

[Artificia](https://doi.org/10.1016/j.ailsci.2023.100071)l[IntelligenceintheLifeSciences3(2023)100071](https://doi.org/10.1016/j.ailsci.2023.100071)

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

DeepmetriclearningfortheclassificationofMALDI-TOFspectral signaturesfrommultiplespeciesofneotropicaldiseasevectors

FernandoMerchana,f,KenjiContrerasa,RolandoA.Gittensb,f,JoseR.Loaizab,c,d,f, JavierE.Sanchez-Galanb,e,f,∗

a*FacultaddeIngenieríadeEléctrica,UniversidadTecnológicadePanamá,CampusVictorLeviSasso,Panama,Panama*   
b*CentrodeBiodiversidadyDescubrimientodeDrogas,InstitutodeInvestigacionesCientíficasyServiciosdeAltaTecnología(INDICASATAIP),CityofKnowledge, Panama,0843-01103,Panama*   
c*SmithsonianTropicalResearchInstitute,Panama,Panama*   
d*ProgramaCentroamericanodeMaestríaenEntomología,UniversidaddePanama,Panama,Panama*   
e*FacultaddeIngenieríadeSistemasComputacionales,UniversidadTecnologicadePanama,CampusVictorLeviSasso,Panama,Panama*   
f*SistemaNacionaldeInvestigación(SNI),SecretariaNacionaldeCiencia,TecnologiaeInnovacion(SENACYT),Panama,Panama*

|  |  |  |
| --- | --- | --- |
| article | info | abstract |
| *Keywords:*  MALDI-TOF  Machinelearning  Metriclearning  Culicidae  Ixodidae  Moleculartaxonomy |  | DeepLearningtechniqueshavesignificantadvantagesformassspectralclassification,suchasparallelizedsignal correctionandfeatureextraction.DeepMetricLearningmodelscombineMetricLearningtodeterminethedegree ofsimilarityordifferencebetweenasetofmassspectrawiththegeneralizationpowerofDeepLearningto improvefeatureextractionevenfurther.Thetwomostpopularofthesemodelscombinemultipleneuralnetworks withidenticalarchitecturesandarecommonlycalledSiamese(SNN)andTripletNeuralNetworks(TNN).Herein, usingbothSNNsandTNNs,weintendedtotaxonomicallycategorizetwosetsofpreviously-validatedmassspectra thatcorrespondedto30speciesofNeotropicalarthropodsintheCulicidaeandIxodidaefamilies,someofwhich arediseasevectors.TheeffectivenessofSNNsandTNNstocorrectlyclassify826spectrafrom12mosquito speciesand310spectrafrom18speciesofhardtickswashighlyeffective,withbothalgorithmsperformingwith minimalaveragelossduringcross-validation.SNNsproducedaccuracyratesforticksandmosquitoesof91.22% and94.46%,respectively,whileaccuracyratesof93%and99%wereobtainedwithTNNs.Ourresultsindicate thatDeepMetricLearningisapracticalmachinelearningtoolforquicklyandpreciselyclassifyingMALDI-TOF-generatedmassspectraofNeotropicalandpublic-health-relevantarthropodspecies. |

**1.Introduction**

Attheforefrontofpublichealthconcernsistheglobalthreatfrom emergingvector-bornezoonoticdiseases,whichisincreasingwiththe continuousgeographicalexpansionofhumansandvectorspecies,and thespreadofnovelpathogens[1].Despitehistoricaleffortstoprevent thespilloverofpathogensfromwildlifeintopeople,theriskofemerging vector-bornediseasesisstillhighinmanyregionsaroundtheworld, particularlyindevelopingcountriessuchasPanamawherelargesocio-economicallyvulnerableindigenouscommunitiesreside[2].

Severalvector-bornepathogenspresentinPanama,including mosquito-transmitted*Plasmodium*(e.g.,malaria)andtick-transmitted pathogenic*Rickettsia*(e.g.,RockyMountainspottedfever),areonly mitigatedbytheeliminationandmonitoringofvectorpopulations [3,4].However,moststudiesonarthropodspeciesinPanamahave concentratedonafewspecies,andthevastmajorityofspeciesre-

mainunstudiedtothisday[5,6].Thisisinpartduetothedifficul-tiesencounteredwhentryingtolab-raiserarespecies,aswellasthe lackofaccurate,reliable,andcost-effectivespeciesidentificationmeth-odsotherthanthetraditionalmorphologicalidentificationtechniques. Moreover,crypticspeciescomplexesareagroupoftaxawithsimilar morphologiesthatbelongtodifferentevolutionaryunitsandareeas-ilyconfusedwithoneanotherwhenrecognizedusingmorphological characters[2,7].

Accordingly,developingnovelidentificationapproachesforarthro-podsthatcanaccuratelyandrapidlycatalogthelargebiodiversityof arthropodsthatexist,isatopicofgreatinterestinPanama[8,9],where severalvector-bornediseasesarestillcommon[10,11].Withthelargest numberofdescribedspeciesofmosquitoesandhardticksinCentral America,Panamaoffersanuniqueopportunitytoassesstheusefulness ofnoveltechniquestotaxonomicallycharacterizethiselevatedbiodi-versity.PreviousworkinPanamahasestablishedthepresenceof265

∗Correspondinga[uthor.](mailto:fernando.merchan@utp.ac.pa)

*E-mailaddresses:*[fernando.merchan@utp.ac.pa](mailto:fernando.merchan@utp.ac.pa)(F.Merchan),[javier.sanchezgalan@utp.ac.pa](mailto:javier.sanchezgalan@utp.ac.pa)(J.E.Sanchez-Galan).

<https://doi.org/10.1016/j.ailsci.2023.100071>  
[Received29December2022;Receivedinrevi](https://doi.org/10.1016/j.ailsci.2023.100071)sedform2April2023;Accepted3April2023   
Availableonline6April2023   
[2667-3185/© 2023TheAuthors.PublishedbyElsevi](http://creativecommons.org/licenses/by-nc-nd/4.0/)erB.V.ThisisanopenaccessarticleundertheCCBY-NC-NDlicense (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

*F.Merchan,K.Contreras,R.A.Gittensetal.*

mosquitospeciesand23Ixodidtickspeciesthatcanparasitizewildlife andhumans[2,7].

MassSpectrometry(MS)isahigh-performanceanalyticaltoolthat canbeemployedtocharacterizeandidentifycomplexbiologicalsam-ples.Thereexistdifferenttechniquestoacquirespectralsignals.Oneof particularinterestinthisworkisknownasMatrix-AssistedLaserDes-orption/Ionization-TimeofFlightorMALDI-TOF,whichisconsidereda two-dimensionalmethodgearedtounderstandcomplexmolecularsam-plesviaionizationandlatertime-of-flightdiscrimination[12,13].Be-sides“true” proteomics,anexcitingandgrowingapplicationofMALDI-TOFMSistheuseofproteinprofiles,or“massfingerprints”,toidentify speciesofmicroorganisms[14],andmorerecentlyofarthropodspeci-mens,fromeggsallthewaytoadults[15–18]   
 AtypicalMALDI-TOFmassfingerprintexperimentgeneratesalarge numberofspectralsignalsthatmustbepre-processedbeforemachine learningalgorithmscanautomaticallyclassifyandidentifyarthropod species[19,20].Massspectralfeaturesallowfortheidentificationof distinctbiomarkersthatcharacterizedifferentbiologicalspecies.Forin-stance,MALDI-TOFhasbeenpreviouslyusedinPanamabyourteam,for speciesidentificationintwoofthemostimportanttaxonomicfamilies ofarthropodsthatcontainvectorsofdisease:*Culicidae*(i.e.,mosquitoes) [2,16,21]and*Ixodidae*(i.e.,hardticks)[7].

Loaizaetal.[2]andGittensetal.[7]usedPrincipalComponent Analysis(PCA)[22]andLinearDiscriminantAnalysis(LDA)[23]to categorizePanama’smosquitoesandticksmassspectra,respectively, acquiredfromsamplesstoredundervariousconditions(e.g.,dry-frozen orinalcoholatroomtemperature).Bothofthesealgorithmsfocuson transformingtheinputdatauntoanorthogonalsub-spaceinwhichnew featuresallowtodiscardirrelevantinformation,improvingclassifica-tionefficiency.Thistypeofmachinelearningframeworkisprecededby apre-processingstageconsistingofsignalsmoothing(Savitzky-Golay filter)andbaselinecorrectiontoremovenoiseandimproveperformance classification.AnotherapproachbyLópez-Fernándezetal.[24]featured asoftwareplatformnamedMass-Upwhichimplementedunsupervised learningalgorithmstogetherwithPCAtofacilitatethepre-processing andanalysisofMALDIspectra.However,thesetypesofalgorithmsmay notbeefficientorrobustenoughtocontinuouslylearnandreliablyclas-sifymassspectrafromvariousarthropodfamilies,collectedandstored withdifferenttechniquesandprocessedandanalyzedwithdifferentpa-rametersandequipment.

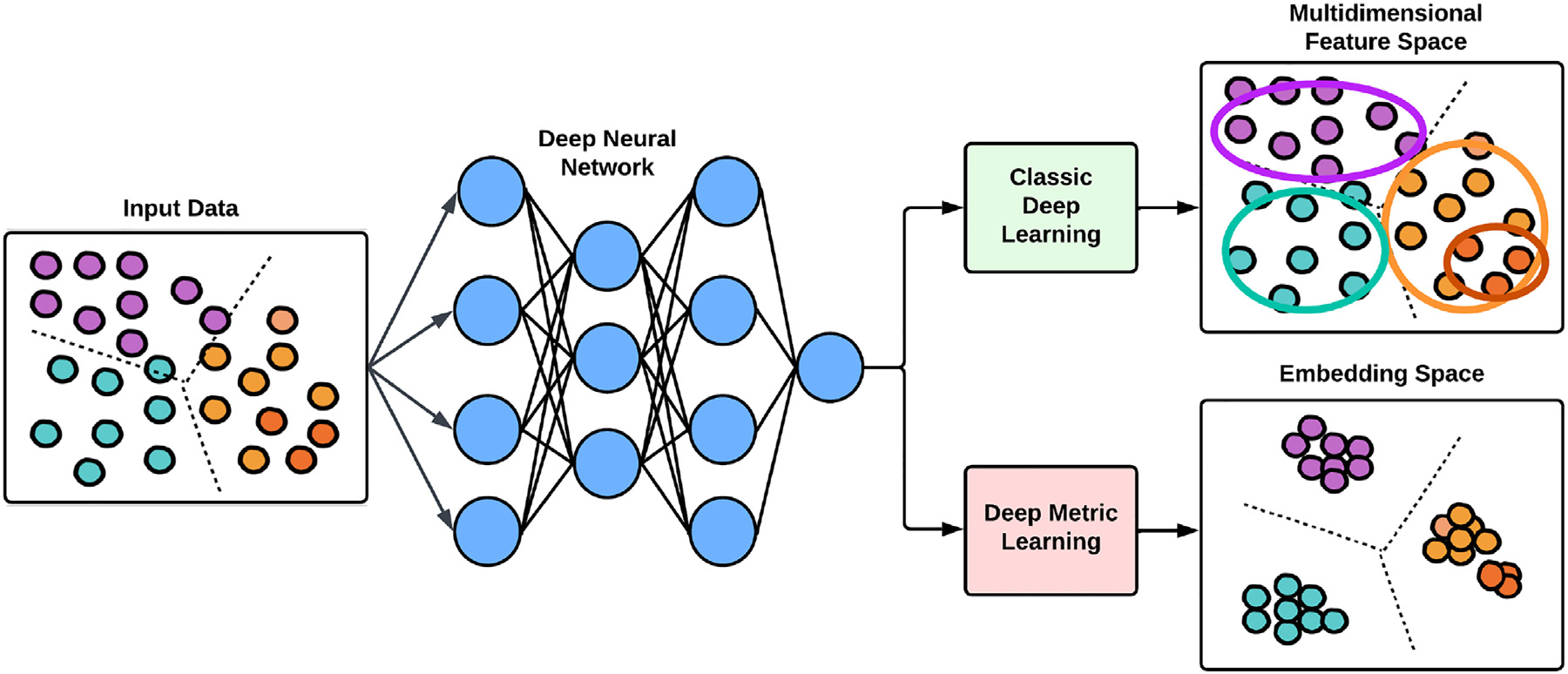
Anotherbranchofmachinelearningthatshowspromiseformass spectraanalysisandclassificationisDeepLearning,whichcomprises allthetechniquesandalgorithmsrelatedtoArtificialNeuralNetworks (ANN).Theinceptionofthesealgorithmswasinspiredbythefunction-ingofneuronsinthehumanbrain.Thesemodelscarryoutanadaptive learningprocessthroughmultiplenonlinearmathematicaltransforma-tionsbasedonmultivariablecalculusandtensoralgebra[25].

DeepMetricLearningisatechniquethatcombinesDeepLearning elementswiththeconceptofMetricLearning(MeL),inwhichamodel attemptstolearnfeaturesfromobjects(suchassignalsandimages)por-trayedasnumericalvectorsintermsoftheirdistanceandlocationin amultidimensionalspace.Thisconceptimpliesthattwovectorsina neighboringspacewillhavesimilarfeatures,whiledistancesbetween themindicatevaryingdegreesofdissimilarity[26].Aclassicexampleof MeLalgorithmscorrespondstoK-NearestNeighbors(K-NN)[27],which performstheclassificationprocessbasedonadistancemetric(e.g.,Eu-clidean,angular[cosine],Manhattan)toareferencesampleandthe*𝑘* neighboringsamplesthatareinthenearbyvectorspace.

TounderstandthedifferencesbetweenDeepMetricLearningand DeepLearning-basedmethods,onemayconsidertheanalogyofaperson (i.e.,anintelligentmodel)learningtodiscriminatebetweendifferentob-jectsinagivendataset.Oneapproachtothistaskisforthepersonto carefullystudythedefiningcharacteristicsofeachclassofobjects(i.e., DeepLearning).Alternatively,thepersonlearnstodeterminethesimi-laritiesbetweenobjectsregardlessoftheirpredeterminedlabel.Further-more,onecanillustratebothmethodologiesbydescribinghowama-

2

*F.Merchan,K.Contreras,R.A.Gittensetal.*  *ArtificialIntelligenceintheLifeSciences3(2023)100071*



**Fig.1.**Graphicalrepresentationcomparingtraditionalfeaturelearningmethodswithmetriclearning.

|  |  |
| --- | --- |
|  | **Fig.2.**Exampleofarawspectrumoftick species*Amblyommanaponense*,collectedasin-tensityvs.m/z(massovercharge),witha moleculesizevaryingfrom0to50,000Da. |

*2.2.Machinelearningarchitecture:Convolutionalneuralnetworks(CNN)*

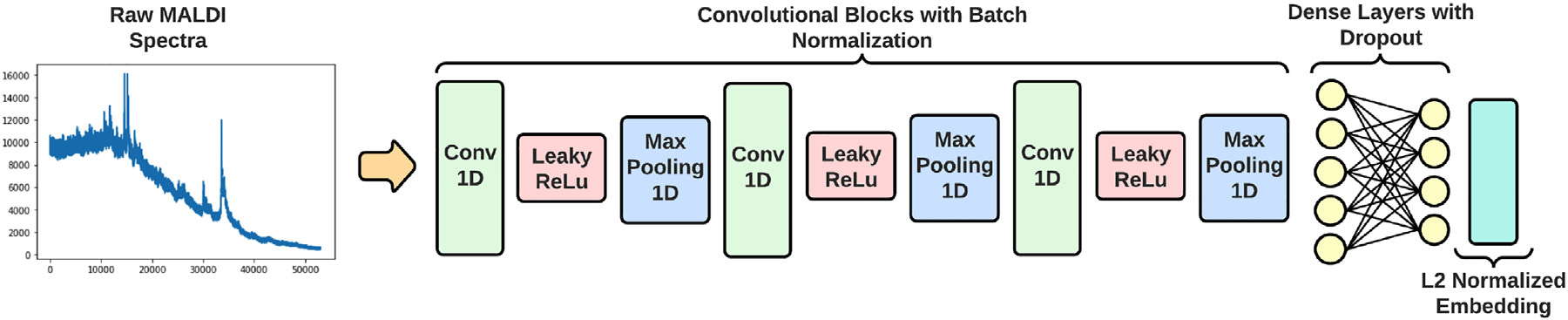
CNNhavebeenthemostcommonlyemployedarchitecturesinrecent yearsforimagerecognition,classification,anddetectiontasks.Thisclass ofneuralnetworkshasthreemaintypesoflayersthatperformtheoper-ationsofconvolution,*pooling*(forfeatureextraction)andclassification bymeansofanANN.Thesearchitecturesaremadeupofmodulesor blocksthatarestackedoneaftertheotherdependingonthecomplexity ofthemodel[30].

Thenetworkarchitectureutilizedinthisstudy(Fig.3)drawsinspira-tionfromtheclassicconvolutionalarchitecture,LeNet[32].Thedesign includesparametersandanactivationfunctioncalledLeakyReLu[33], whichmitigatesthevanishinggradientissuethatcanmaketheneural networkstruggletolearnnewdataduetoverysmallchangesinparam-eteroptimizationduringtraining.

2normalization[34],whichaimstomaximizethedistancebetween Additionally,thelastlayer,whichproducesembeddings,includes

3

*F.Merchan,K.Contreras,R.A.Gittensetal.*  *ArtificialIntelligenceintheLifeSciences3(2023)100071*



**Fig.3.**Diagramoftheunifiedarchitectureusedfortheanalysisofbothdatasets.

|  |  |
| --- | --- |
|  | **Fig.4.**Graphicalrepresentationoftheunified architectureofaSiameseneuralnetwork. |

**Table2**   
IxodidtickspeciesdatasetobtainedfromGittensetal.[7].

|  |  |  |  |
| --- | --- | --- | --- |
| ID | Name | Abreviation | SpectralCount |
| 1  2  3  4  5  6  7  8  9  10  11  12  13  14  15  16  17  18 | *Amblyommacajenense*  *Amblyommacalcarattum*  *Amblyommadissimile*  *Amblyommageayi*  *Amblyommanodosum*  *Amblyommaoblongoguttatum Amblyommaovale*  *Amblyommapecarium*  *Amblyommasaberanae*  *Amblyommavarium*  *Ammblyommannaponense*  *Ammblyommatapirellum*  *Dermacentornitens*  *Haemaphysalisjuxtackochi Ixodesaffinis*  *Ixodesboliviensis*  *Rhipicephalusmicroplus*  *Rhipicephalussanguineus* | AmbCaj AmbCal AmbDis  AmbGea AmbNod AmbObl AmbOva AmbPec AmbSab AmbVar AmbNap AmbTap DermNit HaeJux  IxAff  IxBol  RhiMic  RhiSan | 6  15  9  12  10  8  11  11  11  9  9  56  60  6  9  11  50  6 |

[39](Fig.4),givenbyEq.(1):

*𝐶𝐿𝑜𝑠𝑠*=(1−*𝑦*)1 2(*𝑑*)+(*𝑦*)1 2max{0*,𝛼* −*𝑑*} (1)

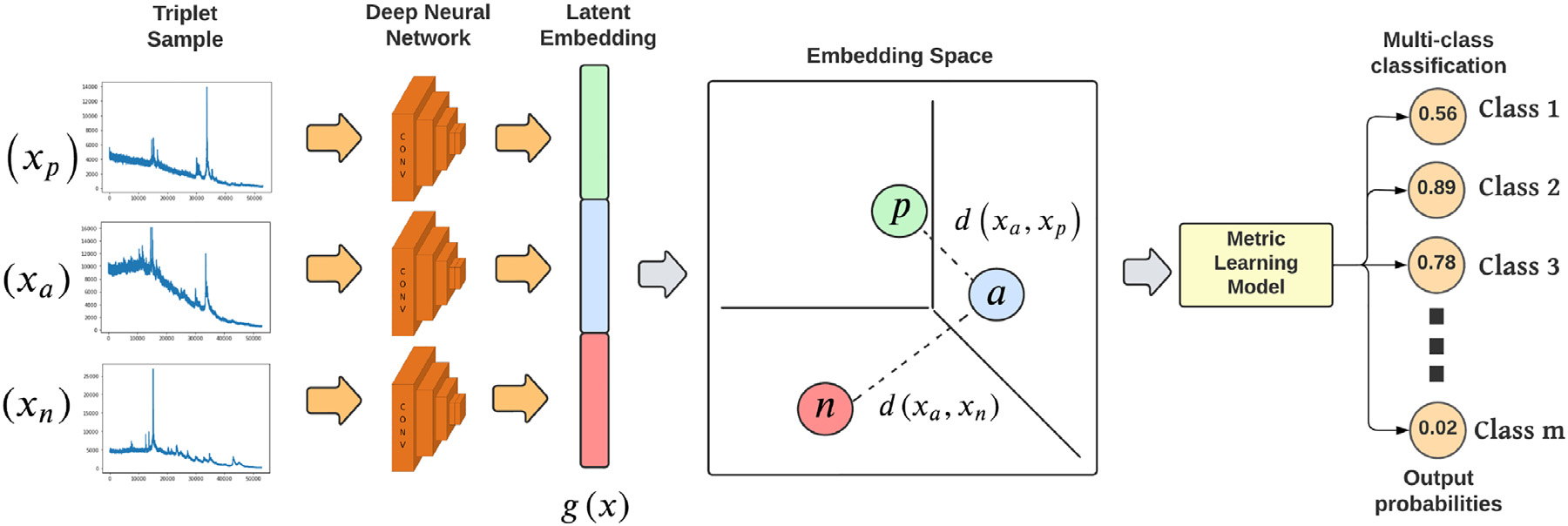
where,   
*𝑑*(*𝑥*1*,𝑥*2 Thecontrastivelossfunction[39](Eq.(1))aimstooptimizetheper-formanceoftheclassificationmodelbymeasuringthesimilaritybe-) =‖‖‖*𝑔*(*𝑥*1 ) −*𝑔*(*𝑥*2 )‖‖‖ (2)

tweenpairsofsamples,denotedbyabinarylabel*𝑦*,where*𝑦*=0denotes similaritybetweensamplesand*𝑦*=1denotesdissimilarity.Thefunc-tiontakesintwoinputsamples,*𝑥*1and*𝑥*2,andcalculatestheEuclidean distancebetweentheirrespectiveembeddings(*i.e.*,mappingfunctions *𝑔*(*𝑥*1)and*𝑔*(*𝑥*2)).Then,thedifferencebetweenembeddingsiscomputed andnormalizedtoformthedistancemeasure,*𝑑*(*𝑥*1*,𝑥*2).

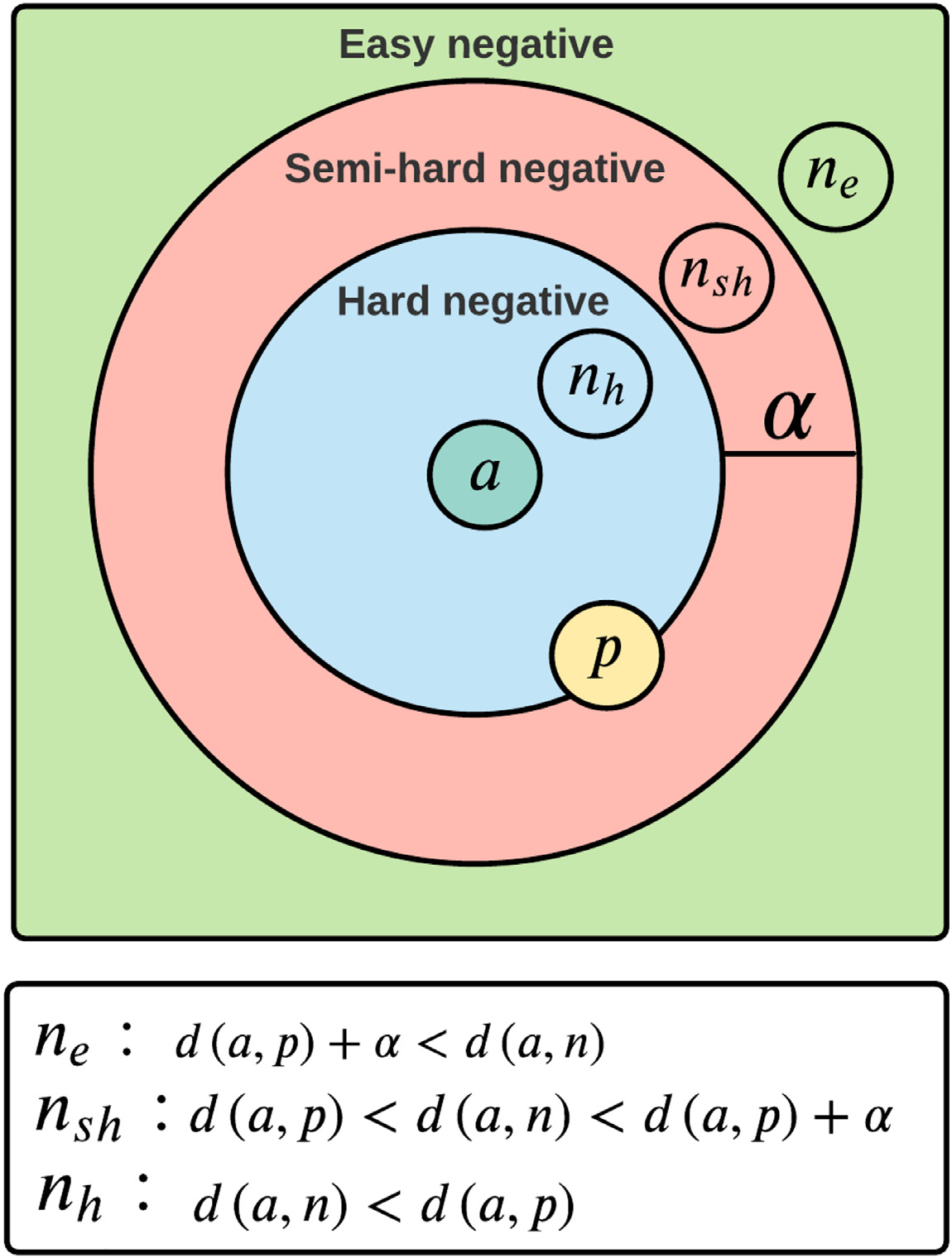
Duringtraining,theparametersof*𝑔*(*𝑥𝑛*)areadjustedtominimize theloss(errorrate)byreducingthedistancemeasurebetweensimilar sampleswhilemaximizingitbetweendissimilarsamples.Thisis,the functionoptimizesanembeddingspacewheresimilarspectraareinthe sameembeddingneighborhoodwhiledissimilaronesaredistant.This iscompletedbysettingathresholdvalue(*𝛼*)suchthatthedistancebe-

4

*F.Merchan,K.Contreras,R.A.Gittensetal.*  *ArtificialIntelligenceintheLifeSciences3(2023)100071*



**Fig.5.**Graphicalrepresentationoftheunifiedarchitectureofatripleneuralnetwork.



**Fig.6.**Descriptionofthetypesoftripletsbasedontheparameter*𝛼*.

|  |  |
| --- | --- |
| [26].Theselectioncriteriaalsodependsonthethresholdparameter*𝛼* (Fig.6).  Moreover,itisnecessarytoremarkthatTNNsareusedasamethod offeatureextraction,unlikeSNNswherebothfeatureextractionand classificationareperformedbythesamearchitecture.Theneuralnet-workswillhavetheirparametersoptimizedduringtrainingtoproduce anewembeddingspacewhereanothermetriclearningmodel(inthis case,K-NN)willperformthemulti-classclassificationstage.  *2.3.Dataanalysis:Experimentalsettings*  DatasetsusedinthisstudyaredescribedinTables1and2with atotalof826mosquitoand310tickspectra,respectively.Thedistri- | butionofsamplesformodeltrainingandtestingwas80%/20%inthe caseofmosquitoes,ashadbeenthecasewithourpreviouslypublished machinelearningalgorithms[2,7],buthadtobeadjustedforticksto 70%/30%aswillbedescribedahead.AsexplainedinSection2.2,the usageofCNNallowedtotrainmodelswithinputrawspectrawithoutthe needforsignificantpre-processingsteps.Therefore,theonlystepbefore trainingwastotake*𝑛*firstfeaturesofeachspectrasample(according tothesmallestspectrumofthedataset),yieldingnumericalarraysof *𝑛*=52906and*𝑛*=52570featuresforticksandmosquitoesrespectively. DuetothespecialtrainingschemethatSNNarchitecturesimply (Section2.2.1),pairsofsampleswerefirstgeneratedrandomly,main-tainingasimilardistributionofsamplesbetweenthedifferentclasses toavoidbiasesduetosampleimbalance(Table3).Afterperforming |

5

*F.Merchan,K.Contreras,R.A.Gittensetal.*

**Table3**   
Distributionofspectrafortrainingandvalidationwith*SNN*.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Sample | #TrainingPairs | #TestingPairs |
| Mosquitoes Ticks | 826  310 | 578  432 | 248  186 |

**Table4**   
Distributionofspectrafortrainingandvalidationwith *TNN*.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Sample | Training(%) | Testing(%) |
| Mosquitoes Ticks | 826  310 | 80%  70% | 20%  30% |

**Table5**   
GenericConfusionMatrix.

Predicted

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Positive | Negative |
| **Actual** | **Positive**  **Negative** | TP  FP | FN  TN |

manualparameterexperimentation,weselectedthethresholdvalueof *𝛼* =0*.*55forbothticksandmosquitoes.

FortheTNN,the*𝛼* wassetto0.4,anditshouldbenotedthatthe bestperformanceforTNNtrainingwasachievedusinganOnlineTriplet MiningschemewithHardNegativeTriplets(Fig.6).Furthermore,this promptedanadjustmentto70%/30%train/testingdatasplittingfor thetick’sdataset,whichhadmoreclasseswithfewersamplesperclass comparedtomosquitoes(Table4).Ultimately,thisadjustmentwasnec-essarytoavoidcomputationalerrorsproducedduringthevalidation stageaftereachtrainingepoch,wheretheminingalgorithmfailedto findtripletsthatfulfillthesimilaritycriteriadescribedinSection2.2.2. Furthermore,inordertoavoidover-fitting,theSNNmodelswere trainedwith*k-fold*crossvalidations,with*𝑘*=5,for30epochswitha *batchsize*of16samples.TheTNNweretrainedwitha*batchsize*of24 samples,usingaK-ShuffleSplitvalidationscheme(alsoknownasMonte CarloCrossValidation)for20iterationsand32epochsineachone.This validationschemeproducesbetterresultswhendealingwithdatasets containinglimitedsamplesperclass[42].

*2.4.Evaluationmetrics*

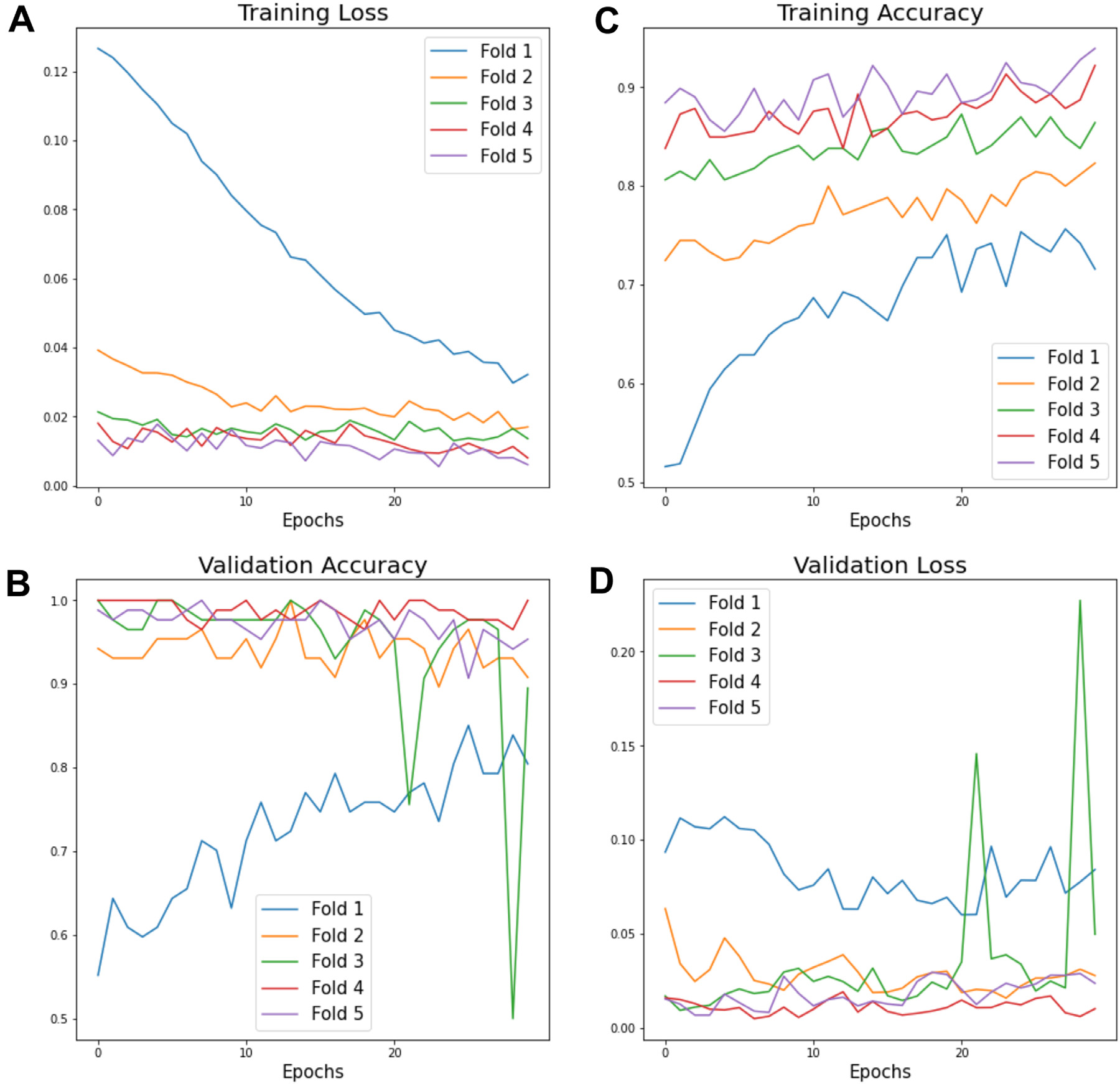
Errormetricswerecalculatedforeachmodelanddatasetcombi-nation.AconfusionmatrixlistingTrueNegatives(TN),TruePositives (TP),FalseNegatives(FN)andFalsePositives(FP)wasalsousedtorep-resentthetotalnumberandproportionofsamplesthatwerecorrectly orincorrectlypredicted(Table5).

Moreover,usingtheitemsoftheconfusionmatrix,othererrormet-ricssuchasAccuracy,Precision,RecallandF1-score(Eqs.4–7,respec-tively)werecomputedandlatercomparedforeachmodel.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *𝐴𝑐𝑐𝑢𝑟𝑎𝑐𝑦*= | | *𝑇𝑃*+*𝑇𝑁*  *𝑇𝑃*+*𝑇𝑁*+*𝐹𝑃*+*𝐹𝑁* | | (4) |
| *𝑃𝑟𝑒𝑐𝑖𝑠𝑖𝑜𝑛*= | | | *𝑇𝑃*  *𝑇𝑃*+*𝐹𝑃* | (5) |
| *𝑅𝑒𝑐𝑎𝑙𝑙*= | *𝑇𝑃*  *𝑇𝑃*+*𝐹𝑁* | | | (6) |
| *𝐹*1−*𝑆𝑐𝑜𝑟𝑒*=2∗*𝑃𝑟𝑒𝑐𝑖𝑠𝑖𝑜𝑛*∗*𝑅𝑒𝑐𝑎𝑙𝑙 𝑃𝑟𝑒𝑐𝑖𝑠𝑖𝑜𝑛*+*𝑅𝑒𝑐𝑎𝑙𝑙*= 2∗*𝑇𝑃*+*𝐹𝑃*+*𝐹𝑁*  2∗*𝑇𝑃*  (7)  Ingeneral,theaccuracycorrespondstothenumberofsamplesthat werecorrectlypredictedintheirparticularclass(*i.e.*,species)among thetotalnumberofsamplesusedinthetest. | | | | |

6

*F.Merchan,K.Contreras,R.A.Gittensetal.*  *ArtificialIntelligenceintheLifeSciences3(2023)100071*



**Fig.7.**LossandAccuracycurvesshownforTraining(A,C)andTestingValidation(B,D)usingthe*SNN*modelforthemosquitosdataset).

**Table8**   
ResultingMetricsfortheSNNmodelusingbothsetsofspec-tra.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Accuracy | Precision | Recall | F1-Score |
| Mosquitoes Ticks | 0.9188 0.9147 | 0.9188 0.9139 | 0.9117 0.9139 | 0.9172 0.9139 |

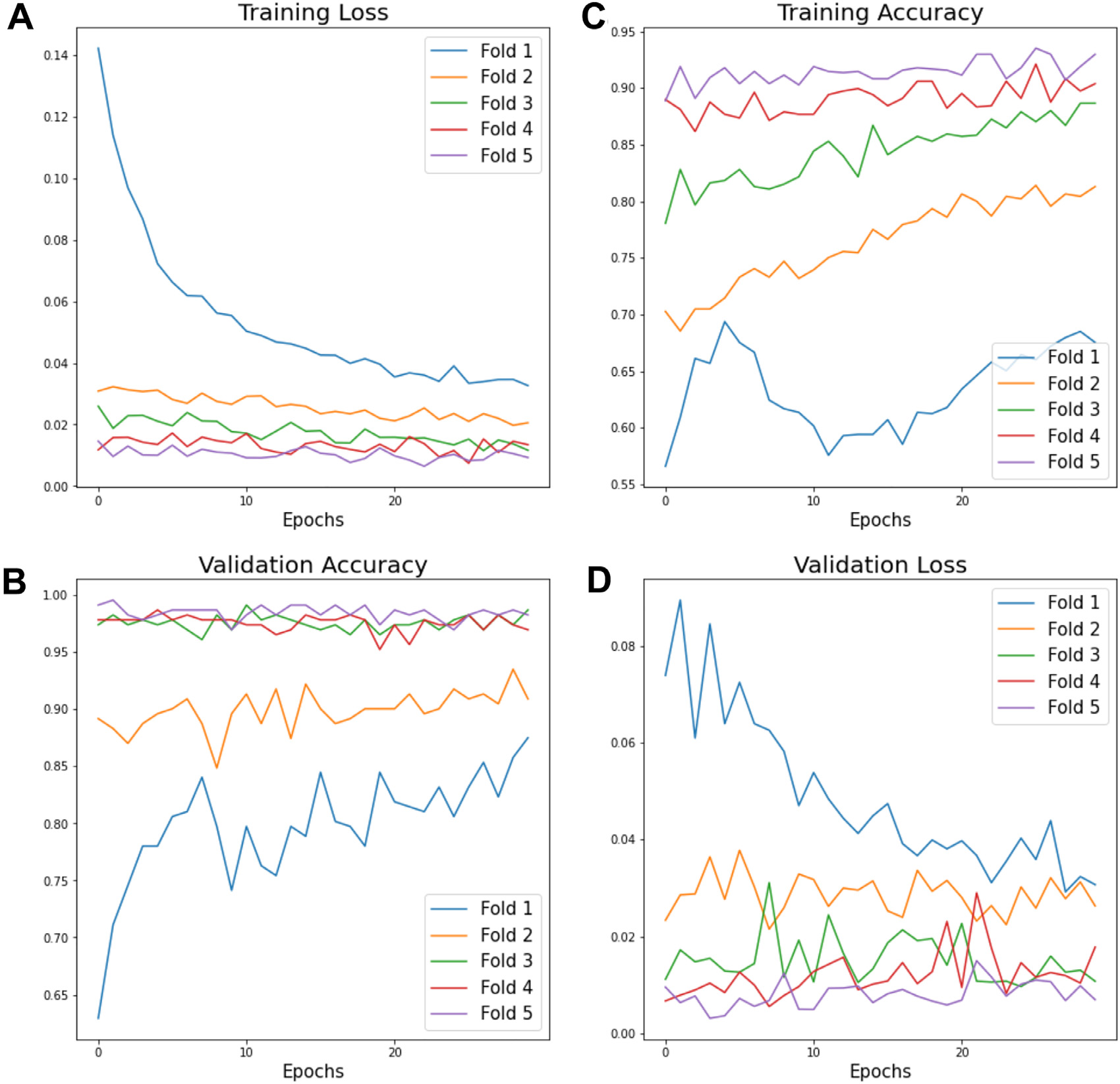
*3.2.Classificationwithtripletneuralnetworks(TNN)*

TNNmodelswereveryeffectiveatclassifyingmassspectradatasets andoutperformedSNNmodels.Attheendoftheentiretrainingand validationcycle,theminimumvalidationlossreachedbytheTNNmod-elswas0.00001formosquitoesand0.0160forticks.Furthermore,the errormetricsofaccuracyand*F1-Score*obtainedforboth*knn*models withtheirrespectivetestsetsreachedupto99.7%forbothparameters inthemosquitodataset,and93.2%and91.4%,respectively,forticks (Table9).

TheConfusionmatricesshowninFig.9,explorewhichspecies areeasiertotopredictforbothmosquitoesandticks.Inthecaseof mosquitoes,thematrixshowedcorrectpredictionsforallexceptone

7

*F.Merchan,K.Contreras,R.A.Gittensetal.*  *ArtificialIntelligenceintheLifeSciences3(2023)100071*



**Fig.8.**LossandAccuracycurvesshownforTraining(A,C)andTestingValidation(B,D)usingthe*SNN*modelfortheticksdataset).

sets.Thesegraphsprovideavisualrepresentationofhowthemodel learnstoseparateand/orjoinsamplesdependingontheirdegreeof similarity.

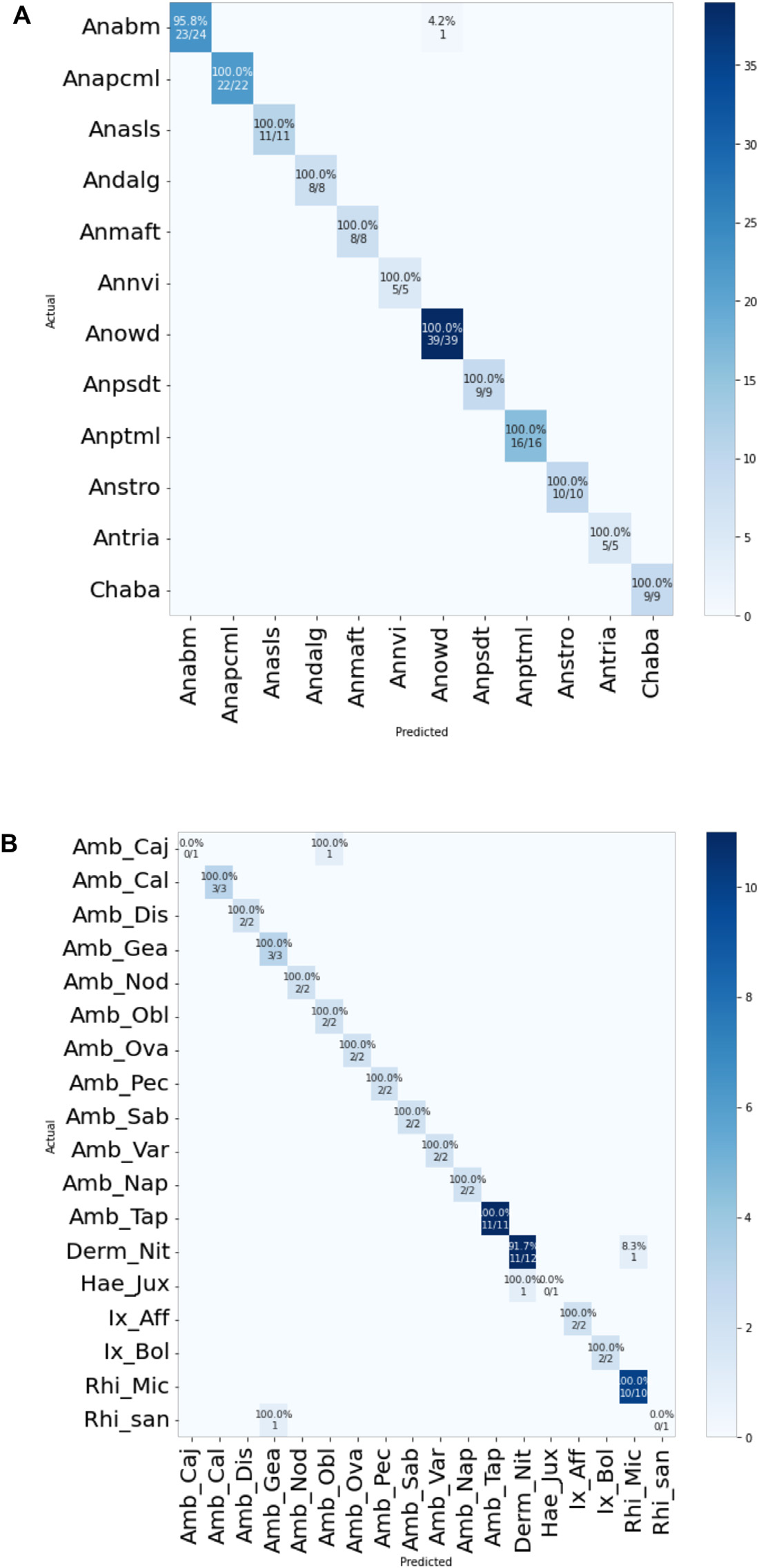
Forexample,theembeddingpriortoTNNtrainingclearlydemon-stratesthatsamplesofonespeciesofmosquito(Fig.10A)orticks (Fig.11A)canbefoundanywhereonthemap,whereasafterthenetwork istrained,samplesofthesamespeciesarecloselyclusteredtogetherand specieswithsimilarspectralcharacteristicsformneighborhoodsorclus-terswhereeachpointrepresentsonespectrum(Figs.10Band11B).

**4.Discussion**

Inourpreviousworkestablishingtheseself-curateddatasetsandan-alyzingthemwithPCAandLDAmachinelearningalgorithms[2,7],we comparedthesmoothedandbaseline-correctedspectrageneratedfrom unidentifiedfield-collectedarthropodsampleswiththemeanspectra fromasubsetofthesamefield-collectedsamplesthathadalreadybeen identifiedusingconventionalmorphologicalmethods.Themaincon-tributionofthesestudieswastoshowawayaroundworkingwithout acommercialreferencelibraryofproteinspectraprofiles,usingwhat wecalledthe“self-curatedreferencelibrary” concept.Becauseweare notusinghigh-quality,lab-rearedspecimenstotrainthealgorithm,this methodologyrequiredsuchpre-processingofthespectratoprovidedan accurate,andunbiasedmethodtoidentifymosquitoandtickspecies,

8

*F.Merchan,K.Contreras,R.A.Gittensetal.*



to99%),anditactuallyreducedthesuccessrateforticks(*i.e.*,from 94.2%to93%),theimprovementsinlackofmanualprep,speedand computingavailabilitymorethanjustifyitsimplementation.

OurresultsshowthatbothSNNsandTNNswerecapableofrec-ognizingandclassifyingMALDIspectrafromroughly50%ofallhard ticktaxaand70%ofallanophelinetaxa(e.g.,bothecologicallydom-inantandrarespecies)reportedforPanama.Theseresultsareequiva-lentandsometimesevensuperiortothosefromLoaizaetal.[2]and Gittensetal.[7].Still,the*pacmap*visualizationoftheembeddingcon-firmedthatthesealgorithmsarenotperfect.Futureresearchwillneed tolookintowhatcausedsomesamplesin*Ana\_bm,Amb\_Caj,Derm\_Nit, Hae\_Jux*,and*Rhi\_San*speciestobemisclassified.Wecanonlyspecu-latethatsomeofthesemisclassificationsmayberelatedtopoorquality

9

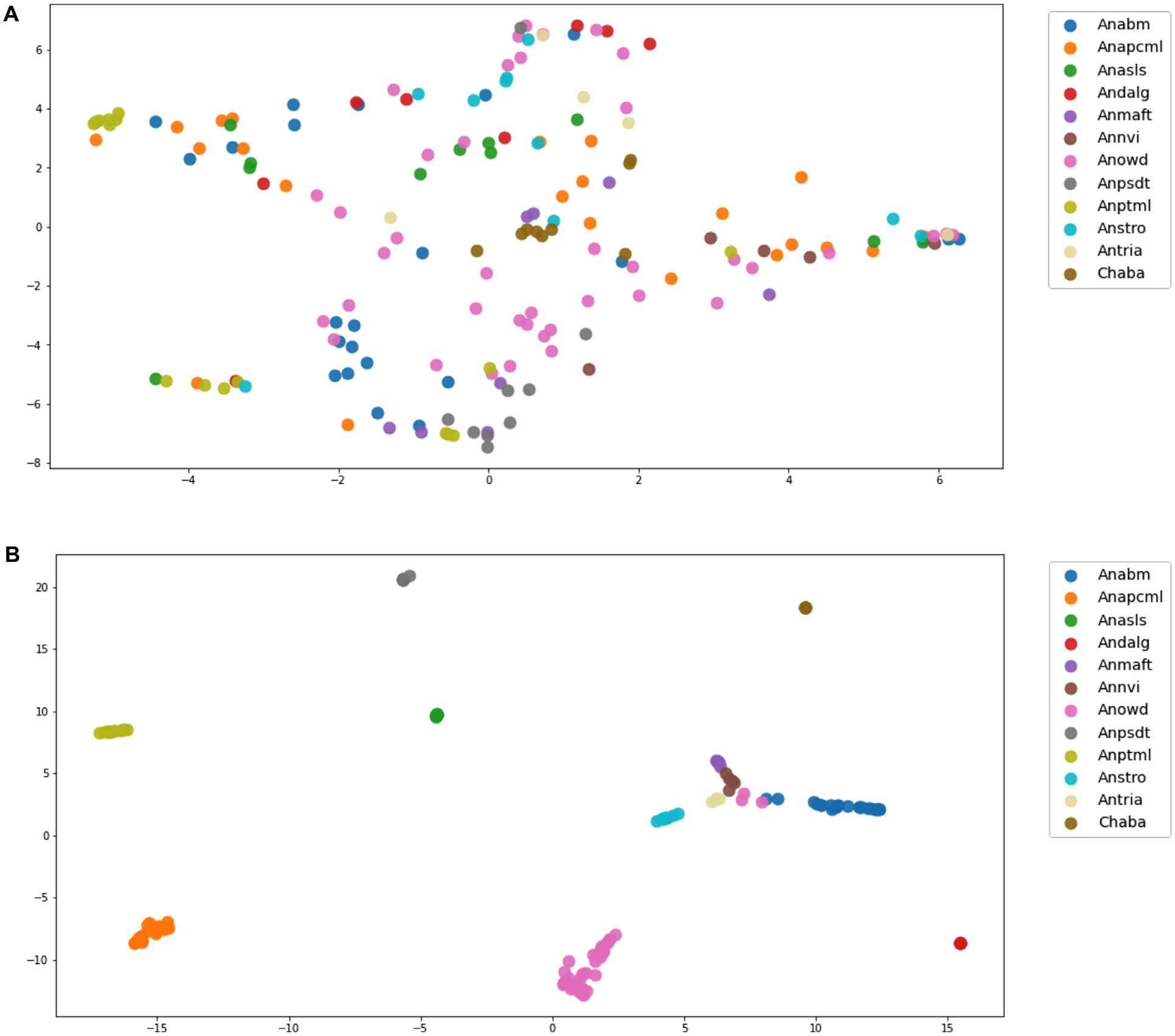
*ArtificialIntelligenceintheLifeSciences3(2023)100071*

**Fig.9.**Confusionmatricesofthe*TNN*modelforthemosquito(A)andticks (B)datasets.

spectraordegradedsamples,andnotnecessarilytobiologically-relevant inferences.

Therearesomeadditionallimitationstothisworkthatneedtobe discussed.Thedatasetsusedinthisworkcomefrombiologicalsam-plesthatweregatheredfromthefieldasopposedtobeingrearedin thelaboratory.Besidesforeigncontaminantsthatarehardtocontrol infield-collectedspecimens(*i.e.*,hostblood,sap,pathogens),whichal-readypresentanimportantchallengetotheclassification,thedatasets alsoallowedustotestthereliabilityofthealgorithmsregardingsam-plestorage.Freshmosquitospecimenshadbeenstoreddry-frozenas opposedtoticksthathadbeenstoredinethanolandfrozenforseveral yearspriortoproteinextractionandsampleprocessingwiththeMALDI-TOF.Ethanolisawell-knownproteindenaturingagentandweknow

*F.Merchan,K.Contreras,R.A.Gittensetal.*  *ArtificialIntelligenceintheLifeSciences3(2023)100071*



**Fig.10.**Visualizationwith*pacmap*ofthe*embedding*producedbythe*TNN*modelbefore(A)andafter(B)trainingforthemosquitosdataset.

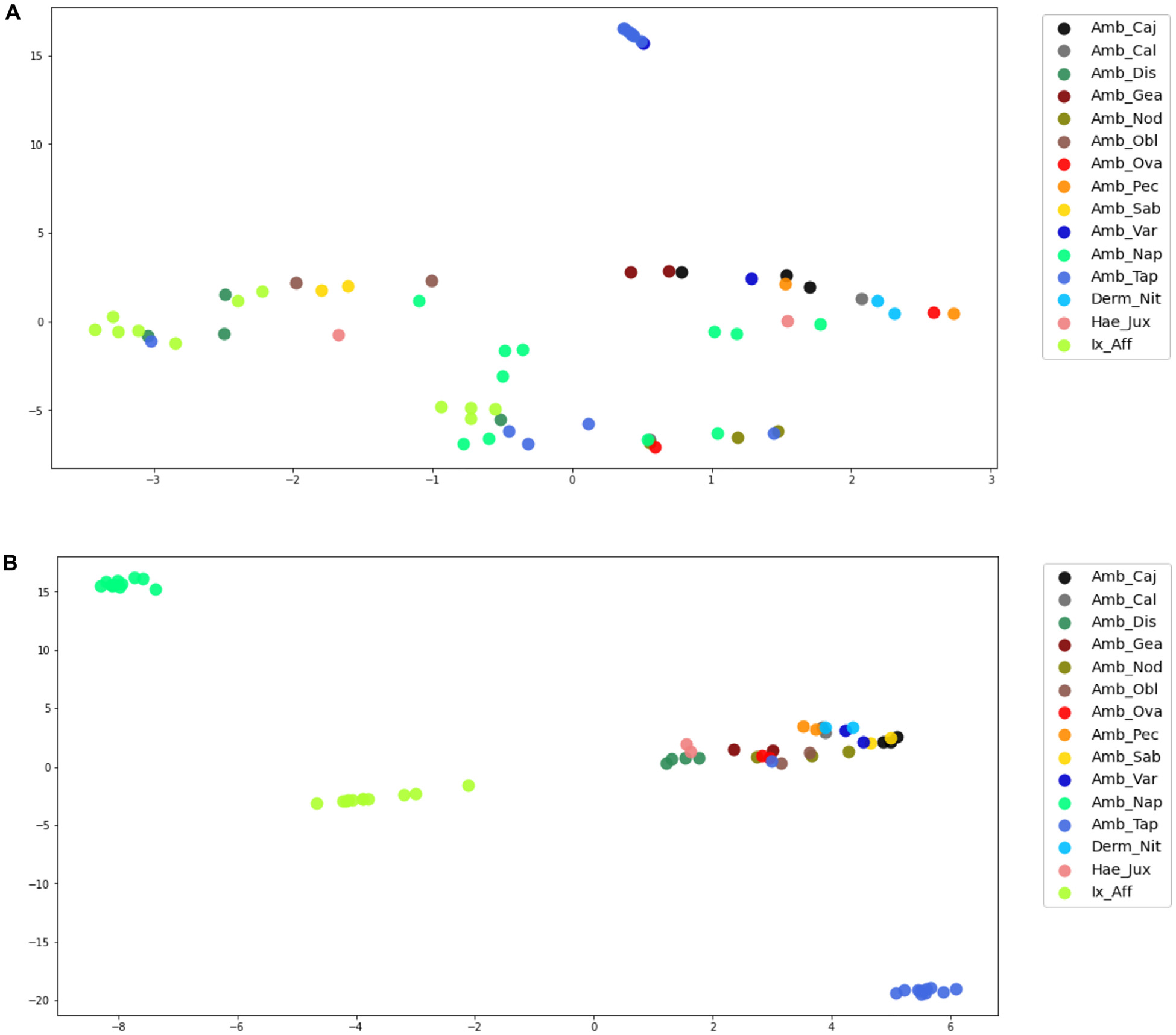
nowthatproteinmassspectraqualityofarthropodspecimensstoredin 70%ethanolisconsiderablyreduced[7,48,49].

Sampledegradation,lowerqualityspectraandlowernumberofspec-imens/spectraperspeciesinthetickdatasetforcedustoadjustthe trainingandtestingdatasplitfortheTNNmodel.Althoughtheini-tialexperimentalplanwastoworkwiththesameratioof80%/20% training/testingsetsformosquitoesandticks,wehadtoadjustthetick datasplitto70%/30%toallowtheTNNmodeltofindtheappropriate tripletsforvalidationandavoidcomputationalerrors.Still,70%/30% or80%/20%arebothtypicaldatasplitsconsidered“adequate” forma-chinelearningalgorithms,especiallywhenusingcross-validationstrate-gies[50–52].

AccordingtoLiuetal.[53],theclassificationofDeepLearningspec-trometrysignalspresentssignificantbenefits,becausetheCNNtraining processactsasaparallelmechanismforpre-processing(*e.g.*,correction ofthesignalbase,curvesmoothing)andfeatureextraction(*e.g.*,peak detection,non-linearfeatures)fromtherawspectra.Therefore,auni-fiedarchitecturewouldeliminatetheneedtoapplythepre-processing schemerequiredwiththeimplementationofclassicalmethodslikePCA andLDA[2,7].Additionally,theeffectivenessofone-dimensionalCNN withaDeepMetricLearning-basedschemawasdemonstratedtomini-mizethedifficultiesintrainingamodelwithareducedorunbalanced dataset[54].Thatis,amodelwithvariableavailabilityofsamplesfor

10

*F.Merchan,K.Contreras,R.A.Gittensetal.*  *ArtificialIntelligenceintheLifeSciences3(2023)100071*



**Fig.11.**Visualizationwith*pacmap*ofthe*embedding*producedbythe*TNN*modelbefore(A)andafter(B)trainingfortheticksdataset.

MALDI-TOFMScoupledwithartificialintelligencehastopredictento-mologicaldriversofdiseasetransmission,providingpotentialnewtools forvectorcontrol.

Tothebestofourknowledge,thisisthefirstattempttoclassify MALDI-TOFdatafromarthropodsfromtheNeotropicsusingDeepMet-ricLearningtechniquesthatemployMetricLearningtechniqueswith neuralnetworksasafeatureextractionmethod,includingmodelsthat combinemultipleneuralnetworkswithidenticalarchitecturescom-monlyknownasSiameseandTripletNeuralNetworks.Ourresultsindi-catethatDeepMetricLearningisyetanotherpracticalmachinelearning toolforquicklyandpreciselyclassifyingMALDI-TOF-generatedmass spectraofpublic-health-relevantarthropodspecies.

**5.Conclusion**

Rapididentificationofarthropodspeciesusingmassspectrafinger-printingcontinuestoofferapromisingtoolforentomologistandpublic-healthspecialist,speciallyinbiodiverseregionssuchastheNeotropics. Oneofthelongtermgoalsofourgroupistodevelopanonline,crowd-sourcedandopen-sourcedatabase,basedonnovelmachinelearning algorithmsthatcancontinuouslygrowandlearnacrossarthropodfam-ilies,andcanimprovetheefficiencyandaccuracyofspeciesidentifica-tionregardlessofsamplepreparationandanalysisparameters.Ourre-

11

*F.Merchan,K.Contreras,R.A.Gittensetal.*

**Dataavailability**

Datawillbemadeavailableonrequest.

**Acknowledgments**

FinancialsupportforthisworkwasprovidedbySENACYTthrough theresearchgrantGRID15-002toJRL,JSG,andRAG.INDICASAT-AIP, UTPandSTRIprovidedadditionaleconomicandlogisticsupport.The SNIsupportsresearchactivitiesbyJRL,JSG,FMandRAG.RAGisalso supportedbySENACYTgrantsFID14-066,ITE15-016.Thefundershad noroleinstudydesign,datacollectionandanalysis,decisiontopublish, orpreparationofthemanuscript.

**References**

[[1]WeaverSC,ReisenWK.Presentandfuturearboviralthreats.AntiviralRes](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0001)  [2010;85(2):328–45.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0002)

[[2]LoaizaJR,AlmanzaA,RojasJC,MejíaL,CervantesND,Sanchez-GalanJE,etal.Ap-plicationofmatrix-assistedlaserdesorption/ionizationmassspectrometrytoiden-tifyspeciesofNeotropicalanophelesvectorsofmalaria.MalarJ2019;18(1):1–12.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0002)

[[3]LainhartW,DutariLC,RoviraJR,SucupiraIM,PóvoaMM,ConnJE,etal.Epidemic](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0002) [andnon-epidemichotspotsofmalariatransmissionoccurinindigenouscomarcas ofPanama.PLoSNeglTropDis2016;10(5):e0004718.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0003)

[[4]MillerMJ,EsserHJ,LoaizaJR,HerreEA,AguilarC,QuinteroD,etal.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0003)

[MolecularecologicalinsightsintoNeotropicalbird–tickinteractions.PLoSONE](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0004) [2016;11(5):e0155989.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0005)

[[5]LoaizaJR,DutariLC,RoviraJR,SanjurOI,LaportaGZ,PecorJ,etal.Disturbance andmosquitodiversityinthelowlandtropicalrainforestofcentralPanama.SciRep](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0005) [2017;7(1):1–13.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0006)

[[6]LoaizaJR,RoviraJR,SanjurOI,ZepedaJA,PecorJE,FoleyDH,etal.Forestdistur-banceandvectortransmitteddiseasesinthelowlandtropicalrainforestofcentral Panama.TropMedIntHealth2019;24(7):849–61.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0006)

[[7]GittensRA,AlmanzaA,BennettKL,MejíaLC,Sanchez-GalanJE,MerchanF,](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0006) [etal.ProteomicfingerprintingofNeotropicalhardtickspecies(Acari:Ixodi-dae)usingaself-curatedmassspectrareferencelibrary.PLoSNeglTropDis 2020;14(10):e0008849.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0007)

[[8]JiménezMG,BabayanSA,KhazaeliP,DoyleM,WaltonF,ReedyE,etal.Prediction](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0007) [ofmosquitospeciesandpopulationagestructureusingmid-infraredspectroscopy andsupervisedmachinelearning.WellcomeOpenRes2019;4.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0008)

[[9]YeS,LuS,BaiX,GuJ.ResNet-Locust-BNnetwork-basedautomaticidentifica](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0008)-[tionofeastasianmigratorylocustspeciesandinstarsfromRGBimages.Insects 2020;11(8):458.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0009)

[[10]GuglielmoneAA,RobbinsRG,ApanaskevichDA,PetneyTN,Estrada-](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0009)Peñ[aA,HorakIG.Thehard](https://doi.org/101007/978-94-007-7497-1)ticksoftheworld.Dordrecht:Springer;2014. [doi](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0011)[:101007/978-94-007-7497-1.](https://doi.org/101007/978-94-007-7497-1)

[[11]Mu](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0011)[rugaiyanJ,RoeslerU.MALD](https://doi.org/101007/978-94-007-7497-1)[I-TOFMSprofiling-advance](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0011)s[inspeciesidentifica](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0011)- [tionofpests,parasites,andvectors.FrontCellInfectMicrobiol2017;7:184.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0011)

[[12]Perez-LaoE,Sanchez-GalanJ,GittensR.Assessingtheperformanceofdifferen](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0011)t[sam-](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0011)[pletargetsforaMALDI-TOFmassspectrometer.In:2014IEEECentralAmericaand PanamaConvention(CONCAPANXXXIV).IEEE;2014.p.1–4.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0012)

[[13]MansillaE.C.,MorenoR.C.,GarcíaM.O.,SánchezB.R.,PérezJ.d.D.C.,Bellido](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0012) J.L.M..Aplicacionesdelaespectrometríademasasmaldi-tofenmicrobiología [clínica.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0013)

[[14]LevasseurM,HebraT,ElieN,GuérineauV,TouboulD,EparvierV.Classifica](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0013)-[tionofenvironmentalstrainsfromordertogenuslevelsusinglipidandprotein MALDI-ToFfingerprintingsandchemotaxonomicnetworkanalysis.Microorganisms 2022;10(4):831.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0013)

[[15]YssoufA,AlmerasL,RaoultD,ParolaP.Emergingtoolsforidentificatio](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0013)n[ofarthro-](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0013) [podvectors.FutureMicrobiol2016;11(4):549–66.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0014)

[[16]BennettKL,GómezMartínezC,AlmanzaA,RoviraJR,McMillanWO,EnriquezV,](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0014) [BarrazaE,DiazM,Sanchez-GalanJE,WhitemanA,etal.Highinfestationofinva-siveAedesmosquitoesinusedtiresalongthelocaltransportnetworkofPanama. ParasitesVectors2019;12(1):1–10.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0015)

[[17]DiemeC,YssoufA,Vega-RúaA,BerengerJ-M,FaillouxA-B,RaoultD,ParolaP,](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0015) [AlmerasL.Accurateidentificationofculicidaeataquaticdevelopmentalstagesby MALDI-TOFMSprofiling.ParasitesVectors2014;7(1):1–14.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0016)

[[18]SevestreJ,LemrabottMAO,BérengerJ-M,ZanDiarraA,OuldMohamedSalem](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0016) [BoukharyA,ParolaP.Detectionofarthropod-bornebacteriaandassessmentof MALDI-TOFMSfortheidentificationoffield-collectedimmaturebedbugsfrom mauritania.Insects2023;14(1):69.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0017)

[[19]BittremieuxW.Spectrum\_utils:aPythonpackageformassspectrometrydatapro-](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0017) [cessingandvisualization.AnalChem2019;92(1):659–61.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0019)

[[20]RàfolsP,VilaltaD,BrezmesJ,CañellasN,DelCastilloE,YanesO,etal.Sig-nalpreprocessing,multivariateanalysisandsoftwaretoolsforMA(LDI)-TOF massspectrometryimagingforbiologicalapplications.MassSpectromRev 2018;37(3):281–306.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0019)

[[21]YssoufA,ParolaP,LindströmA,LiljaT,LAmbertG,BondessonU,etal.Identificatio](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0019)n [ofeuropeanmosquitospeciesbyMALDI-TOFMS.ParasitolRes2014;113(6):2375–8.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0020)

*ArtificialIntelligenceintheLifeSciences3(2023)100071*

[22][TurkM,PentlandA.Eigenfacesforrecognition.JCognNeurosci1991;3(1):](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0021)  [71–86.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0021)

[23][BelhumeurP,HespanhaJ,KriegmanD.Eigenfacesvs.Fisherfaces:recognitionusing](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0021) classsp[ecifi](https://doi.org/10.1109/34.598228)c[linearprojectio](https://doi.org/10.1109/34.598228)n.IEEETransPatternAnalMachIntell1997;19(7):711–[20.doi](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0023)[:10.1109/34.598228.](https://doi.org/10.1109/34.598228)

[24][López-F](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0023)[ernándezH,Santos](https://doi.org/10.1109/34.598228)[HM,CapeloJL,Fdez-RiverolaF,Glez-PeñaD,Reboiro–JatoM.Mass-up:anall-in-oneopensoftwareapplicationforMALDI-TOFmassspec-trometryknowledgediscovery.BMCBioinformatics2015;16(1):1–12.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0023)

[25][LeCunY,BengioY,HintonG.Deeplearning.Nature2015;521(7553):436–44.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0023) [26][KayaM,BilgeHŞ.Deepmetriclearning:asurvey.Symmetry(Basel)](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0025)  [2019;11(9):1066.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0025)

[27][CoverT,HartP.Nearestneighborpatternclassification](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0025).[IEEETransInfTheory](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0025)  [1967;13(1):21–7.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0026)

[28][RakotonirinaA,CaruzzoC,BallanV,KainiuM,MarinM,ColotJ,etal.Wolbachia](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0026) [detectioninAedesaegyptiusingMALDI-TOFMScoupledtoartificialintelligence. SciRep2021;11(1):1–11.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0027)

[29][KhalighifarA,KompE,RamseyJM,Gurgel-GonçalvesR,PetersonAT.Deeplearn-](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0027)[ingalgorithmsimproveautomatedidentificationofchagasdiseasevectors.JMed Entomol2019;56(5):1404–10.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0028)

[30][LeCunY,BengioY,HintonG.Deeplearning.Nature2015;521(7553):436–44.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0028) [31][LiuK,LiS,WangL,YeY,TangH.Full-spectrumpredictionofpeptidestandem](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0029)[mass](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0030)  [spectrausingdeepneuralnetwork.AnalChem2020;92(6):4275–83.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0030)

[32][LeCunY,etal.LeNet-5,convolutionalneuralnetworks.URL:http://yannlecun](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0030)  [com/exdb/lenet2015;20(5):14.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0031)

[33][DubeyS.R.,SinghS.K.,ChaudhuriB.B..Acomprehensivesurveyandper-](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0031)[formanceanalysisofactivationfunctionsindeeplearning.arXivpreprint arXiv:2109145452021;.](http://arxiv.org/abs/arXiv:210914545)

[34][LewkowyczA,Gur-AriG.Onthetrainingdynamicsofdeepnetworkswith](http://arxiv.org/abs/arXiv:210914545)*[𝑙](http://arxiv.org/abs/arXiv:210914545)*[\_2regu-](http://arxiv.org/abs/arXiv:210914545) [larization.AdvNeuralInfProcessSyst2020;33:4790–9.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0032)

[35][ZhangD,LiY,ZhangZ.Deepmetriclearningwithsphericalembedding.AdvNeural](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0032)  [InfProcessSyst2020;33:18772–83.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0033)

[36][ByrdJ,LiptonZ.Whatistheeffec](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0033)t[ofimportanceweightingindeeplearning?.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0033)

In:ChaudhuriK,SalakhutdinovR,editors.Proceedingsofthe36thinternational conferenceonmachinele[arning.ProceedingsofMachineLearningResearch](https://proceedings.mlr.press/v97/byrd19a.html),vol.97.

[PMLR;2019.p.872–81.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0035)<https://proceedings.mlr.press/v97/byrd19a.html>  
[37][SanturkarS,TsiprasD,I](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0035)[lyasA,MadryA.Howdoesbatchnormalizationh](https://proceedings.mlr.press/v97/byrd19a.html)[elpopti-](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0035) [mization?AdvNeuralInfProcessSyst2018;31.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0035)

[38][KingmaD.P.,BaJ..Adam:amethodforstochasticoptimization.arXivpreprint](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0035)  [arXiv:141269802014;.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0036)

[39][HadsellR,ChopraS,LeCunY.Dimensionalityreductionbylearninganinvariant mapping.In:2006IEEEComputersocietyconferenceoncomputervisionandpattern recognition(CVPR’06),Vol.2.IEEE;2006.p.1735–42.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0036)

[40][BujaA,StuetzleW,ShenY.Lossfunctionsforbinaryclassprobabilityestimation](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0036) [andclassification:structureandapplications.In:Workingdraft,Vol.3;2005.p.13.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0037)

[November](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0037)   
[41][Schroff F](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0037),[KalenichenkoD,PhilbinJ.FaceNet:aunifie](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0038)d[embeddingforfacerecogni-](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0038) [tionandclustering.In:ProceedingsoftheIEEEconferenceoncomputervisionand](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0038)  [patternrecognition;2015.p.815–23.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0038)

[42][ShanG.Montecarlocross-validationforastudywithbinaryoutcomeandlimited](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0038)  [samplesize.BMCMedInformDecisMak2022;22(1):1–15.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0039)

[43][AbadiM.,AgarwalA.,BarhamP.,BrevdoE.,ChenZ.,CitroC.,CorradoG.S.,Davis](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0039)  A.,DeanJ.,DevinM.,GhemawatS.,GoodfellowI.,HarpA.,IrvingG.,IsardM.,Jia Y.,JozefowiczR.,KaiserL.,KudlurM.,LevenbergJ.,Mané D.,MongaR.,MooreS., MurrayD.,OlahC.,SchusterM.,ShlensJ.,SteinerB.,SutskeverI.,TalwarK.,Tucker P.,VanhouckeV.,VasudevanV.,ViégasF.,VinyalsO.,WardenP.,WattenbergM., WickeM.,YuY.,ZhengX..TensorFlow:large-scale[machinelearningonheterog](https://www.tensorflow.org/)e- neoussystems.2015.SoftwareavailablefromURL:[https://www.tensorflow.org](https://www.tensorflow.org/)/[.](https://www.tensorflow.org/) [44]VirtanenP,GommersR,OliphantTE,HaberlandM[,ReddyT,CournapeauD,eta](https://www.tensorflow.org/)l.

SciPy1.0:fundament[alalgorithmsforscientifi](https://doi.org/10.1038/s41592-019-0686-2)c[co](https://doi.org/10.1038/s41592-019-0686-2)mputinginPython.NatMethods [2020;17:261–72.doi](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0041)[:10.1038/s41592-019-0686-2.](https://doi.org/10.1038/s41592-019-0686-2)

[45][PedregosaF,Varoqu](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0041)[auxG,GramfortA,MichelV,](https://doi.org/10.1038/s41592-019-0686-2)[ThirionB,GriselO,etal.Scik-](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0041) [it-learn:machinelearninginPython.JMachLearnRes2011;12:2825–30.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0041)

[46][BisongE.BuildingmachinelearninganddeeplearningmodelsonGooglecloud](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0041)  [platform:acomprehensiveguideforbeginners.Apress;2019.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0042)

[47][WangY,HuangH,RudinC,ShaposhnikY.Understandinghowdimensionre-](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0042)ductiontoolswork:anempiricalapproachtodecipheringt-SNE,UMAP,TriMAP, [andPaCMAPfordatavisualization.J](http://jmlr.org/papers/v22/20-1061.html)MachLearnRes2021;22(201):1–73.

<http://jmlr.org/papers/v22/20-1061.html>  
[48][RakotonirinaA,PolM,KainiuM,BarsacE,](http://jmlr.org/papers/v22/20-1061.html)[TutagataJ,KilamaS,OconnorO,Taran-](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0044) [tolaA,ColotJ,Dupont-RouzeyrolM,etal.MALDI-TOFMS:optimizationforfuture](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0044)  [usesinentomologicalsurveillanceandidentificationofmosquitoesfromnewcale-](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0044) [donia.ParasitesVectors2020;13:1–12.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0044)

[49][YssoufA,SocolovschiC,LeulmiH,KernifT,BitamI,AudolyG,etal.Identifi](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0044)-[cationoffleaspeciesusingMALDI-TOF/MS.CompImmunolMicrobiolInfectDis](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0045) [2014;37(3):153–7.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0046)

[50][LiuL,ÖzsuMT.Encyclopediaofdatabasesystems,Vol6.Springer;2009.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0046)

[51][ChiccoD.Tenquicktipsformachinelearningincomputationalbiology](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0046)[.BioData](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0047)  [Min2017;10(1):35.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0047)

[52][AnC,ParkYW,AhnSS,HanK,KimH,LeeS-K.Radiomicsmachinelearningstudy](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0047) [withasmallsamplesize:singlerandomtraining-testsetsplitmayleadtounreliable results.PLoSONE2021;16(8):e0256152.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0048)

[53][LiuJ,GibsonSJ,MillsJ,OsadchyM.Dynamicspectrummatchingwithone-shot](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0048)  [learning.ChemomIntellLabSyst2019;184:175–81.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0050)

[54][LiuJ,OsadchyM,AshtonL,FosterM,SolomonCJ,GibsonSJ.Deepconvolu-tionalneuralnetworksforramanspectrumrecognition:aunifiedsolution.Analyst 2017;142(21):4067–74.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0050)

12

*F.Merchan,K.Contreras,R.A.Gittensetal.*

[[55]KotsiantisS,KanellopoulosD,PintelasP,etal.Handlingimbalanceddatasets:a](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0051)  [review.GESTSIntTransComputSciEng2006;30(1):25–36.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0051)

[[56]NabetC,ChalineA,FranetichJ-F,BrossasJ-Y,ShahmirianN,SilvieO,etal.Predic-](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0051)[tionofmalariatransmissiondriversinanophelesmosquitoesusingartificialintelli-gencecoupledtoMALDI-TOFmassspectrometry.SciRep2020;10(1):1–13.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0052)

*ArtificialIntelligenceintheLifeSciences3(2023)100071*   
[57][RakotonirinaA,CaruzzoC,BallanV,KainiuM,MarinM,ColotJ,etal.Wolbachia detectioninAedesaegyptiusingMALDI-TOFMScoupledtoartificialintelligence. SciRep2021;11(1):1–11.](http://refhub.elsevier.com/S2667-3185(23)00015-6/sbref0053)

13