Molecular Fingerprints Are a Simple Yet Effective Solution to the Drug–Drug Interaction Problem

Yanan Long [1] **Horace Pan** [1] **Chao Zhang** [1] **Hy Truong Song** [1] **Risi Kondor** [1 *] **Andrey Rzhetsky** [1 *]

Abstract

Contemporary machine learning studies tackle the drug–drug interaction forecast problem by featurizing drugs using graph neural networks (GNNs). This automated featurization allows to avoid laborious handcrafting chemical features. We demonstrate here that a simple neural networks using Morgan fingerprints of the drugs outperformed these more complicated GNN models while spending only a small fraction of the time in training. Furthermore, to improve training, we curated and made available a novel dataset with negative drug–drug interaction examples derived from a very large electronic health records dataset.

1. Introduction

Drug–drug interactions (DDIs), or non-additive action of multiple co-administered medications, also known as polypharmacy problem, is a significant source of adverse medical outcomes (Zitnik et al., 2018). The major difficulty in identifying and anticipating DDIs is the combinatorial explosion of potential drug interactions that renders clinical testing of all pairs impractical. Machine learning models trained on expert-curated databases and electronic health records have since provided a tractable solution to identifying potential drug interactions.

Earlier work on DDI (Gottlieb et al., 2012; Vilar et al., 2012; 2014; Cheng & Zhao, 2014) primarily computed similarity measures between a combination of various handpicked chemical properties and structural fingerprints of drugs to predict new drug-drug interactions. One drawback to these DDI approaches is that they use fixed feature representations of the drugs that may not be optimal for predicting drug interactions.

More recent approaches rely on graph neural networks (GNNs) (Battaglia et al., 2018) automatically learning the molecular representation of the drugs to produce the interaction prediction. Zitnik et al. (2018) proposed one of the earlier approaches to use GNNs as encoders to featurize representation of molecules. Similar approaches for learning graph representations for the drugs was proposed by Yin et al. (2022) and Wang et al. (2022). They use a concatenation of handpicked chemical features and the learned features from applying graph neural network layers (Veličković et al., 2018; Kipf & Welling, 2017) layers on the molecular structure of the pair of drugs. Model suggested by Nyamabo et al. (2021) applies multiple graph attention layers on the individual drugs and aggregates their intermediate graph representations with co-attention to produce the output class predictions. Other works cast the problem as a knowledge graph learning problem Lin et al. (2020). The advantage of these GNN approaches is that they are able to learn feature representations for the drugs that are adapted to the classification task at hand without the need for explicit manual feature selection.

While the end-to-end GNN approach has proven to be effective on a variety of DDI benchmark tasks, we believe that fingerprint based approaches still have merit. In our experiments, we show that neural networks that operate on just 2D Morgan fingerprints outperform GNN models and achieve these improved results with simpler model architectures that were much faster than the GNN baselines. Moreover, a central concern about neural networks is that they are opaque and hard to interpret. The benefit to fingerprint based neural networks is that we retain the flexibility of a neural network that can learn useful nonlinear combinations of our features while still retaining a level of intepretability by virtue of using fingerprints as our input representations.

Our contributions in this work are two-fold:

1. We curated a new dataset that augments the DrugBank dataset with negative examples derived from a large commercial insurance dataset;

2. We demonstrated that the use of Morgan fingerprints with simple neural network architectures achieved SOTA performance, outperforming GNN architectures.

[*]Equal contribution [1]University of Chicago. Correspondence to: Yanan Long <ylong@uchicago.edu, yanan.long439@gmail.com>.

The 2022 ICML Workshop on Computational Biology.

2. Problem Setup

To formalize our problem, let D denote the set of drugs encoded by their DrugBank IDs. Our dataset is then a collection of triplets: $\left\{\left(d_i^{(1)}, d_i^{(2)}, l_i\right)\right\}_{i=1}^{n}$, where $l_i \in \{1, 2, \ldots, K\}$ denotes the interaction between the pair of drugs $\left(d_i^{(1)}, d_i^{(2)}\right) \in D \times D$. We treat the learning task as a multi-class classification over K classes. So our objective is to learn a classification function

$$f : D \times D \to \{1, 2, \ldots, K\}$$

where f has low cross entropy loss over the training set. In the DDI literature, we also commonly see the problem cast as a binary classification task over the triplets $\left(d_i^{(1)}, d_i^{(2)}, l_i)\right)$ where one predicts if the given label is in fact a true observed interaction between the pair of drugs. In both the multiclass classification setting and the binary classification setting, practictioners often augment the observed DDI data with negative examples to make the models more robust. This is because we care not just about the presence of interaction but also the absence thereof when screen drugs for "safe" combinations.

2.1. Dataset

The positive DDI examples came from the DrugBank dataset (Wishart et al., 2018), while the negative examples came from IBM MarketScan (IBM Watson Health, 2019), a commercial insurance database containing $>$ 150 million unique subjects. The prescription (RX) records from MarketScan contained information on the identity of the drugs encoded with National Drug Codes (NDCs), start date and end date. Using the dates we identified potential interactions as any overlap in the intervals of validity between any two drugs using an interval tree (Cormen et al., 2022, §17.3). Each row of data in this dataset corresponds to a pair of drugs, represented by their DrugBank IDs and the associated target value, which is a label indicating the type of the resulting interaction between the two drugs. The DrugBank IDs can be mapped to the corresponding SMILES strings, from which we constructed the Morgan fingerprints (FPs) (Morgan, 1965) of the drugs using RDKit (Landrum et al., 2006). For more information about the dataset, see Table 1.

# Drugs with SMILES	3,516
# DDI Pairs	1,457,198
# Interaction Types	256

Table 1. Dataset details

3. Model Architectures

We tested a variety of models on our new drug drug interaction dataset. Our models follow the same general architecture:

1. map the SMILES string to a numerical object e.g. a binary fingerprint or matrices containing atom features and connectivity information)
2. encode the pair of input drugs with an encoder network (e.g.: a multilayer perceptron (MLP) or a graph neural network (GNN))
3. construct a pair drug representation by concatenating or adding the two individual drug representations
4. feed the pair drug representation through a final multilayer perceptron to produce a class output.

We also experimented with variants of the above architecture where the two fingerprint vectors were aggregated before being fed through an MLP instead of after step 2 above.

3.1. Molecular Fingerprint Model

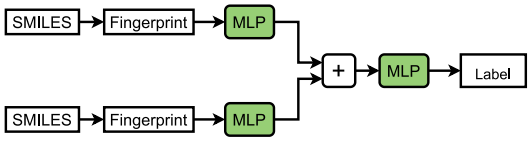


Figure 1. Fingerprint based neural network for predicting the interaction between two input drugs

We first generated the Morgan fingerprints, which is a kind of extended-connectivity fingerprint, from the SMILES strings with RDKit using a radius of 2 (Rogers & Hahn, 2010). The fingerprints are 2048 dimensional binary vectors which are passed through a multilayer perceptron (MLP). The resulting representation vectors are aggregated (either through concatenation or summation) to form a drug-pair representation that is sent through a final multilayer perceptron (MLP) to produce an output label. We applied the LeakyReLU (Zhang et al., 2017) activation function in the MLP layers.

3.2. Graph Neural Network Based Models

In the graph neural network based models, we extract atom features and atom connectivity information from the SMILES strings of the pair of drugs. The atom features, edge features and connectivity are then used as input to a graph neural network (GNN) applies multiple rounds of message passing (Gilmer et al., 2017) to produce a molecular encoding. As in the FP network, the drug encodings are aggregated and then passed through a final MLP to produce the predicted output label.

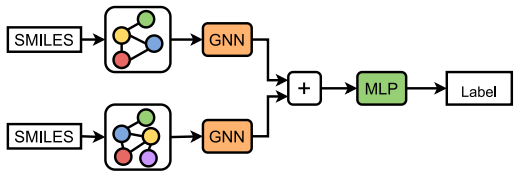


Figure 2. Graph neural network based architecture for predicting the interaction between two input drugs.

3.3. Attention-Based Models

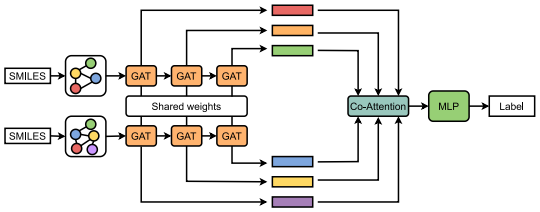


Figure 3. Co-attention architecture based off Nyamabo et al. (2021)'s model with intermediate graph attention layers for predicting the interaction between two input drugs. The number of GAT layers is a tunable hyperparameter.

We also compared our FP model against the SOTA co-attention based model SSI-DDI-v2 that mostly follows the architecture proposed in Nyamabo et al. (2021). This model takes the two molecular graphs with atom features from each drug in a pair as input and uses the Graph Attention Network (GAT) (Veličković et al., 2018; Brody et al., 2021) autoregressively to generate multiple layers of hidden representations for each drug separately, which are subsequently passed to an additive co-attention layer that outputs attention scores for reweighting the hidden representations (cf. Figure 3) during aggregation. The final aggregated hidden representation of the drug pair is then sent to an MLP to produce the predicted output label. Other models making use of self- and/or co-attention include Lin et al. (2022); Nyamabo et al. (2022); Pang et al. (2022).

4. Experiments

We evaluated each of the models on our dataset using $80/10/10$ training/validation/test split. Each of the networks produced a softmax probability distribution over all the DDI classes for predicting the interaction between drug pairs. We used the AdamW optimizer (Loshchilov & Hutter, 2017) to minimize the cross-entropy loss over each minibatch, with an early stopping $\Delta = 5 \times 10^{-3}$ and a tolerance of 4 epochs on validation loss. All models were implemented using Pytorch (Paszke et al., 2019) and Pytorch Geometric (Fey & Lenssen, 2019).

4.1. Hyperparameter Tuning

We manually tuned the hyperparameters of all models trained. Table 2 details the various hyperparameters we used and experimented with.

Parameter	Values
AdamW Learning Rate	$\{10^{-5}, 10^{-4}, 10^{-3}\}$
AdamW Weight Decay	$\{0\}$
MLP Layers	$\{3, 4, 5\}$
GNN Layers	$\{3, 4, 5\}$
Minibatch size	$\{128, 256, 512\}$

Table 2. Hyperparameter Values

5. Results and Discussion

Table 4 shows the results of our experiments. The FP model achieved higher performance on all of the reported metrics and converged in fewer training epochs with less time per epoch than the graph neural network baseline methods (Table 3). Thus, we now focus on more detailed comparison between the FP model and SSI-DDI-v2. Figure 4 shows the values of the metrics among classes that had at least 100 instances in the test dataset (56 classes), with decreasing number of instances per class. Overall, FP and SSI-DDI-v2 had similar performance profile *across* the classes, achieving full accuracy on class 226, corresponding to the interaction string

```
The risk or severity of
tendinopathy can be increased
when Drug1 is combined with
Drug2.
```

Figure 5 shows the difference in accuracy (5a) and weighted-F_1 score (5b) between FP (reference) and SSI-DDI-v2. The largest *positive* difference was found in class 193, corresponding to the interaction string

```
The risk or severity of
myopathy, rhabdomyolysis, and
myoglobinuria can be increased
when Drug1 is combined with
Drug2.
```

On the other hand, the largest *negative* difference was found in class 95, corresponding to the interaction string

```
The protein binding of Drug1 can
be decreased when combined with
Drug2.
```

In this work, we evaluated various neural network models on a novel DDI dataset. Our results showed that the appropriate model architecture built on Morgan Fingerprints outperformed graph neural network models for DDI prediction.

References

Battaglia, P. W., Hamrick, J. B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V., Malinowski, M., Tacchetti, A., Raposo, D., Santoro, A., Faulkner, R., et al. Relational inductive biases, deep learning, and graph networks. *arXiv preprint arXiv:1806.01261*, 2018.

Brody, S., Alon, U., and Yahav, E. How attentive are graph attention networks? *arXiv preprint arXiv:2105.14491*, 2021.

Cheng, F. and Zhao, Z. Machine learning-based prediction of drug–drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties. *Journal of the American Medical Informatics Association*, 21(e2): e278–e286, 2014.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., and Stein, C. *Introduction to Algorithms, 4th edition*. MIT press, 2022.

Fey, M. and Lenssen, J. E. Fast graph representation learning with PyTorch Geometric. In *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019.

Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., and Dahl, G. E. Neural message passing for quantum chemistry. In *International conference on machine learning*, pp. 1263–1272. PMLR, 2017.

Gottlieb, A., Stein, G. Y., Oron, Y., Ruppin, E., and Sharan, R. Indi: a computational framework for inferring drug interactions and their associated recommendations. *Molecular systems biology*, 8(1):592, 2012.

IBM Watson Health. IBM MarketScan research databases for life sciences researchers, 2019.

Kipf, T. N. and Welling, M. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations (ICLR)*, 2017.

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

Lin, S., Wang, Y., Zhang, L., Chu, Y., Liu, Y., Fang, Y., Jiang, M., Wang, Q., Zhao, B., Xiong, Y., et al. Mdf-sa-ddi: predicting drug–drug interaction events based on multi-source drug fusion, multi-source feature fusion and transformer self-attention mechanism. *Briefings in Bioinformatics*, 23(1):bbab421, 2022.

Lin, X., Quan, Z., Wang, Z.-J., Ma, T., and Zeng, X. Kgnn: Knowledge graph neural network for drug-drug interaction prediction. In *IJCAI*, volume 380, pp. 2739–2745, 2020.

Loshchilov, I. and Hutter, F. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

Morgan, H. L. The generation of a unique machine description for chemical structures-a technique developed at chemical abstracts service. *Journal of chemical documentation*, 5(2):107–113, 1965.

Nyamabo, A. K., Yu, H., and Shi, J.-Y. Ssi–ddi: substructure–substructure interactions for drug–drug interaction prediction. *Briefings in Bioinformatics*, 22(6): bbab133, 2021.

Nyamabo, A. K., Yu, H., Liu, Z., and Shi, J.-Y. Drug–drug interaction prediction with learnable size-adaptive molecular substructures. *Briefings in Bioinformatics*, 23 (1):bbab441, 2022.

Pang, S., Zhang, Y., Song, T., Zhang, X., Wang, X., and Rodriguez-Patón, A. Amde: a novel attention-mechanism-based multidimensional feature encoder for drug–drug interaction prediction. *Briefings in Bioinformatics*, 23(1):bbab545, 2022.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S. Pytorch: An imperative style, high-performance deep learning library. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems 32*, pp. 8024–8035. Curran Associates, Inc., 2019. URL http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance.pdf.

Rogers, D. and Hahn, M. Extended-connectivity fingerprints. *Journal of chemical information and modeling*, 50 (5):742–754, 2010.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., and Bengio, Y. Graph Attention Networks. *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=rJXMpikCZ. accepted as poster.

Vilar, S., Harpaz, R., Uriarte, E., Santana, L., Rabadan, R., and Friedman, C. Drug—drug interaction through molecular structure similarity analysis. *Journal of the American Medical Informatics Association*, 19(6):1066–1074, 2012.

Vilar, S., Uriarte, E., Santana, L., Lorberbaum, T., Hripcsak, G., Friedman, C., and Tatonetti, N. P. Similarity-based modeling in large-scale prediction of drug-drug interactions. *Nature protocols*, 9(9):2147–2163, 2014.

Wang, J., Liu, X., Shen, S., Deng, L., and Liu, H. Deepdds: deep graph neural network with attention mechanism to predict synergistic drug combinations. *Briefings in Bioinformatics*, 23(1):bbab390, 2022.

Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., et al. Drugbank 5.0: a major update to the drugbank database for 2018. *Nucleic acids research*, 46(D1): D1074–D1082, 2018.

Yin, Q., Cao, X., Fan, R., Liu, Q., Jiang, R., and Zeng, W. Deepdrug: A general graph-based deep learning framework for drug-drug interactions and drug-target interactions prediction. *bioRxiv*, 2022.

Zhang, X., Zou, Y., and Shi, W. Dilated convolution neural network with leakyrelu for environmental sound classification. In *2017 22nd international conference on digital signal processing (DSP)*, pp. 1–5. IEEE, 2017.

Zitnik, M., Agrawal, M., and Leskovec, J. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*, 34(13):i457–i466, 2018.

Additional Tables and Figures

Model	Training epochs	Avg. time/epoch (min)
FP	**15**	**1.37**
GCNConv (Kipf & Welling, 2017)	34	4.09
GATConv (Veličković et al., 2018)	27	5.00
GATv2Conv (Brody et al., 2021)	19	5.42
SSI-DDI (Nyamabo et al., 2021)	30	7.83
SSI-DDI-v2 (Nyamabo et al., 2021)	25	8.93

Table 3. Training information: number of epochs until convergence and time per epoch. SSI-DDI uses GATConv by default. SSI-DDI-v2 uses GATv2Conv instead. All GNN models used 4 layers.

Model	Accuracy	Macro-F_1	Weighted-F_1	AUROC
FP	**0.9615**	**0.9213**	**0.9612**	**0.9989**
GCNConv	0.8694	0.7144	0.8603	0.9793
GATConv	0.7958	0.6019	0.7763	0.9650
GATv2Conv	0.8696	0.7107	0.8622	0.9880
SSI-DDI	0.9422	0.8903	0.9415	0.9915
SSI-DDI-v2	0.9491	0.9047	0.9488	0.9923

Table 4. Test results on a hold-out set.

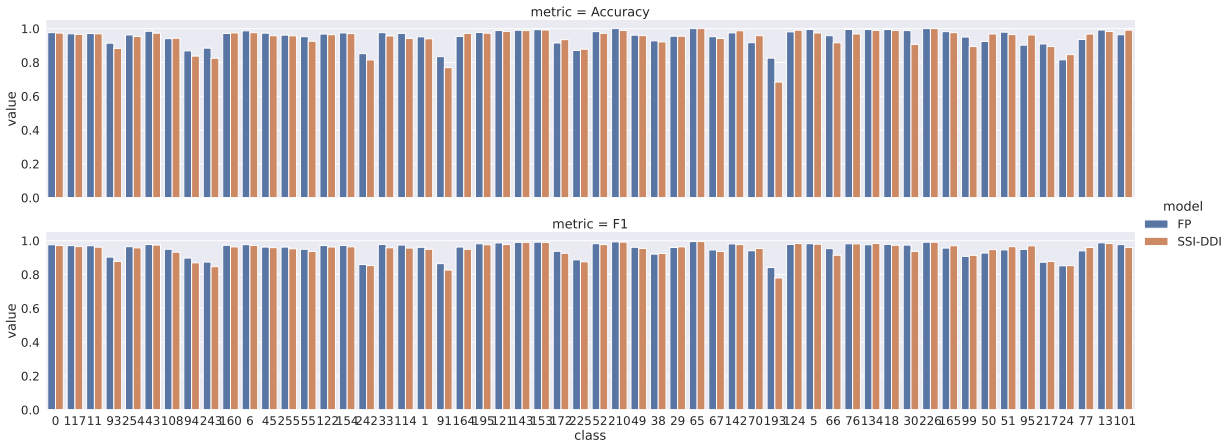
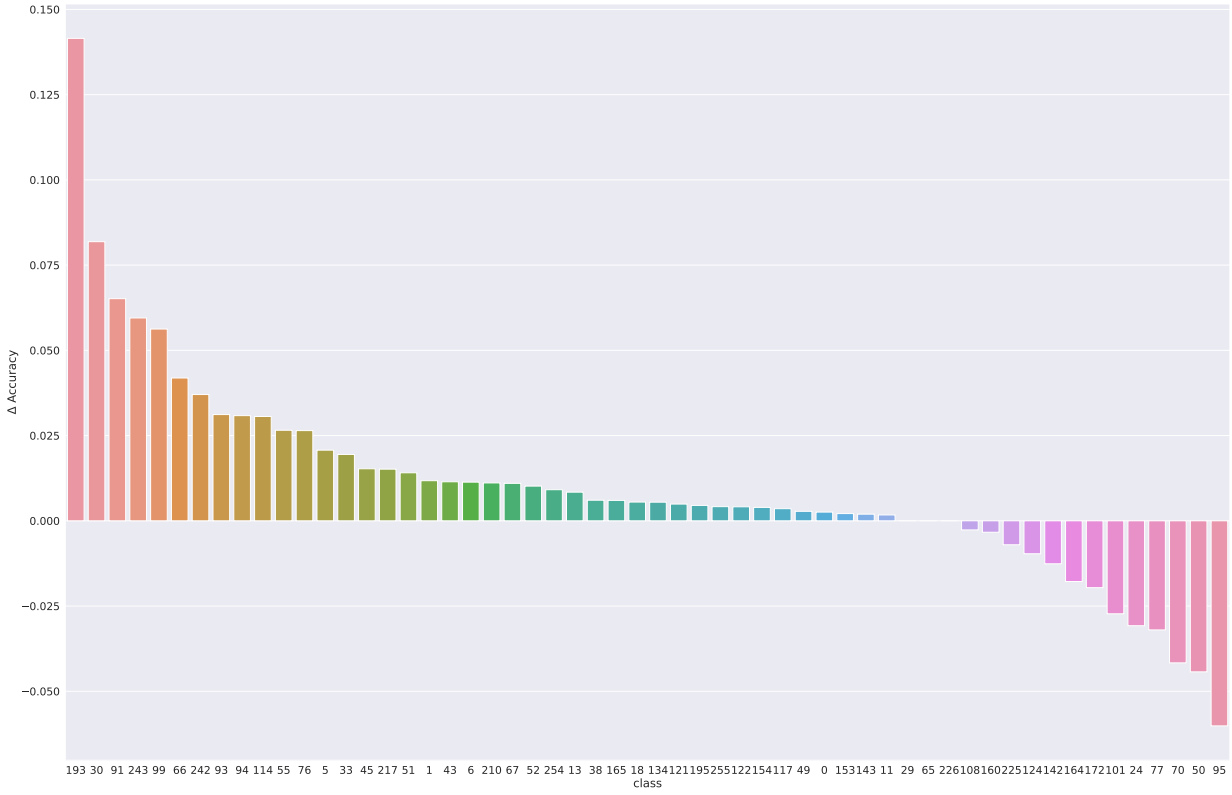
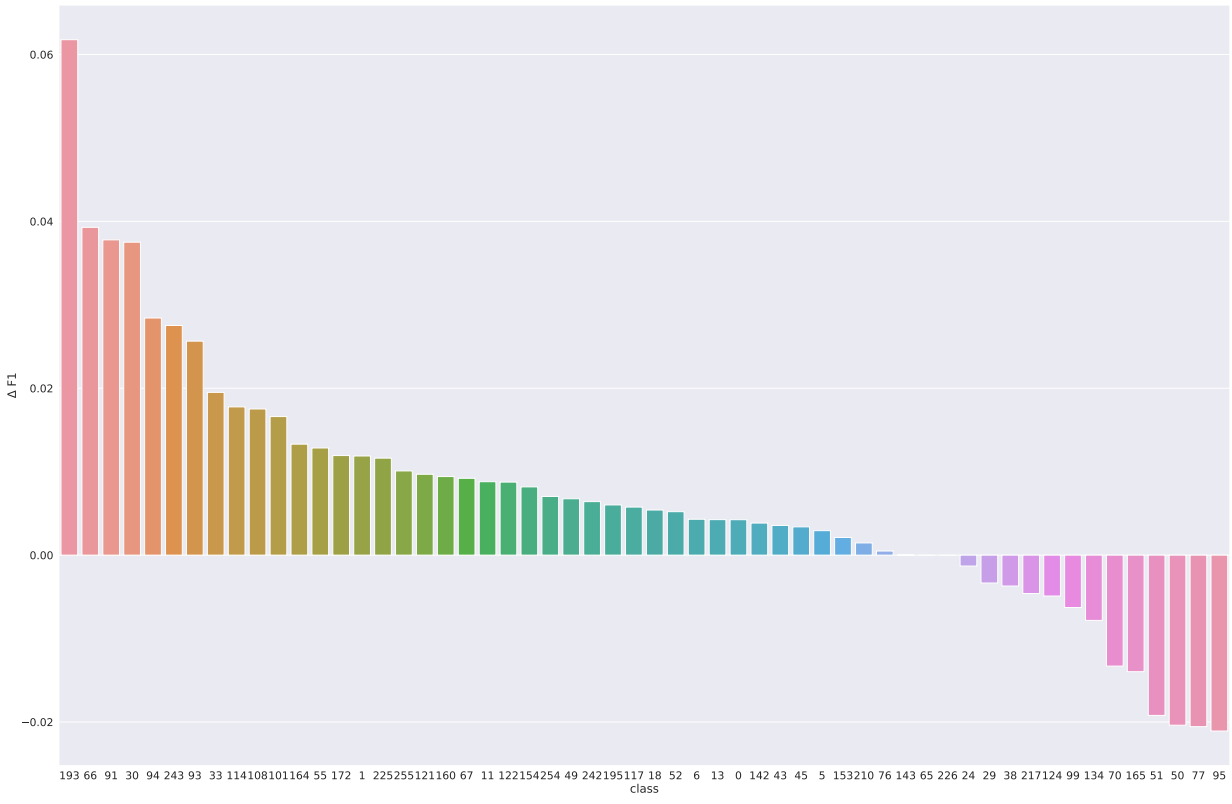


Figure 4. Direct comparison of metrics by class between FP and SSI-DDI-v2 for classes with at least 100 instances in the test dataset, sorted by decreasing number of instances in each class



(a) Accuracy



(b) Weighted-F_1

Figure 5. Differences (FP vs. SSI-DDI-v2) in metrics by class for classes with at least 100 instances in the test dataset