

تکنیک‌های هوش مصنوعی برای تسریع کشف دارو

*محمد مهدی موذنی^۱

۱- کارشناسی ارشد مهندسی صنایع، دانشگاه صنعتی امیرکبیر (پلی تکنیک تهران)
m.moazeni1376@yahoo.com

چکیده

فرآیند سنتی کشف و توسعه دارو، پروسه‌ای بسیار زمان‌بر، پرهزینه و با نرخ شکست بالا است که معمولاً بیش از یک دهه زمان و میلیاردها دلار سرمایه نیاز دارد. با ظهور کلان‌داده‌ها و افزایش قدرت پردازشی، هوش مصنوعی و به ویژه یادگیری ماشین، به عنوان ابزاری تحول‌آفرین برای غلبه بر این چالش‌ها مطرح شده است. هدف این مقاله، ارائه یک مرور ساختاریافته بر تکنیک‌های محاسباتی نوین مورد استفاده در چرخه حیات کشف دارو است. در این پژوهش، رویکردهای مختلف هوش مصنوعی شامل یادگیری نظارت‌شده، یادگیری عمیق و شبکه‌های عصبی برای کاربردهایی نظیر غربالگری مجازی، پیش‌بینی خواص دارو و طراحی دارو از نو مورد بررسی و مقایسه قرار می‌گیرند. همچنین، چالش‌های کلیدی موجود از منظر مهندسی کامپیوتر، از جمله کمبود داده‌های باکیفیت، تفسیرپذیری مدل‌ها و تعمیم‌پذیری الگوریتم‌ها بحث خواهد شد. نتایج بررسی نشان می‌دهد که ادغام مدل‌های محاسباتی با فرآیندهای دارویی، می‌تواند زمان و هزینه‌های تحقیق و توسعه را به طور چشمگیری کاهش دهد.

کلمات کلیدی: هوش مصنوعی، کشف دارو، یادگیری عمیق، شیمی محاسباتی، غربالگری مجازی.

۱- مقدمه

صنعت داروسازی در دهه‌های اخیر با چالش‌های بزرگی در زمینه بهره‌وری و بازگشت سرمایه مواجه شده است. فرآیند سنتی کشف و توسعه یک داروی جدید، پروسه‌ای بسیار پیچیده، زمان‌بر و پرهزینه است. برآوردها نشان می‌دهد که به طور میانگین، توسعه یک داروی جدید از مرحله آزمایشگاه تا ورود به بازار، بین ۱۰ تا ۱۵ سال زمان می‌برد و هزینه‌ای بالغ بر ۲.۶ میلیارد دلار را به خود اختصاص می‌دهد (DiMasi et al., 2016). علاوه بر هزینه‌های سرسام‌آور، نرخ شکست در این صنعت بسیار بالا است؛ به طوری که حدود ۹۰ درصد از نامزدهای دارویی که وارد مرحله آزمایش‌های بالینی می‌شوند، به دلیل عدم اثربخشی یا سمیت بالا شکست می‌خورند (Sun et al., 2022).

این ناکارآمدی در روش‌های سنتی، نیاز به رویکردهای نوین محاسباتی را بیش از پیش آشکار کرده است. با پیشرفت تکنولوژی‌های "توان‌بالا" و توالی‌یابی ژنتیکی، حجم عظیمی از داده‌های بیولوژیکی و شیمیایی تولید شده است که تحلیل آن‌ها از توان روش‌های آماری کلاسیک و ذهن انسان خارج است. این حجم از داده، که اغلب تحت عنوان کلان‌داده شناخته می‌شود، بستر مناسبی را برای به کارگیری تکنیک‌های هوش مصنوعی و یادگیری ماشین فراهم کرده است (Vamathevan et al., 2019).

از دیدگاه مهندسی کامپیوتر، مسئله کشف دارو را می‌توان به عنوان یک مسئله بهینه‌سازی چندهدفه بازنمایی الگودر فضای بسیار بزرگ شیمیایی در نظر گرفت. فضای شیمیایی شامل بیش از 10^{60} مولکول دارویی بالقوه است که جستجوی کورکورانه در آن عملاً غیرممکن است (Polishchuk et al., 2013). الگوریتم‌های هوش مصنوعی، به ویژه روش‌های یادگیری عمیق،

این پتانسیل را دارند که با استخراج ویژگی‌های پیچیده از ساختارهای مولکولی، فضای جستجو را محدود کرده و خواص دارویی را با دقت بالا پیش‌بینی کنند.

در سال‌های اخیر، معماری‌های نوین شبکه‌های عصبی نظیر شبکه‌های عصبی گراف و مدل‌های مولد، پارادایم کشف دارو را از "طراحی مبتنی بر قواعد فیزیکی" به "طراحی مبتنی بر داده" تغییر داده‌اند (Chen et al., 2018). هدف این مقاله، ارائه یک مرور محاسباتی بر تکنیک‌های مختلف هوش مصنوعی است که در مراحل مختلف زنجیره کشف دارو، از جمله غربالگری مجازی، طراحی دارو از نو و پیش‌بینی سمیت مورد استفاده قرار می‌گیرند.

۲- دسته‌بندی روش‌های محاسباتی

برای به‌کارگیری هوش مصنوعی در کشف دارو، ابتدا باید مسئله بازنمایی مولکولی حل شود و سپس مدل مناسب بر اساس حجم داده و نوع مسئله انتخاب گردد. این بخش به بررسی نحوه تبدیل مولکول‌ها به داده‌های قابل پردازش و سپس مقایسه روش‌های کلاسیک و عمیق می‌پردازد.

۲-۱- بازنمایی داده‌ها، از مولکول تا بردار

یکی از چالش‌های اصلی در شیموآنفورماتیک، تبدیل ساختار شیمیایی گسسته به فرمت ریاضی قابل فهم برای الگوریتم‌ها است. به طور کلی، دو رویکرد اصلی برای بازنمایی داده‌ها وجود دارد:

- **بازنمایی مبتنی بر رشته:** در این روش، ساختار مولکول به صورت یک رشته متنی خطی کدگذاری می‌شود (مانند CCO برای اتانول). (این روش برای مدل‌های پردازش زبان طبیعی بسیار کارآمد است (Weininger, 1988)).
- **بازنمایی مبتنی بر گراف:** از دیدگاه ریاضی، یک مولکول به طور طبیعی یک گراف است که در آن اتم‌ها به عنوان گره و پیوندها به عنوان یال در نظر گرفته می‌شوند. این بازنمایی غنی‌ترین اطلاعات ساختاری را حفظ می‌کند و ورودی اصلی شبکه‌های عصبی گراف است.

۲-۲- یادگیری ماشین کلاسیک

تا پیش از ظهور یادگیری عمیق، الگوریتم‌های کلاسیک یادگیری ماشین استاندارد طلایی صنعت بودند. این روش‌ها عمدتاً بر «مهندسی ویژگی دستی» تکیه دارند؛ به این معنی که ابتدا باید خواص فیزیکی-شیمیایی مولکول (مانند وزن مولکولی، تعداد حلقه‌ها و غیره) محاسبه شده و در قالب یک بردار اثرانگشت به مدل داده شود.

- **ماشین بردار پشتیبان:** برای مسائل طبقه‌بندی دوتایی (مانند سمی/غیرسمی) با داده‌های محدود بسیار موثر است.
- **جنگل تصادفی:** این الگوریتم به دلیل قابلیت مدیریت داده‌های نویزی و جلوگیری از بیش‌برازش، یکی از محبوب‌ترین روش‌ها در پیش‌بینی فعالیت دارویی بوده است (Svetnik et al., 2003).

مزیت اصلی روش‌های کلاسیک، تفسیرپذیری بالاتر آن‌ها و عملکرد خوب روی مجموعه داده‌های کوچک (کمتر از چند هزار نمونه) است.

۲-۳- یادگیری عمیق، تغییر پارادایم به یادگیری بازنمایی

با افزایش حجم داده‌های شیمیایی، روش‌های کلاسیک دچار پدیده اشباع عملکرد شدند. یادگیری عمیق با حذف نیاز به مهندسی ویژگی دستی، رویکرد «یادگیری بازنمایی» را معرفی کرد. در این روش، شبکه عصبی خود ویژگی‌های مهم مولکول را از داده‌های خام استخراج می‌کند.

- شبکه‌های عصبی عمیق: برای مدل‌سازی روابط غیرخطی پیچیده بین ساختار شیمیایی و خواص بیولوژیکی استفاده می‌شوند (Hinton et al., 2012).
- شبکه‌های عصبی گراف: این معماری پیشرفته‌ترین ابزار فعلی در کشف دارو است. شبکه‌های عصبی گراف با استفاده از مکانیزم «پاس‌کاری پیام» اطلاعات همسایگی اتم‌ها را جمع‌آوری کرده و یک بردار ویژگی غنی برای کل مولکول می‌سازد. مطالعات نشان داده‌اند که GNN‌ها در پیش‌بینی خواص کوانتومی و برهم‌کنش دارو-پروتئین عملکردی فراتر از تمام روش‌های قبلی دارند (Gilmer et al., 2017).

۳- کاربردهای نوین: از غربالگری تا خلق دارو

با بلوغ روش‌های یادگیری عمیق، کاربرد هوش مصنوعی در کشف دارو از فاز «تحلیلی» به فاز «مولد» تغییر یافته است. این بخش به سه کاربرد اصلی می‌پردازد: طراحی دارو از نو، غربالگری مجازی و پیش‌بینی خواص دفع و سمیت.

۳-۱- طراحی دارو از نو با مدل‌های مولد

روش‌های سنتی کشف دارو معمولاً بر غربالگری کتابخانه‌های موجود متکی بودند، اما فضای شیمیایی شامل بیش از 10^{60} مولکول ممکن است که بررسی همه آن‌ها غیرممکن است. مدل‌های مولد با رویکردی معکوس، به جای جستجو در انبار کاه، خود سوزن را می‌سازند.

- **رمزنگار-رمزگشای خودکار:** این مدل‌ها فضای گسسته شیمیایی را به یک فضای پیوسته تبدیل می‌کنند. با حرکت در این فضای برداری، می‌توان مولکول‌های جدیدی تولید کرد که خواص ترکیبی دارند. گومز-بومبارلی و همکاران نشان دادند که استفاده از می‌تواند فرآیند بهینه‌سازی مولکول را تا ۷۰٪ سریع‌تر کند (Gómez-Bombarelli et al., 2018).
- **شبکه‌های مولد تخصصی:** در این معماری، دو شبکه (تولیدکننده و تفکیک‌کننده) با هم رقابت می‌کنند. تولیدکننده سعی می‌کند مولکول‌های واقعی بسازد و تفکیک‌کننده سعی می‌کند جعلی بودن آن‌ها را تشخیص دهد. نتیجه این رقابت، تولید ساختارهای شیمیایی جدید با اعتبار بالا و خواص دارویی مطلوب است (Kadurin et al., 2016).

۳-۲- غربالگری مجازی هوشمند

غربالگری مجازی فرآیند ارزیابی کامپیوتری میلیون‌ها ترکیب برای یافتن آن‌هایی است که به یک هدف پروتئینی خاص متصل می‌شوند. شبیه‌سازی‌های فیزیکی دقیق (مانند دینامیک مولکولی) بسیار کند هستند. مدل‌های یادگیری عمیق، به‌ویژه شبکه‌های عصبی کانولوشنی، توانسته‌اند این فرآیند را تسریع کنند. در این روش، برهم‌کنش دارو و پروتئین به صورت یک تصویر سه‌بعدی به شبکه داده می‌شود و شبکه احتمال اتصال را پیش‌بینی می‌کند. این رویکرد دقت پیش‌بینی را نسبت به توابع امتیازدهی کلاسیک به طور قابل توجهی افزایش داده است (Wallach et al., 2015).

۳-۳- پیش‌بینی خواص فارماکوکینتیک

یکی از دلایل اصلی شکست داروها در فازهای بالینی، مشکلات مربوط به جذب، توزیع، متابولیسم، دفع و سمیت است. پیش‌بینی این خواص در آزمایشگاه بسیار پرهزینه است. الگوریتم‌های یادگیری چندوظیفه‌ای در اینجا بسیار موفق عمل کرده‌اند. به جای آموزش یک مدل برای هر خاصیت، یک شبکه عصبی بزرگ همزمان چندین خاصیت (مثلاً سمیت کبدی و حلالیت در آب) را یاد می‌گیرد. این کار باعث می‌شود مدل از

اشتراکات بین داده‌ها استفاده کرده و حتی برای خواصی که داده کمی دارند، پیش‌بینی دقیق‌تری ارائه دهد (Ramsundar et al., 2015).

۴- چالش‌های باز و مسیر آینده

علیرغم پیشرفت‌های خیره‌کننده، ادغام کامل هوش مصنوعی در چرخه کشف دارو همچنان با موانع فنی جدی روبرو است. از دیدگاه علوم کامپیوتر، سه چالش اصلی عبارتند از:

۴-۱- کمبود داده و کیفیت آن

برخلاف حوزه‌هایی مانند بینایی ماشین که میلیون‌ها تصویر برچسب‌دار در دسترس است، در کشف دارو داده‌های آزمایشگاهی باکیفیت بسیار کمیاب و گران هستند. بسیاری از داده‌های موجود در پایگاه‌های عمومی دارای نویز هستند یا توزیع نامتوازی دارند (تعداد مولکول‌های فعال بسیار کمتر از مولکول‌های غیرفعال است). راه‌حل‌های آینده: استفاده از تکنیک‌های یادگیری با نمونه‌های کم و یادگیری انتقالی برای آموزش مدل‌های قدرتمند با داده‌های محدود، یکی از داغ‌ترین موضوعات پژوهشی فعلی است (Altae-Tran et al., 2017).

۴-۲- تفسیرپذیری مدل‌ها

شبکه‌های عصبی عمیق اغلب مانند یک «جعبه سیاه» عمل می‌کنند. برای یک شیمیدان دارویی، کافی نیست که مدل پیش‌بینی کند یک مولکول «سمی» است؛ او باید بداند «چرا» و «کدام زیرساختار» باعث سمیت شده است. عدم تفسیرپذیری مانع بزرگی برای اعتماد متخصصان بالینی به هوش مصنوعی است. مسیر آینده: توسعه روش‌های هوش مصنوعی قابل تفسیر که بتوانند تصمیمات مدل را با برجسته‌کردن بخش‌های مهم مولکول (مانند مکانیسم‌های Attention) توجیه کنند، ضروری است (Jiménez-Luna et al., 2020).

۴-۳- تعمیم‌پذیری خارج از توزیع

مدل‌های یادگیری ماشین معمولاً روی داده‌هایی خوب عمل می‌کنند که مشابه داده‌های آموزشی باشند. اما در کشف دارو، هدف کشف مولکول‌های کاملاً جدید در فضاها ناشناخته شیمیایی است. مدل‌های فعلی اغلب در مواجهه با ساختارهایی که از نظر شیمیایی با داده‌های آموزشی فاصله دارند، دچار افت شدید عملکرد می‌شوند (Guedj et al., 2022).

۵- نتیجه‌گیری

این مقاله مروری بر کاربرد تکنیک‌های نوین هوش مصنوعی در تسریع فرآیند کشف دارو داشت. بررسی‌ها نشان می‌دهد که گذار از «روش‌های مبتنی بر آزمون و خطا» به «روش‌های داده‌محور» اجتناب‌ناپذیر است. تکنیک‌هایی نظیر شبکه‌های عصبی گراف و مدل‌های مولد، پتانسیل کاهش زمان توسعه دارو از سال‌ها به ماه‌ها و کاهش هزینه‌ها را دارا هستند. با این حال، موفقیت نهایی در گرو غلبه بر چالش‌های مهندسی نظیر تفسیرپذیری مدل‌ها و مدیریت داده‌های پراکنده است. به نظر می‌رسد آینده این حوزه در گرو همکاری نزدیک بین مهندسان کامپیوتر و دانشمندان علوم زیستی و ایجاد مدل‌های هیبریدی (ترکیب دانش فیزیکی و یادگیری ماشین) باشد.



<https://icaics.ir>
info@icaics.ir

اولین کنفرانس بین المللی هوش مصنوعی
و علوم کامپیوتری نو ظهور: از الگوریتم تا آینده نگری
**First International Conference on Artificial Intelligence
and Emerging Computer Science: From Algorithm to Foresight**

March 17, 2026-GEORGIA

۲۶ اسفند ماه ۱۴۰۴ - گرجستان

مراجع

- Altae-Tran, H., Ramsundar, B., Pappu, A. S., & Pande, V. (2017). Low data drug discovery with one-shot learning. *ACS central science*, 3(4), 283-293.
- Chen, H., Engkvist, O., Wang, Y., Olivecrona, M., & Blaschke, T. (2018). The rise of deep learning in drug discovery. *Drug discovery today*, 23(6), 1241-1250.
- DiMasi, J. A., Grabowski, H. G., & Hansen, R. W. (2016). Innovation in the pharmaceutical industry: new estimates of R&D costs. *Journal of health economics*, 47, 20-33.
- Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., & Dahl, G. E. (2017). Neural message passing for quantum chemistry. International conference on machine learning,

- Gómez-Bombarelli, R., Wei, J. N., Duvenaud, D., Hernández-Lobato, J. M., Sánchez-Lengeling, B., Sheberla, D., Aguilera-Iparraguirre, J., Hirzel, T. D., Adams, R. P., & Aspuru-Guzik, A. (2018). Automatic chemical design using a data-driven continuous representation of molecules. *ACS central science*, 4(2), 268-276.
- Guedj, M., Swindle, J., Hamon, A., Hubert, S., Desvaux, E., Laplume, J., Xuereb, L., Lefebvre, C., Haudry, Y., & Gabarroca, C. (2022). Industrializing AI-powered drug discovery: Lessons learned from the patrimony computing platform. *Expert Opinion on Drug Discovery*, 17(8), 815-824.
- Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A.-r., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., & Sainath, T. N. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal processing magazine*, 29(6), 82-97.
- Jiménez-Luna, J., Grisoni, F., & Schneider, G. (2020). Drug discovery with explainable artificial intelligence. *Nature Machine Intelligence*, 2(10), 573-584.
- Kadurin, A., Aliper, A., Kazennov, A., Mamoshina, P., Vanhaelen, Q., Khrabrov, K., & Zhavoronkov, A. (2016). The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology. *Oncotarget*, 8(7), 10883.
- Polishchuk, P. G., Madzhidov, T. I., & Varnek, A. (2013). Estimation of the size of drug-like chemical space based on GDB-17 data. *Journal of computer-aided molecular design*, 27(8), 675-679.
- Ramsundar, B., Kearnes, S., Riley, P., Webster, D., Konerding, D., & Pande, V. (2015). Massively multitask networks for drug discovery. *arXiv preprint arXiv:1502.02072*.
- Sun, D., Gao, W., Hu, H., & Zhou, S. (2022). Why 90% of clinical drug development fails and how to improve it? *Acta Pharmaceutica Sinica B*, 12(7), 3049-3062.
- Svetnik, V., Liaw, A., Tong, C., Culberson, J. C., Sheridan, R. P., & Feuston, B. P. (2003). Random forest: a classification and regression tool for compound classification and QSAR modeling. *Journal of chemical information and computer sciences*, 43(6), 1947-1958.
- Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., Li, B., Madabhushi, A., Shah, P., & Spitzer, M. (2019). Applications of machine learning in drug discovery and development. *Nature reviews Drug discovery*, 18(6), 463-477.
- Wallach, I., Dzamba, M., & Heifets, A. (2015). AtomNet: a deep convolutional neural network for bioactivity prediction in structure-based drug discovery. *arXiv preprint arXiv:1510.02855*.
- Weininger, D. (1988). SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of chemical information and computer sciences*, 28(1), 31-36.