

[Artificia](https://doi.org/10.1016/j.ailsci.2023.100060)l[IntelligenceintheLifeSciences3(2023)100060](https://doi.org/10.1016/j.ailsci.2023.100060)

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

ResearchArticle   
Combiningmolecularandcellpaintingimagedataformechanismofaction prediction   
GuangyanTiana,1,PhilipJHarrisona,1,∗,AkshaiPSreenivasana,b,JordiCarreras-Puigverta, OlaSpjutha   
a*DepartmentofPharmaceuticalBiosciences,UppsalaUniversity,Sweden*   
b*DepartmentofMedicalSciences,UppsalaUniversity,Sweden*

|  |  |  |
| --- | --- | --- |
| article | info | abstract |
| *Keywords:*  Bioinformatics  Convolutionalneuralnetworks Cheminformatics  Deeplearning  Machinelearning | | Themechanismofaction(MoA)ofacompounddescribesthebiologicalinteractionthroughwhichitproduces apharmacologicaleffect.MultipledatasourcescanbeusedforthepurposeofpredictingMoA,includingcom-poundstructuralinformation,andvariousassays,suchasthosebasedoncellmorphology,transcriptomicsand metabolomics.Inthepresentstudyweexploredthebenefitsandpotentialadditive/synergisticeffectsofcom-biningstructuralinformation,intheformofMorganfingerprints,andmorphologicalinformation,intheform offive-channelCellPaintingimagedata.Forasetof10wellrepresentedMoAclasses,wecomparedtheperfor-manceofdeeplearningmodelstrainedonthetwodatasetsseparatelyversusamodeltrainedonbothdatasets simultaneously.Onaheld-outtestsetweobtainedamacro-averagedF1scoreof0.58whentrainingononly thestructuraldata,0.81whentrainingononlytheimagedata,and0.92whentrainingonbothtogether.Thus indicatingclearadditive/synergisticeffectsandhighlightingthebenefitofintegratingmultipledatasourcesfor MoAprediction. |

**Introduction**

Mechanismofaction(MoA)referstothebiologicalinteraction throughwhichapotentiallytherapeuticsmall-moleculecompoundpro-ducesapharmacologicaleffect,suchasthespecificproteinsthatthe compoundtargetsandthepathwaysthatitmodulates.Uncoveringthe MoAofacompound,althoughasignificantchallengeinchemicalbi-ology[1],providesextremelyusefulinformationforleadcompounds priortoclinicaltrialsandforidentifyingpossibletoxicityorside-effects [2].

Avarietyofdifferentdatasourcescanbeusedtocaptureinfor-mationonacompoundsMoA,includingstructuralinformationfrom thecompound,geneexpressionfromtranscriptomicsdata,proteinin-formationfromproteomicsdata,andmetabolicenzymeactivityfrom metabolomicsdata[2].Recently,cellmorphologydatafromhigh-contentimaginghasprovenusefulforthistask[3].Asignificantbenefit ofmicroscopybasedimageassaysisthattheycanbescaledtohigh-throughputmuchmoreeasilyandlessexpensivelythantranscriptomics andmetabolomicsbasedassays[4].Cellimagingalsoprovidesinfor-mationatthesingle-cellresolutionasopposedtocondensingtheoutput downtomeasuresofpopulationaverages[5].Intermsofthroughput

andefficiencytheL1000[6]geneexpressionassayisperhapscurrently theonlyfeasiblealternativetoimage-basedassays[7]forlargescale datagenerationtosustainpredictivemodeling.

Microscopyimagingcanbeusedtocapturethechangesincell morphologythatarisewhenacellcultureistreatedwithachemi-calcompound[2].TheCellPaintingassayusesfluorescentdyesto paintthecellsinmulti-wellplatesas"richlyaspossible"toilluminate morphologicalchangesineightbroadlyrelevantorganellesandcellu-larsub-compartments(nuclei,mitochondria,cytoskeleton,Golgiappa-ratus,plasmamembrane,cytoplasmicRNA,nucleoliandendoplasmic reticulum)usingsixfluorescentdyesimagedinfivechannels[7]. Acomparativestudyforlibraryenrichmentreportedbetterpredic-tivepowerforHigh-throughputscreeningperformanceusingCellPaint-ingasopposedtoL1000geneexpressionprofiling[8].Whereas,forpre-dictingMoA,Wayetal.[9]foundthatL1000outperformedCellPaint-ing,butthattherewascomplementarity,i.e.someMoAswerebetter predictedbyoneoftheassayscomparedtotheother.Arelatedstudy byLapinsandSpjuth[10]comparedCellPainting,L1000andchemical structurebasedpredictors,andfoundMoAclassesthatwerepredicted betterbyeachofthethreepredictorsrelativetotheothertwo,support-ingtheideaofalikelybenefitthroughcombiningthesedifferentdata

∗Correspondinga[uthor.](mailto:philip.harrison@farmbio.uu.se)

*E-mailaddresses:*[philip.harrison@farmbio.uu.se](mailto:philip.harrison@farmbio.uu.se)(P.J.Harrison),[ola.spjuth@farmbio.uu.se](mailto:ola.spjuth@farmbio.uu.se)(O.Spjuth). 1Theseauthorsco[ntributedequallytothiswork.](mailto:philip.harrison@farmbio.uu.se)

<https://doi.org/10.1016/j.ailsci.2023.100060>  
[Received8December2022;Receivedinrevis](https://doi.org/10.1016/j.ailsci.2023.100060)edform5January2023;Accepted30January2023   
Availableonline17February2023   
2667-3185/© 2023TheAuthors.PublishedbyElsevierB.V.ThisisanopenaccessarticleundertheCCBYlicense(<http://creativecommons.org/licenses/by/4.0/>)

*G.Tian,P.J.Harrison,A.P.Sreenivasanetal.*

sources.AnotherstudypredictingMoA[11],basedondatafromthe ExCAPEdatabase,whichcomparedmodelsbuiltusingimagebasedfea-turestothosebuiltusingchemicalstructuredescriptors,providedfur-thersupportforthecomplementarityofthesetwodatatypes,whereby themodelsperformedsomewhatdifferentlyatanindividualclasslevel. Besidescomparingmodelsbuiltusingdifferenttypesofdataitis alsopossibletocombinethedatasetsandanalyzethemsimultaneously tosearchforadditiveorsynergisticeffects.Forpredictingcytotoxicity andproliferationSealetal.[12]comparedRandomForestmodelsus-ingCellPaintingimagebasedfeatures,molecularfingerprints,andcom-biningbothdatasources.Theyfoundthatthemodelsbasedonimage featuresoutperformedthosebasedonmolecularfingerprints,butthe combinedmodelsperformedbestintenoutoftwelvecases.Another studypredictingthebio-activityofapproximately16,000compounds [13],foundthatmodelsbasedonfeaturesderivedfromCellPainting imagesoutperformedthosebasedonchemicalstructureprofilesfrom graphconvolutionalnetworks(GCNs,[14]),butthatthefusionofthe twodatasetsgaveagaininperformance.

Mosttraditionalimageanalysispipelines,includingthosementioned above,firstextractmorphologicalfeaturesfromthefluorescencestained images,includingmeasuresofsize,shape,intensityandtexturefrom thelabeledcellularcompartments,mostoftenusingtheCellProfiler [15]softwarepackage,andsubsequentlyapplymachinelearningmeth-odstotheextractedfeaturesforthepredictivetaskathand[16].These methodsrequireanaccuratesegmentationalgorithmtoidentifythecel-lularcompartmentspriortofeatureextraction.However,whenconvolu-tionalneuralnetworks(CNNs)areusedontherawimages,featuresare extractedinanautomaticdata-drivenfashion,circumventingtheneed forcellsegmentationandpotentiallyprovidingbetterpredictiveper-formance[3,17].Forinstance,Hofmarcheretal.[18]foundthatCNNs trainedonCellPaintingimagedata,forpredictingactivitylabelsfor over10,000compounds,performedsignificantlybetterthanfullycon-nectedneuralnetworkstrainedonpre-computedimagefeatures.The flexibilityofthearchitecturalchoicesforneuralnetworksalsoprovides asimplemeansofcombiningmultipledatasourcesintothesamemod-elingframework[19].

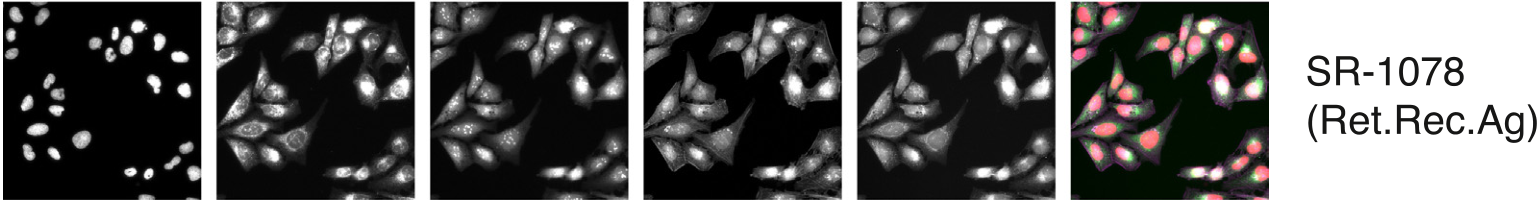
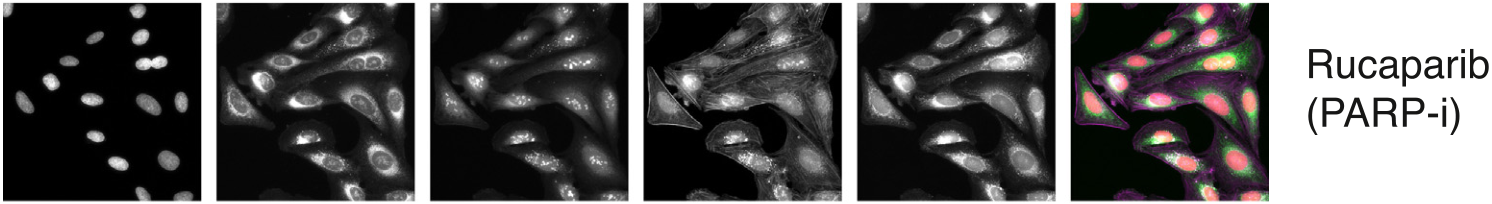
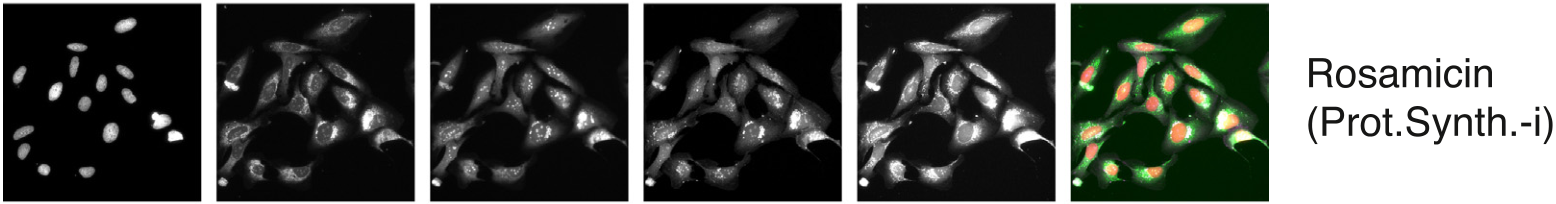
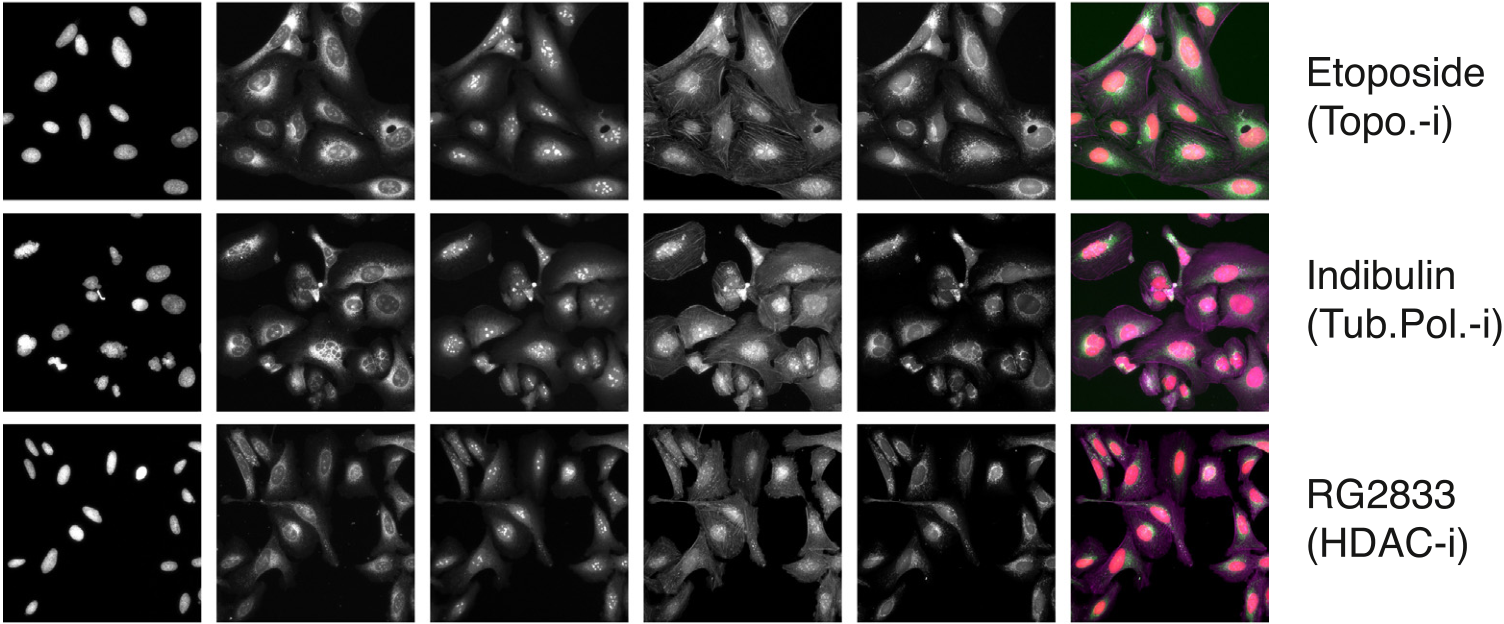
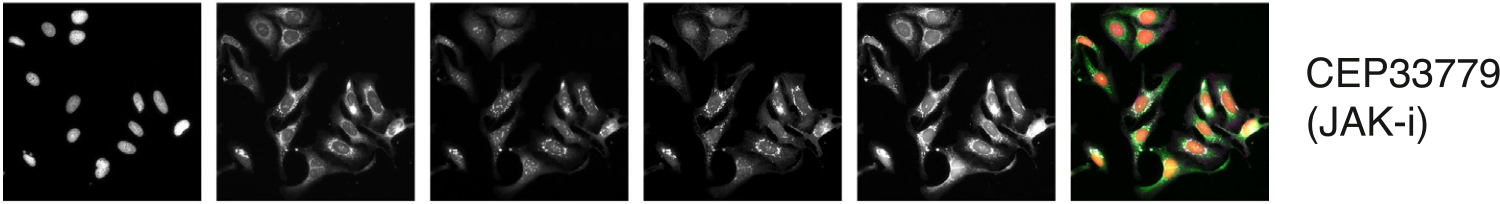
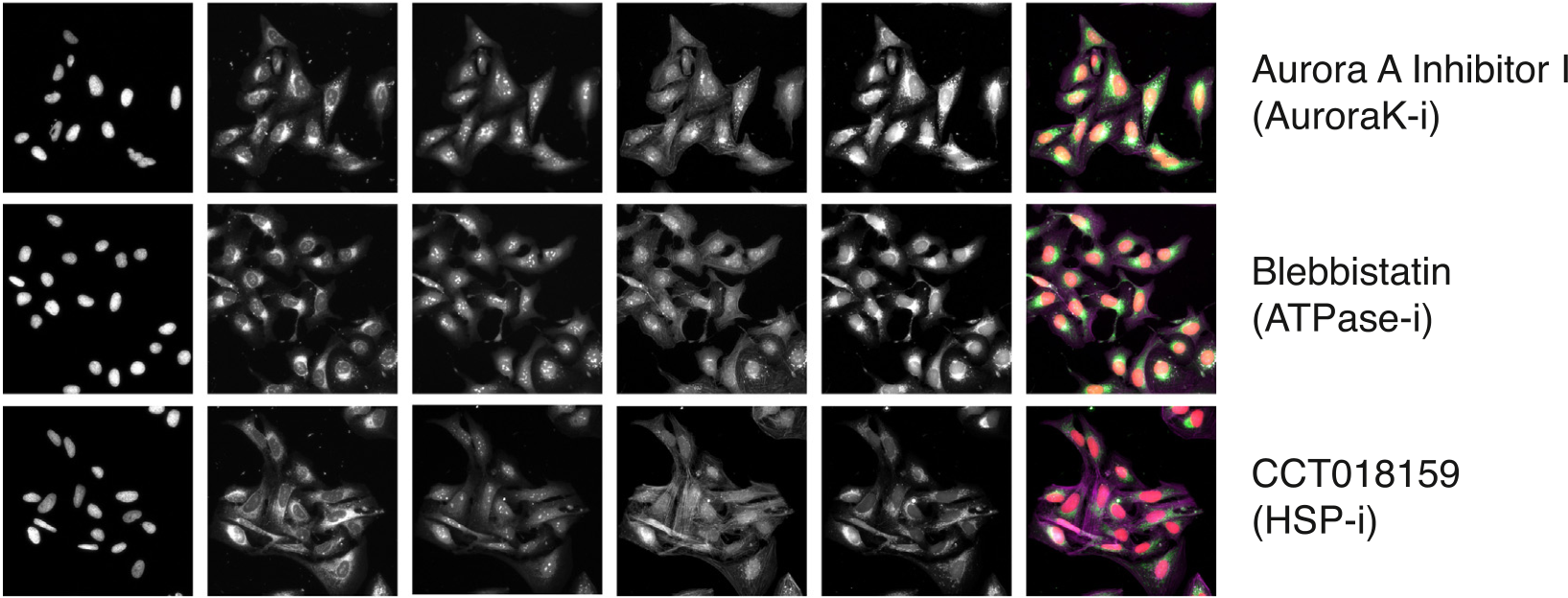
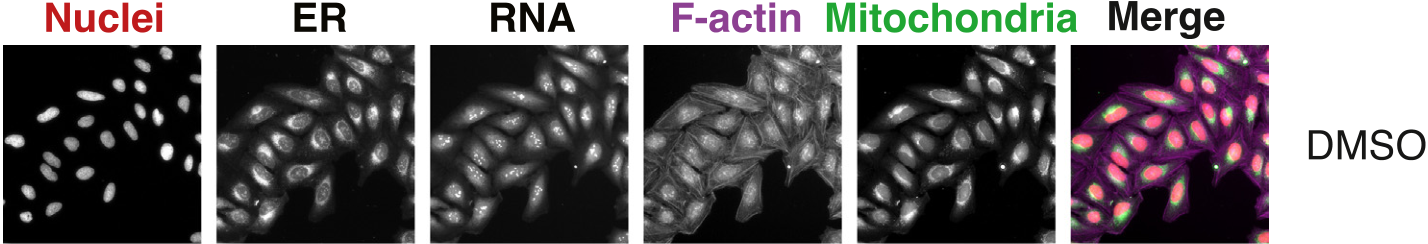
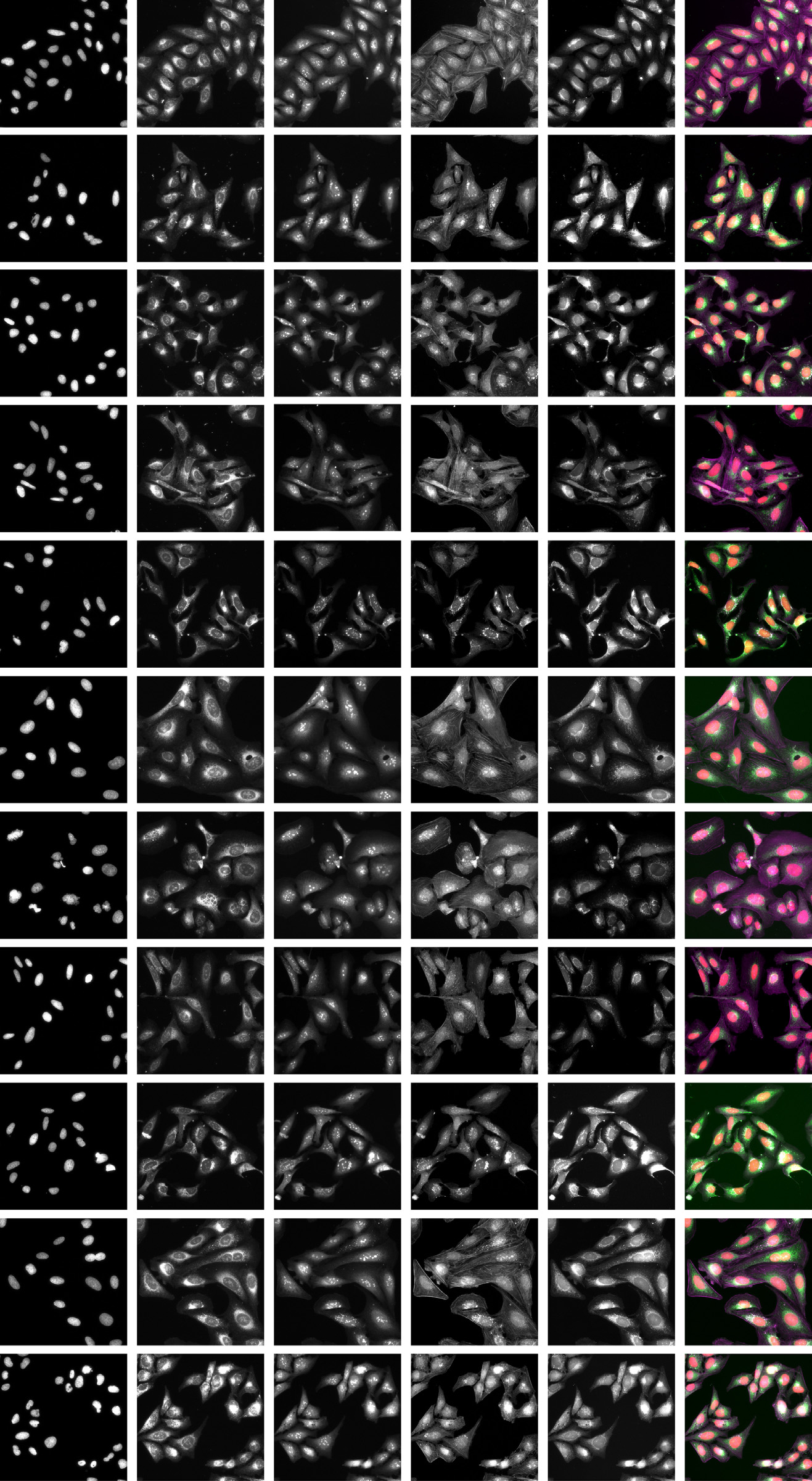
Intheworkpresentedinthismanuscriptwefirstcomparedavariety oftraditionalmachinelearninganddeeplearningmodelsforthepredic-tionofMoAbasedonchemicalstructuredataforupto20MoAclasses. Subsequently,basedonasetof10MoAclasses,wecomparedtheper-formanceofthebestdeeplearningmodelatthecompoundstructural leveltoastate-of-the-artCNNtrainedonCellPaintingimagedatafor thesamesetofcompounds.Weselectedthebestdeeplearningbased compoundstructuremodelsothatwecouldfinallytrainajointmodel fortheMoApredictionbasedonutilizingbothstructuralandimagedata asinput.ExampleCellPaintingimagesforthe10MoAclassescanbe seeninFig.1.Tothebestofourknowledgeourworkrepresentsthefirst combinationoffivechannelCellPaintingimagedataandmolecularfin-gerprintdatatrainedinanend-to-endfashiontopredictMoA,wherein therawimages,asopposedtofeaturesderivedfromtheimages,were usedasinputtothemodels.

**Materialsandmethods**

*Data*

*Moleculardata*   
 Moleculardata(Corselloetal.[20]),intheformofSMILESstrings collectedandprocessedbytheBroadInstitute,wasusedinthisstudy. Thecleanseddatasetcontainsapproximately5500compoundscovering 1300MoAclasses,butmostMoAshaveveryfewcompoundsassociated withthem.ThenumberofcompoundsthateachMoAhasisshownin Fig.2.Asourmodelsshouldperformwellatthecompoundlevel,namely topredicttheMoAforunseencompounds,weusedasubsetofthedata, thetop20MoAs(i.e.the20MoAshavingthemostcompoundsassoci-atedwiththem).

2



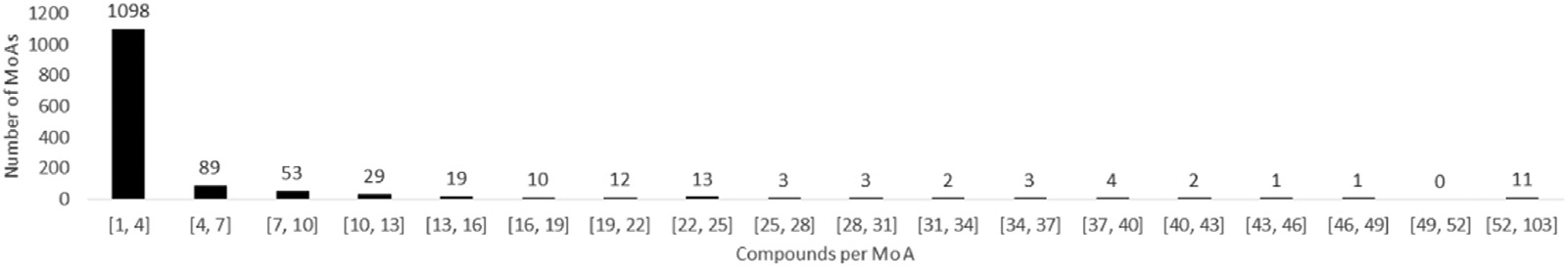
*G.Tian,P.J.Harrison,A.P.Sreenivasanetal.*



**Fig.1.**ExampleCellPaintingimagesforthe10MoAclassesandtheDMSOdatausedforstandardization.Therowtitlesgivethecompoundnamesfortheselected imageswiththeMoAabbreviationinparenthesis,whereistandsforinhibitorandAgforagonist.

3

*G.Tian,P.J.Harrison,A.P.Sreenivasanetal.*  *ArtificialIntelligenceintheLifeSciences3(2023)100060*



**Fig.2.**HistogramrepresentingthecompoundcountsperMoAfordifferentbinningintervals.Notethattherangeoftheintervalforthefinalbinislargerthanthe

others.

isfeasibleasslightlydifferenttokenscanbeproducedbyrandomized SMILES[28].

Forsplittingthedataatthecompound-levelintotraining,valida-tionandtestsetsweusedstratificationbasedontheproportionofcom-poundsforeachMoA.Wesplitout10%ofthedataforthefinalheld-outtestset.Stratifiedsplittingfortheremainingdatawasperformed ninetimes(nineshuffles)fortheSMILESdataintheinitialcomparison andfivetimes(fiveshuffles)fortheimagedataandthecorresponding SMILESsubset.Ineachcase80%ofthedatawasusedfortrainingand 10%forvalidation.

*Modeling*

*Compoundstructurebasedmodels*   
 Weexploredthefollowingdeeplearningmodelsfortheprediction ofMoAusingchemicalstructuredata:MLP,GCN,CNN,andLSTMwith andwithoutdataaugmentation.Forthedeeplearningmodelswede-terminedtheoptimalarchitecturesandparametersthroughmodelex-plorationandparametertuningonthevalidationsets.TheMLPisa basicartificialneuralnetwork[29]thatincludesfullyconnectedinput, hidden,andoutputlayers.OurMLPmodelcontainedoneinputlayer, onehiddenlayerwithdropout(*𝑝*=0*.*85),andonefinalpredictionlayer. GCNsareasubsetofGNNs[30]thatcanprocessnon-Euclideandata, suchasgraphswithnodesandedges[14].OurGCNmodelincluded inputlayersfortheadjacencymatrixandthenodematrixfollowedby threeconvolutionlayerswithdropout(*𝑝*=0*.*5),oneglobalattention poolinglayerandonefinalpredictionlayer.OurCNNmodelcontained oneconvolutionlayer,onemaxpoolinglayerwithdropout(*𝑝*=0*.*8),one flatteninglayerwithdropout(*𝑝*=0*.*8),andonefinalpredictionlayer. OurLSTMmodelincludedanembeddinglayer,abidirectionalLSTM layer,adropoutlayer(*𝑝*=0*.*96)andafinalpredictionlayer.Forthe LSTMwithdataaugmentation,weadjustedthedegreeofaugmentation toensurethateachMoAhadapproximately1000SMILESintheaug-mentedtrainingset.

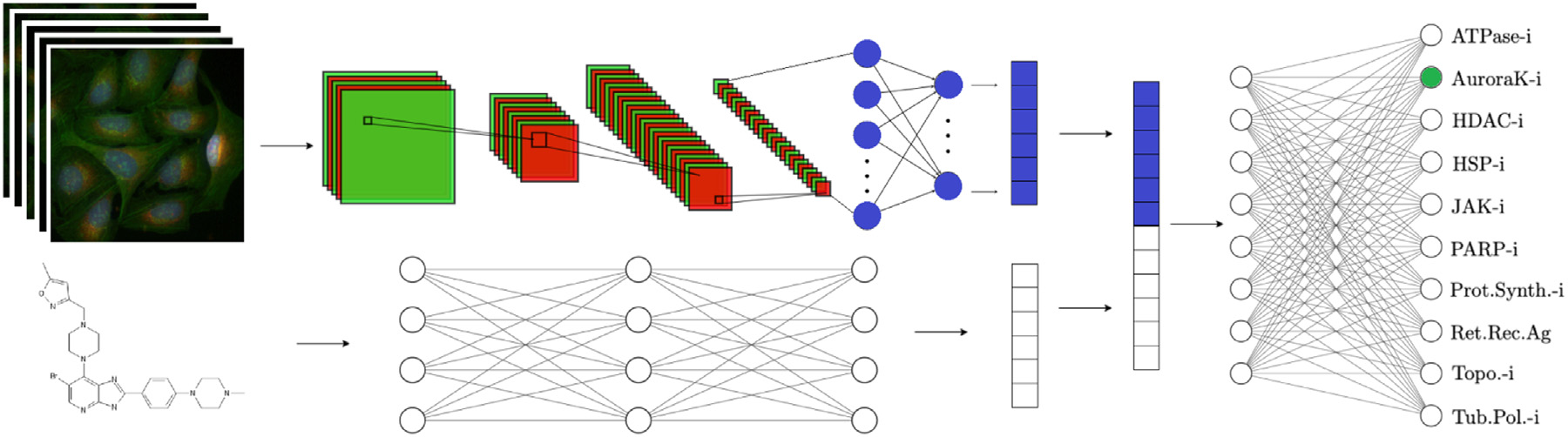
WeusedtheAdamoptimizer[31],sparsecategoricalcross-entropy asthelossfunction,andvalidationlossasthemetricforearlystopping. Toaccommodateforimbalanceofclassesweappliedclassweightingin thelossfunctionstotrainthemodels.

Wealsoexploredmachinelearningalgorithmsthatoperateontab-ulardata(incontrasttothedeepneuralnetworksdescribedabove). Themoretraditionalmachinelearningmodelshaveshowncompet-itiveperformancewithdeeplearningmodelswhendatasetsizesare relativelysmall[32].Forinstance,Jiangetal.[33]showedthatfour descriptor-basedmodelsoutperformedfourgraph-basedmodelsonsev-eralbenchmarkdatasets.Weexaminedfiveindividualmachinelearning algorithmsandfourensemblealgorithms.Theindividualalgorithmsin-cludedrandomforests[34],lightgradientboostingmachines[35],cat boost[36],k-nearestneighborsclassifiers[37],andlogisticregression [38].Theensemblealgorithmsincludedbagging[39],stacking[40], voting[41],andadaboost[42].

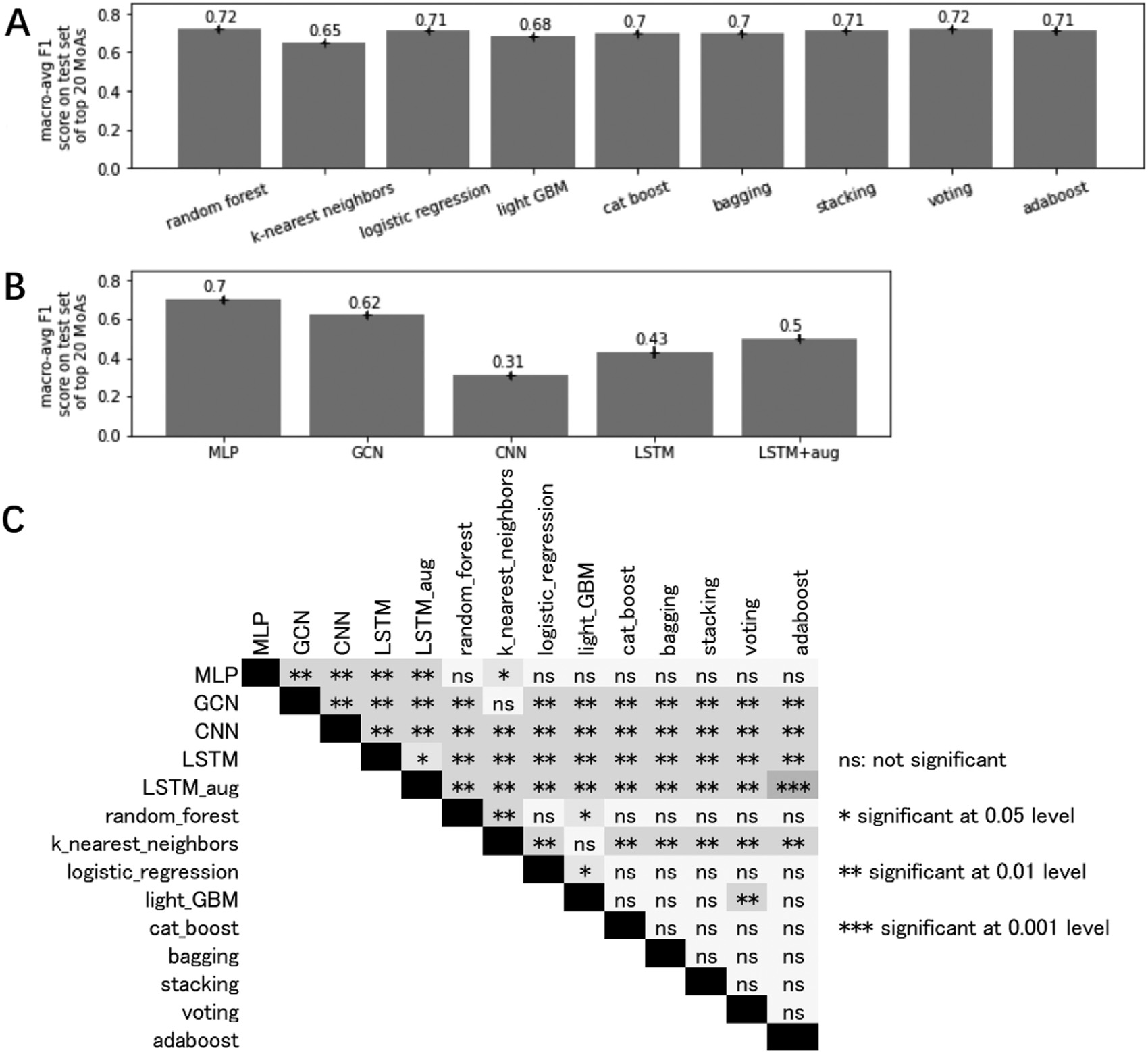
*Cellmorphologybasedmodel*   
 Weappliedthestate-of-the-artCNNmodelEfficientNet[43]topre-dictMoAbasedonthe5-channelCellPaintingimagedata.EfficientNet appliesacompoundscalingmethodtoadjustwidth,depth,andresolu-

4

*G.Tian,P.J.Harrison,A.P.Sreenivasanetal.*  *ArtificialIntelligenceintheLifeSciences3(2023)100060*



**Fig.3.**Thearchitectureoftheglobalmodelwithtwoinputpaths,onefortheCellPaintingimagedataandoneforthechemicalstructuredata.

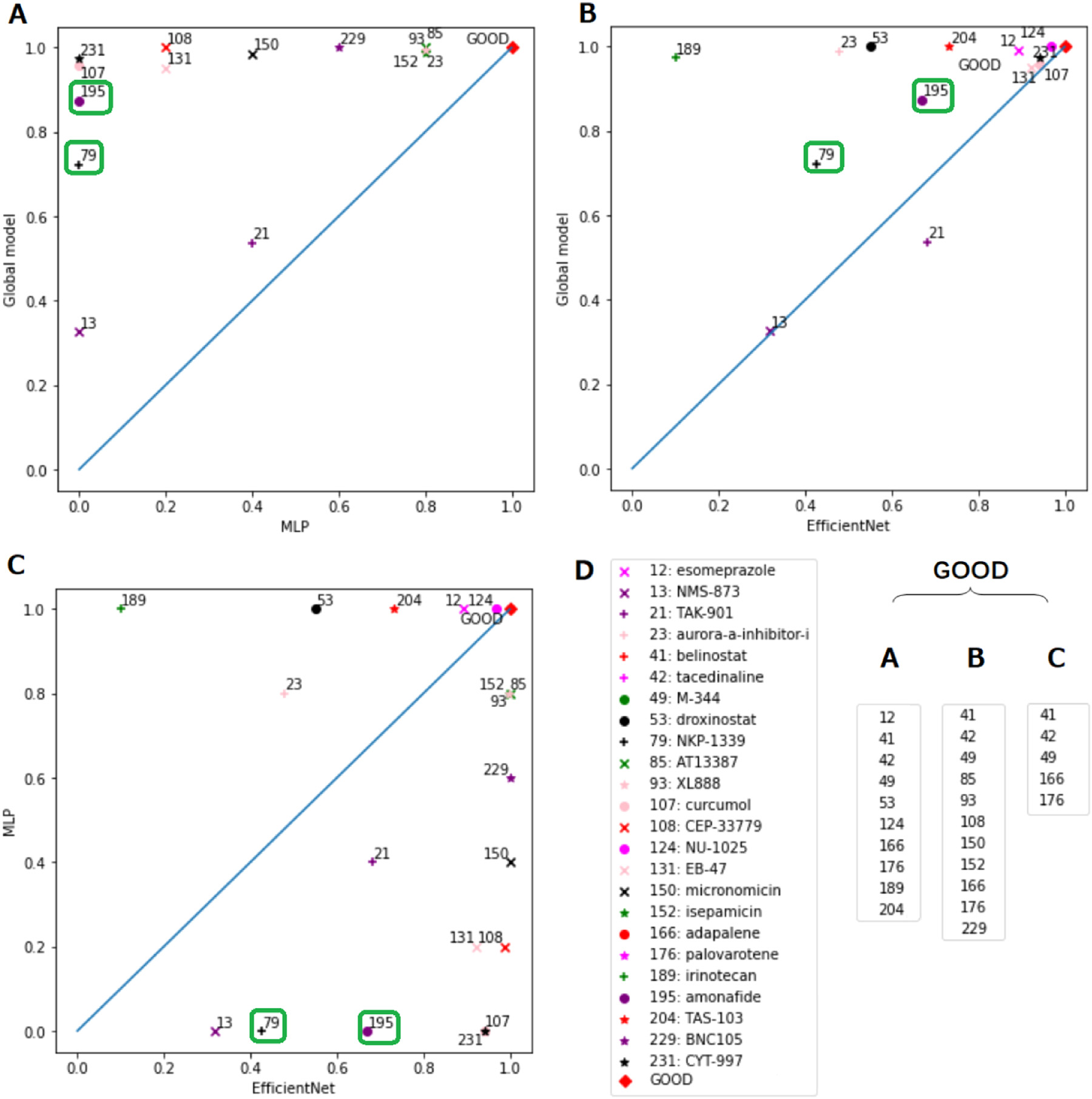


**Fig.4.**A).Comparisonofmacro-averagedF1scoresonthetestsetofthetraditionalmachinelearningmodelsforthetop20MoAs(i.e.theMoAsthatwerebest representedinthedataintermsofthenumberofcompoundstheyhad).B).Comparisonofmacro-averagedF1scoresonthetestsetofthedeeplearningmodels forthetop20MoAs.C).RandomizationtestwithBonferronicorrectionofmacro-averagedF1scoresonthetestsetoftop20MoAs.Theresultsarebasedonthe averagesacrossthenineshufflesofthetrainingandvalidationdata.

receptoragonistcompounds.ForEfficientNettheresultsweresomewhat morestable,rangingfrom0.48fortheAurorakinaseinhibitorsto0.98 forboththeProteinsynthesisinhibitorsandtheRetinoidreceptorag-onists.Fortheglobalmodeltheresultswereevenmorestable,rang-ingfrom0.68fortheATPaseinhibitorsto1.00fortheRetinoidrecep-toragonists.Ourglobalmodel,achievingamacro-averagedF1scoreof 0.92,revealedaclearadditive/synergisticeffectwithanincreaseinF1 scoreof0.11.Thethreedifferentmodelswereallsignificantlydifferent fromoneanotheratthe5%significance-levelbasedonrandomization

5

*G.Tian,P.J.Harrison,A.P.Sreenivasanetal.*  *ArtificialIntelligenceintheLifeSciences3(2023)100060*



**Fig.5.**Comparisonofthepredictionratesforthethreemodelsforeachcompound:A).MLP,trainedsolelyonthechemicalstructuredata,versustheGlobalmodel, trainedonboththechemicalstructureandCellPaintingimagedata;B).EfficientNet,trainedsolelyontheimagedata,versustheGlobalmodel;C).EfficientNet versusMLP;D).’GOOD’clustercontainedcompoundsthatpossessedapredictionrateabove0.97inpanelsA-C.ThecompoundsNKP-1339andamonafidehave beenhighlightedingreenboxesinpanelsA-Castheyshowedagreatersynergisticeffectthantheothercompounds.(Forinterpretationofthereferencestocolour inthisfigurelegend,thereaderisreferredtothewebversionofthisarticle.)

theanalysesweperformedweusedthedefaultparametersettings.In FigureS1A(Supplementarymaterials)weshowaneighborhoodanal-ysiswherecompoundswithatleastonestructurallysimilarneighbor haveconnectinglines.Thesestructurallyconnectedcompoundsareiso-latedinFigureS1B(Supplementarymaterials)andtheircompoundID numbersareshown.ThenamesforthesecompoundsandtheirSMILES stringsareavailableonourGitHubrepository(seeDataandcodeavail-ability).Althoughthegroupingstendtoshowcompoundsbelongingto thesameMoAclass,thereareseveralunconnectedcompounds.InFig-ureS2(Supplementarymaterials)weshowthedistributionsforthecom-pounds,groupedbytheirMoAclasses,foravarietyofphysico-chemical propertiespredictedbyDataWarrior.Wealsohighlightinthisfigurethe sixtestsetcompoundsthatwerepoorlypredictedbyourMLPmodel(see TableS1intheSupplementarymaterials).Fromthiswecanseethatin somecasesthesepoorlypredictedcompoundshavephysico-chemical propertiesinthetailsofthedistributionsfortheirMoAclasses.This ismostevidentforthepropertycLogS,ameasureofaqueoussolubil-

6

*G.Tian,P.J.Harrison,A.P.Sreenivasanetal.*

selvesasmodelinput).Chemicalstructuredata(suchasthatobtained fromMorganfingerprints)cansufferfrom"ActivityCliffs",wherebya smallchangeinstructurecanresultinalargedifferenceinbio-activity, highlightingtheneedtosupplementchemicalstructuredatawithaddi-tionalsourcesofinformation[2],suchastheCellPaintingimagedata usedinthecurrentstudy.

ItshouldbenotedthattheCellPaintingassaywasperformedon asinglecancercellline(U2OS).AlthoughperformingMoAandtarget identificationstudiesusingcancercelllinesasproxyiswell-accepted [47,48],thereisthepossibilitythatusingadifferentcancer(ornon-cancer)cellline,couldyieldslightlydifferentresults.Nevertheless, giventhattheexperimentwasexecutedinacontrolledmanner,and thatallcellsweretreatedequally,weareconfidentintherobustnessof ourresults.

Similarly,thedrugscreeninthisstudywasperformedat10uMfor allcompounds.Thisisaconcentrationtypicallyusedindrugscreensto ensureacellularresponse.However,giventhatthecompoundsinthis workcouldhavedifferentpotency,adoseresponsewouldhaveperhaps aidedamoreaccurateidentificationofthecompounds’activity. Duetodatalimitationsourtestsethadonlytwotofourcompounds perMoA.Thetest-levelpredictiveperformanceundersuchconditions maysufferifbychanceanyofthesecompoundshappenedtobeoutliers fortheirclass.BasedonourcompoundlevelanalysisusingDataWarrior itappearsthatthismayhavebeenthecaseforourtwoJAKinhibitortest compounds,thuspotentiallyexplainingthelowF1scorefortheMLPfor thisMoA.

However,afewcompoundswerebetterpredictedbythechemical structurebasedmodel,relativetotheimagebasedmodel.Perhapsthe maindisadvantageofimagedatasuchasCellPaintingisthatnotall compoundswillnecessarilyproduceamorphologicalchangeorthemor-phologicaleffectsmaybeverysubtlyandpotentiallymaskedbyunac-countedfortechnicalvariationswithinandbetweenplatesduringimage capturing[2].However,inthecurrentstudy,toreducepotentialbias causedbypositionaleffectsinthemicro-wellplates,thecompoundsand controlsweredistributedovertheplatesusingPLAIDandwestandard-izedtheimagesacrosstheplatesbasedonthecontrol/DMSOwells.Itis alsopossiblethatthecompounddoesproduceamorphologicalchange butnotinanyofthecellularcompartmentsororganellescapturedusing theCellPaintingassay.Anotherpossibilityisthatthedoseappliedwas notsufficienttoproduceamorphologicalchange.

Concerningthetraditionalmachinelearningmethodsweexplored forthechemicalstructuredata,theensemblemethodsoutperformedthe individualmethods.Similarly,whencombiningamodelbasedonmul-tipleinputs,withseparatemodelingpathsthatcometogethertomakea finalprediction,wecanpotentiallyachievebetterresultsthanthemod-elsbuiltonjustoneofthedatacategories.Inthisstudy,weshowed thistypeofadditive/synergisticeffectbycombiningMLPforchemical structuredataandEfficientNetforimagedataforMoAprediction. Althoughitwassomewhatsurprisingthatforourmodelsbasedon onlythechemicaldescriptors,thesimplestdeeplearningarchitecture, theMLP,outperformedthemorecomplexnetworksarchitecturesex-plored,asimilarresulthasbeenobtainedinapreviousstudy[49]per-formingdrugtargetpredictiononalargebenchmarkdatasetfromthe ChEMBLdatabase.InourMLParchitecture,weusedanunconvention-allyhighdropoutratetoalleviatetheoverfittingproblemasaresult ofthescarcityofchemicalstructuredata.Wealsotestedotherpossi-blearchitectures,suchasreducingthedropoutrateandincreasingthe numberofhiddenlayers,withfewerneuronsineachlayer.However, thesemodificationsdidnotimprovethemodelperformance.

Itshouldalsobementionedthatthepurposeofthecurrentstudy wastocomparetheaccuracyofmodelstrainedonmorphologicaland structuraldata,andthatthedomainofthemodelshencelimittheir applicabilityoutsideofthisscope(i.e.formakingpredictionswhena testcompound’sMoAdoesnotbelongtoanyofthoseonwhichthe modelwastrained).Thisisduetothefactthatthepredictionsofneural networkmodelsareneitherprobabilisticnorwell-calibrated[50],and

7

*G.Tian,P.J.Harrison,A.P.Sreenivasanetal.*

**Supplementarymaterial**

Supplementarymateri[alassociatedwiththisarticlec](https://doi.org/10.1016/j.ailsci.2023.100060)anbefound,in theonlineversion,atdoi:[10.1016/j.ailsci.2023.100060](https://doi.org/10.1016/j.ailsci.2023.100060).

**References**

[1]HightSK,ClarkTN,KuritaKL,McMillanEA,BrayW,ShaikhAF,HaecklFPJ, Carnevale-NetoF,LaS,LohithA,etal.High-throughputfunctionalannota-tion[ofnaturalprod](https://doi.org/10.1101/748129)uctsbyintegratedactivityprofiling.BioRxiv2022:748129. [doi](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0002)[:10.1101/748129.](https://doi.org/10.1101/748129)

[[2]Tra](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0002)[potsiM-A,Hossei](https://doi.org/10.1101/748129)[ni-GeramiL,BenderA.Computationalanalysesofmechanism ofaction(MoA):data,methodsandintegration.RSCChemBiol2022;3(2):170–200. [3]KensertA,HarrisonPJ,SpjuthO.Transferlearningwithdeepconvolutionalneural](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0002) [networksforclassifyingcellularmorphologicalchanges.SLASDiscovAdvLifeSci R&D2019;24(4):466–75.](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0003)

[[4]CaicedoJC,SinghS,CarpenterAE.Applicationsinimage-basedprofilin](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0003)g[ofpertur-](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0003) [bations.CurrOpinBiotechnol2016;39:134–42.](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0005)

[[5]FengY,MitchisonTJ,BenderA,YoungDW,TallaricoJA.Multi-parameterpheno-typicprofiling:usingcellulareffectstocharacterizesmall-moleculecompounds.Nat RevDrugDiscov2009;8(7):567–78.](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0005)

[[6]PeckD,CrawfordED,RossKN,StegmaierK,GolubTR,LambJ.Amethodforhigh-](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0005) [-throughputgeneexpressionsignatureanalysis.GenomeBiol2006;7(7):1–6.](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0006)

[[7]BrayM-A,SinghS,HanH,DavisCT,BorgesonB,HartlandC,etal.Cellpainting,a](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0006) [high-contentimage-basedassayformorphologicalprofilingusingmultiplexedfluo-rescentdyes.NatProtoc2016;11(9):1757–74.](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0007)

[[8]WawerMJ,LiK,GustafsdottirSM,LjosaV,BodycombeNE,MartonMA,](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0007) [etal.Towardperformance-diversesmall-moleculelibrariesforcell-basedpheno-typicscreeningusingmultiplexedhigh-dimensionalprofiling.ProcNatlAcadSci 2014;111(30):10911–16.](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0008)

[[9]WayGP,NatoliT,AdeboyeA,LitichevskiyL,YangAX,LuX,CaicedoJ,Ci-](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0008)miniBA,KarhohsK,LoganDJ,etal.Morphologyandgeneexpressionprofil-ing[providecomplementaryinf](https://doi.org/10.1101/2021.10.21.465335)ormationformappingcellstate.bioRxiv2021. doi[:10.1101/2021.10.21.465335.](https://doi.org/10.1101/2021.10.21.465335)

[10]Lap[insM,SpjuthO.Evaluation](https://doi.org/10.1101/2021.10.21.465335)ofgeneexpressionandphenotypicprofilingdata asquantitativedescriptor[sforpredictingd](https://doi.org/10.1101/580654)rugtargetsandmechanismsofaction. [Biorxiv2019:580654.doi](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0011)[:10.1101/580654.](https://doi.org/10.1101/580654)

[[11]TrapotsiM-A,MervinLH,](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0011)[AfzalAM,Sturm](https://doi.org/10.1101/580654)[N,EngkvistO,BarrettIP,etal.Compari-sonofchemicalstructureandcellmorphologyinformationformultitaskbioactivity predictions.JChemInfModel2021;61(3):1444–56.](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0011)

[[12]SealS,YangH,VollmersL,BenderA.Comparisonofcellularmorphologicalde-](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0011)[scriptorsandmolecularfingerprintsforthepredictionofcytotoxicity-andprolifera-tion-relatedassays.ChemResToxicol2021;34(2):422–37.](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0012)

[[13]MoshkovN,BeckerT,YangK,HorvathP,DancikV,WagnerBK,ClemonsPA,](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0012) SinghS,CarpenterAE,CaicedoJC.Predictingcompoundactivityfrom phe[notypicprofile](https://doi.org/10.1101/2020.12.15.422887)s[andch](https://doi.org/10.1101/2020.12.15.422887)emicalstructures.bioRxiv2022:2020–112. [doi](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0014)[:10.1101/2020.12.15.422887.](https://doi.org/10.1101/2020.12.15.422887)

[[14]Sca](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0014)[rselliF,GoriM,TsoiAC,Ha](https://doi.org/10.1101/2020.12.15.422887)[genbuchnerM,MonfardiniG.Thegraphneural](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0014)  [networkmodel.IEEETransNeuralNetw2008;20(1):61–80.](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0014)

[[15]CarpenterAE,JonesTR,LamprechtMR,ClarkeC,KangIH,FrimanO,etal.Cell-](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0014)[Profiler:imageanalysissoftwareforidentifyingandquantifyingcellphenotypes. GenomeBiol2006;7(10):1–11.](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0015)

[[16]GrysBT,LoDS,SahinN,KrausOZ,MorrisQ,BooneC,etal.Machinelearningand](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0015) [computervisionapproachesforphenotypicprofiling.JCellBiol2017;216(1):65–71. [17]GuptaA,HarrisonPJ,WieslanderH,PielawskiN,KartasaloK,PartelG,SolorzanoL,](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0016) [SuveerA,KlemmAH,SpjuthO,etal.Deeplearninginimagecytometry:areview. CytometryPartA2019;95(4):366–80.](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0017)

[[18]HofmarcherM,RumetshoferE,ClevertD-A,HochreiterS,KlambauerG.Accurate](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0017) [predictionofbiologicalassayswithhigh-throughputmicroscopyimagesandconvo-lutionalnetworks.JChemInfModel2019;59(3):1163–71.](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0018)

[[19]ChandrasekaranSN,CeulemansH,BoydJD,CarpenterAE.Image-basedprofilin](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0018)g [fordrugdiscovery:dueforamachine-learningupgrade?NatRevDrugDiscov](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0019) [2021;20(2):145–59.](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0020)

[[20]CorselloSM,BittkerJA,LiuZ,GouldJ,McCarrenP,HirschmanJE,etal.Thedrug repurposinghub:anext-generationdruglibraryandinformationresource.NatMed 2017;23(4):405–8.](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0020)

[[21]RodríguezMAF,Carreras-PuigvertJ,Sp](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0020)[juthO.Designingmicroplate](https://doi.org/10.1101/2022.03.31.486595)[layoutsusing](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0020)  [artificia](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0022)l[intelligence.bioRxiv2022.doi](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0022)[:10.1101/2022.03.31.486595.](https://doi.org/10.1101/2022.03.31.486595)

[22][LandrumG.Rdkitdocumentation.Relea](http://refhub.elsevier.com/S2667-3185(23)00004-1/sbref0022)[se2013;1(1–79):4.](https://doi.org/10.1101/2022.03.31.486595)

8