Enhancing Shallow Neural Networks Through Fourier-based Information Fusion for Stroke Classification

Mateus Roder, Gustavo Henrique Rosa,
João Paulo Papa
Departamento de Computação
Universidade Estadual Paulista (Unesp)
Bauru, Brasil
{mateus.roder, gustavo.rosa, joao.papa}@unesp.br

Daniel Carlos Guimarães Pedronette
Departamento de Estatística, Matemática Aplicada e Computacional
Universidade Estadual Paulista (Unesp)
Rio Claro, Brasil
daniel.pedronette@unesp.br

Abstract—Deep learning techniques have been widely researched and applied to several problems, ranging from recommendation systems and service-based analysis to medical diagnosis. Nevertheless, even with outstanding results in some computer vision tasks, there is still much to explore as problems are becoming more complex, or applications are demanding new restrictions that hamper current techniques performance. Several works have been developed throughout the last decade to support automated medical diagnosis, yet detecting neural-based strokes, the so-called cerebrovascular accident (CVA). However, such approaches have room for improvement, such as the employment of information fusion techniques in deep learning architectures. Such an approach might benefit CVA detection as most state-of-the-art models use computer-based tomography and magnetic resonance imaging samples. Therefore, the present work aims at enhancing stroke detection through information fusion, mainly composed of original and Fourier-based samples, applied to shallow architectures (Restricted Boltzmann machines). The whole picture employs multimodal inputs, allowing data from different domains (images and Fourier transforms) to be learned together, improving the model's predictive capacity. As the main result, the proposed approach overpassed the baselines, achieving the remarkable accuracy of 99.72%.

I. INTRODUCTION

Artificial Intelligence (AI) has recently been highlighted in several real-world scenarios, mainly due to its high efficiency in solving particular tasks, such as digital image classification and the high connectivity between cyber-physical systems. As such, different technologies are constantly created and integrated by the search of prominent and autonomous resources, such as AI-based Computer Vision (CV) systems, which take advantage of the exponential increase of data.

It is essential to remark that computer vision is closely related to the human visual system, aiming to produce high-level representations of the real world and enabling intrinsic features to conduct detection and classification understandings [1]. Nevertheless, the real world is tough to be parametrized and poses an arduous task in representing complex applications, especially when dealing with adverse scenarios, such as luminance variation, distinct perspective, and image resolution [2].

Nowadays, most AI techniques are based on deep learning (DL) [3]–[5], which attempts to mimic the human visual processing and the hierarchical learning of features. Such an approach is conducted by creating layers of high-level abstraction responsible for extracting different yet informative features. Some of DL's primary researched techniques are the Convolutional Neural Networks (CNN) [3] and Deep Boltzmann Machines [5]. The former attempts to model hierarchical information through convolutional filters, where the latter models the hierarchical information as a causality model, where few hidden layers extract relevant information. Usually, deep networks require massive data to provide efficient training, which occasionally hinders some applications.

Despite that, there have been remarkable advances in Artificial Intelligence and computer vision techniques, primarily due to the advance of specialized hardware and Graphics Processing Units (GPU). Thus, such advances have allowed the usage of DL-based techniques in a wide range of areas, such as security monitoring [6], [7] and medical diagnosis [8]–[10].

Regarding medical applications, an interesting problem stands for the detection and diagnosis of cerebrovascular accidents. This lesion damages the cerebral tissue and is mainly caused due to a blood supply alteration in brain regions. Additionally, it may permanently impair or reduce brain functions and even lead to patient death [11]. However, the main problem does not concern its mortality, as different and complex sequels accompany it and diminish life's quality of the affected patients. A researched conducted by Rangel et al. [12] concluded that post-accident sequels lead patients to have difficulty with trivial tasks. Authors have observed that 49.6% of the patients presented moderate to severe dependency, while 49.7% appeared to have depressive and dysphoric feelings. Furthermore, the treatment of those people represents a cost of billions of dollars per year [11].

A well-performed detection and classification can aid doctors in asserting the patient's treatment or even when the early diagnosis is uncertain. The computed tomography (CT) images

occasionally decide the final verdict regarding diagnosis and treatment. Besides that, an image-based diagnosis is a powerful tool capable of distinguishing the accident type and the impacted region.

Some recent works have applied Artificial Intelligence to aid medical diagnosis. Pereira et al. [13] employed such a technique to classify images of patients with cerebrovascular accidents. The authors obtained high accuracy rates regarding cerebrovascular accident detection using CT-based images. On the other hand, medical images are hard to obtain and harder to label, which leads to small databases. Thus, the authors also stated that their dataset (300 images) was relatively small to take advantage of the employed techniques' full power.

Even with the significant performance obtained by Pereira et al. [13] and other related works, most of them use deep neural networks and complex systems, which may be challenging to be trained and depend on large volumes of data. Therefore, the main objective of the proposed work is to detect and classify cerebrovascular accidents using a small dataset and less-complex neural networks, such as the Restricted Boltzmann Machines (RBM). Additionally, a new RBM-based methodology has been proposed and used throughout this work, i.e., to employ the frequency information representation and fusion with the spatial domain. As far as the authors are concerned, no single combination of these networks with the Fourier transform or their application on cerebrovascular accidents tasks is reported to date.

The proposed approach employs information fusion from different data domains, i.e., it uses different modalities to feed the classification system. Information fusion uses original CT-based images and their respective Fourier transforms [14]. Thus, there are both spatial and frequency representations, allowing the extraction of different patterns and features. At last, the present work contributes with the following: (i) it introduces a new framework for classification and tomography image processing (without convolutional neural networks); (ii) it proposes the coupling of Fourier transforms and Restricted Boltzmann Machines; and (iii) it provides an effective and efficient application for the medical area.

The remainder of this work 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 the conclusions.

II. THEORETICAL BACKGROUND AND RELATED WORKS

This section presents the theoretical background regarding Restricted Boltzmann Machines, Fourier transform, and cerebrovascular accident. Concomitantly, some of the main works related to the depicted topics are cited.

A. Restricted Boltzmann Machines

Restricted Boltzmann Machines are stochastic neural networks based on physical principles of energy, entropy, and temperature. They are composed of two layers of neurons/units (visible and hidden), capable of modeling problems through unsupervised [15] and supervised learning [16].

In short, an RBM represents a bipartite graph with non-directional connections. Figure 1 describes the architecture of a Restricted Boltzmann Machine, with the visible layer $\mathbf{v}$ having m units, and the hidden layer $\mathbf{h}$ with n neurons. The real-valued matrix $\mathbf{W}$ models the weights (neural connection) between the visible and hidden neurons, with a size of $m \times n$.
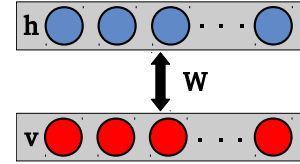

Fig. 1. Standard RBM architecture.

Initially, RBMs were developed using visible and hidden neurons with binary states sampled from a Bernoulli distribution. Later, Welling et al. [17] and Hinton [18] showed variations for the neurons that can be employed in an RBM, such as binomials, rectified linear (ReLU), and Gaussians. In this work, the binary-based version is used and further discussed.

Let $\mathbf{v}$ and $\mathbf{h}$ be the visible and hidden binary units, such that $\mathbf{v} \in \{0, 1\}^m$ and $\mathbf{h} \in \{0, 1\}^n$. The energy of a Bernoulli Restricted Boltzmann Machine is modeled as follows:

$$E(\mathbf{v}, \mathbf{h}) = -\sum_{i=1}^{m} a_i v_i - \sum_{j=1}^{n} b_j h_j - \sum_{i=1}^{m} \sum_{j=1}^{n} v_i h_j w_{ij}, \quad (1)$$

where $\mathbf{a}$ and $\mathbf{b}$ are the values of the biases (bias) of the visible and hidden units, respectively. The probability of a joint configuration $(\mathbf{v}, \mathbf{h})$ is calculated as follows:

$$P(\mathbf{v}, \mathbf{h}) = \frac{e^{-E(\mathbf{v}, \mathbf{h})}}{\sum_{\mathbf{v}, \mathbf{h}} e^{-E(\mathbf{v}, \mathbf{h})}}, \quad (2)$$

where the equation denominator is a normalization factor known as the partition function, which considers all possible configurations involving visible and hidden units, being intractable for high-dimensional spaces, such as images.

Since RBMs are bipartite graphs, the information can flow non-directional, i.e., from the visible to the hidden layer and vice-versa, making it possible to formulate the states' probabilities in a mutually independent way. The Equations 3 and 4 describe the conditional probabilities for each layer of an RBM:

$$P(\mathbf{v}|\mathbf{h}) = \prod_{i=1}^{m} P(v_i|\mathbf{h}) \quad (3)$$

and

$$P(\mathbf{h}|\mathbf{v}) = \prod_{j=1}^{n} P(h_j|\mathbf{v}), \quad (4)$$

where $P(\mathbf{v}|\mathbf{h})$ and $P(\mathbf{h}|\mathbf{v})$ represent the probabilities of the visible layer given the states of the hidden layer, and the hidden layer probabilities given the visible layer states, respectively.

From the Equations 3 and 4, it is possible to obtain the probability of activating a single neuron i given the hidden states and the probability of activating a single neuron j given the visible states. The Equations 5 and 6 describe these activations:

$$P(v_i = 1|\mathbf{h}) = \sigma \left(\sum_{j=1}^{n} w_{ij} h_j + a_i \right) \quad (5)$$

and

$$P(h_j = 1|\mathbf{v}) = \sigma \left(\sum_{i=1}^{m} w_{ij} v_i + b_j \right), \quad (6)$$

where $\sigma(\cdot)$ stands for the logistic function.

In short, the training process of an RBM aims to maximize the observed probabilities of a $P(\mathbf{v})$ configuration, while it is necessary to estimate and adjust the values of $\mathbf{W}$, $\mathbf{a}$ and $\mathbf{b}$ through the optimization of the likelihood cost function. Hinton [15] proposed an efficient method for training, using the input data directly as the probabilities of the visible neurons, known as Contrastive Divergence (CD). Such an approach uses Gibbs sampling to infer the hidden and visible layers through Equations 5 and 6.

In addition to binary neurons, RBMs can supply visible neurons capable of working with non-binary data (continuous), helpful in modeling different input types or signals. Thus, the energy function can be changed as follows:

$$E(\mathbf{v}, \mathbf{h}) = \sum_{i=1}^{m} \frac{(v_i - a_i)^2}{2\sigma_i^2} - \sum_{j=1}^{n} b_j h_j - \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{v_i}{\sigma_i} h_j w_{ij}, \quad (7)$$

where σ_i and σ_i^2 represent the standard deviation and variance, respectively, for each neuron i.

Considering the binary RBM derivatives, the conditional probabilities of the neurons of the visible layer are depicted as follows:

$$P(v_i = 1|\mathbf{h}) \sim \mathcal{N} \left(\sum_{j=1}^{n} w_{ij} h_j + a_i, \sigma_i^2 \right). \quad (8)$$

It is easy to notice that standardized data ($\sigma_i = 1$) leads to a simpler and less complex Equation 8.

B. Fourier Transform

The Fourier transform is a powerful mathematical tool dated from the mid-1800s. Its postulation indicates that any periodic function can be expressed as a sum of sines and cosines of different frequencies, each one weighted by a different coefficient, regardless of the function complexity. If such function is not periodic, it is still possible to apply the decomposition in sines and cosines, but from the integral

under the curves. With the computing advent and technological advances, the fast Fourier transform (FFT) was developed [19].

The Fourier transform can be applied in continuous or discrete domains, where the latter enables efficient approximations computationally fast for calculations. The FFT allows several manipulations of digital signals, mainly in images, such as the application of numerous filters in the frequency domain [19]. In short, the DFT is applied to a given D-dimensional array producing a complex array with the same dimension, in addition to the magnitude (or module) and phase angle. Mathematically, the discrete Fourier transform (DFT) is defined for two dimensions, for example, as follows [19]:

$$\begin{aligned} F(u,v) &= R(u,v) + jI(u,v) \\ &= |F(u,v)|e^{j\phi(u,v)}, \end{aligned} \quad (9)$$

where R and I stand for the real and imaginary components, respectively, u and v represent the two-dimensional matrix coordinates, and ϕ is the phase angle. The magnitude and phase angle of the input signal are defined by Equations 10 and 11, respectively:

$$|F(u,v)| = [R^2(u,v) + I^2(u,v)]^{1/2} \quad (10)$$

and

$$\phi(u,v) = \arctan \left[\frac{I(u,v)}{R(u,v)} \right]. \quad (11)$$

C. Cerebrovascular Accident

A cerebrovascular accident, commonly known as stroke, is characterized by an injury that abruptly strikes the brain tissue. It is mainly caused by changes in the blood supply to a certain region of the brain, resulting in loss or reduction of its functions. The stroke can be classified into Ischemic (most common), characterized by the blockage of a vessel responsible for brain irrigation, and Hemorrhagic, characterized by the rupture of a blood vessel in or around the brain. Figure 2 depicts such differences.
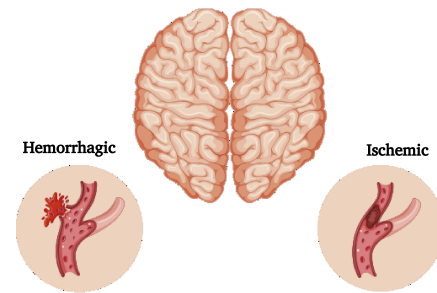

Fig. 2. Stroke types representation. Adapted from: https://abavc.org.br/index.php/sobre-o-avc/

A stroke can occur early due to several risk factors. These are categorized as non-modifiable and modifiable. The first covers age, sex, race, and ethnicity, while the latter factors are related to clinical conditions, such as heart disease and

diabetes mellitus, in addition to lifestyle factors, such as physical inactivity, obesity, malnutrition, tobacco use, and alcohol consumption. Stroke prevention and subsequent sequelae development are closely related to the earlier identification and control of modifiable risk factors, which play a significant role in such diagnosis. The diagnosis involves a detailed medical history, physical and neurological exams, and brain imaging tests [20], making early assumptions difficult.

Millions of people worldwide are affected by strokes every year, where a highly expressive amount ends up dying [11]. However, the biggest stroke-related problem goes beyond mortality, i.e., many survivors have lifelong post-stroke sequelae that are varied and complex. As part of these sequelae, so many patients develop difficulties in trivial daily activities, showing degrees of moderate and severe dependencies for such tasks [12]. In addition to the related work developed by Pereira et al. [13], Medicine has had the support of Artificial Intelligence to assist diagnosis and to detect brain abnormalities by images, such as hemorrhages, thrombosis, minor cranial fractures, and stroke. Such applications generally use images from computed tomography or magnetic resonance imaging (MRI).

D. Related Works

The first work to deal with multimodal Boltzmann Machines was conducted by Srivastava and Salakhutdinov [21], where they used Deep Boltzmann Machines (DBM) to learn several modalities of input data, such as images and text, to compose larger and more robust models. The work carries out extensive experiments and shows that the addition of more than one type of data can favor the learning of Boltzmann Machines considering image classification and image/text retrieval tasks.

Similarly, Wang and Ji [22] used DBMs to handle multimodal data, yet in recognizing events in security videos. The proposed model integrates into levels, attributes, semantics, and high-level contexts, being trained with approximate learning based on mean-field, hence, allowing it to be used directly to infer classes of events through Gibbs sampling. The work was evaluated in two surveillance databases with videos extracted from real environments, reaching promising results. However, the authors found a high complexity to achieve contextual interactions, and they had to use several auxiliary techniques for extracting attributes such as SIFT (Scale-Invariant Feature Transform), BOW (Bag-Of-Words), and STIP (Spatial-Temporal Interest Points).

Regarding Fourier transform-based deep learning, Sharma et al. [23] proposed using discrete Fourier transforms to predict the rain effect in images with the DFT coefficients as input of a deep convolutional neural network, and further remove such an effect from the frequency domain. Ryu et al. [24] proposed a new sampling layer (pooling layer) based on the discrete Fourier transform for convolutional neural networks. The DFT-based sampling replaced the traditional sampling layer (maximum/average) between convolution and fully-connected layers to retain the translation invariance and shape preservation.

Another practical application of the Fourier transform was the digital coloring of high definition infrared images through deep learning [25]. The authors showed it is possible to use unsupervised deep learning coupled with FFT to map images in the infrared spectrum to high-resolution images for histological stains in which there is an enormous difficulty acquiring categorized data.

Concerning AI-based medical diagnosis, Pereira et al. [13] employed convolutional neural networks to detect and classify images of patients with/without stroke. The authors obtained relatively high accuracy rates regarding stroke classification using CT-based images. However, medical images are hard to obtain and harder to label, which may lead to small databases, as [13] pointed. Thus, the authors also stated that their dataset (300 images) was relatively small to take advantage of the employed CNNs' power.

Pedemonte et al. [26] proposed a modified convolutional network (YNet) that used magnetic resonance images as training data based on weakly-supervised learning. The study aimed to segment the images and detect strokes in patients for the correct targeting of treatment therapies. The work was successful in the task; however, the proposed model has a complex architecture and consumed much training time. Lee et al. [27] developed a deep yet comprehensible learning system capable of detecting acute intracranial hemorrhage and five subtypes with unenhanced CT scans. The authors used a small dataset (904 samples) for training the model, which achieved performance similar to specialist radiologists in two independent test datasets containing 200 and 196 cases, respectively. Such a system uses attention maps to understand better the decisions.

Finally, Bacchi et al. [28] used a promising approach to detect ischemic stroke thrombolysis, where they applied two neural networks, i.e., a CNN to process computed tomography images and a 3-layer perceptron network to treat structural data from medical guides (age, gender, etc.). The results were promising and showed an exciting line for the development of robust tools for medicine.

III. FOURIER-BASED MULTIMODAL RESTRICTED BOLTZMANN MACHINES

This work proposes to represent information in different domains through the Fourier transform. We argue that the use of information fusion for the training of Restricted Boltzmann Machines can successfully be applied to the problem of stroke detection and classification of their types (ischemic and hemorrhagic). Additionally, these networks have been chosen due to their simplicity and lack of work involving the researched topics.

Information fusion aggregates data from different domains and is often denoted as multimodal due to the difference in the aggregated distributions. Multimodality can be used with N data types, such as bi and tri-modal. Concerning the proposed work, the data is composed of original computed tomography images and their respective Fourier transforms, where the

latter is composed of two different components: magnitude (Equation 10) and phase angle (Equation 11).

Let MultFRRBM (Multimodal Fourier-based RBM) denotes the proposed approach, where its variations MultFRRBM-P, MultFRRBM-M and MultFRRBM-PM, denote which components are used to compose the multimodality[1]. Formally, data entry representations can be defined as follows: let l be an image with spatial components $(x_1, x_2) \in \Re^2$. Let its 2-D Fourier transform be denoted by F, with module $|F(x_1, x_2)|$ and phase angle $\phi(x_1, x_2)$ defined by Equations 10 and 11, respectively.

Given the three possible MultFRRBM architectures, they are trained under the standard unsupervised learning approach [15] and fine-tuned for the classification task through an additional Softmax layer, which receives the activations of the layers hidden $\mathbf{h^1}$ and $\mathbf{h^2}$ for two types of data, or $\mathbf{h^1}$, $\mathbf{h^2}$ and $\mathbf{h^3}$ for three types of data. Figure 3 illustrates a MultFRRBM-P with two entries, where the first RBM (GaussianRBM) receives the original CT image and the second (FourierRBM) receives the phase angle from the DFT applied to the image.
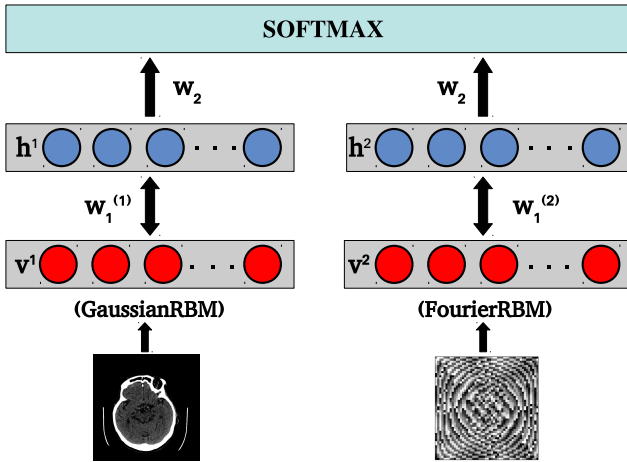

Fig. 3. MultFRRBM-P architecture with two types of inputs, i.e., original images and the phase angle of the DFT.

In this work, the Multimodal RBMs address the problem employing Gaussian visible units, i.e., image pixels as continuous values, for the original grayscale images. Initially, one or more Gaussian-based RBMs are used to process the DFT data, i.e., if the final architecture uses the phase angle (MultFRRBM-P), the data is normalized by a standard Gaussian and fed to the next RBM. The same process occurs for RBMs with magnitude (MultFRRBM-M), where the magnitude of the frequency spectrum is normalized by a standard Gaussian and fed to another Gaussian RBM. Finally, when both DFT components are used (MultFRRBM-PM), phase and magnitude are normalized and fed into two Gaussian simultaneously to a base RBM. Figure 4 illustrates the processing described.

---

[1]P stands for the phase, M stands for the magnitude, and PM stands for both.
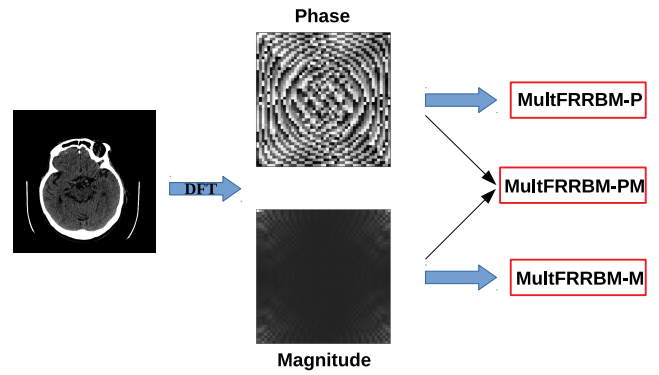

Fig. 4. Scheme of the transformation applied for entry into multimodal models.

Therefore, the proposed approach aggregates information from different domains, i.e., the spatial component originated 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 most suitable for the problem in question and gives greater flexibility to the application.

IV. EXPERIMENTAL METHODOLOGY

This section presents an overall description of the employed dataset, as well as a brief explanation of how the proposed approach was modeled, followed by the experimental setup.

A. Dataset

Pereira et al. [13] proposed the employed dataset, which consists of 25 CT scans of the brain, comprising patients[2] 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, they are composed of irregular patterns, irregular lighting, and different structural characteristics.

Furthermore, Pereira et al. [13] used two additional preprocessing methodologies to image generation, image segmentation to extract only the brain, and brain images generated from the radiological density maps (tissue density atribute map); however, less pre-processing might avoid possible delays in the system in real-world applications. Additionally, their experimental results showed that these pre-processing methodologies did not improve the CNNs performance. Thus, this work only considers the original grayscale images.

Each original image has a resolution of 512×512 pixels, yet, they have been resized to 50×50 to diminish the number of computational resources needed and the RBMs training time. Each image has been grayscaled and normalized according to a standard Gaussian distribution, i.e., zero mean and unit variance. Figure 5 illustrates some samples of CT-based scans from a healthy brain, hemorrhagic stroke, and ischemic stroke, respectively.

---

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

Fig. 5. Tomography of a brain: healthy, with hemorrhagic stroke, and with ischemic stroke, respectively.

B. Experimental Setup

The experimental setup comprises four different RBM architectures, yet with the same training process, i.e., networks trained with the Contrastive Divergence, and with an equal number of epochs for each RBM that composed the multimodal model. The baseline architectures are the Convolutional Neural Networks, CNN_1, and CNN_2, proposed by Pereira et al. [13], while the RBM corresponds to a pure Gaussian-based RBM, and the previously mentioned multimodal variants, i.e., MultFRRBM-P, MultFRRBM-M, and MultFRRBM-PM. The hyperparameter configuration of these architectures is depicted by Table I.

The values described by Table I correspond to empirically chosen hyperparameters, which were selected in preliminary experiments. Also, the number of hidden neurons is based on the anchor model (GaussianRBM), which showed a good performance with $2,000$ units. Such a value is the maximum amount of neurons for all proposed architectures, hence providing a fair comparison between architectures that use one or more RBMs. On the other hand, both learning rates and momentums have been chosen to provide smoother convergences (lower values), while 50 training epochs and mini-batches of size equals to 10 complements a thorough learning procedure.

Considering the employed dataset, we opted to follow the methodology proposed by Pereira et al. [13], 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 ($75/25$). Additionally, 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 [29] with 5% of significance.

Regarding the last adjusting step (fine-tuning), we added a Softmax layer on the top of the RBMs, trained with an Adam [30] optimizer using a learning rate of 0.00001 for RBMs and 0.001 for the Softmax. This additional classification layer had its hyperparameters empirically chosen, while it was trained for 50 epochs using mini-batches with a size equal to 60. Finally, the reported metric stands for the same used by Pereira et al. [13], which is the standard accuracy.

V. EXPERIMENTS

Table II depicts the mean results and their respective standard deviation obtained from the 15 independent runnings, considering both data splits, i.e., $50/50$ and $75/25$. Additionally, the bolded cells represent statistically similar results according to the Wilcoxon signed-rank test with a p-value of 5%. Finally, the result that achieved the highest mean for each split is underlined.

Still glancing at Table II, one can observe that all RBM-based models achieved higher accuracy than the CNNs, considering both data splits. Such results are remarkable as non-convolutional models obtained a better recognition rate in such an important task. Additionally, the employed RBM models have minimal image pre-processing and low architecture complexity, with at most $2,000$ hidden neurons.

Generally speaking, the best results were the GaussianRBM in the $50/50$ split, achieving almost 7% more performance than the best convolutional model CNN_1. Considering the same split, the multimodal models could not perform as well as the unimodal one, indicating that the multimodality might need more training data; however, their results were still superior to the ones obtained by CNN_1 and CNN_2.

Regarding the second split ($75/25$), there has been a significant improvement in the multimodal models, where all approaches obtained a mean accuracy higher than 99%. Amongst these models, the MultFRRBM-P stood out, achieving 99.72% and surpassing both unimodal and baseline models. Additionally, it had the lowest standard deviation, while every other RBM-based model had better mean accuracy and lower standard deviation than the CNNs.

It is worth noticing that MultFRRBM-PM models obtained the lowest mean accuracy among the tested RBM- based models, which led us to believe in the following hypotheses: (i) the magnitude component aggregates less information than the phase component, which can be observed by comparing both MultFRRBM-P and MultFRRBM-M models; and (ii) the combination of magnitude and phase components can generate models that may require more data for training. Additionally, Figures 6 and 7 illustrate the fine-tuning learning procedure, where we can highlight an instability of the MultFRRBM-PM when compared to other models, as shown in the regions with the zoom at the end of the fine-tuning epochs.

Additionally, the number of parameters from the networks was analyzed, to complement a fine piece of information, which gives us the direct comparison between models and enables valuable insights. Such values were calculated based on the models' description in this work for the RBM-based, and based on Pereira et al. [13] for the CNNs. Table III shows the number of parameters for the proposed approach and the baseline architectures, as well as the approximated number of such parameters (to visualization simplicity), and the ratio between the RBM-based models and CNN baselines, which is the parameters number quotient regarding the architectures.

From Table III, one can notice that RBM-based models, i.e., the GaussianRBM and its multimodal versions, have significantly more parameters than CNN_1, almost $43\times$ (here, "$\times$" means times). However, such a greater number was not a barrier for stroke detection and classification since all RBM models outperformed the results from CNN_1. On the

TABLE I
HYPERPARAMETER CONFIGURATION FOR RBM-BASED MODELS.

Architecture	# Hidden Neurons	Learning Rate	Momentum
GaussianRBM	$2,000$	0.0001	0.5
MultFRRBM-P	$500; 1,500$	$0.001; 0.0001$	$0.9; 0.5$
MultFRRBM-M	$500; 1,500$	$0.001; 0.0001$	$0.9; 0.5$
MultFRRBM-PM	$250; 250; 1,500$	$0.001; 0.001; 0.0001$	$0.9; 0.9; 0.5$

TABLE II
MEAN ACCURACY AND ITS RESPECTIVE STANDARD DEVIATIONS FOR THE
TEST SET CONSIDERING BOTH DATA SPLITS.

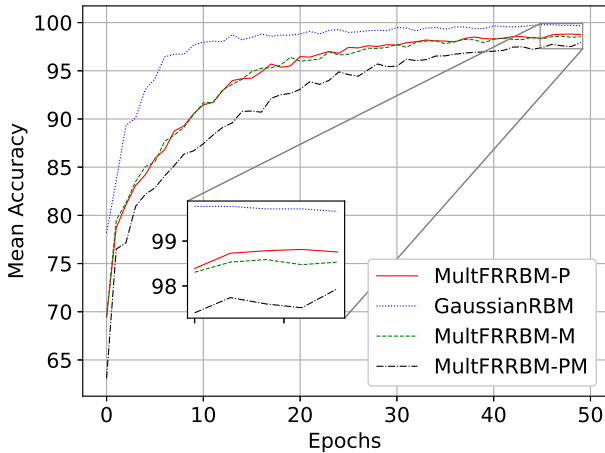| Architecture	50/50	75/25
GaussianRBM	$\mathbf{99.66 \pm 0.52}$	99.66 ± 0.52
MultFRRBM-P	98.76 ± 1.58	$\mathbf{99.72 \pm 0.40}$
MultFRRBM-M	98.53 ± 1.70	$\mathbf{99.49 \pm 0.60}$
MultFRRBM-PM	97.94 ± 1.66	$\mathbf{99.32 \pm 0.71}$
CNN$_1$ [13]	93.46 ± 16.54	97.20 ± 2.45
CNN$_2$ [13]	83.55 ± 13.09	77.33 ± 22.24

Fig. 6. Mean accuracy over the test set, considering the $50/50$ split.

TABLE III
PARAMETERS ANALYSIS OVER THE EMPLOYED MODELS.

	RBM-based	CNN$_1$ [13]	CNN$_2$ [13]
Parameters	$5,006,000$	$116,874$	$60,954,656$
Approximated	$\approx 5M$	$\approx 117K$	$\approx 61M$
Ratio	$-$	$\approx 42.83\times$	$\approx 0.08\times$

other hand, the RBM-based ones have $0.08\times$ the number of the total parameters of CNN$_2$, which is an impressive
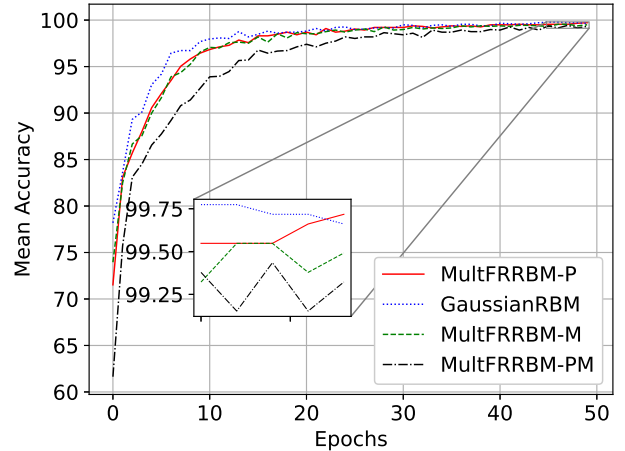

Fig. 7. Mean accuracy over the test set, considering the $75/25$ split.

value, overpassing such CNN in fewer parameters and higher accuracy overall. Finally, one can highlight that convolution-based networks are far more sensitive to data volume, and RBM-based models can be robust where less data is provided.

VI. CONCLUSION

This work explores the use of multimodal data in the context of Restricted Boltzmann Machines applied to the cerebrovascular accident classification task. The employed multimodality used the Fourier transform to provide two additional data distributions, i.e., the magnitude and the phase angle of the transform components.

We argue that the proposed approach can achieve state-of-the-art results and surpasses the previously approaches regarding the experimental tests. Every proposed model has been superior to the baselines proposed by Pereira et al. [13], which led us to the assumption that the evaluated task does not require complex and convolving models nor large volumes of data. Another point that should be highlighted is the potential of multimodal inputs in RBM-based models, especially when combined with the power of the Fourier transform. According to Table II, it is clear that the phase angle contributed more to the model discriminative power than the magnitude, indicating that such an approach is viable and should be more explored.

Finally, regarding future works, we aim to deeply study the possibility to make an auto-data-augmentation with the

employed networks, as suggested by Roder et al. [31], in which the authors generated synthetic images from the original ones with RBMs to increase the data volume, proving to be a promising approach to performance improvement in the medical domain. Also, the authors intend to use convolutional RBMs rather than RBMs.

ACKNOWLEDGMENT

REFERENCES

[1] C. Bishop, *Neural networks for pattern recognition*. Oxford University Press, 1995.

[2] Y. LeCun, K. Kavukcuoglu, and C. Farabet, "Convolutional networks and applications in vision," in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*, 2010, pp. 253–256.

[3] 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.

[4] 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.

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

[6] S. Mohammadi, A. Perina, H. Kiani, and V. Murino, "Angry crowds: Detecting violent events in videos," in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Springer International Publishing, 2016, pp. 3–18.

[7] B. Afonso, L. Cinelli, L. Thomaz, A. Silva, E. da Silva, and S. Netto, "Moving-camera video surveillance in cluttered environments using deep features," 10 2018, pp. 2296–2300.

[8] K. K. Wong, G. Fortino, and D. Abbott, "Deep learning-based cardiovascular image diagnosis: A promising challenge," *Future Generation Computer Systems*, vol. 110, pp. 802–811, 2020.

[9] Q. Hu, H. M. Whitney, and M. L. Giger, "A deep learning methodology for improved breast cancer diagnosis using multiparametric mri," *Scientific reports*, vol. 10, no. 1, pp. 1–11, 2020.

[10] 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.

[11] 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 statistics2020 update: a report from the american heart association," *Circulation*, vol. 141, no. 9, pp. e139–e596, 2020.

[12] 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.

[13] 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.

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

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

[16] 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.

[17] 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.

[18] 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.

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

[20] 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.

[21] N. Srivastava, R. Salakhutdinov *et al.*, "Multimodal learning with deep boltzmann machines." in *NIPS*, vol. 1. Citeseer, 2012, p. 2.

[22] X. Wang and Q. Ji, "Video event recognition with deep hierarchical context model," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4418–4427.

[23] P. K. Sharma, S. Basavaraju, and A. Sur, "Deep learning-based image de-raining using discrete fourier transformation," *The Visual Computer*, pp. 1–14, 2020.

[24] J. Ryu, M.-H. Yang, and J. Lim, "Dft-based transformation invariant pooling layer for visual classification," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 84–99.

[25] 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.

[26] 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.

[27] H. Lee, S. Yune, M. Mansouri, M. Kim, S. H. Tajmir, C. E. Guerrier, S. A. Ebert, S. R. Pomerantz, J. M. Romero, S. Kamalian *et al.*, "An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets," *Nature biomedical engineering*, vol. 3, no. 3, pp. 173–182, 2019.

[28] 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.

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

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

[31] M. Roder, L. A. Passos, L. C. F. Ribeiro, B. C. Benato, A. X. Falcão, and J. P. Papa, "Intestinal parasites classification using deep belief networks," in *International Conference on Artificial Intelligence and Soft Computing*. Springer, 2020, pp. 242–251.