



فرم تعریف پایان نامه دوره کارشناسی ارشد

مدیر محترم گروه آموزشی کامپیوتر

با سلام و احترام

اینجانب سونیا فلاحتی مجاوری به شماره دانشجویی ۴۰۰۴۰۱۱۲۰۴۷ دانشجوی مقطع کارشناسی ارشد رشته هوش مصنوعی و رباتیکز دوره روزانه / نوبت دوم / پردیس خودگردان دانشکده مهندسی برق و کامپیوتر ورودی نیمسال اول سال تحصیلی ۱۴۰۱-۱۴۰۰ که در حال حاضر ۲۴ واحد درسی را با موفقیت گذرانده ام، درخواست تصویب پیشنهادیه پایان نامه خود تحت عنوان "ارائه یک مدل پیش بینی ارتباط دارو-هدف مبتنی بر شبکه عصبی گراف" دارم.

امضاء دانشجو و تاریخ

☐ میدانی

☐ آزمایشگاهی

☒ تئوری: نوع پایان نامه:

استاد(های) راهنما و مشاور

ردیف	نام خانوادگی	نام	سمت در پایان نامه	محل خدمت	مرتبه علمی	امضاء
۱	فاطمه	زمانی	استاد راهنما	دانشگاه صنعتی نوشیروانی	استادیار	
۲						

بسمه تعالی
تحصیلات تکمیلی
تعریف پایان نامه

نام و نام خانوادگی دانشجو: سونیا فلاحتی مجاوری
شماره دانشجویی: ۴۰۰۴۰۱۱۲۰۴۷
رشته تحصیلی: هوش مصنوعی و رباتیکز
مقطع تحصیلی: کارشناسی ارشد
دانشکده: مهندسی برق و کامپیوتر

۱-عنوان پایان نامه

فارسی: ارائه یک مدل پیش بینی ارتباط دارو-هدف مبتنی بر شبکه عصبی گراف

لاتین: A Graph Neural Network model for predicting Drug-Target Affinity

۲-تعریف مساله:

فرایند تولید و توسعه داروهای جدید بسیار طولانی و پرهزینه است. به طور معمول حدود ۱۰ تا ۱۵ سال و ۰.۸ تا ۱.۵ میلیارد دلار هزینه تولید یک داروی جدید است. اخیراً استفاده از هوش مصنوعی^۱ (AI) در کشف دارو بسیار مورد توجه قرار گرفته است. زیرا زمان و هزینه تولید داروهای جدید را به میزان قابل توجهی کاهش می دهد. رویکردهای مبتنی بر شبکه عصبی عمیق^۲ (DNN) در تمام مراحل توسعه دارو مورد استفاده قرار می گیرند[1]. استفاده مجدد از داروها^۳ می تواند از این فرآیند پرهزینه و طولانی جلوگیری کند و مزایایی مثل بازه زمانی کوتاهتر، سرمایه گذاری کمتر و احتمال شکست کمتری دارد[2]. به منظور استفاده مجدد از داروها به طور موثر، دانستن اینکه کدام پروتئین ها توسط کدام داروها هدف قرار می گیرند مفید است. شناسایی ارتباط دارو-هدف^۴ (DTA) گام مهمی در این راستا می باشد. مدل های محاسباتی بسیاری برای پیش بینی ارتباط میان دارو-هدف مطرح شده اند. مدل های یادگیری عمیق دارای بهترین عملکردها در پیش بینی ارتباط دارو-هدف هستند. تحقیقات انجام شده در یادگیری عمیق بیشتر نمایش مولکولهای دارو را به صورت توالی رشته ای در نظر می گیرند که روشی طبیعی برای نشان دادن مولکول ها نیست زیرا هنگام استفاده از رشته ها، اطلاعات ساختاری مولکول از بین می رود که میتواند قدرت پیش بینی یک مدل و کارایی آن را کاهش دهد.

^۱ Artificial Intelligence

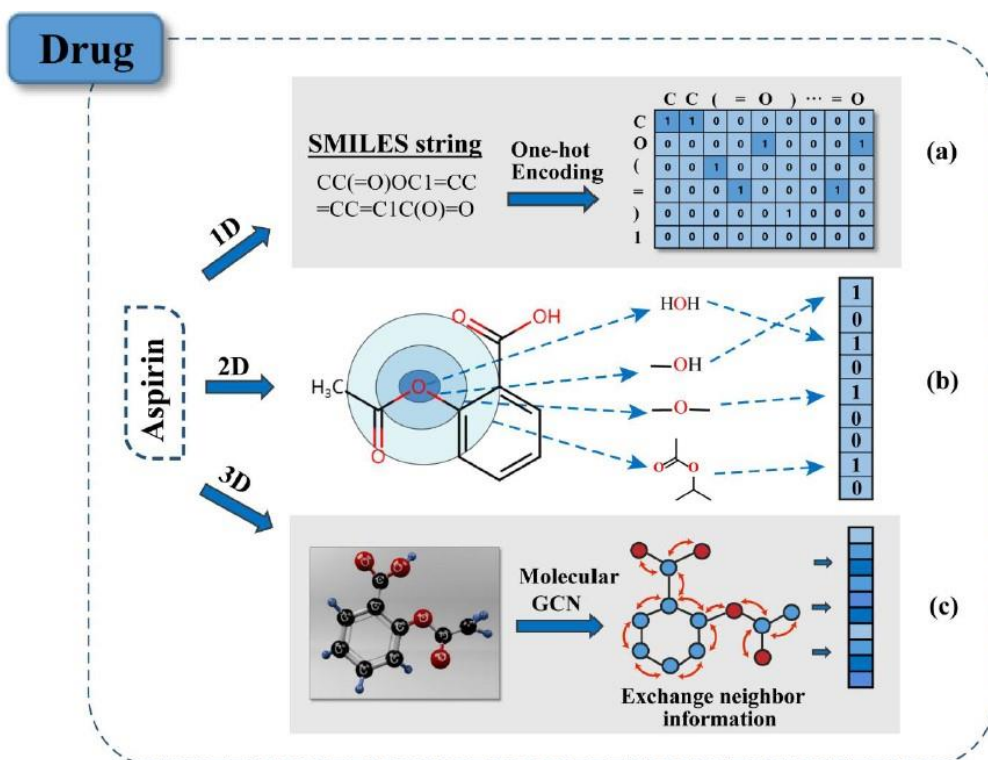
^۲ Deep Neural Network

^۳ Drug Repurposing

^۴ Drug-Target affinity

یکی از رویکردهایی که برای پیش بینی DTA مناسب است، رویکردهای مبتنی بر شبکه عصبی گراف^۵ (GNN) است.

شبکه عصبی گراف از گراف برای نمایش مولکولها استفاده میکند و می تواند ویژگیهای ساختاری مولکولها را بهتر به نمایش بگذارد و مستقیماً داده های گراف را پردازش کند. در شکل های ۱ و ۲ نمونه ای از نمایش دارو و پروتئین به صورت رشته و گراف مشاهده می شود.



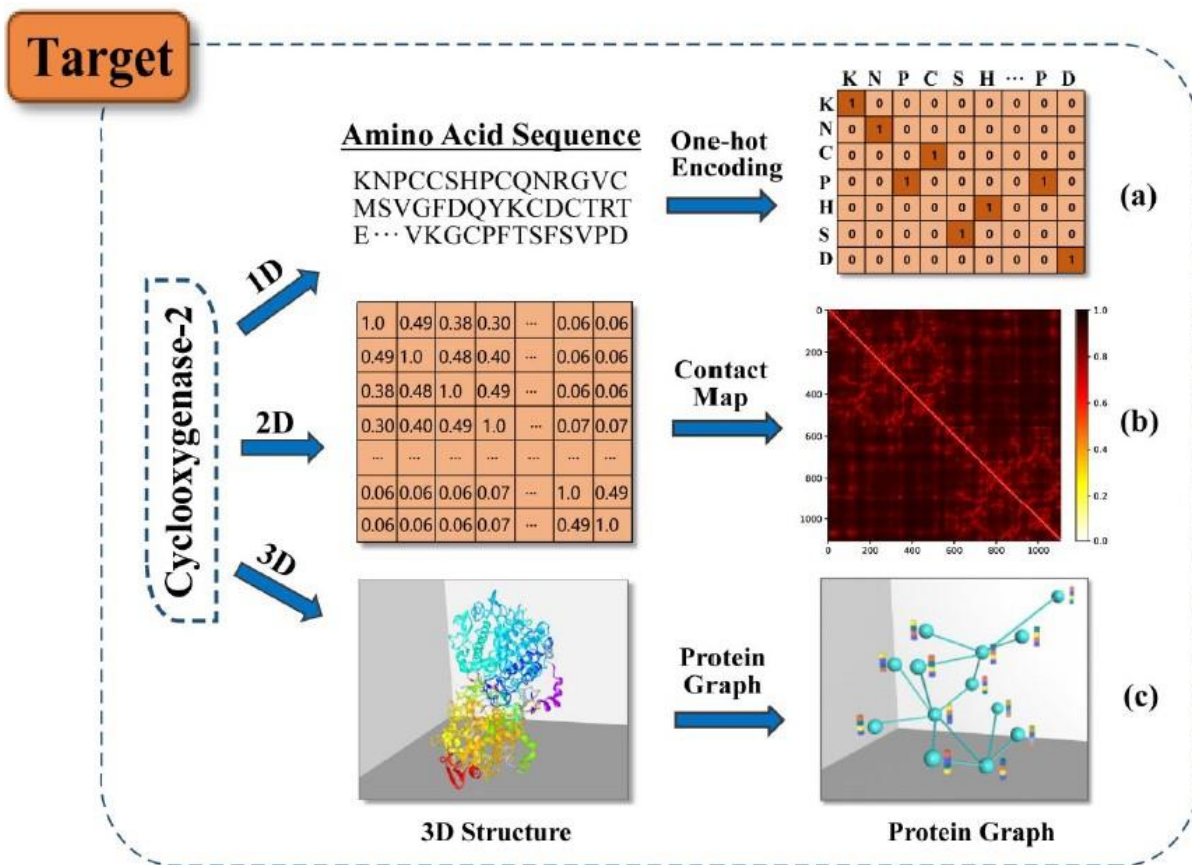
شکل ۱: نمایش دارو بصورت رشته و گراف [1]

SMILES^۶ یک نماد برای ترجمه ساختار ۳ بعدی مولکولهای شیمیایی به صورت رشته است که میتواند برای پردازش در نرم افزار های رایانه ای به کار گرفته شود.

رشته های SMILES دارو را می توان به یک گراف مولکولی تبدیل کرد به طوری که اتمها و پیوندها را به عنوان گره ها و یال های یک گراف نشان دهیم و به عنوان ورودی در ساختار GNN ها استفاده کنیم.

^۵ Graph Neural Network

^۶ Simplified Molecular Input Line Entry System



شکل ۲: نمایش پروتئین بصورت رشته و گراف [1]

مدلها و معماری های زیادی مبتنی بر شبکه عصبی گراف در پیش بینی DTA مطرح شده اند که دارای کارایی بسیار خوبی هستند. هدف ما در این پایان نامه ارائه مدلی شامل ترکیبی از رویکردهای DNN و GNN است که بتواند با تغییر و جایگزینی معماری های مختلف و نمایش ساختار پروتئین و دارو به صورت گراف به نتایج بهتری دست یابد.

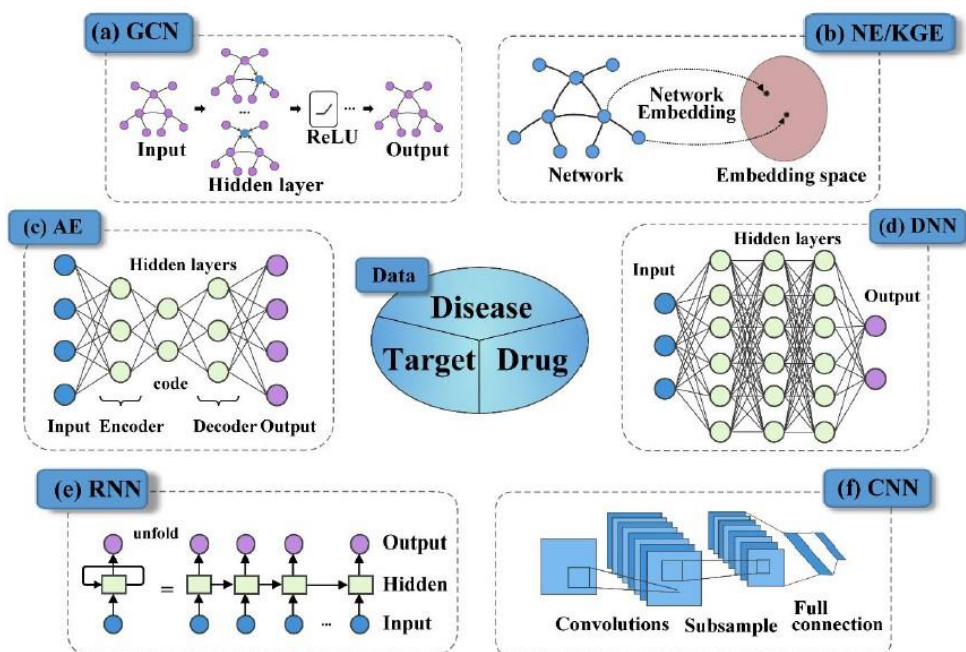
۳- پژوهش های علمی انجام شده در ارتباط با موضوع پایان نامه (بطور مختصر) :

استفاده از روشهای یادگیری عمیق و توانایی آنها در مقایسه با تکنیک های سنتی یادگیری ماشین^۷ در استفاده مجدد از دارو بسیار مورد توجه است. روشهای تولید مجدد دارو میتواند به دو دسته ی مبتنی بر رشته (-sequence based) و مبتنی بر گراف (graph-based) تقسیم شود.

در روشهای مبتنی بر رشته برای ترکیبات مولکولی دارو، یک نمایش متنی برای اطلاعات توپولوژیکی بر اساس قوانین پیوند شیمیایی به نام SMILES (Simplified Molecular Input Line Entry System) موجود است و رشته های پروتئین که به طور کلی از ۲۰ آمینواسید استاندارد تشکیل شده اند [3].

روش های مبتنی بر گراف (graph-based) ، داروها را به عنوان گرافهای مولکولی نشان می دهند.

در روشهای یادگیری عمیق پیش بینی تعاملات میان دارو-هدف بر اساس خواص مولکولی دارو و رشته های پروتئین انجام میشود. معماری های مختلفی از این تکنیک ها در شکل ۳ مشاهده میشود.

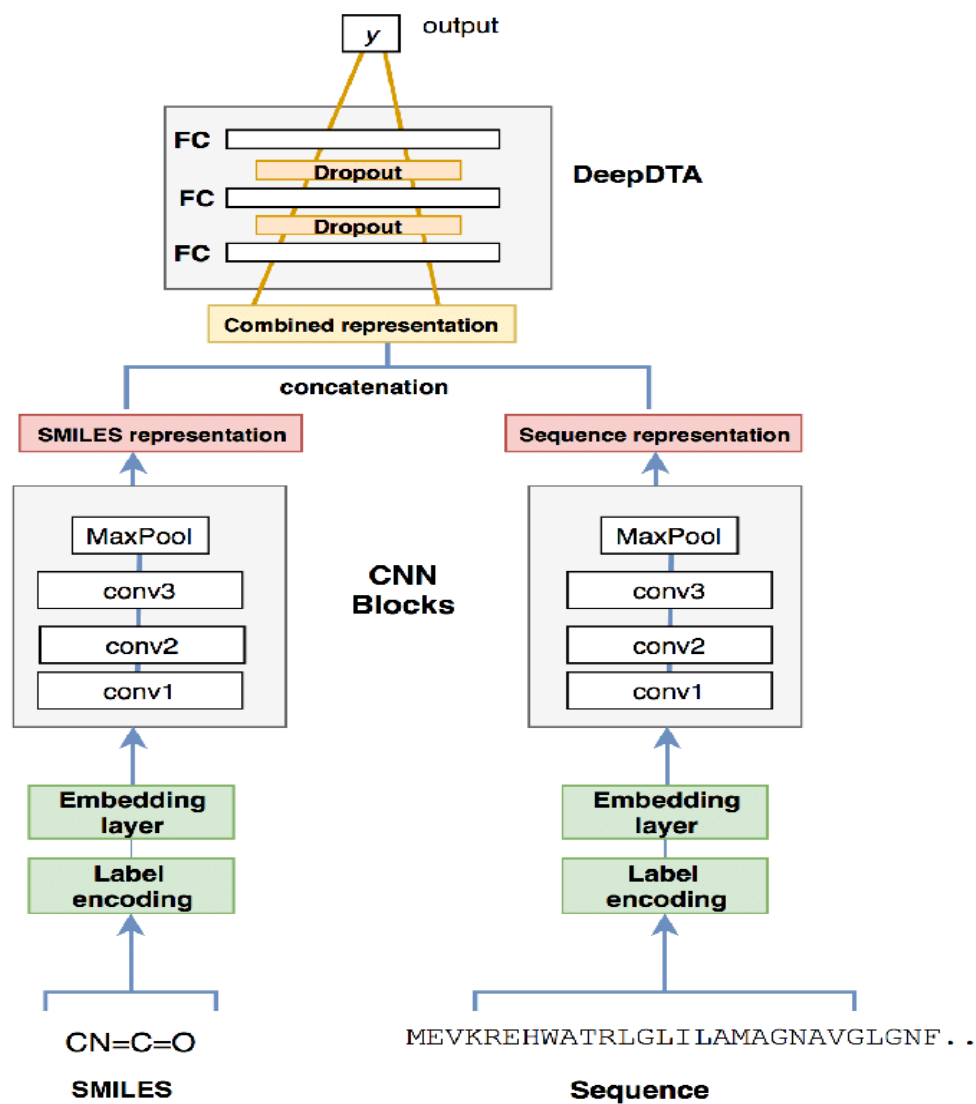


شکل ۳: معماری های مختلف یادگیری عمیق در استفاده مجدد دارو [1]

لی و همکاران در سال ۲۰۱۵ یک روش مبتنی بر جنگل تصادفی^۸ را پیشنهاد کردند [4]. این مدل RF همان سال در KronRLS، نیز مورد استفاده قرار گرفت که از matrix similarity برای بهبود عملکرد پیش‌بینی DTA استفاده کرد [5]. در سال ۲۰۱۷ مدل SimBoost مطرح شد که یک بردار ویژگی را با هر جفت دارو-هدف مرتبط می‌کند از جفت‌هایی با ویژگی‌های اتصال مشاهده‌شده، یک مدل را برای یادگیری آموزش می‌دهد [6]. در سال ۲۰۱۸ مدل DeepDTA [7] معرفی شد، یک مدل مبتنی بر یادگیری عمیق است که از شبکه‌های عصبی کانولوشن (CNN^۹) برای یادگیری و نمایش رشته‌های پروتئین و رشته‌های SMILES استفاده کرد. DeepDTA بهتر از KronRLS و الگوریتم‌های SimBoost عمل کرده است و نمایش‌های حاصل از بلوک‌های CNN را باهم ترکیب کرده تا به یک لایه کاملاً متصل وارد شوند و میل ترکیبی دارو-هدف را پیش‌بینی کند.

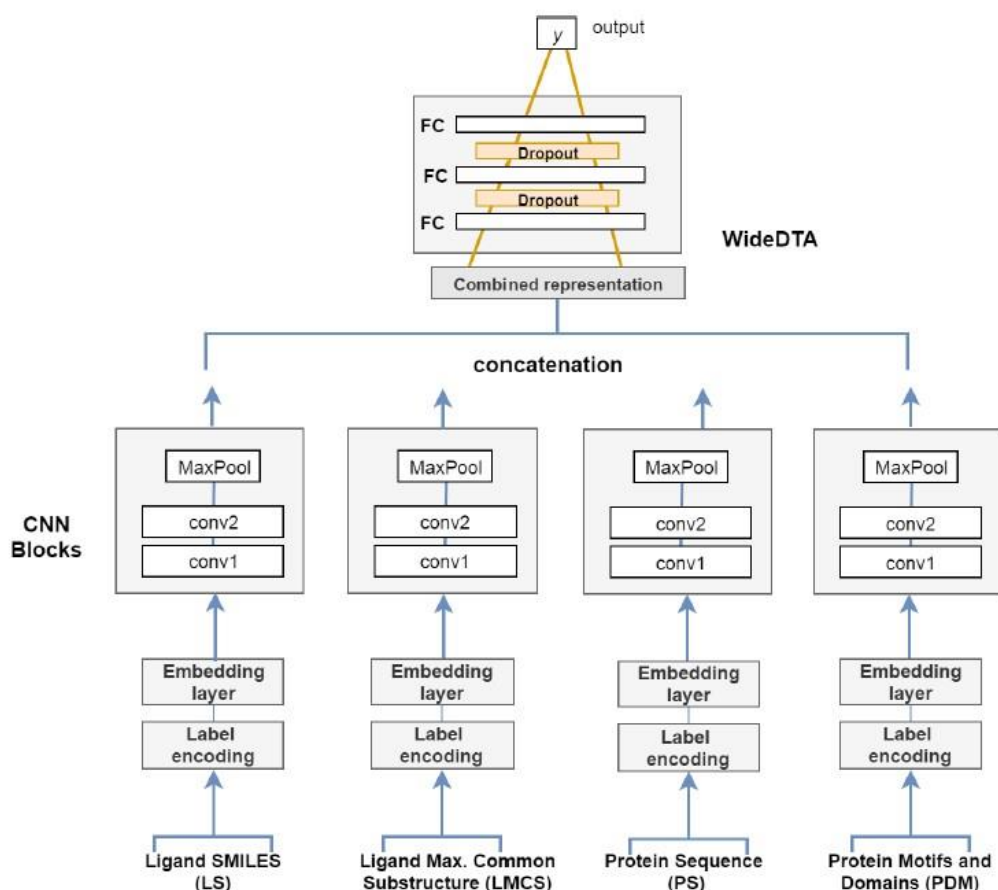
^۸ Random forest

^۹ Convolutional Neural Network



شکل ۴: مدل DeepDTA پیشنهاد شده در سال ۲۰۱۸ [7]

مدل WideDTA [8] در سال ۲۰۱۹ معرفی شده است توسعه ای از DeepDTA است. از CNN برای یادگیری بازنمایی داروها و پروتئین ها استفاده می کند. در WideDTA توالی SMILES و پروتئین به عنوان کلمات نظر گرفته شده اند و به ترتیب با یک دنباله هشت کاراکتری و یک دنباله سه کاراکتری ، نشان داده شدند. WideDTA چهار قطعه متنی مختلف از اطلاعات مربوط به پروتئین ها و لیگاندها را ترکیب می کند.



شکل ۵: مدل پیشنهادی WideDTA در سال ۲۰۱۹ [8]

Method	Proteins	Compounds	CI	MSE
KronRLS	S-W	PubChem Sim	0.871 (0.0008)	0.379
SimBoost	S-W	PubChem Sim	0.872 (0.002)	0.282
DeepDTA	PS (char)	LS (char)	0.878 (0.004)	0.261
WideDTA (best)	PS + PDM	LS + LMCS	0.886 (0.003)	0.262

جدول ۱: مقایسه نتایج حاصل از WideDTA و سایر مدل های مطرح شده

GANsDTA [9] یک مدل نیمه نظارت شده^{۱۰} جدید مبتنی بر شبکه‌های متخاصم مولد^{۱۱} (GANs) برای پیش‌بینی DTA بین رشته‌های SMILES و پروتئین پیشنهاد کرد. DeepGS [10] روش دیگری است که از توصیفگرهای SMILES و رشته‌های پروتئین برای پیش‌بینی DTA استفاده می‌کند. این روش‌ها نشان داده‌اند که شبکه‌های عمیق بهتر از الگوریتم‌های یادگیری ماشین سنتی می‌توانند عمل کنند.

Deep-Affinity [11] از ترکیب CNN و LSTM برای استخراج ویژگی‌ها و پیش‌بینی DTA میان SMILES و رشته‌های پروتئین استفاده کرد. این مدل‌های یادگیری عمیق داروها را به شکل string نمایش میدادند که با ساختار طبیعی مولکولها متفاوت است.

هنگام استفاده از رشته‌ها، اطلاعات ساختاری مولکول از بین می‌رود، که می‌تواند قدرت پیش‌بینی یک مدل و همچنین عملکرد آنرا مختل کند.

در سالهای اخیر شبکه عصبی گراف (GNN) به طور گسترده در زمینه‌های مختلف استفاده شده است. از یک گراف متشکل از گره‌ها و یال‌ها به عنوان ورودی GNN استفاده می‌شود و هیچ محدودیتی برای اندازه گراف ورودی وجود ندارد، که قابلیت انعطاف پذیر برای استخراج اطلاعات مولکول‌ها را ارائه می‌دهد. شبکه کانولوشن گراف^{۱۲} (GCN) و شبکه توجه گراف^{۱۳} (GAT) مدل‌های پرکاربرد GNN هستند و به تدریج در طراحی دارو، مانند پیش‌بینی خواص دارو و همینطور پیش‌بینی DTA استفاده شده‌اند.

GraphDTA [12] نمایش گراف را برای استفاده از اطلاعات ساختار مولکولها معرفی کرد و از GNN برای پیش‌بینی DTA استفاده کرده است. نتایج تأیید می‌کند که در مدل‌های یادگیری عمیق برای پیش‌بینی DTA، نمایش داروها به صورت گراف می‌تواند منجر به بهبود عملکرد مدل شود.

GraphDTA ترکیبات دارویی را به عنوان گرافی از فعل و انفعالات بین اتم در نظر گرفته و مدل را حول این مفهوم پیاده سازی میکند. برای توصیف یک گره در گراف، از مجموعه‌ای از ویژگی‌های اتمی استخراج شده از DeepChem (رامسوندار و همکاران، ۲۰۱۹) استفاده می‌کند. هر گره، یک بردار ویژگی باینری چند بعدی است که شامل پنج مورد اطلاعات زیر است:

نماد اتم،

تعداد اتم‌های مجاور،

تعداد هیدروژن‌های مجاور،

implicit value اتم

و aromatic structure.

در واقع کد SMILES را با استفاده از RDKit و استخراج ویژگی‌های اتمی به گراف مولکولی متناظرش تبدیل میکند. یک مدل پیش‌بینی DTA را بر اساس ترکیبی از شبکه‌های عصبی گراف و CNN پیشنهاد می‌کند. برای پروتئین‌ها از رشته‌ای از کاراکترهای ASCII استفاده می‌کند.

و چندین لایه CNN بعدی را روی آنها اعمال میکند تا مدل بردار نمایش یک رشته را یاد بگیرد.

برای دارو از ۴ مدل مختلف GNN شامل GCN (Graph convolutional network)،

GAT (graph attention

network)

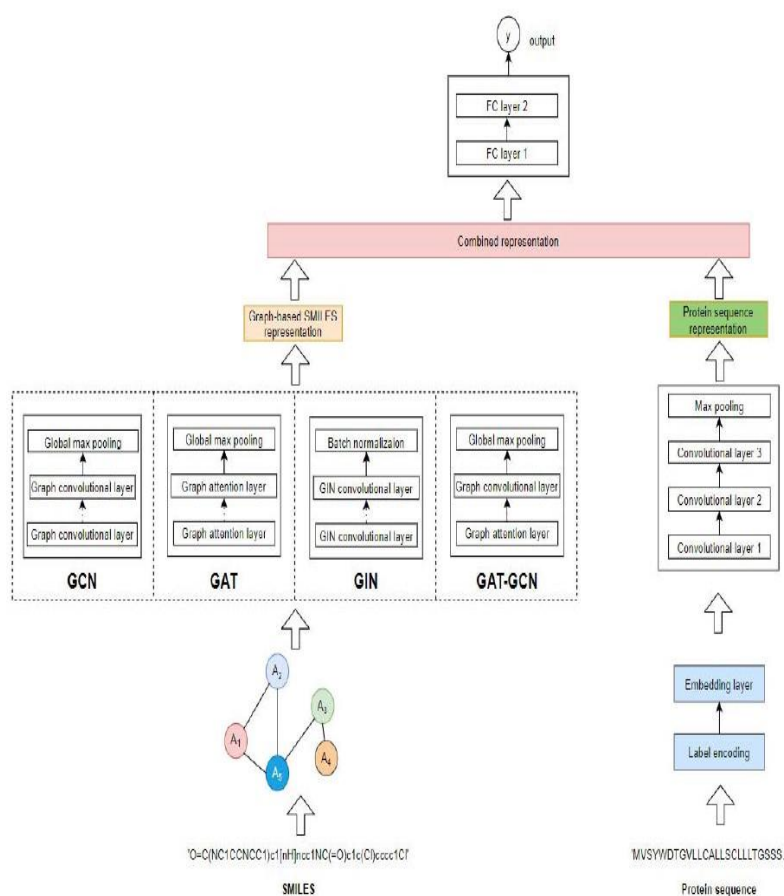
^{۱۰} semi-supervised

^{۱۱} Generative adversarial networks

^{۱۲} Graph Convolutional Network

^{۱۳} Graph Attention Network

GIN (graph isomorphic network) و ترکیبی از GAT-GCN استفاده میکنند.



شکل ۷: معماری مدل پیشنهاد شده GraphDTA [12]

GraphDTA در ۳ مرحله کار می کند: ابتدا کد SMILES یک دارو به یک گراف مولکولی تبدیل می شود و یک الگوریتم یادگیری عمیق، نمایش گراف را می آموزد. در همین حال، رشته پروتئین کدگذاری و تعبیه شده است و چندین لایه کانولوشن ۱ بعدی نمایش آن را یاد می گیرند. در نهایت، دو بردار نمایش با هم ادغام میشوند و از چندین لایه کاملاً متصل عبور می کنند تا خروجی را تخمین بزنند. نتایج حاصل از این مدل روی مجموعه داده های Kiba و Davis در جدول ۲ و ۳ مشاهده می شود.

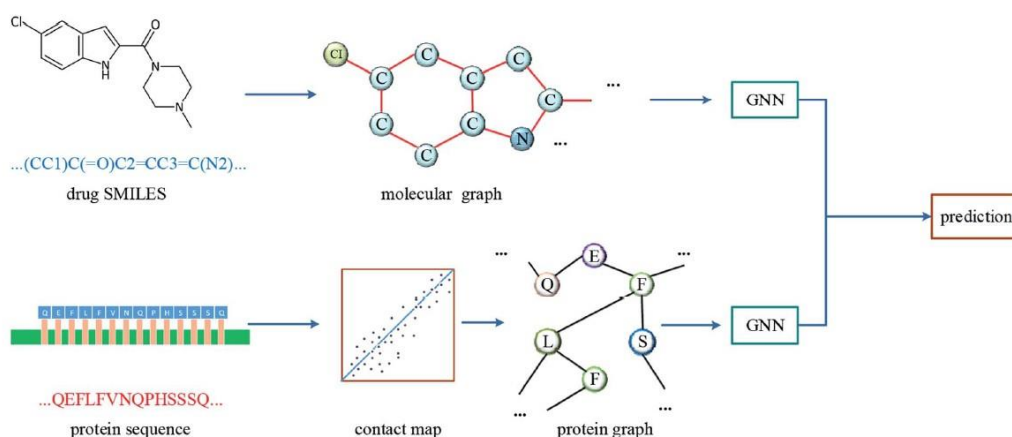
Method	Protein rep.	Compound rep.	CI	MSE
Baseline models				
DeepDTA (Öztürk et al., 2018)	Smith-Waterman	Pubchem-Sim	0.790	0.608
DeepDTA (Öztürk et al., 2018)	Smith-Waterman	1D	0.886	0.420
DeepDTA (Öztürk et al., 2018)	1D	Pubchem-Sim	0.835	0.419
KronRLS (Cichonska et al., 2017, 2018)	Smith-Waterman	Pubchem-Sim	0.871	0.379
SimBoost (He et al., 2017)	Smith-Waterman	Pubchem-Sim	0.872	0.282
DeepDTA (Öztürk et al., 2018)	1D	1D	0.878	0.261
WideDTA (Öztürk et al., 2019)	1D + PDM	1D + LMCS	0.886	0.262
Proposed method - GraphDTA				
GCN	1D	Graph	0.880	0.254
GAT_GCIN	1D	Graph	0.881	0.245
GAT	1D	Graph	0.892	0.232
GIN	1D	Graph	0.893	0.229

جدول ۲: نتایج مدل Graph-DTA روی دیتاست Davis در مقایسه با سایر مدل ها [12]

Method	Protein rep.	Compound rep.	CI	MSE
Baseline models				
DeepDTA (Öztürk et al., 2018)	1D	Pubchem-Sim	0.718	0.571
DeepDTA (Öztürk et al., 2018)	Smith-Waterman	Pubchem-Sim	0.710	0.502
KronRLS (Cichonska et al., 2017, 2018)	Smith-Waterman	Pubchem-Sim	0.782	0.411
SimBoost (He et al., 2017)	Smith-Waterman	Pubchem-Sim	0.836	0.222
DeepDTA (Öztürk et al., 2018)	Smith-Waterman	1D	0.854	0.204
DeepDTA (Öztürk et al., 2018)	1D	1D	0.863	0.194
WideDTA (Öztürk et al., 2019)	1D + PDM	1D + LMCS	0.875	0.179
Proposed method - GraphDTA				
GAT	1D	Graph	0.866	0.179
GIN	1D	Graph	0.882	0.147
GCN	1D	Graph	0.889	0.139
GAT_GCIN	1D	Graph	0.891	0.139

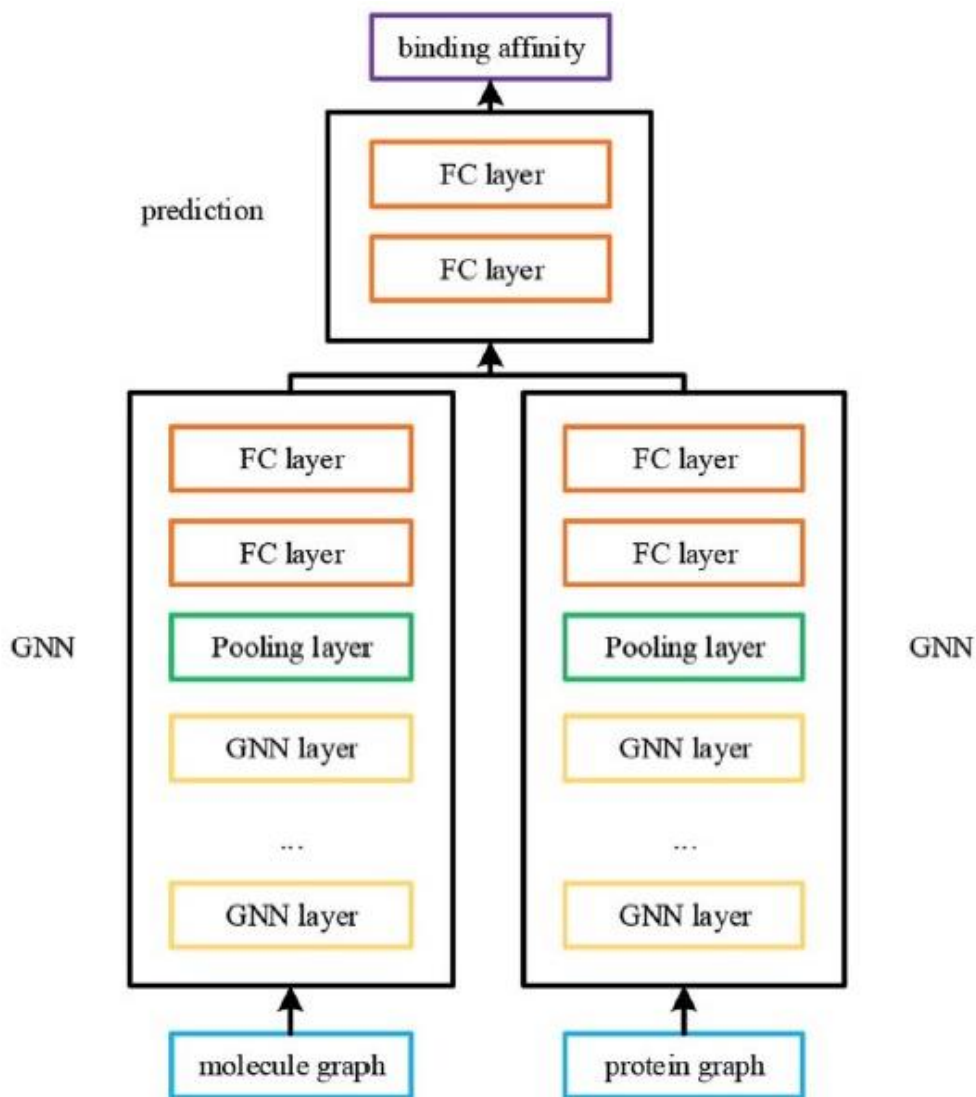
جدول ۳: نتایج مدل Graph-DTA روی دیتاست Kiba در مقایسه با سایر مدل ها [12]

نتایج نشان می دهد که GraphDTA نسبت به سایر مدل های مطرح شده عملکرد بهتری در پیش بینی DTA داشته است. اما چالشی که مطرح میشود این است که در GraphDTA، CNN برای به دست آوردن ویژگی های پروتئینی از طریق رشته ها مورد استفاده قرار گرفت و گرافی برای هر پروتئین ایجاد نشد. DGraphDTA [13] مدلی است که علاوه بر دارو، پروتئین ها را هم به صورت گراف در نظر میگیرد و برای پیش بینی DTA پیشنهاد شده است. با استفاده از GNN، DGraphDTA می تواند رشته های پروتئین را بر اساس contact map به گراف تبدیل کند و می تواند ویژگی های ساختاری پروتئین را با توجه به رشته های آنها پیش بینی کند.



شکل ۸: نمایش دارو و پروتئین در مدل پیشنهادی [13]DGraphDTA

در جدول ۴ مشاهده می شود در مقایسه با DeepDTA، WideDTA و GraphDTA، این مدل پیشنهادی با GCN های سه لایه بهبود عملکرد قابل توجهی دارد.



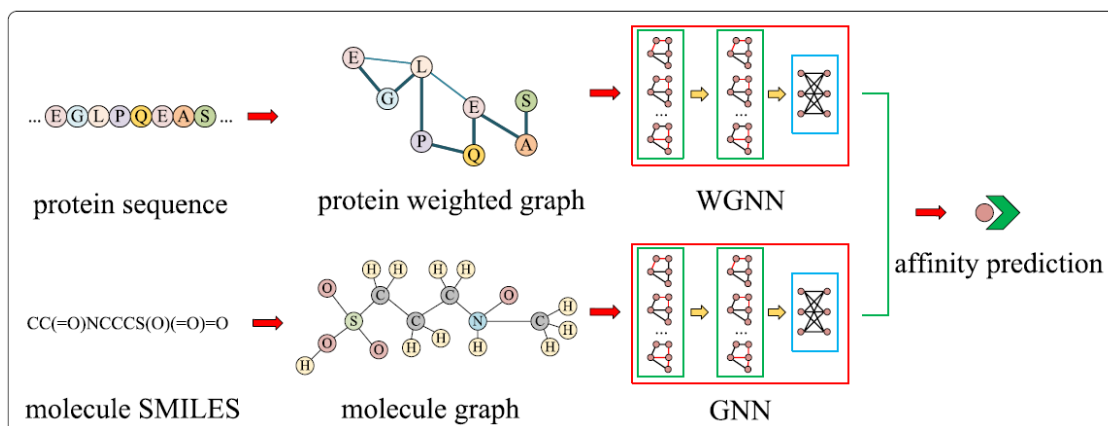
شکل ۹: معماری مدل پیشنهادی [13]DGraphDTA

Method	Proteins and compounds	CI	MSE
KronRLS	S-W & Pubchem Sim	0.871	0.379
SimBoost	S-W & Pubchem Sim	0.872	0.282
DeepDTA	S-W & Pubchem Sim	0.790	0.608
DeepDTA	CNN & Pubchem Sim	0.835	0.419
DeepDTA	S-W & CNN	0.886	0.420
DeepDTA	CNN & CNN	0.878	0.261
WideDTA	PS + PDM & LS + LMCS	0.886	0.262
GraphDTA	GIN & 1D	0.893	0.229
DGraphDTA	GCN & GCN	0.904	0.202

جدول ۴: مقایسه نتایج مدل پیشنهادی DGraphDTA [13]

WGNN-DTA [14] یک مدل گراف وزندار است که می تواند برای پروتئین ها گراف وزندار مناسبی ایجاد کند و آزمایش های مختلف نشان می دهد این مدل برای پیش بینی DTA سریعتر از مدل DGraphDTA است و دقت بالایی دارد.

ورودی های مدل رشته پروتئین و رشته SMILES است. یک گراف وزنی برای پروتئین ساخته شده است که می تواند به طور جامع اطلاعات ساختار پروتئین را شرح دهد و به جای مقدار باینری که در DGraphDTA استفاده می شود، از احتمال تعامل به عنوان وزن یال های گراف استفاده می شود. از دو نوع GNN، (GCN) و (GAT) برای استخراج بردارهای نهفته گراف های پروتئین و دارو، و پیش بینی DTA بر اساس بردارهای نهفته استفاده می کند.



شکل ۱۰: مدل WGNN-DTA [۱۴]

۷-فرض‌های پژوهش:

۷-۱:مجموعه داده: در این پایان نامه مجموعه داده های معیاری که برای ارزیابی عملکرد مدل استفاده می شود شامل دو مجموعه داده Davis و KIBA است که در اکثر مقالات مورد استفاده قرار گرفته است. در ادامه بطور مختصر هر دو مجموعه داده معرفی میشوند.

۷-۱-۱:معرفی مجموعه داده ها:

Davis یک مجموعه داده از Kinase است که توسط (دیویس و همکاران سال ۲۰۱۱) معرفی شد شامل ۲۵۰۴۶ جفت دارو-هدف است.

مجموعه داده Kiba توسط (تانگ و همکاران سال ۲۰۱۴) ارائه شده است. رشته های SMILES از هر دو مجموعه داده از پایگاه داده PubChem و رشته های پروتئین از پایگاه داده پروتئین UniProt استخراج شده است. مشخصات هر دو مجموعه داده در جدول زیر آورده شده است:

	Proteins	Compounds	Interactions
Davis (K_d)	442	68	30056
KIBA	229	2111	118254

جدول ۵ : معرفی مجموعه داده

۷-۲: ارزیابی مدل پیشنهادی

برای ارزیابی مدل از CI (concordance index) و MSE (mean square error) استفاده میشود. CI میزان احتمال شباهت بین مقادیر واقعی و پیش بینی شده را اندازه گیری می کند، فرض کنید δ_i و b_i به ترتیب بیانگر مقدار واقعی و مقدار پیش‌بینی شده نمونه i باشند. بنابراین CI به صورت زیر تعریف می شود:

$$CI = \frac{1}{Z} \sum_{\delta_i > \delta_j} b(b_i - b_j)$$

MSE نشان دهنده تفاوت بین مقادیر پیش بینی شده و مقادیر واقعی است در طول آموزش، یک مدل تلاش می کند تا اختلاف بین مقدار واقعی و پیش بینی شده را کاهش دهد. MSE (به عنوان $loss$ function استفاده میشود، در P بردار پیش بینی است و Y مربوط به بردار خروجی های واقعی است و n تعداد نمونه ها را نشان می دهد:

$$MSE = \frac{1}{n} \sum_{i=1}^n (P_i - Y_i)^2$$

۸- اهداف پژوهش:

هدف از انجام این پژوهش ارائه یک مدل مبتنی بر شبکه عصبی گراف برای پیش بینی ارتباط دارو-هدف است. در این پژوهش از ترکیب شبکه عصبی عمیق و شبکه عصبی گراف استفاده میکنیم و داروها و پروتئین ها را به صورت گراف نمایش میدهیم در راستای نیل به این هدف موارد زیر در نظر گرفته خواهد شد:

- تحقیق درباره ی یافتن مدل مناسب برای نمایش پروتئین ها به صورت گراف و پیاده سازی آن
- تحقیق در رابطه با معماری های جایگزین برای ترکیب شبکه عصبی عمیق و شبکه عصبی گراف و در نهایت پیاده سازی آن.

۹-خلاصه مراحل روش پژوهش و زمان بندی اجرای پایان نامه :

مرحله ۱:

تحقیق در رابطه با نمایش پروتئین بصورت گراف

مرحله ۲:

پیاده سازی مدل پیشنهاد شده

مرحله ۳:

نگارش مقاله کنفرانس

مرحله ۴:

تحقیق در رابطه با معماری های جایگزین برای GNN

مرحله ۵:

پیاده سازی مدل پیشنهاد شده با معماری جایگزین و ترکیبی

مرحله ۶:

نگارش مقاله مجله

مرحله ۷:

نگارش پایان نامه

۹-۱:زمان بندی اجرای پایان نامه:

۱۲	۱۱	۱۰	۹	۸	۷	۶	۵	۴	۳	۲	۱	
												انجام تحقیقات برای نمایش گرافی پروتئین ها
												پیاده سازی مدل پیشنهادی
												نگارش مقاله کنفرانس
												انجام تحقیقات جایگزینی معماری های مختلف
												پیاده سازی مدل
												نگارش مقاله مجله
												نگارش پایان نامه

۱۰- پیاده سازی انجام شده :

با اجرای کد مربوط به مدل مطرح شده GraphDTA و مشاهده نتایج مدل‌های مختلف GNN که بر روی نمایش گراف دارو استفاده شده است.

در تحقیقات قبلی از GCN برای نمایش گراف دارو استفاده شده است و بردارهای ویژگی در سطح گره‌های گراف استخراج شده‌اند و برای تخمین DTA به کار رفته‌اند [12].

لازم به ذکر است در GCN پس از آخرین لایه یک لایه ادغام max pooling اضافه می‌شود.

بر خلاف GCN ها، GAT یک معماری مبتنی بر توجه را برای نمایش گره‌های پنهان در یک گراف با استفاده از مکانیسم توجه پیشنهاد می‌کند.

در این مقاله از یک مدل ترکیبی GAT-GCN استفاده شده که گراف شبکه عصبی

با یک لایه GAT شروع می‌شود و گراف دارو را به عنوان ورودی می‌گیرد و از لایه GCN بعدی عبور می‌کند. هر لایه توسط یک تابع ReLU فعال می‌شود.

معماری GAT-GCN به صورت GAT-GAT تغییر داده شد که در آن GCN با GAT جایگزین شده است. از آموزش مدل تحت این معماری جدید (با ۶۰۰ epoch) روی مجموعه داده Davis به نتایج $CI=0.874$ و $MSE=0.243$ رسیدیم که با نتایج گزارش شده در [12] $MSE=0.245, CI=0.881$ قابل مقایسه می‌باشد.

```
Train epoch: 550 [651000/25046 (82%)] Loss: 0.1124300
Make prediction for 5010 samples...
0.29464105 No improvement since epoch 430 ; best_mse,best_ci: 0.24385601 0.8746497590219549 GAT_GATNet davis
Training on 25046 samples...
Train epoch: 597 [0/25046 (0%)] Loss: 0.112296
Train epoch: 597 [331380/25046 (41%)] Loss: 0.077877
Train epoch: 597 [644400/25046 (82%)] Loss: 0.147819
Make prediction for 5010 samples...
0.27816993 No improvement since epoch 430 ; best_mse,best_ci: 0.24385601 0.8746497590219549 GAT_GATNet davis
Training on 25046 samples...
Train epoch: 598 [0/25046 (0%)] Loss: 0.092846
Train epoch: 598 [329460/25046 (41%)] Loss: 0.143344
Train epoch: 598 [656720/25046 (82%)] Loss: 0.094230
Make prediction for 5010 samples...
0.25496197 No improvement since epoch 430 ; best_mse,best_ci: 0.24385601 0.8746497590219549 GAT_GATNet davis
Training on 25046 samples...
Train epoch: 599 [0/25046 (0%)] Loss: 0.089205
Train epoch: 599 [328580/25046 (41%)] Loss: 0.106614
Train epoch: 599 [660080/25046 (82%)] Loss: 0.087662
Make prediction for 5010 samples...
0.24805175 No improvement since epoch 430 ; best_mse,best_ci: 0.24385601 0.8746497590219549 GAT_GATNet davis
Training on 25046 samples...
Train epoch: 600 [0/25046 (0%)] Loss: 0.074070
Train epoch: 600 [328680/25046 (41%)] Loss: 0.102841
Train epoch: 600 [648680/25046 (82%)] Loss: 0.086804
Make prediction for 5010 samples...
0.25501794 No improvement since epoch 430 ; best_mse,best_ci: 0.24385601 0.8746497590219549 GAT_GATNet davis
(geometric) C:\Users\Aseman\Desktop\GraphDTA-master>pythConcordance Index
```

شکل ۱۱: گزارش نتایج پیاده سازی انجام شده با مدل GAT-GAT

- [1] Pan, Xiaoqin, et al. "Deep learning for drug repurposing: Methods, databases, and applications." *Wiley interdisciplinary reviews: Computational molecular science* 12.4 (2022): e1597..
- [2] Yeu, Yunku, Youngmi Yoon, and Sanghyun Park. "Protein localization vector propagation: a method for improving the accuracy of drug repositioning." *Molecular BioSystems* 11.7 (2015): 2096-2102.
- [3] Weininger, David. "SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules." *Journal of chemical information and computer sciences* 28.1 (1988): 31-36.
- [4]: Li, Hongjian, et al. "Low-quality structural and interaction data improves binding affinity prediction via random forest." *Molecules* 20.6 (2015): 10947-10962.
- [5]: Pahikkala, Tapio, et al. "Toward more realistic drug–target interaction predictions." *Briefings in bioinformatics* 16.2 (2015): 325-337.
- [6]: He, Tong, et al. "SimBoost: a read-across approach for predicting drug–target binding affinities using gradient boosting machines." *Journal of cheminformatics* 9.1 (2017): 1-14.
- [7]: Öztürk, Hakime, Arzucan Özgür, and Elif Ozkirimli. "DeepDTA: deep drug–target binding affinity prediction." *Bioinformatics* 34.17 (2018): i821-i829.
- [8]: Öztürk, Hakime, Elif Ozkirimli, and Arzucan Özgür. "WideDTA: prediction of drug-target binding affinity." *arXiv preprint arXiv:1902.04166* (2019).
- [9]: Zhao, Lingling, et al. "GANsDTA: Predicting drug-target binding affinity using GANs." *Frontiers in genetics* 10 (2020): 1243.
- [10]: Lin, Xuan. "DeepGS: Deep representation learning of graphs and sequences for drug-target binding affinity prediction." *arXiv preprint arXiv:2003.13902* (2020).
- [11]: Karimi, Mostafa, et al. "DeepAffinity: interpretable deep learning of compound–protein affinity through unified recurrent and convolutional neural networks." *Bioinformatics* 35.18 (2019): 3329-3338.
- [12]: Nguyen, Thin, et al. "GraphDTA: predicting drug–target binding affinity with graph neural networks." *Bioinformatics* 37.8 (2021): 1140-1147.
- [13]: Jiang, Mingjian, et al. "Drug–target affinity prediction using graph neural network and contact maps." *RSC advances* 10.35 (2020): 20701-20712.
- [14]: Jiang, Mingjian, et al. "Sequence-based drug-target affinity prediction using weighted graph neural networks." *BMC genomics* 23.1 (2022): 1-17.

۱۲-اطلاعات مربوط به هزینه‌ها

۱۲-۱- آیا برای این طرح از سازمان‌های دیگر نیز درخواست اعتبار شده است؟ خیر

۱۲-۲- فهرست وسایل و موادی که باید از اعتبار این طرح از داخل یا خارج کشور خریداری شود درجدول زیر مشخص نمایید.

نام دستگاه و یا مواد	شرکت سازنده و یا فروشنده	کشور	مصرفی	غیر مصرفی	آیا در ایران موجود است؟	تعداد واحد	قیمت	قیمت کل
جمع هزینه های وسائل ، مواد								ریال

۱۲-۳- هزینه شرکت در سمینار علمی داخلی:

نام کنفرانس	زمان برگزاری کنفرانس	هزینه ثبت نام به ریال
جمع کل هزینه سمینار :		

*هزینه شرکت در سمینار علمی، شامل هزینه ثبت نام سمینار و هزینه بلیط رفت و برگشت به کنفرانس مربوطه می باشد که مبالغ هزینه شده، پس از بازگشت از کنفرانس و با ارائه مدارک لازم، قابل دریافت از امور مالی دانشگاه می باشد.

۱۲-۴- هزینه های تهیه نشریات و آماده سازی پایان نامه :

الف- هزینه های تکثیر اوراق و پرسشنامه ها و غیره :

ب- هزینه تهیه نشریات و کتب مورد لزوم :

پ- هزینه تایپ و تکثیر پایان نامه:ریال

ت- سایر هزینه ها :ریال

جمع هزینه های تهیه نشریات و آماده سازی پایان نامه :ریال

۱۳- جمع هزینه های طرح :

الف- جمع هزینه های وسائل و مواد(ماده ۹) :ریال
ب- جمع هزینه های سمینار علمی داخلی(ماده ۱۰) :ریال
پ- جمع هزینه های تهیه نشریات و آماده سازی پایان نامه(ماده ۱۱) :ریال
جمع کل هزینه :ریال

۱۴- این جانب استاد راهنمای دانشجو، با آگاهی از این امر که کمک هزینه اعطایی برای انجام پایان نامه محدود (حداکثر تا سقف مصوب شورای دانشگاه) خواهد بود، اعلام می نمایم که برای جلوگیری از آسیب در روند اجرای پایان نامه، چنانچه نیاز به صرف بودجه بیشتری باشد از منابع دیگر تامین خواهد شد.

امضاء استاد (ساتید) راهنما
نام و نام خانوادگی با قید تاریخ

۱۵- تعهد نامه دانشجو :

اینجانب متعهد می شوم که در مدت اجرای پایان نامه، بطور تمام وقت انجام وظیفه نمایم، همچنین اطلاع دارم که کلیه نتایج و حقوق حاصله از این پایان نامه، متعلق به دانشگاه بوده و مجاز نیستم بدون موافقت آن، اطلاعاتی را در رابطه با پایان نامه، به دیگری واگذار نمایم ./

نام و امضاء دانشجو

- فرم تعریف پایان نامه در شورای گروه آموزشی مورخ ، مطرح و مورد تصویب قرار گرفت. /

امضاء

۱۶- فرم تعریف پایان نامه در شورای تحصیلات تکمیلی دانشکده مورخ ، مطرح و مورد تصویب قرار گرفت. /

امضاء

۱۷- اعتبار پایان نامه آقای / خانم در شورای تحصیلات تکمیلی دانشگاه
مورخ ، با مبلغ ریال، مورد تصویب قرار گرفت./

امضاء