2023 International Conference on Smart Applications, Communications and Networking (SmartNets) | 979-8-3503-0252-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/SmartNets58706.2023.10215985

A Computational Drug Repositioning Effort using Patients' Reviews Dataset

Ali Akkaya
Department of Computer Engineering
Abdullah Gul University
Kayseri, Turkey
ali.akkaya@agu.edu.tr

Gokhan Bakal
Department of Computer Engineering
Abdullah Gul University
Kayseri, Turkey
gokhan.bakal@agu.edu.tr

***Abstract*—The drug discovery process is one of the core motivations in both medical and, specifically, pharmaceutical disciplines. Due to the nature of the process, it requires an excessive amount of time, clinical experiments, and budget to cover each discovery phase. In this sense, computational drug discovery efforts can shorten the discovery process by providing plausible candidates since many of the attempts fail for several reasons, such as a lack of participants, financial problems, or ineffective results. In this study, the goal is to identify plausible candidate drugs for diseases. To do that, we utilize a personal experience of drugs dataset generated by patients. Beyond the user-generated comments, the users also give a rate between 1 and 10. Since we want to ensure the dataset quality, we first performed sentiment analysis experiments to prove that the reviews/comments are consistent with the given rating score. Then, only the review pairs having an effectiveness rate of 6 or more are selected as pre-filtered drug-disease pairs. We also build a knowledge graph using treatment-related biomedical relations using predications from Semantic Medline Database to identify drug similarities utilizing the Simrank similarity algorithm. As a result, we reported a list of plausible drugs as repurposing/repositioning candidates for further experiments.**

***Index Terms*—computational drug repositioning, sentiment analysis, machine learning, simrank similarity**

I. Introduction

Drug repositioning involves the identification of new therapeutic indications for existing drugs, which can help to overcome the high failure rate and long development times associated with traditional drug discovery phases. This approach has several benefits, including the potential to reduce costs, accelerate development timelines, and leverage the existing knowledge and safety data of the repurposed drugs. However, it also poses particular challenges, such as the need for new clinical trials, regulatory approvals, and the potential unforeseen side effects or drug-drug interactions. Despite these challenges, drug repositioning has the potential to provide significant new treatment options for patients and the pharmaceutical industry. The essential objective of this effort is to generate a list of potential candidate drugs as drug repositioning/repurposing research in the computational (in silico) setting. This candidate list will be created using patient reviews, domain experts' opinions, and computational methods such as machine learning and graph analysis to identify any reasonable similarities between the drugs for new indications. Therefore, the aim is to provide a starting point to help accelerate the drug repositioning process for further computational and clinical investigations. Regarding the consistency of the patients' reviews, we intentionally picked the psychological drug reviews since the impact of drugs can be more perceivable than other cases by the patients. Even though the patients' reviews cannot be unquestionable information, they still can be exploited as strong research indicators.

Based on the experimental drug repurposing experiments, the fundamental contributions of this work are as mentioned below:

- We showed that the user-generated drug reviews are consistent enough to be utilized in further computational efforts in drug repurposing studies.
- We generated knowledge graphs constructed by biomedical entities linked by therapeutically meaningful relations, including *"treats"*, *"disrupts"*, *"prevents"*, *"augments"*, *"stimulates"*, and *"co-exists with"*.
- We built a computational model exploiting the graph-based simrank [1] node (indicating a drug) similarity metric and demonstrated that simple but intuitively potential approaches could be as powerful as more sophisticated models.

The remaining part of the paper is formed as follows. Section II briefly explains relevant literature and background, while Section III introduces the dataset used in the experimental study. Section IV describes the core methodologies employed during the model constructions with the experimental configurations. Section V demonstrates and evaluates the results and finally, Section VI concludes the proposed study and provides an overall summary as well as a couple of potential future research directions.

II. Background

Drug discovery studies for known diseases are rather critical in medical disciplines since clinical investigation phases take an extreme amount of time and investment money. In this sense, computational in-silico analyses play a crucial role in determining the most plausible candidate drug groups before clinical studies. In this way, the required high cost and clinical

979-8-3503-0252-3/23/$31.00 ©2023 IEEE

studies taking a long time can be shortened by focusing on the candidates having a high treatment potential for the target disease [2, 3].

Investigation on an already approved drug for one disease, whether it can be used for another indication in the drug discovery, also significantly expedites the discovery process. These particular investigations are known as drug repositioning/repurposing [4]. At this point, the key advantage here is that the already-approved drugs do not cause any toxic effects on human metabolism, which has been confirmed by previous clinical studies. In general, drug repositioning experiments are conducted based on the similarities in the mechanism of action between the drugs in use and the similarity between the existing diseases [5]. In this regard, interaction-based studies between biomedical entities (such as genes, drugs, organic compounds, proteins, diseases, etc.) have become quite popular in computational drug discovery and repositioning tasks [6].

In this study, the SemMedDB (Semantic Medline Database) knowledge base [7], which is created by the SemRep natural language processing tool [8] by running on the academic studies available on PubMed, is used to generate a therapeutic knowledge graph.

III. Drug Experience Dataset

The initial dataset is patient reviews on drugs, which were generated by scraping the data from the drugs.com website. The dataset contains six data columns, including "*drugName*", "*condition*", "*review*", "*rating*", "*date*", and "*usefulCount*". The dataset was already split into a train and test set, with the train set having 161,297 samples and the test set having 53,766 examples by the dataset creators. Since the review data was not encoded appropriately, we decoded problematic values by semi-automatic ways to use them in the model. Additionally, there were 1,194 missing values discarded from the dataset.

IV. Methodology

In this section, we first explain the sentiment analysis to show the consistency between the user-generated drug reviews and the rating score. Then, we describe the contextual similarity experiments using the SimRank similarity algorithm employed over the generated therapeutic knowledge graph.

A. Sentiment Analysis on User Reviews

Sentiment analysis, also known as opinion mining, intensively exploits natural language processing, text analysis, and computational linguistics to identify and extract subjective information from given textual instances. Technically, it includes determining the emotional tone behind words, such as positive, negative, or neutral, along with identifying and extracting personal information, such as opinions and evaluations [9]. Sentiment analysis can be applied to multiple fields, including social media monitoring, customer feedback analysis, and brand reputation management. The principal goal of sentiment analysis is to understand the attitudes, opinions, and emotions based on the activities of an individual or a particular group. In this work, sentiment analysis is employed to demonstrate the consistency of the patient reviews dataset created by Gräßer et al. [10]. Logistic Regression and Long Short-Term Memory (LSTM) models are implemented to classify the patient reviews based on the rating intervals. The goal is to prove that the patient reviews and their ratings are consistently aligned and in correlation.

1) Data Cleaning: During the data preprocessing phase, the following steps were applied to clean and preprocess the data for better discovery experiments:

- The test and train sets were concatenated to have more instances.
- The rows containing missing values were dropped from the dataset.
- The "review" column values were encoded properly.
- The date and usefulCount columns were dropped.
- Some samples only contained "¡spanX¿" values in the "condition" column due to scraping mistakes, and they were dropped as they did not carry any meaning.
- Each instance was labeled according to the "rating" column. The records having a rating less than or equal to 4 were labeled as -1, the rating greater than 4 and less than 7 were labeled as 0, and ratings greater than or equal to 7 were labeled as 1. The review instance distribution according to the rating score is shown in Figure 1
- Finally, the preprocessed dataset has 212,689 and contains five columns: "*drugName*", "*condition*", "*review*", "*rating*", and "*ratingLabel*".

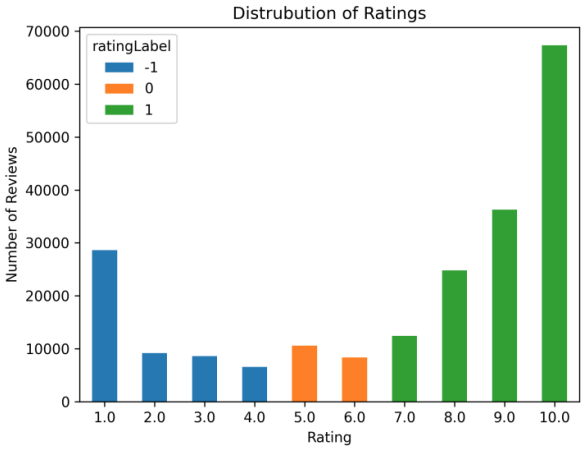


Fig. 1: Rating review instance distribution.

2) Logistic Regression: Logistic regression (LR) is a statistical method employed for predicting binary outcomes, such as the likelihood of an event occurring or not occurring. The model works on the logistic function, which maps the input variables to a probability between 0 and 1, representing the likelihood of the event occurring [11]. The logistic function is a `sigmoid` function formed as an S-shaped curve, and the parameters are estimated using maximum likelihood estimation.

Logistic regression is a supervised learning algorithm in which labeled data is required to train the model, and it can be used for classification and prediction tasks. Plus, the LR model is simple to interpret and widely used in numerous fields such as finance, healthcare, and marketing.

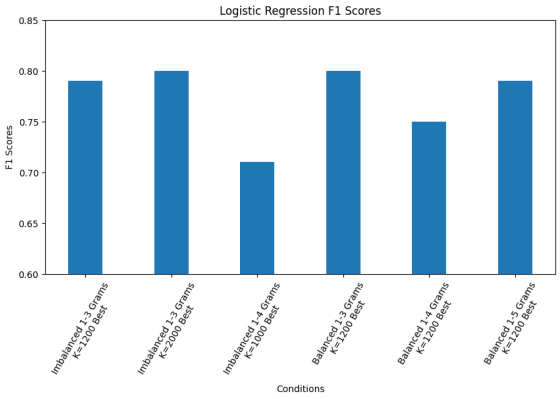
In Equation 1, the LR formula is shown, where "X" is the input value parameter, "a" is the bias term, and "b" is the coefficient for the input parameter.

$$P = \frac{e^{a+bX}}{1 + e^{a+bX}} \tag{1}$$

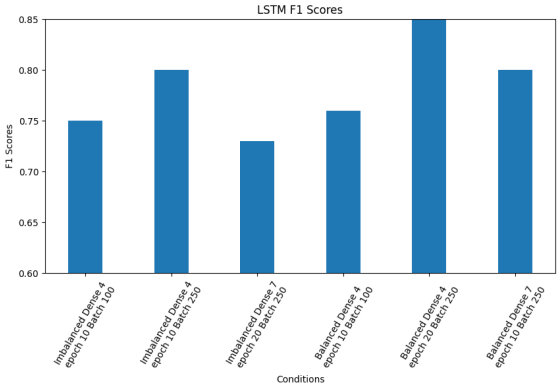
3) Long Short-Term Memory (LSTM) Algorithm: Long Short-Term Memory (LSTM) is a specific Recurrent Neural Network (RNN) type that is particularly well-suited for modeling sequential data. LSTMs are designed to overcome the vanishing and exploding gradient problem, which is common in traditional RNNs [12]. The memory cell value in each unit is controlled by three gates: the `input gate`, `forget gate`, and the `output gate` to regulate the information flow in the cell structure. The input gate controls the new information to be stored, while the forget gate controls the information to be removed from the current cell state. Finally, the output gate manages the information to be passed into the next cell as output. In practice, LSTM models are often used for various natural language processing tasks such as language translation [13], sentiment analysis [14], and text summarization [15].

4) Sentiment Analysis Results: In this subsection, we report the sentiment analysis performance results to indicate the reliability of patient reviews on drugs.com. The study was conducted on reviews for conditions chosen by the domain experts, resulting in a dataset of 50,709 samples. The dataset was split as 80/20 train-test sets with 10-fold cross-validation for the experiments. As shown in Figure 2, the F1 score on the test dataset for both algorithms was between 0.70% and 0.85%. However, the LR algorithm did not achieve any better score than 0.80%, even with various scenarios attempted. In contrast, the LSTM algorithm yielded the F1-score of 0.85% for the best-performing setup.

As can be observed from the performance chart, both models proved that the drug reviews on target psychological diseases are consistent enough for the next step of the computational experiment.



(a) Logistic regression (LR) models' F1 scores



(b) LSTM models' F1 scores

Fig. 2: F1 scores of LR and LSTM models

B. Drug Similarity Experiments on Knowledge Graphs

The general execution flow starts with checking the consistency of the input drug review instances, as illustrated in Figure 3. Afterward, the specific sub-instances are retained as the target diseases, which are psychological conditions. Here, psychological diseases are deliberately selected because we believe that a patient suffering from a psychological problem can feel the treatment effect of the drug taken more than other illnesses. Following by proving the rationality of the drug comment instances, we collected the psychological disease records from the whole dataset under the inspection of psychology domain experts. Based on the previous steps, a new dataset was created containing disease names, the drugs used to treat them, the number of reviews per drug, and the average rating per drug. This dataset was used as the primary data source for further graph-based similarity analysis.

1) SemMedDB Knowledge Base: SemMedDB is a large database containing relationships of biomedical entities and relations in the form of (subject, predicate, and object) obtained from biomedical articles/papers (using titles and abstracts). It is public and made available by the National Library of Medicine (NLM) in the United States for research purposes. It uses the SemRep tool, a rule-based NLP system, to extract "semantic relations"-also called semantic predications, from biomedical text. The relation extraction is performed by the SemRep tool by normalizing the entities (subjects and objects) to single UMLS Metathesaurus concepts (referenced by a unique CUI-Concept Unique Identifier) and basing the predicates on those present in the UMLS semantic web.

2) Therapeutic Knowledge Graph Generation: To obtain the similarities between drugs/nodes, we first generated a therapeutic knowledge graph through the biomedical predications available in the SemMedDB knowledge base in the form of $(entity_i, relation_k, entity_j)$. Since there is a plethora of irrelevant predications potentially causing noises and the running-time complexity concerns, we only obtain the predications containing specific relations, including *"treats"*, *"disrupts"*,

"prevents", *"augments"*, *"stimulates"*, and *"co-exists with"* to build individual knowledge graphs. A publicly available graph library in Python language, NetworkX [16], is used to build the graph structures and compute the simrank similarity scores of the drugs.

In the initial dataset, there were 846 unique conditions. Since the principal scope of this study is limited to psychological disorders, domain experts were consulted to review all the distinct conditions in the dataset and identify the psychological conditions. In the end, the experts listed 35 different psychological diseases. These conditions are listed as shown in Appendix A. After the expert inspection, there were 50,709 samples comprised of 35 unique psychological conditions, 377 unique drug names, and 918 distinct drug-condition pairs.

To match the drug names with the corresponding node entities in the knowledge graph through the Concept Unique Identifiers (CUIs), we utilized the UMLS API to obtain the top three CUIs related to each drug. In addition, each drug-CUI combination was represented as a distinct row in the dataset not to miss any potential similarities.

3) Simrank Similarity: SimRank similarity is a metric to measure the similarity between two nodes in a graph structure. The similarity between two nodes is computed as the similarity of their neighbors connected recursively. Here, the fundamental assumption is that two nodes are similar if they are connected to identical neighbor nodes [1]. For instance, the similarity score between two nodes, u and v, is computed by the formula in Equation 2:

$$s(u,v) = \frac{C}{|I(u)||I(v)|} \Sigma_{i=1}^{|I(u)|} \Sigma_{j=1}^{|I(v)|} s(I_i(u), I_j(v)) \qquad (2)$$

where C is a decay factor, $I(u)$ represents the in-neighbors set of node u, and $|I(u)|$ indicates the number of nodes linked to node u. The rules for calculating the similarity score are following:

- if nodes u and v are identical, $s(I_i(u), I_j(v)) = 1$,
- if either u or v has no neighbor, $s(I_i(u), I_j(v)) = 0$,
- All other cases, $s(I_i(u), I_j(v))$ is computed iteratively based on the given formula in the Equation 2.

The NetworkX library's "`simrank_similarity`" function was operated to compute the simrank similarity score for the drugs in the dataset. However, this method uses a recursive approach and has a memory complexity of $O(n^2)$, making it impractical to use on a process that needs to calculate comparisons where n indicates the number of nodes. A new method was adopted that uses matrix manipulation to overcome this limitation. A separate graph was created for each predicate type, and the Simrank similarity score was calculated for each node concerning the remaining nodes in the graph structure. The calculated scores were then used to select candidate-relevant drug combinations; however, some of the CUIs in the dataset were not present in the graph, and this situation made it impossible to calculate the Simrank similarity score for some drug combinations.

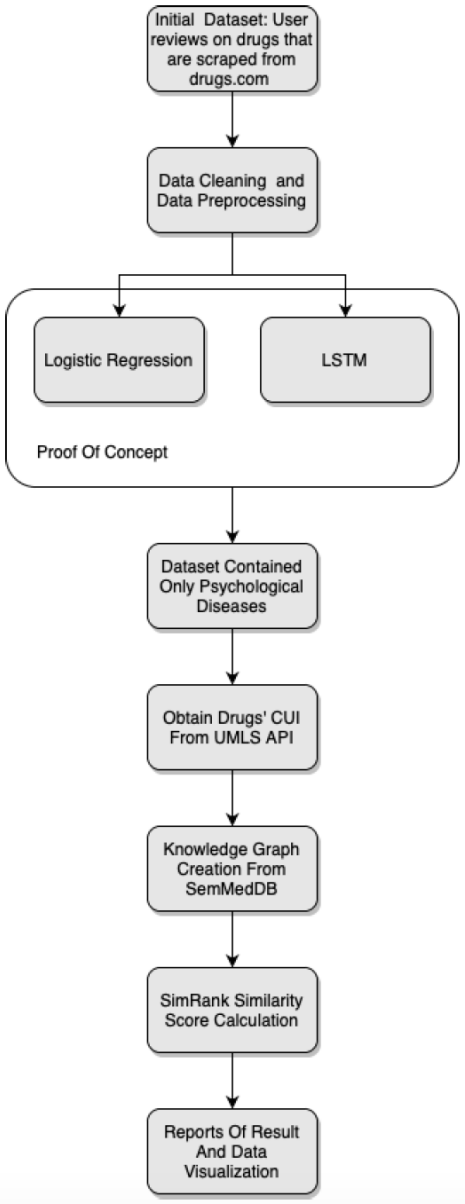


Fig. 3: Overall execution flow of the proposed approach

V. Results and Evaluations

Multiple graphs were created by connecting biomedical entities using potentially informative relations (predicates) related to the treatment concept. These graphs were built using the predications (triples) obtained from these relations and were used to identify similar drugs through simrank similarity scores. However, similarity scores were not notable in the graphs constructed by *"treats"*, *"disrupts"*, *"prevents"*, and *"augments"* relations. Therefore, we only reported the similarity results from the graphs constructed by ***"stimulates"***, and ***"co-exists with"*** relations, where we observed promising drug similarities for further manual investigations.

To construct the stimulative graph using the stimulates predicate, we used 46,576 SemMedDB predications where

7,544 distinct nodes were available. Similarly, we employed 99,626 predications using the co-exists predicate to create the coexistence graph, where 15,982 unique nodes were present.

By running the simrank similarity algorithm over the stimulative graph, the 86 unique drug names with CUI info samples were obtained, illustrating the relationships between these distinct samples in an 86×86 heatmap similarity matrix as shown in Figure 4.

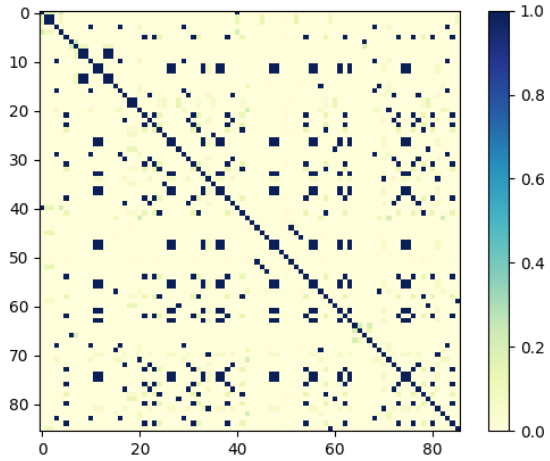


Fig. 4: Similarity heatmap matrix for *stimulates* relation

Based on the results obtained, we have multiple candidate pairs having non-zero similarity values implying potential similarities in the cells. Similar to the stimulative graph, we also executed the simrank algorithm on the coexistence graph, and 144 distinct drug names with their CUIs were collected, signifying the associations between drug entities in a 144×144 heatmap similarity matrix as visualized in Figure 5.

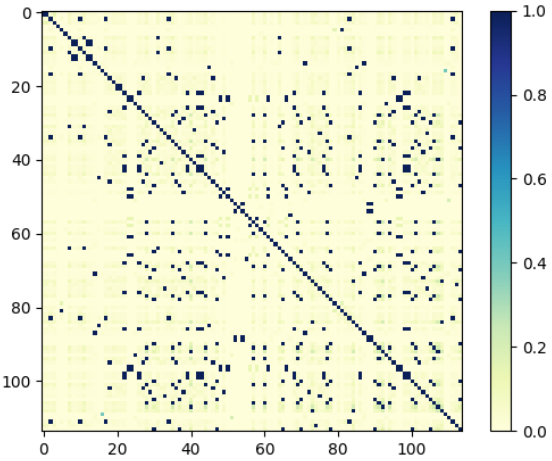


Fig. 5: Similarity heatmap matrix for *co-exists with* relation

As can be seen from the heatmap illustrations, the associations between drugs through stimulates and coexists_with relations can be discovered by the simrank similarity algorithm running over the biomedical knowledge graphs. Since this work is an initial experimental study exploiting a basic intuitive idea, the principal aim is to prove that even a simple approach can discover potentially-plausible candidate drug pairs. To validate the potential power of the proposed work, we shared the discovered plausible pairs in Table I.

Relation Type	Drug-1	Drug-2	Relevant Publications
STIMULATES	Yohimbine	Desipramine	- *Effects of desipramine and yohimbine on alpha 2- and beta-adrenoreceptor sensitivity* - *Desipramine-yohimbine combination treatment of refractory depression. Implications for the beta-adrenergic receptor hypothesis of antidepressant action*
	Baclofen	Naltrexone	- *Alcohol dependence - neurobiology and treatment- A comparative study on the safety and efficacy of naltrexone versus baclofen versus acamprosate in the management of alcohol dependence* - *Baclofen and naltrexone effects on alcohol self-administration: Comparison of treatment initiated during abstinence or ongoing alcohol access in baboons*
	Phenytoin	Valproic Acid	- *Evaluating appropriateness of digoxin, carbamazepine, valproic acid, and phenytoin usage by therapeutic drug monitoring* - *Interaction between phenytoin and valproic acid: plasma protein binding and metabolic effects* - *Time-dependent interaction between phenytoin and valproic acid*
	Aripiprazole	Lithium	- *How to treat mania* - *Clinical picture and treatment of bipolar affective disorder in children and adolescents* - *Canadian Network for Mood and Anxiety Treatments (CANMAT) and International Society for Bipolar Disorders (ISBD) 2018 guidelines for the management of patients with bipolar disorder*
	Buprenorphine	Methadone	- *Monitoring prenatal exposure to buprenorphine and methadone* - *Analytical approaches for the determination of buprenorphine, methadone and their metabolites in Biological matrices* - *Perioperative management of patients on buprenorphine and methadone: A narrative review*
	Clonidine	Prazosin	- *Best practice guide for the treatment of nightmare disorder in adults* - *Effects of prazosin and clonidine on sympathetic and baroreflex function in patients with essential hypertension*
	Lamotrigine	Clozapine	- *The efficacy of lamotrigine in clozapine-resistant schizophrenia: a systematic review and meta-analysis* - *Lamotrigine and clozapine for bipolar disorder* - *Rapid-onset agranulocytosis in a patient treated with clozapine and lamotrigine* - *Lamotrigine and Clozapine for Bipolar Disorder* - *Effects of clozapine plus lamotrigine on phencyclidine-induced hyperactivity*
	Propranolol	Prazosin	- *Effects of propranolol and atenolol on immobilization stress-induced hypertension and down-regulation of central beta-adrenoceptors in rats* - *Emerging treatments for PTSD* - *Evaluation of the efficacy of prazosin versus propranolol as initial antihypertensive therapy*
COEXISTS_WITH	Clozapine	Quetiapine	- *Clozapine, quetiapine and olanzapine among addicted schizophrenic patients: towards testable hypotheses* - *How to treat mania* - *A meta-analysis of neuropsychological change to clozapine, olanzapine, quetiapine, and risperidone in schizophrenia* - *Quetiapine and clozapine in parkinsonian patients with dopaminergic psychosis*
	Clozapine	Chlorpromazine	- *How to treat mania* - *Differing Prevalence and Correlates of Metabolic Syndromes Between Chlorpromazine and Clozapine: A 10-year Retrospective Study of a Male Chinese Cohort* - *The efficacy and safety of clozapine versus chlorpromazine in geriatric schizophrenia*
	Naltrexone	Buprenorphine / naloxone	- *Medication Treatment of Opioid Use Disorder* - *The Opioid Epidemic: Crisis and Solutions* - *Anxiety, Depression, and Insomnia Among Adults With Opioid Dependence Treated With Extended-Release Naltrexone vs Buprenorphine-Naloxone: A Randomized Clinical Trial and Follow-up Study* - *Comparative effectiveness of extended-release naltrexone versus buprenorphine-naloxone for opioid relapse prevention (X-BOT): a multicentre, open-label, randomised controlled trial*
	Clozapine	Olanzapine	- *Clozapine, quetiapine and olanzapine among addicted schizophrenic patients: towards testable hypotheses* - *The Importance of Conduct Disorder in the Treatment of Violence in Schizophrenia: Efficacy of Clozapine Compared With Olanzapine and Haloperidol* - *The effects of clozapine, risperidone, and olanzapine on cognitive function in schizophrenia* - *Clozapine and olanzapine in the treatment of the psychotic disorders in Parkinson's disease*
	Haloperidol	Chlorpromazine	- *Effects of clozapine, chlorpromazine and haloperidol on schedule-controlled behavior* - *Haloperidol versus chlorpromazine for schizophrenia* - *Effect of melperone, chlorpromazine, haloperidol, and diazepam on experimental anxiety in normal subjects* - *Haloperidol, clopenthixol, and chlorpromazine in chronic schizophrenia*
	Methadone	Clonidine	- *Methadone, Buprenorphine, and Clonidine Attenuate Mitragynine Withdrawal in Rats* - *Opiate addiction and the locus coeruleus. The clinical utility of clonidine, naltrexone, methadone, and buprenorphine* - *Efficacy of clonidine and of methadone in the rapid detoxification of patients dependent on heroin* - *Craving and drug reward: a comparison of methadone and clonidine in detoxifying opiate addicts*
	Carbamazepine	Valproic Acid	- *Carbamazepine Verses Valproic Acid as Monotherapy in Epileptic Patients* - *The newer antiepileptic drugs: carbamazepine and valproic acid* - *Use of lithium, carbamazepine, and valproic acid in a state-operated psychiatric hospital*
	Clozapine	Haloperidol	- *Effectiveness of clozapine, haloperidol and chlorpromazine in schizophrenia during a five-year period* - *The Importance of Conduct Disorder in the Treatment of Violence in Schizophrenia: Efficacy of Clozapine Compared With Olanzapine and Haloperidol* - *Effects of haloperidol and clozapine on synapse-related gene expression in specific brain regions of male rats* - *Clozapine and haloperidol in moderately refractory schizophrenia: a 6-month randomized and double-blind comparison*

TABLE I: Manual investigation results

Considering the candidate pairs, our domain experts, who are in the psychology field, manually inspected the result pairs yielded by the model. Based on the manual inspection, we shared the medical research articles published in the relevant discipline as supportive evidence.

VI. Conclusion & Future Directions

In this work, we demonstrated that the drug comments reflecting personal experiences on diseases could be a valuable resource for computational drug repositioning efforts. First, we employed sentiment analysis experiments to prove that personal comments are consistent. In the sentiment analysis work, we obtained decently high F1 scores relying on between 0.70 and 0.85 by utilizing logistic regression and LSTM models. Then, we built multiple knowledge graphs comprising essential biomedical relations in the SemMedDB and entities taken from the drug-disease pair entities in the comment

dataset. Following, we ran the simrank similarity algorithm on the knowledge graphs and identified the most similar drugs to each other in terms of connectivity context in the graph. Plus, we practiced manual inspections on the promising candidate drug pairs by domain experts and shared the inspection results supported by research articles. Consequently, we showed that similarity metrics running on graphs also could be a powerful technique for pre-filtering purposes in drug repositioning experiments. As a future research direction, we plan to apply a custom translating embeddings approach using the same predications we used in the graphs to filter the candidate pairs further before the clinical experiments.

ACKNOWLEDGMENT

We are grateful to Google Cloud Services for providing us with academic credit support to use conduct this research. We also thank our domain experts Ayşe Gökçen Gündoğdu and İdil Kalay for their manual investigation and interpretation efforts. Besides, this study is funded by TUBITAK 2209-A Research Project Support Program for Undergraduate Students and partially supported by TUBITAK 3501 Career Development Program through grant 122E103.

REFERENCES

[1] G. Jeh and J. Widom, "Simrank: a measure of structural-context similarity," in *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2002, pp. 538–543.

[2] J. Li, S. Zheng, B. Chen, A. J. Butte, S. J. Swamidass, and Z. Lu, "A survey of current trends in computational drug repositioning," *Briefings in bioinformatics*, vol. 17, no. 1, pp. 2–12, 2016.

[3] G. Bakal, H. Kilicoglu, and R. Kavuluru, "Non-negative matrix factorization for drug repositioning: experiments with the repodb dataset," in *AMIA Annual Symposium Proceedings*, vol. 2019. American Medical Informatics Association, 2019, p. 238.

[4] T. N. Jarada, J. G. Rokne, and R. Alhajj, "A review of computational drug repositioning: strategies, approaches, opportunities, challenges, and directions," *Journal of cheminformatics*, vol. 12, no. 1, pp. 1–23, 2020.

[5] J. Li and Z. Lu, "A new method for computational drug repositioning using drug pairwise similarity," in *2012 IEEE international conference on bioinformatics and biomedicine*. IEEE, 2012, pp. 1–4.

[6] Z. Gao, P. Ding, and R. Xu, "Kg-predict: a knowledge graph computational framework for drug repurposing," *Journal of biomedical informatics*, vol. 132, p. 104133, 2022.

[7] H. Kilicoglu, D. Shin, M. Fiszman, G. Rosemblat, and T. C. Rindflesch, "Semmeddb: a pubmed-scale repository of biomedical semantic predications," *Bioinformatics*, vol. 28, no. 23, pp. 3158–3160, 2012.

[8] H. Kilicoglu, G. Rosemblat, M. Fiszman, and D. Shin, "Broad-coverage biomedical relation extraction with semrep," *BMC bioinformatics*, vol. 21, pp. 1–28, 2020.

[9] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams engineering journal*, vol. 5, no. 4, pp. 1093–1113, 2014.

[10] F. Gräßer, S. Kallumadi, H. Malberg, and S. Zaunseder, "Aspect-based sentiment analysis of drug reviews applying cross-domain and cross-data learning," in *Proceedings of the 2018 international conference on digital health*, 2018, pp. 121–125.

[11] J. C. Stoltzfus, "Logistic regression: a brief primer," *Academic emergency medicine*, vol. 18, no. 10, pp. 1099–1104, 2011.

[12] Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: Lstm cells and network architectures," *Neural computation*, vol. 31, no. 7, pp. 1235–1270, 2019.

[13] M.-T. Luong and C. D. Manning, "Stanford neural machine translation systems for spoken language domains," in *Proceedings of the 12th International Workshop on Spoken Language Translation: Evaluation Campaign*, 2015, pp. 76–79.

[14] Y. Wang, M. Huang, X. Zhu, and L. Zhao, "Attention-based lstm for aspect-level sentiment classification," in *Proceedings of the 2016 conference on empirical methods in natural language processing*, 2016, pp. 606–615.

[15] S. Song, H. Huang, and T. Ruan, "Abstractive text summarization using lstm-cnn based deep learning," *Multimedia Tools and Applications*, vol. 78, pp. 857–875, 2019.

[16] A. Hagberg, P. Swart, and D. S Chult, "Exploring network structure, dynamics, and function using networkx," Los Alamos National Lab.(LANL), Los Alamos, NM (United States), Tech. Rep., 2008.

APPENDIX A
PSYCHOLOGICAL CONDITIONS LIST

"Depression", "Anxiety", "Bipolar Disorder", "Insomnia", "ADHD", "Anxiety and Stress", "Major Depressive Disorder", "Panic Disorder", "Generalized Anxiety Disorder", "Opiate Dependence", "Schizophrenia", "Obsessive-Compulsive Disorder", "Schizoaffective Disorder", "Social Anxiety Disorder", "Alcohol Dependence", "Opiate Withdrawal", "Post Traumatic Stress Disorder", "Premenstrual Dysphoric Disorder", "Alcohol Withdrawal", "Mance Anxiety", "Borderline Personality Disorder", "Autism", "Alzheimer's Disease", "Chronic Fatigue Syndrome", "Postpartum Depression", "Anorexia", "Binge Eating Disorder", "Sexual Dysfunction, "SSRI Induced", "Hypersomnia", "Tourette's Syndrome", "Shift Work Sleep Disorder", "Bulimia", "Mania", "Paranoid Disorder", "Agitated State".