Scopus

ELSEVIER

Journal of Vibration Engineering

ISSN:1004-4523

Registered

- SCOPUS
- GOOGLE SCHOLAR
- DIGITAL OBJECT IDENTIFIER (DOI)
- IMPACT FACTOR 6.1

Our Website
www.jove.science

"AnalysisOfActivationFunctionInTheConvolutionNeuralNetwork Model"(BestActivationFunction SigmoidOrTanh)

Archana Tomar[1],HarishPatidar[2]

[1]*DepartmentofComputerScience,MandsaurUniversity,India*

E-mail:archana.simmy@gmail.com

[2]*DepartmentofComputerScience,MandsaurUniversity,India*

Abstract

The activation functions, which are mathematical formulae, governhow a neural network responds. A particular sort of function is calledan activation function. In ANN, the Activation function performs theconversiontaskoftheinputsignalintoanoutputsignal..Thesucceedinglayerofthestackthenusesthisoutputsignalasinput.Everyneuron in the network is connected to the function, and it controlswhetherornotaneuronfiresaccordingtohowcrucialitsinputistotheprediction the model is making.In this paper, an explanation of themost used Activation Function which is sigmoid and Tanh. With thehelp of the activation functions, which also serve a secondary functionintheprocess,Eachneuron'soutputcanbenormalizedto fallintooneof two ranges: either 0 to 1 or -1 to 1. For prediction best activation insigmoid and tanh applied in multiple dataset like MANIST, FER etcdata set with different epochs and identify the result of both Activationfunctions. The sigmoid, tanh, and Relu activation functions—the threemostcommonactivationfunctions—haveallbeensubjected"toanalysis.

Keywords:

Activationfunctions,Neurons,CNN,RELU,Sigmoid,Tanh

1. INTRODUCTION

InArtificialNeuralnetworksweneedmorecomputationtotrainlargesize image dataset and it is also not suitable for fetching all detailsfromimages,CNNisamodelinNNwhichisusedforfetchingsomespecial dependent information and it is also used in classification ofdata.

In the CNN model perform multiple filtration and in this filtrationprocess gain different types of extraction it may be size or shape oranything which provides information for determining the input orpicture.ThemainpropertiesofCNNishowtoidentifypicturescorrectly andperformclassification.soinpictureidentificationmultiplelayerswork together.So,thelogicbehindtheCNNis,Firstly,providesomeimagesasan inputandpassthemtotheinputlayer.After that in this stage, feature extraction or filtration is appliedininputwhichcoverstheconvolutionlayerandActivationfunction.

This function controls how a neuron reacts to the information it hasreceivedfromthelayerbeneathit.Theconstructionofneuralnetworks has involved the use of several activation functions ofvarioustypes.Thethreefunctionsthatmakeuptheclassicalactivation functions are the step function, sigmoid function, and tanhfunction. The sigmoid functions smoothness makes it easier to createlearning algorithms

After receiving the result of activation function application somepooling layers are applied in this model and the result of this layer isoptimized feature map and this is used for preventing and optimizingthetime.Nowheretheclassifierworksbyapplyingthefullyconnected layer which is performed as multilayer processing and theuse of a fully connected layer is to classify the picture. and at the lastperformanoutputlayerforprovidingaccurateprediction.This process is done again and again in CNN model to train our data forpredicting accurateresults.

ThemajorityofexpertsagreethattheCNNnetworksactivation

functionhasanimpactonclassificationaccuracyaswellashowlongittakestotrainthenetwork.AlthoughtheperformanceofCNNcouldbe impacted by the findings of this study, that possibility is not beingfully explored. To address a vacuum in the pertinent research, weexamine how activation functions affect the CNN model[4] datasetspredictionaccuracy.Duringthisphase,aconvolutionalneuralnetwork(CNN)isconstructedanditslayersareeachgivenadifferentactivation function.

2. LITERATURESURVEY

In this article Author Albawi[1], explained in detail about the CNNand their issues with all parameters or elements and also explainedabouttheirworking whichareusedinCNN.

convolutionlayerplaysanimportantroleintheCNNnetworkmodeland this layer work like time taken so whole CNN process will focusonconvolutionlayerworkingherewholenetworkortheirperformancedependentontheirworkingsoinCNNmultilayeroperation perform for prediction accurate result and for that multiplelayerworktogetherbutduetomultilayeringworkingtimealsoincreases.

AuthorR.Chauhan[2]implementsCNNonMNISTdatasetforimagedetectionandrecognition.After evaluation, CNN model accuracycalculated inMNIST dataset is about 99 percent and in CIFAR-10dataset it is around 80 percent. The author also explains that forcalculating the accuracy of the CIFAR-10dataset it uses real-timedata and dropout on CPU units. and for final conclusion authorevaluate the better result of both dataset like MNIST and CIFAR-10and he analyze that MNIST provide better performer result than theCIFAR-10 and if we trained our model with GPU unit so we willdefinitely improvedtheirperformance

Author Ajit A.[3]explained about the feature extraction from thegivendataandhealsodefinedthatnowadaysforanyobjectdetectionor any image recognition CNN has been used. He also defines theworking of algorithms in a step like in this model including multipleprocedurewhichinvolve

1. Backpropagationconcept
2. ConvolutionalLayers
3. Featureformation
4. Pooling.

In this article the author also explains about the different types ofarchitecture of CNN[23] like LeNet,AlexNet,VGG Net, Google Netand Microsoft ResNet with detailed explanation about how CNNworksinall these architectures.

AuthorKaramA.-F.[5],involveasupervisedlearningconceptinthisarticle. During research the author started with techniques which aregenerally involved in model construction which provide the highaccuracyandfinalizethebesttechniqueforimageclassification.authoralsoincludeoptimizationtechniqueinthisresearchwhich involve

Learning Rate

ReductionTraining

Generator:

Preparingdata:

FittingtheModel:

Heretheauthorconcludedthat

If you are working with zoomed pictures, here add more parametersfor predicting accurate results and in this to ignore or avoid the over-fitting model data augmentation technique are used.so here authorpredict the accuracy of ADAM is 0. 95 percent and this is better thanSGDandRMSprop.

Feng et al[6] author analyzes the performance of multiple activationfunctionsinANN.Inthispaperauthordefinethepartofvarioussortsofactivationfunctions,justastheirindividualpreferencesanddetriments and applicable fields are additionally talked about, so thatindividuals can pick the most suitable activation functions to get thesuperior performance of ANNs.Theauthoranalyzesvariousactivationfunctionattributesandclarifiestheirfocalpointsandimpediments.Everyactivationfunctionhas its own qualities, so the author can only with significant effortstatewhichoneisthemoresuitable,theonlythingitcandoispickinganappropriateactivationfunctiondependentonourpointandexplicitorganizationstructure.0

Rasamoelina. et al[7] Review of Activation Functions for ArtificialNeuralNetworksInthispaper,theauthorspresentthemostcommonly used state-of-the-art trend activation functions commonlyusedinhiddenlayersofartificialneuralnetworks.Showsareviewofthefunction.Basically,theauthordescribedhertwoclassesofactivationfunctions:derivatives,zero-centeredsaturatedmonotonicity. Next, the author also confirmed her 8 major activationfunctions. According to the author, the SELU, Swish, and Mishexperimental results landscapes provide smoother output. Note alsothat an automatic search gives satisfactory results for finding theactivation function.

etal[8]authorproposedanddesignedanewAF(activationfunction)called RELu.Sointhisfunctiontheauthorintroducestwoparameters,one is constant and The second threshold parameter smoothes thefunction, makes it non-monotonic, and introduces non-linearity intothe network. The results of the experiments show that the Relu-Memristor-Like Activation Function for Deep Learning outperformsReLU and other activation functions on deeper models and across avarietyofchallengingdatasets.Asinthispaper,theauthorconductedpracticalexperimentsbytrainingandclassifyingamulti-layerperceptron(MLP)onbenchmarkdataliketheWisconsinbreastcancer,MNIST, Iris and Carevaluation. The RMAF is based on the properties of the memristive windowfunction,whichcanadapttodeepnetworks.andthefunction'sthresholdparameterpallowsnegativerepresentationstoflowthroughthe network during forward propagation and is capable of scaling toanygivennetwork.TheseRMAFpropertiesallownetworkstobenefitfrom negative representation.

Sharma. et al[9] ACTIVATION FUNCTIONS IN NEURAL NETWORKSInthisarticle,brieflydescribevariousactivationfunctionsusedinthefieldofdeeplearninganddiscusstheimportanceof activationfunctionsin developingeffectiveandefficientdeeplearning models. It also describes gender. Performance of artificialneural networks. In this article, the authors emphasize the need foractivation functions and the need for nonlinearity in neural networks.First, the author describes activation functions, then briefly discussestheneedforactivationfunctionsandtheneedfornonlinearityinneuralnetworks. Next, we discuss different types of activation functionscommonly used in neural networks. Therefore, the authors focus hereontheuseofactivationfunctionsinthecomputational(hidden)layersofneuralnetworksandtheiractualplacementin differentdomains.

Relu AF (activation function) and Tanh AF have certain correlatedcharacteristics, according to Li. et al[10] (activation function). TheTanh Activation function's results might increase the qualities Reluunits activate and decrease the qualities they cut. The author alsodemonstrateshowthenetworkscouldsignificantlyimprove.byconverting the Tanh activation function into the weighted sum of theRelu activation function. So the author ran a series of experiments onsome datasets, and the results show that this method could improveResNet's accuracy. When the Relu unit's output is positive, the yieldoftheTF(Tanhfunction)isalsocertain,whichcouldimprovetheRelu unit's outcome. When the result of the Relu unit is zero, the result ofthe Tanh function is negative, which may introduce the informationdiscardedbyRelu.TheweightedsumoftheReluandTanhactivationfunctionscouldimprovethepopulardeepconvolutionalneuralnetworks.

3. CNNMODEL

In the CNN[22] model the first layer works as an input layer. In thislayer we collect all the images and input in the form of RGB or graycolorscale.

3.1 TYPESOFLAYERS:

INPUTLAYER:
Firstly collect all images and after that arrange in a pixel format andthe range of this format is 0-255 and before calculation in a machinethis range is represented or transformed into 0 or 1 format du etomachinecalculation.
thisisanexampleof4*4 imagewith threecolorchannel

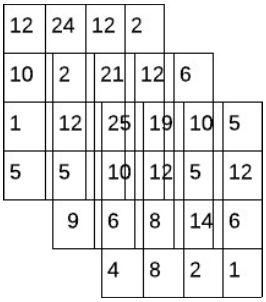

Figure 1: 4*4 image color

channelCONVOLUTIONLAYER:
For extracting the features of images in this layer filtration is appliedand extract information from this. and this filtration is perform manytimesinconvolutionlayer

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix} * \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & -4 & -4 & 0 \\ 0 & -4 & -4 & 0 \\ 0 & -4 & -4 & 0 \\ 0 & -4 & -4 & 0 \end{bmatrix}$$

Figure2:Filtration in6*6 imageset

Thisis1stepoffiltrationbutinthismodelthesameprocessisappliedmultipl e times.in this process filtration is applied like image value setandfilteredvaluelike1*0+andinthesameperformcalculationand design afinalmap.

POOLINGLAYER:
after detecting a final map its converted to reduced or optimizedfeature map or it can be a max or average pooling here we select avgor max value from a final feature map like from set of 4 values 0,-4,-4,0 so the max value is 0,0,0,0 and the average value like-2,-2,-2,-2.and this is a final optimized image set and now apply final connectedlayer or we can say here perform classification using this layer andretrieveoutputbasedon Activationfunction[16].

3.2 ACTIVATION

FUNCTIONSSIGMOIDACTIVATION

FUNCTION

The"activationfunction[14]istheonethatismostusuallyusedsinceit is a non-linear function. Values between 0 and 1 can be changedwiththeSigmoid function.Itmay be"characterisedasfollows:

f(x)= 1/e-x

The "sigmoid function[12] is continuously differentiable and has theshapeofanS-curve.Theequationforthederivativeof"thefunction,

f'(x)=1-sigmoid(x)

Additionally,because"thesigmoidfunctionisnotsymmetricaroundzero, allofthevaluesgeneratedbyneuronswillhavethesamesigns.Thesigmoid functioncanbescaledasonesolutiontothisproblem.

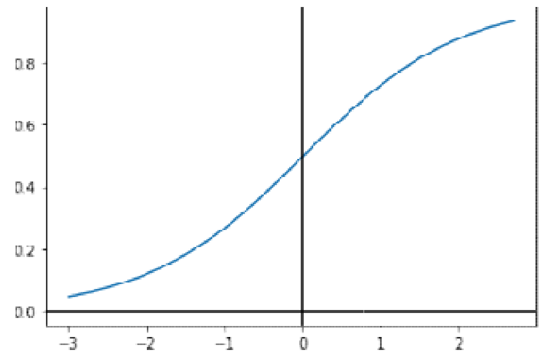

Figure3:SigmoidFunction

TANHFUNCTION

Theappropriatefunctionisthehyperbolictangent.TheTanhfunctionis symmetric to the origin unlike the sigmoid function, which issimilartoit.Thisaltersthesignsoftheoutputsofthelayersthatcamebefor e it, which are then utilized as inputs for the layer that follows.Itmight bestated"asfollows:

f(x)=2sigmoid(2x)-1

The "Tanh[11]" function may take on any value between -1 and 1,and it is continuous and differentiable. The gradient of the Tanhfunctionismoreabruptwhencomparedtothegradientofthesigmoid function. Tanh is preferred over the sigmoid function because it iszero-centered and hasgradientsthatcan alterin any direction.

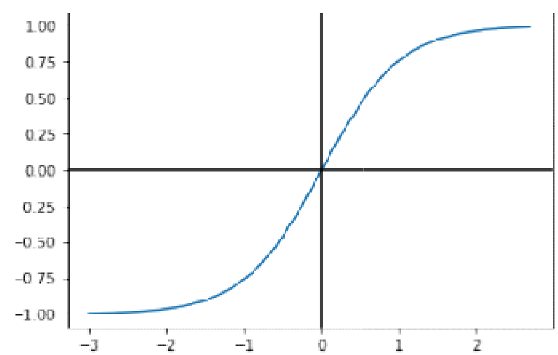

Figure4:TanhFunction

ACTIVATION FUNCTIONS BASED ON LOGISTIC SIGMOIDANDTANH

In "the early phases of neural networks, conventional AFs[19] likeLogisticSigmoidandTanhwereoftenemployed.Despitetheirenorm ousoutput,theseAFsmadedeepneuralnetworktrainingdifficult. There have also been other efforts to enhance these AFs forothernetworks.Theparametric,monotonic,smooth,andlimited features of the Tanh based and Logistic Sigmoid AFs are contrasted.The limitations of Tanh's[13] output range and zero gradient areaddressedbyusingsTanh,ascaledversionofthehyperbolic"tangent.

$sTanh(x)=A\times Tanh(B\times x)$ .... (1)

comprising"the[A,A]outputrange.Theparametricsigmoidfunction (PSF) has been defined as a continuous, differentiable, andbounded"function.[17]

$PSF(x)=\frac{1}{(1+e^{-x})^m}$ ……(2)

The "gradient flow is shown to get better when m is increased to alargervalue,whichisahyperparameterThesymmetryoftherecentlyformed features is preserved by investigating the sum of the shiftedlog sigmoidforuseasan AF

SOFTMAXFUNCTION

AnothertypeofActivationfunctionissoftmax.It'sbasicallyamathematical function which is used in neural networks output layertocalculateandevaluatetheprobabilitydistribution.Anditalsoused inclassificationproblemresolutions.AndTheoutputofthisfunction is in the range of 0 and 1.This function is applied ineveryclassneuron. Thesoftmax functionisintheformof:

f(x)= exp(x)/Σexp(x)

MISHACTIVATIONFUNCTION

TheMishactivationfunction,anuniqueself-regularizednon-monotonicactivationfunctionthathasthemathematicalformula

f(x)=xtanh(softplus(x)).

As a result, this function has many similarities to Swish and GELU,includingtheunboundedpositivedomain,boundednegativedomain,non-monotonicshape,andsmoothderivative.Accordingtotheauthor[24]ofthisstudy,MishthereforeyieldsbetterempiricalresultsthanSwish,ReLU,andLeakyReLUunderthemajorityofexperimentalconditions.

SWISH

Swish[15]activationfunctionsapplythepropertiesofonesidedboundednessatzerowithsmoothnessandalsoapplynon-monotonicity, which may assume a role in the noticed adequacy ofSwish and comparable AF(activation capacities). In the experimentalresult the author[15] shows that Swish will in general work in a waythat is better than ReLU on more profound models across varioustesting datasets.

f(x)=x.sigmoid(x)

RECTIFIEDLINEARUNIT(RELU)FUNCTION

The rectified linear measure (ReLU) function, one of the mostpopular AFs in DL models, is a fast-learning AF that promises todelivercutting-edge performance withstellarresults. Indeeplearning,ReLUfunctionsoutperformotherAFssuchasthesigmoidandtanhfunctionsintermsofperformanceandgeneralization. Formakingiteasiertooptimizebyusinggradientdescentmethodwe use Relu function and it is a linear function that holds thepropertiesof linear models.

TheReLUfunctionappliesineachinputelementandtheoutputofthisfunction isthe sameastheinputvaluesifitisgreaterthan zerootherwiseitissetas0.So,theReLUis

f(xj)=max(0,xj)xj
,if xj>=0
0,if xj<0

Table.1.Thefollowingtableprovidesasummaryoftheactivationfunctionsthatarebased ontheRectified LinearUnit

Name	Parametric	Monotonic	Smooth	Bounded
Rectified LinearUnit(ReLU) ,2010	Notpresent	Present	Notpresent	ForNegativeInputs
Leaky ReLU(LReLU) ,2013	Notpresent	Present	Notpresent	Notpresent
Parametric ReLU(PReLU), 2015	**	Present	Notpresent	Notpresent
RandomizedReLU (RReLU),2015	Notpresent	Present	Notpresent	Notpresent
ConcatenatedReLU (CReLU),2016	Notpresent	Present	Notpresent	ForNegativeInputs
Bounded ReLU(BReLU) ,2016	Notpresent	Present	Notpresent	Yes
Parametric TanhLinearUnit (PTELU),2017	Present	Present	Present	ForNegativeInputs
Flexible ReLU(FReLU) ,2018	Present	Present	Notpresent	ForNegativeInputs
Elastic ReLU(EReLU) ,2018	Notpresent	Present	Notpresent	ForNegativeInputs
RandomlyTranslation ReLU(RTReLU), 2018	Notpresent	Present	Notpresent	ForNegativeInputs
Dual ReLU(DualReLU) ,2018	Notpresent	Present	Notpresent	Notpresent
Paired ReLU (PairedReLU), 2018	Present	Present	Notpresent	Notpresent
Average BiasedReLU (ABReLU),2018	Notpresent	Present	Notpresent	ForNegativeInputs

2019				
Lipschitz ReLU(L-ReLU),2020	Present	Dependsuponϕand η	Depend supon ϕandη	Depend supon ϕandη

Relu 2010 Not even a smooth graph predictor and parametric butitismonotonicsimilarly totheremaining Functionsdetail.

There are no. of Relu transformation functions also like dynamicRectifiedLinearUnit[18],integralRelu[20],ReS[21].Asperthe analysisreportReluFunctionprovidesbetterresultforpositivevaluesbutstillaproblemwithnegativevaluessointhisanalysisfirstlydetailfocusonsigmoidandtanhfunctionforevaluatingtheirperformance.

DIFFERENCES BETWEEN SIGMOID AND THE TANHACTIVATIONFUNCTION

The“waythetwofunctions'gradientsbehaveisakeydistinction betweenthem.

The“gradientsofthesigmoid(red)andtanh(blue)activationfunctionsare plotted”below:

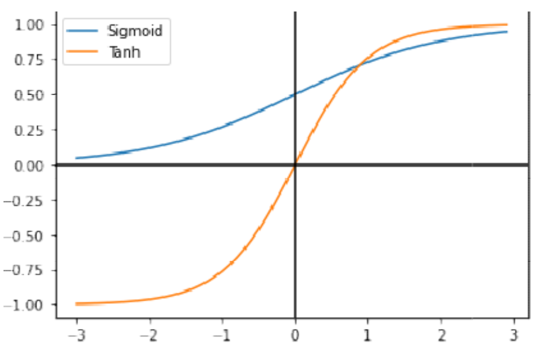

Figure5.Thesigmoidandthetanhactivationfunction

Thedatatendtoclusteraroundthenumberzerowhenweapplytheseactivationfunctionstoa neuralnetwork. Thisisso becausethenetwork's origin is zero. As a result, we must concentrate the vastmajorityofoureffortsonhoweachgradientbehaveswhenitisclose”to zero.

It “has been pointed out to us that the gradient of the tanh function isaround four times bigger than the gradient of the sigmoid function.Throughoutthewholetrainingphase,usingthetanhactivationfunction produces higher gradient values and larger updates to thenetworkweights.Therefore,thetanhactivationfunctionistheoneweshouldemployifwewantourgradientstobereliableandourlearningstepstobesignificant.Anotherdistinctionbetweenthesetwomethods is the output of tanh, which is symmetric about zero andleads to a faster convergence. Another distinction between the twomethods”isthis.

4. *COMPARATIVE PARAMETERIZED RESULTS OF SIGMOIDANDTANHDATASET*

For implementing here we used numeric or mnist dataset. which are inthe form of images some sample images shown below in figure 6.Thedatasetis groupedtogetherto identify numbers.

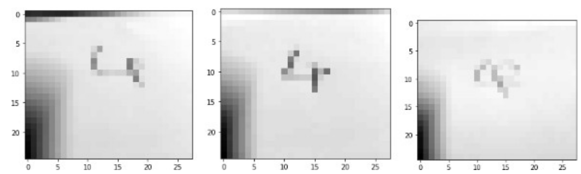

Figure6:Sampleimageofdataset

In the FEB data set, 48 by 48 pixels is the resolution of each image.The effectiveness of the model will be affected by the comparison ofseveraloptimizersinthetest'ssubsequentstage.Itassistsinidentifying the optimizer with the best all-around performance. Thevalidationaccuracyandvalidationlossfiguresarecalculatedtoassesstheefficiencyofthedifferentoptimizerandhyper-parametersettings.In order to assess how effective our proposed constrained CNN is atspottinginstancesofpicturetampering,weranаnumberofexperimentsandransome”analyses.

Figure7:SampleimagesfromFERdataset

4.1RESULTSANDDISCUSSION

The“technicalfeasibilityoftheproposedmixedactivationfunction wasinvestigatedthroughanexperimentalinquiry.

FirstlydiscussaboutMNISTdatasetresult
so as a resultant in CNN first convolution layer use Tanh Activationfunctionandchecktheaccuracyofclassificationoverthismodelandafterthat us Relu and tanh in fully connected and output layer and check theaccuracy of classification here input shape is used as 28*28 also uses thesameoptimizeras Adaminthismodel.

and then compared with sigmoid with relu Activation in CNN model onsameinput shape

so theresultofthisexperiment is:

Table2:Thefindingsobtainedafteranalyzingtherelative effectivenessoftheactivationfunctions

Activation Function	Accuracy	Validationloss
Sigmoid	0.98	0.04

Tanh	0.11	2.3

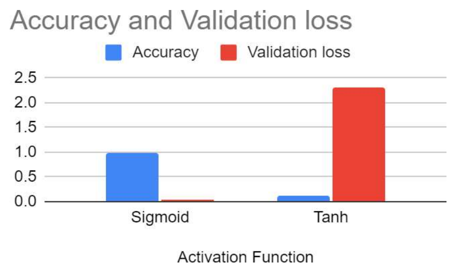

Figure8:Graphofbothactivation function

InthisresultdifferentActivationshowsthevalidationaccuracyonchangingactivation overthe5trainingepochsforthemodel.

Table3:resultingtableofvalidationaccuracyonchangingactivation overthecourseof10 training epochsforthemodel

Activation Function	Accuracy	Validationloss
Sigmoid	0.98	0.05
Tanh	0.23	2.3

as the result shows that comparare with tanh and sigmoid functionbetterresultprovidesigmoidwithminimumlosswithhighaccuracy

Here accuracy and loss checked on different epochs first checkedwith 2 epochs and after that checked on 10 epochs and its showslike minor changes are discovered here in the accuracy of tanhwhileincreasingtheepochs.

SecondlydiscussaboutFEBdatasetresult

The CNN "layer's activation functions were alternated in the firststage of the experiment while the remaining hyper-parameterswereleftat theirdefault values.

Table.4. The findings obtainedafteranalyzingtherelativeeffectivenessof the activationfunctions

ActivationFunction	**TrainAccuracy(%)**	**Validation Accuracy(%)**	**Validationloss**
Sigmoid	43.12	51.34	2.23
Tanh	33.36	42.01	2.111

Using ten boards for different activation functions, we displayedthree separate graphs for validation accuracy, validation loss, andtraining accuracy. From looking at these graphs, we can see thatthe Sigmoid activation function had the highest accuracy amongtheTanhactivationfunctions.Buttheproblemariseswithvalidationloss.

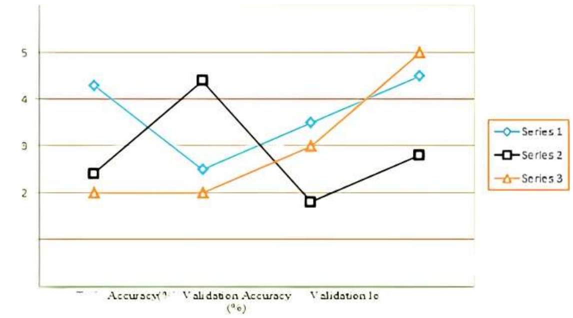

Figure 5 shows the resulting graph of validation accuracy onchangingactivationoverthecourseof30trainingepochsforthemodel.

These"varianceshaveasignificantimpactonthemodel'saccuracyandthevalueitlosesduetovariationsinactivationfunctions. If we pay attention to the Tanh and Sigmoid activationfunctions, we might be able to practically detect the difference inaccuracy"andloss value.

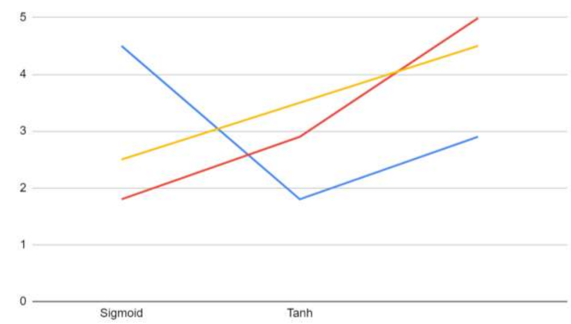

Figure 6 shows the resulting graph of validation loss as a function ofchangingactivationwhenthemodelwasbeingtrainedfor30epochs.

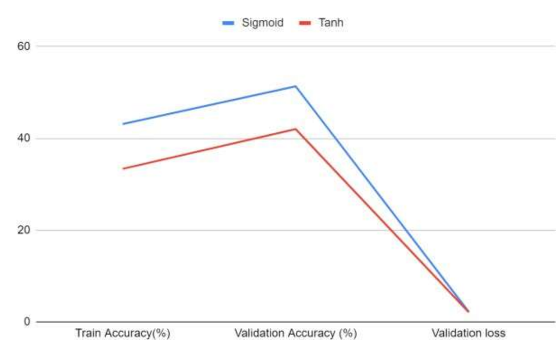

Figure7showstheresultinggraphoftrainingaccuracyonchangingactivation overthecourseof30training epochsforthemodel.

The "optimizer's hyper-parameter must be changed in the next stepwhilekeepingtheactivationlevelconstant.Threedifferentoptimizers—Adam, RMSprop, and SGD—are used in this test, andthey are each applied to one of three possible activation functions.Table 3 shows the accuracy as well as the loss value that was seenwhilekeepingtheactivationfunctionSwishandchanging"theoptimizers.

Table.3.The results that were obtained after comparing severaloptimizers while maintaining the Sigmoid activation function foratotal of30epochs.

Optimizers	TrainAccuracy(%)	Validation Accuracy(%)	Validation loss
Adam	72.64	81.90	1.112
RMSprop	72.46	81.30	1.034
SGD	38.94	43.29	2.234

In conclusion, table 4 below shows the accuracy and loss value thatwereseenwhilekeepingtheSigmoidactivationfunctionandchanging the"optimizers.

Table.4.The results obtained during the comparison of severaloptimizers while maintaining the tanh activation function for a total of30 epochs

Optimizers	TrainAccuracy(%)	Validation Accuracy(%)	Validation loss
Adam	75.43	78.51	1.104
RMSprop	70.99	73.72	1. 757
SGD	35.28	46.22	2.215

Thedatainthethreetables aboveshowthat, depending ontheactivation function they are connected to, various optimizers havevarying consequences. The findings were enhanced when RM SpropandtheSigmoidactivationfunctionwerecombined,andthevalidation loss was estimated to be 1.034. However, if Adam isfactored into the tanh computation, the loss value is considerablybetter,comingout"at 1.104.

4. CONCLUSION

Thisarticleprovidesa comprehensive overviewofthe CNNmodel, which is often used for image classification applications.Thecomplexityofthismodelleavesalotofroomforimprovement, and a number of researchers have put up a varietyofideastoboosttheaccuracyofdifferentCNNmodels.the capacity of the activation functions toenhance pattern learninginthedatajustifiestheiruseinthehiddenlayersofneuralnetworks,whereautomatedfeaturerecognitionisperformed.Theactivation functionscanalsobeusedforcategorizationinanumberofmachinelearningfields.Theresultantobservedthattheaccuracy ofSigmoid is better than the Tanh in both MNIST andFEB dataset with different epochs like 10 and 30 training epochs.The activation function is a crucial element of convolutional neuralnetworks,whichcanmapoutnon-linear"properties

REFERENCES

1. S. Albawi, et al., "Zawi, S," Understanding of a ConvolutionalNeural Network International Conference on Engineering andTechnology (ICET),vol.2017,2017.
2. R. Chauhan et al., "Convolutional neural network (CNN) forimagedetectionandrecognition"FirstInternationalConference on Secure Cyber Computing and Communication (ICSCCC),vol.2018,2018[doi:10.1109/ICSCCC.2018.8703316].
3. A. Ajit et al., "A review of convolutional neural networks"International Conference on Emerging Trends in InformationTechnologyandEngineering(ic-ETITE),vol.2020,2020[doi:10.1109/ic-ETITE47903.2020.049].
4. D. Dai, "An introduction of CNN: Models and training onneuralnetworkmodels,"InternationalConferenceonBigData, Artificial Intelligence and Risk Management (ICBAR), vol.2021,2021[doi:10.1109/ICBAR55169.2021.00037].
5. J. Feng and S. Lu, "Performance analysis of various activationfunctionsinartificialneuralnetworks,"J.Phys.Conf.Ser.,vol.1237,no.2,p.022030,2019[doi:10.1088/1742-6596/1237/2/022030].
6. A.-F. Karam et al., "Applying convolutional neural networksforimagedetection"InternationalConferenceonSmartApplications, Communications and Networking (SmartNets),vol.2019,2019[doi:10.1109/SmartNets48225.2019.9069750].
7. A. Rasamoelina et al., "A review of activation function forartificial neural network,SAMI" 18th World Symposium onAppliedMachineIntelligenceandInformatics,vol.2020.IEEE,2020.
8. Y. Yu, et al., "Rmaf: Relu-memristor-like activation functionfordeeplearning,"IEEEAccess,vol.8,pp.72727-72741,2020[doi:10.1109/ACCESS.2020.2987829]
9. S.Sharmaetal.,"Activationfunctionsinneuralnetworks," TowardsDataSci.,vol.6,no.12,pp.310-316,2017.
10. X.Lietal.,CombineReLUWithTanh,vol.1.IEEE.IEEE,2020[doi:10.1109/ITNEC48623.2020.9084659].
11. B.L.KalmanandS.C.Kwasny,"Why tanh: Choosingasigmoidal function." [Proceedings, 1992], IJCNN InternationalJointConferenceonNeuralNetworks,vol.4.IEEE,1992[doi:10.1109/IJCNN.1992.227257].
12. H. K. Kwan, "Simple sigmoid-like activation function suitablefor digital hardware implementation," Electron. Lett., vol. 28,no.15,pp.1379-1380,1992[doi:10.1049/el:19920877].
13. S. Marra et al., "Tanh-like activation function implementationfor high-performance digital neural systems," Physiol. IEEE,237-240,2006 [doi:10.1109/RME.2006.1689940].
14. B.KarlikandA.VehbiOlgac,"Performanceanalysisofvariousactivation functions in generalized MLP architectures of neuralnetworks," Int. J. Artif. Intell. Expert Syst., vol. 1, no. 4, pp.111-122,2011.
15. Ramachandran,Prajit,B.Zoph,andQuocv.Le,'Swish:aself-gated activation function.' arXiv Preprint ArXiv:1710.059417.1,vol.5,2017.
16. B. Ding et al., "Activation functions and their characteristics indeep neural networks,", 2018, Chinese Control and DecisionConference (CCDC). IEEE[doi:10.1109/CCDC.2018.8407425].
17. C.Nwankpa,etal.,'Activationfunctions:Comparisonoftrendsinpracticeandresearchfordeeplearning.'arXivPreprintArXiv:1811.03378,2018.
18. Sietal.,"AdynamicReLUonneuralnetwork"13thDallas CircuitsandSystemsConference(DCAS).IEEE.IEEE,2018.
19. M.M.LauandK.H.HannLim,"Reviewofadaptiveactivationfunction in deep neural network." IEEE-EMBS Conference onBiomedical Engineering and Sciences (IECBES). IEEE, 2018[doi:10.1109/IECBES.2018.8626714].

20. K. C. Kirana, et al., “Improved neural network using integral-RELUbasedpreventionactivationforfacedetection”International ConferenceonElectrical,ElectronicsandInformationEngineering (ICEEIE),vol.6.IEEE,2019[doi:10.1109/ICEEIE47180.2019.8981443].
21. O.Sharma,‘Anewactivationfunctionfordeepneuralnetwork.’2019InternationalConferenceonMachineLearning,Big Data, Cloud and Parallel Computing (COMITCon). IEEE,2019.
22. Y. Guo, et al., “Algorithm research on improving activationfunctionofconvolutionalneuralnetworks,”,2019,ChineseControland Decision Conference(CCDC).IEEE.
23. Z.Qiumeietal.,“Improvedconvolutionalneuralnetworkbasedon fast exponentially linear unit activation function,” IEEEAccess,vol.7,pp.151359-151367,2019 [doi:10.1109/ACCESS.2019.2948112].
24. D.Misra,‘Mish:Aselfregularizednon-monotonicactivation function.’arXivPreprintArXiv:1908.08681,2019.