Düzce University
Journal of Science & Technology

Research Article

Study on Drug Repurposing for ALS Treatment Using Pre-trained Knowledge Graph Embeddings: Methods and Findings

Selcan YALKIZIMI [a,*], Ümit ŞENTÜRK [b]

[a] *Department of Computer Engineering, Faculty of Engineering, Bolu Abant İzzet Baysal University, Bolu, TURKEY*
[b] *Department of Computer Engineering, Faculty of Engineering, Bolu Abant İzzet Baysal University, Bolu, TURKEY*
** Corresponding author's e-mail address: 2210711001@ogrenci.ibu.edu.tr*
DOI: 10.29130/dubited.1507832

ABSTRACT

In this study, research has been conducted using pre-trained knowledge graph embedding for drug repurposing in treating ALS (Amyotrophic Lateral Sclerosis), and its results have been presented. Drug repurposing studies for ALS have been carried out through two main methods: disease-drug relationship and genes-drug relationship. Drug repurposing recommendations for ALS have been provided by predicting connections between disease and drug entities on the DRKG (Drug Repurposing Knowledge Graph). The findings obtained from the study have been evaluated by comparing them with the list of clinical trial drugs obtained from DrugBank. DRKG has been utilized as a comprehensive biological knowledge graph containing genes, compounds, diseases, biological processes, side effects, and symptoms. This graph has proven to be an effective resource for extracting information related to ALS disease. In drug repurposing studies, drugs obtained through disease-drug relationships have been compared with the list of clinical trial drugs associated with ALS, yielding significant results. Additionally, interactions between genes associated with ALS and drugs related to these genes have been examined in studies conducted through gene-drug relationships. The results obtained from the study demonstrate that DRKG is an effective resource for identifying drugs with potential therapeutic effects in the treatment of ALS. The findings suggest that drug repurposing studies could offer new and effective solutions for the treatment of ALS, marking a significant step forward in this regard.

Keywords: *Drug repurposing, ALS (Amyotrophic Lateral Sclerosis), Pre-trained knowledge graph embedding, Clinical trial drugs*

Önceden Eğitilmiş Bilgi Grafik Gömme Yöntemleri Kullanılarak ALS Tedavisi için İlaç Yeniden Kullanımı Üzerine Bir Çalışma: Yöntemler ve Bulgular

ÖZ

Bu çalışmada, ALS hastalığının tedavisinde ilaç yeniden kullanımı amacıyla önceden eğitilmiş bilgi grafik yerleştirmesi kullanılarak bir çalışma yapılmış ve sonuçları sunulmuştur. İki ana yöntemle, yani hastalık ve ilaç ilişkisi ile genler ve ilaçlar üzerinden, ALS için ilaç yeniden kullanımı çalışmaları gerçekleştirilmiştir. DRKG (Drug Repurposing Knowledge Graph) üzerinde hastalık ve ilaç varlıkları arasındaki bağlantılar tahmin edilerek ALS için ilaç yeniden kullanımı önerileri sunulmuştur. Çalışma sonucunda elde edilen bulgular, DrugBank üzerinden elde edilen klinik deneme ilaçları listesi ile karşılaştırılarak değerlendirilmiştir. DRKG, genleri, bileşikleri, hastalıkları, biyolojik süreçleri, yan etkileri ve semptomları içeren geniş kapsamlı bir biyolojik bilgi

Received: 03/07/2024, Revised: 01/10/2024, Accepted: 20/10/2024

grafiği olarak kullanılmıştır. Bu grafik, ALS hastalığı ile ilgili bilgilerin çıkarılmasında etkili bir kaynak olmuştur. İlaç yeniden kullanımı çalışmalarında, hastalık-ilaç ilişkisi üzerinden elde edilen ilaçlar, ALS ile ilişkilendirilmiş klinik deneme ilaçları listesiyle karşılaştırılmış ve önemli sonuçlar elde edilmiştir. Ayrıca, gen- ilaç ilişkisi üzerinden yapılan çalışmalarda, ALS ile ilişkilendirilmiş genler ve bu genlerle ilişkilendirilmiş ilaçlar arasındaki etkileşimler incelenmiştir. Çalışmanın elde ettiği sonuçlar, DRKG' nin ALS tedavisinde potansiyel terapötik etkilere sahip ilaçları belirlemede etkili bir kaynak olduğunu göstermektedir. Elde edilen bulgular, ilaç yeniden kullanımı çalışmalarının ALS hastalığının tedavisinde yeni ve etkili çözümler sunabileceği konusunda önemli bir adım olarak değerlendirilebilir.

Anahtar Kelimeler: *İlaç yeniden kullanımı, ALS (Amyotrofik lateral skleroz), Önceden eğitilmiş bilgi grafik gömme, Klinik deneme ilaçları*

I. INTRODUCTION

Amyotrophic Lateral Sclerosis (ALS), first described by Charcot in the nineteenth century, is defined as a progressive neurodegenerative disease with a survival of three to five years. This disease, the treatment and etiology of which are unknown, is characterized by permanent loss of function of upper and lower motor neurons. ALS, which is diagnosed in 1500 to 4500 people every year in our country, is known as a fatal disease characterized by progressive loss of muscle function. Although many symptomatic and therapeutic drug researches continue, only FDA-approved riluzole and edaravone are included in the treatment protocol [1]. There is currently no treatment that will completely cure the disease in motor neuron disease. Symptomatic and supportive therapies help to improve the quality of life and life span of patients [2]. The prognosis of the disease is not good. Patients die 2-5 years after diagnosis [3]. The effect of Riluzole, which is paid for by the Social Security Institution (SGK) in Türkiye and used for the treatment of ALS patients, is also controversial. Although it is claimed that this drug slows down the progression of the disease, it has been observed that some patients who took Riluzole regularly did not stop the progression of the disease and could not prevent the worsening of the prognosis [3]. Although the etiology of ALS is not known exactly, it is suggested that many different factors play a role in its pathogenesis. Genetics, oxidative stress, glutamate excitotoxicity, mitochondrial dysfunction, axonal transport disorder, neuroinflammation, and RNA disorders are the main factors [4]. Familial ALS cases constitute approximately 10% of all ALS cases and are phenotypically and genetically heterogeneous [5]. It is usually inherited as Autosomal Dominant (OD; inheritance pattern in which one copy of a gene causes the disease). The association with a mutation in Superoxide Dismutase 1 (SOD-1; an enzyme that neutralizes free radicals) was first proposed in 1993. Subsequent developments have shown that some genes cause ALS, while others increase the risk of ALS or may affect the disease process [4].

Drug repurposing is an innovative approach to exploring new therapeutic uses for existing drugs. It focuses on research to understand how existing drugs can effectively treat different diseases beyond the purposes for which they were originally designed. It is based on identifying new interactions between biological entities such as genes and compounds. Traditional approaches to do this rely on costly and time-consuming experimental methodologies. As a result, several approaches have been developed that aim to leverage the various types of information that already exist about drugs, their targets, and diseases to reduce cost and accelerate drug reuse. Among these, approaches that represent existing knowledge in the form of a knowledge graph and utilize graph-based machine learning techniques based on graph neural networks and knowledge graph embedding models have gained popularity [6].

The drug discovery process ranges from reading and analyzing already existing literature to testing the ways in which potential drugs interact with targets. The preclinical development phase of drug discovery involves testing potential drug targets in animal models. Using artificial intelligence at this stage can help trials run smoothly [7]. It can enable researchers to more quickly and successfully predict how a drug might interact with an animal model. After going through the preclinical development phase and receiving approval from the United States Food and Drug Administration (FDA or USFDA), researchers began testing the drug with human participants. In general, this is a four-step process and is generally

considered to be the longest and most expensive stage of the manufacturing journey [7]. A representative drug discovery step is schematized in Figure 1. The drug discovery process comprises several major steps that include identifying compounds by screening compound collections via primary assays, such as high through-put screening in vitro, and secondary assays that include counter-screens and ADMET (absorption, distribution, metabolism, excretion, and toxicity) studies. Structure–activity relationship (SAR) and in silico studies in combination with cellular functional tests are used in an iterative cycle to improve the functional properties of the drug candidates. New drug candidates with desired characteristics are synthesized via organic synthesis. The selected drug candidate which has now passed all preclinical tests successfully is given to human patients in a clinical trial [8].

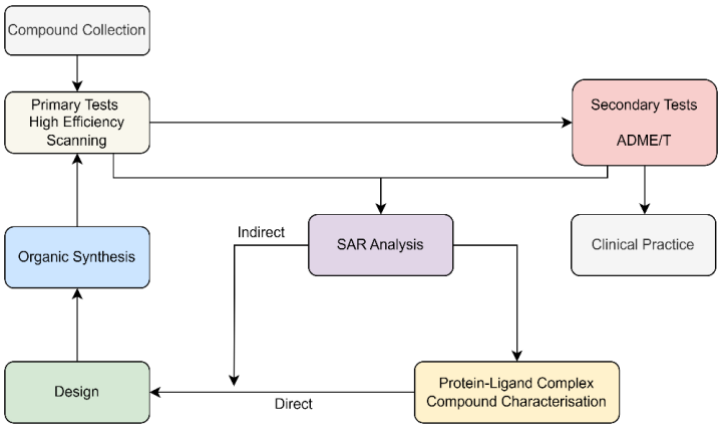

Figure 1. *A representative diagram highlighting the steps and cyclical nature of the drug discovery process [8].*

Most drugs that enter clinical trials fail, often due to a poor understanding of the mechanisms governing drug response. Machine learning techniques hold tremendous promise for better drug response predictions, but most have not reached clinical application due to their lack of interpretability and focus on monotherapies [9]. The methods to be used for the rapid screening of drug molecules are the HTS(High-Throughput Screening) method and the virtual screening method. In the HTS method, thousands of molecules are rapidly screened whether they show activity against a specific receptor or enzyme. In virtual screening, drug molecules can be rapidly classified or ranked in activity order by using machine learning methods. For this purpose, machine learning methods such as support vector machines (SVM) and random forest (RF) are frequently used in the literature for the detection of active molecules. However, these algorithms do not perform well in unbalanced datasets. For this reason, the data sets in the literature where these algorithms are used are mostly composed of balanced datasets[10]. The enormous amount of PubChem bioassay data, PubChem, which is updated daily, constitutes a publicly available big data resource for compounds with various target response information, including most drugs and drug candidates. Similar to PubChem, ChEMBL is a database containing binding, functional, ADME, and toxicity data for a large number of compounds. Compared to PubChem, ChEMBL contains a large amount of manually compiled data from the literature. DrugBank [11], one of the data sources designed specifically for drugs and drug candidates, is a publicly available database containing all approved drugs with their mechanisms, interactions, and relevant targets [12].

With the increase in the data size in the associated databases, it has become necessary to use new methods for virtual screening. In recent years, deep neural networks (DNNs) have shown very good performances in many areas and have surpassed the performances of machine learning methods such as SVM and RF [10]. In a 2014 study, Gramatica et al. presented a new methodology for directing existing drugs to diseases that were not initially targeted using biomedical knowledge. This methodology involves graphically representing and automatically analyzing knowledge using computational linguistics and graph theory [13]. In 2016, Udrescu et al. presented a new approach based on complex network science techniques, starting from the assumption that the analysis of drug-drug interactions can lead to the development of new drug discovery tools. In this study, they revealed functional drug categories and relationships, linked network clusters to relevant pharmacological properties, and

validated 85% of predicted properties by cross-checking with various databases [14]. In 2020, Zhou et al. published a study presenting powerful network-based methodologies for the rapid identification of candidate reusable drugs and potential drug combinations targeting 2019-nCoV/SARS-CoV-2 [15]. In 2020, Ioannidis et al. proposed an RGCN(Relational Graph Convolutional Network) model that greatly outperforms GCN(Graph Convolutional Network) and state-of-the-art KGE (Knowledge Graph Embedding) models on low-connected learning tasks using effective deep graph learning (DGL) methods and confirmed the identification of several drugs used in clinical trials as possible drug candidates [16]. In addition, in 2020, Ioannidis et al. constructed a comprehensive biological knowledge graph between genes, compounds, diseases, biological processes, side effects, and symptoms to assist such machine learning techniques, and this graph was named Drug Repurposing Knowledge Graph (DRKG) [6]. In 2022, Xiangxiang Zeng and his team developed an artificial intelligence model called ImageMol, which was pretrained on 10 million unlabeled molecules. This model successfully predicted molecular targets, as well as properties such as drug metabolism, toxicity, and brain permeability, and identified anti-SARS-CoV-2 candidate molecules. It has been demonstrated that this system could be effective in accelerating the drug discovery process for neurodegenerative diseases like Alzheimer's and diseases like COVID-19 [17]. In the same year, Wang and his team introduced new methods for molecular property prediction and drug discovery using graph and sequence-based neural networks. These methods demonstrated significant improvements in ROC-AUC and PRC-AUC metrics in COVID-19 drug discovery tasks during the AI Cures open challenge [18]. In 2023, Kang-Lin Hsieh and colleagues conducted a study on drug repurposing for Alzheimer's disease using DRKG, showing that molecular profiles integrated into a knowledge graph could systematically identify potentially reusable drugs [19]. Similarly, in 2024, Yunguang Qiu and Feixiong Cheng summarized artificial intelligence-supported drug discovery methodologies focused on the complex molecular structure of Alzheimer's disease, emphasizing the importance of AI-based drug repurposing strategies in identifying new indications for Alzheimer's disease [20].

In contrast to approaches in the existing literature, this study addresses drug reuse for ALS disease through a bidirectional analysis examining both disease-drug and gene-drug associations. DRKG's extensive network of biological knowledge underpins the methods used, enabling a more comprehensive assessment of the potential therapeutic implications of gene-drug and disease-drug linkages in ALS. One of the main contributions of the study is that this two-pronged approach enables a broader perspective of potential drug candidates for ALS. While the literature often focuses on a single type of association, the analysis here of both disease-drug and gene-drug associations provides an opportunity to more accurately determine the therapeutic effects of drugs. The use of DRKG and advanced analysis methods enables prediction of associations between disease and drug entities with higher accuracy through pre-trained knowledge graph overlays and provides drug reuse recommendations. The identification of new treatment strategies for ALS by evaluating gene-drug associations reflects an innovative approach. Furthermore, the comparison of the results with ALS-associated clinical trial drugs provides an important contribution in assessing the clinical relevance of the findings. These contributions draw attention to the development of new approaches to drug reuse in ALS treatment and the importance of knowledge graphs in this process.

In the remaining sections of this study, Chapter 2 provides a detailed explanation of the dataset and methods used in the project, while Chapter 3 presents the findings of research on drug repurposing for ALS disease. Chapter 4 extracts drug repurposing results for ALS disease using DRKG embeddings and summarizes the findings of the study. Subsequently, Chapters 5 and 6 will present a discussion on the overall outline of the study and potential future research directions. Additionally, an evaluation will be conducted on how the findings could contribute to clinical applications and the implications of these findings on ALS treatment. This evaluation may play an important role in shaping future research aimed at developing new therapeutic strategies for the treatment of ALS disease.

II. METHODS

In this section, the dataset and methods used in the paper are described in detail. Using the pre-trained knowledge graph embedding of the dataset DRKG (Drug Reuse Knowledge Graph), a drug reuse study was performed to provide a recommendation for the treatment of ALS disease. Drug reutilisation is studied in two ways: through disease and drug associations and through genes and drugs interactions. By predicting the links between disease entities and drug entities in the DRKG, a drug reuse recommendation for ALS is presented. Firstly, a list of ALS-related diseases was extracted from the DRKG. Diseases are coded with identification numbers on the DRKG. This was done by scanning the identification code of ALS disease on DRKG. A list of drugs on Drugbank was used as candidate drugs in our study. There are 8104 drugs in this drug list. In order to evaluate the drugs for reuse, they were compared with the list of clinical trial drugs associated with ALS disease collected from Drugbank. Figure 2 provides a Graphical Summary of the Data Set and Methods of the Drug Reuse Study for ALS Disease.

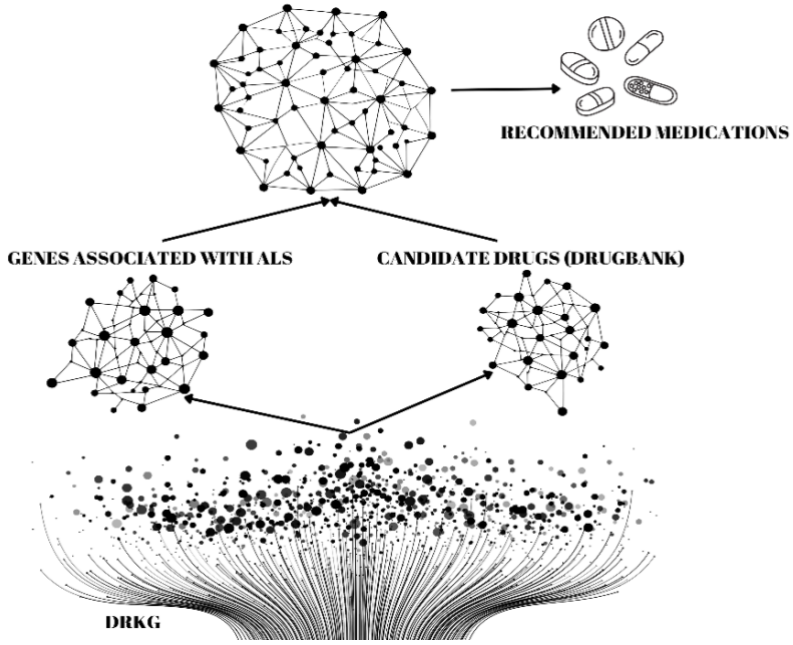

Figure 2. *Drug Reuse Study for ALS Disease: Graphical Summary of Data Set and Methods*

A. DATASET

DRKG is a comprehensive biological knowledge graph that associates genes, compounds, diseases, biological processes, side effects, and symptoms. Figure 3 visualizes the possible interactions between entity type pairs in the DRKG.

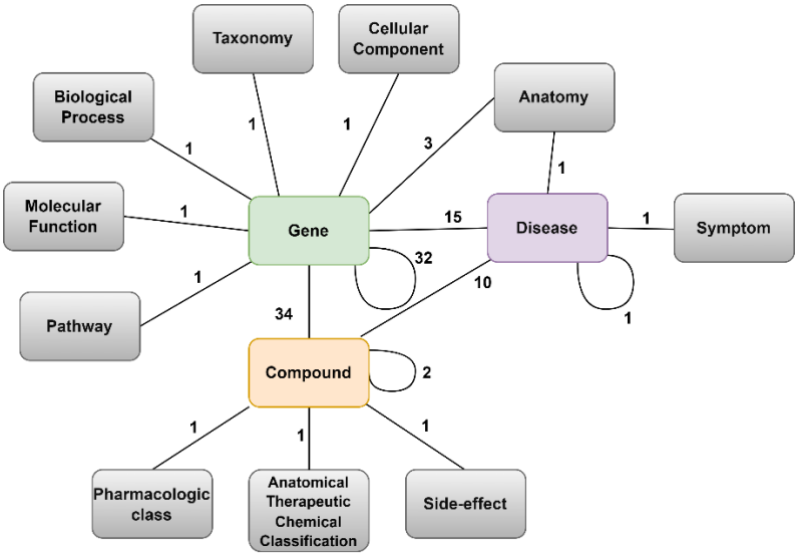

Figure 3. *Representation of the DRKG. The number next to an edge indicates the number of relationship types between the corresponding entity types in the DRKG [21].*

It contains information from six existing databases, including DRKG, DrugBank, Hetionet, GNBR, String, IntAct, and DGIdb, as well as data collected from recent publications specifically related to COVID-19. This database contains 97,238 entities belonging to 13 entity types and 5,874,261 triples belonging to 107 edge types. These 107 edge types represent a type of interaction between one of the 17 entity type pairs, as shown in Table 1 (more than one type of interaction is possible between the same pair of entities)[21]. The type-wise distribution of entities in DRKG and the original data sources are shown in Table 1.

Table 1. *Number of nodes per node type in the data sources in the DRKG [21].*

Entity type	Drugbank	GNBR	Hetionet	STRING	IntAct	DGIdb	Bibliography	Total Entities
Anatomy	-	-	400	-	-	-	-	400
Atc	4,048	-	-	-	-	-	-	4,048
Biological Process	-	-	11,381	-	-	-	-	11,381
Cellular Component	-	-	1,391	-	-	-	-	1,391
Compound	9,708	11,961	1,538	-	153	6,348	6,25	24,313
Disease	1,182	4,746	257	-	-	-	33	5,103
Gene	4,973	27,111	19,145	18,316	16,321	2,551	3,181	39,22
Molecular Function	-	-	2,884	-	-	-	-	2,884
Pathway	-	-	1,822	-	-	-	-	1,822
Pharmacologic Class	-	-	345	-	-	-	-	345
Side Effect	-	-	5,701	-	-	-	-	5,701
Symptom	-	-	415	-	-	-	-	415
Tax	-	215	-	-	-	-	-	215
Total	19,911	44,033	45,279	18,316	16,474	8,899	9,464	97,238

Table 2 shows examples of the number of triples between different pairs of entity types in DRKG for DRKG and various data sources.

Table 2. *Some of the numbers of interactions in DRKG and in data sources [21].*

Entity-type pair	Drugbank	GNBR	Hetionet	STRING	IntAct	DGIdb	Bibliography	Total interactions
(Gene, Gene)	-	66,722	474,526	1,496,708	254,346	-	58,629	2,350,931
(Compound, Gene)	24,801	80,803	51,429	-	1,805	26,29	25,666	210,794

(Disease, Gene)	-	95,399	27,977	-	-	-	461	123,837
(Compound, Compound)	1,379,271	-	6,486	-	-	-	-	1,385,757
(Compound, Disease)	4,968	77,782	1,145	-	-	-	-	83,895

DRKG is available in its final form on Git Hub. An image created by selecting samples from the data on the DRKG database is shown in Table 3.

Table 3. *Sample DRKG database file image.*

Gene::2157	bioarx::HumGenHumGen:Gene:Gene	Gene::2157
Compound::DB09080	bioarx::DrugHumGen:Compound:Gene	Gene::154
Compound::DB00669	DRUGBANK::ddi-interactor-in::Compound:Compound	Compound::DB13064
Gene::10959	GNBR::Te::Gene:Disease	Disease::MESH:D006509
Gene::2742	Hetionet::GpBP::Gene:Biological Process	Biological Process::GO:0006821
Gene::6202	Hetionet::GiG::Gene:Gene	Gene::26156
Anatomy::UBERON:0000057	Hetionet::AuG::Anatomy:Gene	Gene::113791
Compound::DB01115	Hetionet::CcSE::Compound:Side Effect	Side Effect::C0344232
Gene::3007	Hetionet::GpPW::Gene:Pathway	Pathway::PC7_2529

In the DRKG study, data from the DrugBank, Hetionet, GNBR, String, IntAct, and DGIdb databases were filtered and a list of triples (head-entity, relationship type, tail-entity) of data from each dataset was first extracted. In this process, an entity type was associated with an identifier of the entity followed by an identification number that makes it unique. For example Gen::229475. In the representation of relationships, the database name, the relationship name, and the types of head and tail entities were used. Example: DGIDB::INHIBITOR::Gene:Compound. Different ids can be used to represent entities such as genes, compounds, and diseases from different data sources, these ids are mapped to a common ID to remove the incompatibility here. Finally, relations with less than 50 edges were excluded by removing relations with insufficient data [21].

B. GRAPH NEURAL NETWORK(GNN)

Graph neural network (GNN) is a deep learning method that belongs to the artificial neural network (ANN) family and performs information extraction from graphs. It was first used in 2008 and its development started in 2014 and after [22]. GNNs, which were developed in response to the failure of convolutional neural networks (CNNs) to meet the expected performance on visuals, are used in many fields such as physics, chemistry, biology, and cyber security [22]. Graphs are a type of data structure that models a set of objects (nodes) and their relationships (edges). Recently, research on analyzing graphs with machine learning has received increasing attention due to the great expressive power of graphs [23]. A graph G = (V, E) consists of two sets: V the set of nodes (also called vertices), and E the set of edges (also called arcs). Each edge connects a pair of nodes, indicating a relationship between them [6]. Graph analysis, a unique non-euclidean data structure for machine learning, focuses on tasks such as node classification, link prediction, and clustering. Graph neural networks (GNNs) are deep learning-based methods that operate on the graph domain. Due to its convincing performance, GNN has recently become a widely applied graph analysis method [23]. Studies in this field are increasing day by day in Türkiye, and graph-based methods such as GNN are used in many national theses and research projects. However, this type of network is still developing in our country compared to global applications.

In general, the GNN model structure consists of four steps:
1- Finding the structure of the graph,
2- Determining the graph type and scale,
3- Determine the design loss function,
4- Building the model using computational modules [23].

III. FINDINGS

In this section, the findings from the research on drug repurposing for amyotrophic lateral sclerosis (ALS) are presented. The study conducted two separate analyses based on disease-drug and gene-drug relationships to obtain results for both approaches. First, using the **mesh code** for ALS, the connections between disease and drug entities in the DRKG database were predicted. In this process, scores representing the disease-drug relationship were calculated, and drugs with the highest scores were identified. These drugs were then compared with a list of 158 clinical trial drugs related to ALS obtained from DrugBank to derive results. Secondly, gene-drug relationships were examined by assessing the connections between disease-related genes and drugs. Similar formulas and algorithms were employed to determine the relationships between genes and drugs. The findings from both methods were completed by comparing the repurposed drugs with the clinical trial drugs related to ALS.

Mesh code (Medical Subject Headings) is a frequently used indexing term in medical literature. Mesh terms are a standard list of terms used to categorize and index medical subjects. Mesh codes are a unique identifier of each Mesh term and are used to categorize and index research in medical literature. These codes are used in scientific papers, references in medical literature, and databases. Here, we aimed to access all ALS disease entity links in our database by using the mesh code "D000690", which is the mesh code of ALS disease.

$$d = \gamma - \|h + r - t\|_2 \quad (1)$$

$$score = \log\left(\frac{1}{1 + \exp(-d)}\right) \quad (2)$$

In the formula provided in Equation 1[6]:
d: is a score representing the drug-disease relationship.
γ: is a constant coefficient (which may depend on a specific dataset or model parameter).
h: is a vector representing the disease.
r: is a vector representing the relationship between the drug and the disease.
t: is a vector representing the drugs.
‖h + r - t‖_2: This term represents the L2 norm of the vectorial difference between the disease and the drug, which measures how close the drug and the disease are.
This equation measures the similarity between the drug and the disease in vector space and determines how strong the treatment effect is.

In the formula for Log-Sigmoid Score Calculation provided in Equation 2[6]:
score: is defined as the final score of the relationship between the drug and the disease.
log-sigmoid function: brings the score into a specific range, making all scores negative. This function is the logarithm of the sigmoid function, which ensures that the scores remain within a more manageable range.
This equation normalizes the relationship score calculated by d (the value obtained from the first equation) through a logarithmic sigmoid function. The log-sigmoid function compresses the scores into a range between 0 and 1, thus producing a score that represents the strength of the relationship. At the same time, this function is used to bring all scores into a range below 0.

In our study, the ['Hetionet::CtD::Compound:Disease','GNBR::T::Compound:Disease'] connection was used as the edge connection type for the drug-disease relationship. This connection represents the treatment relationship between a specific drug and a disease. When repurposing drugs, only treatment-related edge connections should be used. The formulas provided in Equation 1 and Equation 2 are used to calculate the edge score. Here, the log-sigmoid function must be used to ensure all scores are < 0 [6]. After calculating the edge scores, a list of the drugs with the highest scores, i.e., those most associated

with the disease, was generated. These drugs were then compared with the list of 158 clinical trial drugs related to ALS collected from DrugBank, and a result list was obtained.

In the study conducted through gene-drug relationships, drug repurposing for ALS was investigated by predicting the connections between host gene entities related to the disease and drug entities in DRKG. The connection relationships of target gene entities associated with the disease were listed in a file, and for the drugs, a list entirely obtained from DrugBank was used. The edge connection types 'GNBR::N::Compound:Gene', 'DRUGBANK::target::Compound:Gene, DGIDB::INHIBITOR::Gene:Compound' were used in separate trials. The formulas given in Equation 1 and Equation 2 for calculating the edge scores and the pathway followed in the disease-drug relationship were also applied here. Equation 1 was used to calculate the relationship score between the gene and the drug, thereby measuring the vectorial distance between the genes and drugs to determine the connection strength. Equation 2 used the log-sigmoid function to normalize this score and obtain the edge score. This score was used to identify and rank the relationship between genes and drugs. After calculating the edge scores, the repurposed drugs were compared with clinical trial drugs related to ALS, and a list was created.

IV. RESULTS

In this section, we obtain drug repurposing results for ALS disease by using pretrained DRKG embeddings. In the study based on disease-drug relationships, drug repurposing for ALS was conducted by predicting the connections between disease entities and drug entities in DRKG. In the study based on gene-drug relationships, trials were carried out using inhibitory connections between genes and drugs. Separate results were obtained from the drug repurposing analysis conducted with both methods.

A. DRUG REPURPOSING THROUGH THE DISEASE-DRUG LINK

Among the first 100 drugs found for the treatment of ALS disease based on the disease-drug relationship, 18 drugs, which are also on the drugbank clinical drug trial list, are listed in Table 4 starting from the most related one according to the edge score.

Table 4. *The 18 drugs and their scores are also included in the Drugbank clinical drug trial list.*

Order No	Drug Name	Edge Score
[0]	Trehalose	-0.1777
[1]	Arimoclomol	-0.1847
[3]	Pridopidine	-0.2015
[6]	Testosterone	-0.2211
[8]	Mecobalamin	-0.2360
[13]	Sirolimus	-0.2563
[23]	Minocycline	-0.2878
[25]	Quinidine	-0.2961
[26]	Mycophenolate mofetil	-0.2963
[28]	Ubidecarenone	-0.2986
[31]	Colchicine	-0.3003
[56]	Methylprednisolone	-0.3561
[59]	Cimetidine	-0.3604
[62]	Tamoxifen	-0.3657
[66]	Tretinoin	-0.3686
[68]	Thalidomide	-0.3704
[73]	Capsaicin	-0.3746
[91]	Ceftriaxone	-0.3956

The first 10 of the first 100 drugs listed for the treatment of ALS disease, regardless of whether they were used in clinical trials on the disease-drug relationship, are listed in Table 5 in order of relationship priority.

Table 5. *The top 10 drugs are listed regardless of whether they have been used in clinical trials.*

Order No	Drug Name	Identity Code	Edge Score
1	Trehalose	DB12310	-0.1777
2	Arimoclomol	DB05025	-0.1847
3	Glutathione	DB00143	-0.1869
4	Pridopidine	DB11947	-0.2015
5	Melatonin	DB01065	-0.2073
6	Cholesterol	DB04540	-0.2127
7	Testosterone	DB00624	-0.2211
8	Dexamethasone	DB01234	-0.2328
9	Mecobalamin	DB03614	-0.2360
10	Estradiol	DB00783	-0.2475

When evaluating drugs associated with ALS, the reason for considering not only the drugs used in clinical studies but also those that have not yet been implemented in clinical practice is that some drugs, despite being insufficiently researched, may provide potential benefits and pave the way for future studies. Furthermore, the evaluation of these drugs can contribute to innovative approaches in the management of ALS, offering a broader treatment perspective.

When comparing the drugs listed in Table 5, which we obtained as output, with the clinical trial list related to ALS collected from DrugBank in Table 4, it is observed that the five drugs (Trehalose, Arimoclomol, Pridopidine, Testosterone, Mecobalamin) that have the highest edge scores and are most associated with the disease are common in the clinical trial list.

B. DRUG REPURPOSING THROUGH GENE-DRUG LINKAGE

In this section, experiments were performed by selecting inhibitory association links appearing in three data sources: 'GNBR::N::Compound:Gene', 'DRUGBANK::target::Compound:Gene' and 'DGIDB::INHIBITOR::Gene:Compound'. In the study for the gene-drug association for ALS disease treatment, in order to predict the links between genes and drugs to act as inhibitors, a list of biological gene entities associated with ALS disease was taken on DRKG and it is seen that there are 550 associated genes. Then, using the pre-trained knowledge graph placements and the equation used to calculate the edge score, we found the 100 drugs with the highest score and ranked these drugs per target gene. Thus, according to the type of inhibitory relationship we chose, the gene with the highest drug-gene association and the drugs associated with it were found. Then, in order to evaluate the connection of our prediction with the clinical drugs used in treatment, the overlap and the number of hits between these 100 predicted drugs and the drugs used in clinical trials for treatment were checked. Here, the number of hits is related to determining how many of the total number of genes the predicted drug is associated with. If a drug is associated with all genes, the number to be seen here will be equal to the total number of genes we found associated with ALS, i.e. 550. The same study was repeated for the three types of association mentioned above.

Within the scope of the 'GNBR::N::Compound::Gene' relationship, when we listed the genes associated with ALS disease in the list of genes associated with this disease, Gene::627 and Gene::9217 genes associated with 16 drugs were determined as the genes with the most drug associations. Among the drugs associated with these genes, those with the most drug associations are listed first and these associations can be seen in Table 6.

Table 6. *List of drugs Gene::627 and Gene::9217 related to 16 drugs by 'GNBR::N::Compound:Gene' relationship.*

Gene::627	16	Gene::9217	16
[9]	Tretinoin	[5]	Sirolimus
[12]	Capsaicin	[21]	Ubidecarenone
[17]	Testosterone	[24]	Quinidine
[32]	Sirolimus	[30]	Tamoxifen
[33]	Tamoxifen	[32]	Capsaicin
[44]	Methylprednisolone	[53]	Bosutinib
[50]	Scopolamine	[56]	Thalidomide
[59]	Thalidomide	[59]	Methylprednisolone
[69]	Dextromethorphan	[60]	Tretinoin
[79]	Trazodone	[65]	Prednisone
[81]	Minocycline	[67]	Cimetidine
[83]	Dronabinol	[77]	Minocycline
[84]	Prednisone	[81]	Colchicine
[87]	Pimozide	[93]	Testosterone
[89]	Atropine	[97]	Mycophenolate mofetil
[91]	Midazolam	[99]	Atropine

Table 7 shows the top 10 of the 100 drugs with the highest score listed according to the 'GNBR::N::Compound:Gene' relationship and the top 10 drugs that overlap with the list of ALS clinical trial drugs and how many genes these drugs are associated with, that is, the number of hits. For example, the drug named Sirolimus coded with the id 'DB00877', which is also included in the clinical trial list related to ALS Disease, is associated with 538 genes related to ALS disease and ranks first, while the drug named Tretinoin coded with the id 'DB00755' is associated with 522 genes and ranks second. The drug named Tamoxifen with the ID code 'DB00675' is in the 3rd place with 478 hits.

Table 7. *The top 100 highest-scoring drugs listed by 'GNBR::N::Compound:Gene' association overlap with the list of ALS clinical trial drugs and the number of genes with which these drugs are associated.*

Identity Code	Drug Name	Hit Count
DB00877	Sirolimus	538
DB00755	Tretinoin	522
DB00675	Tamoxifen	478
DB00624	Testosterone	398
DB01041	Thalidomide	360
DB00908	Quinidine	307
DB06616	Bosutinib	257
DB06774	Capsaicin	256
DB00635	Prednisone	198
DB01394	Colchicine	179

Within the scope of the 'DRUGBANK::target::Compound::Gene' relationship, Gene::283, Gene::5551, and Gene::6647 genes, which were found to be associated with 13 drugs, were identified as the genes with the most drug associations in the list of genes associated with ALS disease. Among the drugs associated with these genes, those with the most drug associations were ranked first and these associations can be seen in Table 8. When the drugs associated with three different genes that ALS disease interacts within the current association type are analyzed, it is seen that there are common drugs in all three gene types.

Table 8. *List of drugs Gene::5551 and Gene::283 related to 13 drugs by association with 'DRUGBANK::target::Compound:Gene'.*

Gene::283	**13**	**Gene::5551**	**13**	**Gene::6647**	**13**
[8]	Bosutinib	[7]	Sirolimus	[0]	Arimoclomol
[11]	Capsaicin	[11]	Prednisone	[7]	Cannabidiol
[19]	Colchicine	[13]	Tretinoin	[23]	Medical Cannabis
[26]	Cannabidiol	[16]	Thalidomide	[30]	Bosutinib
[36]	Tretinoin	[22]	Minocycline	[33]	Sirolimus
[49]	Trametinib	[36]	Methylprednisolone	[51]	Dolutegravir
[54]	Sirolimus	[39]	Mycophenolate mofetil	[69]	Minocycline
[60]	Tenofovir alafenamide	[44]	Colchicine	[70]	Colchicine
[66]	Tamoxifen	[56]	Dextromethorphan	[77]	Tretinoin
[80]	Testosterone	[68]	Quinidine	[78]	Mycophenolate mofetil
[81]	Medical Cannabis	[89]	Bosutinib	[81]	Quinidine
[91]	Thalidomide	[91]	Testosterone	[83]	Tauroursodeoxycholic acid
[95]	Dolutegravir	[94]	Tacrolimus	[84]	Capsaicin

Table 9 shows the top 10 of the 100 drugs with the highest score listed according to the 'DRUGBANK::target::Compound::Gene' relationship and the top 10 drugs that overlap with the ALS clinical trial drugs list and how many genes these drugs are associated with, that is, the number of hits. Here, it was revealed that the drug named Bosutinib coded with the ID code ‘DB06616’, which is also included in the clinical trial list related to ALS Disease, ranked first with 446 genes related to ALS disease, the drug named Dolutegravir coded with the id code ‘DB08930’ ranked second with 379 genes, and the drug named Sirolimus coded with the id code ‘DB00877’ ranked third with 360 genes.

Table 9. *The top 100 highest-scoring drugs listed by 'DRUGBANK::target::Compound::Gene' are associated with the list of ALS clinical trial drugs and the number of genes with which these drugs are associated.*

Identity Code	**Drug Name**	**Hit Count**
DB06616	Bosutinib	446
DB08930	Dolutegravir	379
DB00877	Sirolimus	360
DB00755	Tretinoin	320
DB00908	Quinidine	218
DB09061	Cannabidiol	211
DB06774	Capsaicin	187
DB00624	Testosterone	152
DB08911	Trametinib	152
DB00675	Tamoxifen	120

Within the scope of the 'DGIDB::INHIBITOR::Gene::Compound' relationship, when we list the genes associated with ALS disease in the list of genes associated with this disease, Gene::4283, which is associated with 14 drugs, was determined as the gene with the most drug associations. Among the drugs associated with this gene, those with the most drug associations are listed first and these relationships can be seen in Table 10.

Table 10. *List of drugs related to Gene::4283 which is related to 14 drugs by association with 'DGIDB::INHIBITOR::Gene:Compound'.*

Gene::4283	14
[7]	Olanzapine
[12]	Methylprednisolone
[33]	Prednisone
[46]	Ciprofloxacin
[57]	Sirolimus
[61]	Corticotropin
[64]	Tretinoin
[68]	Tacrolimus
[73]	Fingolimod
[76]	Testosterone
[85]	Pimozide
[93]	Thalidomide
[94]	Atropine
[98]	Trazodone

Table 11 shows the top 10 of the 100 drugs with the highest score listed according to the 'DGIDB::INHIBITOR::Gene::Compound' relationship and the top 10 drugs that overlap with the ALS clinical trial drugs list and how many genes these drugs are associated with, that is, the number of hits. It was revealed that the drug named Tretinoin coded with the ID code 'DB00755', which is also included in the clinical trial list related to ALS Disease, ranked first in relation to 455 genes related to ALS disease, the drug named Sirolimus coded with the id code 'DB00877' ranked second in relation to 445 genes and the drug named Bosutinib coded with the id code 'DB06616' ranked third in relation to 425 genes.

Table 11. *The top 100 highest-scoring drugs listed by 'DGIDB::INHIBITOR::Gene:Compound' association overlap with the list of ALS clinical trial drugs and the number of genes with which these drugs are associated.*

Identity Code	Drug Name	Hit Count
DB00755	Tretinoin	455
DB00877	Sirolimus	445
DB06616	Bosutinib	425
DB00864	Tacrolimus	314
DB00908	Quinidine	291
DB00624	Testosterone	266
DB08911	Trametinib	230
DB01041	Thalidomide	194
DB00688	Mycophenolate mofetil	150
DB00675	Tamoxifen	144

In this study, drug reuse studies were performed separately on the disease-drug relationship and gene-drug relationship and separate results were obtained for both studies. A list of 158 drugs in clinical trials related to ALS disease from clinical drug studies on Drugbank was compiled and the results were compared with this list.

Among the first 100 drugs found through the disease-drug relationship for the treatment of ALS disease, 18 drugs that are also on the drugbank clinical drug trial list were found and listed. When we compared the clinical trial list compiled from DrugBank with the first 10 drugs selected among the top 100 drugs with the highest edge score, i.e. the most associated with the disease, which we obtained regardless of whether they were used in clinical studies on disease-drug association for the treatment of ALS disease, it was seen that the 5 drugs with the highest edge score in the clinical trial list (Trehalose, Arimoclomol, Pridopidine, Testosterone, Mecobalamin) were the same.

In the study conducted for the gene-drug association, 550 gene information was collected on ALS-related DRKG, and using the inhibitory links between these genes and drugs, a drug reuse study was

performed separately according to each selected inhibitory linkage type associated with ALS disease, and the 100 drugs with the highest score were found and these drugs were ranked per target gene. As a result, the gene with the most drug-gene connections and the drugs associated with it were listed according to the type of inhibitory linkage we selected. In addition, we listed the 100 drugs with the highest scores that overlapped with the list of ALS clinical trial drugs, that is, how many drugs are currently registered as trial drugs for ALS disease on the Drugbank and how many genes these drugs are associated with (number of hits). The same study was repeated for three different inhibitory association types and separate results were obtained. According to these results, the prominent findings for each association type are as follows; According to the 'GNBR::N::Compound::Gene' Association, the genes associated with the most drugs are Gene::627 and Gene::9217 and these genes are associated with 16 drugs. According to this association type, Sirolimus (DB00877) is associated with 538 genes, Tretinoin (DB00755) is associated with 522 genes, and Tamoxifen (DB00675) associated with 478 genes are among the prominent drugs overlapping with the clinical trial drug list. According to the 'DRUGBANK::target::Compound::Gene' Association, the genes associated with the most drugs are Gene::283, Gene::5551, and Gene::6647, which are associated with 13 drugs. Drugs that stand out by overlapping with the clinical trial drug list according to this association type include Bosutinib (DB06616), which is associated with 446 genes, Dolutegravir (DB08930), which is associated with 379 genes, and Sirolimus (DB00877), which is associated with 360 genes. According to the 'DGIDB::INHIBITOR::Gene::Compound' association, Gene::4283 is the gene associated with the most drugs, and this gene is associated with 14 drugs. The prominent drugs that overlap with the clinical trial drug list according to this association type are Tretinoin (DB00755), which is associated with 455 genes, Sirolimus (DB00877), which is associated with 445 genes, and Bosutinib (DB06616), which is associated with 425 genes. These results show interactions between ALS disease-associated genes and their associated drugs for all three types of associations. The drug Tretinoin (DB00755) shows interactions with ALS disease-associated genes under both the 'GNBR::N::Compound:Gene' association and the 'DGIDB::INHIBITOR::Gene:Compound' association. Bosutinib (DB06616) stands out by showing interaction with ALS disease-associated genes under 'DRUGBANK::target::Compound::Gene' and 'DGIDB::INHIBITOR::Gene::Compound' association types. In addition, Sirolimus (DB00877) interacts with genes associated with ALS disease under all three association types. These common drugs were associated with genes associated with ALS disease under different association types in the gene-drug association study. On the other hand, Sirolimus (DB00877) was prominent both in the gene-drug association study and in the disease-drug association study. These results indicate that these drugs may have potential therapeutic effects on ALS and may influence the prioritization of these drugs in relevant studies. However, as stated at the beginning of the study, although the FDA-approved drugs riluzole and edaravone, which are included in the current treatment protocol for ALS disease, were included in our clinical trial drug list, they were not listed as recommendations as a result of our training. However, the results of this study show that many drugs used in clinical trials were discovered as possible drug candidates. This study by no means recommends specific drugs but provides a deep learning methodology to prioritize some of the available drugs for research.

V. DISCUSSION

The results of this study represent an important step in the discovery of potential ALS treatment candidates through disease-drug and gene-drug linkages. In the disease-drug association study, when the drugs with the highest scores were identified, they were found to be in common with some clinical trial drugs. This suggests that the disease-drug association approach may be effective in identifying the potential reuse of drugs currently being trialed in clinical trials.

On the other hand, the gene-drug association study examined the interactions between genes associated with ALS disease and drugs associated with these genes using different types of inhibitory relationships. This study highlights the potential of a gene-drug association-based treatment strategy for ALS disease. In particular, the identified partner drugs were associated with ALS disease-associated genes under different association types, suggesting that these drugs may have potential therapeutic effects on ALS.

The results of this study may contribute to the development of new approaches for the discovery of potential treatment candidates for ALS. However, the limitations of this study should also be considered. For example, the lack of data sources used, or the limited datasets may affect the generalisability of the results. Therefore, it is important for further research to confirm and extend these findings.

VI. CONCLUSION

In this study, we focused on the identification of potential therapeutic drugs for the treatment of ALS disease. For this purpose, separate studies on disease-drug association and gene-drug association were carried out. In addition, drugs in the clinical trial phase in the DrugBank database were also analyzed and compared with the results obtained.

Firstly, in the study on the disease-drug association, drug reuse studies for ALS disease were performed using data to estimate the links between disease and drug entities in DRKG. Using the mesh code 'D000690', access to all entity links for ALS disease was provided. Then, the drugs with the highest scores on the disease-drug association were identified and these drugs were compared with ALS disease-related clinical trial drugs collected from drugbank. Similarly, in the gene-drug association study, the links between genes and drugs for ALS disease were predicted. Separate trials were conducted for three different relationship types and the drugs with the highest scores were identified. Finally, the results obtained through disease-drug association and gene-drug association may play an important role in identifying potential treatment candidates for ALS disease.

VII. REFERENCES

[1] Z. Yildirim, D. N. Toktaş, Ö. Demir, G. Zülfiye, and B. Ş. Utsukarçi, "Amyotrofik Lateral Skleroz Patofizyolojisi ve Tedavi Yaklaşımları," *Hacettepe University Journal of the Faculty of Pharmacy,* vol. 43, no. 1, pp. 50-69, 2023.

[2] F. Kamel, D. M. Umbach, L. Stallone, M. Richards, H. Hu, and D. P. Sandler, "Association of lead exposure with survival in amyotrophic lateral sclerosis," *Environmental Health Perspectives,* vol. 116, no. 7, pp. 943-947, 2008.

[3] F. Özcan, A. Kaya, and M. E. Yayla, "Amyotrofik lateral skleroz hastalığı ve aile hekiminin rolü," *Celal Bayar Üniversitesi Sağlık Bilimleri Enstitüsü Dergisi,* vol. 3, no. 3, pp. 431-435, 2016.

[4] A. D. Marangoz and Ç. Erdoğan, "Amyotrofik lateral skleroz (ALS) hastalığının patogenezi," *Pamukkale Medical Journal,* vol. 13, no. 2, pp. 477-484, 2020.

[5] S. Byrne *et al.*, "Rate of familial amyotrophic lateral sclerosis: a systematic review and meta-analysis," *Journal of Neurology, Neurosurgery & Psychiatry,* 2010.

[6] V. N. Ioannidis *et al.*, "Drkg-drug repurposing knowledge graph for covid-19," *arXiv preprint arXiv:2010.09600,* 2020. [Online]. Available: https://github.com/gnn4dr/DRKG/tree/master?tab=readme-ov-file#drkg-dataset.

[7] İ. N. Çelik, F. K. Arslan, R. Tunç, and İ. Yıldız, "İlaç Keşfi ve Geliştirilmesinde Yapay Zekâ," *Journal of Faculty of Pharmacy of Ankara University,* vol. 45, no. 2, pp. 400-427, 2021.

[8] H. S. Chan, H. Shan, T. Dahoun, H. Vogel, and S. Yuan, "Advancing drug discovery via artificial intelligence," *Trends in pharmacological sciences,* vol. 40, no. 8, pp. 592-604, 2019.

[9] B. M. Kuenzi *et al.*, "Predicting drug response and synergy using a deep learning model of human cancer cells," *Cancer cell,* vol. 38, no. 5, pp. 672-684. e6, 2020.

[10] H. Kanberiz, "Derin öğrenme ile ilaç moleküllerinin aktivitelerinin sınıflandırılması," Sağlık Bilimleri Enstitüsü.

[11] DrugBank. "DrugBank Online." https://go.drugbank.com (accessed 27 January 2024, 2024).

[12] H. Zhu, "Big data and artificial intelligence modeling for drug discovery," *Annual review of pharmacology and toxicology,* vol. 60, pp. 573-589, 2020.

[13] R. Gramatica, T. Di Matteo, S. Giorgetti, M. Barbiani, D. Bevec, and T. Aste, "Graph theory enables drug repurposing–how a mathematical model can drive the discovery of hidden mechanisms of action," *PloS one,* vol. 9, no. 1, p. e84912, 2014.

[14] L. Udrescu *et al.*, "Clustering drug-drug interaction networks with energy model layouts: community analysis and drug repurposing," *Scientific reports,* vol. 6, no. 1, p. 32745, 2016.

[15] Y. Zhou, Y. Hou, J. Shen, Y. Huang, W. Martin, and F. Cheng, "Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2," *Cell discovery,* vol. 6, no. 1, p. 14, 2020.

[16] V. N. Ioannidis, D. Zheng, and G. Karypis, "Few-shot link prediction via graph neural networks for covid-19 drug-repurposing," *arXiv preprint arXiv:2007.10261,* 2020.

[17] X. Zeng *et al.*, "Accurate prediction of molecular properties and drug targets using a self-supervised image representation learning framework," *Nature Machine Intelligence,* vol. 4, no. 11, pp. 1004-1016, 2022/11/01 2022, doi: 10.1038/s42256-022-00557-6.

[18] Z. Wang *et al.*, "Advanced graph and sequence neural networks for molecular property prediction and drug discovery," *Bioinformatics,* vol. 38, no. 9, pp. 2579-2586, 2022, doi: 10.1093/bioinformatics/btac112.

[19] K.-L. Hsieh, G. Plascencia-Villa, K.-H. Lin, G. Perry, X. Jiang, and Y. Kim, "Synthesize heterogeneous biological knowledge via representation learning for Alzheimer's disease drug repurposing," *Iscience,* vol. 26, no. 1, 2023.

[20] Y. Qiu and F. Cheng, "Artificial intelligence for drug discovery and development in Alzheimer's disease," *Current Opinion in Structural Biology,* vol. 85, p. 102776, 2024.

[21] V. N. a. S. Ioannidis, Xiang and Manchanda, Saurav and Li, Mufei and Pan, Xiaoqin and Zheng, Da and Ning, Xia and Zeng, Xiangxiang and Karypis, George. "DRKG - Drug Repurposing Knowledge Graph for Covid-19." https://github.com/gnn4dr/DRKG/ (accessed 15 January 2024, 2024).

[22] H. T. Gümüş and C. Eyüpoğlu, "Grafik Sinir Ağlarına Genel Bir Bakış," *EMO Bilimsel Dergi,* vol. 13, no. 2, pp. 39-56, 2023.

[23] J. Zhou *et al.*, "Graph neural networks: A review of methods and applications," *AI open,* vol. 1, pp. 57-81, 2020.