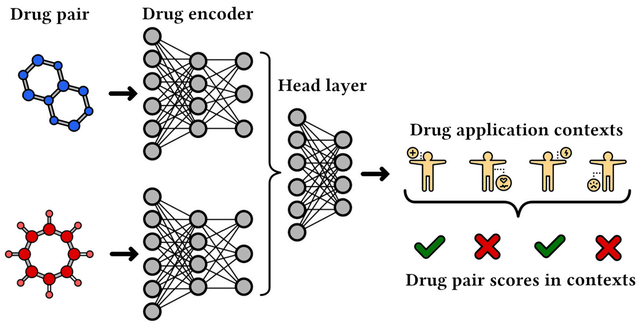
ChemicalX: A Deep Learning Library for Drug Pair Scoring

BENEDEK ROZEMBERCZKI[ψ], CHARLES TAPLEY HOYT[⋆], ANNA GOGLEVA[ψ], PIOTR GRABOWSKI[ψ], KLAS KARIS[⋆], ANDREJ LAMOV[ψ], ANDRIY NIKOLOV[ψ], SEBASTIAN NILSSON[ψ], MICHAEL UGHETTO[ψ], YU WANG[†], TYLER DERR[†], BENJAMIN M. GYORI[⋆]

HARVARD MEDICAL SCHOOL [⋆] VANDERBILT UNIVERSITY [†] ASTRAZENECA [ψ]

The Pair Scoring Task



The Training Algorithm

Data: $\mathcal{X}_{\mathcal{D}}$ - Drug feature set.
$\mathcal{X}_{\mathcal{C}}$ - Context feature set.
$\mathcal{B}$ - Labeled drug pair - context batch.
Result: $\mathcal{L}$ - The cost for the batch.

$\mathcal{L} \leftarrow 0$
for $(d, d', c, y^{d,d',c}) \in \mathcal{B}$ **do**
  $\mathbf{h}^{d} \leftarrow f_D(\mathbf{x}^{d}, \mathcal{G}^{d}, \mathbf{X}_N^{d}, \mathbf{X}_E^{d})$
  $\mathbf{h}^{d'} \leftarrow f_D(\mathbf{x}^{d'}, \mathcal{G}^{d'}, \mathbf{X}_N^{d'}, \mathbf{X}_E^{d'})$
  $\mathbf{h}^{c} \leftarrow f_C(\mathbf{x}^{c})$
  $\widehat{y}^{d,d',c} \leftarrow f_H(\mathbf{h}^{d}, \mathbf{h}^{d'}, \mathbf{h}^{c})$
  $\mathcal{L} \leftarrow \mathcal{L} + \ell(y^{d,d',c}, \widehat{y}^{d,d',c})$
end

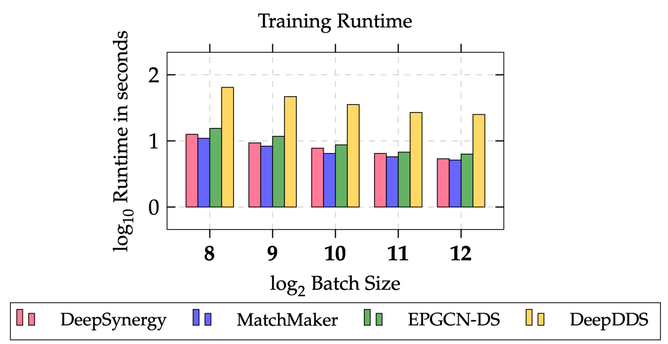
The Library Design

```python
from chemicalx.data import DrugCombDB, BatchGenerator

loader = DrugCombDB()

context_set = loader.get_context_features()
drug_set = loader.get_drug_features()
triples = loader.get_labeled_triples()

generator = BatchGenerator(batch_size=1024,
                           context_features=True,
                           drug_features=True,
                           drug_molecules=False,
                           context_feature_set=context_set,
                           drug_feature_set=drug_set,
                           labeled_triples=triples)
```

```python
import torch
from chemicalx.models import DeepSynergy

model = DeepSynergy(context_channels=112,
                    drug_channels=256)

optimizer = torch.optim.Adam(model.parameters())
model.train()
loss = torch.nn.BCELoss()

for epoch in range(200):
    for batch in generator:
        optimizer.zero_grad()
        prediction = model(batch.context_features,
                           batch.drug_features_left,
                           batch.drug_features_right)
        loss_value = loss(prediction, batch.labels)
        loss_value.backward()
        optimizer.step()
```

Experimental Validation

	AUROC	AUPR	F_1
DeepDDI	.929 ± .001	.907 ± .001	.848 ± .009
DeepSynergy	**.940 ± .001**	**.919 ± .001**	**.887 ± .001**
MR-GNN	.937 ± .002	.917 ± .001	.875 ± .002
SSI-DDI	.823 ± .002	.800 ± .003	.756 ± .001
EPGCN-DS	.855 ± .003	.834 ± .002	.785 ± .004
DeepDrug	.923 ± .004	.904 ± .002	.857 ± .002
GCN-BMP	.709 ± .003	.694 ± .002	.592 ± .003
DeepDDS	.915 ± .002	.898 ± .002	.839 ± .003
MatchMaker	.912 ± .002	.892 ± .001	.849 ± .001



References

Rozemberczki et al., *A Unified View of Relational Deep Learning for Drug Pair Scoring. IJCAI, 2022.*

Rozemberczki et al., *MOOMIN: Deep Molecular Omics Network for Anti-Cancer Drug Combination Therapy. CIKM, 2022.*