ISSN 1870-4069

Parkinson's Detection Using Convolutional Neural Networks on Handwritten Wave Images

Carolina Rosas-Alatriste, Noé Oswaldo Rodríguez-Rodríguez, Paola Itzel Delena-García, Antonio Alarcón-Paredes

Instituto Politécnico Nacional,
Centro de Investigación en Computación,
Mexico

{crosasa2023, nrodriguezr2023, pdelenag2023, aalarcon}@cic.ipn.mx

Abstract. Parkinson's disease is a neurodegenerative condition for which the early detection is a very challenging activity for the medical community. Although traditional methods for Parkinson's disease diagnosis involve the use of EEG (electroencephalographic) activity, previous works have proposed to analyze sketches of guided spirals and waves drawn by a patient versus those drawn by healthy people. In this work, we made use of the same dataset, employing data augmentation techniques for enriching the diversity of the images. Besides, architectures such as ResNet50 and VGG19 demonstrated promising results using transfer learning. Results reported in this manuscript are comparable with those of the stateof-the-art, but also have the potential to improve the accuracy in the near future.

Keywords: Parkinson, CNN, deep learning, machine learning, classification, supervised learning.

1 Introduction

Parkinson's disease is a common neurological condition that can significantly disrupt a patient's ability to lead a normal life. It is a progressive neurodegenerative disorder that is often challenging to detect in its early stages. Traditional methods of diagnosing Parkinson's disease using EEG (electroencephalogram) data involve laborious and time-consuming manual feature extraction. To address this issue, in this article, we propose a diagnostic method that can be conducted in a medical office assisted by convolutional neural networks (CNN).

Taking into consideration the clinical presentation of the disease, as discussed in previous references, it is possible to diagnose Parkinson's disease by analyzing the dynamics of sketching guided spirals and waves drawn by a patient on a sheet of paper. We used the database published by Zham et al. (2017) [12], which is composed by a set of images of sketches labeled by healthcare professionals into categories of healthy patients and patients with Parkinson's disease.

Table 1. Summary table of the results found in the literature for the detection of Parkinson's disease. The table specifies the techniques and specific tasks reported in each case.

Reference	Tasks	Methodology	ACC
Vatsaraj and Nagare [11]	Spiral and wave analysis	CNN	96.7%
Zham et al. [12]	Guided spiral	Naïve Bayes with features in handwriting	93.3%
Gallicchio et al. [3]	Spirals and stability movement	Deep Echo State Networks	89.3%
Gil Martín et al. [4]	Spirals and stability movement	CNN	96.5%
Khatamino et al. [7]	Spirals and stability movement	CNN	72.5%
Chakraborty et al. [1]	Spirals and wave analysis	CNN	93.3%

1.1 Related Work

The prevailing method extensively employed by medical professionals in clinical settings for diagnosing Parkinson's Disease involves assessing and evaluating patients through a review of their medical history. This assessment often leads to the assignment of a rating scale based on the patient's performance. The predominant rating system in use to date is the Unified Parkinson Disease Rating Scale (UPDRS), as introduced by Goetz and Stebbins in 2004. [5]. Based on the work done by Goets and Stebbins, Sa, W. et al. in 2003 [9] argued that bradykinesia and other motor symptoms play a fundamental roll in opportune clinic Parkinson´s diagnosis.

Numerous studies have highlighted the direct connection between Parkinson's disease and symptoms related to motor function disorders, such as rigidity, tremors, and bradykinesia. Rigidity and bradykinesia are often evident in the early stages of the disease and affect a patient's ability to write and sketch. Research indicates that an individual's handwriting is influenced by factors like education, knowledge, and language proficiency (Zham et al., 2017) [12].

In contrast, the sketching of spiral and wave drawings serves as independent and noninvasive measures. Extracting features from handwritten sketches can be dynamic, facilitating real-time and dependable analysis. This approach also allows for the development of applications capable of extracting these features through online patient assessments.

Some related work that makes use of Machine Learning related techniques are referenced in Table 1 1. One of the most outstanding results found in the state-of-the-art is the research made by Gil Martín et al. (2019) [4]. This study contributes to this endeavor by examining the application of a convolutional neural network (CNN) for the detection of Parkinson's disease based on drawing movements. The CNN comprises two key components: feature extraction, which involves convolutional layers, and classification, implemented through fully connected layers.

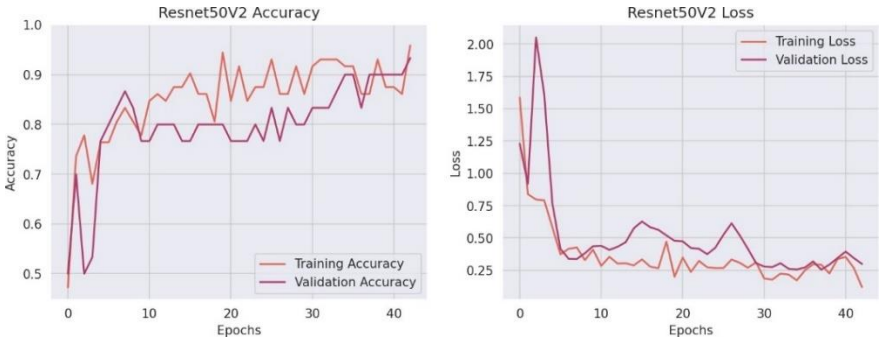

Fig. 1. Results for ResNet50.

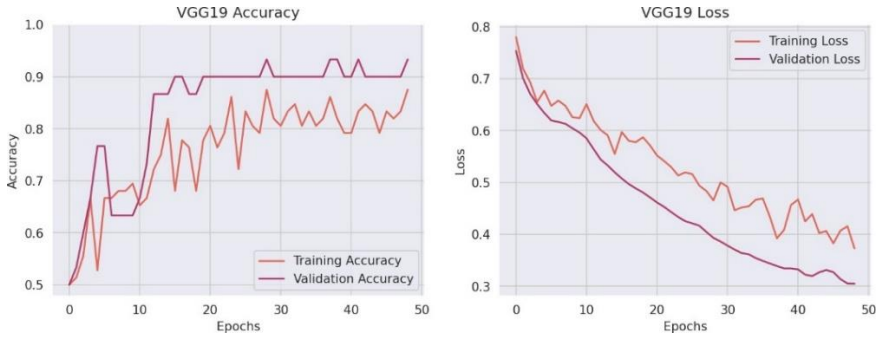

Fig. 2. Results for VGG19.

The inputs to the CNN consist of the Fast Fourier Transform module, focusing on frequencies within the 0 Hz to 25 Hz range.We assessed the discriminative capacity of various directions during drawing movements and found that the X and Y directions yielded the most favorable outcomes. This analysis was conducted using a publicly available dataset: the Parkinson Disease Spiral Drawings Using Digitized Graphics Tablet dataset. The most noteworthy results obtained in this study demonstrate an accuracy of 96.5%. In other hand, the top result found in literature is obtained by Vatsaraj and Nagare (2021) [11].

In this research, the team explore this biomarker by scrutinizing the sketching patterns evident in spiral and wave drawings produced by both healthy subjects and Parkinson's disease patients. Additionally, this study introduces optimizations to algorithms for feature extraction and classification. Notably, the proposed model exhibits an accuracy of 96.67%, alongside a precision of 93.33% and a recall of 100%. Additionally, Parkinson's disease (PD) is a neurodegenerative condition characterized by frequently changing motor symptoms.

The effective monitoring of these symptoms is crucial for tailoring treatment to individual patients. Donié et al. (2023) [9] mention that traditional time series classification (TSC) and deep learning methods have limitations when applied to PD symptom monitoring using data from wearable accelerometers. This is primarily due to the complexity of PD movement patterns and the limited size of datasets. In the state-of-the-art, we will find some achievements boarding various techniques. The

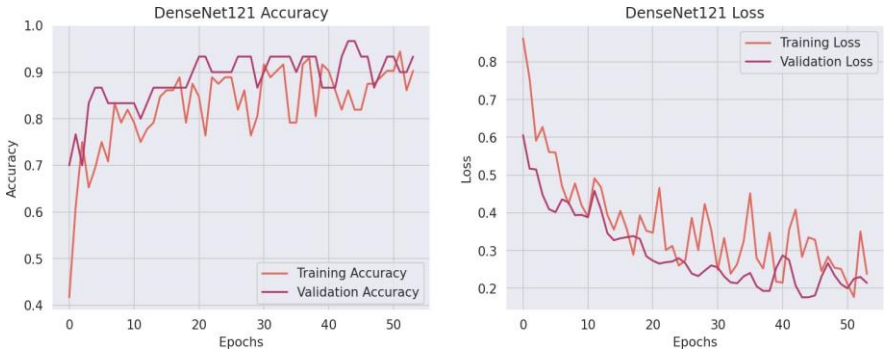

Fig. 3. Results for DenseNet121.

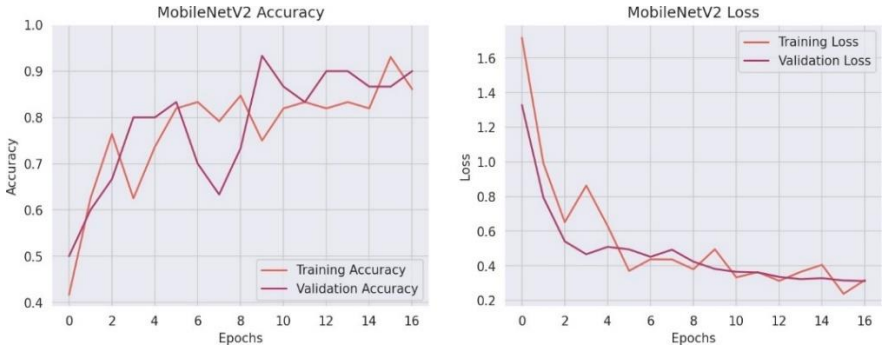

Fig. 4. Results for MovileNetV2.

mentioned techniques take advantage of specific aspects seen on the Parkinson's clinical picture. For instance, Taylan et al. (2020) [10], mention that there has been a growing promise in the use of various statistical regression models for diagnosing neurodegenerative conditions such as Parkinson's disease.

Nevertheless, when experimental data includes outlier observations that significantly deviate from the rest of the data points, traditional and widely recognized statistical regression models can yield inaccurate results for neurodegenerative disease diagnosis. In that sense, Dabbabi et al. (2023) [2], considered vocal cord disorders that are often considered a prominent contributor to Parkinson's disease in many individuals, with speech impairments serving as one of the initial indicators of this condition.

In their study, they propose a model using VOT-MFCC as the primary feature and employs a Fully-Connected Deep Neural Network (FC-DNN) as the classifier; with encouraging results. Also, in the matter of Parkinson's detection via Machine Learning, in the state-of-the-art we might find promising results. Kumar et al. (2023) [8] reported Deep Learning techniques proposed as a means to streamline the Parkinson's detection process and enhance accuracy.

YAMNet, a computationally efficient deep-learning model designed for audio categorization, was employed to extract features from a speech signals dataset related to Parkinson's disease. The study assessed the effectiveness and precision of

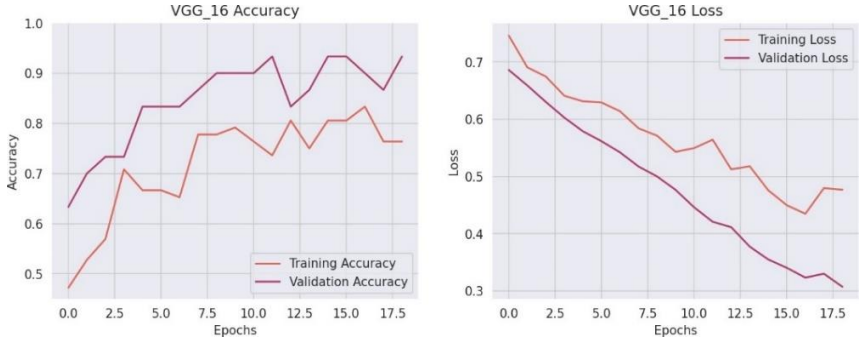

Fig. 5. Results for VGG 16.

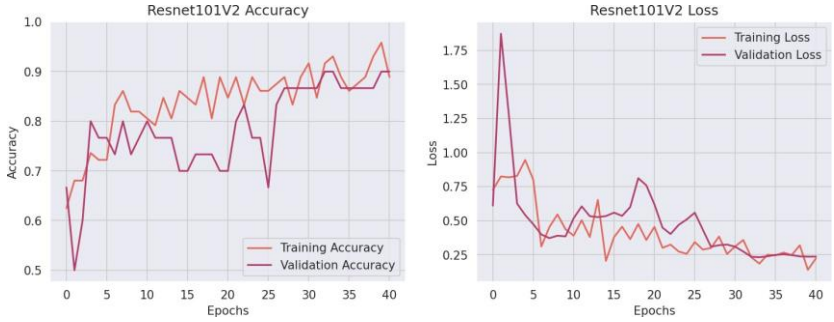

Fig. 6. Results for ResNet 101.

the predictions. By analyzing speech signals, the research aimed to develop a precise and efficient tool for early detection and management of the disease, achieving an accuracy rate near to 82%. This underscores the potential of using speech signals as a diagnostic tool for Parkinson's disease.

In the same way, Gomez et al. (2023) [6], mentioned that Patients afflicted with Parkinson's disease (PD) commonly exhibit reduced facial movements, so in their study, they use three distinct approaches are explored to model the facial expressions of individuals with PD: (i) facial analysis using single images as well as sequences of images, (ii) employing transfer learning from facial analysis to recognize action units, and (iii) implementing triplet-loss functions to enhance the automated classification of PD patients and healthy subjects. The investigators also reported 82% as their best accuracy.

2 Methods

2.1 Data Augmentation

The dataset used in this work corresponds to the results achieved by Zhan et al. (2017) [12]. In this research, the study introduces a novel approach by utilizing the Composite Index of Speed and Pen-pressure (CISP) from sketching as a potential feature for

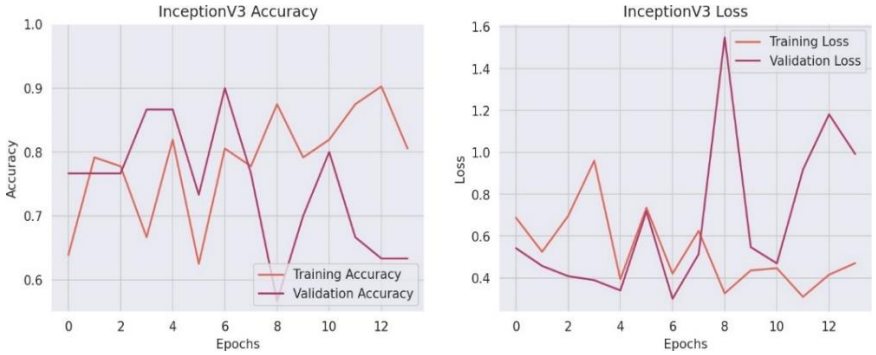

Fig. 7. Results for Inception V3

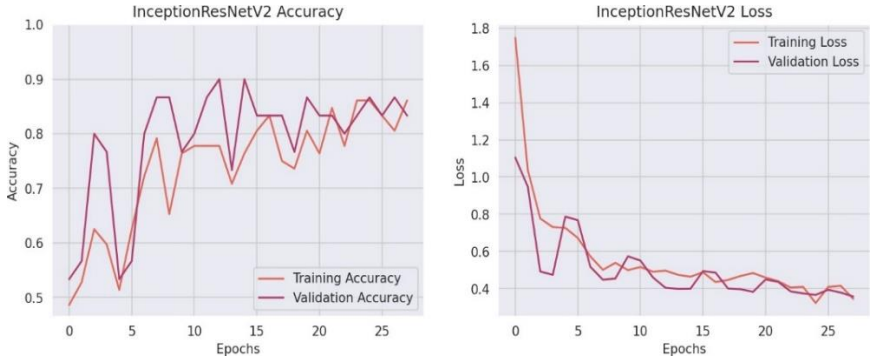

Fig. 8. Results for Inception ResNet.

assessing the severity of Parkinson's disease (PD). The study involved a total of 55 participants, comprising 28 individuals in the control group (CG) and 27 PD patients. The results include a public database composed of 102 images of sketched waves categorized in two classes: Parkinson and healthy. Given the limited size of the database, data augmentation techniques in the context of image data preprocessing for deep learning were applied.

The first step involves rescaling the pixel values of the images to a standardized range of [0, 1], a routine pre-processing step to ensure that the neural network receives data in a consistent format. The rotation range parameter allows each image to be randomly rotated within a range of 40 degrees in both clockwise and counterclockwise directions. This augments the dataset by introducing variations in object orientations, simulating the real-world diversity of image capture.

Furthermore, the width shift range and height shift range parameters permit horizontal and vertical shifts of up to 20% of the total image width and height, respectively. These mimics the effects of different object placements within the frame, making the model more adaptable to such changes.

Shear transformations are introduced with the shear range parameter, allowing for slanting along the horizontal axis within a 20% range. This emulates perspective changes and adds to the dataset's diversity. Zooming in and out, a common source of

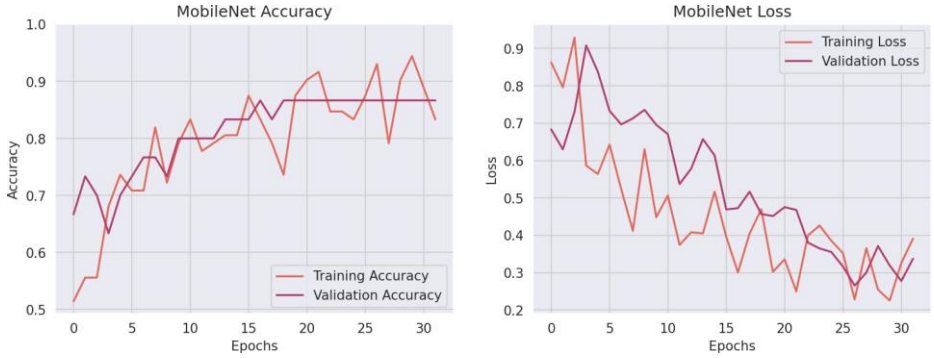

Fig. 9. Results for MobileNet.

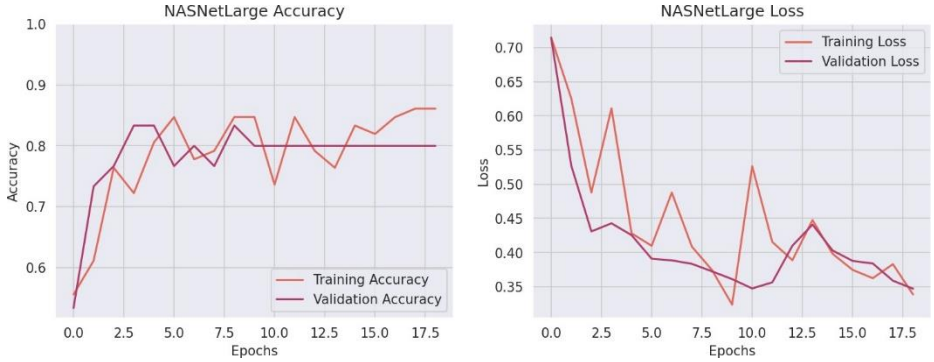

Fig. 10. Results for NasNet large.

variation in real-world images, is achieved with the zoom range parameter set to 20%. This helps the model handle variations in object size and distance. Finally, horizontal flip is enabled, enabling random horizontal flipping of images.

This is useful for scenarios where objects can appear in a mirrored orientation. In summary, these data augmentation techniques are invaluable for enhancing the performance and robustness of deep learning models, particularly when faced with limited training data. By simulating various real-world conditions and image variations during training, the model becomes better equipped to generalize its learning to unseen data, leading to more accurate and reliable predictions in practical applications.

2.2 Transfer Learning with Early Stop Implementation for Binary Classification

A deep learning model tailored for binary classification has been carried out using transfer learning. It leverages the architectures ResNet50, VGG19, DenseNet121, MobileNetV2, VGG16, ResNEt101, InceptionV3, InceptionResNEtV2, MobileNEt, NasNetLarge and ConvNeXtLarge. Those arquitectures come pre-equipped with weights from ImageNet. In all cases, the last 30% layers of the respective architecture were unfrozen for the training stage, while the other layers remained frozen. Additionally, "top" layers were added at the end of the network for binary

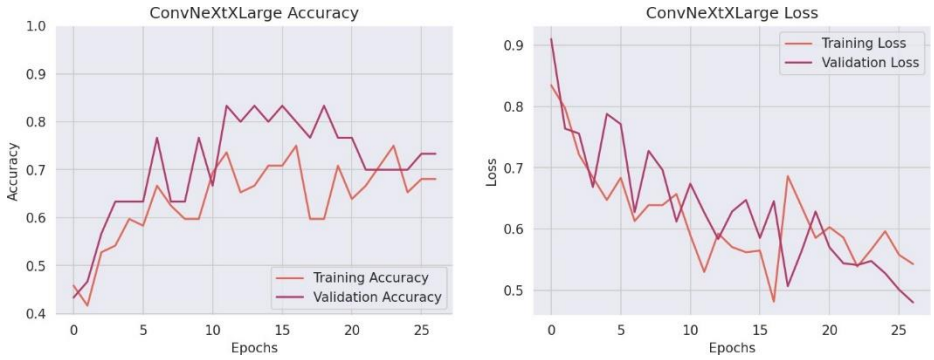

Fig. 11. Results for ConvNext large.

classification. The purpose of this model is to serve as a feature extractor. The base model is then subjected to a process known as fine-tuning, wherein a selective set of its last 30% layers is unfrozen for further training.

This selective approach allows the upper layers to adapt to the task's specific requirements while keeping the remaining layers frozen, preserving the knowledge learned from ImageNet. To form the complete classification model, the base model is integrated into a Sequential architecture. This is achieved through a series of operations, commencing with a Flatten layer, which reshapes the feature maps obtained from the base model into a one-dimensional vector. This is followed by the introduction of two.

Dense layers; the first one incorporates 256 units and a ReLU activation function, while the final layer consists of a single unit with a sigmoid activation function, suitable for binary classification. The Adam optimizer was set with a specified learning rate, the utilization of binary cross-entropy as the loss function, and the adoption of accuracy as the evaluation metric. Finally, an early stopping mechanism is implemented, serving as a safeguard against overfitting by monitoring the validation loss and terminating the training process if the loss fails to improve over a predetermined number of epochs. The mechanism ensures that the model is restored to its most optimal state during training.

3 Results

In Table 2, a summary of the numerical results is found. The performance graphs for loss and accuracy during training and validation stages are presented. The obtained results evaluate the performance of various Convolutional Neural Network (CNN) algorithms for a specific task. The task involved the classification of data, and the metrics used for evaluation included both training and validation accuracy, as well as training and validation loss, along with the corresponding best epoch.

4 Discussion

In table 2, we find the results obtained for train accuracy, validation accuracy, train loss, validation loss, and best epoch. We observe that the best results for validation

Table 2. Results obtained after the model implementation for different model architectures.

CNN Algorithms	Train ACC	Validation ACC	Train Loss	Validation Loss	Best Epoch
ResNet50	0.9538	0.9333	0.1202	0.2997	38
VGG19	0.8750	0.9333	0.3729	0.3049	29
DenseNet121	0.8194	0.9333	0.4092	0.2418	24
MobileNetV2	0.7500	0.9333	0.4959	0.3821	10
VGG16	0.7361	0.9333	0.5642	0.4207	12
ResNet101	0.8889	0.9000	0.2230	0.2359	34
Inceptionv3	0.8056	0.8999	0.4205	0.3009	7
InceptionResNetV2	0.7778	0.8999	0.4959	0.4038	13
MobileNet	0.7917	0.8666	0.4480	0.4726	17
NastNetLarge	0.7222	0.8333	0.6110	0.4428	4
ConvNeXtXLarge	0.7361	0.8333	0.5299	0.6267	12

accuracy were achieved after implementing the model in conjunction with the ResNet50 architecture and the specifications described in the methodology section (2). Regarding the state-of-the-art, Table 1 shows that the best result obtained for the wave analysis task is reported by Vatsaraj and Nagare (2021) [11] with a 96.7% accuracy.

In comparison to the results achieved in this study, the absolute difference between the two results corresponds to 3.27%. Additionally, the same absolute difference is observed after the implementation of the model combined with the architectures VGG19, DenseNEt121, MobileNetV2 and VGG16. Considering the dataset's size, the results reveal consistency in the model. Nevertheless, it is necessary to highlight the fact that the best results in this study were obtained after implementing fewer complex architectures, which is consistent with what was reported by Vatsaraj and Nagare (2021) [5].

Therefore, for future research, it is suggested to use or build networks with fewer convolutional layers in order to get closer to the 96.7% reported in the state-of-the-art. ResNet50, which is known for its depth, demonstrated the highest training accuracy at 95.38%. This suggests that the model learned the training data effectively and can capture complex patterns. However, it exhibited a slightly lower validation accuracy of 93.33%, indicating that it may have encountered some overfitting, as the validation accuracy is slightly lower than the training accuracy.

The training and validation losses of ResNet50 were 0.1202 and 0.2997, respectively, and the best epoch was achieved at 38. These results indicate a good balance between model complexity and generalization, making ResNet50 a strong candidate for this task. VGG19, another deep architecture, demonstrated a relatively high training accuracy of 87.50% but reached an even higher validation accuracy of 93.33%. This suggests that VGG19 achieved good generalization performance. The training and validation losses were 0.3729 and 0.3049, respectively, and the best epoch occurred at 29.

These results indicate that VGG19 managed to generalize well without overfitting,

making it a promising choice for this task. DenseNet121 exhibited an 81.94% training accuracy and a 93.33% validation accuracy. Its training and validation losses were 0.4092 and 0.2418, and the best epoch was at 24. DenseNet121 demonstrated robust generalization performance with a slightly lower training accuracy but consistent validation accuracy, suggesting its effectiveness in capturing relevant features.

MobileNetV2, VGG16, and ResNet101 showed varying performance with training and validation accuracies of 75.00% to 88.89%. MobileNetV2 had the highest training loss, indicating room for improvement in its ability to capture features effectively. On the other hand, VGG16 and ResNet101 showed better balance in terms of losses. Inceptionv3, InceptionResNetV2, MobileNet, NastNetLarge, and ConvNeXtXLarge exhibited performance below the 90% accuracy threshold. These models may require further optimization or modifications to improve their classification capabilities.

The results provide insights into the suitability of different CNN architectures for the given task. ResNet50 and VGG19 demonstrated strong performance, while other models showed varying degrees of success. The choice of the best model depends on the specific trade-off between training and validation accuracy, as well as considerations of overfitting and generalization. Further investigations and fine-tuning may be necessary to enhance the performance of some models.

5 Conclusion

Parkinson's disease is a common neurological condition that can significantly disrupt a patient's ability to lead a normal life. It is a progressive neurodegenerative disorder that is often challenging to detect in its early stages. Traditional diagnostic methods, particularly those using electroencephalogram (EEG) data, are often time-consuming and challenging to apply in the early stages of the disease. Parkinson's disease is associated with distinctive motor symptoms, we introduced an approach that analyzes sketching patterns in guided spirals and waves drawn by patients. By applying CNNs, we aimed to facilitate a more efficient and accessible diagnostic process.

This work was based on a dataset provided by Zham et al. (2017) [12] that included sketches categorized as Parkinson's or healthy. To address the limited size of the dataset, we employed data augmentation techniques, preparing the images for deep learning analysis. Transfer learning was then applied using pre-trained CNN architectures, where the top layers were tailored for binary classification. Regarding the results obtained, ResNet50 and VGG19 exhibited strong performance.

ResNet50 achieved the highest training accuracy. This suggests the model's ability to capture complex patterns yet, the validation accuracy was slightly lower, indicating the possibility of overfitting. On the other hand, VGG19 showed a good balance between training and validation accuracy.

By another hand, DenseNet121 demonstrated robust generalization performance, making it a promising choice for the Parkinson´s diagnosis. Also, the experimentation showed that MobileNetV2, VGG16, and ResNet101 showed varying performance but indicated potential for improvement. Finally; Inceptionv3, InceptionResNetV2, MobileNet, NastNetLarge, and ConvNeXtXLarge had accuracies below the desired threshold. This suggests that these models may require further optimization or modifications to enhance their classification capabilities.

Considering the previous and comparing the found results to the state-of-theart, we found a discrepancy of 3.27% in accuracy. Notably, Vatsaraj and Nagare (2021) [11] achieved an accuracy of 96.7%. To approach this level of accuracy in future research, it is suggested that models with fewer convolutional layers be explored.

References

1. Chakraborty, S., Aich, S., Jong-Seong-Sim, Han, E., Park, J., Kim, H.: Parkinson's disease detection from spiral and wave drawings using convolutional neural networks: A multistage classifier approach. In: Proceedings of the 22nd International Conference on Advanced Communication Technology, pp. 298–303 (2020) doi: 10.23919/icact48636.2020.9061497
2. Dabbabi, K., Kehili, A., Cherif, A.: Parkinson detection using VOT-MFCC combination and fully-connected deep neural network (FC- DNN) classifier. In: Proceedings of the IEEE International Conference on Advanced Systems and Emergent Technologies, pp. 1–6 (2023) doi: 10.1109/ic_aset58101. 2023.10150791
3. Gallicchio, C., Micheli, A., Pedrelli, L.: Deep echo state networks for diagnosis of parkinson's disease. In: Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, pp. 1–6 (2018) doi: 10.48 550/ARXIV.1802.06708
4. Gil-Martín, M., Montero, J. M., San-Segundo, R.: Parkinson's disease detection from drawing movements using convolutional neural networks. Electronics, vol. 8, no. 8, pp. 907 (2019) doi: 10.3390/electronics8080907
5. Goetz, C. G., Stebbins, G. T.: Assuring interrater reliability for the UPDRS motor section: utility of the UPDRS teaching tape. Movement Disorders, vol. 19, no. 12, pp. 1453–1456 (2004) doi: 10.1002/mds.20220
6. Gomez, L. F., Morales, A., Fierrez, J., Orozco-Arroyave, J. R.: Exploring facial expressions and action unit domains for parkinson detection. PLOS ONE, vol. 18, no. 2, pp. e0281248 (2023) doi: 10.1371/journal.pone.0281248
7. Khatamino, P., Canturk, I., Ozyilmaz, L.: A deep learning-CNN based system for medical diagnosis: an application on parkinson's disease handwriting drawings. In: Proceedings of the 6th International Conference on Control Engineering and Information Technology, pp. 1–6 (2018) doi: 10.1109/ceit.2018.8751879
8. Kumar, S. A., Sasikala, S., Arthiya, K. B., Sathika, J., Karishma, V.: Parkinson's speech detection using YAMNet. In: Proceedings of the 2nd International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation, pp. 1–5(2023) doi: 10.1109/icaeca56562.2023. 10200704
9. Williams, S., Wong, D., Alty, J. E., Relton, S. D.: Parkinsonian hand or clinician's eye? Finger tap bradykinesia interrater reliability for 21 movement disorder experts. Journal of Parkinson's Disease, vol. 13, no. 4, pp. 525–536 (2023) doi: 10.3233/jpd-223256
10. Taylan, P., Yerlikaya-Özkurt, F., Uçak, B. B., Weber, G.: A new outlier detection method based on convex optimization: application to diagnosis of parkinson's disease. Journal of Applied Statistics, vol. 48, no. 13-15, pp. 2421–2440 (2020) doi: 10.1080/02664763. 2020.1864815
11. Vatsaraj, I., Nagare, G.: Early detection of Parkinson's disease using contrast enhancement techniques and CNN. International Journal of Engineering Research and Technology, vol. 10, no. 5, pp. 295–298 (2021) doi: 10.17577/IJERTV 10IS050187
12. Zham, P., Kumar, D. K., Dabnichki, P., Arjunan, S. P., Raghav, S.: Distinguishing different stages of Parkinson's disease using composite index of speed and pen-pressure of sketching a spiral. Frontiers in Neurology, vol. 8 (2017) doi: 10.3389/fneur.2017.00435