Hyper-sinh-Convolutional Neural Network for Early Detection of Parkinson's Disease from Spiral Drawings

LUCA PARISI,
Department of Computer Science,
Faculty of Engineering and Informatics,
University of Bradford
Bradford, BD7 1DP, United Kingdom
University of Auckland Rehabilitative Technologies
Association (UARTA)
University of Auckland
Auckland, 1010, NEW ZEALAND
luca.parisi@ieee.org

RENFEI MA,
University of Auckland Rehabilitative Technologies
Association (UARTA)
University of Auckland
Auckland, 1010, NEW ZEALAND
renfei.ma@auckland.ac.nz

AMIR ZAERNIA,
Department of Biomedical and Electronics Engineering,
Faculty of Engineering and Informatics,
University of Bradford
Bradford, BD7 1DP, UNITED KINGDOM
a.zaernia@bradford.ac.uk

MANSOUR YOUSEFFI,
Department of Biomedical and Electronics Engineering,
Faculty of Engineering and Informatics,
University of Bradford
Bradford, BD7 1DP, UNITED KINGDOM
m.youseffi@bradford.ac.uk

Abstract—Modern developments in the state-of-the-art open-source activation functions for Convolutional Neural Networks (CNNs) have broadened the selection of benchmark activations for Deep Learning (DL)-aided classification. Nevertheless, achieving discrimination of non-linear input image data in CNN is still not straightforward and it is unclear how such novel activation functions can have translational applications with tangible impact. hyper-sinh, made freely available in TensorFlow and Keras, was demonstrated as a benchmark activation function on five (N=5) datasets in its ground-breaking paper. Measuring the value from deploying this activation in a specific application is pivotal to supply the required evidence of its performance on real-life supervised DL-based image classification tasks. In this study, a CNN was for the first time combined with hyper-sinh to aid early detection of Parkinson's Disease (PD) from discriminating pathophysiological patterns extracted from spiral drawings. Thus, the hyper-sinh activation was deployed to maximise the separability of the input features from spiral drawings via automated pattern recognition. We demonstrate the accuracy and reliability of hyper-sinh-CNN to aid early diagnosis of PD, evaluated against other gold standard activation functions, including the recent Quantum ReLU (QReLU) and the modified Quantum ReLU (m-QReLU) that solved the 'dying ReLU' problem for the first time in the literature of DL. Two (N=2) benchmark datasets from the database of the Botucatu Medical School, São Paulo State University in Brazil, scaled to be in 28 by 28 pixels as the MNIST benchmark data, were used to discriminate between input image patterns of 158 subjects (53 healthy controls and 105 patients with PD) from spirals drawn on graphics tablets. Overtraining was avoided via early stopping and the models were developed and tested in TensorFlow and Keras (Python 3.6). The supervised model (hyper-sinh-CNN) could detect early Parkinson's Disease with 81% and 91% classification accuracy from the two datasets respectively (F1-scores: 73% and 91% correspondingly). Furthermore, the model achieved high sensitivity (81% and 91%). Thus, this study validates the application of hyper-sinh to aid real-life supervised DL-based image classification, in particular early diagnosis of PD from spiral drawings.

Luca Parisi, Amir Zaernia, Renfei Ma, Mansour Youseffi

1. Introduction

Millions of patients live with Parkinson's disease (PD) worldwide, which is a long-term neurodegenerative disorder resulting from damage to neurons that synthesise dopamine. As PD develops, tremor in one hand and bodily stiffness may be experienced; nevertheless, the main four symptoms are: tremor, stiffness in the limbs, decreased movement and coordination, difficulty with fine movements to accomplish the activities of daily living, such as eating and drinking. Whilst the underlying aetiology of PD remains unknown, there is no consensus on which biomarkers could be leveraged to provide a definite diagnosis of PD. To improve function, alternate treatments, such as exercise and full body vibration using painless vibrating machines to stabilise multiple muscles and posture have been investigated with promising results [1]. Advanced computational techniques have leveraged feature-driven Machine Learning (ML) to accomplish early detection of PD from patterns used as indicators of its related speech impairments [2].

Impaired handwriting in PD patients may appear years before a clinical diagnosis is made and, thus, could be the first signs of possible PD [3]. By assessing handwriting via a digitiser, which measures the mean pressure and mean velocity, as well as the spatial and temporal characteristics of each stroke, previous research showed that 97.5% of the patients at risk of PD had significantly different writing with smaller letters ('micrographia'), less force being applied and a significantly longer writing time [4]. More recent studies have stated the impairment in writing can be analysed easier via a computerised analysis of spirals [5, 6].

The development of the open-source kernel function hyperbolic sinh or 'hyper-sinh' has provided a further benchmark kernel for Deep Learning (DL)-aided classification [7]. When coupled with the shallow 'Fully-Connected Neural Network' (FC-NN) and the deep 'Convolutional Neural Network' (CNN), its reliability and computational efficiency have been demonstrated on five benchmark datasets for aiding both image and text classification tasks [7]. However, besides its reproducibility, additional studies need to verify its replicability for other applications. In fact, CNN may be unable to achieve linear separability of noisy image data, due to activation functions incurring vanishing [8] or exploding [9] gradients leading to poor or lack of convergence.

In these scenarios, failure of the activation functions in the CNN to converge may lead to misclassifications and, thus, result in misdiagnoses when deployed in a clinical setting. Convergence during training can be aided by selecting an optimal activation function to facilitate learning from non-linear data to ensure they can be linearly separable [2, 7, 10-14]. As the traditional activation function involved in the training of CNN, i.e., the Rectified Linear Unit ('ReLU'), is affected by the 'dying ReLU' problem [15, 16], which, despite appropriate He initialisation [17], is characterised by negative summed inputs with large weight updates. Recently, the Quantum ReLU ('QReLU') and modified Quantum ReLU ('m-QReLU') were developed in TensorFlow and Keras [18] and validated against nine ReLU-based AFs on seven benchmark datasets as the first activation functions to resolve this long-standing issue in the literature of DL [16]. In fact, both the QReLU and the m-QReLU were demonstrated to aid complex medical diagnostics, such as detecting COVID-19 from ultrasound images and early PD from spiral drawings [16].

As an alternative approach, the modified arcsinh ('m-arcsinh') developed as a generalised kernel and activation function in scikit-learn [19] was validated on 15 benchmark datasets from scikit-learn and the University California Irvine (UCI) ML repository [10]. Building on the m-arcsinh, the hyperbolic sinh or 'hyper-sinh' was developed to generalise its formulation to be accurate and reliable from shallow to deep neural networks in TensorFlow and Keras [20] and validated on 5 benchmark datasets, involving both image and text classification tasks [7].

This study's novelty lies in assessing the replicability of the hyper-sinh in a clinical application, along with assessing its performance when coupled with a CNN against gold standard activation functions, including the recent QReLU and m-QReLU [16]. In this research work, such a supervised DL technique was leveraged to perform a classification task that involves noisy data with overlapping features, i.e., image data of spirals drawn on a graphics tablet to discern patterns as indicators of healthy and PD [16], to aid diagnosis of PD from the underlying image-based patterns associated with its characteristic subtle tremor at early stages.

2.   Automated detection of parkinson's disease from image data

2.1 Data and Software for Modelling

To demonstrate the accuracy and reliability of early detection of PD from image patterns from spiral drawings on graphics tablets via the proposed algorithm, the following two datasets with image features from 53 healthy subjects and 105 patients with PD were obtained from the database of the Botucatu Medical School, São Paulo State University in Brazil:

- Spiral HandPD dataset (data format: .jpg) [21] on 92 subjects (74 PD patients, 18 healthy controls).
- NewHandPD dataset (data format: .jpg) [22] on 66 additional subjects (31 PD patients, 35 healthy controls).

Both datasets were represented as one-dimensional (1D) vectors of 784 features for each image (28*28 pixels); training and testing datasets were used as per the proportions already established from the two benchmark datasets used.

The data were reshaped to match the picture format (height * width * channel) and then, to a four-dimensional (4D) tensor input (batch size, height, width, channel). The image datasets underwent min-max normalisation to prepare them for the subsequent CNN-based classification.

The CNN model was developed and evaluated via the Keras library in Python, whilst the hyper-sinh [7] activation function was developed in both TensorFlow and Keras [20]. All other activation functions evaluated (ReLU, Leaky ReLU, sigmoid and tanh) are available in Keras, except for the custom QReLU and m-QReLU [16] also developed and available in both TensorFlow and Keras [18].

2.2 hyper-sinh-based CNN Algorithm

CNN is a deep neural network-based classifier that leverages convolutional layers and filters to achieve automated pattern recognition without the need for extrinsic feature engineering and separability of the input image data based on the activation function deployed [16, 20, 23]. These input data must be labelled, since the CNN is a supervised DL-based learning classifier, thus the class which each input image belongs to is preliminarily known.

The choice of which activation function to use within convolutional layers impacts the accuracy and reliability of the classification of such input data into the target classes [16, 20, 24]. Thus, in this study, CNN was assessed with various activation functions, including the recently developed hyper-sinh [7], which scales from shallow to deep neural networks, such as the CNN.

To avoid overtraining or overfitting, thus preventing the model from 'remembering' the features seen during training and enabling learning that yields generalisation, early stopping was applied [23, 24]. The Adam optimiser was used for training, whilst the loss or probability error to minimise was the sparse softmax cross-entropy between logits and labels, as suitable in the discrete classification task of interest in which the classes (healthy and PD-related spiral drawings) are mutually exclusive (each image is either characteristic of healthy or PD-related patterns).

Given its ability to cope with non-linearly separable data, the hyper-sinh (**Fig. 1**) activation function was coupled with CNN and evaluated against other gold standard activations (ReLU, Leaky ReLU, sigmoid, tanh, QReLU and m-QReLU [16]).

The equations for the hyper-sinh activation function for positive values of input (x) data, and negative values and zero, are available in (1) and (2) respectively.

$$y = \frac{\sinh x}{3} \; if \; \mathrm{x} > 0 \,, \tag{1}$$

$$where \; x \; is \; the \; input \; data$$

$$y = \frac{x^3}{4} \; if \; \mathrm{x} \leq 0 \tag{2}$$

A flattening layer was applied to flatten the input data to a one-dimensional (1D) vector for the fully connected layer. The output layer had two neurons, as per the number of classes to discriminate (healthy vs PD-related image patterns). In **Table I**, these and other hyperparameters used in the CNN model tested in this study are listed, which

were leveraged to classify both benchmark datasets, regardless of the activation functions evaluated.

TABLE I. LEARNING HYPERPARAMETERS FOR THE CNN MODEL USED FOR BOTH BENCHMARK DATASETS AND REGARDLESS OF THE ACTIVATION FUNCTIONS DEPLOYED.

Hyperparameter	Value
Learning rate	0.001
Maximum number of epochs or training iterations	15
Batch size	128
Dropout rate prior to output layer	0.25
Number of convolutional layers	2
Number of max pooling (down-sampling) layers	2
Number of convolutional filters in first convolutional layer	32
Kernel size of first convolutional layer	5
Number of convolutional filters in second convolutional layer	64
Kernel size of second convolutional layer	3
Number of strides in max pooling layer	2
Kernel size of max pooling layer	2
Number of neurons in fully connected/dense layer	1,024
Number of neurons in final output layer	2

2.3 Performance Measures to Validate Model Accuracy and Reliability

The classification performance of the CNN was evaluated quantitatively based on the trade-off between accuracy and reliability on the test sets [7, 10-14, 16, 20] whilst applying early stopping to avoid overfitting or overtraining; the latter metric was assessed via multiple quantifiers, i.e., precision, recall and F1-score as their harmonic mean, thus representing their trade-off.

In binary classification problems, e.g., discriminating between healthy and PD-related image patterns, classes can be defined as 'true' or 'false', resulting in four scenarios (**Table II**):

- true positive (TP), the occurrence of pathophysiological image patterns correctly detected (e.g., due to PD).
- true negative (TN), physiological/healthy speech image correctly classified.
- false positive (FP), physiological/healthy image patterns misclassified as if they were pathophysiological instead.
- false negative (FN), the occurrence of a pathophysiological image patterns misclassified as if they were physiological/healthy instead.

TABLE II. CONFUSION MATRICES RELATING THE ACTUAL AND PREDICTED IMAGE PATTERNS AS TO WHETHER THEY ARE INDICATORS OF PATHOPHYSIOLOGICAL TREMOR (E.G., DUE TO EARLY PARKINSON'S DISEASE).

		Actual	
		True (pathophysiological image patterns)	False (physiological image patterns)
Predicted	Positive (pathophysiological image patterns)	TP	FP
Predicted	Negative (physiological image patterns)	FN	TN

Besides the classification accuracy, the classification performance of the CNN with the various activation functions tested were assessed via two further performance metrics as indicators of reliability, i.e., precision and recall (**Tables III** and **IV**).

As a more comprehensive quantifier of the model reliability, the F1-score was computed too (**Tables III** and **IV**), which corroborates or disproves the test classification-related outcomes.
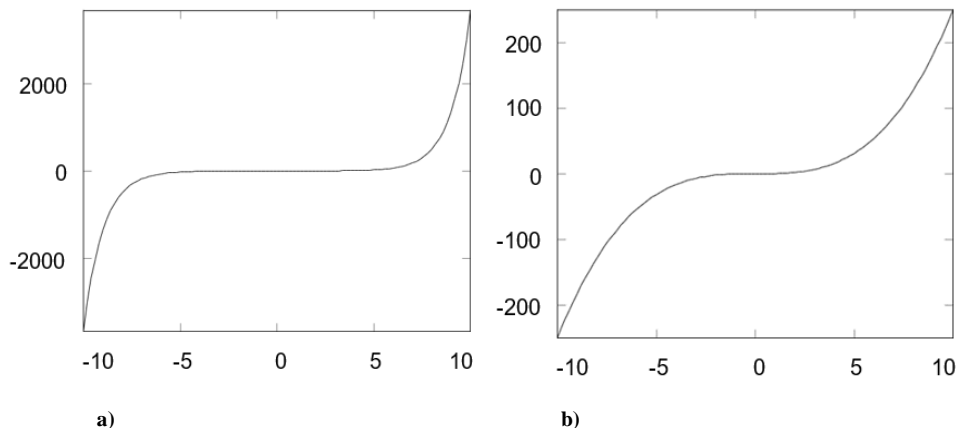
Figure 1. The hyper-sinh activation function for positive values of input data (**a**)) and for negative values or zero of input data (**b**)).

III. HYPER-SINH WITH A CONVOLUTIONAL NEURAL NETWORK FOR SUPERVISED DEEP LEARNING-BASED AUTOMATED DETECTION OF PARKINSON'S DISEASE

A quantitative analysis of classification accuracy and reliability was performed for automated early detection of PD from image patterns from spiral drawings when considering both benchmark datasets (**Tables III** and **IV**):

Both accuracy and reliability metrics have been reported in **Tables III** and **IV**, as well as their 95% confidence intervals.

TABLE III. COMPARISON OF CLASSIFICATION PERFORMANCE OF THE CONVOLUTIONAL NEURAL NETWORK WITH ALL ACTIVATION FUNCTIONS TESTED, EVALUATED ON THE SPIRAL HANDPD BENCHMARK DATASET [21].

Activation function	Accuracy (0-1)	Precision (0-1)	Recall (0-1)	F1-score (0-1)
ReLU	0.81	1.00	0.81	0.90
	(0.76-0.87)	(0.98-1.00)	(0.78-0.83)	(0.85-0.93)
Leaky ReLU	0.82	0.83	0.82	0.83
	(0.79-0.85)	(0.81-0.85)	(0.79-0.84)	(0.80-0.84)
Sigmoid	0.85	0.88	0.85	0.86
	(0.80-0.88)	(0.85-0.90)	(0.82-0.87)	(0.83-0.88)
Tanh	0.82	0.92	0.82	0.86
	(0.81-0.85)	(0.90-0.94)	(0.79-0.84)	(0.82-0.88)
QReLU	0.88	0.95	0.88	0.90
	(0.87-0.90)	(0.93-0.97)	(0.87-0.90)	(0.88-0.92)
m-QReLU	0.92	0.95	0.92	0.93
	(0.90-0.95)	(0.94-0.96)	(0.90-0.94)	(0.91-0.95)
hyper-sinh	0.81	0.66	0.81	0.73
	(0.78-0.83)	(0.65-0.69)	(0.79-0.84)	(0.72-0.75)

QReLU: Quantum ReLU; m-QReLU: modified Quantum ReLU; hyper-sinh: hyperbolic sinh.

TABLE IV. COMPARISON OF CLASSIFICATION PERFORMANCE OF THE CONVOLUTIONAL NEURAL NETWORK WITH ALL ACTIVATION FUNCTIONS TESTED, EVALUATED ON THE NEWHANDPD BENCHMARK DATASET [22].

Activation function	Accuracy (0-1)	Precision (0-1)	Recall (0-1)	F1-score (0-1)
ReLU	0.75	0.76	0.75	0.76
	(0.73-0.76)	(0.74-0.78)	(0.73-0.77)	(0.74-0.77)

Leaky ReLU	0.83	0.83	0.83	0.83
	(0.82-0.84)	(0.81-0.84)	(0.81-0.84)	(0.81-0.84)
Sigmoid	0.57	1.00	0.57	0.72
	(0.53-0.59)	(0.73-1.00)	(0.54-0.59)	(0.69-0.75)
Tanh	0.94	0.94	0.94	0.94
	(0.93-0.95)	(0.92-0.95)	(0.92-0.95)	(0.92-0.95)
QReLU	0.83	0.83	0.83	0.83
	(0.82-0.85)	(0.81-0.86)	(0.82-0.84)	(0.82-0.85)
m-QReLU	0.79	0.79	0.79	0.79
	(0.78-0.81)	(0.77-0.82)	(0.78-0.80)	(0.78-0.81)
hyper-sinh	0.91	0.91	0.91	0.91
	(0.89-0.94)	(0.88-0.94)	(0.89-0.94)	(0.89-0.94)

QReLU: Quantum ReLU; m-QReLU: modified Quantum ReLU; hyper-sinh: hyperbolic sinh.

3. Discussion and Conclusion

The supervised DL-based algorithm alone (CNN) could detect early PD with 81% and 91% classification accuracy, and 73% and 91% F1-score (**Tables III** and **IV**) via the recent hyper-sinh activation function on the PD-related Spiral HandPD [21] and the NewHandPD [22] image datasets, respectively. Thus, the hyper-sinh [7], when coupled with a CNN, led to the highest classification performance (**Tables III** and **IV**). The QReLU and m-QReLU also performed better than the other activation functions when classifying both datasets (88% and 83% classification accuracy, and 90% and 83% F1-score, **Tables III** and **IV**), except for the tanh function for the second dataset. The sigmoid activation functions led to the worst classification performance on the NewHandPD dataset (57% classification accuracy, **Table IV**), despite a good reliability (72% F1-score, **Table IV**).

This study validates the use of hyper-sinh when coupled with a CNN for supervised DL-based classification to aid early detection of PD from image features of spiral drawings from graphics tablets, which is an inexpensive and non-invasive tool to facilitate its early diagnosis.

Future work is underway to understand if hyperparameter optimisation can level out differences in overall classification performance of the CNN when leveraging the activation functions 'hyper-sinh', 'QReLU' and 'm-QReLU'. Additional training data will be required minimise the overlapping regions of the transformed data associated with physiological and pathophysiological tremor more reliably.

Acknowledgment

References

[1] Chakravarthy, K. V., Chaturvedi, R., Agari, T., Iwamuro, H., Reddy, R., & Matsui, A. (2020). Single arm prospective multicenter case series on the use of burst stimulation to improve pain and motor symptoms in Parkinson's disease. *Bioelectronic medicine*, *6*(1), 1-8.

[2] Parisi, L., RaviChandran, N., & Manaog, M. L. (2018). Feature-driven machine learning to improve early diagnosis of Parkinson's disease. *Expert Systems with Applications*, *110*, 182-190.

[3] Carmeli, E., Patish, H., & Coleman, R. (2003). The aging hand. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, *58*(2), M146-M152.

[4] Rosenblum, S., Samuel, M., Zlotnik, S., Erikh, I., & Schlesinger, I. (2013). Handwriting as an objective tool for Parkinson's disease diagnosis. *Journal of neurology*, *260*(9), 2357-2361.

[5] Drotár, P., Mekyska, J., Rektorová, I., Masarová, L., Smékal, Z., & Faundez-Zanuy, M. (2016). Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease. *Artificial intelligence in Medicine*, *67*, 39-46.

[6] San Luciano, M., Wang, C., Ortega, R. A., Yu, Q., Boschung, S., Soto-Valencia, J., ... & Saunders-Pullman, R. (2016). Digitized spiral drawing: a possible biomarker for early Parkinson's disease. *PloS one*, *11*(10), e0162799.

[7] Parisi, L., Ma, R., RaviChandran, N., & Lanzillotta, M. (2020). hyper-sinh: An Accurate and Reliable Function from Shallow to Deep Learning in TensorFlow and Keras. *arXiv preprint arXiv:2011.07661*.

[8] Clevert, D. A., Unterthiner, T., & Hochreiter, S. (2015). Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289*.

[9] Pascanu, R., Mikolov, T., & Bengio, Y. (2012). Understanding the exploding gradient problem. *CoRR, abs/1211.5063*, *2*(417), 1.

[10] Parisi, L. (2020). m-arcsinh: An Efficient and Reliable Function for SVM and MLP in scikit-learn. *arXiv preprint arXiv:2009.07530*.

[11] Parisi, L., RaviChandran, N., & Manaog, M. L. (2020). A novel hybrid algorithm for aiding prediction of prognosis in patients with hepatitis. *Neural Computing and Applications*, *32*(8), 3839-3852.

[12] Parisi, L., & RaviChandran, N. (2020). Evolutionary Denoising-Based Machine Learning for Detecting Knee Disorders. *Neural Processing Letters*, *52*(3), 2565-2581.

[13] Parisi, L., & RaviChandran, N. (2020). Evolutionary feature transformation to improve prognostic prediction of hepatitis. *Knowledge-Based Systems*, *200*, 106012.

[14] Parisi, L., RaviChandran, N., & Manaog, M. L. (2018). Decision support system to improve postoperative discharge: A novel multi-class classification approach. *Knowledge-Based Systems*, *152*, 1-10.

[15] Dahl, G. E., Sainath, T. N., & Hinton, G. E. (2013, May). Improving deep neural networks for LVCSR using rectified linear units and dropout. In *2013 IEEE international conference on acoustics, speech, and signal processing* (pp. 8609-8613). IEEE.

[16] Parisi, L., Neagu, D., Ma, R., & Campean, F. (2020). QReLU and m-QReLU: Two novel quantum activation functions to aid medical diagnostics. *arXiv preprint arXiv:2010.08031*.

[17] Glorot, X., Bordes, A., & Bengio, Y. (2011, June). Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics* (pp. 315-323). JMLR Workshop and Conference Proceedings.

[18] Parisi, L. (2020). QReLU and m-QReLU in TensorFlow and Keras. *GitHub repository*. https://github.com/luca-parisi/QReLU_m-QReLU_TensorFlow_Keras.

[19] Parisi, L. (2020). m-arcsinh in scikit-learn. *GitHub repository*. https://github.com/luca-parisi/m-arcsinh_scikit-learn.

[20] Parisi, L. (2020). hyper-sinh in TensorFlow and Keras. *GitHub repository*. https://github.com/luca-parisi/hyper-sinh_TensorFlow_Keras.

[21] Pereira, C. R., Pereira, D. R., Silva, F. A., Masieiro, J. P., Weber, S. A., Hook, C., & Papa, J. P. (2016). A new computer vision-based approach to aid the diagnosis of Parkinson's disease. Computer Methods and Programs in Biomedicine, 136, 79-88.

[22] Pereira, C. R., Weber, S. A., Hook, C., Rosa, G. H., & Papa, J. P. (2016). Deep learning-aided Parkinson's disease diagnosis from handwritten dynamics. In 2016 29th SIBGRAPI Conference on Graphics, Patterns, and Images (SIBGRAPI), 340-346. IEEE.

[23] LeCun, Y., & Bengio, Y. (1995). Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, *3361*(10), 1995.

[24] Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., & Chen, T. (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, *77*, 354-377.