Multi-View Breast Cancer Classification via Hypercomplex Neural Networks

Eleonora Lopez, *Graduate Student Member, IEEE*, Eleonora Grassucci, *Graduate Student Member, IEEE*, Martina Valleriani, and Danilo Comminiello, *Senior Member, IEEE*

***Abstract*—Traditionally, deep learning-based methods for breast cancer classification perform a single-view analysis. However, radiologists simultaneously analyze all four views that compose a mammography exam, owing to the correlations contained in mammography views, which present crucial information for identifying tumors. In light of this, some studies have started to propose multi-view methods. Nevertheless, in such existing architectures, mammogram views are processed as independent images by separate convolutional branches, thus losing correlations among them. To overcome such limitations, in this paper we propose a novel approach for multi-view breast cancer classification based on parameterized hypercomplex neural networks. Thanks to hypercomplex algebra properties, our networks are able to model, and thus leverage, existing correlations between the different views that comprise a mammogram exam, thus mimicking the reading process performed by clinicians. As a consequence, the proposed method is able to handle the information of a patient altogether without breaking the multi-view nature of the exam. Starting from the proposed hypercomplex approach, we define architectures designed to process two-view exams, namely PHResNets, and four-view exams, i.e., PHYSEnet and PHYSBOnet, with the ability to grasp inter-view correlations in a wide range of clinical use cases. Through an extensive experimental evaluation conducted with two publicly available datasets, CBIS-DDSM and INbreast, we demonstrate that our parameterized hypercomplex models clearly outperform real-valued counterparts and also state-of-the-art methods, proving that breast cancer classification benefits from the proposed multi-view architecture. Full code and pretrained models for complete reproducibility of our experiments are freely available at https://github.com/ispamm/PHBreast.**

***Index Terms*—Hypercomplex Neural Networks, Breast Cancer Classification, Multi-View Deep Learning, Hypercomplex Algebra**

I. Introduction

Among the different types of cancer that affect women worldwide, breast cancer alone accounts for almost one-third, making it by far the cancer with highest incidence among women [1]. For this reason, early detection of this disease is of extreme importance and, to this end, screening mammography is performed annually on all women above a certain age [2], [3]. During a mammography exam, two views of the breast are taken, thus capturing it from above, i.e., craniocaudal (CC) view, and from the side, i.e., mediolateral oblique (MLO) view. More in detail, the CC and MLO views of the same breast are known as ipsilateral views, while the same view of both breasts as bilateral views. Importantly, when reading a mammogram, radiologists examine the views performing a double comparison, that is comparing ipsilateral views along with bilateral views, as each comparison provides valuable information. Such multi-view analysis has been found to be essential in order to make an accurate diagnosis of breast cancer [4], [5].

Recently, many works are employing deep learning (DL)-based methods in the medical field and, especially, for breast cancer classification and detection with encouraging results [6]–[17]. Inspired by the multi-view analysis performed by radiologists, several recent studies try to adopt a multi-view architecture in order to obtain a more robust and performing model [18]–[29].

Such approaches often implement multi-view frameworks based on multi-path architectures. However, there are several issues associated with such method in the context of multi-view learning. As a matter of fact, a recent study demonstrates that a simple multi-path network can suffer from several problems [19]. To begin with, it is observed that the model can favour one of the two views during learning, thus relying mostly on that single view and not truly taking advantage of the multi-view input. This may often result in some form of overfitting. Additionally, the model might fail entirely in leveraging the correlated views and actually worsen the performance with respect to its single-view counterpart [19], [30]. Thus, it is made clear that improving the ability of deep networks to truly exploit the information contained in multiple views is still a largely open research question. To address and overcome these problems, we leverage a novel technique in deep learning, which relies in exotic algebraic systems, such as quaternions and, more in general, hypercomplex ones.

In recent years, quaternion neural networks (QNNs) have gained a lot of interest in a variety of applications [31]–[36]. The reason for this being the particular properties that characterize these models. As point of fact, thanks to the quaternion algebra rules on which these models are based on (e.g., the Hamilton product), quaternion networks possess the capability of modeling interactions between input channels, thus capturing internal latent relations within them and additionally reducing the total number of parameters by 75%, while still attaining comparable performance to its real-valued counterparts. Furthermore, built upon the idea of QNNs, the recent parameterized hypercomplex neural networks (PHNNs) generalize hypercomplex multiplications as a sum of Kronecker products, going beyond quaternion algebra, thus improving previous shortcomings by making these models applicable to any n-dimensional input (instead of just

Authors are with the Department of Information Engineering, Electronics and Telecommunications (DIET), Sapienza University of Rome, Italy. Corresponding author's email: eleonora.lopez@uniroma1.it.

3D/4D as the quaternion domain) thanks to the introduction of the parameterized hypercomplex multiplication (PHM) and convolutional (PHC) layer [37], [38].

Motivated by the aforementioned problems and the benefits of hypercomplex models, we propose a novel framework of multi-view learning for breast cancer classification based on PHNNs, taking a completely different approach with respect to existing methods in the literature. More in detail, we propose a family of parameterized hypercomplex ResNets (PHResNets) able to process ipsilateral views corresponding to one breast, i.e., two views. We also design two parameterized hypercomplex networks, namely PHYSBOnet and PHYSEnet, involving respectively a Shared Bottleneck and a Shared Encoder. Such architectures are aimed to process ipsilateral views of both sides, i.e., four views, capturing correlated information contained in ipsilateral views through PHC layers and additionally exploiting bilateral information by sharing parameters between views.

The advantages of our approach are manifold. Firstly, instead of handling the mammographic views independently, which results in losing significant correlations, our models process them as a unique component, without breaking the original nature of the exam. Secondly, thanks to hypercomplex algebra properties, the proposed models are endowed with the capability of preserving existing latent relations between views by modeling and capturing their interactions, thus mimicking the examination process of radiologists in real-life settings. Thirdly, our parameterized hypercomplex networks are characterised by the number of free parameters *halved* with respect to their real-valued counterparts. Finally, the proposed approach is portable and flexible. Indeed, PHC layers are easily integrated in any convolutional network by simply swapping convolutional layers with parameterized hypercomplex ones. On top of that, hypercomplex models not only possess the ability of processing multiple-views as a unique entity by modeling their correlations, but they are also capable of processing an arbitrary number of views, ranging from 1 up to n such as 2 or 4 and so on. Being parameterized by n, the user can set such parameter according to its needs and the dataset of choice. Thus, our proposed architectures can be easily applied for other types of multi-view or multimodal problems, such as multimodal brain tumor segmentation or multi-view chest X-ray disease classification, and smoothly integrated with any other backbone neural model.

We evaluate the effectiveness of our approach on two publicly available benchmark datasets of mammography images, namely CBIS-DDSM [39] and INbreast [2]. We conduct a meticulous experimental evaluation that demonstrates how our proposed models, owing to the aforementioned abilities, possess the means for properly leveraging information contained in multiple mammographic views and thus far exceed the performance of both real-valued baselines and state-of-the-art methods.

More concretely, our contributions are:

1) We present a novel approach for multi-view breast cancer classification, which processes mammographic views as a single exam, gaining advantages from the different views to output the diagnosis, as radiologists actually do. We expound how, exploiting hypercomplex algebra properties, our method is able to leverage correlations between the views, providing more accurate predictions.
2) We propose different architectures for both two-view and four-view exams, endowed with the ability of exploiting information contained in ipsilateral views and, in the latter case, bilateral views as well.
3) We introduce two multi-level models, namely PHYSBOnet and PHYSEnet, for the four-view scenario, that involve two steps focusing on a breast-level analysis first and on a patient-level analysis then and *vice versa*, in order to unveil the best way to process four-view exams.
4) We show that our method is flexible to be adapted to any model and to diverse exams comprised of one or multiple views, simply replacing the user-defined hyperparameter n.
5) We evaluate the validity of our assumptions on two publicly available benchmarks, performing experiments with two and with four-view exams. In each test we conduct, our parameterized hypercomplex models far exceed state-of-the-art methods while also demonstrating more robust results, proving the effectiveness of the proposed parameterized hypercomplex multi-view framework.

The rest of the paper is organized as follows. Section II gives a detailed overview of the multi-view approach for breast cancer analysis, delving into why it is important to design a painstaking multi-view method, Section III provides theoretical aspects and concepts of hypercomplex models, and Section IV presents the proposed method. The experiments are set up in Section V and evaluated in Section VI. A summary of the proposed models with specific data cases is provided in Section VII, while conclusions are drawn in Section VIII.

II. Multi-View Approach in Breast Cancer Analysis

Several types of imaging modalities exist for the detection process of breast cancer, such as mammography, ultrasound, biopsy and so on. Among these, mammography is considered the best imaging method for breast cancer screening and the most effective for early detection [3]. A mammography exam comprises four X-ray images produced by the recording of two views for each breast: the craniocaudal (CC) view, which is a top to bottom view, and a mediolateral oblique (MLO) view, which is a side view. The diagnosis procedure adopted by radiologists consists in looking for specific abnormalities, the most common being: masses, calcifications, architectural distortions of breast tissue, and asymmetries (when comparing the two breasts and the two views) [2]. During the reading of a mammography exam, employing multiple views is crucial in order to make an accurate diagnosis as they retain highly correlated characteristics. Admittedly, comparing ipsilateral views (CC and MLO views of the same breast) helps to detect eventual tumors, as sometimes they are visible only in one of the two views, and additionally helps to analyze the 3D structure of masses. Whereas, studying bilateral views (same view of both breasts) helps in locating masses as asymmetries

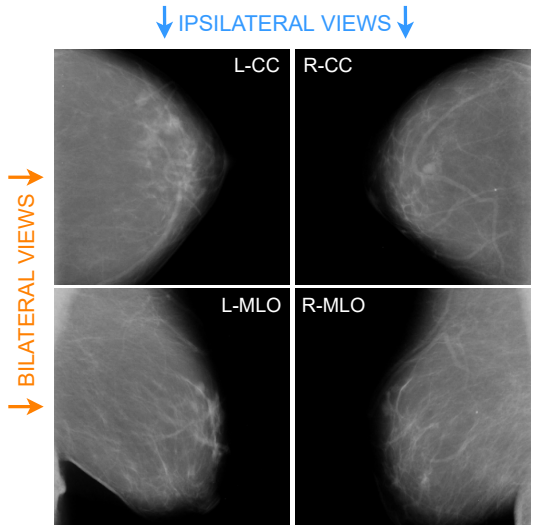

Fig. 1. Example of a mammography exam from the CBIS-DDSM dataset with the four views: left CC (L-CC), right CC (R-CC), left MLO (L-MLO) and right MLO (R-MLO). Horizontal couples are bilateral views (orange), while vertical couples are ipsilateral views (blue).

between them are an indicating factor [5]. An example of a complete mammogram exam with ipsilateral and bilateral views is shown in Fig. 1.

Given the multi-view nature of the exam and the multi-view approach employed by radiologists, many works are focusing on utilizing two or four views for the purpose of classifying breast cancer, with the goal of leveraging the information coming from ipsilateral and/or bilateral views. In point of fact, studies have shown that a model employing multiple views can learn to generalize better compared to its single-view counterpart, increasing its discriminative power, while also reducing the number of false positives and false negatives [6], [9], [18].

On one hand, some works show the advantages of leveraging multiple views adopting simple approaches, such as a method consisting in just taking the average of the predictions given by the same model when fed the two ipsilateral views [6]. On the other hand, more complex approaches consist in designing a multi-view architecture that should exploit correlations for learning and not just at inference time. Recent techniques for leveraging multiple views from a mammography exam propose architectures comprised of multiple convolutional neural networks (CNN) paths or *columns*, where each column processes a different view and its output is then concatenated together and fed to a number of fully connected layers to obtain the final output [18]–[25]. A recent study proposes an approach to leverage the complete mammography exam, thus comprised of four views, and focuses on studying several ways of combining information given by the different CNNs, i.e., view-wise (CC with MLO) or breast-wise (left with right) etc., finding that it is indeed an important aspect and proving that models are sensible to how such information is fused together [18]. Alternatively, a number of works [20]–[22], [26] adopt the same idea of having multiple columns but instead focus on using just two ipsilateral views. In particular, the reason for which ipsilateral views are employed most often, rather than bilateral ones, is that they correspond to the same breast and thus provide different information regarding the same tumor. Thus, they aid in both the detection and classification process, as they help to locate and discern malignant from benign findings thanks to the provided 3D view. Instead, bilateral views play more of an auxiliary role as they help in the detection process only.

Even though multimodal methods have been employed in a variety of recent deep learning works for breast cancer analysis, not much attention has been paid to how a DL model actually leverages information contained in the multiple views. Indeed, a recent study shows that such multimodal approaches suffer from the fact that the model might actually fail to exploit these information, leading to a counter-intuitive situation in which the single-view counterpart outperforms the multi-view one. Therefore, just employing an architecture with multiple columns for each view is not enough to really leverage the knowledge coming from the correlated inputs [19]. Even in other applications of multimodal learning (e.g., involving speech or text), it is a common phenomenon that DL networks not properly modelled fail to utilize information contained in the different input modalities [30]. Therefore, when processing multi-view exams, a painstaking and meticulous method has to be developed. To this end, we propose to leverage a novel method based on hypercomplex algebra, whose properties are described in the following section.

III. Quaternion and Hypercomplex Neural Networks

Quaternion and hypercomplex neural networks have their foundations in a hypercomplex number system $\mathbb{H}$ equipped with its own algebra rules to regulate additions and multiplications. Hypercomplex numbers generalize a plethora of algebraic systems, including complex numbers $\mathbb{C}$, quaternions $\mathbb{Q}$ and octonions $\mathbb{O}$, among others. A generic hypercomplex number is defined as

$$h = h_0 + h_i\hat{\imath}_i + \ldots + h_n\hat{\imath}_n, \qquad i = 1, \ldots, n \qquad (1)$$

whereby $h_0, \ldots, h_n$ are the real-valued coefficients and $\hat{\imath}_i, \ldots, \hat{\imath}_n$ the imaginary units. The first coefficient h_0 represents the real component, while the remaining ones compose the imaginary part. Therefore, the algebraic subsets of $\mathbb{H}$ are identified by the number of imaginary units and by the algebraic rules that govern the interactions among them. For instance, a complex number has just one imaginary unit, while a quaternion has three imaginary units and the vector product in this domain is not commutative so the Hamilton product has been introduced to multiply two quaternions. Interestingly, a real number can be expressed through eq. (1) by setting $i = 0$ and considering the real part only. It is important to note that subset domains exist solely at pre-defined dimensions, i.e., $n = 2, 4, 8, 16, \ldots$ while no algebra rules have been discovered yet for other values.

The addition operation is performed through an element-wise addition of terms, i.e.,, $h + p = (h_0 + p_0) + (h_i + p_i)\hat{\imath}_i +$

$\ldots + (h_n + p_n)\hat{\imath}_n$. Similarly, the scalar product can be formulated with $\alpha h = \alpha h_0 + \alpha h_i \hat{\imath}_i + \ldots + \alpha h_n \hat{\imath}_n$ by multiplying the scalar to real-valued components. Nevertheless, more specific formulas need to be introduced to model vector multiplication because of imaginary units interplays. Indeed, as an example, quaternions and octonions products are not commutative due to imaginary units properties for which $\hat{\imath}_1 \hat{\imath}_2 \neq \hat{\imath}_2 \hat{\imath}_1$. Therefore, quaternion convolutional neural network (QCNN) layers are based on the Hamilton product which organizes the filter weight matrix $\mathbf{W} = \mathbf{W}_0 + \mathbf{W}_1 \hat{\imath}_1 + \mathbf{W}_2 \hat{\imath}_2 + \mathbf{W}_3 \hat{\imath}_3$ to perform convolution with the input $\mathbf{x} = \mathbf{x}_0 + \mathbf{x}_1 \hat{\imath}_1 + \mathbf{x}_2 \hat{\imath}_2 + \mathbf{x}_3 \hat{\imath}_3$ as:

$$\mathbf{W} * \mathbf{x} = \begin{bmatrix} \mathbf{W}_0 & -\mathbf{W}_1 & -\mathbf{W}_2 & -\mathbf{W}_3 \\ \mathbf{W}_1 & \mathbf{W}_0 & -\mathbf{W}_3 & \mathbf{W}_2 \\ \mathbf{W}_2 & \mathbf{W}_3 & \mathbf{W}_0 & -\mathbf{W}_1 \\ \mathbf{W}_3 & -\mathbf{W}_2 & \mathbf{W}_1 & \mathbf{W}_0 \end{bmatrix} * \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \end{bmatrix}. \quad (2)$$

Processing multidimensional input with QCNNs has several advantages. Indeed, due to the filter submatrices $\mathbf{W}_i$, $i = 1, \ldots, 4$ reusing in eq. (2), QCNNs are defined with $1/4$ free parameters with respect to real-valued counterparts with the same architecture structure. Moreover, sharing the filter submatrices among input components allows QCNNs to capture internal relations in input components and to preserve correlations among them. However, this approach is limited to 4D inputs, thus various knacks are usually employed to apply QCNNs to different 3D inputs, such as RGB color images. In these cases, a padding channel is concatenated to the three-channel image to build a 4D image adding, however, useless information. Recently, novel approaches proposed to parameterize hypercomplex multiplications and convolutions to maintain QCNNs and hypercomplex algebras advantages, while extending their applicability to any nD input [37], [38]. The core idea of these methods is to develop the filter matrix $\mathbf{W}$ as a parameterized sum of Kroncker products:

$$\mathbf{W} = \sum_{i=0}^{n} \mathbf{A}_i \otimes \mathbf{F}_i, \quad (3)$$

whereby n is a tunable or user-defined hyperparameter that determines the domain in which the model operates (i.e., $n = 4$ for the quaternion domain, $n = 8$ for octonions, and so on). The matrices $\mathbf{A}_i$ encode the algebra rules, that is the filter organization for convolutional layers, while the matrices $\mathbf{F}_i$ enclose the weight filters. Both these elements are completely learned from data during training, thus grasping algebra rules or adapting them if no algebra exists for the specific value of n directly from inputs. The parameterized hypercomplex convolutional (PHC) layer is malleable to operate in any nD domain by easily setting the hyperparameter n, thus extending QCNNs advantages to every multidimensional input. Indeed, PHC-based networks can process color images in their natural domain ($n = 3$) without adding any uninformative channel (as previously done for QCNNs), while still exploiting hypercomplex algebra properties and preserving correlations and latent relations between channels. Moreover, due to the data-driven fashion in which this approach operates, PHC-based models with $n = 4$ outperform QCNNs both in terms of prediction accuracy and training as well as inference time [37], [38].

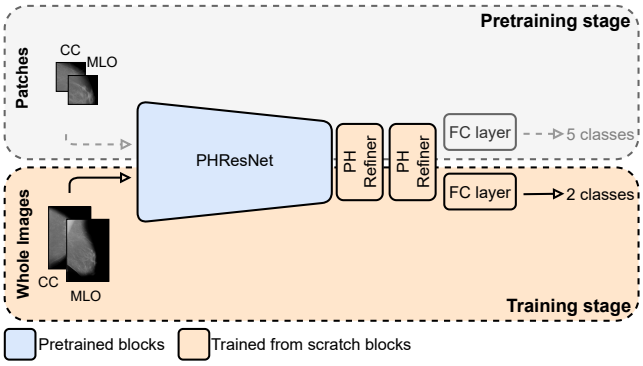

Fig. 2. Training pipeline and PHResNet overview. During the pretraining stage the PHResNet is trained on patches with five classes. At training stage on whole images, two PH convolutional refiner blocks are added.

Furthermore, PHC layers employ $1/n$ free parameters with respect to their real-valued counterpart, so the user can govern both the domain and the parameters reduction by simply setting the hyperparameter n.

Therefore, when input data show some correlation properties, it is possible to equip common convolutional neural networks with PHC layers by replacing real-valued operations with hypercomplex ones and to choose the proper domain depending on the input structure.

IV. Proposed Method

In the following section, we expound the proposed approach and we delineate the structure of our models and the training recipes we adopt. More in detail, we design *ad hoc* networks for two-view and four-view exams, i.e., a complete mammography exam. Thanks to hypercomplex algebra properties, our models are able to process the original exams as a unique component, without breaking its nature, grasping correlations among the different mammographic views and catching more information for classification.

A. Multi-view parameterized hypercomplex ResNet

The core idea of our method is to leverage information contained in multiple views through parameterized hypercomplex convolutional (PHC) layers in order to obtain a more performant and robust classifier for breast cancer.

ResNets are among the most widespread models for medical image classification [6], [9], [15], [18], [19], [40], [41]. They are characterized by residual connections that ensure a proper gradient propagation during training. A ResNet block is typically defined by:

$$\mathbf{y} = \mathcal{F}(\mathbf{x}) + \mathbf{x}, \quad (4)$$

where $\mathcal{F}(\mathbf{x})$ is usually composed by interleaving convolutional layers, batch normalization (BN) and ReLU activation functions. When equipped with PHC layers, real-valued convolutions are replaced with PHC to build parameterized hypercomplex ResNets (PHResNets), therefore $\mathcal{F}(\mathbf{x})$ becomes:

$$\mathcal{F}(\mathbf{x}) = \mathrm{BN}\left(\mathrm{PHC}\left(\mathrm{ReLU}\left(\mathrm{BN}\left(\mathrm{PHC}\left(\mathbf{x}\right)\right)\right)\right)\right), \quad (5)$$

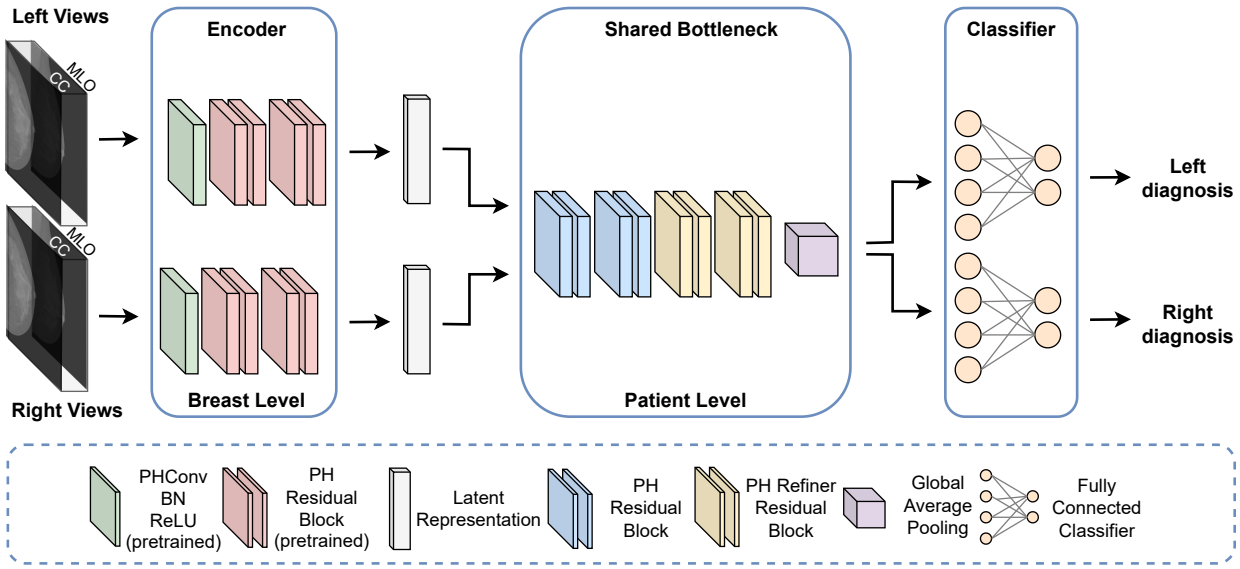

Fig. 3. PHYSBOnet architecture. The model first performs a breast-level analysis by taking as input two pairs of ipsilateral views that are handled by two pretrained PH encoder branches with $n = 2$. The learned latent representations are then processed in a patient-level fashion by four shared PH residual blocks with $n = 2$. Finally, the outputs from the two branches are fed to a separate final fully connected layer after a global average pooling operation.

in which $\mathbf{x}$ can be any multidimensional input with correlated components. It is worth noting that PH models generalize real-valued counterparts and also monodimensional inputs. As a matter of fact, a PH model with $n = 1$ is equivalent to a real-valued model receiving 1D inputs.

B. Parameterized hypercomplex architectures for two-view analysis

The proposed multi-view architecture in case of two views is straightforward: we employ parameterized hypercomplex ResNets (PHResNets) and we fix the hyperparameter $n = 2$. The model is depicted in Fig. 2 together with the training strategy we deploy. The two views of the same breast, i.e., ipsilateral views, are fed to the network as a multidimensional input (channel-wise) enabling PHC layers to exploit their correlations. We adopt ipsilateral views instead of bilateral views, as they are two views of the same breast and thus contain information that helps the model in both the detection process and the classification one. Ultimately, the model produces a binary prediction indicating the presence of either a malignant or benign/normal finding (depending on the dataset). In such manner, the model is able to process the two mammograms as a unique entity, mimicking the diagnostic process of radiologists. Indeed, PHC layers possess the capability of modeling the interactions between the highly correlated views, that would otherwise be lost, and thus leverage information present in ipsilateral views in order to make a more accurate prediction.

C. Parameterized hypercomplex architectures for four-view analysis

The case of four views is not as straightforward as the two views. In fact, even if a mammography exam corresponds to a single patient, in general, two breasts obviously present different findings and consequently have different labels. However, in screening settings, radiologist operate with all four mammographic views as there are additional indicating factors for tumors found by analysing views of different sides, that is bilateral views. Therefore, we propose two novel architectures meticulously designed for processing a whole mammography exam without breaking its nature by exploiting the properties of the hypercomplex domain.

The first model we propose is shown in Fig. 3, named PH Shared Bottleneck network (PHYSBOnet). PHYSBOnet is based on an initial breast-level focus and a consequent patient-level one, through the following components: two encoder branches for each breast side with $n = 2$, a shared bottleneck with the same hyperparameter n to build a broader focus on the patient, and two final classifier layers. Each encoder takes as input two views (CC and MLO) and has the objective of learning a latent representation of the ipsilateral views. The learned latent representation is then processed by a component of the model, the bottleneck, which is shared between the two side branches, as in [19]. Finally, for each side we have a classification layer, which produces the binary prediction relative to the corresponding side. Moreover, we additionally deploy a variant of such network to further exploit hypercomplex algebra by setting the hyperparameter $n = 4$ in the shared bottleneck. Thus, in this case, the latent representations learned by the two encoder branches are merged together in order for the PHC layers to work their magic. Finally, to further show the flexibility of the PH approach, we employ the same architecture with $n = 1$, i.e., in this case the PH model is able to simulate its real-valued counterpart or alternatively to operate with a single view exam. Hence, we demonstrate how the hyperparameter n can be readily set according to users' needs, showing three possible use cases: $n = 1$ for a one-view exam or to simulate the equivalent real model, $n = 2$ for a

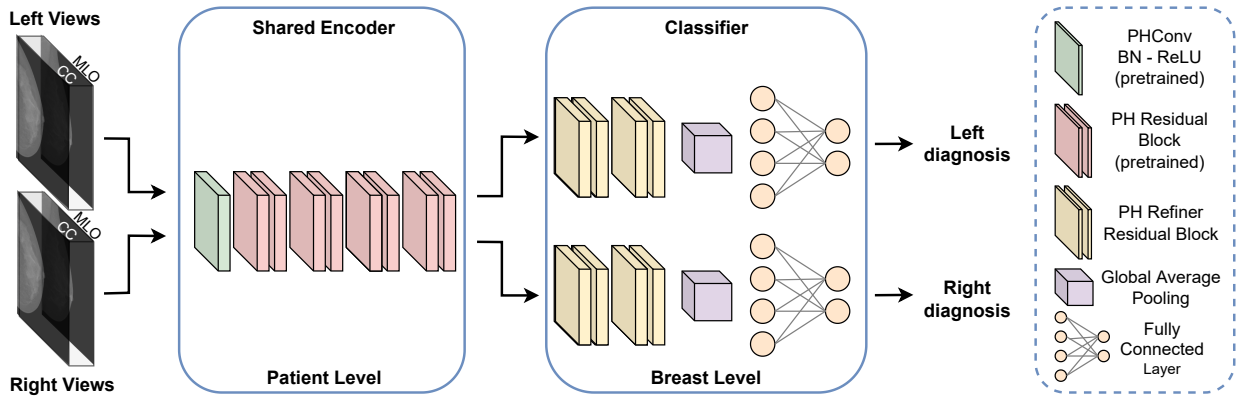

Fig. 4. PHYSEnet architecture. The model comprises an initial deep patient-level framework that takes as input two pairs of ipsilateral views which are processed by a pretrained shared PHResNet18 with $n = 2$ that serves as encoder. Ultimately, the two learned latent representations are fed to the respective classification branch composed of PH refiner residual blocks with global average pooling and the final fully connected layer to perform the breast-level learning.

two-view analysis and $n = 4$ to process the merged latent representations.

The second architecture we proffer, named Parameterized Hypercomplex Shared Encoder network (PHYSEnet), is inspired by the approach of [18] and it is depicted in Fig. 4. It has a broader focus on the patient-level analysis through an entire PHResNet18 with $n = 2$ as the encoder model, whose weights between the two sides (left and right breast) are shared. Then, two final classification branches perform a breast-level analysis. The input of the encoder comprises two mammographic views corresponding to one breast (CC and MLO), i.e., ipsilateral views, whilst its weights are shared between the two sides to jointly analyze the whole information of the patient. With this procedure, the model is able to simultaneously exploit information contained in ipsilateral views, thanks to PHC layers, and, additionally, information from bilateral views thanks to the shared weights of the entire encoder model [19]. In the end, the two learned latent representations are fed to two classification branches consisting of residual blocks and the final fully connected layer responsible for predicting the presence of a malignant tumor or benign/normal finding, for each side.

D. Training procedure

Training a classifier from scratch for this kind of task is very difficult for a number of reasons. For starters, the task itself is much more challenging with respect to the classification of natural images: mammogram images containing benign or malignant tumors present very little differences not noticeable by inexpert eyes and distinguishable only by trained and skilled clinicians. Additionally, in general, and even more so for such a challenging task, neural models require huge volumes of data for training. However, there are only a handful of publicly available datasets for breast cancer and, on top of that, they are characterized by an extremely limited number of examples. Finally, and most importantly, a lesion occupies only a tremendously small portion of the original image, thus making it arduous to be detected by a model [6], especially if the image is reduced in quality, which is often necessary for memory constraints.

TABLE I
DATA DISTRIBUTION FOR SPLITS OF CBIS-DDSM AND FOR INBREAST WHEN CONSIDERING TWO AND FOUR VIEWS. IN THE WHOLE TABLE, TWO VIEWS OF THE SAME BREAST ARE COUNTED AS ONE INSTANCE.

CBIS-DDSM				
	Mass split		**Mass-calc split**	
	Malignant	**Benign**	**Malignant**	**Benign/Normal**
Train	254	255	480	582
Val	59	83	104	149
INbreast				
	Two views		**Four views**	
	Malignant	**Benign**	**Malignant**	**Benign/Normal**
Train	34	89	33	85
Val	14	39	8	22

To overcome these challenges, we deploy an *ad hoc* pretraining strategy divided in two main steps and illustrated in Fig. 2. First, we pretrain the model on patches of mammograms and then we involve the pretrained weights to initialize the network for training on whole images. Indeed, pretraining the network on patches of images that contain either a region of interest (ROI) or a background/normal part of the image, allows the model to learn the features that distinguish malignant from benign tumors [6], [18]. Subsequently, these features are exploited for the training on whole images by initializing the network weights with the patch classifier weights. Such training strategy plays a determining role in boosting the performance of the models as we demonstrate in the experimental Section VI.

V. EXPERIMENTAL SETUP

In this section we describe the experimental setup of our work that comprises the datasets we consider, the metrics employed for evaluation, model architectures details and training hyperparameters.

A. Data

We validate the proposed method with two publicly available datasets of mammography images, whose sample summary is presented in Table I. The first one is the Curated Breast Imaging Subset of DDSM (CBIS-DDSM) [39] and it is an updated and standardized version of the Digital Database for Screening Mammography (DDSM). It contains 2478 scanned film mammography images from 1566 women, decompressed and converted to the standard DICOM format. The dataset provides pixel-level annotations for the regions of interest (ROI) and biopsy-proven pathology labels (benign or malignant). Furthermore, for each lesion the type of abnormality is reported: calcification (753 cases) or mass (891 cases). It is important to notice how the dataset does not contain healthy cases but only positive ones, where for the majority of them a biopsy was requested by the radiologist in order to make a final diagnosis, meaning that the dataset is mainly comprised of the most difficult cases. Additionally, the dataset provides the data divided into splits containing only masses and only calcifications, respectively, in turn split into official training and test sets, characterized by the same level of difficulty. The second dataset employed in this study, INbreast [2], is a database of full-field digital mammography (FFDM) images. It was acquired at the Breast Centre in CHSJ, Porto, between April 2008 and July 2010, and contains 115 mammography exams for a total of 410 images. It includes several types of lesions, among which masses and calcifications. In this case, the data splits are not provided, thus they are manually created by splitting the dataset patient-wise in a stratified fashion, using 20% of the data for validation. Finally, INbreast does not provide pathological confirmation of malignancy but BI-RADS labels. Therefore, in the aim of obtaining binary labels we consider BI-RADS categories 4, 5 and 6 as positive and 1, 2 as negative, whilst ruling out category 3 following the approach of [6].

For both datasets the same preprocessing is applied: the images are resized [6] to 600×500 and are augmented in the aim of improving the generalization of the model. The random transformations applied are the following: rotation between -25 and $+25$ degrees, horizontal and vertical flips. It is relevant to notice that the views of the same exam are transformed in the same way. Finally, when considering four views, the right views are flipped horizontally such that the breast faces the same direction in all the images of a given exam.

Regarding models optimization, the patch classifier training is carried out with the entirety of CBIS-DDSM, while the training on whole images of two-view models is fulfilled with both CBIS-DDSM and INbreast. More in detail, we experiment both with the mass split of CBIS-DDSM as well as the union of the mass and calcification splits. Instead, for the optimization of four-view networks we utilize only INbreast. The reason for the latter being that to consider four views for CBIS-DDSM, the official training/validation splits do not contain enough full-exam cases and creating different splits would result in data leakage between patches and whole images.

B. Evaluation metrics

We adopt AUC (Area Under the ROC Curve) as the main performance metric to evaluate our models, as it is one of the most common metrics employed in medical imaging tasks [6], [18]–[20]. The ROC curve summarizes the trade-off between True Positive Rate (TPR) and False Positive Rate (FPR) for a predictive model using different probability thresholds. The AUC is a value between 0 and 1 representing model capabilities. To further assess network performance, we additionally evaluate our models in terms of classification accuracy.

C. Training details

We train our models minimizing the binary cross-entropy loss using Adam optimizer [42], with a learning rate of 10^{-5}. Regarding two-view models, the batch size is 8 for PHResNet18 and 2 for PHResNet50. Instead, for four-view models the batch size is set to 4 and, to overcome the problem of imbalanced data, we utilize a weighted loss assigning a weight to positive examples equal to the number of benign/normal examples divided by the number of positive examples. We adopt two regularization techniques: weight decay and early stopping. Specifically, we set the L2 penalty multiplier to $10^{-3.3}$ and we early-stop the training when the validation AUC does not improve for 20 epochs, with the maximum number of epochs set to 100. Finally, we select as final model the one corresponding to the epoch with best validation AUC [18].

1) Two-view architectures: We perform the experimental validation with two ResNet backbones: ResNet18 and ResNet50. Herein, we describe the details of the architectures, together with the details of the vanilla real-valued equivalents adopted for comparison. The networks have the same structure as in the original paper [43] with slight variations. In the first convolutional layer the number of channels in input is set to be equal to the hyperparameter $n = 2$ and we omit the max pooling operation since we apply a global average pooling operation, after which we add 4 refiner residual blocks constructed with the bottleneck design [43]. Ultimately, the output of such blocks is fed to the final fully connected layer responsible for classification. The backbone network is initialized with the patch classifier weights, while the refiner blocks and final layer are trained from scratch, following the approach of [6]. In both hypercomplex and real domain, the models take as input two views as if they were a single multi-dimensional (channel-wise) entity.

2) Four-view architectures: Here, we expound the architectural details of the proposed models when considering as input the whole mammogram exam. Also in this case, we use as baselines the respective real-valued counterparts, the shared bottleneck model (SBOnet) and the shared encoder network (SEnet), for which the structural details are the same as in the hypercomplex domain. For both the proposed models we start from a PHResNet18 instead of a PHResNet50 because, from the results of the experiments conducted with two views, we can see that although on CBIS-DDSM our PHResNet50 achieves better results than PHResNet18, on INbreast the performance of the two models is comparable

because it contains less samples and the PHResNet50 may result overparameterized. Additionally, considering that the subset of complete exams further reduces the amount of examples, we believe that the PHResNet18 is the more suitable choice for four-view architectures, as it has less parameters, thus being also more reproducible and accessible.

The first proposed model is the parameterized hypercomplex shared bottleneck model (PHYSBOnet), and the idea is to start from a PHResNet18 and divide its blocks in such a way that the first part of the network serves as an encoder for each side and the remaining blocks compose the shared bottleneck. Thus, the encoders comprise a first 3×3 convolutional layer with a stride of 1, together with batch normalization and ReLU, and the first 4 residual blocks of ResNet18, as shown in Fig. 3. Then, the shared bottleneck is composed of 8 residual blocks and a global average pooling layer, with the first 4 residual blocks being the standard remaining blocks of ResNet18, and the last 4 the refiner residual blocks employed also in the architecture with two views. Finally, the two outputs are fed to the respective fully connected layer, each responsible to produce the prediction related to its side. The second proposed model, parameterized hypercomplex shared encoder model (PHYSEnet), presents as shared encoder network a whole PHResNet18, while the two classifier branches are comprised of the 4 refiner blocks with a global average pooling layer and the final classification layer.

At training time, both for PHYSBOnet and PHYSEnet, the encoder portions of the networks are initialized with the patch classifier weights, while the rest is trained from scratch. Rather, in a second set of experiments conducted only with PHYSEnet, the whole architecture is initialized with the weights of the best whole-image two-view classifier trained on CBIS-DDSM.

VI. Experimental evaluation

In this section, we present an exhaustive evaluation of our method, firstly investigating preliminary experiments without pretraining and then in detail with two multi-view scenarios. We first consider half of a mammography exam, thus consisting of two ipsilateral views, and, secondly, we consider a complete mammography exam, thus comprised of four views. For all experiments, the mean AUC and accuracy over 3 runs is reported together with the standard deviation.

We validate our proposed framework by comparing it against real-valued counterparts and state-of-the-art models for multi-view breast cancer classification. We show how our novel approach outperforms such methods in both multi-view scenarios on two publicly available datasets. Indeed, thanks to the hypercomplex algebra properties that allow to capture and model latent relations present in multi-dimensional inputs, our models are able to exploit the highly correlated nature of the views that make up a mammography exam and thus achieve superior performance.

A. Preliminary experiments

1) Whole-image without pretraining: We conduct preliminary experiments to evaluate the ability of the models to learn from data without any form of pretraining (first part of Table III). We test four different architectures on whole mammograms of CBIS-DDSM considering the mass data split. We employ PHResNet18 and PHResNet50 compared against their real-valued counterparts ResNet18 and ResNet50, respectively. Herein, it is evident that all the models involved are not able to learn properly and struggle to discriminate between images containing a benign lesion from images containing a malignant lesion. In point of fact, distinguishing the malignancy of an abnormality from the whole image only is extremely challenging because the lesion itself occupies a minuscule portion of the entire image. Nevertheless, even with poor performance, it is already evident that the proposed PHResNets are able to capture more information contained in the correlated views and thus reach a higher AUC and accuracy with respect to the real-valued models in all the experiments. Indeed, the best performing model, i.e., the PHResNet50, achieves an AUC of 0.70 and an accuracy of 70.657%. Although the networks reach good results considering the limited number of examples for training, to overcome the challenge of learning from whole mammograms, all further experiments exploit the pretraining strategy described in Subsection IV-D.

TABLE II
Results for patch classifiers on CBIS dataset containing both mass and calc. The parameterized hypercomplex ResNets (PHResNets) outperform real-valued counterparts.

Model	Accuracy (%)
ResNet18	74.942
PHResNet18	**76.825**
ResNet50	75.989
PHResNet50	**77.338**

2) Patch classifier: Preliminary experiments also include the pretraining phase of patch classifiers, which is carried out with the purpose of extracting crucial information concerning the lesions. Aforesaid, detecting and classifying lesions from the entire mammogram is quite challenging as the abnormality is only a small fraction of the image. Furthermore, in resizing the images, the fine-grained detail that characterizes high-resolution mammograms is lost. Therefore, pretraining the models on patches of mammograms is a way to exploit this information that would otherwise be lost and that is crucial especially for the purpose of discriminating between a malignant lesion and a benign one.

Specifically, for each lesion present in the dataset, 20 patches are taken: 10 of background or normal tissue and 10 around the ROI in question. Aiming to utilize this classifier for the training of whole mammograms with two views, we also require two views at the patch-level. The definition of two views for patches is straightforward. For all lesions that are visible in both views of the breast, patches around that lesion are taken for both views, thus each of the 10 patches taken around a ROI from the CC view have a respective patch taken from the respective MLO view. Then, the patch classifier takes as input two-view 224×224 patches of the original mammogram, concatenated along the channel dimension, and classifies them into one of the following five classes: *back-*

TABLE III
RESULTS FOR TWO-VIEW MODELS. FOR CBIS DATASET - MASS SPLIT, WE PERFORM EXPERIMENTS WITH NO PRETRAINING AND WITH PRETRAINING ON PATCHES. FOR CBIS DATASET - MASS AND CALC SPLIT, EXPERIMENTS ARE CONDUCTED WITH MODELS PRETRAINED ON PATCHES. FOR INBREAST DATASET, WE PRETRAIN THE MODELS ON PATCHES AND THEN ON WHOLE CBIS IMAGES (PATCHES + CBIS). OUR METHODS FAR EXCEED REAL-VALUED BASELINES AND STATE-OF-THE-ART MODELS IN EACH TEST WE CONDUCT.

Dataset	Model	Params	Pretraining	AUC	Accuracy (%)
CBIS (mass)	ResNet18	11M	✗	0.646 ± 0.008	64.554 ± 2.846
	ResNet50	16M		0.663 ± 0.011	67.606 ± 1.408
	PHResNet18 (ours)	5M		0.660 ± 0.020	67.371 ± 2.846
	PHResNet50 (ours)	8M		**0.700** ± 0.002	**70.657** ± 1.466
CBIS (mass)	ResNet18	26M	Patches	0.710 ± 0.018	70.892 ± 3.614
	ResNet50	32M		0.724 ± 0.007	73.474 ± 1.076
	Shared ResNet [19]	12M		0.735 ± 0.014	72.769 ± 2.151
	Breast-wise-model [18]	23M		0.705 ± 0.011	69.484 ± 2.151
	DualNet [44]	13M		0.705 ± 0.018	69.719 ± 1.863
	PHResNet18 (ours)	13M		0.737 ± 0.004	74.882 ± 1.466
	PHResNet50 (ours)	16M		**0.739** ± 0.004	**75.352** ± 1.409
CBIS (mass and calc)	ResNet18	26M	Patches	0.659 ± 0.012	66.271 ± 1.271
	ResNet50	32M		0.659 ± 0.013	65.217 ± 3.236
	PHResNet18 (ours)	13M		**0.677** ± 0.005	**68.116** ± 1.388
	PHResNet50 (ours)	16M		0.676 ± 0.014	67.062 ± 0.995
INbreast	ResNet18	26M	Patches + CBIS	0.832 ± 0.008	84.277 ± 2.882
	ResNet50	32M		0.836 ± 0.013	84.906 ± 3.773
	PHResNet18 (ours)	13M		0.837 ± 0.018	**86.163** ± 2.882
	PHResNet50 (ours)	16M		**0.848** ± 0.006	85.535 ± 1.089

ground, benign calcification, malignant calcification, benign mass and malignant mass. Also in this case, we train our proposed models and their real-valued counterparts, of which the results are reported in Table II. We can observe that at the patch-level there is a great gap in performance between our parameterized hypercomplex models and real ones, with PHResNet50 yielding 77.338% accuracy, further exhibiting the ability of PHC layers to model latent relations between multi-dimensional inputs and, as a consequence, improving networks capability of learning and generalizing.

B. Experiments with two views

1) State-of-the-art methods for comparison: We first compare the proposed PHResNets against the respective real-valued baseline models (ResNet18 and ResNet50) and thereafter against three state-of-the-art multi-view architectures [18], [19], [44]. Actually, a variety of multi-view approaches for breast cancer can be found in the literature [20]–[26]. However, as detailed in Section II, all such methods have the same core design, so we establish to use the most recent and advanced works for comparison. On one hand, the multi-view architectures proposed in [18], [19] are designed for mammography exams, with two views and four views, respectively. In these studies, models were trained using a private dataset, thus aiming to compare these approaches with our own, we train the networks with CBIS-DDSM, adopting the same pretraining strategy used for our models. Additionally, original architectures employ as backbone a variation of the standard ResNet50, namely ResNet22, which was designed specifically for the purpose of handling high-resolution images. However, in our case the mammograms are preprocessed and resized as explained in Subsection V-A, thus we straightforwardly use the more proper ResNet18 instead. Finally, since [18] proposes networks designed to handle four views, to compare this approach with ours, we employ the proposed breast-wise-model by trivially considering only *half* of it: instead of having four ResNet columns (one per each view and side) we consider only two columns for the CC and MLO views of one side only. On the other hand, the method proposed in [44] is for processing frontal and lateral chest X-rays, with the same idea as the other studies but, instead of using as backbone model a ResNet, they employ DenseNet121.

2) Results: The results of the experiments we conduct in the two-view scenario are reported in Table III, together with the number of parameters for each model we train: as explained in Section III, PHNNs with $n = 2$ operate with the number of free parameters *halved* with respect to their real-valued counterparts.

Firstly, the advantages of the employed pretraining strategy is clear by comparing the results obtained by our proposed models and their respective real-valued equivalents in the top part of the table with the part corresponding to the pretraining on patches. Most importantly, the center of the table reveals that PHResNets clearly outperform both baseline counterparts implemented in the real domain and all other state-of-the-art methods, with PHResNet50 yielding 0.739 AUC and 75.352% accuracy in the mass split. As well, our approach achieves the best results also in the mass and calc split, with an AUC equal to 0.677 and an accuracy of 68.116%. Herein, the overall performance is reduced with respect to the aforementioned experiments performed with the mass split only. In fact, most works in the literature focus only on mass detection/classification [5], [13], [14], [27], [28]. Nonetheless, thanks to PHC layers, our models are able to grasp essential information that real networks are not able to model and thus outperform them once again. In both the splits, the two most highest AUC and accuracy values are obtained by the two PHResNets, demonstrating the advantages of our approach.

TABLE IV
RESULTS FOR FOUR-VIEW MODELS ON INBREAST. WE PRETRAIN MODELS ON CBIS PATCHES AS WELL AS PATCHES AND THEN WHOLE CBIS IMAGES FOR A FURTHER FINE TUNING. WE ALSO TEST THE PH NETWORK WITH $n = 4$, CONCATENATING (CONCAT) THE LATENT REPRESENTATIONS OF THE VIEWS, AS A SLIGHT MODIFICATION OF THE ARCHITECTURE IN FIG. 3, AND WITH $n = 1$ TO FURTHER DEMONSTRATE THE FLEXIBILITY OF OUR APPROACH.

Model	Params	Pretraining	AUC	Accuracy (%)
SBOnet	26M	Patches	0.769 ± 0.017	70.000 ± 8.819
SBOnet concat	27M		0.750 ± 0.030	71.111 ± 10.183
SEnet	41M		0.799 ± 0.029	<u>82.222</u> ± 1.924
View-wise-model [18]	24M		0.714 ± 0.071	75.556 ± 5.092
Breast-wise-model [18]	24M		0.786 ± 0.014	74.445 ± 6.939
PHYSBOnet $n = 1$ (ours)	26M		0.758 ± 0.052	64.444 ± 7.698
PHYSBOnet $n = 2$ (ours)	13M		0.782 ± 0.022	77.778 ± 9.622
PHYSBOnet $n = 4$ concat (ours)	7M		<u>0.830</u> ± 0.010	78.889 ± 1.924
PHYSEnet (ours)	20M		**0.833** ± 0.044	**83.334** ± 3.335
SEnet	41M	Patches + CBIS	0.805 ± 0.026	81.111 ± 6.939
PHYSEnet (ours)	20M		**0.843** ± 0.016	**86.667** ± 5.774

Surely, hypercomplex algebra allows to process the views as a unique component as it is done by radiologists in real-life scenarios. Additionally, thanks to the characteristic sharing of filter submatrices, the models are able to capture information related to how the views are correlated with each other and exploit such modeled relations for learning. On top of that, it is interesting to notice how our proposed models present the lowest standard deviation on AUC and accuracy over different runs, thus being less sensible to initialization and generally more robust and stable. Ultimately, compared with the baseline real-valued networks, PHC-based models possess half the number of free parameters and, nonetheless, are still able to outperform them. This further proves how the hypercomplex approach truly takes advantage of the correlations found in mammogram views.

Then, we show the advantages of our approach by conducting a further set of experiments with the INbreast dataset. In this case, the models are initialized with the weights of the best whole-image classifier trained with CBIS-DDSM. Despite INbreast contains a very limited number of images, the features learned on CBIS-DDSM are transferred effectively so that the models are able to attain even better performance, i.e., an AUC of 0.848 and 85.535% accuracy obtained by PHResNet50. Also in this case, both PHResNets exceed the respective real-valued baselines, proving once more the ability of PHC layers to draw on the correlations of mammographic views in the aim of achieving more accurate predictions as radiologist do to produce a more accurate diagnosis. Even in this case, the standard deviation over the experiment runs is reduced by our approach, further proving the robustness of the proposed method.

C. Experiments with four views

1) State-of-the-art methods for comparison: We first compare the proposed models against the respective baseline models implemented in the real domain (SBOnet and SEnet), and further against two state-of-the-art approaches designed for breast cancer [18]. Specifically, we employ as comparison their best performing model, view-wise-model, along with the breast-wise-model, since we already consider it for the case of two views. The same considerations mentioned in Subsection VI-B1 apply herein as well: ResNet22 is replaced with ResNet18, which is trained with the same dataset and pretraining procedure as our models to guarantee a fair comparison.

2) Results: In Table IV, we can observe the number of free parameters for each model and the results on the INbreast dataset with four views. Thanks to the properties of parameterized hypercomplex networks, we are able to propose very complex architectures whilst still having a reasonable number of parameters, with the most complex having 41M in the real domain and only 20M in the hypercomplex domain, thus being more appropriate to a small dataset such as the one considered.

The proposed PH models, including both the variants of PHYSBOnet, reach higher accuracy and AUC with respect to the equivalent real-valued baselines, achieving so with half the number of free parameters, thus highlighting the role that correlations between mammographic views play in making the right prediction if exploited properly, as our hypercomplex models are able to do. Importantly, our PHYSEnet largely outperforms all other models, attaining an AUC of 0.833 and an accuracy of 83.334%, proving how the simultaneous exploitation of ipsilateral views, related to PHC layers, and the shared weights between bilateral views of the encoder model results in a more performing classifier. Indeed, PHYSEnet has an initial deep patient-level block that encodes view embeddings by sharing weights between the two sides, providing a more powerful representation with respect to the ones given by the breast-level encoders of PHYSBOnet. Interestingly, our second proposed network, the PHYSBOnet with $n = 2$, is also able to compete with the best models in the state-of-the-art. In particular, it surpasses the view-wise model and matches the same AUC as the breast-wise-model, while achieving a higher accuracy, with almost half the number of parameters, thus being lighter and more efficient. Even more notable is the performance of PHYSBOnet in the variant with $n = 4$, which exceeds state-of-the-art methods both in terms of accuracy and AUC, attaining the second highest AUC 0.830 with $1/3$ of free parameters with respect to the best performing model, i.e., PHYSEnet. Thus, we once more prove the benefits of parameterized hypercomplex networks, showing their ability to be adapted in different frameworks and exams. Thanks to hy-

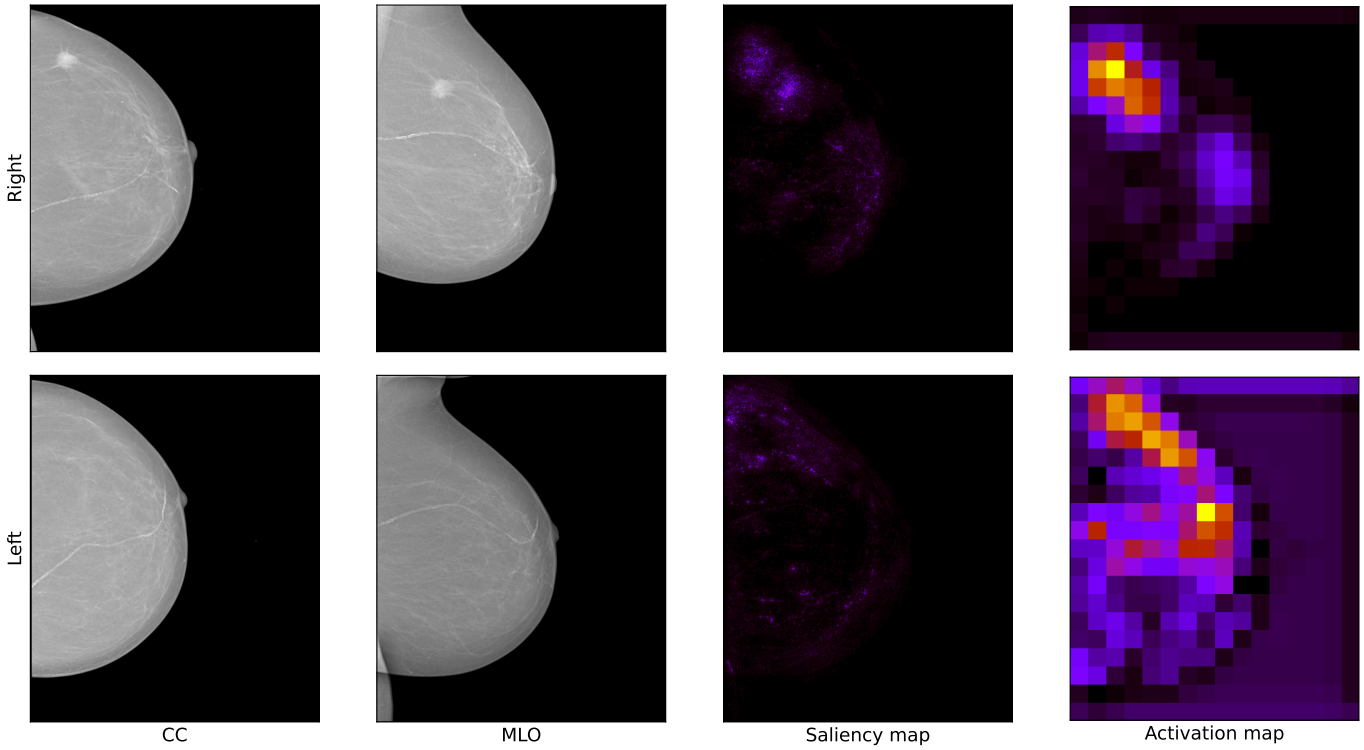

Fig. 5. Visual explanation learned from PHYSEnet on a sample from the INbreast dataset. The first line is the right side that contains a malignant mass, while the second line is the benign left side. Saliency maps are computed at a breast-level, while activation maps leverage patient-level information.

percomplex algebra properties, they are able to surpass state-of-the-art models possessing double and triple the number of parameters. The versatility of our hypercomplex approach is further demonstrated by the experiments conducted with $n = 1$, in which the PH network is able to simulate its real-valued counterpart. In addition, by simply setting $n = 1$, we can easily adapt our approach also to single-view exams, further showing the great portability of this method.

Ultimately, we also test the most performing architecture, that is PHYSEnet, together with its equivalent implemented in the real domain, using as initialization the weights of the best whole-image two-view model trained with CBIS-DDSM. Both models benefit from the learned features on whole images instead of patches, gaining a boost in performance. However, the greater improvement is attained by the PHC version of the network, which once again surpasses the performance of all other experiments with an AUC of 0.843 and accuracy of 86.667%.

3) Visualizing multi-view learning: To further assess the crucial role of multi-view processing in breast cancer classification, here we provide a visual explanation of PHYSEnet results.

We consider a patient from INbreast with a benign finding in the left breast and a malignant mass in the right breast. We compute the saliency maps for each side separately, using gradients of the class-dependent output over the input. With this procedure we are able to visualize which area of the two-view input is most influential for the model decision. Additionally, we take the output at the end of the PH refiner blocks and we compute the average of each activation map in order to obtain a visual representation of what the model has learned. Figure 5 displays the sample views and the computed maps. Firstly, it is worth noting that the model provides one saliency and one activation map for the two views, therefore compacting the double information in a single, more impactful representation. Regarding the saliency maps, for the right side we can observe the highlighted pixels that indeed correspond to the malignant mass visible in the CC and MLO views. This means that this region has a great impact for the final prediction of the right breast, which in fact is accurately classified by PHYSEnet. Meanwhile, the left-side map is darker and no pixel region is particularly accentuated since there are no malignant findings in the left breast, and surely PHYSEnet produces the correct prediction. While saliency maps provide an explanation for the model decision, activation maps provide a visual depiction of the learned latent representation. Also in this case, the pixels corresponding to the mass present in the right breast are activated, while on the left side the ones corresponding to the benign calcification.

D. Ablation Study

The most relevant ablation studies for the proposed architectures are actually already included in previous experiments reported in Section VI. As an example, an interesting ablation study is to remove significant parts from the different networks. In that sense, the most crucial network modules are represented by the PHC layers. Removing the potential of parameterized hypercomplex operations, the model turns out

TABLE V
MODELS USE CASES SUMMARY. THE TICKS ✓ MEAN THAT THE MODEL CAN BE ADOPTED IN THE SPECIFIC CASE IF OUR PRETRAINED WEIGHTS ARE EMPLOYED. THIS IS BECAUSE WITHOUT ROI ANNOTATIONS IT IS NOT POSSIBLE TO TRAIN A PATCH CLASSIFIER, THUS THE PRETRAINING STAGE CANNOT BE PERFORMED.

Use Cases	PHResNet18	PHResNet50	PHYSBOnet	PHYSEnet
Single View	✓	✓	✗	✗
Two Views	✓	✓	✗	✗
Four Views	✗	✗	✓	✓
ROI available	✓	✓	✓	✓
ROI not available	✓	✓	✓	✓
Light memory	✓	✗	✓	✗
Small size dataset	✓	✓	✓	✓
Other problems	✓	✓	✓	✓

to be equivalent to its real-valued counterpart. However, we have shown in Section VI that the real-valued counterpart is greatly outperformed by the respective hypercomplex network for each proposed model. Those results prove how the gain in performance is owed to the capabilities to exploit multi-view correlations that are endowed to our methods thanks to the introduction of hypercomplex algebra in a convolutional layer.

In Section VI, we have additionally tested the impact of the deployed pretraining strategy: randomly initializing networks weights leads to poor performance, because tumor masses or calcifications are arduous to detect, and even more to classify, from the original mammogram, as they occupy a tiny portion of the image. On the contrary, thanks to the pretraining recipe we adopt, the fine-grained detail characteristic of high-resolution mammograms is leveraged as a result of the patch classifier, and effectively transferred for learning with entire mammogram images.

VII. HOW DO I CHOOSE THE BEST MODEL FOR MY DATA?

In this section, we aim at answering the reader question *Which model do I employ on my data and for my problem?*

We propose several different networks, often specific for the kind of problem. Therefore, here we analyze the framework, the structure of available data and the cases in which each model fits at best. We consider multiple scenarios, including datasets with 1, 2 or 4 views, whether ROI annotations are available or not, the memory constraints, the size of the dataset and if proposed models are easily exportable to other kind of medical imaging problems. Table V reports the applicability of each network we propose in the above-mentioned scenarios. Importantly, our approach is easily scalable and exportable to other multi-view medical problems such as chest X-ray disease classification and multimodal brain tumor segmentation. Moreover, while we prove the crucial role of pretraining for breast cancer classification tasks, especially when scarce data is available, this is only applicable in case of datasets with provided ROI annotations. Nevertheless, our methods can be applied even in dataset where ROI observations are missing, since we provide pretrained models and weights[1] to overcome this limitation. We believe that Table V may help the reader clarifying which model better fits in specific cases and may increase the usability of our approach in future researches and in different medical fields.

[1]Weights are freely available at: https://github.com/ispamm/PHBreast

VIII. CONCLUSIONS

In this paper, we have introduced an innovative approach for breast cancer classification that handles multi-view mammograms as a radiologist does, thus leveraging information contained in ipsilateral views as well as bilateral. Thanks to hypercomplex algebra properties, neural models are endowed with the capability of capturing and truly exploiting correlations between views and thus outperform state-of-the-art models. At the same time, our approach is flexible to be adopted in any existing neural network as well as for different imaging exams with an arbitrary number of views. On a thorough evaluation on publicly available benchmark datasets, we have showed the improved AUC and accuracy values of the proposed approach with respect to state-of-the-art methods, together with a greater robustness of such results. We believe that our approach is a real breakthrough for this research field and that it may pave the way to novel methods capable of processing medical imaging exams with techniques closer to radiologists and to human understanding.

REFERENCES

[1] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, "Cancer statistics, 2022," *CA: A Cancer Journal for Clinicians*, vol. 72, no. 1, pp. 7–33, 2022.

[2] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, "INbreast: Toward a full-field digital mammographic database," *Academic Radiology*, vol. 19, no. 2, pp. 236–248, 2012.

[3] S. Misra, N. L. Solomon, F. L. Moffat, and L. G. Koniaris, "Screening criteria for breast cancer," *Adv. Surg.*, vol. 44, pp. 87–100, 2010.

[4] D. Gur, G. S. Abrams, D. M. Chough, M. A. Ganott, C. M. Hakim, R. L. Perrin, G. Y. Rathfon, J. H. Sumkin, M. L. Zuley, and A. I. Bandos, "Digital breast tomosynthesis: Observer performance study," *American Journal of Roentgenology*, vol. 193, no. 2, pp. 586–591, 2009.

[5] Y. Liu, F. Zhang, C. Chen, S. Wang, Y. Wang, and Y. Yu, "Act like a radiologist: Towards reliable multi-view correspondence reasoning for mammogram mass detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, no. 01, pp. 1–1, 2021.

[6] L. Shen, L. Margolies, J. Rothstein, E. Fluder, R. McBride, and W. Sieh, "Deep learning to improve breast cancer detection on screening mammography," *Sci. Rep.*, vol. 9, 2019.

[7] N. Wu, Z. Huang, Y. Shen, J. Park, J. Phang, T. Makino, S. Kim, K. Cho, L. Heacock, L. Moy, and K. J. Geras, "Reducing false-positive biopsies using deep neural networks that utilize both local and global image context of screening mammograms," *Journal of Digital Imaging*, vol. 34, pp. 1414 – 1423, 2021.

[8] G. Murtaza, L. Shuib, A. Wahid, G. Mujtaba, H. Nweke, M. Al-Garadi, F. Zulfiqar, G. Raza, and N. Azmi, "Deep learning-based breast cancer classification through medical imaging modalities: state of the art and research challenges," *Artificial Intelligence Review*, vol. 53, 03 2020.

[9] D. Abdelhafiz, C. Yang, R. Ammar, and S. Nabavi, "Deep convolutional neural networks for mammography: advances, challenges and applications," *BMC Bioinformatics*, vol. 20, 2019.

[10] S. S. Aboutalib, A. A. Mohamed, W. A. Berg, M. L. Zuley, J. H. Sumkin, and S. Wu, "Deep Learning to Distinguish Recalled but Benign Mammography Images in Breast Cancer Screening," *Clinical Cancer Research*, vol. 24, no. 23, pp. 5902–5909, 2018.

[11] I. Sechopoulos and M. R. Mann, "Stand-alone artificial intelligence - the future of breast cancer screening?" *The Breast*, vol. 49, pp. 254–260, 2020.

[12] M. A. Al-antari, M. A. Al-masni, and T.-S. Kim, "Deep learning computer-aided diagnosis for breast lesion in digital mammogram," *Adv. Exp. Med. Biol.*, vol. 1213, pp. 59–72, 2020.

[13] R. Agarwal, O. Diaz, X. Lladó, M. H. Yap, and R. Martí, "Automatic mass detection in mammograms using deep convolutional neural networks," *Journal of Medical Imaging*, vol. 6, no. 3, pp. 1–9, 2019.

[14] J. Niu, H. Li, C. Zhang, and D. Li, "Multi-scale attention-based convolutional neural network for classification of breast masses in mammograms," *Medical Physics*, vol. 48, no. 7, pp. 3878–3892, 2021.

[15] G. Zhao, Q. Feng, C. Chen, Z. Zhou, and Y. Yu, "Diagnose like a radiologist: Hybrid neuro-probabilistic reasoning for attribute-based medical image diagnosis," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2021.

[16] H. Pinckaers, B. van Ginneken, and G. Litjens, "Streaming convolutional neural networks for end-to-end learning with multi-megapixel images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 3, pp. 1581–1590, 2022.

[17] C. Lian, M. Liu, J. Zhang, and D. Shen, "Hierarchical fully convolutional network for joint atrophy localization and alzheimer's disease diagnosis using structural MRI," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 4, pp. 880–893, 2020.

[18] N. Wu, J. Phang, J. Park, Y. Shen, Z. Huang, M. Zorin, S. Jastrzebski, T. Fevry, J. Katsnelson, E. Kim, S. Wolfson, U. Parikh, S. Gaddam, L. Lin, K. Ho, J. Weinstein, B. Reig, Y. Gao, H. Toth, K. Pysarenko, A. Lewin, J. Lee, K. Airola, E. Mema, S. Chung, E. Hwang, N. Samreen, S. Kim, L. Heacock, L. Moy, K. Cho, and K. Geras, "Deep neural networks improve radiologists' performance in breast cancer screening," *IEEE Trans. Med. Imaging*, vol. 39(4), 2020.

[19] N. Wu, S. Jastrzębski, J. Park, L. Moy, K. Cho, and K. Geras, "Improving the ability of deep neural networks to use information from multiple views in breast cancer screening," in *Proc. of the Third Conf. on Med. Imaging with Deep Learning*, vol. 121. PMLR, 2020, pp. 827–842.

[20] C. Zhang, J. Zhao, J. Niu, and D. Li, "New convolutional neural network model for screening and diagnosis of mammograms," *PLoS ONE*, vol. 15(8), 2020.

[21] T. Kyono, F. J. Gilbert, and M. van der Schaar, "Mammo: A deep learning solution for facilitating radiologist-machine collaboration in breast cancer diagnosis," *arXiv preprint: arXiv:1811.02661*, 2018.

[22] ——, "Triage of 2d mammographic images using multi-view multi-task convolutional neural networks," *ACM Trans. Comput. Healthcare*, vol. 2, no. 3, 2021.

[23] ——, "Multi-view multi-task learning for improving autonomous mammogram diagnosis," in *Proceedings of the 4th Machine Learning for Healthcare Conference*, ser. Proceedings of Machine Learning Research, vol. 106. PMLR, 2019, pp. 571–591.

[24] L. Sun, J. Wang, Z. Hu, Y. Xu, and Z. Cui, "Multi-view convolutional neural networks for mammographic image classification," *IEEE Access*, vol. 7, pp. 126 273–126 282, 2019.

[25] J. Song, Y. Zheng, M. Z. Ullah, J. Wang, Y. Jiang, C. Xu, Z. Zou, and G. Ding, "Multiview multimodal network for breast cancer diagnosis in contrast-enhanced spectral mammography images," *International Journal of Computer Assisted Radiology and Surgery*, vol. 16, pp. 979 – 988, 2021.

[26] G. Carneiro, J. C. Nascimento, and A. P. Bradley, "Unregistered multiview mammogram analysis with pre-trained deep learning models," in *Int. Conf. on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2015.

[27] J. Ma, X. Li, H. L., Ruixuan, W. Ruixuan, B. Menze, and W. Zheng, "Cross-View relation networks for mammogram mass detection," in *2020 25th International Conference on Pattern Recognition (ICPR)*, 2021, pp. 8632–8638.

[28] Z. Yang, Z. Cao, Y. Zhang, Y. Tang, X. Lin, R. Ouyang, M. Wu, M. Han, J. Xiao, L. Huang, S. Wu, P. Chang, and J. Ma, "MommiNet-v2: Mammographic multi-view mass identification networks," *Medical Image Analysis*, vol. 73, p. 102204, 2021.

[29] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, "Multimodal machine learning: A survey and taxonomy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 2, pp. 423–443, 2019.

[30] W. Wang, D. Tran, and M. Feiszli, "What makes training multi-modal classification networks hard?" in *2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 12 692–12 702.

[31] C. Gaudet and A. Maida, "Deep quaternion networks," in *IEEE Int. Joint Conf. on Neural Netw. (IJCNN)*, Rio de Janeiro, Brazil, Jul. 2018.

[32] T. Parcollet, M. Morchid, and G. Linarès, "A survey of quaternion neural networks," *Artif. Intell. Rev.*, Aug. 2019.

[33] T. Parcollet, M. Ravanelli, M. Morchid, G. Linarès, C. Trabelsi, R. De Mori, and Y. Bengio, "Quaternion recurrent neural networks," in *Int. Conf. on Learning Representations (ICLR)*, New Orleans, LA, May 2019, pp. 1–19.

[34] D. Comminiello, M. Lella, S. Scardapane, and A. Uncini, "Quaternion convolutional neural networks for detection and localization of 3D sound events," in *IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, Brighton, UK, May 2019, pp. 8533–8537.

[35] M. Ricciardi Celsi, S. Scardapane, and D. Comminiello, "Quaternion neural networks for 3D sound source localization in reverberant environments," in *IEEE Int. Workshop on Machine Learning for Signal Process.*, Espoo, Finland, Sep. 2020, pp. 1–6.

[36] C. Brignone, G. Mancini, E. Grassucci, A. Uncini, and D. Comminiello, "Efficient sound event localization and detection in the quaternion domain," in *IEEE Trans. on Circuits and Systems II: Express Brief (Early Access)*, 2022, pp. 1–5.

[37] A. Zhang, Y. Tay, S. Zhang, A. Chan, A. T. Luu, S. C. Hui, and J. Fu, "Beyond fully-connected layers with quaternions: Parameterization of hypercomplex multiplications with $1/n$ parameters," *Int. Conf. on Machine Learning (ICML)*, 2021.

[38] E. Grassucci, A. Zhang, and D. Comminiello, "PHNNs: Lightweight neural networks via parameterized hypercomplex convolutions," *arXiv preprint: arXiv:2110.04176*, 2021.

[39] R. Lee, F. Gimenez, A. Hoogi, M. Kanae, M. Gorovoy, and D. Rubin, "A curated mammography data set for use in computer-aided detection and diagnosis research," *Sci. Data*, vol. 4, 2017.

[40] M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio, "Transfusion: Understanding transfer learning for medical imaging," in *Proceedings of the 33rd International Conference on Neural Information Processing Systems (NIPS)*, 2019, p. 3347–3357.

[41] K. Geras, S. Wolfson, S. Kim, L. Moy, and K. Cho, "High-resolution breast cancer screening with multi-view deep convolutional neural networks," *arXiv preprint: arXiv:1703.07047*, 2017.

[42] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Y. Bengio and Y. LeCun, Eds., 2015. [Online]. Available: http://arXiv.org/abs/1412.6980

[43] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[44] J. Rubin, D. Sanghavi, C. Zhao, K. Lee, A. Qadir, and M. Xu-Wilson, "Large scale automated reading of frontal and lateral chest x-rays using dual convolutional neural networks," *arXiv preprint: arXiv:1804.07839*, 2018.