MobileNet-v2 Enhanced Parkinson's Disease Prediction with Hybrid Data Integration

Sameer Bhat
Multimedia Systems Department, Gdansk University of Technology
Gdansk 80-233, Poland *sameer.bhat@pg.edu.pl*

Piotr Szczuko
Multimedia Systems Department, Gdansk University of Technology
Gdansk 80-233, Poland *piotr.szczuko@pg.edu.pl*

Abstract

This study investigates the role of deep learning models, particularly MobileNet-v2, in Parkinson's Disease (PD) detection through handwriting spiral analysis. Handwriting difficulties often signal early signs of PD, necessitating early detection tools due to potential impacts on patients' work capacities. The study utilizes a three-fold approach, including data augmentation, algorithm development for simulated PD image datasets, and the creation of a hybrid dataset. MobileNet-v2 is trained on these datasets, revealing a higher generalization or prediction accuracy of 84% with hybrid datasets. Future research will explore the impact of high variability synthetic datasets on prediction accuracies and investigate the MobileNet-v2 architecture's memory footprint for timely inferences with low latency.

Keywords: Handwriting analysis, Archimedean Spiral, Parkinson's Disease, MobileNet-v2, Hybrid datasets.

1. Introduction

Parkinson's Disease (PD) is a neurodegenerative disorder that gradually deteriorates over time, primarily affecting individuals aged 55 and above. It is characterized by various motor and non-motor symptoms used for diagnosis and various clinical markers. PD affects over one million individuals in the U.S., a number projected to reach 1.2 million by 2030, impacting more than 10 million globally, with about 60,000 new diagnoses annually in the United States alone [1]. PD, characterized by neurodegenerative changes affecting limb behaviour and, in severe cases, cognitive functions, typically affects the elderly and arises from the death of dopaminergic neurons in the brain's substantia nigra [5]. Handwriting difficulties often signal early signs of PD, affecting millions worldwide. Recent studies have investigated handwriting changes given its widespread nature [9, 36, 39, 5]. While clinical diagnosis traditionally relies on symptom examination and history, the evolving healthcare landscape necessitates early detection tools, especially as initial signs may appear before retirement, impacting patients' work capacities. Various studies have focused on PD diagnosis through handwriting spiral analysis, highlighting the role of machine learning (ML) and deep learning (DL). ML demonstrates promise in PD identification through such analyses, aiding in early detection, monitoring, and improving patient outcomes [12, 20, 3]. Although not a PD diagnostic criterion, handwriting reflecting cognitive, visual, and motor abilities often prompt medical attention. Studies explore ML's application to diagnose PD severity, propose enhanced algorithms, and leverage MRI scans and DL for accurate classification [10, 37, 7, 17]. Other approaches use ML to predict PD with non-motor symptoms, extract features from digital

drawing tests, or apply convolutional neural networks (CNNs) for end-to-end processing of handwriting images [36, 35, 11]. The Unified Parkinson's Disease Rating Scale (UPDRS) and its Movement Disorder Society revision (MDS-UPDRS) serve as standards for clinical PD assessments [13, 23]. However, clinical diagnosis accuracy varies, leading to exploration of alternative methods such as computer-based handwriting analysis [9, 5, 38, 26, 10, 2, 16]. Handwriting alterations in PD, notably micrographia, show potential as biomarkers [33, 6].

Recent research in PD diagnosis shows significant advancements in leveraging DL and artificial intelligence (AI) techniques for improved accuracy. Studies such as [18] demonstrate the effectiveness of DL and AI, especially in classifying datasets with high accuracy. Unique perspectives, such as the kinematic and geometric features emphasized by [25], contribute to disease detection methodologies. Additionally, novel techniques like the multi-pooling approach proposed by [34] offer lightweight yet accurate classification methods addressing gender and PD classification. Advancements in deep CNNs have shown superior performance in PD detection, as evidenced by studies such as [24]. Modalities such as handwriting and hand-drawn images, highlighted by [8], play pivotal roles in leveraging computer vision and machine learning techniques for detection. Furthermore, research exploring various modalities, such as EEG signals analyzed by [22], contributes to advancing PD detection methodologies. Collectively, these studies underscore diverse approaches and modalities that surpass previous benchmarks, highlighting the potential of innovative technologies in improving diagnostic accuracy [9, 4]. Recent strides emphasize the potential of ML and innovative technologies for diagnosis and severity assessment, as demonstrated by [37]. Additionally, research by [7] and [17] introduce improved algorithms and classification models. In contrast, Gazda et al. [10] present an ensemble of deep-learning architectures achieving remarkable accuracy for spiral drawing tasks. Researchers have also explored neurocognitive features, as demonstrated by the studies [31] and [35], revealing promising avenues for automated neurodegenerative disease detection. Novel methods for early detection, such as the end-to-end CNN approach proposed by [11] and the Continuous Convolution Network (CC-Net) introduced by [21], further advance PD diagnostics. The aforementioned studies underscore the evolving landscape of PD detection, showcasing the effectiveness of ML across various modalities. The proposed methods demonstrate potential improvements in accuracy, accessibility, and early diagnosis, reflecting a promising direction for future research in the field [38, 5]. As ML and DL algorithms continue to evolve, they hold substantial promise in PD detection, offering precise disease identification and opportunities for dynamic disease monitoring and treatment evaluation [12, 20, 3].

1.1. Research Motivation

ML and DL algorithms show promise in PD detection through handwriting analysis. However, achieving high accuracy requires extensive datasets for training DL models, posing a challenge for ML techniques. While ML methods accurately distinguish PD patients from healthy individuals based on handwriting, their predictive capabilities are limited by dataset size and complexity. In contrast, DL models excel at uncovering intricate patterns but necessitate large datasets for optimal accuracy. Transitioning to larger datasets for training the DL models is crucial for maximizing their potential in PD diagnosis. These datasets enable deep learning algorithms to learn from a broader range of data patterns, enhancing prediction accuracy. Combining the DL with extensive datasets can enhance PD detection capabilities, potentially surpassing conventional ML approaches. Therefore, investing in large-scale data collection efforts is essential to advance PD diagnosis research.

1.2. Study outline

This study explored training the MobileNet-v2 model on small datasets of hand-drawn and synthetically generated spiral pattern images, revealing challenges in achieving high performance. Data augmentation expanded the Parkinson's dataset and trained the MobileNet-v2 model [28], offering insights across three datasets. An algorithm simulated hand-drawn spiral patterns to generate larger datasets, reducing errors and improving accuracies. A hybrid dataset, combining real Parkinson's samples with simulated data, enhanced model performance, demonstrating its adaptability.

1.3. Research contribution

The study uses synthesized Parkinson's datasets to train the MobileNet-v2 model, showing different accuracies on test datasets. It proposes an algorithm to simulate hand-drawn spiral patterns, creating large training datasets. This approach achieves high accuracies of 99.67% on simulated test datasets and 84.45% on hybrid datasets. By training and evaluating with hybrid datasets, the study demonstrates improved prediction performance of MobileNet-v2 models compared to previous methods, particularly in handling variations across larger datasets.

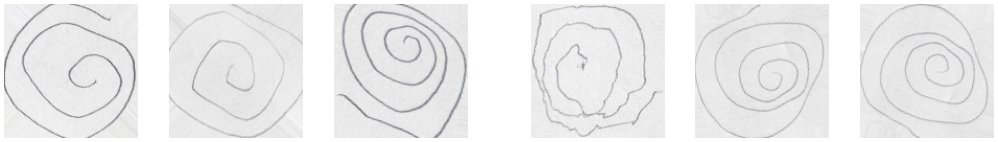
1.4. Study organization

Section 2 outlines the research approach and data collection methods, while Section 2.4 discusses the rationale for selecting the MobileNet-v2 architecture. In Section 2.5, the configuration and design of the MobileNet-v2 model are detailed, followed by an explanation of data augmentation techniques in Section 2.1. Section 2.2 analyzes strategies for distributing datasets, while Section 2.3 explains the generation of simulated data. Model training procedures are outlined in Section 2.6, and results are presented in Section 3. Section 4 interprets the findings, and Section 5 summarizes key insights and future directions.

2. Method

The study employs a three-fold approach to evaluate the DL model performance based on the Parkinson's Image data set (PID) [40], hereafter referred to as PID, as the input to the model. The PID consists of 72 training and 30 testing image samples of spiral and wave curves. This study only employs an image set consisting of spiral curves and leaves out a wave data set for future works.

The first fold employs the PID data set referred to as Pd_aug_ds or D_{au} containing training and test images. Due to the low sample count, the data set remains inadequate for training the MobileNet-v2 model. As such, we employ data augmentation to increase the dataset size; the datasets are balanced regarding training and testing samples. The second fold includes developing an algorithm for generating a simulated PD image dataset, hereafter referred to as sim_aug_ds or D_{sm}. The third fold includes generating a mixed or hybrid dataset hereafter referred to as D_{au+sm} comprising images from D_{au} and D_{sm}. Next to creating the three datasets, we split them into training, validation, and testing datasets for evaluating MobileNet-v2 model's performance. We train the MobileNet-v2 model using the three different data sets D_{au}, D_{sm} and D_{au+sm}, and report the model training validation and test accuracies.



(a) Healthy (b) Parkinson's

Fig. 1. Samples of real spiral images from augmented dataset (D_{au}) representing hand-drawn spirals of healthy and Parkinson's classes.

(a) Healthy (b) Parkinson's

Fig. 2. Samples of spiral images from synthetically generated dataset (D_{sm}) representing hand-drawn spirals of healthy and Parkinson's classes.

2.1. Data Augmentation

During model training, the TensorFlow Keras ImageDataGenerator augmented image data, enhancing dataset diversity and size. Various transformations like rotation, shifting, flipping, and resizing improved model robustness and generalization, especially with limited labelled data. Figures 1 and 2 depict real Parkinson's dataset and synthetically created spiral images, respectively. Parameters were adjusted for effective augmentation: 360-degree rotation ($rotation_range = 360$), horizontal and vertical flipping ($horizontal\_flip = True$, $vertical\_flip = True$), and 0.1 horizontal and vertical shifts ($width\_shift\_range = 0.1$, $height\_shift\_range = 0.1$). Although not used, *brightness_range* could adjust brightness. Varying intensity, rather than constant grayscale, enabled the model to recognize patterns. These settings boosted dataset augmentation, improving model adaptability and generalization.

2.2. Data Distribution

The data distribution process partitioned datasets into training, validation, and test sets. For the augmented Parkinson's dataset, both training and test sets were balanced, with 2556 samples per class in training and 465 in the test set. Similarly, the simulated dataset had 7100 samples per class in training and 620 in the test set. The hybrid dataset combined samples from both datasets, resulting in a balanced training set with 4686 samples and a test set with 775 samples per class. This approach ensured that the model was trained on diverse and representative data. The training data was split into training and validation using a 90:10 ratio. Table 1 displays the data split.

Table 1. Data distribution into training, validation, and testing datasets.

Dataset	Training set	Testing set	Split (Training : Validation)
D_{au}	5112	930	4600 : 512
D_{sm}	14200	1240	12780 : 1420
D_{au+sm}	9372	1550	8435 : 937

2.3. Synthetic data generation

Algorithm 1 generates spiral images, with its functionality contingent on the tuning of various parameters, thereby influencing the characteristics of the synthesized dataset. Adjusting parameters like *size*, *num_points*, and coefficients a, b, and k, enabled control over the size, complexity, and shape of the generated spirals. Moreover, alter-

Algorithm 1 Generate Spiral Images

Require: $size, num\_points, a, b, k, noise\_level, irregularity\_level, padding,$
$dynamic\_variation, min\_dynamic\_var, max\_dynamic\_var$
Ensure: img
Initialize img as a matrix of size $size \times size$ with all elements set to 255 (white background)
Compute the center of the image as $center = size//2$
Generate an array of equally spaced values for t from 0 to 10π with num_points points
Compute the radius variation as a random value between -0.1 and 0.1
Compute the x and y coordinates of the Archimedes spiral using the formula:
    $x = a + b \cdot t^k + (\text{radius\_variation}) \cdot \sin(t)$
    $y = a + b \cdot t^k + (\text{radius\_variation}) \cdot \cos(t)$
if dynamic_variation is True and the spiral is for Parkinson's disease **then**
    Generate a random value for dynamic_var between min_dynamic_var and max_dynamic_var
    Add dynamic_var to the noise_level
    Add Gaussian noise to the y coordinates with mean 0 and standard deviation noise_level
end if
Add Gaussian noise to the x and y coordinates with mean 0 and standard deviation irregularity_level
Map the coordinates to the image space with padding:
    $x\_img = \text{round}(x \times (size/(4 + \text{padding})) + \text{center})$
    $y\_img = \text{round}(y \times (size/(4 + \text{padding})) + \text{center})$
for i from 0 to $length(x\_img) - 1$ **do**
    if $0 \leq x\_img[i] < size$ **and** $0 \leq y\_img[i] < size$ **then**
        Draw a line from $(x\_img[i], y\_img[i])$ to $(x\_img[i+1], y\_img[i+1])$ in black color
    end if
end for
return The generated image img

ing *noise_level* and *irregularity_level* introduced variability and imperfections akin to real-world data, enhancing the dataset's robustness. The *padding* parameter allows for adjustments to the spatial arrangement of the spiral within the image frame.

2.4. Model Selection

Convolutional Neural Networks have garnered significant attention in image processing for their economic potential and high accuracy rates. Several popular CNN architectures, such as AlexNet [19], InceptionV3 [30], VGG16 [29], ResNet [15], and MobileNet-v2 [27], dominate the field of image processing and classification. While convolution operations play a crucial role in computer vision tasks, the large and deep structures of networks like AlexNet, VGG16, InceptionV3, and ResNet often lead to increased processing time and costs. However, MobileNet-v2 stands out due to its inverted residual structure and linear bottleneck design, which reduce convolution calculations and make it memory-efficient; as such, we adopt the MobileNet-v2 for our PD detection problem and test it on the three different datasets D_{au}, D_{sm} and D_{au+sm}. The model accuracy can be computed by the following expression: Model accuracy $= \frac{TP+TN}{TP+TN+FP+FN}$; where TP, TN, FP, and FN denote the number of true positives, true negatives, false positives, and false negatives, respectively.

2.5. MobileNet-v2–Network Architecture

The MobileNet-v2 [28, 27] architecture represents a significant advancement in efficient CNNs, tailored explicitly for deployment on resource-constrained devices such as mobile phones and embedded systems. MobileNet-v2 employs a novel inverted residual structure at its core, which deviates from the traditional residual models seen in deeper networks like ResNet. Inverted residual blocks consist of thin bottleneck layers at the input and output ends, contrary to the expanded representations typically found in conventional residual architectures. This design choice helps reduce the computational burden while preserving the representational power of the network. The MobileNet-v2 model, outlined by [14], adopts an inverted residual structure featuring thin bottleneck

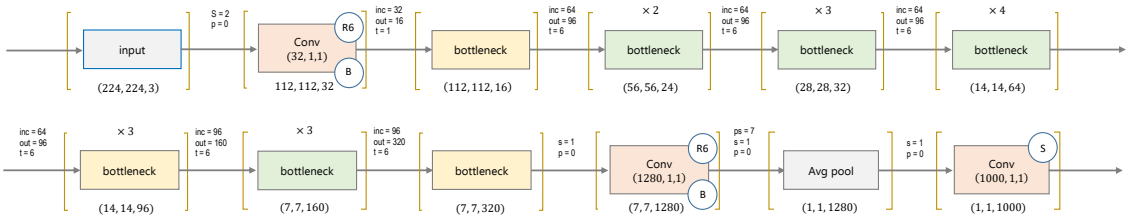


Fig. 3. Architecture of MobileNet-v2 [14]. In the figure, $R6$ denotes the ReLU6 activation function, B denotes batch Normalization, and S denotes Softmax activation function.

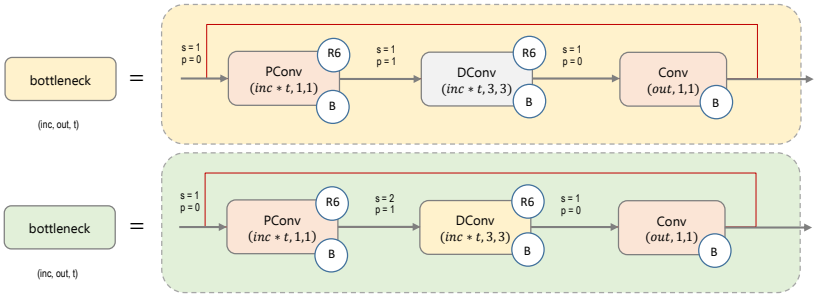


Fig. 4. Structure of bottleneck layer. The legends $DConv, PConv, inc$, and t denote depthwise convolution, pointwise convolution, input channel, and expansion factor respectively [14].

layers at both input and output stages, deviating from traditional residual models. The model employs lightweight depthwise convolutions within these inverted residual blocks to filter features in intermediate expansion layers, capturing spatial dependencies within channels while minimizing parameters and computational complexity. This design choice optimizes model efficiency without sacrificing performance, supported by Figures 3 and 4. Non-linearities are strategically removed from narrow layers to preserve representational power, enhancing computational efficiency. MobileNet-v2 thus strikes a notable balance between efficiency and performance, making it suitable for resource-constrained applications in diverse real-world scenarios.

2.6. Model Training, Validation and Testing

The model training, validation and testing were conducted using Google Colab's TPU-4 with a 2GB RAM configuration. Model testing was specifically carried out with checkpoints during the training and validation process. The loaded checkpoints were tested for high validation accuracies to generate model predictions on datasets (unknown or not used during the training and validation process) reserved for testing only.

3. Results

The study reports the model training and validation results based on the three learning rates used for training the MobileNet-v2 DNN model. The model was evaluated using the cross-entry loss [32] as a performance metric. The CSV file recorded observations enabled by the models trained for 50 Epochs. The models were tuned by varying the learning rates to 0.01, 0.001, and 0.0001, chosen to compare model performance. Figure 5 depicts the results of training and validation losses incurred by the MobileNet-v2 model trained on datasets D_{au}, D_{sm}, and D_{au+sm}. The training and validation loss curves show that the model trained on the hybrid dataset exhibited reduced training losses, indicating lower generalization error estimation. Tuning the model for prediction accuracy within the range of 90% to 100% led to overfitting issues in non-hybrid datasets while achieving a validation accuracy of 0.998 using the hybrid dataset improved generalization to unseen data. Consequently, the model's prediction accuracy on the testing

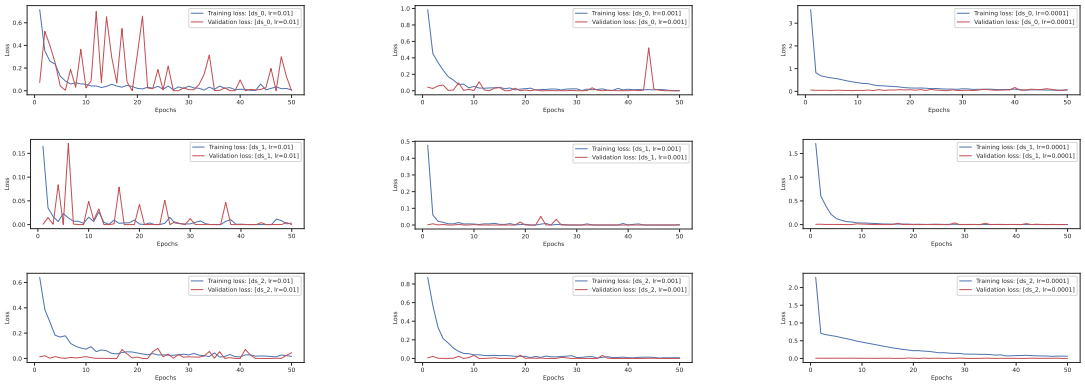


Fig. 5. Training and validation losses for datasets – D_{au}, D_{sm}, and D_{au+sm}. The acronyms ds_0, ds_1 and ds_2 in images correspond to the datasets D_{au}, D_{sm}, and D_{au+sm} respectively. Symbol lr in the legends represent the learning rate for the MobileNet-v2 model.

Table 2. Results of MobileNet-v2 models trained on D_{au}, D_{sm}, and D_{au+sm}.

Input dataset	Hyper Parameter	Training accuracy		Test accuracy on trained models		
Augmented	learning-rate	training	validation	Model-D$_{au}$	Model-D$_{sm}$	Model-D$_{au+sm}$
Parkinsons (D_{au})	0.01	99.89	99.80	74.51	0.50	0.50
	0.001	100.00	99.80	72.58	0.50	75.14
	0.0001	97.60	71.29	59.46	0.50	63.87
Simulated (D_{sm})	0.01	99.89	99.86	53.87	99.35	97.82
	0.001	99.99	100.00	50.32	100.00	99.67
	0.0001	99.98	99.64	54.43	99.83	82.17
Hybrid (D_{au+sm})	0.01	72.73	70.22	68.32	69.80	68.64
	0.001	99.77	99.67	63.48	70.00	84.45
	0.0001	99.34	82.17	57.48	69.67	71.93

set reached approximately 83% to 84%, striking a balance between model complexity and generalization. Figure 6 shows the confusion matrix plots that visualize the training and validation accuracies and test prediction coded on purple and green colour intensities, respectively. Although the training and the validation accuracy appear to be almost the same for the model trained on the three datasets, the testing accuracies vary considerably with the three datasets as input to the trained model. Table 2 shows the prediction results observed on all the three datasets used for training, validation, and testing the MobileNet-v2 network model.

4. Discussion

The MobileNet-v2 architecture exhibited varying behaviours while training on the three datasets. As visualized in the confusion matrix plots, when trained on each dataset, the model architecture showcased different prediction behaviours, except during the training and validation process, wherein the model achieved the best prediction accuracies. Trained models to be applicable in real-world contexts require models to generate highly accurate predictions when testing them against unknown samples. We observed significant differences in the generated inference when testing our model on test datasets of unknown samples. The MobileNet-v2 DNN architecture demonstrated a considerable performance improvement when tested on the simulated (D_{sm})and the Hybrid dataset (D_{au+sm}). In contrast, lower accuracy was observed in the case of the augmented Parkinson's dataset (D_{au}). Since the simulated dataset included a large dataset consisting of spiral image curve traces with a regular increasing radius of the spiral pattern compared to the irregular patterns inherent to the actual D_{au} dataset, the model showed nearly 100% prediction accuracy during training, validation, and testing. Moreover, we com-

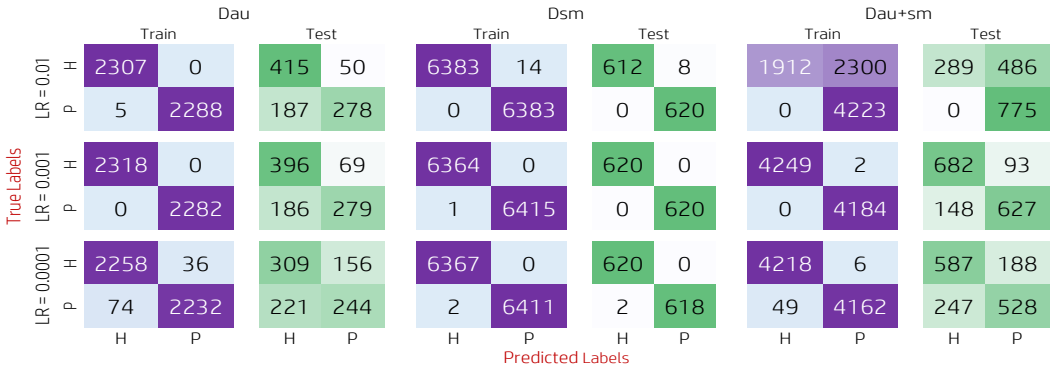


Fig. 6. MobileNet-v2 network model evaluation results on the training datasets, and the testing datasets. H and P denote the healthy and Parkinson's class, respectively. LR represents the learning rate.

pare our study with the work by Wang et al. [38], reporting model accuracy of 89.3% on the custom-developed spiral dataset. Although the reported accuracy is higher than the accuracy of our model, the CNN model used has been tested on 10% of unknown samples, implying that the CNN model has been tested for a low sample size. In contrast, we reserve and test the MobileNet-v2 model on 15% of unknown samples to highlight the robustness and generalization offered by our dataset and the model used. Besides the low sample size, PD prediction accuracy reported in several studies is typically high, reaching above 98% to 99%. Our findings reveal that even after increasing the dataset size, the predictions generated by the model on the hybrid dataset are low compared to predictions generated using only the simulated dataset. Authors in [10] also report that the high accuracy reported in several studies could be a false claim. As such, this study emphasizes reporting the prediction results based on the hybrid datasets.

Key takeaways: Variability in image spiral image characteristics is crucial in the DL model training, directly influencing their ability to generalize to unseen data. More specifically, variations in image attributes such as colour (irregular grayscale patterns), texture, lighting, and orientation contribute to the model's capacity to learn robust representations. When training the DL models, a diverse dataset with a wide range of variations allows the model to learn invariant features that are essential for generalization. In the context of MobileNet-v2 training, image characteristics variability affected the model's ability to extract meaningful features from the data. Images with consistent attributes, such as uniform lighting and simplistic patterns, may lead to model overfitting, where the model memorizes specific training examples without genuinely understanding the underlying patterns. Conversely, introducing variabilities, such as changes in lighting conditions, colour distributions, or geometric transformations, challenges the model to learn more abstract and invariant representations. This variability enhances the model's adaptability to real-world scenarios and improves its performance on unseen data. To mitigate the observed potential overfitting to simulated data, despite the application of data augmentation, we will further enhance our approach by integrating regularization techniques such as L1 and L2 regularization. These methods introduce penalty terms to the model's loss function, discouraging it from excessively fitting noise in the training data and promoting more generalized patterns. Additionally, robust cross-validation methods will be employed, involving data partitioning into multiple subsets for training and validation. By validating our model across diverse real-world datasets, we aim to ensure its reliability and applicability in various practical scenarios. These measures strengthen our efforts to combat overfitting and enhance the model's robust performance.

5. Conclusion

The synthetic dataset generated played a crucial role in MobileNet-v2's training, especially concerning the colour characteristics of the generated images. Initially, we trained the model on synthetic images containing dark black-coloured spiral curves (data points with constant lightning intensity), and the model exhibited low responsiveness. However, by introducing grayscale variations and manipulating lighting conditions of spiral curves, significant impacts on the model's training dynamics were observed. The study only reports the results based on datasets with variability introduced in the images. This adjustment allowed MobileNet-v2 to capture better features related to curve shapes and patterns, ultimately leading to improved performance during training and inference.

This study reports the optimized model behaviours only and leaves out the discussion on training the MobileNet-v2 model on datasets with a higher degree of simulated variability for future research. Initially, we generated multiple synthetic datasets with multi-level variability in terms of noise, radius, and irregularity in the simulated spiral patterns; however, the model exhibited a high amount of training, validation and testing losses (to be reported in future works). In future works, we aim to investigate the impact of high variability synthetic datasets on the prediction accuracies. Moreover, leaving for future works, investigating the impact of the memory footprint of the MobileNet-v2 architecture to generate timely inferences with low latency needs exploration. Since the MobileNet-v2 network can help generate predictions enabling the detection of healthy or Parkinson's patients, how the model learns features by focusing on specific regions of image patches exhibiting the level of variability learned by the model needs further exploration considering the explainability demands in AI. In addition, future research using MobileNet-v2 would be testing on open datasets such as PaHaW and DraWritPD for performance evaluation and generalization.

Furthermore, variability in image characteristics helped prevent the model from learning spurious correlations that may exist in the training data; however, do not generalize to new instances. Exposing the model to a diverse range of image variations during training makes it more robust to noise and irrelevant features, resulting in better performance on unseen data. For future investigations, exploring the impact of dynamic variations, represented by *dynamic_variation*, could yield insights into the sensitivity of diagnostic models to temporal fluctuations in handwriting patterns. Additionally, augmenting the algorithm to accommodate different spiral shapes or incorporating domain-specific knowledge could enhance its applicability across diverse datasets. Lastly, evaluating the algorithm's performance under varying parameter settings could elucidate optimal configurations for specific diagnostic tasks, fostering advancements in automated disease detection systems.

5.1. Acknowledgements

This work was supported by the Faculty of Electronics, Telecommunications, and Informatics, Gdansk University of Technology, Poland.

References

[1] Awards, S.: *Parkinson's foundation better lives. Together. Psa.* `https://rb.gy/vff5qr`. Accessed on March 17, 2024.

[2] Bernardo, L. S., Quezada, A., Munoz, R., Maia, F. M., Pereira, C. R., Wu, W., and Albuquerque, V. H. C. de: Handwritten pattern recognition for early Parkinson's disease diagnosis. In: *Pattern Recognition Letters* 125.0 (2019), pp. 78–84.

[3] Chandra, J., Muthupalaniappan, S., Shang, Z., Deng, R., Lin, R., Tolkova, I., Butts, D., Sul, D., Marzouk, S., Bose, S., Chen, A., Bhaskar, A., Mantena, S., and Press, D. Z.: Screening of Parkinson's Disease Using Geometric Features Extracted from Spiral Drawings. In: *Brain Sciences* 11.10 (2021).

[4] Das, A., Das, H. S., Choudhury, A., Neog, A., and Mazumdar, S.: Detection of Parkinson's Disease from Hand-Drawn Images Using Deep Transfer Learning. In: *Intelligent Learning for Computer Vision*. Ed. by Sharma, H., Saraswat, M., Kumar, S., and Bansal, J. C. Singapore: Springer Singapore, 2021, pp. 67–84.

[5] Das, A., Das, H. S., Neog, A., Bharat Reddy, B., and Swargiary, M.: Performance Analysis of Different Machine Learning Classifiers in Detection of Parkinson's Disease from Hand-Drawn Images Using Histogram of Oriented Gradients. In: *Applications of Artificial Intelligence in Engineering*. Ed. by Gao, X.-Z., Kumar, R., Srivastava, S., and Soni, B. P. Singapore: Springer Singapore, 2021, pp. 205–215.

[6] Dixit, S., Bohre, K., Singh, Y., Himeur, Y., Mansoor, W., Atalla, S., and Srinivasan, K.: A Comprehensive Review on AI-Enabled Models for Parkinson's Disease Diagnosis. In: *Electronics* 12.4 (2023).

[7] Fang, Z.: Improved KNN algorithm with information entropy for the diagnosis of Parkinson's disease. In: *2022 International Conference on Machine Learning and Knowledge Engineering (MLKE)*. Guilin, China: IEEE, 2022, pp. 98–101.

[8] Ferdib-Al-Islam and Akter, L.: Early Identification of Parkinson's Disease from Hand-drawn Images using Histogram of Oriented Gradients and Machine Learning Techniques. In: *2020 Emerging Technology in Computing, Communication and Electronics (ETCCE)*. Bangladesh: IEEE, 2020, pp. 1–6.

[9] Fratello, M., Cordella, F., Albani, G., Veneziano, G., Marano, G., Paffi, A., and Pallotti, A.: Classification-Based Screening of Parkinson’s Disease Patients through Graph and Handwriting Signals. In: *Engineering Proceedings* 11.1 (2021), p. 49.

[10] Gazda, M., Hires, M., and Drotár, P.: Ensemble of convolutional neural networks for Parkinson's disease diagnosis from offline handwriting. In: *IGS2021: The 20th Conference of the International Graphonomics Society*. Las Palmas de Gran Canaria, Spain: Springer, 2022, pp. 1–5.

[11] Gazda, M., Hireš, M., and Drotár, P.: Multiple-Fine-Tuned Convolutional Neural Networks for Parkinson's Disease Diagnosis From Offline Handwriting. In: *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 52.1 (2022), pp. 78–89.

[12] Gil-Martín, M., Montero, J. M., and San-Segundo, R.: Parkinson's Disease Detection from Drawing Movements Using Convolutional Neural Networks. In: *Electronics* 8.8 (2019).

[13] Goetz, C. G., Tilley, B., Shaftman, S. R., Stebbins, G. T., Fahn, S., Martinez-Martin, P., Poewe, W., Sampaio, C., Stern, M. B., Dodel, R., Dubois, B., Holloway, R., Jankovic, J., Kulisevsky, J., and Anthony, N. L.: *MDS-Unified Parkinson's Disease Rating Scale (MDS-UPDRS)*. `https://www.movementdisorders.org/MDS/MDS-Rating-Scales/MDS-Unified-Parkinsons-Disease-Rating-Scale-MDS-UPDRS.htm`. Accessed on March 20, 2024.

[14] HackMD: *MobileNet-V2: Summary and Implementation*. `https://hackmd.io/@machine-learning/ryaDuxe5L`. Accessed on April 01, 2024.

[15] He, K., Zhang, X., Ren, S., and Sun, J.: Deep Residual Learning for Image Recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2016.

[16] Kan, P.-J., Lin, C.-H., Su, C.-S., Lin, H.-Y., Chen, W.-L., and Liang, C.-K.: Polar Expression Feature of Digitized Handwritten Pattern for Automated- Parkinson's-Disease Screening Using Perceptual Color Representation-Based Classifier. In: *IEEE Access* 7.0 (2019), pp. 61738–61755.

[17] Kaplan, E., Altunisik, E., Ekmekyapar Firat, Y., Datta Barua, P., Dogan, S., Baygin, M., Burak Demir, F., Tuncer, T., Palmer, E., Tan, R.-S., Yu, P., Soar, J., Fujita, H., and Rajendra Acharya, U.: Novel nested patch-based feature extraction model for automated Parkinson's Disease symptom classification using MRI images. In: *Computer Methods and Programs in Biomedicine* 224.0 (2022), p. 107030.

[18] Khatamino, P., Cantürk, İ., and Özyılmaz, L.: A Deep Learning-CNN Based System for Medical Diagnosis: An Application on Parkinson's Disease Handwriting Drawings. In: *2018 6th International Conference on Control Engineering & Information Technology (CEIT)*. Istanbul, Turkey: IEEE, 2018, pp. 1–6.

[19] Krizhevsky, A., Sutskever, I., and Hinton, G. E.: Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems* 25 (2012).

[20] Kurt, I., Ulukaya, S., and Erdem, O.: Classification of Parkinson's Disease Using Dynamic Time Warping. In: *2019 27th Telecommunications Forum (TELFOR)*. 2019, pp. 1–4.

[21] Li, Z., Yang, J., Wang, Y., Cai, M., Liu, X., and Lu, K.: Early diagnosis of Parkinson's disease using Continuous Convolution Network: Handwriting recognition based on off-line hand drawing without template. In: *Journal of Biomedical Informatics* 130.0 (2022), p. 104085.

[22] Loh, H. W., Ooi, C. P., Palmer, E., Barua, P. D., Dogan, S., Tuncer, T., Baygin, M., and Acharya, U. R.: GaborPDNet: Gabor Transformation and Deep Neural Network for Parkinson's Disease Detection Using EEG Signals. In: *Electronics* 10.14 (2021), p. 1740.

[23] Movement Disorder Society Task Force on Rating Scales for Parkinson's Disease: The Unified Parkinson's Disease Rating Scale (UPDRS): Status and recommendations. In: *Movement Disorders* 18.7 (2003), pp. 738–750.

[24] Naseer, A., Rani, M., Naz, S., Razzak, M. I., Imran, M., and Xu, G.: Refining Parkinson's neurological disorder identification through deep transfer learning. In: *Neural Computing and Applications* 32.3 (2020), pp. 839–854.

[25] Nõmm, S., Bardõš, K., Toomela, A., Medijainen, K., and Taba, P.: Detailed Analysis of the Luria's Alternating SeriesTests for Parkinson's Disease Diagnostics. In: *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*. 2018, pp. 1347–1352.

[26] Radmard, S., Ortega, R. A., Ford, B., Vanegas-Arroyave, N., McKhann, G. M., Sheth, S. A., Winfield, L., Luciano, M. S., Saunders-Pullman, R., and Pullman, S. L.: Using computerized spiral analysis to evaluate deep brain stimulation outcomes in Parkinson disease. In: *Clinical Neurology and Neurosurgery* 208 (2021), p. 106878.

[27] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C.: MobileNetV2: Inverted Residuals and Linear Bottlenecks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2018.

[28] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C.: *MobileNetV2: Inverted Residuals and Linear Bottlenecks*. 2019. arXiv: `1801.04381 [cs.CV]`.

[29] Simonyan, K. and Zisserman, A.: *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 2015. arXiv: `1409.1556 [cs.CV]`.

[30] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z.: Rethinking the Inception Architecture for Computer Vision. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2016.

[31] Templeton, J. M., Poellabauer, C., and Schneider, S.: Classification of Parkinson's disease and its stages using machine learning. In: *Scientific Reports* 12.1 (2022), p. 14036.

[32] The Linux Foundataion: *PyTorch*. `https://pytorch.org/docs/stable/generated/torch.nn.CrossEntropyLoss.html`. Accessed on April 15,2024.

[33] Thomas, M., Lenka, A., and Kumar Pal, P.: Handwriting Analysis in Parkinson's Disease: Current Status and Future Directions. In: *Movement Disorders Clinical Practice* 4.6 (2017), pp. 806–818.

[34] Tuncer, T. and Dogan, S.: A novel octopus based Parkinson's disease and gender recognition method using vowels. In: *Applied Acoustics* 155.0 (2019), pp. 75–83.

[35] Valla, E., Nõmm, S., Medijainen, K., Taba, P., and Toomela, A.: Tremor-related feature engineering for machine learning based Parkinson's disease diagnostics. In: *Biomedical Signal Processing and Control* 75.0 (2022), p. 103551.

[36] Varalakshmi, P., Tharani Priya, B., Anu Rithiga, B., Bhuvaneaswari, R., and Sakthi Jaya Sundar, R.: Diagnosis of Parkinson's disease from hand drawing utilizing hybrid models. In: *Parkinsonism & Related Disorders* 105.0 (Dec. 2022), pp. 24–31.

[37] Al-Wahishi, A., Belal, N., and Ghanem, N.: Diagnosis of Parkinson's Disease by Deep Learning Techniques Using Handwriting Dataset. In: *Advances in Signal Processing and Intelligent Recognition Systems*. Ed. by Thampi, S. M., Krishnan, S., Hegde, R. M., Ciuonzo, D., Hanne, T., and Kannan R., J. Singapore: Springer Singapore, 2021, pp. 131–143.

[38] Wang, Y., Yang, J., Cai, M., Liu, X., Lu, K., Lou, Y., and Li, Z.: Application of optimized convolutional neural networks for early aided diagnosis of essential tremor: Automatic handwriting recognition and feature analysis. In: *Medical Engineering & Physics* 113.0 (2023), p. 103962.

[39] Yang, T.-L., Lin, C.-H., Chen, W.-L., Lin, H.-Y., Su, C.-S., and Liang, C.-K.: Hash Transformation and Machine Learning-Based Decision-Making Classifier Improved the Accuracy Rate of Automated Parkinson's Disease Screening. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 28.1 (2020), pp. 72–82.

[40] Zham, P., Kumar, D. K., Dabnichki, P., Poosapadi Arjunan, S., and Raghav, S.: Distinguishing Different Stages of Parkinson's Disease Using Composite Index of Speed and Pen-Pressure of Sketching a Spiral. In: *Frontiers in Neurology* 8 (2017), pp. 1–7.