing values, or are recorded in unstructured formats such as handwritten notes and free-text entries in electronic health records (EHRs). These issues make it difficult to preprocess and annotate the data accurately. Additionally, the process of labelling data—such as classifying the efficacy of a drug on a spe- cific disease—is a time consuming and resource-intensive task that typically requires expert domain knowledge from pharmacologists, clinicians, or biomedical researchers. This reliance on expert curation further compounds the data scarcity problem. The limited availability of labelled data also affects the generalizability of AI models. Models trained on small, biased, or incomplete datasets may not perform well when exposed to new or diverse clinical scenarios. For instance, a model trained on data from a particular population may not translate effectively to another demographic group due to genetic, environmental, or socio-economic differences. This issue becomes particularly critical in the context of global health, where drug repurposing efforts must consider diverse patient populations. To deal with these issues, researchers have turned to solutions like semi-supervised learning, transfer learning, and augmentation of data. Semi-supervised learning allows models to learn from a small amount of labelled data supplemented with a larger pool of unlabeled data, thereby im- proving performance. Transfer learning leverages knowledge from pretrained models on related tasks, which can then be finetuned for drug repurposing applications with limited data. Data augmentation methods, such as generating synthetic sam- ples using generative adversarial networks (GANs), have also shown promise in enriching training datasets. Collaborative efforts and open-access initiatives have emerged to address the data scarcity problem. Public repositories like DrugBank, ChEMBL, PubChem, and clinical trial databases provide valuable resources for AI-driven drug repurposing. These platforms aggregate biochemical, pharmacological, and clin- ical information, enabling researchers to access standardized datasets. In addition, initiatives like The Cancer Genome Atlas (TCGA) and Genotype-Tissue Expression (GTEx) provide large-scale genomic and transcriptomic data, which are crucial for discovering new drug-disease relationships. Despite these efforts, the need for high-quality, diverse, and well-annotated datasets remains a significant barrier. Moving forward, it is imperative to develop frameworks for secure data sharing, adopt unified data standards, and encourage interdisciplinary collaborations to improve data quality and availability. Ethical considerations, including patient privacy and data ownership, must also be addressed through robust governance models. Ultimately, overcoming the challenge of limited labelled data is essential for unlocking the full potential of AI in drug repurposing and accelerating the discovery of new therapeutic uses for existing drugs.

D. Need for Experimental Validation of AI Predictions

While AI has demonstrated great promise in predicting drug-disease associations and identifying potential candidates for drug repurposing, these computational insights must un- dergo rigorous experimental validation before clinical applica- tion. Without empirical confirmation, AI-generated predictions remain theoretical and cannot be translated into actionable medical treatments. This validation process is essential to confirm the safety, efficacy, and biological relevance of the proposed drug mechanisms. Experimental validation typically involves in vitro assays, in vivo animal models, and eventually, clinical trials. These stages help determine whether a predicted drug can effectively modulate the target pathway, alleviate disease symptoms, or produce the intended therapeutic out- comes. However, transitioning from computational prediction to laboratory testing introduces several challenges. Firstly, in vitro and in vivo studies are time-consuming, expensive, and labor-intensive. They also require access to specialized equipment and technical expertise, which may not always be available to computational research teams. Also, not all predicted drug-target interactions are biologically viable. Some may lack the necessary pharmacokinetics or pharmacodynam- ics to function in living organisms. Others may have off- target effects, toxicity, or unfavourable drug interactions that were not captured in the computational model. Therefore, experimental validation must be comprehensive and rigorous, encompassing a broad range of assays to assess both the efficacy and safety of repurposed drugs. Another significant bottleneck is the limited scalability of experimental methods. While AI can rapidly generate hundreds or thousands of drug repurposing hypotheses, experimental validation can only handle a small subset of these predictions due to resource constraints. This discrepancy necessitates the development of prioritization frameworks to identify the most promising candidates for further testing. Strategies such as multicri- teria decision analysis, network-based ranking, and expert curation can help filter and select high-value predictions for validation. Collaborative partnerships between computational scientists,

pharmacologists, and clinical researchers are critical to bridging the gap between AI predictions and experimental validation. Integrating experimental feedback into AI models can create a virtuous cycle where validated results are used to refine and improve prediction accuracy. Moreover, the estab- lishment of shared infrastructures—such as high-throughput screening facilities, automated testing platforms, and open ac- cess validation datasets—can streamline the validation process and reduce the time from prediction to proof-of-concept. Ulti- mately, experimental validation is not merely a procedural step but a cornerstone of responsible AI-driven drug repurposing. It ensures that computational insights translate into tangible health benefits and helps build trust among stakeholders, in- cluding regulatory agencies, healthcare providers, and patients. Future progress in this area will depend on interdisciplinary collaboration, innovation in experimental design, and sustained investment in translational research infrastructures.

E. Regulatory Barriers and Ethical Concerns

As Artificial Intelligence (AI) continues to revolutionize drug repurposing, it also presents significant regulatory and ethical challenges. These challenges must be addressed sys- tematically to ensure that AI-driven methodologies not only accelerate drug discovery but also maintain safety, trans- parency, fairness, and accountability in clinical and pharma- ceutical domains. The integration of AI into drug repurposing pipelines brings into focus issues such as data privacy, model accountability, clinical applicability, legal liability, and the need for evolving regulatory frameworks. Regulatory agencies such as the Food and Drug Administration (FDA), the Euro- pean Medicines Agency (EMA), and other national authorities are increasingly facing pressure to update their guidelines for assessing and regulating AI-driven tools and platforms. One of the most pressing challenges is the lack of clear regulatory frameworks that specifically cater to AI applications in drug repurposing. Traditional regulatory pathways are often based on deterministic models where mechanisms and decision processes can be clearly outlined. In contrast, AI models, particularly those utilizing deep learning and other complex algorithms, often operate as "black boxes." Their internal decision-making logic is not readily interpretable, making it difficult for regulators to understand, validate, and trust the rationale behind the AI generated drug repurposing predictions. This opacity can lead to significant hesitation in approving AI- generated outputs for clinical use, particularly when patient safety is on the line. As a result, drug candidates identified by AI may still need to undergo traditional preclinical and clinical testing, potentially limiting the speed and cost-effectiveness benefits AI promises. In addition to structural regulatory limi- tations, the standardization and validation of AI models remain another critical area of concern. Regulatory agencies currently lack universally accepted benchmarks and validation protocols for evaluating the safety and efficacy of AI systems used in drug repurposing. This makes it challenging to compare the performance of different AI tools or establish confidence in their outputs. There is a growing need for the establishment of internationally recognized standards for data collection, algorithm training, validation, and reporting that can support regulatory review processes. Without such standards, AI-based tools may face inconsistencies in approval across countries, limiting their global scalability and utility. Data governance and patient privacy also raise ethical concerns. AI models for drug repurposing often require vast amounts of biomed- ical data—including genomic data, electronic health records (EHRs), clinical trial datasets, and literature text—which can be sensitive and personally identifiable. The collection, storage, and processing of such data must adhere to strict privacy regulations such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States or the General Data Protection Regulation (GDPR) in the European Union. Breaches of these regulations, intentional or otherwise, can lead to legal consequences and loss of public trust. Additionally, issues around data ownership and consent need to be clarified. Patients may not be fully informed that their anonymized data could be used in AI algorithms to discover new drug uses, raising questions about informed consent and transparency. Another ethical dimension lies in bias and equity in AI-driven drug repurposing. The performance of AI models depends entirely on the quality of the data they are trained with. If training datasets are skewed toward certain popula- tions—such as individuals from developed countries, specific ethnicities, or particular age groups—then the AI system's predictions may not be generalizable or fair across diverse global populations. This could lead to the recommendation of repurposed drugs that are less effective or even harmful to underrepresented populations, thereby exacerbating existing healthcare disparities. Ethical AI development necessitates the use of diverse, representative datasets and robust fairness auditing to ensure equitable treatment outcomes. Furthermore, the accountability of AI systems remains a gray area. If an AI system recommends a repurposed drug that results in adverse outcomes for patients, the question arises: who is legally liable? Is it the AI developer, the healthcare professional who used the AI tool, the pharmaceutical company, or the regulatory body that approved it? These legal ambiguities must be addressed proactively to encourage the safe and responsible deployment of AI in drug repurposing. Legal frameworks need to evolve in parallel with technological advances to ensure clear delineation of responsibilities and protection for all stakeholders.

V. CASE STUDIES AND APPLICATIONS

The integration of artificial intelligence into drug repur- posing has moved beyond theoretical frameworks to achieve significant real-world impact, as evidenced by multiple suc- cessful case studies and industry applications. Notable ex- amples include the identification of Baricitinib for COVID-19 by Benevolent AI, the revival of Thalidomide for treat- ing multiple myeloma, and the emerging use of Metformin in cancer therapy—each demonstrating the power of AI to uncover new therapeutic potentials in existing drugs. Leading pharmaceutical companies and AI firms such as Benevolent AI, Insilico Medicine, and DeepMind are revolutionizing drug discovery through AI-driven platforms capable of mining complex biomedical data, predicting drug-target interactions, and designing novel molecules. Industry trends are also shifting towards AI-powered high-throughput screening and drug repurposing for rare and neglected diseases, opening new avenues in personalized and equitable healthcare. These advancements illustrate how AI is not only transforming tradi- tional pharmaceutical practices but also offering scalable, cost- effective, and life-saving solutions across the global healthcare landscape.

A. Real-World Success Stories

The field of AI-driven drug repurposing is rapidly evolving, and a number of real-world success stories underscore its transformative potential. These instances not only validate the role of AI in drug discovery but also serve as blueprints for future applications. They demonstrate how the intersection of

technology, pharmacology, and clinical insights can yield life-saving therapies more efficiently and economically than traditional approaches. One of the most high-profile examples is the identification of Baricitinib for treating COVID-19. Initially approved for rheumatoid arthritis, Baricitinib was not on the radar for antiviral applications until BenevolentAI employed its advanced AI-driven platform to analyse potential drugs that could inhibit the SARS-CoV-2 virus. The AI sys- tem processed extensive data, including biomedical literature, gene expression data, and known drug-target relationships, to hypothesize that Baricitinib could block the virus from entering human cells by inhibiting the AP2-associated protein kinase 1 (AAK1). Moreover, its anti-inflammatory effects made it a dual-action candidate—capable of both preventing viral entry and reducing the cytokine storm associated with severe COVID-19 cases. The drug was quickly moved into clinical testing and, after positive outcomes, received Emer- gency Use Authorization by the FDA. This rapid repurposing, accomplished in a matter of weeks, would have taken years through conventional methods. Another poignant success story involves Thalidomide, a drug that once represented one of the darkest chapters in pharmaceutical history. Originally marketed as a sedative in the 1950s, Thalidomide was pulled from the market once its teratogenic effects were discovered, resulting in thousands of birth defects. Decades later, re- searchers using computational modelling and AI-based molec- ular docking discovered that Thalidomide had strong anti- angiogenic properties. This revelation led to its re-evaluation as a treatment for multiple myeloma, a type of blood cancer characterized by abnormal blood vessel formation within bone marrow.AI played a key role in optimizing the therapeutic window and minimizing risks by modeling safe dosage levels and pinpointing patient subgroups most likely to benefit from the treatment. Today, Thalidomide and its derivatives are FDA-approved for specific cancers and leprosy, turning a once-vilified compound into a valuable therapeutic option. Metformin, a widely prescribed antidiabetic drug, also stands out as a repurposing success. Long-term observational studies had noted lower cancer rates among diabetic patients on Metformin. AI algorithms helped validate these observations by mining vast datasets from electronic health records, ge- nomics, and transcriptomics. Machine learning models sug- gested that Metformin's activation of AMP-activated protein kinase (AMPK) and inhibition of the mTOR pathway could suppress tumor growth. Clinical trials are now underway to explore its efficacy in treating various cancers, including breast and pancreatic cancer. Metformin's case illustrates how AI can identify new mechanisms of action and unlock unexpected therapeutic uses of old drugs. These cases underscore the growing credibility of AI-powered platforms in discovering repurposing opportunities that would otherwise remain hidden. The use of AI not only cuts down research time but also significantly reduces RD costs. Importantly, it allows the healthcare system to respond swiftly during global health emergencies, as seen with COVID-19. As AI capabilities expand, the frequency and success rate of such repurposing breakthroughs are expected to rise, transforming the landscape of modern medicine.

B. AI in the Pharma Industry

The pharmaceutical industry has experienced a significant transformation with the incorporation of AI throughout the drug discovery process. From target identification to lead optimization and clinical trials, AI is revolutionizing how pharmaceutical companies operate. With its ability to process and interpret vast quantities of structured and unstructured data, AI helps reduce the time, cost, and risk associated with traditional RD processes. Several pioneering companies—such as BenevolentAI, Insilico Medicine, and DeepMind—are at the forefront of this transformation, setting new standards in drug repurposing and discovery. BenevolentAI is arguably one of the most prominent examples. The company utilizes a biomedical knowledge graph powered by natural language processing (NLP) and machine learning to connect billions of relationships among genes, proteins, drugs, and diseases. The knowledge graph is continuously updated with infor- mation from scientific literature, clinical trials, patents, and biochemical data. BenevolentAI's system hypothesized that Baricitinib could be repurposed for COVID-19, and the subse- quent clinical validation cemented its place as a leader in AI-powered drug discovery. Beyond COVID-19, the company is exploring neurological and autoimmune disorders, using AI to streamline and personalize the drug discovery process. Insilico Medicine has developed an end-to-end AI platform combining PandaOmics for target identification and Chemistry42 for drug generation. By integrating omics data, phenotypic screening results, and text mining, Insilico's systems identify novel targets and design compounds that meet ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) criteria. Notably, they published the discovery of a novel molecule for idiopathic pulmonary fibrosis that progressed from idea to preclinical candidate in just 46 days—far quicker than the multi-year timelines of conventional methods. Insilico also partners with major pharmaceutical firms, illustrating the commercial viability and collaborative potential of AI in the industry. DeepMind, through its AlphaFold project, has redefined structural biology. The AI system predicts protein structures with near-experimental accuracy, solving a 50-year grand challenge in biology. Accurate protein structures are critical for understanding disease mechanisms and identifying druggable sites. Although DeepMind doesn't directly repur- pose drugs, AlphaFold's contributions have laid the ground- work for researchers to perform AI assisted docking studies, design inhibitors, and repurpose drugs with greater confidence. Researchers now use AlphaFold models to discover how existing compounds interact with new targets, accelerating the repurposing process. Furthermore, companies such as Recur- sion Pharmaceuticals and Exscientia are using high-throughput phenotypic screening, generative chemistry, and AI-powered trial design to enhance RD. These firms represent a new breed of "techbio" companies where data science and biology coexist symbiotically. Pharmaceutical giants like Novartis, Pfizer, and AstraZeneca have also embraced AI, forming partnerships with tech firms to embed machine learning into their pipelines. As more pharmaceutical companies adopt AI, the competitive landscape is shifting. The convergence of biological data, computing power, and machine learning expertise is giving rise to more targeted, efficient, and patient- centric drug development. AI has transitioned from being a support tool to becoming a core enabler of innovation in pharma, especially for drug repurposing initiatives that require agile, data-intensive approaches.

C. Industry and Research Trends

The current wave of AI integration in drug repurposing is being shaped by several noteworthy trends, each reinforcing the technology's role as a cornerstone of future pharmaceutical RD. These trends span from innovations in high-throughput screening to broader ethical considerations and regulatory adaptations that ensure the responsible use of AI in medicine. Adoption of AI-powered high-throughput screening (HTS) is one of the most notable trends. Hundreds of compounds must be physically tested against biological targets as part of the time-consuming and costly conventional HTS procedure. AI now allows researchers to simulate and evaluate compound- target interactions in silico, significantly reducing costs and time lines. Machine learning models trained on large datasets can predict binding affinities, off-target effects, and toxicity with high accuracy. For example, deep learning

techniques such as graph convolutional networks and transformer models are being used to process molecular graphs and predict phar- macological activity, enabling more precise prioritization of re purposing candidates. Another significant trend is the growing interest in rare and neglected diseases, where traditional phar- maceutical investment is lacking due to low expected returns. AI can help by identifying repurposable drugs from existing pharma companies as that may have therapeutic effects in these areas. Databases like Orphanet and the NIH's GARD (Genetic and Rare Diseases Information Center) provide foundational data, which AI algorithms analyse for pattern recognition and mechanistic insights. Through transfer learning and few-shot learning, AI models are able to generate hypotheses even from limited data, offering hope to patient populations previously overlooked by mainstream drug development. Collaborative ecosystems are also gaining momentum. Initiatives like The Open Targets Platform, Drug Repurposing Hub, and the LINCS Project facilitate data sharing across academia, in- dustry, and regulatory bodies. These platforms integrate multi omics data, phenotypic screening results, and literature mining to create comprehensive datasets. Researchers use these re- sources to train and validate AI models, democratizing access to tools and knowledge previously limited to large pharma companies. Open science and precompetitive collaborations are expected to play a vital role in maximizing the reach and impact of AI in drug repurposing. Regulatory innovation is beginning to catch up with technological advances. Agencies like the FDA and EMA are actively exploring guidelines for the validation and approval of AI-generated drug candidates. Programs to evaluate AI in clinical decision-making and medi- cation development have been started by the FDA's Center for medication Evaluation and Research (CDER). These efforts include sandbox environments, regulatory science pilots, and adaptive trial designs to accommodate AI-generated evidence. The emergence of such frameworks is crucial for instilling trust and enabling smoother translation of AI discoveries to the clinic. Ethical and social dimensions are becoming integral to AI research. The concept of explainable AI (XAI) is increas- ingly emphasized to ensure that decisions made by AI systems are transparent and interpretable by human experts. Especially in healthcare, where lives are at stake, black-box models pose risks. Ensuring fairness, accountability, and equitable access to AI-driven therapies is vital, particularly when repurposing drugs for underrepresented populations. All of these tenden- cies point to a strong, forward-thinking environment where artificial intelligence (AI) serves as a fundamental pillar of pharmaceutical innovation rather than merely a tool. A new age of drug discovery is being ushered in by the convergence of science, technology, ethics, and regulation. This new era will be quicker, more intelligent, and more sensitive to the demands of healthcare around the world.

VI. FUTURE DIRECTIONS AND SOLUTIONS

As the intersection of artificial intelligence (AI) and drug repurposing continues to evolve, the future promises ground breaking innovations and transformative impacts on pharma- ceutical research. However, several challenges remain, includ- ing the need for improved model performance, enhanced trans- parency, experimental validation, and data privacy. Targeted approaches that not only improve upon current techniques but also introduce cutting-edge frameworks and technology are needed to address these issues. This section explores the key future directions and proposed solutions to overcome current limitations in AI-driven drug repurposing.

A. Enhancing Model Generalizability and Performance

A critical aspect of advancing AI-driven drug repurpos- ing lies in improving model generalizability—the ability of AI systems to perform accurately across diverse, real-world datasets. Currently, many AI models are trained on narrow or biased datasets, which limits their applicability to broader clinical scenarios. The first step in improving generalizability is to collect a diverse range of data types, such as clinical, proteomic, metabolomics, transcriptomic, genomic, and other data from different populations. This diversity ensures that models are exposed to multiple biological and environmental contexts, allowing them to detect complex patterns that might otherwise be overlooked. Integrating multi-omics data not only enhances predictive power but also supports the development of robust models capable of identifying novel drug-disease associations. Using current data for new, unexplored treatment areas can be facilitated by transfer learning, a technique that applies information from one activity to another. For example, a model trained on oncology data can be adapted to study neurodegenerative diseases, provided the underlying biological features share some overlap. Moreover, ensemble learning, which combines multiple models to produce more accurate results, can mitigate individual model biases and reduce overfitting. Techniques such as cross-validation and bootstrapping further improve reliability by rigorously testing model performance on unseen data. Graph-based deep learn- ing approaches like graph neural networks (GNNs) offer a promising direction for modelling molecular interactions and complex biological networks. These models capture relational dependencies between entities, such as protein-protein or drug target interactions, which are essential for repurposing efforts. Continuous model retraining is also essential. As new data becomes available from clinical trials, literature, or real-world evidence (RWE), models must be updated to incorporate this knowledge. This dynamic learning approach ensures models remain relevant and accurate in an ever-changing biomed- ical landscape. Investments in large-scale, annotated, and inter operable datasets—coupled with collaboration between academia, industry, and regulatory bodies—will be crucial in driving forward the generalizability and performance of AI in drug repurposing.

B. Developing AI Explainability Techniques

The transparency of most AI systems, especially deep learn- ing models, resulting in worries over their interpretability and transparency. Within the field of drug repurposing, explainabil- ity is not merely desirable—it is necessary to obtain trust from regulators, clinicians, and researchers. Explainable AI (XAI) seeks to provide insights into how models arrive at specific predictions, enabling human users to understand, validate, and act upon these recommendations. Several methods have been created to render AI predictions more understandable. One of the most commonly employed is SHAP (Shapley Additive explanations), that gives importance values to all the inputs feature, identifying the role of each gene, protein, or molecular descriptor to the model's output. Similarly, LIME (Local Interpretable Model-agnostic Explanations) ap- proximates the behaviour of complex models with simpler interpretable ones in the vicinity of a specific prediction. Another promising technique involves attention mechanisms, particularly in transformer-based models. These mechanisms allow models to "focus" on the most relevant features of a dataset, such as specific binding sites on a protein or structural motifs in a compound. Visualizing these attention weights can offer biological insights and suggest potential mechanisms of action. Counterfactual

explanations—what would happen if a certain input feature were changed—can also aid in hypothesis generation. For example, if removing a methyl group from a compound significantly reduces predicted efficacy, researchers might explore derivatives with similar modifications. From a regulatory perspective, explainability supports compliance with guidelines that require documentation of how clinical decisions are made. This is especially important in drug repurposing, where off-label use can have legal and ethi- cal implications. By fostering a deeper understanding of AI decisions, XAI techniques empower researchers to critically assess predictions, reduce bias, and improve model refinement. Ultimately, closing the gap between computer forecasts and practical biological applications will depend on the use of explainability frameworks.

C.Bridging AI Predictions with Experimental Validation

While AI has shown tremendous promise in identifying repurposing candidates, its true value is realized only when predictions are experimentally validated. Bridging computa- tional insights with laboratory and clinical testing ensures that AI models contribute meaningfully to drug development pipelines. The first step in this integration is high-throughput screening (HTS), where AI-selected drug candidates are tested against various biological assays. AI can prioritize compounds with the highest predicted efficacy, reducing the experimental burden and cost. In addition, advanced screening platforms such as organoids and organ-on-chip technologies provide more physiologically relevant models for testing drug efficacy and toxicity. Collaborative frameworks between computational scientists and experimental researchers are essential. Initiatives like the NIH NCATS program and the Open Targets platform exemplify successful models where AI predictions are rapidly transitioned into experimental pipelines. Clinical validation remains a significant hurdle due to time, cost, and regulatory constraints. However, adaptive clinical trial designs—where treatment arms are modified based on interim results—can expedite the validation of repurposed drugs. AI can also assist in patient stratification, identifying subgroups that are more likely to respond to a given therapy. Real-world evidence (RWE) from electronic health records (EHRs), patient reg- istries, and pharmacovigilance databases offers another avenue for validation. AI models can analyse treatment outcomes and adverse events in large patient cohorts, providing evidence for the safety and efficacy of repurposed drugs. Furthermore, AI-driven retrosynthetic analysis and virtual screening tools support medicinal chemistry efforts to optimize repurposed compounds. These techniques help in improving bio avail ability, reducing toxicity, and enhancing target specificity. By establishing a feedback loop between AI predictions and experimental data, the field can evolve towards a more iterative and integrative approach to drug repurposing. This synergy will accelerate the translation of computational discoveries into clinically approved therapies.

D. Privacy-Preserving AI and Federated Learning

In the era of big data, privacy concerns are paramount, particularly when handling sensitive biomedical and patient information. AI models trained on such data must adhere to regulations like the Health Insurance Portability and Ac- countability Act (HIPAA) in the US, the General Data Pro- tection Regulation (GDPR) in Europe, and other frameworks worldwide. This poses a problem for cooperative research and information exchange. Federated learning (FL) offers a transformative solution. Rather than compiling information in a single location, FL allows AI models to be trained locally on distributed datasets. Only model parameters are transmitted to a central server; raw data is never exported from the company. This preserves privacy while enabling collaborative learning across hospitals, research centres, and pharmaceutical companies. For example, multiple hospitals across different countries could train a common model to predict drug efficacy in COVID-19 patients without violating local privacy laws. FL enables the model to benefit from the combined knowledge of all institutions while main training data sovereignty. By adding statistical noise to shared data, privacy-preserving strategies like differential privacy provide an extra degree of protection by making it nearly hard to reverse-engineer specific patient information. Secure multi- party computing (SMPC) and homomorphic encryption fur- ther safeguard data while it is being computed and sent. In addition to protecting privacy, FL improves the general- izability and robustness of the model. Training on diverse datasets from multiple sources reduces bias and improves performance across varied populations and clinical settings. However, implementing FL comes with technical challenges, including communication overhead, system heterogeneity, and the need for synchronization across different nodes. Research is ongoing to develop more efficient algorithms, compression techniques, and decentralized architectures to address these issues. To fully realize the potential of privacy-preserving AI, clear governance structures, standardized protocols, and ethical guidelines are essential. Collaborative efforts such as the Open Mined community and the Personal Genome Project are paving the way for secure, transparent, and scalable implementations. As federated learning matures, it is poised to become a cornerstone of AI-driven drug repurposing, enabling large scale collaboration without compromising individual privacy or institutional integrity.

VII. CONCLUSION

The integration of Artificial Intelligence (AI) into the realm of drug repurposing has revolutionized the traditional paradigms of drug discovery, offering a faster, cost-effective, and more efficient pathway to identify new therapeutic uses for existing drugs. Throughout this research, we have explored how AI-based methodologies—including machine learning, deep learning, natural language processing, graph neural net works, and reinforcement learning—have emerged as powerful tools to overcome the bottlenecks of conventional repurposing techniques. These advanced approaches enable precise pre diction of drug-target interactions, automate the extraction of biomedical insights from unstructured literature, and uncover hidden relationships across complex biological networks. A key insight from our investigation is the significance of leveraging large-scale, multi-modal datasets—ranging from electronic health records using proteomics and genomics and biomedical publications—for training and validating AI mod- els. These datasets empower AI systems to generate predictive outputs with higher accuracy and relevance, particularly when combined with robust computational frameworks. Addition ally, the real-world case studies, such as the AI-assisted repur- posing of Baricitinib for COVID-19 and Metformin's evolving role in cancer therapy, exemplify the practical utility and life saving potential of these innovations. However, the research has also brought to light several pressing challenges, including data heterogeneity, lack of interpretability in black-box AI models, limited availability of annotated biomedical data, the critical need for experimental validation, and regulatory as well as ethical concerns. These obstacles pose significant barriers to the widespread clinical adoption of AI in drug repurposing, but they also offer avenues for future innovation. Strategies such as enhancing model transparency through Explainable AI

(XAI), deploying privacy-preserving federated learning techniques, and fostering cross-disciplinary collaboration between AI ex- perts and biomedical researchers can help mitigate these chal- lenges. Looking ahead, the future of AI in drug repurposing appears exceedingly promising. With ongoing advancements in computational power, algorithmic sophistication, and data integration capabilities, AI is poised to become a central pillar in the global pharmaceutical landscape. As the technology matures, its applications will likely extend beyond common diseases to target rare, neglected, and emerging conditions that have historically been underserved. AI's potential to personalize treatment strategies based on patient-specific bio logical profiles also hints at a paradigm shift toward precision medicine. To realize this vision, further research should focus on the development of standardized benchmarks for model evaluation, increased accessibility to high-quality biomedi- cal datasets, and comprehensive regulatory frameworks that support ethical innovation. Moreover, fostering collaborations between government agencies, business, and academics will be essential to translate AI breakthroughs from theoretical models into real-world clinical solutions.

VIII.REFERENCES

[1] T. T. Ashburn and K. B. Thor, "Drug repositioning: identifying and developing new uses for existing drugs," Nat. Rev. Drug Discov., vol. 3, no. 8, pp. 673–683, 2004.

[2] S. Pushpakom et al., "Drug repurposing: progress, challenges and recommendations," Nat. Rev. Drug Discov., vol. 18, no. 1, pp. 41–58, 2019.

[3] G. Jin and S. T. C. Wong, "Toward better drug repositioning: prioritizing and integrating existing methods into efficient pipelines," Drug Discov. Today, vol. 19, no. 5, pp. 637–644, 2014.

[4] M. Sirota et al., "Discovery and preclinical validation of drug indications using compendia of public gene expression data," Sci. Transl. Med., vol. 3, no. 96, pp. 96ra77, 2011.

[5] Y.-Y. Ke et al., "Artificial intelligence approach fighting COVID-19 with repurposing drugs," Biomed. J., vol. 43, no. 4, pp. 355–362, 2020.

[6] P. Richardson et al., "Baricitinib as potential treatment for 2019-nCoV acute respiratory disease," The Lancet, vol. 395, no. 10223, p. e30, 2020.

[7] J. M. Stokes et al., "A deep learning approach to antibiotic discovery,"Cell, vol. 180, no. 4, pp. 688–702.e13, 2020.

[8] W. Zhang, Y. Chen, D. Li, and J. Xu, "Artificial intelligence in drug design," Front. Pharmacol., vol. 11, p. 848, 2020.

[9] M. Zitnik, M. Agrawal, and J. Leskovec, "Modeling polypharmacy side effects with graph convolutional networks," Bioinformatics, vol. 34, no. 13, pp. i457–i466, 2018.

[10] Q. Wang et al., "Graph neural networks for drug repurposing: a survey," Brief. Bioinform., vol. 23, no. 1, p. bbab383, 2022.

[11] Y. Peng, S. Yan, J. Xu, Y. Zhang, and F. Wang, "A review of biomedical literature mining for drug repurposing," Brief. Bioinform., vol. 22, no. 4, p. bbaa324, 2021.

[12] J. Lee et al., "BioBERT: a pre-trained biomedical language representa- tion model for biomedical text mining," Bioinformatics, vol. 36, no. 4, pp. 1234–1240, 2020.

[13] Y. Gu et al., "Domain-specific language model pretraining for biomed- ical natural language processing," ACM Trans. Comput. Healthcare, vol. 3, no. 1, pp. 1–23, 2021.

[14] J. X. Yu, Y. Wang, G. X. Xu, and Z. Xu, "A reinforcement learning framework for drug discovery using graph neural networks," Nat. Mach. Intell., vol. 4, pp. 1006–1015, 2022.

[15] M. Olivecrona, T. Blaschke, O. Engkvist, and H. Chen, "Molecular de novo design through deep reinforcement learning," J. Cheminform., vol. 9, no. 1, p. 48, 2017.

[16] E. Gawehn, J. A. Hiss, and G. Schneider, "Deep learning in drug discovery," Mol. Inform., vol. 35, no. 1, pp. 3–14, 2016.

[17] J. Vamathevan et al., "Applications of machine learning in drug dis- covery and development," Nat. Rev. Drug Discov., vol. 18, no. 6, pp. 463–477, 2019.

[18] E. Topol, "High-performance medicine: the convergence of human and artificial intelligence," Nat. Med., vol. 25, no. 1, pp. 44–56, 2019.

[19] Y.-C. Lo, S. E. Rensi, W. Torng, and R. B. Altman, "Machine learning in chemoinformatics and drug discovery," Drug Discov. Today, vol. 23, no. 8, pp. 1538–1546, 2018.

[20] Z. Zhavoronkov et al., "Deep learning enables rapid identification of potent DDR1 kinase inhibitors," Nat. Biotechnol., vol. 37, no. 9, pp. 1038–1040, 2019.

[21] European Medicines Agency, "Guideline on computerised systems and electronic data in clinical trials," EMA, 2022.

[22] U.S. Food and Drug Administration, "Artificial Intelligence and Machine Learning in Software as a Medical Device (SaMD)," FDA, 2021.

[23] A. Holzinger, A. Carrington, and H. Mu¨ller, "Measuring the quality of explanations: the system causability scale (SCS)," KI - Ku¨nstl. Intell., vol. 34, pp. 193–198, 2020.

[24] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you?: Explaining the predictions of any classifier," in Proc. 22nd ACM SIGKDD, pp. 1135–1144, 2016.

[25] T. Ching et al., "Opportunities and obstacles for deep learning in biology and medicine," J. R. Soc. Interface, vol. 15, no. 141, p. 20170387, 2018.

[26] D. Amodei et al., "Concrete problems in AI safety," arXiv preprint arXiv:1606.06565, 2016.