IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 9 August 2023, accepted 20 August 2023, date of publication 23 August 2023, date of current version 30 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3308075

RESEARCH ARTICLE

Enhanced Parkinson's Disease Diagnosis Through Convolutional Neural Network Models Applied to SPECT DaTSCAN Images

HAJER KHACHNAOUI[1], BELKACEM CHIKHAOUI[2], NAWRES KHLIFA[1], AND ROSTOM MABROUK[3]

[1]Laboratoire de Biophysique et Technologies Médicales, Institut Superieur des Technologies Medicales de Tunis, Université de Tunis El Manar, Tunis 1006, Tunisia
[2]LICEF Research Center, Université TÉLUQ, Montreal, QC H2S 3L5, Canada
[3]Department of Computer Sciences, Bishop's University, Sherbrooke, QC J1M 1Z7, Canada

Corresponding author: Hajer Khachnaoui (hajerkhachnaoui69@gmail.com)

ABSTRACT Convolutional Neural Networks (CNNs) are highly regarded in Deep Learning (DL) and have shown promising results in medical image analysis, making them a leading model for Computer-Aided Diagnosis (CAD) systems. Their efficacy extends to the diagnosis of neurological disorders, including Parkinson's Disease (PD), which is typically identified through Single Photon Emission Computed Tomography (SPECT) scans. However, relying solely on visual inspection of SPECT images during medical examinations can introduce inaccuracies due to subjective factors. We propose a CAD system for automatic PD diagnosis using pre-trained CNN models, Transfer Learning (TL) technique, and the Bilinear Pooling method to address this issue. The study employs several CNN architectures, specifically Efficient-Net B0, and Mobile-Net V2 models, and a custom CNN architecture. These pre-trained architectures were originally trained on ImageNet and adapted to the current task using the TL technique. These architectures are leveraged with a bilinear pooling form, resulting in three Bilinear CNN (BCNN) models. These models are applied to pre-processed SPECT image data of PD patients and Healthy Controls (HC), categorized into three distinct datasets. The proposed method is evaluated on a total of 2720 SPECT images (1360 PD and 1360 HC subjects) obtained from the Parkinson's Progression Marker Initiative (PPMI) dataset. The findings show that the BCNN EfficientNet-B0-MobileNet-V2 model achieved the highest accuracy score of 98.47%, surpassing other developed CNN models and outperforming existing methods. In conclusion, the proposed CAD system provides an efficient diagnostic tool to assist physicians in making accurate PD diagnoses, independent of subjective factors.

INDEX TERMS Bilinear pooling, convolutional neural networks, deep learning, Parkinson's disease, SPECT.

I. INTRODUCTION

Parkinson's Disease (PD) is a chronic neurodegenerative condition that affects millions of people worldwide. It is characterized by the progressive loss of dopaminergic neurons in the midbrain's substantia nigra, leading to a reduction of dopamine in the striatum and destruction of dopamine transporters (DaT) [1]. This loss of dopaminergic neurons can result in various motor and non-motor symptoms such as tremors, postural imbalances, slowed movement, sleep disorders, and depression [2]. Currently, various neuroimaging modalities, such as Magnetic Resonance Imaging (MRI), Position Emission Tomography (PET), and Single Photon Emission Computed Tomography (SPECT), are used to diagnose PD [1]. SPECT DaTSCAN functional imaging, which utilizes [123]-Ioflupane (also known as FP-CIT) as a radioligand to bind to dopamine transporters in the striatum region, is the most commonly used modality for PD diagnosis due to

The associate editor coordinating the review of this manuscript and approving it for publication was Humaira Nisar.

its high sensitivity and specificity [1]. Scans from PD subjects show a smaller and rounded striatum region because of the deficiency of dopaminergic neurons. On the contrary, healthy control subjects' scans preserve their highly illuminated C-shape. Traditionally, SPECT images are interpreted manually where the diagnosis result may be subject to the risk of human error. Thus, Semi-quantitative analysis of striatal FP-CIT uptake, measured by the specific binding ratio (SBR), is commonly used to supplement visual interpretation [3]. The SBR value gives neutral measures of dopaminergic function. Nonetheless, this semi-quantitative measure is affected by variability of SPECT image characteristics caused by differences in acquisition and reconstruction protocols and by scan-specific variations such as head motion and varying radius of rotation of the camera heads. In addition, the striatal uptake shape information and particular pattern are not reflected in the SBR results [3]. To overcome this, several Computer-Aided Diagnosis (CAD) systems based on Deep Learning (DL) approaches have been developed for automatic PD diagnosis in SPECT DaTSCAN images. DL is the key instrument in Artificial Intelligence applications [4]. It has shown promising results in various domains, including image and signal processing, facial and voice recognition, and objects detection [5], [6], [7], [8]. In medical imaging, the better availability of reliable and public datasets with ease of access in online platforms, and the advancement in computational processing power and storage has helped to increase the use of DL methods, especially Convolutional Neural Networks (CNN) [5]. CNN models have shown tremendous potential in the analysis and interpretation of medical images for diagnosing various diseases, including PD [9]. These models offer a significant benefit in that they solve tasks end-to-end, with the network itself learning the salient features of the data rather than requiring pre-engineered human knowledge [10]. Furthermore, they can represent very complex functions and learn features at multiple levels of abstraction, for example, edges at the lower layers, forms at the middle layer, and task-specific features at the deeper layers [10]. This property, make the CNN models suitable for Transfer Learning (TL) techniques [5]. TL involves transferring pre-trained CNN architectures' weights to perform new models, effectively reducing the time and effort required to design a new model and optimize its parameters [5]. Pre-trained models, including VGG, Inception, ResNet, and MobileNet, have become increasingly popular in medical imaging, particularly in the diagnosis of diseases using SPECT images, such as PD [10], [5].

Therefore, the aim of this study is to develop a CAD system based on DL approach to provide a unified measurement standard for PD diagnosis. Indeed, different CNN models were developed based on eights hottest striatal slices in different form selected from the original SPECT, as the SBR value is computed from these slices. This automated approach aims to assist physicians in efficiently and accurately classifying subjects as having PD or not in their SPECT scans, providing a reliable second opinion diagnostic tool.

The remaining parts of the paper are assembled as follows: In section two, we review previous works related to our study. Section three describes the dataset characteristics, including image preprocessing, and provides a detailed explanation of the proposed approach. We introduce the concept of pre-trained CNN models, TL technique, and the bilinear pooling method. Section four presents the experimental results, while section five provides a discussion of these results. Finally, section six presents the conclusions and suggests possibilities for future work.

II. LITERATURE REVIEW

The pace of advancement in intelligent and automated diagnosis tools has accelerated rapidly [1]. While Hand-Crafted Machine Learning (ML) techniques involving feature extraction have been acknowledged as a valuable means of enhancing the diagnosis of numerous medical conditions, they necessitate expertise and domain specialists [9]. On the other hand, important features are generated automatically through end-to-end DL methods. Therefore, DL models, especially CNN, the most prevalent and extensively employed image classification technique, are widely embraced in medical imaging diagnosis. Thus, various researchers [11], [12], [13], [14], [15] have developed studies employing DL approaches to assist physicians and reduce subjective decisions in PD diagnosis.

For instance, the study of Shih-Yen et al. [11] employs six pre-trained CNN models, namely AlexNet, GoogleNet, VGG19, ResNet 50, ResNet101, and DenseNet, to classify multiple stages of PD based on 99mTc-TRODAT-1 SPECT images. A total of 1010 images were analyzed from 202 cases, with each case contributing five slices for analysis, including the maximum active slice of the striatum along with its previous and subsequent slices. The cases were classified into HC, early, middle, and late stages based on the Hoehn and Yahr Scale standards (HYS). Additionally, another dataset with six classes, including HC and the five HYS stages, was created. The highest accuracy score of 85.5% was obtained with DenseNet in the four-class database. These results demonstrated the potential of the proposed method to assist clinicians in the early diagnosis and treatment of PD. However, the study's limitations include a small sample size and an imbalanced dataset.

Additionally, in 2022, the research work of Magesh et al. [12] presents a TL technique of the pre-trained VGG16 model to accurately classify DaTSCANs as either PD or non-PD. The pre-trained model was modified by freezing all layers except the last two, and by adding two dropout layers and a dense layer with a sigmoid activation function at the end of the model. A total of 642 DaTSCAN SPECT images were used, with 430 images labelled as PD and 212 images labelled as non-PD, and the slice 41 was selected for analysis from each subject. These images were sourced from the PPMI dataset. The model's design aims to provide clear and understandable insights into its decision-making process, improving diagnostic accuracy and facilitating treatment decisions.

The proposed method achieves an accuracy rate of 95.2%. To explain the model's predictions, the study uses Local Interpretable Model-Agnostic Explainer (LIME) methods that generate visual indicators in the form of super-pixels on the DaTSCANs. By prioritizing interpretability, this model has the potential to aid medical workers in the early detection of PD. However, the dataset size was limited and the data was imbalanced.

In the same year, Jigna et al. [13] presented a DL method that aimed to classify PD patients based on dopamine levels using SPECT imaging data and a CNN model. To achieve this, the PPMI dataset was pre-processed and balanced with data amplification. The CNN model contained 14 layers and classified input SPECT images into four categories: PD, HC, Scans Without Evidence of Dopaminergic Deficit (SWEDD), and GenReg PD (Persons suffering from PD because of their genetics). This model was built with 1761 SPECT images and surpassed other classification models, achieving an impressive accuracy of 88.9%. However, the study's generalizability is limited due to its relatively small dataset and the lack of comparison with other commonly used methods for PD classification.

Furthermore, an ensemble of DL models was proposed by Ankit et al. [14] to predict PD using DaTSCAN images. Four pre-trained CNN models including VGG16, ResNet50, Inception-V3, and Xception, were initially used to classify PD, followed by a fuzzy fusion logic-based ensemble approach to improve the overall classification performance. The proposed model was evaluated on 645 images divided into 432 PD and 213 non-PD provided by the PPMI database, where it achieved an accuracy of 98.45%, all of which are higher than individual model accuracies.

Recently, the research study of Juan et al. [15] proposed a Bayesian DL-based multi-level ensemble classification system for achieving high performance in multiclass classification while also providing the uncertainty of each classification decision. The system combines information from various architectures based on the uncertainty of their predictions to achieve high accuracy in different real scenarios. To accomplish this, several CNNs with identical architecture but different kernel sizes are trained, and their predictions are combined based on their reliability. A decision tree is employed to divide the multiclass problem into binary sub-problems. The proposed system differentiates between various pulmonary pathologies with 98.19% accuracy and diagnoses PD from HC using 269 DaTSCAN images with 95.31% accuracy, without requiring extensive pre-processing. This system's ability to provide additional information from the uncertainty of predictions makes it useful for a wide range of medical scenarios.

From these studies [11], [12] [13], [14], [15], it is found that the DL approaches achieve significant results for PD diagnosis from SPECT images. However, these results can be improved further. Thus, the development of a novel CNN model for PD diagnosis is the main motivation for this work.

III. MATERIALS AND METHODS

A. CONTRIBUTIONS

Motivated by the literature studies that identify PD patients, this research work provides a binary classification model to distinguish PD patients from HC subjects in SPECT imaging based on DL and particularly CNN. The diagram of the proposed approach is shown in Figure 1. The main contributions and objectives of this work are summarized as follows:

a) A diagnostic tool based on DL methods is proposed to improve the performance of PD diagnosis.

b) The PPMI dataset was used in this study, as it is a large database that includes healthy and unhealthy subjects from various locations, which adds diversity to the dataset and makes the proposed approach robust. The proposed method used eight slices selected from the DaTSCAN SPECT imaging, specifically the most active slices of the striatum (putamen and caudate) from 37 to 44, which were employed to calculate the SBR values. These slices were pre-processed in three distinct formats, each with a resolution of 224×224. The first format involved separating the slices into eight individual images. In the second format, the slices were combined into a single image. Lastly, the third format entailed merging the slices into an image.

c) Several CNN architectures were developed, including EfficientNet-B0 and MobileNet-V2 models, as well as three Bilinear CNN (BCNN) models (EfficientNet-B0-EfficientNet-B0, MobileNet-V2-MobileNet-V2, and EfficientNet-B0-MobileNet-V2), along with a custom CNN architecture. The EfficientNet-B0 CNN model, which was pre-trained on the ImageNet dataset and is considered cutting-edge in the field, was selected for use in this study. Additionally, the pre-trained MobileNet-V2 architecture is used as it is compatible in input and output with EfficientNet-B0 for the BCNN model. The custom CNN model is trained on a series of eight slices combined into one image, while the other models are trained using the two other forms of data. In the first form, eight slices selected from the SPECT image are associated in one image. The second form is that each CNN model is trained eight times with the eight slices (each slice is used as an input), and the predictions of the eight CNN models are associated and classified using the Adaboost algorithm.

d) This study presents made original contributions to the field of PD diagnosis through the use of innovative techniques. Specifically, the study employed BCNN for PD diagnosis, which represents a novel application of this technique. BCNN allows for the extraction of spatial information from images and can capture interactions between different Regions Of Interest (ROI), making it a promising tool for PD diagnosis. Furthermore, the study combined pre-trained EfficientNet-B0 and MobileNet-V2 architectures in the BCNN for classification, which, to the best of our knowledge, has not been explored before in the literature. These contributions highlight the potential of BCNN and the feasibility of combining pre-trained CNN models for improved

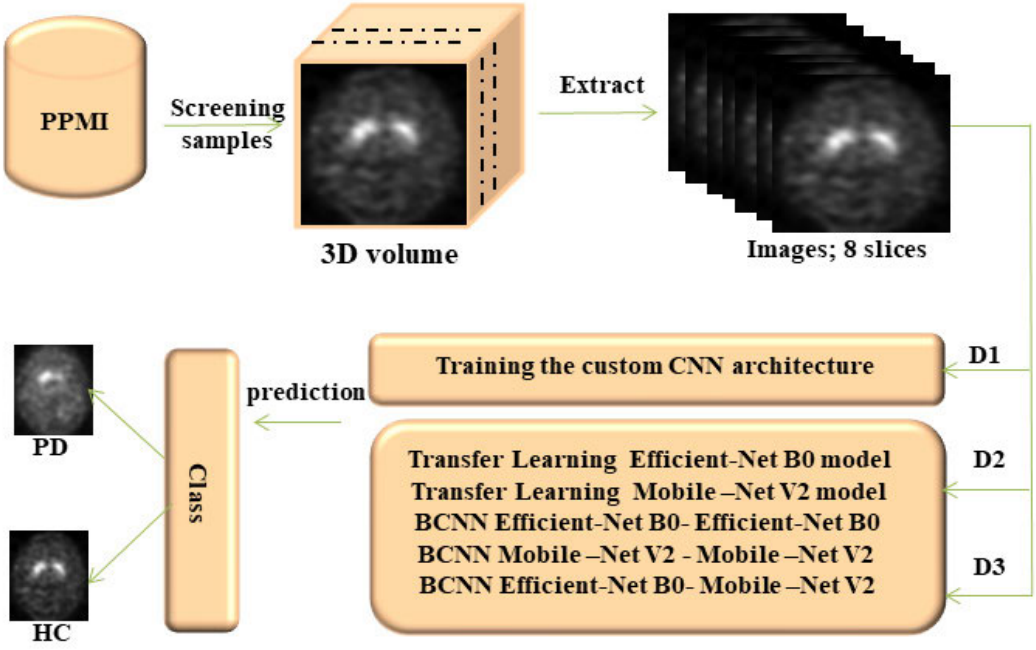


FIGURE 1. The proposed method workflow for PD diagnosis using three image datasets: D1, D2, and D3 with a resolution of 224 × 224. D1 combines eight slice images into one image, D2 associate the same eight slices into a single image, and D3 presents each slice as a separate image.

classification of PD, paving the way for future research in this area.

e) The proposed approach underwent an evaluation process where several performance metrics such as precision, recall, accuracy, F1 score, and ROC-AUC were measured.

B. PREPARATION OF DATASETS

1) DATASET IMAGE PREPROCESSING

This research work conducted its experiments on data obtained from Parkinson's Progression Marker Initiative (PPMI) dataset. The PPMI is a multimodal, prolonged study of radiomic feature observations, neuroimaging, and biological markers in both PD patients and healthy controls (HC) [16]. This extensive searchable archive was built through the collaborative efforts of scientists, researchers, sponsors, and study populations from various industries to facilitate PD research and therapies by identifying progression biomarkers. The imaging data used in this study comprised 2720 SPECT DaTSCAN images obtained from 340 cases. These images are classified into two equal classes namely PD and HC. Figure 2 illustrates samples of SPECT DaTSCAN images of PD patients and HC subjects.

2) IMAGE PREPROCESSING

The raw SPECT DaTSCAN images taken at PPMI had undergone some preprocessing before they were added to the online database [16]. Initially, attenuation correction was carried out using phantoms procured from the same time the subject was imaged. Additionally, the images were reconstructed and spatially normalized to eliminate any differences in shape or size against several unique subjects. This alignment was done by the Montreal Neurological Institute (MNI) accepted, the standard coordinate system for medical imaging. The resulting SPECT DaTSCAN images were presented as a 3D volume space with dimensions of $91 \times 109 \times 91$(representing 91 slices with each slice being 91×109 pixels) in DICOM and NIFTI format. We used the eight slices from 37 to 44 where the striatum structure volume is delineated to define our ROIs for the study development as they have been used to calculate the SBR values.

Using the eight slices, data were stratified into three different datasets, namely D1, D2, and D3, each containing images with a resolution of 224 × 224. The first dataset, D1, consists of images that are composed of a series of eight slices. On the other hand, the second dataset, D2, contains images that combine the eight slices from D1 into one image. Finally, the third dataset, D3, contains the same eight slices as D1, but each slice is presented as a separate image. Examples from each dataset are presented in Figure 3.

In order to address the limitation of a relatively small dataset, which comprised 2720 SPECT DaTSCAN images obtained from 340 cases, with each case providing 8 images, we implemented data augmentation technique. By applying various transformations, such as horizontal and vertical flipping, translation, noise addition, zooming, and rotation at different degrees, we generated nine additional images for each original sample. As a result of this augmentation process, we generated nine additional images for each original sample. This approach significantly enriched and expanded the dataset, resulting in a total of 27200 images, with 24480 new

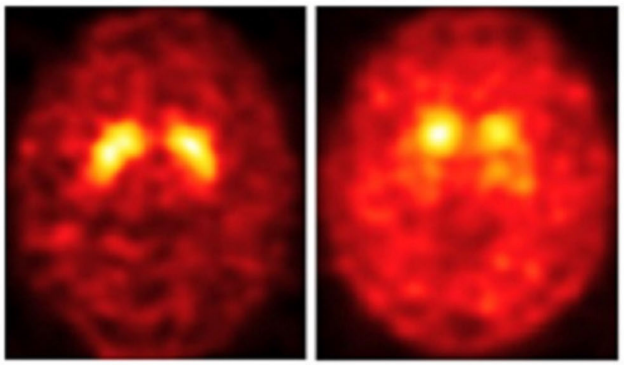


FIGURE 2. An illustrative representation of SPECT DaTSCAN images showing high contrast markings of the putamen and caudate regions in both PD patients and HC subjects.

images created through the application of these augmentation techniques. Next, we carefully divided the dataset into training and test sets. To facilitate effective learning, 75% of the images from each dataset were randomly assigned to the training set, enabling the models to learn from the patterns and underlying features present in the images. Additionally, we created a validation set from a portion of the training data, which served as a checkpoint to monitor the models' performance on unseen data and prevent over-optimization on the training set. The remaining 25% of images were reserved as the test set. Special attention was given to ensuring that there was no overlap between the training and test subsets. Specifically, we took precautions to avoid situations where samples generated from a single image would be present in both sets. This strict separation guaranteed that the test set was entirely independent from the training process, ensuring an impartial evaluation of the models' performance on unseen data.

C. FINE-TUNING OF THE PRE-TRAINED CNN MODELS: EFFICIENTNET-B0 AND MOBILENET-B0

CNN is widely used in computer vision tasks due to its ability to automatically extract features and create hierarchical representations, where initial layers can learn general features that can be reused in different problem domains, while later layers learn task-specific patterns [10]. However, training CNNs from scratch requires a significant amount of data, computational resources, and time. To address this issue, TL has emerged as a promising technique that utilizes pre-trained CNN models' knowledge to enhance the performance of new CNN models [5]. This technique has two main approaches, Feature extractor TL and Fine tuning TL. In Feature extractor TL, only the dense network on top is trained to classify the learned features, while Fine tuning TL involves Fine-Tuning the entire or a portion of the CNN model's layers. One of the main advantages of using TL is that it significantly reduces the amount of data and time required for training by avoiding the need to train the entire layer from scratch. Additionally, the parameters that need to be updated are fewer, making the training process more efficient.

In this study, we utilize TL on the EfficientNet-B0 and MobileNet-V2 models, which are trained on ImageNet, a standard benchmark dataset. Specifically, we fine-tune the last 17 layers of EfficientNet-B0 architecture and the last 13 of MobileNet-V2. This approach allows the first layers of the CNN models to extract generalizable options from a larger dataset, while the later layers take on the specifics of the smaller dataset. To determine the optimal number of layers to fine-tune, we executed multiple experiments, evaluating the performance of the models with different numbers of fine-tuned layers. By fine-tuning only a portion of the layers, we avoided overfitting and ensured that the pre-trained models remained robust. In addition, we were able to leverage the feature extraction capabilities of the remaining layers, which contributed to the models' overall performance.

The proposed pre-trained CNN models were fine-tuned using the SGD optimizer with a momentum of 0.8, decay = 1e-6, and a learning rate of 0.001. The epoch number was set at 40, while the batch size was set to 16. As we perform a binary classification task, the loss function was set to binary cross-entropy, and the activation function in the output layer was set to Sigmoid. Notably, these hyper-parameters were selected based on multiple experiments, which confirmed their ability to produce the best model performance, and as such, they were fixed for all subsequent evaluations.

1) THE PRE-TRAINED CNN MODEL: EFFICIENTNET-B0

CNNs are a popular choice for image classification tasks, but their development often involves a fixed resource cost, which can limit their accuracy [10]. Scaling techniques have been used to improve accuracy when more resources become available, but these methods are often arbitrary and require manual tuning, resulting in little or no improvement in performance. To address this issue, EfficientNet-B0 was developed using a compound scaling method, which involves a grid search to determine the relationship between different scaling dimensions of the baseline network under a fixed resource constraint [17]. This approach represents a significant improvement over traditional scaling methods and has enabled the development of seven different EfficientNet models. These models were developed using the AutoML MNAS framework, and use a mobile inverted bottleneck convolution (MBConv) with an additional Squeeze and Excitation (SE) block [17]. The models perform well on ImageNet and have achieved state-of-the-art accuracy on other datasets using TL technique. In this study, we use EfficientNet-B0, which is a mobile-sized baseline network in the EffichientNet family consisting of seven MBConv blocks. The network architecture of EfficientNet-B0 is shown in Figure 4.

2) THE PRE-TRAINED CNN MODEL: MOBILENET-V2

MobileNet-V2 is a widely adopted pre-trained CNN that has shown outstanding performance in various vision

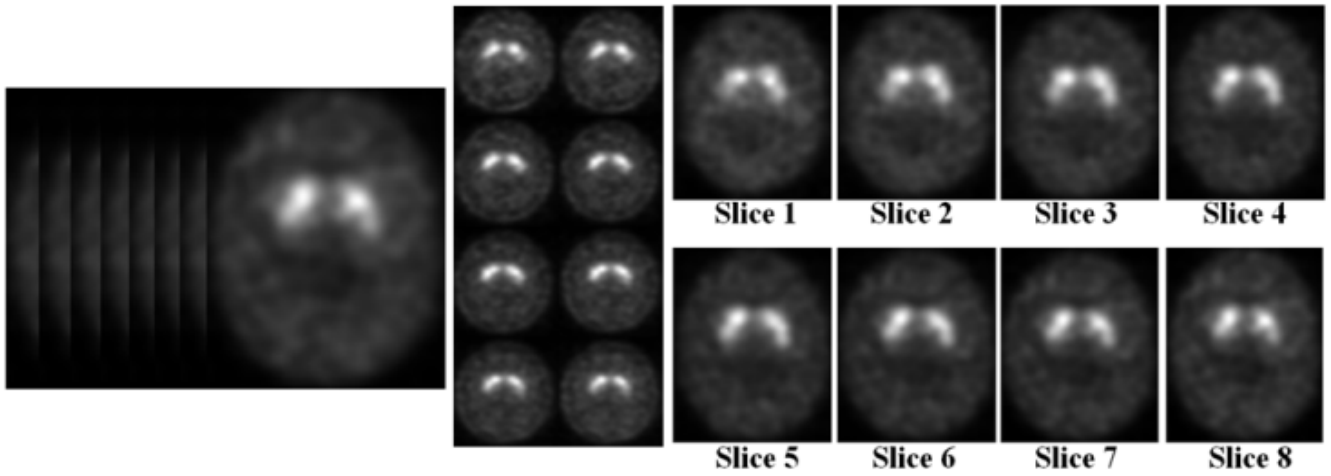


FIGURE 3. Sample Images from the three datasets: D1, D2, and D3, each containing images with a resolution of 224 × 224. D1 consists of a series of eight slices, while D2 combines these slices into one image. In contrast, D3 has the same eight slices as D1, but each slice is presented as a separate image.

applications, including classification, object detection, and semantic segmentation [18]. This model is highly efficient and has achieved state-of-the-art results on many visual recognition tasks. It builds on the concepts introduced in MobileNet-V1, utilizing depthwise separable convolutions as efficient building blocks, while also introducing inverted residual blocks between bottleneck layers. The intermediate expansion layer employs lightweight depthwise convolutions to filter features as a source of non-linearity.

The pre-trained MobileNet-V2 architecture comprises an initial fully convolutional layer with 32 filters, followed by 19 residual bottleneck layers [18]. This design significantly reduces the number of operations and memory needed while maintaining higher accuracy than the original MobileNet-V1. Larger image sizes lead to better performance, and MobileNet-V2 supports any input size greater than 32 × 32.

Overall, MobileNet-V2 is a powerful and efficient model that has demonstrated exceptional performance in image classification tasks. The use of inverted residual blocks and lightweight depthwise convolutions has improved the model's accuracy while maintaining its efficiency, making it an invaluable tool for vision-based applications. The network architecture of Mobile-Net V 2 is illustrated in Figure 5.

D. THE BILINEAR POOLING (BCNN) MODEL

The BCNN model is an architecture designed for recognizing and classifying visual objects by focusing on specific details [19]. It employs two equal or different CNN models in parallel as feature extractors, with their output feature maps combined through an outer product function to generate an image descriptor. Compared to conventional CNN models, the BCNN architecture produces a significant amount of information, thanks to its ability to model local pairwise feature interactions in a translationally invariant manner. This capability is particularly useful in fine-grained categorization, as the outer product captures pairwise correlations between feature channels and can model interactions between different parts of an object. Moreover, the BCNN simplifies gradient calculations, making it easier to train compared to other models. Figure 6 depicts a bilinear architecture with two CNN models, A and B.

The BCNN approach operates as follows:

a) An input image I is fed into two CNN models A and B, which can be either identical or different networks.

b) Feature maps are generated from both A and B, with the idea being that each network learns distinct features from the input image.

c) The Bilinear pooling layer combines the two independent output features from A and B, producing a fixed-length high-dimensional feature representation. The intuition behind the bilinear pooling layer is that the feature interactions allow for the detection of more specific details in the image.

d) Finally, the output is passed to a fully connected layer to obtain predictions.

The Bilinear model for image classification is described by the following formula.

$$\mathrm{M} = F(f_A, f_B, P, C) \tag{1}$$

where M is the bilinear model, f_A and f_B are two feature exactors of CNN_A and CNN_B, P is the pooling function and C is the classification function.

The feature extractor produces a mapping function $f : JxL \rightarrow \mathrm{R}^{cxD}$ that considers an image J and a location (position and scale) L and outputs a feature of size cxD.

In order to obtain bilinear features, the matrix outer product at each position is used to combine the output of f_A and f_B as $(\mathrm{B}\,(l, J, f_A, f_B) = f_A\,(l, J)^T f_B(l, J))$.

where the feature dimension c of f_A and f_B must be compatible. Bilinear features at all positions of the image are

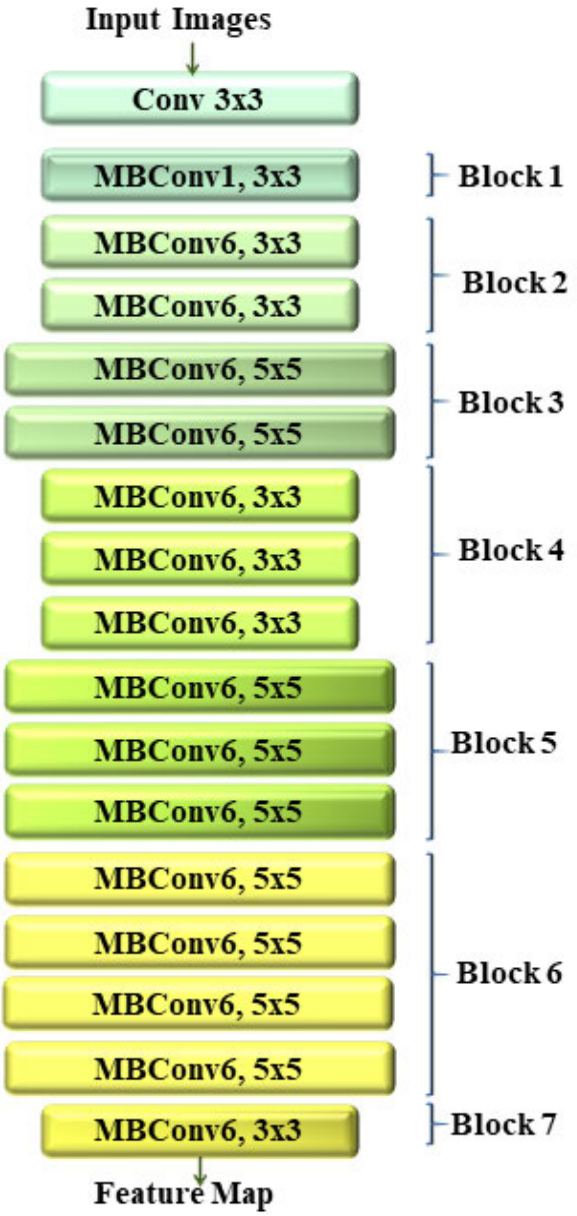


FIGURE 4. The Architecture of EfficientNet-B0: a pre-trained deep neural network designed for image classification tasks and trained on the ImageNet dataset [17].

gathered by the pooling function (P). One way of P is to sum the bilinear features of all positions:

$$\mathrm{P}(J) = \sum_{l \in L} B(l, J, f_A, f_B) \tag{2}$$

After that P(J) is reshaped to a 1-dimensional bilinear vector V(J) and pass through signed square root step as:

$$x = \mathrm{sign}((\mathrm{V}(\mathrm{J}))\sqrt{|V(\mathrm{J})|} \tag{3}$$

And Y(J) follow by d_2 normalization as $(Z(J) = \frac{\mathrm{Y(J)}}{||\mathrm{Y(J)}||_2})$ to improve performance.

The Z is fed to the classifier (C) to produce scores for different input categories.

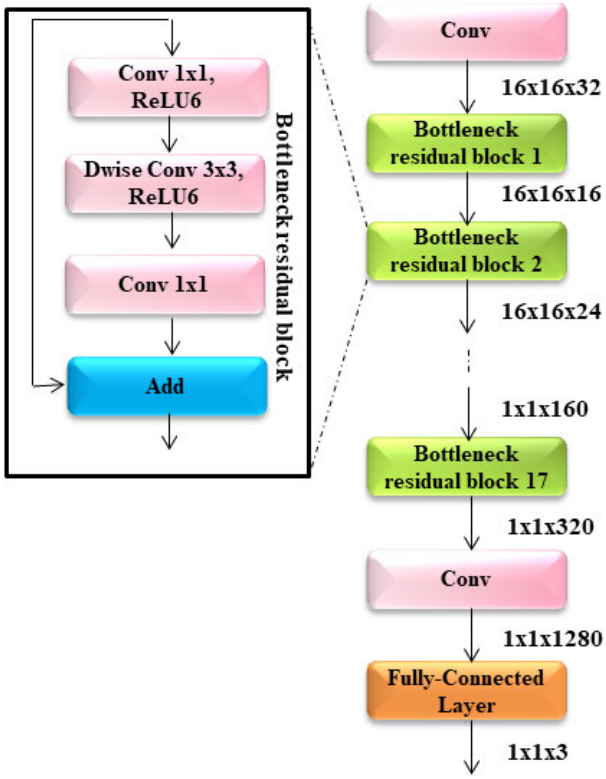


FIGURE 5. The Architecture of MobileNet-V2: a pre-trained deep neural network designed for image classification tasks and trained on the ImageNet dataset [18].

E. THE PROPOSED BCNN MODEL

The BCNN models proposed for PD patients and HC subjects classification utilized two pre-trained CNN architectures, EfficientNet-B0 and MobileNet-V2. These architectures were selected for their high performance in image classification tasks and compatibility with each other. Three different BCNN models were designed using these architectures. The first model combines two EfficientNet-B0 streams denoted by [EfficientNet-B0-EfficientNet-B0], the second model consists of two MobileNet-V2 streams indicated by [MobileNet-V2-MobileNet-V2], and the third model combines streams denoted by [EfficientNet-B0-MobileNet-V2]. The TL fine-tuning technique was employed for each pre-trained CNN architecture in the entire BCNN model, training only the last 17 layers of EfficientNet-B0 architecture and the last 13 of MobileNet-V2. The same optimized hyper-parameters used in EfficientNet-B0 and Mobile-Net V2 are employed in the proposed BCNN models.

Since the Bilinear form simplifies gradient computation at the pooling layer, we built the proposed BCNN model in an end-to-end manner [19], [20]. Indeed, we considered two matrices H and D of size MxL and LxN respectively, to represent the two CNN feature maps. It is worth noting that both matrices were of the same size, with a value of 320 × 320. The pooled bilinear features x=H^T D of size MxN are obtained by the pooling step as shown in formula (4).

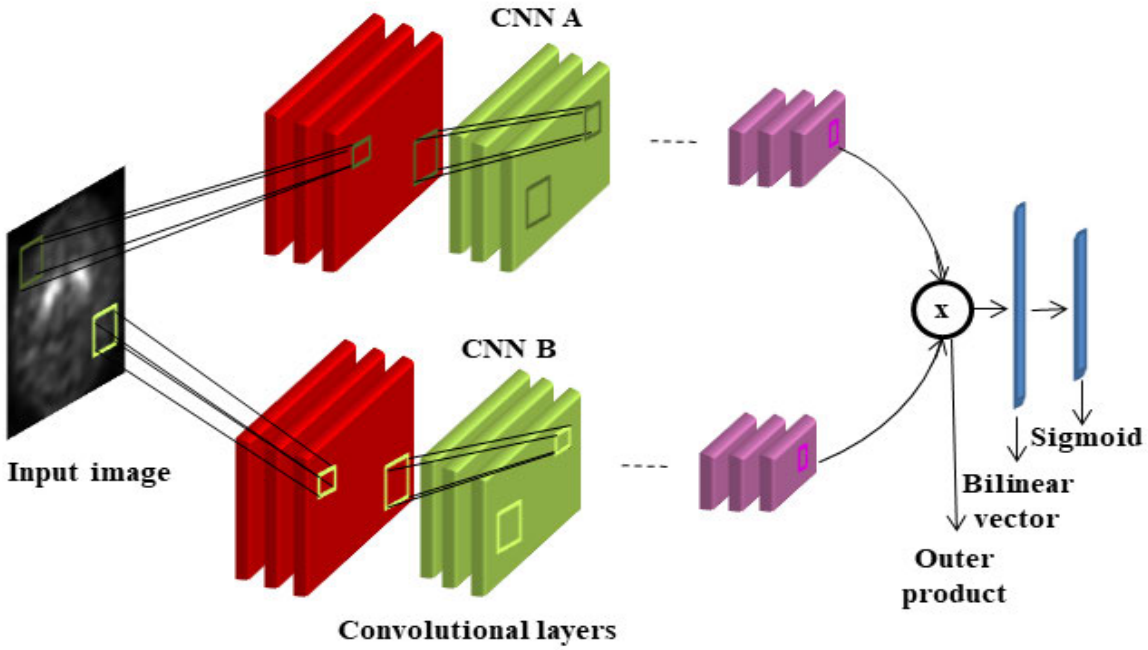


FIGURE 6. A detailed overview of the bilinear pooling (BCNN) model for image classification [19].

As the classification gradient and normalization layer is straightforward, we efficiently obtained the backward propagation of the bilinear pooling layer by the following chain rule.

$$\frac{\mathrm{dl}}{\mathrm{d}\mathbf{H}} = D\left(\frac{\mathrm{dl}}{\mathrm{d}\mathbf{Z}}\frac{\mathrm{dz}}{\mathrm{d}\mathbf{Y}}\frac{\mathrm{dy}}{\mathrm{d}\mathbf{X}}\right)^{T}, \quad \frac{\mathrm{dl}}{\mathrm{d}\mathbf{D}} = \mathrm{H}\left(\frac{\mathrm{dl}}{\mathrm{d}\mathbf{Z}}\frac{\mathrm{dz}}{\mathrm{d}\mathbf{Y}}\frac{\mathrm{dy}}{\mathrm{d}\mathbf{X}}\right) \quad (4)$$

where l represents the model loss.

The global scheme describing the end-to-end gradient computation in the BCNN model is illustrated in Figure 7.

F. CUSTOM CNN ARCHITECTURE

CNN architectures have demonstrated outstanding performance in image pre-processing, which has been the primary driving factor behind their popularity [5], [10]. It typically consists of two main components which are a convolutional base and a classifier. The convolutional base comprises a series of convolutional and pooling layers that generate features from the image, while the classifier usually consists of fully connected layers that classify the image based on these extracted features. To classify SPECT images, we designed a custom CNN architecture consisting of four convolutional layers, with 16, 32, 64, and 64 nodes in the first, second, third, and fourth convolutional layers, respectively. To improve the learning capacity of the proposed model, a Rectified Linear Unit (ReLU) activation function was added. Then, four global max-pooling layers were added, which helped to minimize the risk of overfitting and reduce the number of parameters in the model. Subsequently, to prevent network overfitting and divergence, two dropout layer with a dropout rate of 20% was incorporated. Following this, a dense layer was used to transform the output of the convolutional layers into the final classification output. Finally, a fully connected output layer is added with two nodes, one for each category, and with Sigmoid as the activation function. This layer enables the model to predict output images.

This model is trained over 40 epochs using the Stochastic Gradient Decent (SGD) optimizer with a learning rate of 0.001, decay of 1e-3, momentum of 0.9, and binary cross-entropy loss. It is trained, validated, and tested on a set of 2D SPECT images, comprising a sequence of eight slices with a resolution of 224 × 224 without resorting to any generated images via data augmentation techniques. The detailed structure of the custom CNN architecture is given in Figure 8.

IV. RESULTS

This section presents the experimental results obtained from using various CNN models to identify PD patients from HC subjects using three distinct datasets: D1, D2, and D3, each containing images with a resolution of 224 × 224. D1 consists of a series of eight slices, while D2 combines these slices into one image. In contrast, D3 has the same eight slices as D1, but each slice is presented as a separate image. The datasets were randomly divided into training and test sets, with 75% of the images allocated for training and the remaining 25% for testing. The experiments were conducted using the NVIDIA GeForce GTX 1080 Ti with 11 GB of graphics memory and Tensorflow libraries. The execution time of each model depends on its complexity. Interestingly, the custom CNN model exhibits the shortest execution time, taking only a few minutes to complete.

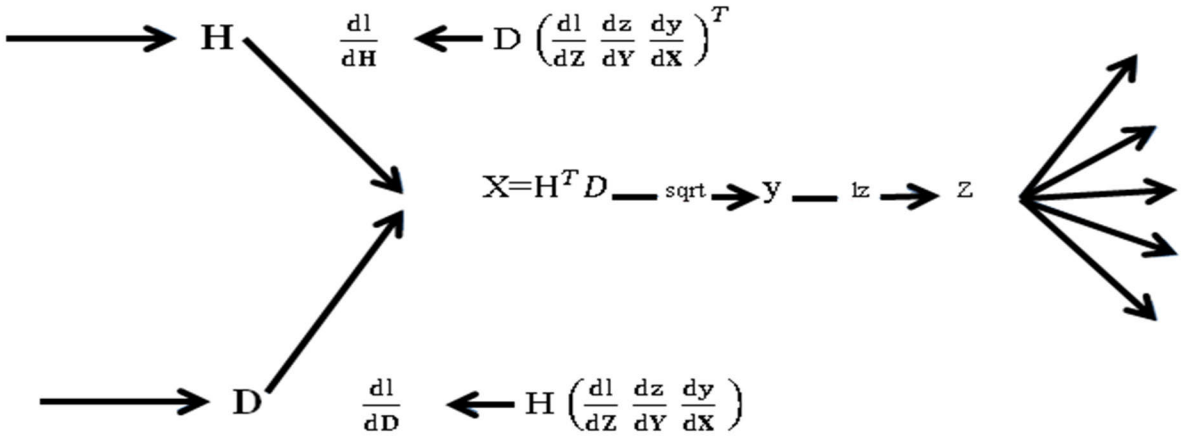


FIGURE 7. The end-to-end gradient computation at the bilinear pooling (BCNN) model for image classification [19].

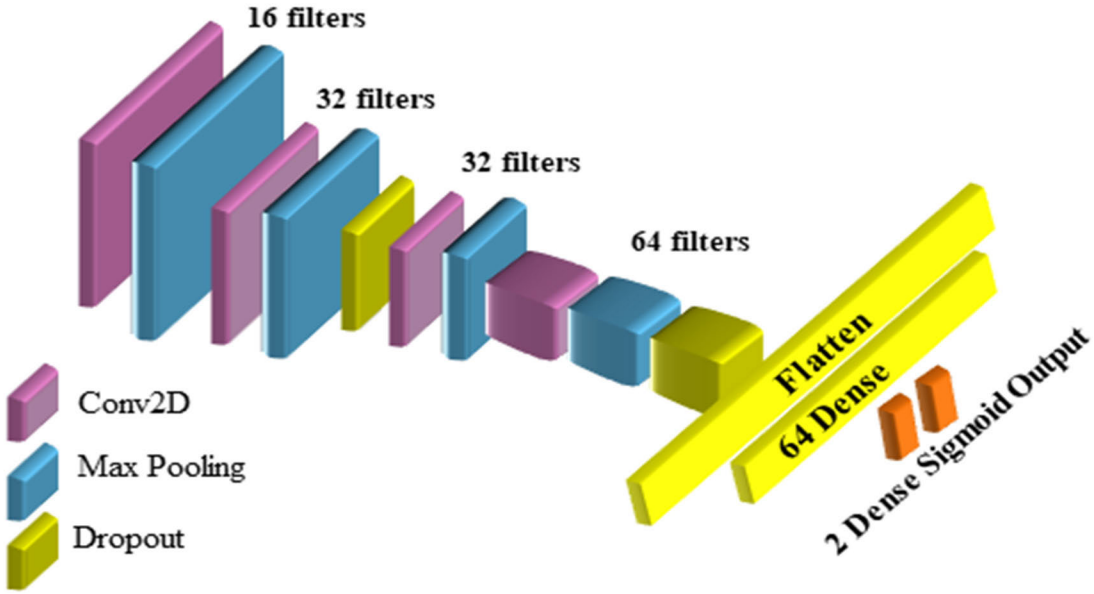


FIGURE 8. A detailed overview of the custom CNN architecture designed for PD diagnosis.

TABLE 1. Experimental results on dataset D1.

	Accuracy (%)	Precision (%)	F1 score (%)	ROC-AUC (%)	Recall (%)
Custom CNN (training)	98.74	99.22	98.83	98.77	98.45
Custom CNN (testing)	97.06	97.96	96.97	97.04	96.00

Subsequently, the CNN models executed on the D2 database come next, with both pre-trained models (EfficientNet-B0 and MobileNet-V2) exhibiting negligible time differences between them. However, the execution of the different BCNN models using the D2 database takes more time compared to EfficientNet-B0 and MobileNet-V2. On the other hand, the CNNs utilizing the D3 database require more time, as each algorithm is executed eight times (for each slice).

The performance metrics, including accuracy, precision, F1 score, and recall evaluated using the PPMI dataset, were computed for each CNN model, as shown in Table 1, 2 and 3. These performance metrics are calculated on training and testing data and based on confusion matrix parameters.

TABLE 2. Experimental results on dataset D2.

	Accuracy (%)	Precision (%)	F1 score (%)	ROC-AUC (%)	Recall (%)
Training					
EfficientNet-B0	96.47	96.10	96.48	96.47	96.86
MobileNet-V2	96.47	99.66	96.32	96.44	93.20
EfficientNet-B0-EfficientNet-B0	97.02	96.37	97.05	97.02	97.73
MobileNet-V2-MobileNet-V2	97.57	98.00	97.53	97.56	97.07
EfficientNet-B0-MobileNet-V2	**99.14**	**99.29**	**99.13**	**99.13**	**98.97**
Testing					
EfficientNet-B0	96.12	95.81	96.15	96.11	96.49
MobileNet-V2	92.59	99.21	92.25	92.74	86.21
EfficientNet-B0-EfficientNet-B0	94.59	95.41	94.50	94.58	93.60
MobileNet-V2-MobileNet-V2	91.88	92.99	92.02	91.90	91.07
EfficientNet-B0-MobileNet-V2	95.29	95.19	95.41	95.28	95.63

TABLE 3. Experimental results on dataset D3.

	Accuracy (%)	Precision (%)	F1 score (%)	ROC-AUC (%)	Recall (%)
Training					
EfficientNet-B0	96.94	95.73	96.99	96.94	98.23
MobileNet-V2	97.76	97.64	97.76	97.76	97.87
EfficientNet-B0-EfficientNet-B0	97.80	97.43	97.81	97.80	98.19
MobileNet-V2-MobileNet-V2	97.88	97.28	97.89	97.88	98.50
EfficientNet-B0-MobileNet-V2	99.10	98.90	99.09	99.10	99.29
Testing					
EfficientNet-B0	96.94	94.81	97.00	96.95	99.29
MobileNet-V2	97.65	97.67	97.67	97.65	97.67
EfficientNet-B0-EfficientNet-B0	97.88	96.57	97.91	97.88	99.29
MobileNet-V2-MobileNet-V2	97.65	96.10	97.67	97.64	98.36
EfficientNet-B0-MobileNet-V2	**98.47**	**97.51**	**98.51**	**98.45**	**99.54**

A probability curve, known as the Receiver Operating Curve (ROC), displays the classification model's performance at various classification thresholds by plotting True Positive (TP) i.e., sensitivity against False Positive (FP) i.e., specificity values. However, in this particular study, our primary focus lies in the classification decisions rather than the probabilities themselves. Hence, we have opted for an approximation of the ROC curve (resulting in two line segments). The ROC curve that exhibits the highest area under the curve (AUC) value is depicted in Figures 9- 11.

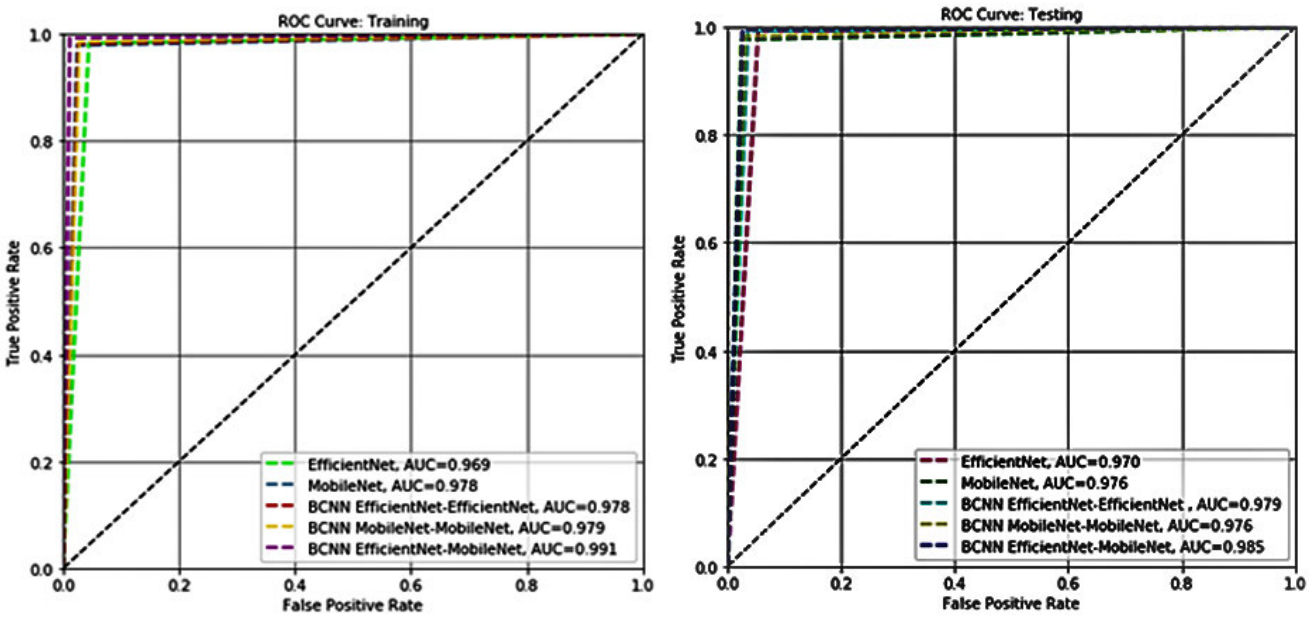
Gradient-weighted Class Activation Mapping (Grad-CAM) is an interpretability technique that enables the visualization of crucial parts of images that contribute most to classification models' predictions. It achieves this by extracting gradients from the final convolutional layer of a CNN and utilizing this information to emphasize regions responsible for the model's predicted probability of an image belonging to a predefined class. Figure 12 to 14 illustrate the comparison between the original images of a HC in SPECT and the corresponding images with superimposed attention maps, created using the Grad-CAM algorithm. These attention maps were generated for the various CNN models on the three datasets.

The learning curves provide valuable insights into the model's learning process and its ability to generalize on both the training and validation data.

Figures 15 and 16 showcase two examples of learning curves for both the training and validation processes of two CNNs.

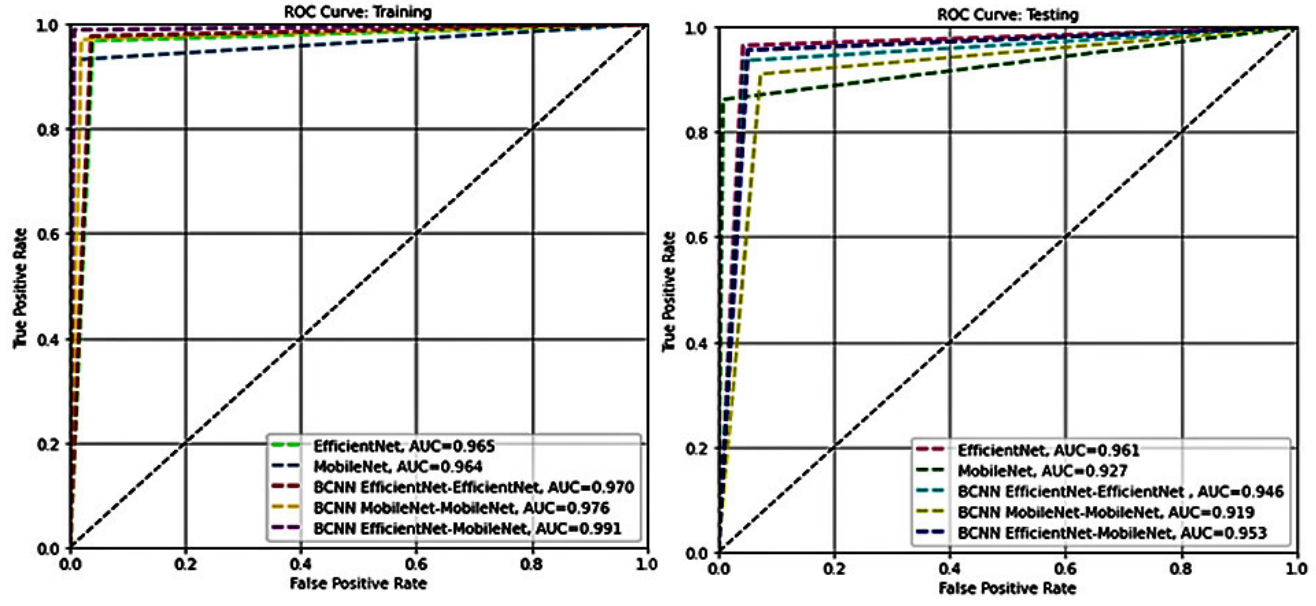
V. DISCUSSION

This study proposes various CNN models to differentiate between PD and HC in DaTSCAN SPECT images. These models include the pre-trained EfficientNet-B0 and MobileNet-V2, three BCNN models (EfficientNet-B0- EfficientNet-B0, MobileNet-V2-MobileNet-V2, and EfficientNet-B0-MobileNet-V2), and a custom CNN. The use of a pre-trained architecture from ImageNet for medical imaging classification is a practical approach to benefit from the learned high-level features and transfer knowledge from a large dataset to a smaller medical imaging dataset. In fact, EfficientNet-B0 and MobileNet-V2 were employed in this study, leveraging the pre-trained CNN models to achieve high performance without the need for extensive architecture redesign. These pre-trained models were fine-tuned to adapt the specialized features to work with the new dataset by retraining the last layers of these networks. They were also used to build the three BCNN models, where we proposed the use of bilinear pooling to enhance performance. Bilinear pooling is a second-order aggregation of CNN activations that model high-order statistical information and extract more image features compared to classical first-order aggregation methods like sum or max [19], [20]. We carefully selected the EfficientNet-B0 and MobileNet-V2 architectures for their strong performance in image classification tasks and compatibility in input and output. In the proposed BCNN models, we extracted feature maps from the same input images using the two parallel CNNs, merged them using the outer



a) ROC curve of training b) ROC curve of testing

FIGURE 9. ROC curves and AUC of the developed CNN architectures trained and tested on D3 dataset for PD diagnosis.



a) ROC curve of training b) ROC curve of testing

FIGURE 10. ROC curves and AUC of the developed CNN architectures trained and tested on D2 dataset for PD diagnosis.

product operation, and applied two normalization layers. Subsequently, a single fully connected layer was adjusted for the classification task. In the final layer of all developed CNN models, the Sigmoid activation function was employed due to its effectiveness in binary classification tasks. However, pre-trained models and BCNN models are extremely complex in terms of the number of parameters. For instance, the BCNN EfficientNet-B0-MobileNet-V2 model has over 12 million parameters. To address this issue, we have developed a custom CNN that contains only 864,242 parameters. While designing the custom CNN architecture, the optimal number and type of layers were determined by conducting extensive research into various networks and performing different executions.

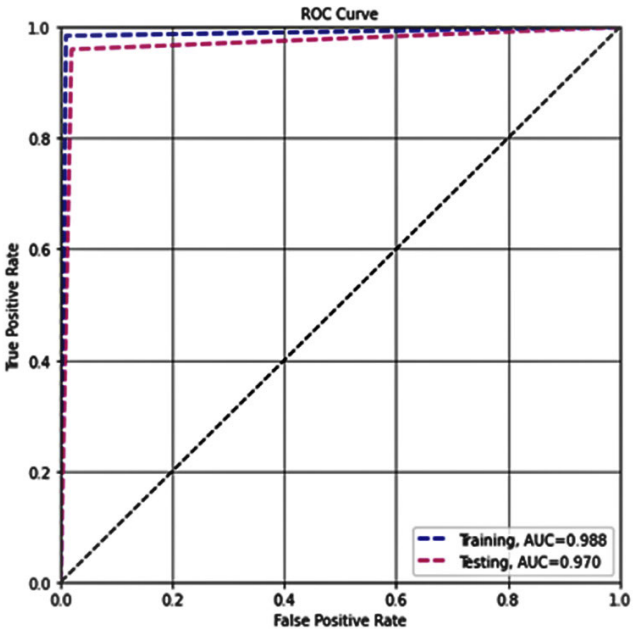


FIGURE 11. ROC curves and AUC of the developed custom CNN architecture trained and tested on D1 dataset for PD diagnosis.

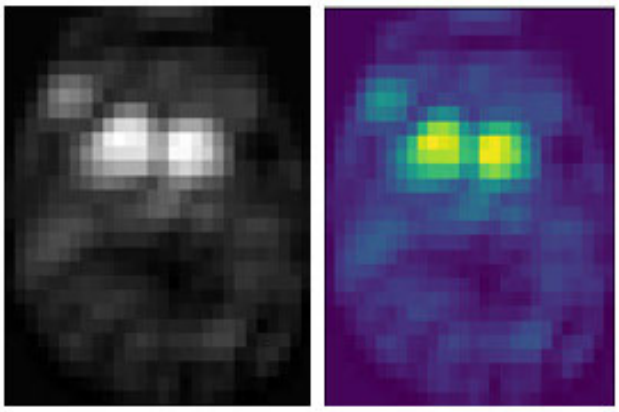


FIGURE 12. Comparison of SPECT Images of HC from D1 with attention maps generated using Grad-CAM Algorithm for the custom CNN.

We performed a series of experiments to determine the most appropriate optimizer and hyper-parameters for our CNN models. Various optimizers were evaluated, and we selected the hyper-parameters that consistently yielded desirable results. For the proposed pre-trained CNN models, EfficientNet-B0 and MobileNet-V2. For EfficientNet-B0, we trained only the last 17 layers, and for MobileNet-V2, we trained the last 13 layers. we fine-tuned them using the SGD optimizer with specific settings, including a momentum of 0.8, decay of 1e-6, and a learning rate of 0.001. We conducted training over 40 epochs with a batch size of 16, using binary cross-entropy as the loss function and Sigmoid as the activation function in the output layer. The same optimized hyper-parameters used in EfficientNet-B0 and MobileNet-V2 were employed in the BCNN models. Additionally, the custom CNN architecture underwent training for 40 epochs with the SGD optimizer, using a learning rate of 0.001, a decay of 1e-3, a momentum of 0.9, and a binary cross-entropy loss function.

The developed CNN models were trained on the preprocessed PPMI data, which was split into appropriate sets. In order to preprocess the volumetric DaTSCAN SPECT images, we have selected eight individual slices, which were processed as sequences of slices in the custom CNN. These eight slices were then combined into a single image and used in both the EfficientNet-B0 and MobileNet-V2 architectures, as well as in the three BCNN models. Consequently, instead of using a single data stream, we used the same data stream in different forms.

The results, as shown in Table 1, Table 2, Table 3, and Figure 9-11, demonstrated high performance across all models trained and tested on a specific dataset, with performance metrics greater than 90% across all datasets. Indeed, we can observe that all models perform well on the training set, achieving accuracy scores ranging from 96.47% to 99.14%. The EfficientNet-B0-MobileNet-V2 (D2) model achieved the highest accuracy score of 99.14%, while the MobileNet-V2 (D2) and EfficientNet-B0 (D2) models achieved the lowest accuracy score of 96.47%. Additionally, EfficientNet-B0-MobileNet-V2 (D2) achieved the highest F1 score and ROC-AUC scores of 99.13%. and 96.13%, respectively. However, in terms of precision, the Mobile-Net (D2) model achieved the highest score of 99.66%. On the other hand, the EfficientNet-B0-MobileNet-V2 (D3) model achieved the highest recall score of 99.29%.

When evaluating the models on the testing set, we can observe that the performance of the models is lower than on the training set. The accuracy scores on the testing set range from 91.88% to 98.47%. The EfficientNet-B0-MobileNet-V2 (D3) model achieved the highest accuracy score of 98.47%, while the MobileNet-V2-MobileNet-V2 (D2) model achieved the lowest accuracy score of 91.88%. Moreover, in terms of F1 score, ROC-AUC, and Recall, the EfficientNet-B0-EfficientNet-B0 (D3) model achieved the highest scores of 98.51%, 98.45%, and 99.54%, respectively. However, Mobile-Net (D2) achieved the highest precision score of 99.21%.

The performance of the various CNN models was evaluated also using the ROC curve, with the best model represented by the curve closest to the upper left corner of the figure. The AUC, which measures the area under the curve and ranges from 0 to 1, was used to quantify the diagnostic ability of each model, with a higher AUC indicating better performance. In this regard, all the developed CNN models exhibited a robust separation capacity, as reflected by the vertices approaching angles of 90 degrees. Notably, a right angle at the point of curvature indicates a perfect diagnostic system. Nevertheless, the BCNN EfficientNet-B0-MobileNet-V2 (D3) outperformed the other developed CNN

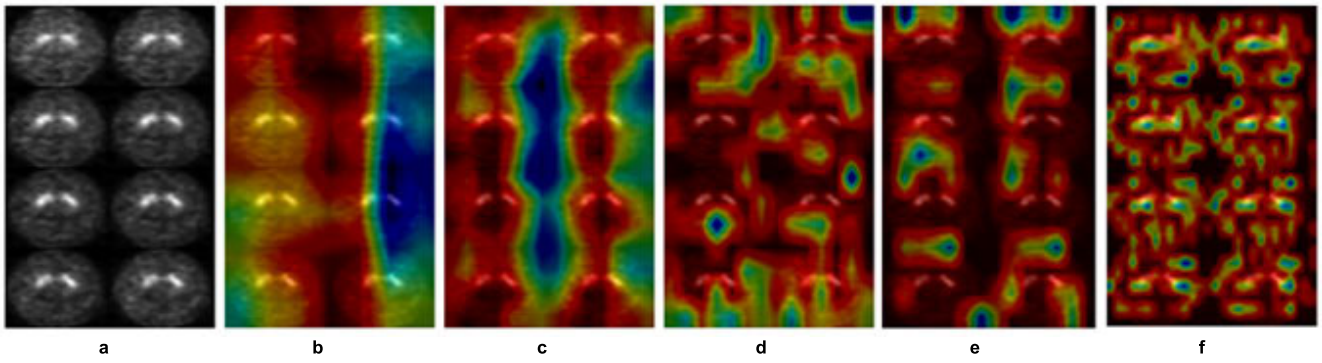


FIGURE 13. Comparison of SPECT Images of HC from D2 with attention maps generated using Grad-CAM algorithm for the various CNN models; a) Original image, b) EfficientNet-B0, c) MobileNet-V2, d) BCNN EfficientNet-B0-EfficientNet-B0 e) BCNN MobileNet-V2-MobileNet-V2 f) BCNN MobileNet-V2-EfficientNet-B0.

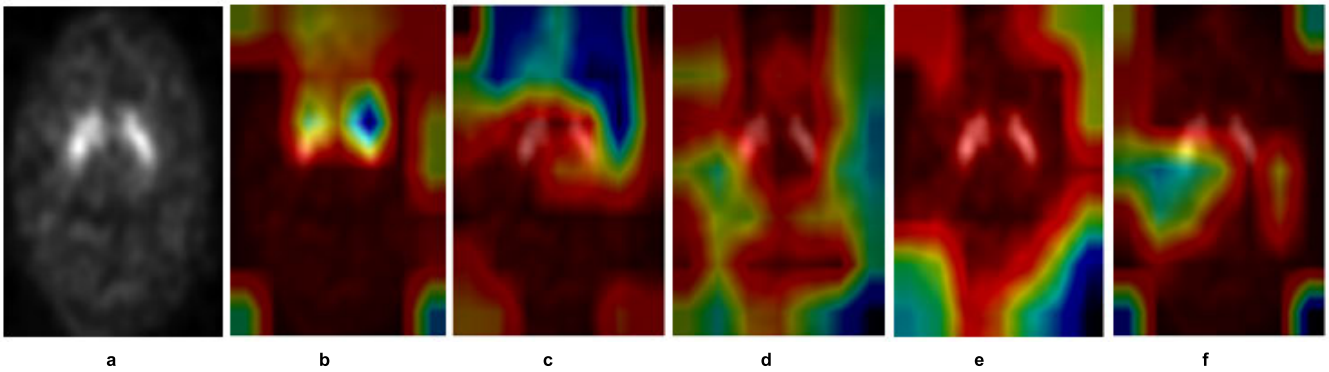


FIGURE 14. Comparison of SPECT Images of HC from D3 with attention maps generated using Grad-CAM algorithm for the various CNN models; a) Original image, b) EfficientNet-B0, c) MobileNet-V2, d) BCNN EfficientNet-B0-EfficientNet-B0 e) BCNN MobileNet-V2-MobileNet-V2 f) BCNN MobileNet-V2-EfficientNet-B0.

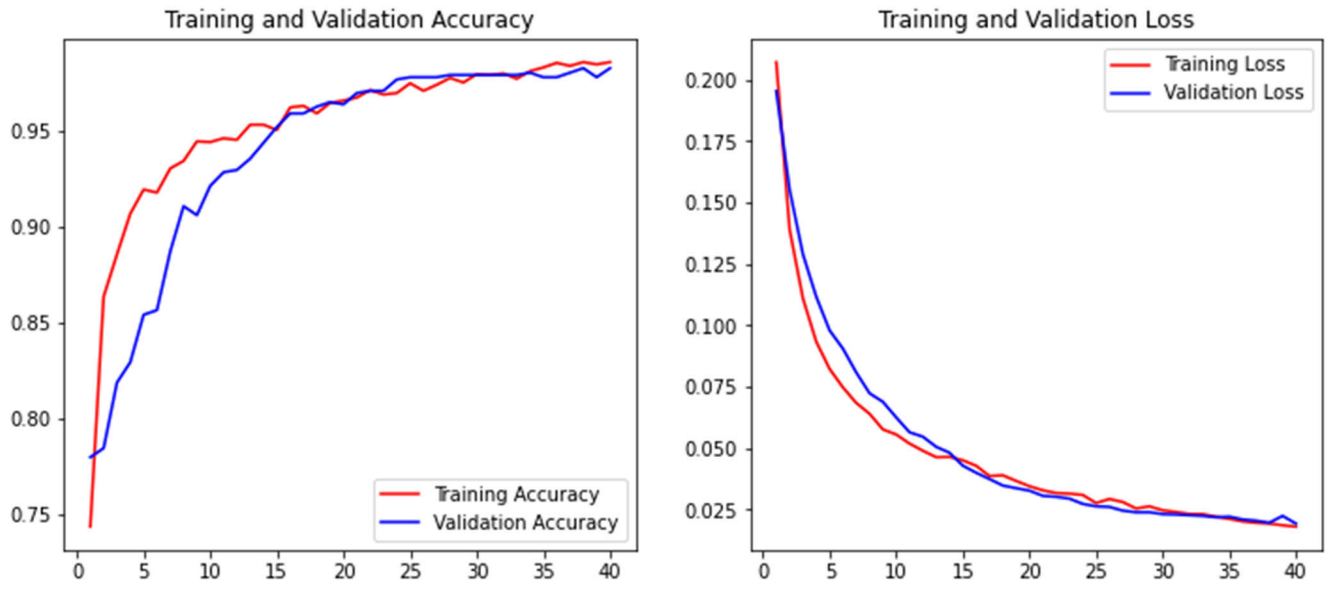


FIGURE 15. Learning and validation curves of BCNN MobileNet-V2-EfficientNet-B0 for D3 (Slice 1).

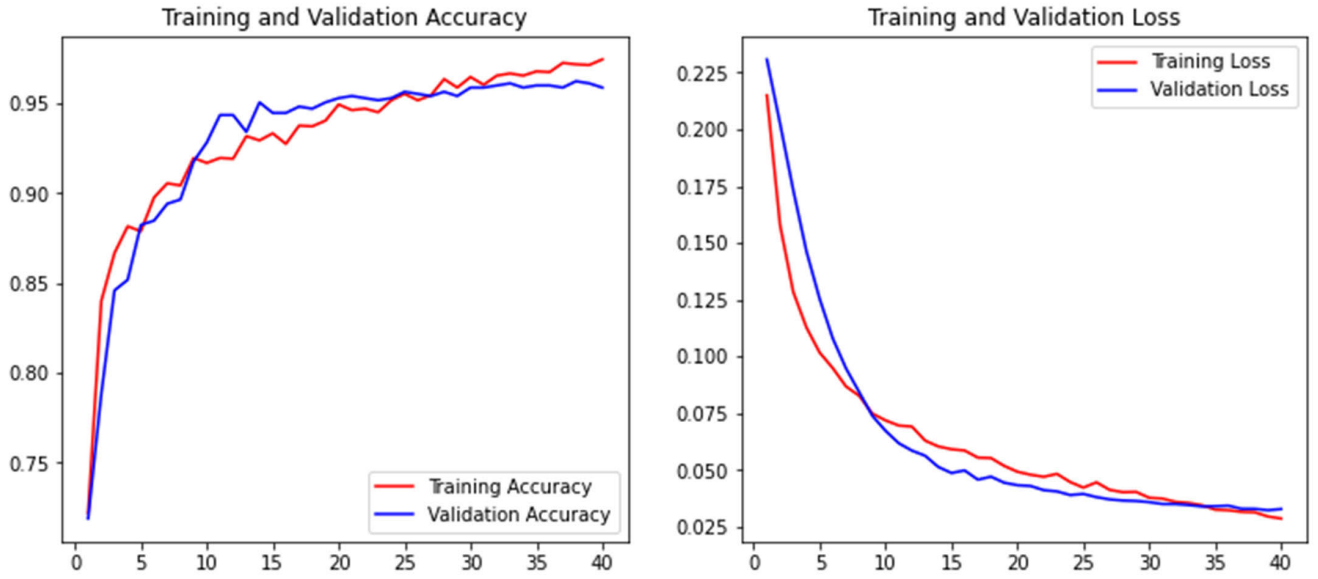


FIGURE 16. Learning and validation curves of BCNN MobileNet-V2-EfficientNet-B0 for D3 (Slice 2).

models, achieving an AUC of 99.29% and covering more surface area.

Overall, we can observe that the EfficientNet-B0-MobileNet-V2 (D2) performed well on the training set, and EfficientNet-B0-MobileNet-V2 (D3) models performed well on the testing sets, achieving high scores in most of the performance metrics. This resulted in an accuracy of 98.47% with only 40 epochs, highlighting the power of bilinear pooling and TL methods, as well as the practicality of leveraging pre-trained models for image classification. Notably, using a BCNN model with two different architectures led to even better performance, potentially due to the dissimilar outputs of feature extractors with different structures. Additionally, the BCNN model can acquire more distinguishing representations and capture diverse information at various levels, further enhancing its efficacy. Ultimately, these findings suggest that incorporating BCNN with two different architectures may be an effective strategy for enhancing the discriminating ability and improving classification performance in image recognition tasks.

It is worth noting that although the same images were used in all three datasets, they were organized in different ways. This variation in image organization had an impact on the performance of the models, with some models performing better on certain datasets than others. These findings suggest that the way images are presented and organized can have an effect on the performance of CNN models in diagnosing PD using SPECT DaTSCAN images.

During the training phase, we closely monitored the loss values and accuracy rate. The loss values steadily decreased as the models iteratively adjusted their parameters to minimize the error between predicted and actual values. This decreasing loss indicates that the models were gradually improving their ability to make accurate predictions on the training data. Additionally, the accuracy performance rate displayed a favorable trend throughout the training process. Initially, the accuracy rate might have been relatively low, but as the models iteratively learned from the data, it steadily increased over the epochs. The fact that the accuracy rate consistently increased and reached a high level after several epochs shows that the models were able to capture the underlying patterns and features in the training data effectively. This progress is crucial for ensuring that the models can generalize well and make accurate predictions on new, unseen data.

The Grad-CAM-generated attention map reveals the most influential portions of the SPECT image that impact the model's predictions. Interestingly, the various CNN models developped on the three datasets demonstrate a focus on distinct areas of the brain. This observation holds significance, where the PD affects multiple brain regions rather than being restricted solely to the striatum.

The ANOVA statistical test was employed in this research work to assess the statistical significance of performance differences among the developed models. For the training performance, statistically significant differences were identified in the CNN models concerning precision, F1 score, ROC-AUC, and recall, with the exception of accuracy. The p-value of 0.000 was lower than the predetermined significance level of 0.05 ($p<0.05$). Regarding the test performance, statistically significant differences were found in the CNN models for precision, and ROC-AUC, while no significant differences were found for accuracy, F1 score, and recall values.

In the literature, several state-of-the-art studies [11], [12], [13], [14], [15] that use DL for PD diagnosis, also utilize the same SPECT image dataset as this study, albeit with a varying

TABLE 4. Comparison of the proposed model with state-of-the art studies for pd diagnosis using spect images.

References	Datasets	Method Used	Accuracy
Shih-Yen Hsu et al. [11]	202 99mTc-TRODAT-1 SPECT images	Six pre-trained CNN models for the classification of multiple stages of PD	85.5% (DenseNet)
Pavan Rajkumar Magesh et al. [12]	642 DaTSCAN SPECT images	TL technique of the pre-trained VGG16 model to distinguish PD from non-PD	95.2%
Hathaliya Jigna et al. [13]	78256 SPECT images	Custom CNN model contained 14 layers and classified input SPECT images into four categories: PD, HC, SWEDD, and GenReg PD.	88.9%
Kurmi Ankit et al. [14]	645 DaTSCAN images	Four pre-trained CNN models followed by a fuzzy fusion logic-based ensemble approach	98.45%
Juan E. Arco et al. [15]	269 DaTSCAN images	Bayesian DL-based multi-level ensemble classification system	95.31%
Proposed study	2720 DaTSCAN images	BCNN EfficientNet-B0-MobileNet-V2 (D3)	98.47%

number of subjects. This means that comparing the results of this study to the previous ones is not straightforward. Nonetheless, as an indication, we present a comparison of our proposed model's performance with existing classification models in Table 4 which demonstrates that our proposed method achieves a significantly higher accuracy of 98.47%. Thus, by fine-tuning pre-trained CNN models and utilizing the bilinear pooling method, we achieved state-of-the-art performance. These findings underscore the potential of DL models in binary classification tasks and the importance of selecting appropriate architectures for different datasets. Overall, the proposed CNN model has the potential to serve as a diagnostic tool for PD, enhancing diagnosis performance and reducing dependence on subjective clinical evaluations.

While the proposed method has demonstrated promising findings, some limitations need to be addressed. Firstly, the dataset includes only 2720 SPECT images obtained from 340 cases, which may not be representative of the overall population of PD patients, and lacks diversity as it only consists of images from PD patients and HC subjects. Moreover, the study only focuses on SPECT imagery, and other types of imaging modalities that could aid in the diagnosis of PD are not explored. Therefore, a more extensive and diverse dataset is necessary to enhance the generalizability of the proposed method. Additionally, while bilinear pooling operation is suitable for fine-grained classification, high-dimensional bilinear features are impractical and can significantly increase memory usage, making them unsuitable for further analysis. Additionally, in the BCNN model, the dimensionality of feature vectors must be the same for both outputs, which can be a limitation for multi-model bilinear pooling. While generalizations like trilinear pooling have been proposed to consider triples of features, this can make the model more difficult to train and susceptible to overfitting.

VI. CONCLUSION

The current study aims to address the challenges in diagnosing PD in clinical practice by proposing an automated approach that utilizes DaTSCAN imaging and DL to classify subjects as having PD or not. Various CNN models, including EfficientNet-B0, MobileNet-V2 models, three BCNN models, and a custom CNN architecture, were introduced to accomplish this task. Experimental findings revealed that the BCNN EfficientNet-B0-MobileNet-V2 (D3) model achieved the highest accuracy score of 98.47%, outperforming other developed CNN architectures and previous studies. This success can be attributed to the use of pre-trained CNN models, TL technique, and bilinear pooling method. Given the results achieved, the proposed approach has the potential to be a valuable tool for assisting physicians in PD diagnosis. In future work, various algorithms, such as parallel Strength Pareto Evolutionary Algorithm and memetic differential evolution, will be used to tune the hyper-parameters of the proposed method, as the optimal selection of hyper-parameters was not considered in this work. Additionally, the proposed approach will be tested on different datasets with a larger number of samples and classes. Finally, an unsupervised DL model will be developed for early PD prediction within the SWEDD group.

REFERENCES

[1] H. Khachnaoui, R. Mabrouk, and N. Khlifa, "Machine learning and deep learning for clinical data and PET/SPECT imaging in Parkinson's disease: A review," *IET Image Process.*, vol. 14, no. 16, pp. 4013–4026, Dec. 2020, doi: 10.1049/iet-ipr.2020.1048.

[2] J. Jankovic, "Parkinson's disease: Clinical features and diagnosis," *J. Neurol., Neurosurg. Psychiatry*, vol. 79, no. 4, pp. 368–376, Apr. 2008, doi: 10.1136/jnnp.2007.131045.

[3] J. C. Taylor and J. W. Fenner, "Comparison of machine learning and semi-quantification algorithms for (I123)FP-CIT classification: The beginning of the end for semi-quantification?" *EJNMMI Phys.*, vol. 4, no. 1, pp. 1–20, Dec. 2017, doi: 10.1186/s40658-017-0196-1.

[4] A. V. L. N. Sujith, G. S. Sajja, V. Mahalakshmi, S. Nuhmani, and B. Prasanalakshmi, "Systematic review of smart health monitoring using deep learning and artificial intelligence," *Neurosci. Informat.*, vol. 2, no. 3, Sep. 2022, Art. no. 100028, doi: 10.1016/j.neuri.2021.100028.

[5] H. E. Kim, A. Cosa-Linan, N. Santhanam, M. Jannesari, M. E. Maros, and T. Ganslandt, "Transfer learning for medical image classification: A literature review," *BMC Med. Imag.*, vol. 22, no. 1, p. 69, Dec. 2022, doi: 10.1186/s12880-022-00793-7.

[6] N. Yadav, S. M. Alfayeed, A. Khamparia, B. Pandey, D. N. H. Thanh, and S. Pande, "HSV model-based segmentation driven facial acne detection using deep learning," *Exp. Syst.*, vol. 39, no. 3, Mar. 2022, Art. no. e12760, doi: 10.1111/exsy.12760.

[7] P. Pareek and A. Thakkar, "A survey on video-based human action recognition: Recent updates, datasets, challenges, and applications," *Artif. Intell. Rev.*, vol. 54, no. 3, pp. 2259–2322, Mar. 2021, doi: 10.1007/s10462-020-09904-8.

[8] X. Li, F. Dong, S. Zhang, and W. Guo, "A survey on deep learning techniques in wireless signal recognition," *Wireless Commun. Mobile Comput.*, vol. 2019, pp. 1–12, Feb. 2019, doi: 10.1155/2019/5629572.
[9] H. Khachnaoui, N. Khlifa, and R. Mabrouk, "Machine learning for early Parkinson's disease identification within SWEDD group using clinical and DaTSCAN SPECT imaging features," *J. Imag.*, vol. 8, no. 4, p. 97, Apr. 2022, doi: 10.3390/jimaging8040097.
[10] L. Chen, S. Li, Q. Bai, J. Yang, S. Jiang, and Y. Miao, "Review of image classification algorithms based on convolutional neural networks," *Remote Sens.*, vol. 13, no. 22, p. 4712, Nov. 2021, doi: 10.3390/rs13224712.
[11] S.-Y. Hsu, L.-R. Yeh, T.-B. Chen, W.-C. Du, Y.-H. Huang, W.-H. Twan, M.-C. Lin, Y.-H. Hsu, Y.-C. Wu, and H.-Y. Chen, "Classification of the multiple stages of Parkinson's disease by a deep convolution neural network based on 99 mTc-TRODAT-1 SPECT images," *Molecules*, vol. 25, no. 20, p. 4792, Oct. 2020, doi: 10.3390/molecules25204792.
[12] P. R. Magesh, R. D. Myloth, and R. J. Tom, "An explainable machine learning model for early detection of Parkinson's disease using LIME on DaTSCAN imagery," *Comput. Biol. Med.*, vol. 126, Nov. 2020, Art. no. 104041, doi: 10.1016/j.compbiomed.2020.104041.
[13] J. Hathaliya, R. Parekh, N. Patel, R. Gupta, S. Tanwar, F. Alqahtani, M. Elghatwary, O. Ivanov, M. S. Raboaca, and B.-C. Neagu, "Convolutional neural network-based Parkinson disease classification using SPECT imaging data," *Mathematics*, vol. 10, no. 15, p. 2566, Jul. 2022, doi: 10.3390/math10152566.
[14] A. Kurmi, S. Biswas, S. Sen, A. Sinitca, D. Kaplun, and R. Sarkar, "An ensemble of CNN models for Parkinson's disease detection using DaTSCAN images," *Diagnostics*, vol. 12, no. 5, p. 1173, May 2022, doi: 10.3390/diagnostics12051173.
[15] J. E. Arco, A. Ortiz, J. Ramírez, F. J. Martínez-Murcia, Y.-D. Zhang, and J. M. Górriz, "Uncertainty-driven ensembles of multi-scale deep architectures for image classification," *Inf. Fusion*, vol. 89, pp. 53–65, Jan. 2023, doi: 10.1016/j.inffus.2022.08.010.
[16] K. Marek et al., "The Parkinson progression marker initiative (PPMI)," *Prog. Neurobiol.*, vol. 95, no. 4, pp. 629–635, Dec. 2011, doi: 10.1016/j.pneurobio.2011.09.005.
[17] H. Alhichri, A. S. Alswayed, Y. Bazi, N. Ammour, and N. A. Alajlan, "Classification of remote sensing images using EfficientNet-B3 CNN model with attention," *IEEE Access*, vol. 9, pp. 14078–14094, 2021, doi: 10.1109/ACCESS.2021.3051085.
[18] J. P. Gujjar, H. R. P. Kumar, and N. N. Chiplunkar, "Image classification and prediction using transfer learning in Colab notebook," *Global Transitions Proc.*, vol. 2, no. 2, pp. 382–385, Nov. 2021, doi: 10.1016/j.gltp.2021.08.068.
[19] T.-Y. Lin, A. RoyChowdhury, and S. Maji, "Bilinear CNN models for fine-grained visual recognition," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1449–1457, doi: 10.1109/ICCV.2015.170.
[20] R. Mastouri, N. Khlifa, H. Neji, and S. Hantous-Zannad, "A bilinear convolutional neural network for lung nodules classification on CT images," *Int. J. Comput. Assist. Radiol. Surgery*, vol. 16, no. 1, pp. 91–101, Jan. 2021, doi: 10.1007/s11548-020-02283-z.
[21] Y. Zhang, M. Ni, C. Zhang, S. Liang, S. Fang, R. Li, and Z. Tan, "Research and application of AdaBoost algorithm based on SVM," in *Proc. IEEE 8th Joint Int. Inf. Technol. Artif. Intell. Conf. (ITAIC)*, May 2019, pp. 662–666, doi: 10.1109/ITAIC.2019.8785556.

HAJER KHACHNAOUI received the B.S. degree from the Higher Institute of Medical Technologies, Université de Tunis El Manar, in 2016, and the M.S. degree in medical image processing from Université de Tunis El Manar, in 2018. She is currently pursuing the Ph.D. degree with the Department of Medical Image Processing, BTM Laboratory, which is also affiliated with Université de Tunis El Manar. Her research interests lie in the fields of medical image processing and artificial intelligence.

BELKACEM CHIKHAOUI received the M.Sc. and Ph.D. degrees from the University of Sherbrooke, QC, Canada, in 2009 and 2013, respectively. He is currently a Full Professor of computer science with the Department of Science and Technology, TELUQ University, Montreal, QC, Canada. He has been the Canada Research Chair in multimodal data mining, since 2023. He is also an Adjunct Professor with the University of Sherbrooke, and a Senior Researcher and the Laboratory Director with the CRIUGM Research Institute, University of Montreal. His current research interests include data mining, pattern recognition, machine learning, neural networks, social networks and web, recommendation systems, health informatics, and the Internet of Things.

NAWRES KHLIFA received the Ph.D. degree from the National Engineering School of Tunis (ENIT). She is an Engineer. She is currently a Professor with the Higher Institute of Medical Technologies, Université de Tunis El Manar. She coordinates the TIMEd Team: Medical Image Processing, BTM Laboratory. Her research work focuses on artificial intelligence and CAD design in medical imaging, emotion recognition, and gaze tracking.

ROSTOM MABROUK received the M.Sc. degree from the National Engineering School of Tunis, Tunisia, and the Ph.D. degree from the University of Sherbrooke, QC, Canada. He is currently an Adjunct Professor with the Department of Computer Science, Bishop's University, Sherbrooke, QC, Canada. His current research interests include medical image analysis, machine learning, and health informatics.