

[Artificia](https://doi.org/10.1016/j.ailsci.2022.100052)l[IntelligenceintheLifeSciences3(2023)100052](https://doi.org/10.1016/j.ailsci.2022.100052)

|  |  |  |
| --- | --- | --- |
|  | Contentslistsavailableat[ScienceDirect](http://www.ScienceDirect.com) |  |
| ArtificialIntelligenceintheLifeSciences |
| journalhomepage:[www.elsevier.com/locate/ailsci](http://www.elsevier.com/locate/ailsci) |

Editorial

Specificcontributionsofartificialintelligencetointerdisciplinarylife scienceresearch– exploringandcommunicatingnewopportunities

JürgenBajorath

*DepartmentofLifeScienceInformaticsandDataScience,B-IT,LIMESProgramUnitChemicalBiologyandMedicinalChemistry,Rheinische Friedrich-Wilhelms-Universität,Friedrich-Hirzebruch-Allee5/6,BonnD-53115,Germany*

|  |  |
| --- | --- |
| article | info |

*Keywords:*   
Artificialintelligence   
Deeplearning   
Lifesciences   
Drugdiscovery   
Interdisciplinaryresearch

**Artificialintelligencebetweenhypeandreality**

Artificialintelligence(AI)islaudedasanauspiciousproblemsolver inmanyareas.However,theunderstandingofAImethodsisoftenlim-ited.Hence,anauraofmystery–andalsoconcern– mightbegenerated aroundAI.Forexample,theexpectationthatmachineswould“think”independentlyandreachautonomousdecisionsbeyondhumanreason-ingisnotfactual.Inscience,thepopularityandpromiseofAImostly originatefromnotableadvancesinafewfields,butarealsoinfluenced bybusiness-drivenhypeandunrealisticexpectations.

As*ArtificialIntelligenceintheLifeSciences(AILSCI)*iscompletingits secondyear,thiscontributionaimstoputdevelopmentsinAIthatare particularlyrelevantforthejournalintoscientificperspective.Itisbased upon–andfurtherextends– tworecentopenaccesspublicationsaddress-ingawidelifescienceaudience[1,2].

Incomputerscience,variousdisciplinesarecoveredundertheterm AI[3].Amongthese,deeplearning(DL)usingdeepneuralnetworks (DNNs),asub-disciplineofmachinelearning(ML),hasbeenresponsible forrecentprogressinareassuchascomputervision(imageanalysis)or naturallanguageprocessing.Theseadvanceshavegreatlycontributedto thepopularityofAIinscience.Robotics,anotherAIdiscipline,isamain-stayinindustryandalsoplaysanimportantroleinlaboratoryautoma-tion.Furthermore,expertandrecommendersystems,whichalsobelong totheAIspectrum,areexploredindifferentscientificfields.Similar toAI-drivendevelopmentsinphysics,theoreticalbiology,orquantum chemistry,AIisbeginningtoimpactthelifesciencesincludingearly-phasedrugdiscoveryonalargerscale.Here,thetermAIisforthemost partsynonymouslyusedwithDLwhenappliedatinterfacesbetween computationandexperiment[2,4–6].Inmedicine,DLisemployedin

*E-mailaddress:*[bajorath@bit.uni-bonn.de](mailto:bajorath@bit.uni-bonn.de)

differenttherapeuticareas[7]suchasradiologyoroncology[8,9].In clinicalpractice,medicalimageanalysisrepresentsaprimegrowtharea forDL[8,10].

Forthesedevelopmentsinthegreaterlifesciencearena,*AILSCI*rep-resentsawell-positionedpublicationvenue.Oneof*AILSCI’s*corevalues isensuringhighscientificstandardsofpublications,includingmethod developmentandAIapplications.

**Characteristicsofdeepmachinelearning**

Generally,MLusesalgorithmsfortheextractionoffeaturepatterns fromtrainingdatatoclassifytestobjectsoraddressregressiontasks. Hence,MLmethodsarestatisticalinnatureandderivepredictivemod-elscapturinglinearornon-linearinstance-featurerelationshipsbasedon inferencefromdata.DNNsarewellsuitedforfeatureextractionfrom largevolumesofunstructureddata(suchaspixelsinimages)andfor learningnewobjectrepresentations.DLreliesonsystematiccorrelation offeaturepatternsandknownclasslabelsandderivesmodelswithde-cisionfunctionsthatarenotpre-programmed.Hence,thereisnothing mysteriousaboutthistypeofsupervised“machineintelligence”. ShallowNNswerepopularduringtheearlystagesofMLinbiology, chemistry,anddrugdiscovery,butwerelargelyreplacedbyotherap-proachessuchasdecisiontreemethods(randomforest,gradientboost-ing),Bayesianmodeling,orsupportvectormachines.Thiswaslargely duetoageneraltendencyofshallowNNstooverfitmodelstotrain-ingdataandtheirhighsensitivitytovaryingparametersettings.The increasinglypopularsecond-generationDNNsrepresenthighlyversa-tilecomputationalarchitectures.Incomputerscience,agreatvarietyof DNNsandassociatedlearningstrategieshavebeenintroduced,some-

<https://doi.org/10.1016/j.ailsci.2022.100052>  
[Received7December2022;Accepted9Dece](https://doi.org/10.1016/j.ailsci.2022.100052)mber2022   
Availableonline11December2022   
[2667-3185/© 2022TheAuthors.PublishedbyElsevi](http://creativecommons.org/licenses/by-nc-nd/4.0/)erB.V.ThisisanopenaccessarticleundertheCCBY-NC-NDlicense (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

*J.Bajorath*  *ArtificialIntelligenceintheLifeSciences3(2023)100052*

timesdescribedwithtermslikeanetwork“jungle” or“zoo” [11].This architecturalvarietyhasprosandcons.Formanyapplications,alterna-tiveDNNscanbeconsidered,butfindingpreferredsolutionsisnotnec-essarilystraightforward.Moreover,complexDNNmodelsareoftende-rivedwithoutdemonstratingthattheircomplexityisindeedrequiredfor thepredictionstasksathand.ComparedtootherMLapproaches,DNNs areparticularlyrichinhyper-parametersandderivationofDNNmodels requiressubstantialknowledge,skills,andexperience.Accordingly,al-thoughpublicdomainsoftwareisavailableforconstructingDNNs,DL isnotanapproachthatisreadilyaccessibletonon-experts.Thereis astrongdiscrepancybetweenthemechanicsofmodelbuilding,which mightbehandledbylessexperiencedusers,andtheevaluationofre-sultsandrecognitionofpotentialcaveatsormodelerrors,whichre-quiresmuchmoreexpertise.Importantly,similartoother–butnotall–MLmethods,DNNshavenotorious“blackbox” character[12],meaning thatitisnottransparenthowthesemodelsreachtheirdecisions.The blackboxofDNNsisamajorissueinlifescienceanddrugdiscovery applications,asfurtherdiscussedbelow.

**Dataheterogeneity**

Lifescienceanddrugdiscoverydataarehighlyheterogeneousin termsofvolumes,composition,andcomplexity.Early-phasedrugdis-coveryconcentratingontargetvalidation,bioassays,compounds,and activityassessmentisnotadata-richdisciplinecomparedtootherar-easwhereDLhasmadeastrongimpact.Inearly-phasedrugdiscovery, datasetsfrommedicinalchemistryaretypicallyconfinedtotestresults forcompoundseriesandthereforelimitedinsize.Thisalsoappliesto datasetsfrom,forexample,probeinvestigationsinchemicalbiology, timeseriesexperimentsinbiology,orconfirmatoryassaysinbiological screening.Aconsequenceofdataheterogeneityandsparsenessisthat sufficientlylargedatasetsfor“hungry” DNNsareoftennotavailable. Moreover,informaticsapproachesinthelifescienceshavetraditionally employedpre-definedobject(forexample,targetorcompound)repre-sentations(descriptors)andnotreliedonrepresentationlearning.To furthercomplicatematters,effectiveML/DLmodelscanalsobegener-atedonthebasisofverysmalldatasets[13].Hence,therearemany incentivestoundertakeexpeditionsintotheDNNjungle,furtherana-lyzelearningcharacteristicsofdifferentmethods,andcompareML/DL modelsofdifferentcomplexity.

**Artificialintelligenceinmedicinalchemistry**

Asanexemplaryfieldwithmanyopportunitiesforpracticalapplica-tions,onemayhaveacloserlookatthestate-of-theartofAIinmedicinal chemistry,acoredisciplineofearly-phasedrugdiscovery.

MLalreadyhasalonghistoryinmedicinalchemistry,atraditionally conservativediscipline.Formorethantwodecades,MLmethodshave beenusedforcompoundpropertypredictionsandotherapplications. Inmedicinalchemistry,propertiesofinterestforcomputationalstud-iesinclude,firstandforemost,biologicalactivitiesofsmallmolecules, butalsophysiochemical(e.g.solubility)or*invivo*properties(suchas metabolicstabilityortoxicity).Predictionsofsuchpropertiesaimto supportthekeytaskinthepracticeofmedicinalchemistry,thatis,de-cidingwhichcompound(s)tosynthesizenext.Overtime,shallowNNs thatwerepopularearlyonforpropertypredictionswereforthemost partreplacedbyotherMLmethods,asdiscussedabove.Importantly,in medicinalchemistry,chemicalintuition,experience,andsubjectivede-cisionscontinuetoplayamajorrole.Accordingly,blackboxpredictions thatcannotbeexplainedinchemicaltermsworkagainsttheacceptance ofMLforpracticalapplications.However,thepopularityofDNNsand highexpectationsassociatedwithDLarealsochangingcomputational medicinalchemistry.

2

*J.Bajorath*  *ArtificialIntelligenceintheLifeSciences3(2023)100052*

asbigdata.However,formedicinalchemistry,whichistraditionally notdata-driven,thesedatavolumesarechallenging.Ontheotherhand, data-drivenapproachesprovidenewopportunitiesforthefurtherdevel-opmentofmedicinalchemistryasascientificdiscipline[24].

Thesituationisdifferentforpredictivemodeling.Inmedicinalchem-istry,MLismostlyappliedatthelevelofindividualtarget-directed projectsusingrelativelysmalldata.Eachoftheseprojectsprovidesa specificcontextformodeling.Indatascience,thecontextdependence ofdatastructuringandanalysisisknowntoworkagainstgeneraliza-tionofknowledgeextraction,whichrequiresabstractionfromproject-baseddatasetsandproject-specificanalysiscriteria[25].Bycontrast,in medicinalchemistry,projectfocustakescenterstageandconfinesthe applicabilityofML.Furthermore,thepredominantsmalldataframe-workinmedicinalchemistryalsosuggestsalternativestrategiesforML. Ratherthanheavilyinvestigatingmethodologieswhosestrengthsde-penduponlargedatavolumes,approachessuchastransferlearning [26]oractivelearning[27,28]canbeappliedthatarecapableofpre-dictingmolecularpropertiesornewcompoundsonthebasisofsparse data.Transferlearningmakesitpossibletousedatafromrelatedpre-dictiontasks(targets)formodeling;activelearningderivespredictive modelsfromminimalsetsofinformativetraininginstances.Inmedici-nalchemistry,theseapproachesareparticularlyrelevantforaddressing noveltargetswithinterestingdiseasebiologyforwhichonlylimited compoundinformationisavailable.Allinall,thereismuchroomfor furthercomputationaldevelopmentswithpracticalutilityformedicinal chemistry.

**Modelimpactandacceptance**

Returningtothegreaterlifesciencearena,thereareotherareas whereDNNshaveachievedunprecedentedadvancessuchasindenovo proteinstructureprediction[29].Regardless,DLwillultimatelyonly becomeanintegralpartofinterdisciplinarylifescienceresearchifit measurablyimpactsexperimentalprograms.Importantly,furtherestab-lishingDLininterdisciplinarysettingsisonlypossibleiflifesciencein-vestigatorsanddrugdiscoverypractitionersagreetorelyonpredictions forexperimentaldesign.Thisrequiresincreasingmodelacceptancein interdisciplinaryresearch.Asisthecasewithanynewtechnology,time willberequireduntilDLcanrealizeitspotentialinthisarea.However, therearespecificrequirementsthatmustbemettofurtherincreasethe confidenceofexperimentalistsinpredictivemodels.

**Rationalizingpredictions**

Experimentalistsarenaturallyreluctanttorelyonpredictionsthat aredifficultorimpossibletounderstand.Giventheblackboxnatureof DNNs,thispresentsamajorobstaclefortheacceptanceofsuchmodels forexperimentaldesign.Therefore,increasingattentionisbeingpaidto approachesfor“explainableAI” (XAI)thatmakeitpossibletorational-izetheresultsofML/DLmodelsandinterpretpredictionsinchemicalor biologicalterms[30,31].Amongothers,theseincludemethodsforthe identificationoffeaturesmakinglargestcontributionstoindividualpre-dictionsorthedeterminationoffeaturesetsthatareminimallyrequired toproduceanaccurateprediction.CloselyrelatedtoXAIapproaches aremethodstoquantifytheuncertaintyofpredictions[32–34].Obtain-inguncertaintyestimatesalsohelpstobuildconfidenceinpredictive modeling.AlthoughthereareMLapproachesthatyieldpredictionun-certainties,forexample,probabilistic(Bayesian)modeling[34],most methodsincludingDNNsproduceendpointswithoutuncertaintyesti-mates,whicharesubjecttofurtheranalysis.

**Prospectiveapplications**

TheultimateassessmentofthepotentialofML/DLforthelifesci-encesdependsonprospectiveapplications,thatis,predictionsleading

3

*J.Bajorath*  *ArtificialIntelligenceintheLifeSciences3(2023)100052*

[[8]HosnyA,ParmarC,QuackenbushJ,SchwartzLH,AertsH.Artificia](http://refhub.elsevier.com/S2667-3185(22)00022-8/sbref0008)l[intelligencein](http://refhub.elsevier.com/S2667-3185(22)00022-8/sbref0008)  [radiology.NatRevCancer2018;18:500–10.](http://refhub.elsevier.com/S2667-3185(22)00022-8/sbref0008)

[[9]FarinaE,NabhenJJ,DacoregioMI,BataliniF,MoraesFY.Anoverviewofartificia](http://refhub.elsevier.com/S2667-3185(22)00022-8/sbref0008)l [intelligenceinoncology.FutureSciOA2022;8:FSO787.](http://refhub.elsevier.com/S2667-3185(22)00022-8/sbref0010)

[[10]ShenD,WuG,SukHI.Deeplearninginmedicalimageanalysis.AnnRevBiomed](http://refhub.elsevier.com/S2667-3185(22)00022-8/sbref0010)  [Eng2017;19:221–48.](https://www.asimovinstitute.org/neural-network-zoo/)

[[11]vanVenn,F.Theneuralnetworkzoo(2016).https://www.asimovinstitute.](https://www.asimovinstitute.org/neural-network-zoo/)  [org/neural-network-zoo/.](https://www.asimovinstitute.org/neural-network-zoo/)

[[12]CastelvecchiD.CanweopentheblackboxofAI?Nature2016;538:20–3.](https://www.asimovinstitute.org/neural-network-zoo/)

[[13]](http://refhub.elsevier.com/S2667-3185(22)00022-8/sbref0013)[SiemersFM,FeldmannC,BajorathJ.Minimaldatarequirementsforaccu](http://refhub.elsevier.com/S2667-3185(22)00022-8/sbref0012)[ratecom-poundactivitypredictionusingmachinelearningmethodsofdifferentcomplexity.](http://refhub.elsevier.com/S2667-3185(22)00022-8/sbref0013) [CellRepPhysSci2022;3:101113.](http://refhub.elsevier.com/S2667-3185(22)00022-8/sbref0014)

[[14]RudinC.Stopexplainingblackboxmachinelearningmodelsforhighstakesdeci-](http://refhub.elsevier.com/S2667-3185(22)00022-8/sbref0014) [sionsanduseinterpretablemodelsinstead.NatMachIntell2019;1:206–15.](http://refhub.elsevier.com/S2667-3185(22)00022-8/sbref0014)

[[15]MeyersJ,FabianB,BrownN.Denovomoleculardesignandgenerativemodels.](http://refhub.elsevier.com/S2667-3185(22)00022-8/sbref0014)  [DrugDiscovToday2021;26:2707–15.](http://refhub.elsevier.com/S2667-3185(22)00022-8/sbref0015)

[[16]TongX,LiuX,TanX,LiX,JiangJ,XiongZ,XuT,JiangH,QiaoN,ZhengM.](http://refhub.elsevier.com/S2667-3185(22)00022-8/sbref0015) [Generativemodelsfordenovodrugdesign.JMedChem2021;64:14011–27.](http://refhub.elsevier.com/S2667-3185(22)00022-8/sbref0016)

[[17]DeAlmeidaAF,MoreiraR,RodriguesT.Syntheticorganicchemistrydrivenbyar-](http://refhub.elsevier.com/S2667-3185(22)00022-8/sbref0016) [tificialintelligence.NatRevChem2019;3:589–604.](http://refhub.elsevier.com/S2667-3185(22)00022-8/sbref0017)

[[18]StrubleTJ,AlvarezJC,BrownSP,ChytilM,CisarJ,DesJarlaisRL,EngkvistO,](http://refhub.elsevier.com/S2667-3185(22)00022-8/sbref0017) [FrankSA,GreveDR,GriffinDJ,HouX,JohannesJW,KreatsoulasC,LahueB, MatheaM,MogkG,NicolaouCA,PalmerAD,PriceDJ,RobinsonRI,SalentinS, XingL,JaakkolaT,GreenWH,BarzilayR,ColeyCW,JensenKF.Currentandfu-turerolesofartificialintelligenceinmedicinalchemistrysynthesis.JMedChem](http://refhub.elsevier.com/S2667-3185(22)00022-8/sbref0018) [2020;63:8667–82.](http://refhub.elsevier.com/S2667-3185(22)00022-8/sbref0019)

[[19]BortW,BaskinII,GimadievT,MukanovA,NugmanovR,SidorovP,MarcouG,Hor-vathD,KlimchukO,MadzhidovT,VarnekA.Discoveryofnovelchemicalreactions bydeepgenerativerecurrentneuralnetwork.SciRep2021;11:1–15.](http://refhub.elsevier.com/S2667-3185(22)00022-8/sbref0019)

[[20]StokesJM,YangK,SwansonK,JinW,Cubillos-RuizA,DonghiaNM,McNairCR,](http://refhub.elsevier.com/S2667-3185(22)00022-8/sbref0019) [FrenchS,CarfraeLA,Bloom-AckermannZ,TranVM,Chiappino-PepeA,BadranAH, AndrewsIW,ChoryEJ,ChurchGM,BrownED,JaakkolaTS,BarzilayR,CollinsJJ. Adeeplearningapproachtoantibioticdiscovery.Cell2020;180:688–702.](http://refhub.elsevier.com/S2667-3185(22)00022-8/sbref0020)

[[21]WaltersWP,MurckoM.AssessingtheimpactofgenerativeAIonmedicinalchem-](http://refhub.elsevier.com/S2667-3185(22)00022-8/sbref0020) [istry.NatBiotechnol2020;38:143–5.](http://refhub.elsevier.com/S2667-3185(22)00022-8/sbref0021)

4