Multimodal Convolutional Deep Belief Networks for Stroke Classification with Fourier Transform

Mateus Roder, Nicolas Gomes, Arissa Yoshida, João Paulo Papa
Computing Department
São Paulo State University (UNESP)
Bauru, Brasil
{mateus.roder, nicolas.gomes, arissa.yoshida, joao.papa}@unesp.br

Fumie Costen
School of Electrical and Electronic Engineering
The University of Manchester
Manchester, United Kingdom
fumie.costen@manchester.ac.uk

***Abstract*—Several studies have investigated the vast potential of deep learning techniques in addressing a wide range of applications, from recommendation systems and service-based analysis to medical diagnosis. However, even with the remarkable results achieved in some computer vision tasks, there is still a vast scope for exploration. Over the past decade, various studies focused on developing automated medical systems to support diagnosis. Nevertheless, detecting cerebrovascular accidents remains a challenging task. In this regard, one way to improve these approaches is to incorporate information fusion techniques in deep learning architectures. This paper proposes a novel approach to enhance stroke classification by combining multimodal data from Fourier transform with Convolutional Deep Belief Networks. As the main result, the proposed approach achieved state-of-the-art results with an accuracy of 99.94%, demonstrating its effectiveness and potential for future applications.**

I. Introduction

Nowadays, most artificial intelligence techniques are based on deep learning (DL) [1]–[3], which attempts to mimic the human brain and its visual processing to learn hierarchical features. Given the advances, DL has recently been highlighted in several real-world scenarios, mainly due to its high efficiency in solving particular tasks, such as image classification [4].

Concerning medical applications, it is consensus that cerebrovascular accidents, aka stroke, and classification (complete diagnosis) represent an important problem to be supported by artificial intelligence. This lesion damages the cerebral tissue and it may permanently impair or reduce brain functions, also may leading to patient death [5]. Additionally, stroke may leave long-life sequels and curtail the life's quality of the affected patients, costing billions of dollars per year worldwide to treat those people [5].

Pereira et al. [6] dealt with stroke classification proposing standard Convolutional Neural Networks (CNNs) to classify brain-computed tomography (CT) images. In a similar manner, Bacchi et al. [7] developed a promising approach to detect ischemic stroke thrombolysis. Moreover, Roder et al. [8] employed Restricted Boltzmann Machines (RBMs) with Fourier transform to classify images of patients with cerebrovascular accidents. The authors obtained high accuracy rates regarding the problem using CT-based images. On the other hand, medical images are hard to obtain and harder to label, which leads to small databases. The authors in [8] also stated that the database used in their work (approximately 300 images) was relatively small, and the employed techniques should be improved with more data.

The main objective of this paper is to detect and classify stroke using a dataset of CT images provided by Pereira et al. [6], with a new information fusion methodology for Convolutional Deep Belief Networks (CDBNs) [9] able to overpass the previous state-of-the-art results on such a database [8]. The proposed approach employs the frequency information representation from Fourier transform [10] and fusion it with the spatial domain (original CT images) in new multimodal CDBN architecture. As far as the authors are concerned, no single combination of CDBNs with the Fourier transform or their application on cerebrovascular accident tasks is reported to date. At last, the present work has three main contributions: (i) it introduces a new framework for classification and image processing with Fourier transform and convolutions in Deep Belief Networks; (ii) it proposes the coupling of Fourier transform and Convolutional Deep Belief Networks; and (iii) it provides an effective application for the medical area.

The remainder of this paper is structured as follows: Section II presents the theoretical background and related works concerning the researched topics. Section III introduces the proposed approach, while Section IV presents a brief explanation regarding the employed dataset and the experimental methodology. Finally, Section V presents the experimental results and Section VI states conclusions.

II. Theoretical Background and Related Works

In this section, we present the main concepts of CDBNs, Fourier Transform, and cerebrovascular accident. Additionally, related works to the previous topics are presented.

A. Convolutional Deep Belief Networks

A Deep Belief Network [2] is a generative model based on stacking two or more Restricted Boltzmann Machines. RBMs are energy-based stochastic neural nets based on physical principles, composed of two layers of neurons (visible and hidden), capable of modeling the input data probability in its hidden layer. The DBNs/RBMs can be employed in unsupervised and supervised paradigms, since they can be trained under both

situations [11], [12], or mixed with unsupervised pre-training and supervised fine-tuning [2], [13].

Initially, RBMs and DBNs were developed using visible and hidden layers with binary states sampled from a Bernoulli distribution. Both layers are connected via a real-valued matrix corresponding to the neural connections. Later, Welling et al. [14] and Hinton [15] showed variations for the neurons in an RBM, such as binomials, rectified linear (ReLU), and Gaussians. Additionally, Lee et al. [9] proposed the Convolutional-based RBM (CRBM) and DBN (CDBN), representing an outstanding breakthrough in energy-based models.

To model the CDBN, one must formulate the CRBM first. Let $\mathbf{D} \in \Re^{C \times N \times N}$ be an input image with $N \times N$ pixels and C channels, and $\mathbf{v} \in \Re^{C \times N \times N}$ and $\mathbf{h} \in \Re^{K \times M \times M}$ the visible and hidden layers, respectively, where K stands for the number of $P \times P$ convolutional filters. The visible and hidden layers are connected by the matrix $\mathbf{w} \in \Re^{K \times C \times P \times P}$, with biases $\mathbf{b} \in \Re^{C}$ and $\mathbf{a} \in \Re^{K}$ associated to the visible and hidden layers, respectively. The energy of a CRBM is described as follows:

$$E(\mathbf{v}, \mathbf{h}) = -\sum_{k=1}^{K} \mathbf{h}_k \otimes \left(\sum_{c=1}^{C} (\mathbf{w}_{kc} * \mathbf{v}_c) \right) - \sum_{k=1}^{K} a_k \left(\sum_{i,j}^{M^2} h_{kij} \right) - \sum_{c=1}^{C} b_c \left(\sum_{i,j}^{N^2} v_{ij} \right), \quad (1)$$

where $*$ and $\otimes$ stand for the convolutional operator and the dot product, respectively, $\mathbf{w}_{kc}$ denotes the filter k concerning channel c, and $\mathbf{v}_c$ corresponds to the image channel c.

Since CRBM variants are bipartite graphs and the probability of a joint configuration depends on the given energy, one can calculate the visible and hidden probabilities as in a standard RBM. Equations 2 and 3 describe the conditional probabilities for a neuron in the hidden and visible layers of a CRBM, respectively:

$$P(\mathbf{h}_k = 1 | \mathbf{v}) = \sigma \left(\sum_{c=1}^{C} (\mathbf{w}_{kc} * \mathbf{v}_c) + a_k \right) \quad (2)$$

and

$$P(\mathbf{v}_c | \mathbf{h}) = \sigma \left(\sum_{k=1}^{K} (\mathbf{w}_{kc}^T * \mathbf{h}_k) + b_c) \right), \quad (3)$$

where $\sigma(\cdot)$ denotes the logistic function. Equation 2 computes the probability of all hidden neurons from the convolutional filter k, and Equation 3 uses the transposed weight matrix to map back the hidden activations to the visible layer. In short, training a CRBM aims to maximize the log-likelihood under the convolutional parameters ($\mathbf{w}$, $\mathbf{b}$, and $\mathbf{a}$). The Contrastive Divergence [11] is applied to approximate the model gradients and update its parameters according to the ascending stochastic gradient, as shown by Lee et al. [9]. As such, it becomes straightforward to change the activation of visible neurons to a Gaussian distribution or even apply the ReLU activation function.

Since the CRBM was defined, it is easy to extrapolate such concepts to deeper versions, such as CDBN. This model represents a stack of two or more CRBMs greedily trained, i.e., one block of CRBM each time, which composes the final CDBN model with L hidden layers (each hidden layer is approximated by a CRBM). To connect each CRBM block, we just forward the feature map activation from the convolutional kernels from the $l-1^{th}$ to the l^{th} hidden layer, and train such CRBM with contrastive divergence. Also, one can apply probabilistic max-pooling [9] or the standard max-pooling [16].

B. Fourier Transform

The Fourier transform indicates that any periodic function can be expressed as a sum of sines and cosines of different frequencies, weighted by its coefficient. If an arbitrary function is not periodic, one can still decompose it in sines and cosines, but from the integral under the curves. Moreover, with the computing advent and technological advances, the fast Fourier transform (FFT) was developed [17], democratizing its use in continuous or discrete domains. With the FFT, one can employ several manipulations on digital signals (as images, for instance), by applying filters in the frequency domain. The Fourier transform in a D-dimensional array produces a complex array with the same dimension in addition to the magnitude (or module) and phase angle. The discrete Fourier transform (DFT) is usually defined for two dimensions, as follows [17]:

$$\begin{aligned} F(x, y) &= R(x, y) + jI(x, y) \\ &= |F(x, y)| e^{j\phi(x,y)}, \end{aligned} \quad (4)$$

where R and I represent the real and imaginary components, respectively, u and v stand for the two-dimensional matrix coordinates, and ϕ is the phase angle. The magnitude and phase angle of the input signal is defined by Equations 5 and 6, respectively:

$$|F(x, y)| = [R^2(x, y) + I^2(x, y)]^{1/2} \quad (5)$$

and

$$\phi(x, y) = \arctan \left[\frac{I(x, y)}{R(x, y)} \right]. \quad (6)$$

C. Cerebrovascular Accident

A cerebrovascular accident, commonly known as a stroke, is an injury that abruptly strikes the brain tissue. Its main causes rely on the blood supply abnormalities to a certain brain region, resulting in the loss or reduction of its functions. The stroke can be classified into ischemic, characterized by the blockage of a vessel responsible for brain irrigation, and hemorrhagic, represented by the rupture of a blood vessel in or around the brain. A stroke can occur early due to several risk factors (non-modifiable and modifiable factors), such as age, heart disease, physical inactivity, tobacco use, and alcohol consumption, to cite a few. To reach the final diagnosis,

professionals need a detailed medical history, physical and neurological exams, and brain imaging tests [18].

Virani et al. [5] pointed out that millions of people worldwide are affected by strokes every year, with a high death rate. Nevertheless, other expressive stroke-related problem stands for the lifelong post-stroke sequelae, varying in many ways and degrees [19]. Nonetheless, several studies employing artificial intelligence to support and assist the medical diagnosis have been doing, as showed by Pereira et al. [6] and Roder et al. [8], focusing on detecting brain abnormalities by images from computed tomography or magnetic resonance imaging.

D. Related Works

Fourier transforms have a wide range of applications, ranging from classical image processing methods to deep learning-based ones. Lotfollahi et al. [20], for instance, employed the Fourier transform on digital image coloring of high-definition infrared images through deep learning. The authors explored a domain expressively difficult to acquire categorized data, the histological stains. They showed that is possible to employ unsupervised deep learning coupled with a fast Fourier transform to map images in the infrared spectrum to high-resolution images of histological stains.

Chen et al. [21] introduced an end-to-end deep neural network model, named Fourier Image Network (FIN). The framework does not require preprocessing steps and takes two or more raw holographic images of different types, succeeding in external generalization. FIN uses global spatial frequency information processed by trained modules of spatial Fourier transform (SPAF), which uses DFT in one of its modules, similar to what happens in [22]. The proposed approach's effectiveness was experimentally demonstrated by training the model with image samples of human lung tissue and blindly testing with images of other human body tissues, such as prostate, salivary gland, and cervical smear samples. According to the author [21], the speed of FIN is 27 times faster than iterative wave propagation-based algorithms for holographic image reconstruction. Furthermore, compared to existing CNN-based models, FIN exhibits great generalization performance and is much faster in its inference speed.

Regarding deep learning applications to support medical diagnosis, Pereira et al. [6] employed convolutional neural networks to detect and classify images of patients with/without stroke. The authors achieved promising results and relatively high accuracy rates, ranging from 77% to 97%, regarding stroke classification using CT-based images. Nevertheless, it is well-known that medical images are hard to obtain, and harder to label, which may lead to small databases, as pointed out in their work.

Bacchi et al. [7] developed a promising approach to detect ischemic stroke thrombolysis. The authors applied two Neural Networks to process images and text, where the first was a CNN fed with computed tomography images, while the second was a 3-layer perceptron network to deal with structural data from medical guides (age, gender, etc.). The results were promising on their private dataset and showed an exciting line for the development of robust tools for medicine.

Finally, Roder et al. [8] addressed the stroke classification problem employing RBMs and Fourier transform. They used the same database from [6], i.e., a public small-sized stroke database to classify ischemic, hemorrhagic, and healthy brains from CT images. The authors proposed the information fusion with the original images and their transformed versions on the frequency domain from the discrete Fourier transformation, obtaining different RBM architectures as the input data varies. The authors achieved more than 98% of accuracy in almost all experiments, outstanding results even without convolutions.

III. Fourier-based Multimodal Convolutional Deep Belief Networks

It is well-known that information can be represented in different domains, such as space or frequency, through the Fourier transforms, for instance. Given that, this work proposes to enhance the power of Convolutional Deep Belief Networks by applying different data modalities as input, in a multimodality fashion, employing the Fourier transform to improve the performance of the stroke classification problem. Such networks have been chosen due to their simplicity and lack of work involving the researched topics.

The information fusion in multimodality aggregates data from different domains. Such an approach can employ as many types of data as available, such as bi and tri-modal, for instance. Regarding the proposed approach, the multimodal data is composed of original computed tomography images and their respective Fourier transforms, where the latter is composed of two different components: magnitude (Equation 5) and phase angle (Equation 6).

Let FCDBN (Multimodal Fourier-based CDBN) denotes the proposed approach, where its variations FCDBN-P, FCDBN-M, and FCDBN-PM, denote which components are used to compose the multimodality, i.e., P stands for the phase, M stands for the magnitude, and PM for both.

These architectures are trained under the standard unsupervised learning approach [2], [9] in which the second hidden layer receives the kernels activations from all Convolutional RBMs that process the input data, highlighting that all hidden layers are max-pooled after trained. This approach is depicted in Figure 1, in which $\mathbf{h^1}$ and $\mathbf{h^2}$ stand for the convolutional kernels from the two input models (first hidden layer). After that, they are fine-tuned for the classification task through an additional Softmax layer, which receives the last hidden layer activations from the convolutional kernels, $\mathbf{h^3}$ (or $\mathbf{h^L}$ for more hidden layers), and outputs the three class probabilities. Figure 1 illustrates an FCDBN-P with two hidden layers and two entries, where the first Convolutional RBM (ConvRBM) receives the original CT image and the second (Fourier-ConvRBM) receives the phase angle from the DFT applied to the image.

In this work, the Multimodal CDBNs address the problem of employing Gaussian visible units for convolutions, i.e., image pixels as continuous values from a Gaussian distribution from

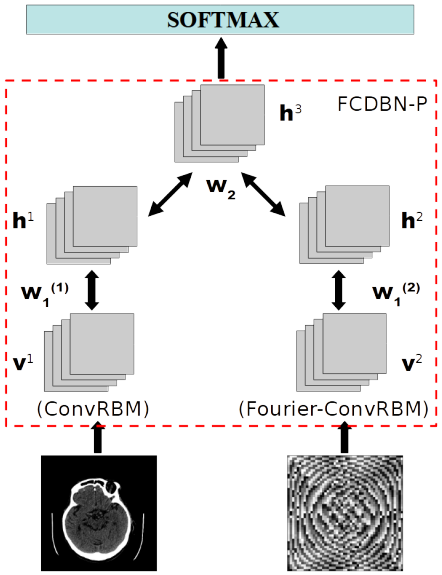

Fig. 1. FCDBN-P architecture with two types of inputs, i.e., original images and the phase angle of the DFT.

the original grayscale images. For the proposed approach, Gaussian-based CRBMs are employed to process the DFT data, i.e., if a given architecture uses the phase angle (FCDBN-P), the two data modalities are normalized by a standard Gaussian and fed to two CRBMs. The same process occurs for FCDBNs with magnitude (FCDBN-M), where the magnitude of the frequency spectrum is normalized by a standard Gaussian and fed to another CRBM. Finally, when both DFT components are used (FCDBN-PM), phase and magnitude are normalized and fed into two Gaussian simultaneously to a base CRBM (with CT images).

In summary, by applying the multimodal approach, the models aggregate information from the spatial component originating from the original images and the frequency originated from the components of the Fourier transform. Additionally, the employed multimodality enables the analysis of which configuration is more suitable for the problem in question and gives greater flexibility to the application.

IV. Experimental Methodology

In this section, we present an overall description of the dataset and the experimental setup regarding the proposed approach is explained, with hyperparameters details.

A. Dataset

The dataset employed in this study was introduced by Pereira et al. [6], which consists of 25 brain CT scans, featuring patients [1]with (a) healthy brain, (b) ischemic stroke, and (c) hemorrhagic stroke. The total number of images accounts for 300, where 100 depicts healthy brains and 200 a cerebrovascular accident (100 hemorrhagic and 100 ischemic). In addition, the images exhibit irregular patterns, varying lighting conditions, and diverse structural characteristics.

[1]Patients' identification have been omitted for ethical reasons.

Each original image in the dataset is in grayscale and has a resolution of 512×512 pixels. However, to reduce the number of computational resources needed and to speed up the CDBNs training, they have been resized to 150×150. The images have also been normalized according to a standard Gaussian distribution, i.e., zero mean and unit variance.

B. Experimental Setup

The experimental setup comprises four different CDBN architectures, yet with the same training process, i.e., networks trained with the Contrastive Divergence, and with an equal number of epochs for each ConvRBM that composed the deep model. The baseline architectures are the Neural Networks proposed by Pereira et al. [6], CNN_1, and CNN_2, and the ones proposed by Roder et al. [8], i.e., GaussianRBM, MultFRBM-P, MultFRBM-M, and MultFRBM-PM. The architectures proposed in this paper stand for FCDBN-P, FCDBN-M, and FCDBN-PM, with the CDBN being the base model. The hyperparameter configuration is presented in Table I.

TABLE I
Hyperparameter configuration for the Fourier-based CDBN models.

Architecture	# Kernel Shape	# Kernels	Learning Rate	Momentum
CDBN	$(5, 5)$	32	10^{-6}; 10^{-7}	0.5
FCDBN-P	$(5, 5)$	$32, 32$	10^{-5}; 10^{-6}; 10^{-6}	0.5
FCDBN-M	$(5, 5)$	$32, 32$	10^{-5}; 10^{-6}; 10^{-6}	0.5
FCDBN-PM	$(5, 5)$	$32, 32, 32$	10^{-6}; 10^{-6}; 10^{-6}; 10^{-7} [2]	0.5

From Table I, it is important to highlight that all hyperparameters were empirically chosen in exploratory experiments, leading to the best configurations in terms of convergence and stability. The kernel shape for convolutions is equal to all layers and input data to simplify this analysis, and the same rationale is employed on the kernel number, keeping the same as the anchor model (CDBN) for each additional RBM added. Also, the learning rate stands for the values that provided stability in the learning procedure. Moreover, all models have a momentum of 0.5 on stochastic gradient descent for all layers. Finally, only one hidden layer for all CDBN models was employed to show how promising the proposed approach is.

Regarding the training and evaluation methodology for the employed dataset, we opted to follow the works of Pereira et al. [6] and Roder et al. [8], which provided two distinct data configurations: a half-and-half split between training and testing sets $(50/50)$ and a quarter-third of training and testing splits $(75/25)$. We opted to employ an additional partition, to analyze the robustness of the proposed methodology, being 25% for training and 75% for testing. Moreover, the experiments have been evaluated throughout 15 independent executions to reduce the models' stochasticity, to remove any virtual split-based data bias, and to provide enough executions to conduct a statistical test, i.e., Wilcoxon signed-rank test [23] with 5% of significance.

[2]This model required lower learning rates to convergence.

Concerning the training pipeline, all models were first trained in an unsupervised manner (pre-training phase), as CDBNs were proposed [9], for 50 epochs, with a mini-batch of 50 samples. After the pre-training, we added a Softmax layer on top of all CDBNs (multimodal or not) to compose the classification models, in which the Adam [24] optimizer was employed with a learning rate of 10^{-4} to the Softmax layer, 50 fine-tuning epochs, and 60 samples per mini-batch. Since the dataset has equal class distribution, we opted to keep the standard accuracy, as Pereira et al. [25] and Roder et al. [8], as performance measurement.

V. Experiments

Table II presents the mean results and their respective standard deviation obtained from the 15 independent executions, considering the three data splits, i.e., $25/75$, $50/50$, and $75/25$. Additionally, the Wilcoxon signed-rank test with a p-value of 5% was applied to the proposed models following the rationale: select the best mean accuracy in a split and test against all other Fourier-based models to test (one-versus-all), enabling tagging with $(=)$ the statistically similar results, and with $(!=)$ the different ones. Moreover, the highest mean achieved by a given model in each split is underlined.

TABLE II
Mean accuracy and its respective standard deviations for the test set considering the data splits.

Architecture	25/75	50/50	75/25
CDBN	97.86 ± 2.41 $(=)$	99.41 ± 0.50 $(!=)$	99.83 ± 0.33 $(=)$
FCDBN-P	$\underline{97.99 \pm 2.19}$	$\underline{99.71 \pm 0.60}$	$\underline{99.94 \pm 0.21}$
FCDBN-M	97.88 ± 2.38 $(=)$	99.48 ± 0.65 $(!=)$	99.83 ± 0.34 $(=)$
FCDBN-PM	97.41 ± 2.21 $(!=)$	99.57 ± 0.79 $(=)$	99.72 ± 0.50 $(=)$
GaussianRBM [8]	-	99.66 ± 0.52 $(=)$	99.66 ± 0.52 $(!=)$
MultFRBM-P [8]	-	98.76 ± 1.58 $(=)$	99.72 ± 0.40 $(=)$
MultFRBM-M [8]	-	98.53 ± 1.70 $(=)$	99.49 ± 0.60 $(!=)$
MultFRBM-PM [8]	-	97.94 ± 1.66 $(!=)$	99.32 ± 0.71 $(!=)$
CNN_1 [6]	-	93.46 ± 16.54	97.20 ± 2.45
CNN_2 [6]	-	83.55 ± 13.09	77.33 ± 22.24

From Table II and the first data split $(25/75)$, one can observe that all CDBN-based models achieved impressive mean accuracy, around 97%, better than expected since the number of training samples was pretty low. However, the standard deviation was high, showing that the proposed models may require more data to become stable. Also, the better model so far was FCDBN-P, have been statistically different from FCDBN-PM, a more complex version.

Regarding the $50/50$ split, the best model so far was the FCDBN-P, achieving a mean accuracy of 99.71% surpassing the multimodal models from Roder et al. [8] and CNNs of Pereira et al. [6]. Additionally, such a model was statistically different from CDBN, and FCDBN-M. Still glancing at the proposed approaches, they all surpass the multimodal RBMs from [8] in more than 1 point on mean accuracy and low standard deviation, pointing out that our approach is robust even with low data volume. However, our best result was statistically similar to three of four models from Roder et al. [8].

Finally, concerning the third split $(75/25)$, there has been a significant improvement in the multimodal CDBN models, where all approaches obtained a mean accuracy higher than 99.72%. Over all models, the FCDBN-P stood out, achieving 99.94% and surpassing both unimodal (CDBN) and baseline models. Additionally, it had the lowest standard deviation (0.21) compared to all models. Such results corroborate the positive effects regarding the proposed approach, in which all models overpass the non-convolutional and shallow [8] versions, in which the FCDBN-P was statistically better than three of four models from Roder et al. [8]. Moreover, the impressive performance improvement given more data indicates that such a methodology may be even more efficient for applications with more data available to train.

One important observation regarding the proposed approach is that, looking at the computer vision state-of-the-art, the employed models for different domains stands for huge models like ResNet [26] and Vision Transformers [27], for instance, which are well-proved to be data-hungry. On the other hand, FCDBN- models achieved outstanding results with low volume of data compared to the CNNs from Pereira et al. [6], which evince the proposed models relevance. Additionally, Figures 2, 3 and 4 illustrate the fine-tuning learning procedure.

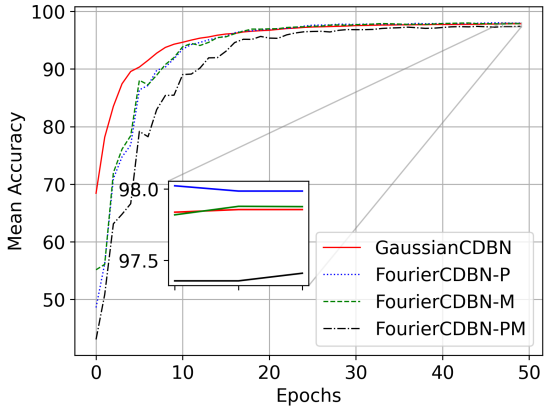

Fig. 2. Mean accuracy over the test set considering the $25/75$ split.

VI. Conclusion

This work explores the use of multimodal data in the context of Convolutional Deep Belief Networks applied to the stroke classification task. The proposed multimodality employs the Fourier transform to provide two additional data distributions, i.e., the magnitude and the phase angle.

We argue that the proposed approaches achieved state-of-the-art results and surpass the previously approaches regarding the experimental tests. Every proposed model has been superior to the baselines employed by Roder et al. [8] and Pereira et al. [6], which led us to the assumption that our models may perform in a stable, and accurate, manner even with a low volume of data available. Another crucial point is the

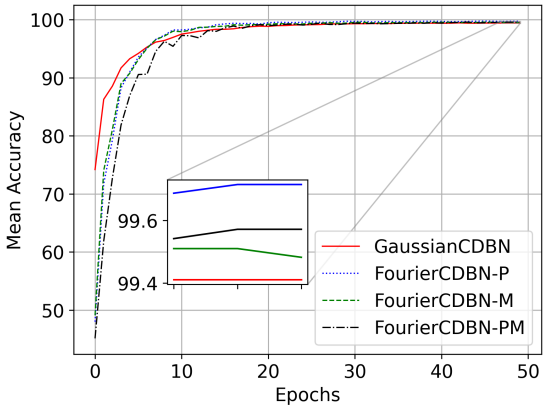

Fig. 3. Mean accuracy over the test set considering the $50/50$ split.

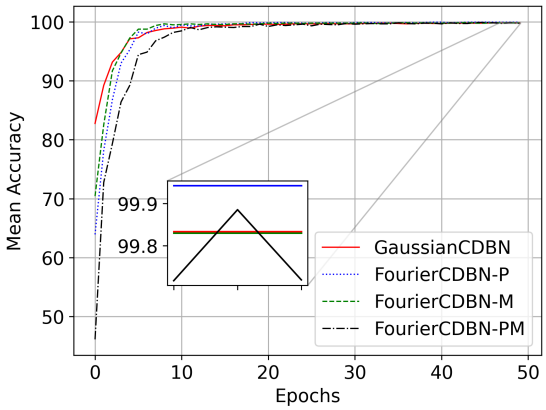

Fig. 4. Mean accuracy over the test set considering the $75/25$ split.

potential of multimodal CDBNs, especially when combined with the power of the Fourier transform. According to Table II, it is clear that the phase angle contributed more to model discrimination than the magnitude, indicating that such an approach is viable and should be extended to other datasets.

Concerning future works, we aim to deeply study the proposed approach on other tasks in the medical domain. Also, the authors intend to evaluate the idea of making a large pre-trained model with this approach to open-use.

References

[1] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[2] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[3] R. Salakhutdinov and G. E. Hinton, "An efficient learning procedure for deep boltzmann machines," *Neural Computation*, vol. 24, no. 8, pp. 1967–2006, 2012.

[4] C. Bhatt, I. Kumar, V. Vijayakumar, K. U. Singh, and A. Kumar, "The state of the art of deep learning models in medical science and their challenges," *Multimedia Systems*, pp. 1–15, 2020.

[5] S. S. Virani, A. Alonso, E. J. Benjamin, M. S. Bittencourt, C. W. Callaway, A. P. Carson, A. M. Chamberlain, A. R. Chang, S. Cheng, F. N. Delling *et al.*, "Heart disease and stroke statistics—2020 update: a report from the american heart association," *Circulation*, vol. 141, no. 9, pp. e139–e596, 2020.

[6] D. R. Pereira, P. P. Reboucas Filho, G. H. de Rosa, J. P. Papa, and V. H. C. de Albuquerque, "Stroke lesion detection using convolutional neural networks," in *2018 International joint conference on neural networks (IJCNN)*. IEEE, 2018, pp. 1–6.

[7] S. Bacchi, T. Zerner, L. Oakden-Rayner, T. Kleinig, S. Patel, and J. Jannes, "Deep learning in the prediction of ischaemic stroke thrombolysis functional outcomes: a pilot study," *Academic radiology*, vol. 27, no. 2, pp. e19–e23, 2020.

[8] M. Roder, G. H. Rosa, J. P. Papa, and D. C. Guimarães Pedronette, "Enhancing shallow neural networks through fourier-based information fusion for stroke classification," in *2021 34th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, 2021, pp. 378–385.

[9] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations," in *Proceedings of the 26th annual international conference on machine learning*, 2009, pp. 609–616.

[10] H. J. Nussbaumer, "The fast fourier transform," in *Fast Fourier Transform and Convolution Algorithms*. Springer, 1981, pp. 80–111.

[11] G. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Computation*, vol. 14, no. 8, pp. 1771–1800, 2002.

[12] H. Larochelle and Y. Bengio, "Classification using discriminative restricted boltzmann machines," in *Proceedings of the 25th international conference on Machine learning*. ACM, 2008, pp. 536–543.

[13] M. Roder, J. Almeida, G. H. De Rosa, L. A. Passos, A. L. Rossi, and J. P. Papa, "From actions to events: A transfer learning approach using improved deep belief networks," in *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2021, pp. 01–08.

[14] M. Welling, M. Rosen-zvi, and G. Hinton, "Exponential family harmoniums with an application to information retrieval," in *Advances in Neural Information Processing Systems 17*, L. Saul, Y. Weiss, and L. Bottou, Eds. MIT Press, 2005, pp. 1481–1488.

[15] G. Hinton, "A practical guide to training restricted boltzmann machines," in *Neural Networks: Tricks of the Trade*, ser. Lecture Notes in Computer Science, G. Montavon, G. Orr, and K.-R. Müller, Eds. Springer Berlin Heidelberg, 2012, vol. 7700, pp. 599–619.

[16] Y.-L. Boureau, J. Ponce, and Y. LeCun, "A theoretical analysis of feature pooling in visual recognition," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 111–118.

[17] R. C. Gonzalez and R. E. Woods, *Processamento de imagens digitais*. Editora Blucher, 2000.

[18] M. S. Sirsat, E. Fermé, and J. Câmara, "Machine learning for brain stroke: A review," *Journal of Stroke and Cerebrovascular Diseases*, vol. 29, no. 10, p. 105162, 2020.

[19] E. S. S. Rangel, A. G. S. Belasco, and S. Diccini, "Qualidade de vida de pacientes com acidente vascular cerebral em reabilitação," *Acta paulista de enfermagem*, vol. 26, no. 2, pp. 205–212, 2013.

[20] M. Lotfollahi, S. Berisha, D. Daeinejad, and D. Mayerich, "Digital staining of high-definition fourier transform infrared (ft-ir) images using deep learning," *Applied spectroscopy*, vol. 73, no. 5, pp. 556–564, 2019.

[21] H. Chen, L. Huang, T. Liu, and A. Ozcan, "Fourier imager network (fin): A deep neural network for hologram reconstruction with superior external generalization," *Light: Science & Applications*, vol. 11, no. 1, p. 254, 2022.

[22] Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar, "Fourier neural operator for parametric partial differential equations," *arXiv preprint arXiv:2010.08895*, 2020.

[23] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics Bulletin*, vol. 1, no. 6, pp. 80–83, 1945.

[24] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR*, 2015.

[25] S. Pedemonte, B. Bizzo, S. Pomerantz, N. Tenenholtz, B. Wright, M. Walters, S. Doyle, A. McCarthy, R. R. De Almeida, K. Andriole *et al.*, "Detection and delineation of acute cerebral infarct on dwi using weakly supervised machine learning," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 81–88.

[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE CVPR*, 2016, pp. 770–778.

[27] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.