DETECTING SIDE EFFECTS OF ADVERSE DRUG REACTIONS THROUGH DRUG-DRUG INTERACTIONS USING GRAPH NEURAL NETWORKS AND SELF-SUPERVISED LEARNING

G. KEERTHY[1], K. RADHIKA[2], A. SWATHI[3], D. JAMUNA[4]

PG SCHOLAR[1], ASSOCIATE PROFESSOR[2], ASSISTANT PROFESSOR[3], ASSOCIATE PROFESSOR[4],

JAYAPRAKASH NARAYAN COLLEGE OF ENGINEERING, DHARMAPUR, MAHABUBNAGAR

ABSTRACT

Adverse Drug Reactions(ADRs) due to drug-drug interactions present a public health problem worldwide that deserves attention due to its impact on mortality, morbidity, and healthcare costs. There have been major challenges in healthcare with the ever-increasing complexity of therapeutics and an aging population in many regions. At present, no standard method to detect such adverse drug reactions exists until otherwise reported by patients after the drug is released to the market. Further, several studies show that it is extremely challenging to detect these rare cases during clinical trials held before the drug is released. Therefore, a reliable and efficient technique to predict such side effects before the release of the drug to the market is the need of the hour. Through the power of Graph Neural Networks and the knowledge representation abilities of self-supervised learning, we designed an effective framework to model drug-drug interactions by leveraging the spatial and physical properties of drugs by representing them as molecular graphs. Through this approach, we developed a technique that resembles the dynamics of a chemical interaction. On training and testing this approach on the TwoSIDES Polypharmacy Dataset by Therapeutic Data Commons(TDC), we achieve a state of the art results by obtaining a precision of 75% and an accuracy of 90% on the test dataset. Further, we also perform a case study on the DrugBank dataset and compare our results on the interaction type prediction task to validate our approach on the drug-drug interaction domain and achieve excellent results with precision, F1, and accuracy of 99%. Our study and experimental approaches lay the groundwork for further research on side-effect prediction through drug-drug interaction and the use of Graph Neural Networks in the field of Molecular Biology.

KEYWORDS: Adverse drug reaction, drug-drug interaction, side effect prediction, graph neural network, self-supervised learning, scientific machine learning

INTRODUCTION

Computational methods hold great promise for mitigating the health and financial risks of drug development by predicting possible side effects before entering into the clinical trials. Several learning based methods have been proposed for predicting the side effects of drugs based on various features such as: chemical structures of drugs [1, 8, 9, 2, 5], drug-protein interactions [8, 2, 15,] protein-protein interactions (PPI) [8, 9], activity in metabolic networks pathways, phenotype information and gene annotations [8]. In parallel to the above-mentioned approaches, recently, deep learning models have been employed to predict side effects: (i) uses biological, chemical and semantic information on drugs in addition to clinical notes and case reports and (ii) [4] uses various chemical fingerprints extracted using deep architectures to compare the side effect prediction performance. While these methods have proven useful for predicting adverse drug reactions (ADRs – used\ interchangeably with drug side effects), the features they use are solely based on external knowledge about the drugs (i.e., drug-protein interactions, etc.) and are not cell or condition (i.e., dosage) specific. To address this issue ,Wang et al. (2016) utilize the data from the LINCS L1000 project. This project profiles gene expression changes in numerous human cell lines after treating them with a large number of drugs and small-molecule compounds. By using the gene expression profiles of the treated cells, provides the first comprehensive, unbiased, and cost-effective prediction of ADRs. The paper formulates the problem as a multi-label classification task. Their results suggest that the gene expression profiles provide context-dependent information for the side-effect prediction task. While the LINCS dataset contains a total of 473,647 experiments for 20,338 compounds, their method utilizes only the highest quality experiment for each drug to minimize noise. This means that most of the expression data are left unused, suggesting a potential room for improvement in the prediction performance. Moreover, their framework performs feature engineering by transforming gene expression features to enrichment vectors of biological terms. In this work, we investigate whether the incorporation of gene expression data along with the drug structure data can be leveraged better in a deep learning framework without the need for feature engineering. In this study, we propose a deep learning framework, Deep Side, for ADR prediction. Deep Side uses only (i) in vitro gene expression profiling experiments (GEX) and their experimental meta data (i.e., cell line and dosage - META), and (ii) the chemical structure of the compounds (CS). Our models train on the full LINCS L1000 dataset and use the SIDER dataset as the ground truth for drug - ADR pair labels [13]. We experiment with five architectures: (i) a multi-layer perceptron (MLP), (ii) MLP with residual connections (Res MLP), (iii) multi-modal neural net- works (MMNN. Concat and MMNN.

Sum), (iv) multi-task neural network (MTNN), and finally, (v) SMILES convolutional neural network (SMILES Conv). We present an extensive evaluation of the above-mentioned architectures and investigate the contribution of different features. Our experiments show that CS is a robust predictor of side effects. The base MLP model, which uses CS features as input, produces _11% macro-AUC and _2% micro- AUC improvement over the state-of-the-art results provided in [32], which uses both GEX (high quality) and CS features. The multi-modal neural network model, which uses CS, GEX and META features and uses summation in the fusion layer (MMNN. Sum) achieves 0:79 macro-AUC and 0:877 micro-AUC which is the best result among MLP based approaches. We also find out that when the chemical structure features are fully utilized in a complex model like ours, it overpowers the information that is obtained from the GEX dataset. The convolutional neural network that only uses the SMILES string representation of the drug structures achieves the best result among all the proposed architectures with provides 13:0% macro-AUC and 3:1% micro-AUC improvement over the state-of-the-art algorithm. Finally, inspecting the confident false positives predictions reveal side effects that are not reported in the ground truth dataset, but are indeed reported in the literature.

LITERATURE REVIEW

1. X. Zhu, J. Liu, J. Zhang, Z. Yang, F. Yang, and X. Zhang, ‘‘FingerDTA: A fingerprint-embedding framework for drug-target binding affinity prediction,’’ Big Data Mining Analytics, vol. 6, no. 1, pp. 1–10, Mar. 2023.

It introduces a novel framework, FingerDTA, for predicting drug-target binding affinities. The challenge of accurately predicting the interaction between small molecules (drugs) and target proteins is crucial in drug discovery, as it helps identify potential drug candidates with high efficacy. Current methods often struggle with the complexity of molecular structures and their interactions. To address this, FingerDTA employs a fingerprint-based approach combined with deep learning techniques. The framework uses molecular fingerprints to represent drugs and proteins, capturing key features that are essential for predicting binding affinity. It then embeds these fingerprints into a neural network model to assess the likelihood of successful binding. The authors demonstrate that their model outperforms existing methods in terms of prediction accuracy, showcasing the power of fingerprint embeddings in handling the intricate task of drug-target interaction prediction. Through extensive experiments on benchmark datasets, FingerDTA is shown to achieve high performance, with the potential to accelerate drug discovery processes by providing reliable predictions for drug-target binding affinity. The work emphasizes the importance of leveraging advanced machine learning methods to address challenges in computational biology, particularly in the context of precision medicine and the identification of novel therapeutic agents. The study highlights FingerDTA's effectiveness in improving the prediction of drug-target interactions, thus advancing the field of drug discovery.

2. J. Zhang and M. Xie, ‘‘NNDSVD-GRMF: A graph dual regularization matrix factorization method using non-negative initialization for predicting drug-target interactions,’’ IEEE Access, vol. 10, pp. 91235–91244, 2022.

It proposes a novel approach to predict drug-target interactions (DTIs) by leveraging matrix factorization techniques combined with graph dual regularization. Accurate prediction of DTIs is critical for advancing drug discovery and understanding the molecular mechanisms behind drug efficacy and safety. Traditional methods for predicting DTIs often suffer from issues such as sparse data and limited interaction information. To address these challenges, the authors introduce the NNDSVD-GRMF method, which incorporates a non-negative matrix factorization (NMF) model initialized with singular value decomposition (SVD) to ensure non-negativity and enhance interpretability. The core innovation of the NNDSVD-GRMF method lies in the application of dual regularization, which combines both graph-based regularization and matrix factorization to better capture the underlying relationships between drugs and targets. This dual regularization improves the model’s ability to incorporate both structural and interaction data, thus enhancing the prediction accuracy. The method is tested on several benchmark datasets and demonstrates superior performance compared to existing DTI prediction methods, showing its potential to predict previously unknown drug-target interactions. Through extensive evaluation, the authors confirm that NNDSVD-GRMF provides a promising solution for the DTI prediction problem, offering improved accuracy and reliability in drug discovery applications. This approach highlights the effectiveness of combining matrix factorization with graph regularization techniques for solving complex biological prediction problems.

3. K. KC, R. Li, F. Cui, and A. R. Haake, ‘‘Predicting biomedical interactions with higher-order graph convolutional networks,’’ IEEE/ACM Trans. Comput. Biol. Bioinf., vol. 19, no. 2, pp. 676–687, Mar. 2022.

It introduces an innovative method for predicting biomedical interactions using higher-order graph convolutional networks (GCNs). Biomedical interaction prediction, particularly in the context of drug discovery and disease understanding, remains a challenging problem due to the complex and interconnected nature of biological systems. Traditional methods, including first-

order GCNs, often fail to capture the rich relational information between nodes in a graph, such as drug-target interactions or protein-protein interactions, which are essential for accurate prediction. To overcome this limitation, the authors propose the use of higher-order graph convolutions, which extend the traditional GCN model by incorporating not only direct neighbors of a node but also indirect connections through multiple hops in the graph. This approach enables the model to learn more complex, multi-level dependencies among nodes, thus improving prediction accuracy. The proposed method is validated on several biomedical datasets, demonstrating superior performance over existing techniques that rely on first-order GCNs or other conventional methods. By capturing higher-order dependencies, the model achieves better generalization and can uncover subtle relationships between biological entities. The results suggest that higher-order GCNs offer significant potential in predicting biomedical interactions, advancing the field of computational biology, and contributing to drug discovery, disease modeling, and personalized medicine. This work emphasizes the importance of utilizing advanced graph-based models to handle the complexity of biomedical networks and interactions.

4. M. Chen, Y. Jiang, X. Lei, Y. Pan, C. Ji, and W. Jiang, ‘‘Drugtarget interactions prediction based on signed heterogeneous graph neural networks,’’ Chin. J. Electron., vol. 33, no. 1, pp. 231–244, Jan. 2024.

It presents a novel approach for predicting drug-target interactions (DTIs) using signed heterogeneous graph neural networks (HGNNs). Accurate DTI prediction is a critical task in drug discovery and personalized medicine, as understanding the interactions between drugs and their target proteins can help identify potential therapeutic agents and improve drug efficacy. The challenge lies in the complexity and heterogeneity of biological networks, where drug-target relationships are often influenced by various factors such as the type of drug, target protein, and interaction context. To address this, the authors propose a signed heterogeneous graph model that integrates multiple types of biological entities (e.g., drugs, targets, and their relationships) into a unified graph. The signed nature of the graph captures both positive and negative interactions, providing a more nuanced representation of the biological network. The method utilizes graph neural networks to effectively model these interactions, learning to predict DTIs by propagating information through the graph structure. The model is evaluated on multiple benchmark datasets, showing significant improvements in prediction accuracy compared to traditional DTI prediction methods. The study demonstrates the effectiveness of signed heterogeneous graph neural networks in capturing complex, multi-relational interactions, highlighting their potential to accelerate drug discovery processes. This work contributes to the growing body of research exploring advanced graph-based methods for biomedical applications and provides new insights into the prediction of drug-target interactions.

5. F. Castiglione, C. Nardini, E. Onofri, M. Pedicini, and P. Tieri, ‘‘Explainable drug repurposing approach from biased random walks,’’ IEEE/ACM Trans. Comput. Biol. Bioinf., vol. 20, no. 2, pp. 1009–1019, Mar. 2023.

It introduces a novel method for drug repurposing using explainable biased random walks. Drug repurposing, which involves identifying new therapeutic uses for existing drugs, is a promising strategy to expedite drug discovery and reduce development costs. However, one of the challenges in this area is the ability to explain why certain drugs may be effective against diseases for which they were not originally intended. To address this, the authors propose an approach based on biased random walks over biomedical networks that integrate drug, protein, and disease information. By performing random walks with a bias towards nodes that share similar characteristics, the model identifies potential drug-disease relationships in a more targeted way. The approach is explainable, allowing researchers to trace the steps taken by the model to make predictions, thus offering transparency and interpretability in the repurposing process. The proposed method is validated on several benchmark datasets, demonstrating its ability to predict plausible drug-disease associations while providing valuable insights into the reasoning behind these predictions. The results show that the approach outperforms traditional drug repurposing methods, offering a more efficient and explainable solution. This work highlights the potential of biased random walks in uncovering hidden relationships within complex biological networks and offers a practical framework for advancing drug repurposing efforts in computational biology and medicine.

EXISTING SYSTEM

The presence of Artificial Intelligence(AI) is apparent throughout all realms of science. Nevertheless, AI in chemistry has come up to be one of the most researched areas. Reference [14] successfully modeled QSAR which predicts the carcinogenic potency of aromatic amines and an FDA/OTR MultiCASE model predicting the carcinogenicity of pharmaceuticals. Similarly, its worth noting that [15] employed 2D similarity fingerprints for its efficiency and simplicity of computation for drug-drug interactions to prevent adverse effects.

- The advent of Machine Learning(ML) and Neural Networks(NN) has raised interest in drug discovery. Reference [16] uses genomic features from the cell lines and chemical information from drugs and shows

that it is possible to build in-silico-based multi-drug models to impute missing IC50 values with non-parametric machine learning algorithms. Reference [17] calculates drug-drug pair similarities using four features namely: drug phenotypic, therapeutic, chemical, and genomic properties with applied predictive models namely Naive Bayes, Decision Tree, k-nearest neighbor, logistic regression, and support vector machine(SVM). Reference [18] discusses a two stage hybrid approach model that identifies the positive instances using a feature based binary classifier, and then a Long Short Term Memory (LSTM) based classifier to classify positive instances while [19] uses Discriminative Vector Machines for accurate prediction of protein-protein interactions.

- The Graph Neural Network(GNN) Model [20] approach is known to perform well in chemical graphs inclusive of but not limited to analyzing molecular structures, protein-protein interaction networks, and drug discovery. Graph Attention Network (GAT) [10] has been applied in various tasks and applications such as [21] and [22]. Reference [23] uses the Graph Convolution operation which allows the nodes to aggregate information from their immediate neighbors in the graph. Further, [24] makes use of spectral graph convolutions that leverage the eigen values and eigenvectors of the graph Laplacian matrix to define convolutional operations, enabling the propagation of information through the graph and [25] discussed the formulation of spectral graph convolution networks for directed graphs. GraphSAGE [11] was used in [26] and [27] and has proved to make a significant improvement in performance in transductive-based prediction tasks for large and elaborate graph networks.

Disadvantages

- The complexity of data: Most of the existing machine learning models must be able to accurately interpret large and complex datasets for Detecting Side Effects of Adverse Drug Reactions.
- Data availability: Most machine learning models require large amounts of data to create accurate predictions. If data is unavailable in sufficient quantities, then model accuracy may suffer.
- Incorrect labeling: The existing machine learning models are only as accurate as the data trained using the input dataset. If the data has been incorrectly labeled, the model cannot make accurate predictions.

PROPOSED SYSTEM

• The study aims to reduce the number of clinical hospitalizations due to Adverse Drug Effects. Our proposed framework predicts the side effects of drug-drug interactions based on the molecular structure of the drug and their interaction dynamics. This enables a holistic and early determination of side effects caused by such adverse reactions and ensures that variables such as quality of the physician, subjectivity of the reports, and quality of healthcare are eliminated. Instead, a scientific and probabilistic view is provided to predict side effects. Our study is among the first to explore the underlying mathematical relations between drug-drug interaction and side effect prediction through deep learning.

• We present a novel approach to side effect prediction due to drug-drug interaction using Graph Neural Networks by representing the drugs involved in the reaction as molecules. The nodes of the graph are represented by atoms while the edges are represented by chemical bonds, thus making use of the spatial and physical properties of molecules by their graphical representation. Approaches that have been designed so far to predict side effects by drug-drug interaction neither use the structure of the chemicals involved nor leverage Graph Neural Networks to represent the physical and bonding properties of chemical compounds.

• To realize the dynamics of a chemical reaction and effectively leverage the spatial and physical features of both the reactants, We develop a Dual Input Graph Neural Network Hybrid Model with a 2-stage training phase. The Dual Input framework is designed to model the dynamics of a chemical reaction by sharing features between both reactants. We achieve a stable model training curve despite the complexity of the system. Further, each reactant is pre-trained to ensure a knowledge base derived from the properties of the reacting drugs is transferred to the side effect prediction task. Such a framework despite being only in the initial stages of research shows state-of-the-art results in precision and accuracy.

• The dataset that we used for our study is the Two Sides poly-pharmacy side effects dataset from Therapeutic Data Commons(TDC). Since this is among the only datasets for side-effect prediction through Drug-Drug interaction, our experiments act as a baseline benchmark. Therefore, we aim for our study to be a base for side effect prediction and aim to open doors to further research into the field of Molecular Graph Neural Networks and Drug-Drug Interactions.

• Since the TwoSides poly-pharmacy side effects dataset is the only available dataset for side effects prediction and owing to a lack of previous benchmarks on the same, we provide an

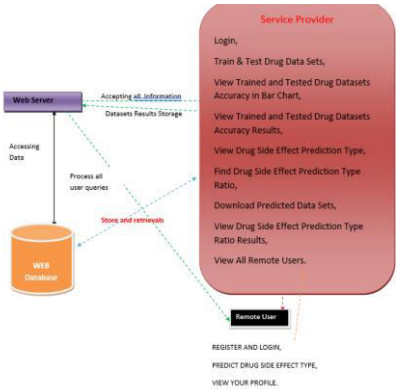
experimental validation of our proposed framework on the DrugBank Dataset. Though the final task for the DrugBank dataset is the prediction of the drug-drug interaction type, by comparing our results with related works, we achieve state-of-the-art results for our proposed framework on the drug-drug interaction type prediction task as well. We show that our proposed framework effectively models drug-drug interactions and despite the change in the prediction task, it can model the dynamics of a chemical interaction and lays the platform for further research on the Drug-Drug Interaction Domain and Bioinformatics.

Advantages

- We propose a novel framework to model drug-drug interactions using Graph Neural Networks by modeling the properties of drug molecules and leveraging the pattern recognition abilities of self-supervised learning and the potential of ensemble learning. Since chemical molecules can be represented as a molecular graph, we can leverage the complex representations and operations offered by Graph Neural Networks to model their atomic and bond-level properties.
- Recent developments through Graph Convolution Network, Graph Attention Network , and SAGE and their outstanding results on graph-level tasks have increased the popularity of Graph Neural Networks thus making it a reliable option in various fields. Despite some of the pitfalls of Graph Neural Networks such as under-fitting and the need to generate high-quality vertex embeddings, fine tuning pre-trained weights and ensemble approaches raise the performance of Deep Learning models by a great extent unlocking the door for further research in Graph Neural Networks, and their applications on drug research.

IMPLEMENTATION

SYSTEM ARCHITECTURE



MODULES

SERVICE PROVIDER

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Train & Test Drug Data Sets, View Trained and Tested Drug Datasets Accuracy in Bar Chart, View Trained and Tested Drug Datasets Accuracy Results, View Drug Side Effect Prediction Type, Find Drug Side Effect Prediction Type Ratio, Download Predicted Data Sets, View Drug Side Effect Prediction Type Ratio Results, View All Remote Users.

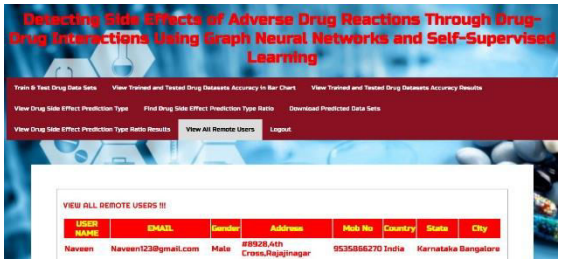
VIEW AND AUTHORIZE USERS

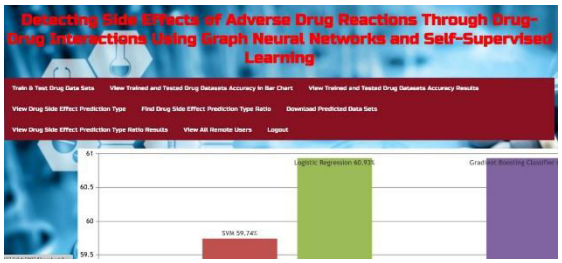
In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

REMOTE USER

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN,PREDICT DRUG SIDE EFFECT TYPE, VIEW YOUR PROFILE.

RESULTS





CONCLUSION

The pharmaceutical drug development process is a long and demanding process. Unforeseen ADRs that arise at the drug development process can suspend or restart the whole development pipeline. Therefore, the a priori prediction of the side effects of the drug at the design phase is critical. In our Deep Side framework, we use context-related (gene expression) features along with the chemical structure to predict ADRs to account for conditions such as dosing, time interval, and cell line. The proposed MMNN model uses GEX and CS as combined features and achieves better accuracy performance compared to the models that only use the chemical structure (CS) finger-prints. The reported accuracy is noteworthy considering that we are only trying to estimate the condition-independent side effects. Finally, SMILES Conv model outperforms all other approaches by applying convolution on SMILES representation of drug chemical structure.

REFERENCES

1. Atias, N., Sharan, R.: An algorithmic framework for predicting side e_ects of drugs. Journal of Computational Biology 18(3), 207{218 (2011)

2. Bresso, E., Grisoni, R., Marchetti, G., Karaboga, A.S., Souchet, M., Devignes, M.D., Sma_l-Tabbone, M.: Integrative relational machine-learning for understanding drug side-e_ect pro_les. BMC bioinformatics 14(1), 207 (2013)

3. Cai, M.C., Xu, Q., Pan, Y.J., Pan, W., Ji, N., Li, Y.B., Jin, H.J., Liu, K., Ji, Z.L.: Adrecs: an ontology database for aiding standardization and hierarchical classi_cation of adverse drug reaction terms. Nucleic acids research 43(D1), D907{D913 (2014)

4. Dey, S., Luo, H., Fokoue, A., Hu, J., Zhang, P.: Predicting adverse drug reactions through interpretable deep learning framework. BMC Bioinformatics 19 (12 2018). https://doi.org/10.1186/s12859-018-2544-0

5. Dimitri, G.M., Li_o, P.: Drugclust: A machine learning approach for drugs side effects prediction. Computational Biology and Chemistry 68, 204 { 210 (2017). https://doi.org/https://doi.org/10.1016/j.compbiolchem.2017.03.008

6. Groopman, J.E., Itri, L.M.: Chemotherapy-induced anemia in adults: incidence and treatment. Journal of the National Cancer Institute 91(19), 1616{1634 (1999)

7. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2016)

8. Huang, L.C., Wu, X., Chen, J.Y.: Predicting adverse side e_ects of drugs. BMC genomics 12(5), S11 (2011)

9. Huang, L.C., Wu, X., Chen, J.Y.: Predicting adverse drug reaction pro_les by integrating protein interaction networks with drug structures. Proteomics 13(2), 313{324 (2013)

10. Io_e, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167 (2015)

11. Kim, S., Thiessen, P.A., Bolton, E.E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B.A.,et al.: Pubchem substance and compound databases. Nucleic acids research 44(D1), D1202{D1213 (2015)

12. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classi_cation with deep convolutional neural networks. In: Advances in neural information processing systems. pp. 1097{1105 (2012)

13. Kuhn, M., Letunic, I., Jensen, L.J., Bork, P.: The sider database of drugs and side e_ects. Nucleic acids research 44(D1), D1075{D1079 (2015)

14. Landrum, G., et al.: Rdkit: Open-source cheminformatics (2006)

15. Lee, W., Huang, J., Chang, H., Lee, K., Lai, C.: Predicting drug side e_ects using data analytics and the integration of multiple data sources. IEEE Access 5, 20449{20462 (2017). https://doi.org/10.1109/ACCESS.2017.2755045