



# An innovative approach for parkinson's disease diagnosis using CNN, NCA, and SVM

Yahya Dogan<sup>1</sup>

Received: 16 January 2024 / Accepted: 29 July 2024

© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2024

## Abstract

Parkinson's disease (PD) is a prevalent neurodegenerative disorder affecting millions of people globally, with substantial health risks and economic burdens. This study aims to introduce an innovative hybrid approach combining deep learning and machine learning algorithms to improve the diagnosis of PD using handwriting dynamics indicative of Parkinson's symptoms. The proposed approach integrates hybrid feature extraction using nine fine-tuned transfer learning models, i.e., InceptionV3, DenseNet201, EfficientNetB0, ResNet50, MobileNetV2, VGG16, Xception, NASNetMobile, and InceptionResNetV2. Initially, features from these models are used individually or in binary and ternary combinations. Given the limited sample size in PD datasets, some extracted features through fine-tuning may lack significance, and fully connected layers can lead to overfitting. To address this issue, Neighborhood Component Analysis (NCA) is employed to refine these features, retaining only the most informative ones. Finally, the selected features are classified using Support Vector Machines (SVM) maximizing the margin between classes and reducing the risk of overfitting. The proposed hybrid model achieves a state-of-the-art accuracy of 99.39% on the Parkinson Hand Drawings dataset. The combination of features extracted from DenseNet201, Xception, and NASNetMobile models, processed using NCA and SVM methods, has been identified as the most efficient model, balancing high accuracy with computational efficiency. Qualitative assessments further confirm the accuracy and reliability of the approach.

**Keywords** Parkinson's disease · Convolutional neural network · Feature extraction · Machine learning · Support vector machine · Neighborhood component analysis

## Abbreviations

PD	Parkinson's disease	DST	Dynamic spiral test
NCA	Neighborhood component analysis	SST	Static spiral test
SVM	Support vector machine	HOG	Histogram of oriented gradients
CNN	Convolutional neural network	MLP	Multi-layer perceptron
FP	Fixed-point	BiGRU	Bidirectional gated recurrent unit
CIT	Carbamide-Isoquinoline-tracer	FRCNN	Faster region-based convolutional neural network
PPMI	Parkinson's progression markers initiative	LFDO	Levy flight distribution optimizer
SNUH	Seoul National University Hospital	F1-Score	Harmonic mean of precision and recall
SAE	Stacked auto-encoder	t-SNE	T-distributed stochastic neighbor embedding
KNN	K-nearest neighbor	GANs	Generative adversarial networks
LOSO CV	Leave-one-subject-out cross-validation		
FP-CIT	Fluoropropyl-carbamide-isoquinoline-tracer		
ROC	Receiver operating characteristic		
AUC	Area under the curve		

✉ Yahya Dogan  
yahyadogan@siirt.edu.tr

<sup>1</sup> Computer Engineering, Engineering Faculty, Siirt University, 56100 Siirt, Turkey

## 1 Introduction

PD is a prominent age-related neurodegenerative disorder, ranking as the second most common after Alzheimer's disease. PD is characterized by a range of motor and non-motor symptoms including speech difficulties, altered

movement patterns, and tremors. These symptoms not only significantly impair the quality of life for those affected but also present a considerable challenge to healthcare providers due to the lack of a definitive cure. Accurate prediction and effective management strategies for PD are essential in alleviating the impact of the disease on individuals. The global prevalence of PD underscores the importance of these strategies. Approximately 1% of the population over the age of 50 and 2.5% of those over the age of 70 are affected by this debilitating condition [1]. The progressive nature of PD means that its symptoms worsen over time, which further complicates treatment and management efforts. Worldwide, more than 10 million people are living with PD. This high prevalence creates a significant economic burden on healthcare systems and societies. The cost of managing PD is substantial, with estimates suggesting that it can amount to approximately \$23,000 per patient annually. These costs encompass a range of expenses including medical treatments, long-term care, and loss of productivity.

Early diagnosis of PD is critically important for several reasons. Firstly, early detection allows for timely intervention, which can significantly slow the progression of the disease and improve the quality of life for patients. Early treatment can help manage symptoms more effectively and delay the onset of severe disability. Furthermore, early diagnosis allows patients and their families more time to plan for the future, make necessary lifestyle adjustments, and seek appropriate support and resources.

Individuals with PD often struggle with controlling bodily movements due to neural changes, affecting fine motor skills like writing. Consequently, handwriting alterations can be early indicators of PD. In the initial stages, these changes are subtle and may go unnoticed, but detecting them is crucial as they can signal preclinical PD. Symptoms such as constricted handwriting (micrographia) or rapid changes in writing size may indicate early-stage PD. Handwriting in PD patients is often affected by slow movement, tremors, impaired balance, and stiffness. Numerous studies have highlighted handwriting impairments as a significant biomarker for early PD detection. Handwriting analysis has emerged as a promising diagnostic tool for PD, offering a cost-effective and time-efficient alternative to traditional neurological evaluations and brain imaging scans. The intricate connection between motor control and handwriting makes it a valuable window into the neurological health of individuals. Expert handwriting analysis is gaining popularity as a viable way to detect early signs of PD by identifying subtle changes like tremors, micrographia, and anomalies in pen pressure and stroke patterns.

Recently, machine learning and deep learning models have been increasingly utilized for PD diagnosis [2–4].

These technologies can automate the analysis process, assisting experts in decision-making by recognizing complex patterns indicative of PD with high accuracy. This automation ensures consistent and reliable results, reducing the reliance on individual expertise. Additionally, these algorithms can uncover new biomarkers and subtle indicators of PD, improving early diagnosis and treatment outcomes for individuals with PD. Recent successes in using CNNs for automatic feature extraction from images are well-documented. However, these models require large datasets, and the sample size in Parkinson's datasets is typically quite small. To address this issue, transfer learning approaches are generally employed. Yet, even with transfer learning, not all features may be meaningful due to the limited number of samples. Another challenge is that the fully connected layers used in the final stages of transfer learning models contain numerous parameters. Training these parameters with a small number of samples can lead to overfitting. Therefore, this study aims to use various transfer learning models to extract a combination of features. Afterward, the non-significant features are eliminated, and the final classification is performed using a classifier with a limited number of parameters.

The motivation behind this study is to contribute to the field of PD diagnosis by proposing a novel hybrid approach that leverages the strengths of both deep learning and machine learning algorithms. For the first time, this approach systematically utilizes features extracted from various transfer learning models [5–13] to enhance diagnostic accuracy. However, not all features extracted through transfer learning are meaningful, necessitating an efficient feature selection process. This study leverages NCA [14] to eliminate redundant or less informative features, thereby enhancing the discriminative power of the model. By intelligently selecting and preserving only the most informative features, NCA significantly contributes to the performance improvement of the diagnostic model. Finally, the classification stage utilizes SVM [15] to leverage the enriched feature vectors obtained through the preceding steps. SVMs are less prone to overfitting, especially when dealing with small datasets, as they focus on maximizing the margin between classes. Additionally, SVMs are effective in high-dimensional spaces and can work well with a clear margin of separation. They also perform well with a limited number of parameters, making them suitable for scenarios with limited data. This comprehensive approach combines the strengths of deep learning for feature extraction, NCA for dimensionality reduction, and SVM for classification, aiming to achieve robust and accurate diagnostic performance. The proposed methodology achieves a state-of-the-art accuracy of 99.39% on the Parkinson Hand Drawings dataset. This success highlights the practical utility of the developed

methodology in real-world diagnostic scenarios. Additionally, the t-distributed Stochastic Neighbor Embedding (t-SNE) method demonstrates clear and distinct groupings among the test samples, further supporting the model's applicability.

The remainder of the article is organized to include: Sect. 2 summarizes prior studies on diagnosing PD. Section 3 provides a detailed discussion of the proposed methodology, including datasets, evaluation metrics, and training details. Section 4 presents experimental studies and results. The paper concludes by discussing the findings in Sect. 5 and potential future work in Sect. 6.

## 2 Related works

Diagnosing PD requires a multifaceted approach, encompassing physical exams, neurological evaluations, and imaging techniques such as MRI or CT scans to identify brain abnormalities. Additional tools include dopamine transporter imaging and genetic testing. Laboratory tests are employed to rule out other conditions with similar symptoms. In recent years, the integration of artificial intelligence has progressively enhanced the diagnostic process, with deep learning and machine learning techniques significantly improving Parkinson's diagnosis by analyzing various indicators, including speech [16], handwriting disorders [17], EEG signals [18], hand tremors [19], nocturnal breathing patterns [20], smell signatures [21], MRI brain scans [22], urine biomarkers [23], and sketching patterns like spirals and waves [2].

Patients with PD often struggle with motor tasks like writing and drawing due to altered neuronal control of body movements. This has led researchers to create datasets based on handwriting and drawing to identify patterns distinguishing PD patients from healthy individuals, aiding early diagnosis. Among these datasets, notable examples include PaHaW, HandPD, and Parkinson Hand Drawings (also known as the Spiral/Wave dataset), each contributing valuable insights into the motor impairments associated with PD. The PaHaW dataset includes handwriting and drawing samples from 75 individuals (37 PD, 38 healthy), collected in collaboration with St. Anne's University Hospital and Masaryk University. The HandPD dataset features spiral and meander sketches from 92 individuals (74 PD, 18 healthy) collected at Botucatu Medical School in Brazil. It contains 368 samples with each drawing repeated four times. The Parkinson Hand Drawings dataset consists of 204 samples (102 PD, 102 healthy) involving Archimedean spirals and Sinusoidal waves [24]. There are other datasets available as well; however, since these datasets are more extensively studied in the literature, this section will focus on recent studies utilizing these datasets,

organized by year. The methods, benefits, parameters, levels of difficulty, and accuracy values of recent studies using these datasets are detailed in Table 1. The exact parameters of the models in these studies are not explicitly stated; therefore, aspects such as the transfer learning model used are mentioned. This information provides researchers in the field with insights into model capacity. An assessment has been made based on the estimated number of parameters to determine the complexity of these models. Approximately, models with parameters suitable for use on mobile devices are classified as 'moderate', efficient models with fewer than 5M parameters are classified as 'moderate to high', and models with more than 5M parameters are classified as 'high'.

Drotar et al. [25, 26] have made significant contributions to the diagnosis of PD using the PaHaW dataset. In their studies, they evaluated the effects of different handwriting modalities on the diagnosis of PD. Factors such as on-surface movement, in-air movement, and pressure applied to the tablet surface, which are rarely considered, were analyzed. These factors have been shown to provide valuable information for diagnosing PD through handwriting. In addition to traditional kinematic and spatiotemporal features, new features based on the entropy of the handwriting signal and empirical mode decomposition were introduced. Using the Mann–Whitney U test filter and the Relief algorithm allowed for a more accurate and effective feature selection process, enabling the precise identification of disease-specific features. Kinematic features such as speed, acceleration, velocity, and jerk, along with pressure and spatiotemporal features, were extracted and classified using machine learning algorithms like KNN, AdaBoost, and SVM. These results demonstrate that handwriting can be used as a biomarker for PD and that classification performance achieved a high accuracy of 89%. These studies highlight the effectiveness of handwriting dynamics and different algorithms in the early detection and identification of PD. Impedovo [27] investigates various velocity-based features extracted from handwriting signals, such as the sigma log-normal model, Maxwell-Boltzmann distribution, Fourier, and Cepstrum transforms. The study demonstrates that combining these novel velocity-based features with traditional ones enhances the classification performance, achieving a notable accuracy of 98.44% on the PaHaW dataset. This improvement highlights the efficacy of these features in distinguishing handwriting patterns between individuals with PD and healthy controls. Additionally, the research shows that these features effectively utilize the potential of different tasks, including the Archimedes spiral task, which was previously considered less impactful for classification purposes. However, the extraction of these handcrafted features is time-consuming and labor-intensive. Naseer

**Table 1** Comparison of datasets, years, benefits/methods, parameters, levels of complexity, and accuracy of related works in literature

Study	Dataset	Year	Methods/Benefits	Params	Levels of Complexity	Acc(%)
[2]	Spiral-Wave	2024	Presents a high-accuracy hybrid method for Parkinson's diagnosis by combining spiral and wave CNN-KNN architectures. This approach merges CNN's feature extraction with KNN's classification. An Ensemble Voting Classifier averages sub-classifier probabilities enhancing accuracy and performance by leveraging both models' strengths preventing overfitting and capturing nuanced data relationships	Custom CNN layers	High	96.67
[37]	Spiral-Wave	2023	Proposes the Diplin model which combines WGAN transfer learning and EfficientNetV2 to achieve outstanding performance in image classification, particularly in predicting associated disease risk. This approach merges the powerful sample generation and feature extraction capabilities of DL with the efficiency of lightweight neural networks providing a robust framework for disease risk prediction	EfficientNetV2 layers	Moderate to High	98.0
[36]	Spiral-Wave	2023	By combining FRCNN and BiGRU in parallel the model extracts both kinematic and spatiotemporal features allowing for better symptom detection. Additionally, the LFDO algorithm auto-tunes the hyperparams enhancing the model's performance across various datasets	Faster RCNN, BiGRU params	High	98.82
[35]	Spiral-Wave	2023	Modifies the low-parameter MobileNetV2 model to create an effective lightweight solution for real-time Parkinson's detection on mobile and edge devices without intensive computational resources	MobileNetV2 layers	Moderate	98.0
[38]	Spiral-Wave	2022	Integrates CNN and ANN with GridSearchCV to effectively utilize drawing and acoustic features for Parkinson's detection	CNN and ANN layers, GridSearchCV params	High	98.0
[34]	Spiral-Wave	2022	Develops a high-accuracy hybrid method for Parkinson's diagnosis by integrating CNN and traditional HOG for feature extraction with classical classifiers like SVM, LR, KNN and Bayes. This approach combines the powerful feature extraction of DL with the effective classification of traditional ML algorithms	ResNet50 layers, SVM params	High	89.0
[39]	Spiral-Wave	2021	Utilizes the ResNet50 model with transfer learning to accurately diagnose PD with spiral and wave images preprocessed using data thinning and data augmentation resulting in enhanced model performance and more reliable detection of the disease	ResNet50 layers	High	96.67
[33]	Spiral-Wave	2021	Incorporates multiple PD handwriting datasets and employs DL-based algorithms to effectively address the high variability in handwritten material for improved Parkinson's detection performance	GoogLeNet layers	High	90.0
[32]	Spiral-Wave	2020	Combines prediction probabilities from two CNN architectures using LR and RFC meta-classifiers with the final prediction made through average voting	Custom CNN layers, LR and RFC params	High	93.3
[28]	PAHAW	2020	Models handwriting features using AlexNet and employs transfer learning to address limited sample sizes. Utilizes pre-trained AlexNet models on ImageNet and MNIST exploring freezing and fine-tuning methods. Demonstrates that the fine-tuned ImageNet approach efficiently extracts features	AlexNet layers	High	98.28
[27]	PAHAW	2019	Investigates various velocity-based features extracted from handwriting signals including the sigma log-normal model, Maxwell-Boltzmann distribution, Fourier and Cepstrum transforms. Combines new velocity-based features with classic ones to improve classification performance	SVM params	Moderate	98.44
[25]	PAHAW	2015	Proposes analyzing features like in-air movement and tablet surface pressure rarely examined for diagnosing PD. Introduces novel features based on handwriting signal's entropy and empirical mode decomposition uncovering complex hidden information relevant to PD diagnosis	SVM params	Moderate	89.0

**Table 1** (continued)

Study	Dataset	Year	Methods/Benefits	Params	Levels of Complexity	Acc(%)
[31]	HandPD	2018	Employs a DL-oriented approach utilizing CNNs to analyze handwriting dynamics enhancing PD diagnosis accuracy. Provides a comprehensive signal-based dataset composed of features related to handwritten dynamics supporting the use of handwriting as a biomarker for PD. Propose an ensemble of CNNs to distinguish PD patients from the control group better	Custom CNN layers	High	95.0
[30]	HandPD	2016	Introduces a CNN-based method to analyze handwriting dynamics from smartpen signals enhancing PD diagnosis accuracy. Provides a comprehensive dataset of sensor data (pressure, tilt, acceleration) from handwriting exams, supporting handwriting dynamics as a PD biomarker	Custom CNN layers	High	84.42

et al. [28] model handwriting features using AlexNet, employing transfer learning to manage limited sample sizes. They use pre-trained AlexNet models on ImageNet and MNIST, exploring both freezing and fine-tuning methods. Their fine-tuned ImageNet approach effectively extracts features, achieving 98.28% accuracy on the PaHaW dataset. Their deep convolutional neural network classifier, enhanced with transfer learning and data augmentation techniques, efficiently identifies handwriting impairments related to Parkinson's without traditional feature extraction.

Pereira et al. [29] have significantly advanced PD diagnosis with the development of the HandPD dataset, a unique collection of images featuring spirals and meanders extracted from digitized handwritten exams. Their proposed pipeline addresses the challenge of learning from non-registered images, demonstrating that meanders are more informative than spirals due to the latter's complex contours. Despite the high variability of the dataset, which includes patients in the early stages of PD, their approach shows promise for improving diagnostic accuracy. They evaluated three pattern recognition techniques: Naïve Bayes, optimum-path forest, and SVM with radial basis function, optimizing SVM kernel parameters through cross-validation. Using meander images, their approach achieved a recognition rate of approximately 67%, demonstrating its potential effectiveness in aiding PD diagnosis. In their subsequent work, Pereira et al. [30] further enhanced PD diagnosis by introducing a CNN-based method to analyze handwriting dynamics from smartpen signals. This method leverages a deep learning-oriented approach to automatically identify features from signals extracted during handwriting exams. The study also provided a comprehensive dataset of sensor data (pressure, tilt, acceleration), supporting the use of handwriting dynamics as a reliable biomarker for PD. The proposed approach outperformed traditional methods, achieving high

classification accuracy by effectively distinguishing between healthy individuals and PD patients. Specifically, they achieved an average overall accuracy of 84.42% over the test set considering the meander dataset and 83.77% over the test set considering the spiral dataset. The experimental results, involving different CNN architectures and image resolutions, demonstrated the method's robustness and potential for early-stage PD detection, with accuracy rates significantly higher than previous techniques. Building on these findings, Pereira et al. [31] have made further advancements by employing a deep learning-oriented approach utilizing CNNs to analyze handwriting dynamics. This method significantly enhances the accuracy of PD diagnosis, leveraging CNNs to learn features from images produced by handwritten dynamics, capturing critical information during individual assessments. The primary benefits of this study include the innovative use of CNNs, which provide a more accurate and reliable diagnostic tool compared to traditional methods. Notably, this approach achieved accuracy rates close to 95% in the context of early-stage detection. Furthermore, proposing an ensemble of CNNs to better distinguish PD patients from the control group ensures a robust and comprehensive analysis, significantly improving diagnostic accuracy. Overall, the study demonstrates that analyzing handwritten dynamics using deep learning techniques is a promising approach for the early and accurate identification of PD, outperforming traditional handcrafted features and methods.

Recent observations in the diagnosis of PD indicate an increased use of the Parkinson Hand Drawings dataset, which consists of Archimedean spirals and sinusoidal waves, due to its simplicity and effectiveness in capturing motor impairments associated with the disease. Chakraborty et al. [32] proposed a comprehensive system design for analyzing spiral and wave drawing patterns to detect PD, leveraging two distinct CNNs for each drawing type. By utilizing prediction probabilities from the CNN



architectures, the authors trained meta-classifiers, namely logistic regression (LR) and random forest classifier (RFC), which provided weighted predictions through ensemble voting. The model achieved an overall accuracy of 93.3%. This multistage classification approach improved the precision of detecting PD and demonstrated that specific models worked better for certain samples, emphasizing the need for a combined decision-making process. The system's ability to perform accurate and precise predictions at the onset of the disease marks a significant contribution to the field, potentially leading to more effective clinical interventions.

Kamran et al. [33] suggested a comprehensive method for the early diagnosis of PD using handwriting samples by incorporating multiple PD datasets and employing deep transfer learning-based algorithms. By combining different PD datasets and applying various data augmentation techniques, the researchers effectively addressed the high variability in handwritten material, which resulted in improved diagnostic performance. Their approach, which leverages fine-tuned CNN architectures, achieved a remarkable accuracy of 99.22% when using combined datasets. They specifically achieved a 90% accuracy score using GoogleNet with only the Parkinson Hand Drawings dataset. This high level of accuracy demonstrates the system's potential to enhance early detection of PD, thereby contributing to more effective clinical interventions and better management of the disease's progression.

MORALES-CASTRO et al. [34] developed a high-accuracy hybrid method for Parkinson's diagnosis by integrating CNN and traditional HOG for feature extraction with classical classifiers like SVM, LR, KNN, and Bayes. This innovative approach harnesses the powerful feature extraction capabilities of deep learning while leveraging the effective classification potential of traditional machine learning algorithms. As a result, the best classification scenario was achieved using the ResNet50 neural network, which outperformed the HOG method. This approach achieved an impressive accuracy of nearly 90% on both spiral and wave drawings. Additionally, SVM emerged as the top classifier in both scenarios, demonstrating robustness as the test set remained independent of the training set, thus ensuring unbiased category assignments. This methodology significantly advances PD detection by combining advanced feature extraction with reliable classification techniques, achieving high diagnostic accuracy and enhancing early detection capabilities.

Kumar and Bansal [35] introduced a modified MobileNetV2 model designed for real-time PD detection on mobile and edge devices, emphasizing lightweight architecture and efficiency without heavy computational demands. Utilizing spiral and wave hand drawings, their approach demonstrates significant contributions to early

PD diagnosis by achieving a remarkable accuracy of 97.70%. The innovative use of fewer parameters while maintaining high accuracy underscores the model's efficiency. Furthermore, their research highlights the therapeutic benefits of artistic activities for Parkinson's patients, suggesting that despite the challenges posed by the disease, creative expression through drawing can alleviate symptoms like depression and anxiety and provide a sense of accomplishment. The use of a balanced dataset and the application of transfer learning for feature extraction from hand drawings of both healthy and Parkinson's individuals further validate the robustness of their model. The study concludes that hand drawings are a valuable diagnostic tool, offering a non-invasive, efficient, and accurate method for PD detection.

Krishnsmoorthy et al. [36] presented a significant advancement in the early detection of PD through the development of the levy flight optimized hybrid weighted faster recurrent network (Lf-HWFRNet). This system integrates the faster region-based convolutional neural network (FRCNN) and the bidirectional gated recurrent unit (BiGRU) in parallel, effectively extracting both kinematic and spatiotemporal features. The unique levy flight distribution optimizer (LFDO) auto-tunes hyperparameters, enhancing the model's performance across diverse datasets. This innovative approach utilizes handwritten samples from databases such as PaHaW, NewHandPD, and Parkinson's hand drawing dataset. It achieves an impressive accuracy of 98.82% on the Parkinson's drawing dataset. The primary contributions of this work include enhancing image quality through preprocessing and augmentation techniques, using a weighted average ensemble method for optimal feature relevance, and significantly improving prediction performance with the LFDO algorithm. These advancements enhance classification accuracy and reduce processing time to 0.045 s per image. This demonstrates the model's potential as an effective clinical tool for early PD diagnosis.

Zhou et al. [37] proposed the Diplin model, which combines WGAN, transfer learning, and EfficientNetV2 to achieve outstanding performance in image classification, particularly in predicting associated disease risk. The foundation of the model lies in WGAN, utilizing the Wasserstein distance to ensure accuracy between real and generated sample distributions. The implementation process involves constructing and training a WGAN-based sample generation model to produce high-quality samples, followed by a sample feature preprocessing model to enhance discriminative capabilities. Transfer learning and EfficientNetV2 are then integrated to build and train a classification model, leveraging pre-trained models to accelerate training and improve performance. Emphasis is placed on optimizing sample feature extraction and

classification modules, leading to remarkable results. This model achieved an impressive accuracy rate of 98% on the validation set, demonstrating its effectiveness in image classification for disease risk prediction. The experimental results show that in the application scenario of nursing homes, the Diplin model can provide practical support for predicting the health risks of the elderly, and this model can be run directly on a tablet. These results indicate that the Diplin model significantly advances disease risk prediction, offering a practical and efficient solution for use in environments without professional medical equipment, such as nursing homes.

Saleh et al. [2] presented a high-accuracy hybrid method for Parkinson's diagnosis by combining spiral and wave CNN-KNN architectures. This approach merges CNN's powerful feature extraction with KNN's effective classification, leveraging an ensemble voting classifier to average sub-classifier probabilities. This method captures nuanced data relationships and prevents overfitting, leading to a robust and reliable system. The primary focus is predicting PD through hand tremors, evident in varying speed and pen pressure between healthy and affected individuals during sketching. The authors proposed this ensemble classifier to enhance medical services, improve quality of life, and enable early detection. Unlike traditional CNNs, this architecture offers flexibility with small and imbalanced datasets, automating feature extraction and classification. By optimizing data augmentation parameters, the model's robustness and generalization are improved without data deformation. Additionally, the study addresses the critical issue of misclassification in the medical field, proposing solutions to mitigate potential fatal errors. The proposed system achieves 96.67% accuracy, demonstrating its effectiveness and potential for real-world application in PD diagnosis.

Deep learning models require large datasets. However, the sample sizes of Parkinson's datasets are generally small, which creates challenges in training and generalizing the model. To address these challenges, transfer learning models are used. Yet, even with transfer learning, not all features obtained are meaningful. This study proposes a hybrid model consisting of CNN, NCA, and SVM components to overcome these limitations. The proposed approach identifies the most distinctive feature set by utilizing single, pairwise, and triple combinations of nine different transfer learning models. Among these features, the significant ones are selected using the NCA method, and the classification process is carried out with SVM. This architecture aims to improve the early diagnosis of PD by combining the strengths of transfer learning models, the precision of NCA for feature selection, and the classification capabilities of SVM.

## 3 The material and method

### 3.1 Datasets

Parkinson Hand Drawings dataset [40] consists of hand-written samples, comprising 204 original images. These images are divided into two distinct classes: Healthy and PD. The dataset includes 102 examples of Spiral and 102 examples of Waves. From the 102 images of each type, 51 samples are derived from participants with PD, while the remaining 51 samples are obtained from healthy participants. Random sample images from both classes are depicted in Fig. 1.

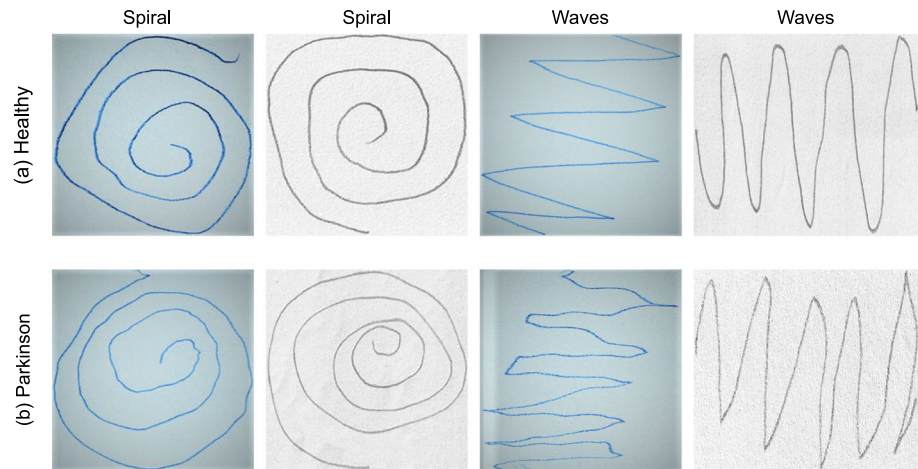
Figure 1 illustrates representative samples from the Parkinson Hand Drawings dataset used in this study. The dataset includes two types of images from both healthy individuals and PD patients: spirals and waves. Handwriting impairment is assessed using computational spiral analysis, which evaluates spatial, dynamic, and kinematic anomalies and markers of motor function and dysfunction. Digitally enhanced spiral drawings correlate with motor scores and may be more sensitive to early changes than subjective judgments. Spirals are particularly effective in capturing kinematic features because subjects tend to traverse the figure in a 360° rotation, making them a strong discriminator between PD patients and healthy controls. Wave drawing patterns are also significant in assessing handwriting impairments. They provide a distinct set of kinematic features, such as amplitude, frequency, and consistency. These features can reveal fine motor control issues and movement irregularities indicative of PD. Waves are beneficial for detecting tremors and rhythmic disturbances in the handwriting of PD patients. The repetitive nature of wave patterns allows for a detailed analysis of movement smoothness and control, offering additional markers for early diagnosis and differentiation between PD patients and healthy individuals.

### 3.2 Proposed approach

The proposed method for PD identification comprises three stages: feature extraction, feature selection, and classification.

**Feature extraction:** This section aims to represent the relevant example with a few features using transfer learning models. Transfer learning is typically employed to apply pre-learned knowledge from deep learning models trained on large and diverse datasets to new, smaller, or more specialized datasets. Examples represented with a few features are critical for facilitating this transfer because they provide a more general and inclusive representation. Among the algorithms used in this field are methods such

**Fig. 1** Sample images from the Parkinson Hand Drawings dataset



as domain adaptation, transfer learning, and multi-task learning. These algorithms aim to enhance performance by effectively utilizing the knowledge obtained from the source task on examples with few features in the target task.

In this study, nine different transfer learning models, namely InceptionV3 [5], DenseNet201 [6], EfficientNetB0 [7], ResNet50 [8], MobileNetV2 [9], VGG16 [10], Xception [11], NASNetMobile [12], and InceptionResNetV2 [13], were used to determine the feature vectors that best represent PD. The output of global average pooling (GAP) layers was used as feature vectors. These layers summarize the feature vectors after the convolutional layers of transfer learning models. These outputs contain visual or linguistic cues learned by the models and are typically high-dimensional. The feature vector obtained from the  $i$ -th model, denoted by  $V_i$ , is presented in Eq. (1):

$$V_i = f_i(x) \quad (1)$$

where the function  $f_i$  represents the  $i$ -th transfer learning model and  $x$  represents the input image. The relevant transfer learning models are those that exhibit superior performance on the ImageNet dataset. One of the main contributions of this study is to investigate how combining feature vectors from different transfer learning models affects model performance. Specifically, the focus is on determining which models' feature vector combinations enhance overall performance and the effectiveness of this combination method. In this context, experimental research has been conducted to identify the most important features using individual models, binary combinations, and triple combinations. In multiple model combinations, features extracted from each model are merged. At this stage, pre-trained weights from the ImageNet dataset [41] have been used for individual models, followed by fine-tuning on the Parkinson Hand Drawings dataset. Combinations of feature vectors obtained from various transfer learning models are

utilized to create a more comprehensive and representative feature set. The feature combination is formulated in Eq. (2):

$$V_{concat} = \phi(V_1, V_2, \dots, V_n) \quad (2)$$

where the function  $\phi$  combines different feature vectors to create a new vector. The resulting feature vector,  $V_{concat}$ , obtained at this stage will serve as the basis for feature selection and classification in subsequent stages. The feature extraction process is a critical step that enables more accurate identification of PD and directly impacts the model's overall performance.

**Feature selection:** In the second stage, the NCA method was used to select the most meaningful feature vectors obtained and use them during the classification phase. NCA is a powerful metric learning algorithm strategically crafted to elevate the classification performance of a stochastic nearest neighbors rule. Its primary objective is to maximize Leave One Out (LOO) classification accuracy by acquiring a supervised linear transformation within the feature space. This approach diverges from traditional methods that solely focus on similarity metrics, making NCA particularly distinctive in its application.

The fundamental premise of NCA lies in its ability to optimize LOO performance for future test data. In LOO classification, the KNN algorithm endeavors to predict a single point by measuring distances within the feature space. Rather than relying on pre-defined or random distance metrics, NCA takes a unique approach. It aims to learn an effective distance measurement through a linear transformation of the input data.

The essence of NCA is to learn a suitable distance metric to keep data points together. This metric brings points belonging to similar classes closer while pushing those belonging to different classes apart. NCA performs a space transformation by reweighting points in the feature



space, ensuring that examples from the same class are closer to each other and those from different classes are farther apart. This approach makes relationships between classes more distinct and enhances classification accuracy. NCA utilizes two fundamental components: the training data and classification labels. The training data is represented as  $X = x_1, x_2, \dots, x_n$ , where there are  $N$  data points, and each  $x_i$  represents the feature vector corresponding to the respective example. Classification labels are used for each example, denoted as  $y_i$ .  $y_i$  typically takes values from 1 to  $K$ , where  $K$  represents the total number of classes.

NCA performs a space transformation by reweighting points (examples) in the feature space, ensuring that examples from the same class are closer to each other while those from different classes are farther apart. This transformation is achieved using a weight matrix  $A$ . NCA learns a distance function based on this weight matrix. This distance function computes the distance between examples using the learned metric. The distance between examples  $d_{ij}$  is defined in Eq. (3):

$$d_{ij} = \|A(x_i - x_j)\|^2 \quad (3)$$

where  $x_i$  and  $x_j$  are two different examples, and  $A(x_i - x_j)$  represents the norm of the transformed difference vector of these two examples by the matrix  $A$ . NCA models the neighborhood of an example among all other examples with a probability distribution. For each  $x_i$ , the probability of another  $x_j$  example being a neighbor of  $x_i$  is calculated in Eq. (4):

$$p_{ij} = \frac{\exp(-\|A(x_i - x_j)\|^2)}{\sum_{k \neq i} \exp(-\|A(x_i - x_k)\|^2)} \quad (4)$$

where  $p_{ij}$  represents the probability of selecting the  $x_j$  example as a neighbor of the  $x_i$  example. In total, these probabilities are calculated for all  $k$  where  $k \neq i$ . The goal of NCA is to learn the matrix  $A$  that ensures examples with the same class labels are closer to each other. The objective function used for this purpose is to maximize the sum of probabilities of having neighbors with correct class labels for all examples. This objective function is presented in Eq. (5):

$$L(A) = \sum_{i=1}^n \sum_{j \neq i} \delta(y_i, y_j) p_{ij} \quad (5)$$

where the function  $\delta(y_i, y_j)$  checks whether  $y_i$  and  $y_j$  class labels are the same; if they are the same, it takes the value 1, otherwise, it takes the value 0. Thus, maximizing the function  $L(A)$  implies ensuring that examples with similar class labels are closer to each other. Gradient ascent is commonly used to update the values of the matrix  $A$ . The

gradient is calculated by taking the derivative of the objective function  $L(A)$  concerning  $A$ , and at each step,  $A$  is updated in the direction of this gradient.

It is crucial to note that NCA operates as a supervised learning algorithm, unlike unsupervised techniques such as principal component analysis. The key differentiator is the incorporation of target values during the estimation process. This supervised nature enables NCA to tailor its linear transformation to the specific characteristics of the training data, leading to enhanced predictive capabilities. In summary, NCA stands out as a metric learning method that goes beyond traditional similarity-based approaches. By incorporating supervised learning principles and focusing on a linear transformation of input data, NCA strives to optimize LOO classification performance, ultimately contributing to improved accuracy in predicting outcomes for unseen test data.

**Classification:** In the third stage, the SVM from machine learning algorithms has been used to classify the obtained feature vectors. SVM is recognized as a robust supervised learning algorithm, adept at classification, regression, and outlier detection tasks. The core objective of SVM revolves around identifying a hyperplane in an  $N$ -dimensional space where  $N$  represents the number of features which effectively segregates data points from distinct classes. The hallmark of SVM lies in its emphasis on maximizing the margin, which denotes the maximum separation between data points of different classes.

SVM exhibits several advantages, particularly in high-dimensional spaces. It is effective even in scenarios where the number of features exceeds the number of samples, demonstrating its versatility. Memory efficiency is another strength, as SVM employs only a subset of training points, known as support vectors, in the decision function.

Moreover, SVM can accommodate various kernel functions, including custom ones, to refine the decision function. This adaptability allows it to excel across different data types and structures. However, the potential for overfitting is a significant drawback, particularly when the number of features greatly exceeds the number of samples. This underscores the importance of careful selection of kernel functions and regularization terms.

The operation of SVM involves selecting a hyperplane that maximizes the margin, ensuring optimal separation between classes. This selection is guided by support vectors, which are the data points closest to the hyperplane. These points are critical as they determine the position and orientation of the hyperplane. The margin is conceptualized as a region bounded by two parallel hyperplanes. The mathematical formulation of the SVM objective is to maximize this margin. The decision function and its corresponding hyperplane can be described by Eq. (6):

$$w \cdot x - b = 0 \quad (6)$$

where  $w$  is the weight vector,  $x$  is an input feature vector, and  $b$  is the bias. The hyperplanes that define the margin are given by Eq. (7):

$$w \cdot x - b = 1 \text{ and } w \cdot x - b = -1 \quad (7)$$

The margin, defined as the distance between these two hyperplanes, is  $\frac{2}{\|w\|}$ . Hence, the problem of maximizing the margin translates to minimizing  $\|w\|$ . The hinge loss, used as the loss function in SVM, penalizes misclassifications and data points that fall within the margin. The hinge loss for a data point  $(x_i, y_i)$  is defined as in Eq. (8):

$$\max(0, 1 - y_i(w \cdot x_i - b)) \quad (8)$$

where  $y_i$  are the class labels. This loss function contributes to the objective function, which also includes a regularization term to prevent overfitting, as shown in Eq. (9):

$$\text{minimize } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \max(0, 1 - y_i(w \cdot x_i - b)) \quad (9)$$

where  $C$  is the regularization parameter, balancing the trade-off between maximizing the margin and minimizing the hinge loss. Gradient updates in SVM involve adjusting the weights ( $w$ ) based on the gradients of the loss function. These updates are contingent on correctly classifying a data point and its distance from the margin. In conclusion, SVM emerges as a formidable solution in complex scenarios, effectively handling classification tasks by focusing on margin maximization and strategic use of support vectors. Figure 2 illustrates the best-performing model combination based on the experimental results, showcasing the practical application of SVM in enhancing predictive performance.

Figure 2 shows the proposed architecture combining deep learning and machine learning techniques for PD diagnosis. In the proposed architecture, different CNN models' single, paired, and triple combinations were tested, and the combination with the highest performance was identified and used. Accordingly, each CNN model was utilized to extract different features from the data. After the feature extraction process, the features collected from each model's GAP layer were utilized. This layer averages each feature map to produce a more compact feature vector. The features from each CNN model were then combined into a single feature vector through a concatenation process. This stage allows the model to integrate different features obtained from various CNN architectures. The concatenated feature vectors were selected and dimensionally reduced using the NCA method. This stage ensures that the model focuses on the most important and distinctive features. The selected and dimensionally reduced features

were classified using an SVM model. SVM is a powerful machine learning algorithm capable of effectively classifying nonlinear data. This architecture aims to diagnose PD with higher accuracy by combining the strengths of various deep learning models.

### 3.3 Evaluation metrics

In this study, several evaluation metrics were used to assess the performance of the proposed model. These metrics include accuracy, precision, specificity, sensitivity, F1-score, and AUC. Each metric provides a different perspective on the model's performance, allowing for a comprehensive evaluation.

**Accuracy:** Accuracy measures the proportion of correctly classified instances out of the total instances. It is a general measure of the model's overall performance. In the context of PD diagnosis, accuracy indicates how well the model correctly identifies both patients with the disease and healthy individuals. It is presented in Eq. (10):

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

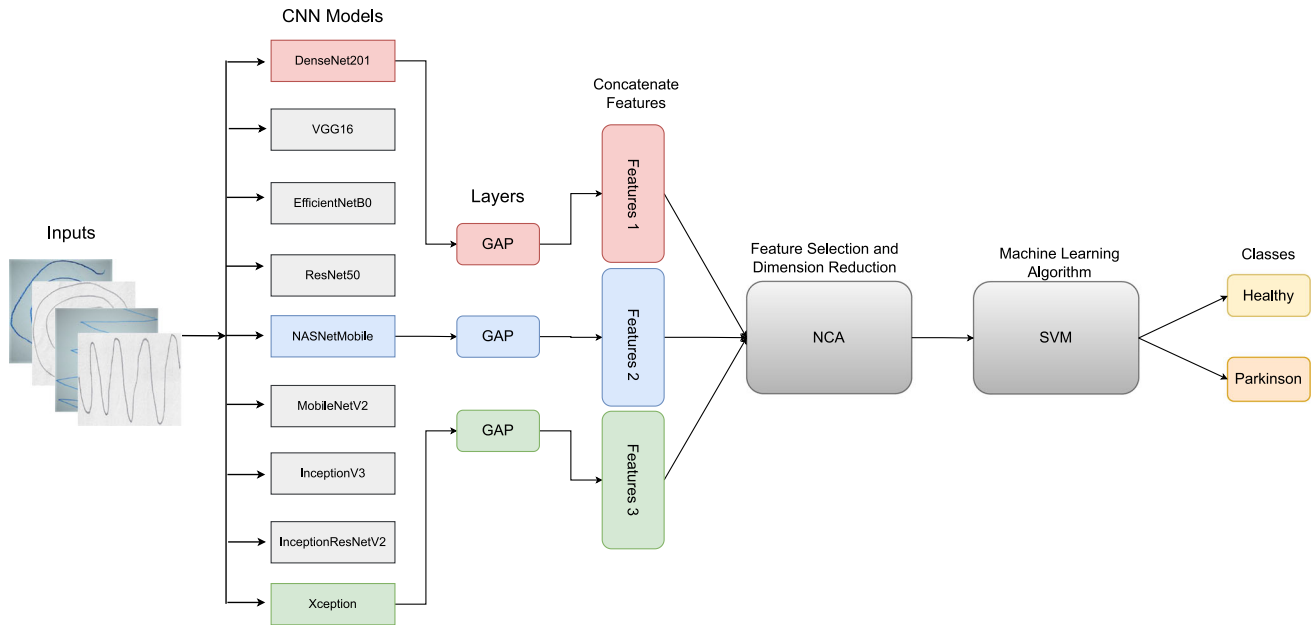
where  $TP$  represents true positives (correctly identified patients with PD),  $TN$  represents true negatives (correctly identified healthy individuals),  $FP$  represents false positives (healthy individuals incorrectly identified as having PD), and  $FN$  represents false negatives (patients with PD incorrectly identified as healthy). High accuracy suggests that the model performs well in distinguishing between healthy and diseased individuals.

**Precision:** Precision, also known as positive predictive value, measures the proportion of true positives out of the total predicted positives. It is particularly important in medical diagnosis as it reflects the reliability of a positive diagnosis. In the context of PD, high precision means that when the model predicts an individual has PD, it is very likely to be correct. It is presented in Eq. (11):

$$\text{Precision} = \frac{TP}{TP + FP} \quad (11)$$

Precision is crucial in reducing the number of false positives, thereby minimizing the misdiagnosis of healthy individuals as having PD.

**Specificity:** Specificity, also known as the true negative rate, measures the proportion of true negatives out of the total actual negatives. It indicates how well the model identifies healthy individuals. In the context of PD, high specificity means the model is effective at correctly identifying individuals who do not have the disease. It is presented in Eq. (12):



**Fig. 2** Proposed hybrid model for the diagnosis of PD

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (12)$$

High specificity is important to avoid false positives, ensuring that healthy individuals are not incorrectly diagnosed with PD.

**Sensitivity:** Sensitivity, also known as recall or the true positive rate, measures the proportion of true positives out of the total actual positives. It indicates how well the model identifies patients with PD. High sensitivity means that the model is effective at correctly identifying individuals who have the disease. It is presented in Eq. (13):

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (13)$$

High sensitivity is critical for early diagnosis and treatment, ensuring that most patients with PD are correctly identified.

**F1-Score:** The F1-score is the harmonic mean of precision and sensitivity, providing a balance between the two metrics. It is particularly useful when dealing with imbalanced datasets, where the number of healthy individuals may significantly differ from the number of patients with PD. It is presented in Eq. (14):

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \quad (14)$$

The F1-score is a comprehensive measure that considers both false positives and false negatives, making it valuable in medical diagnosis scenarios where both types of errors have significant consequences.

**Area under the curve (AUC):** AUC measures the ability of the model to distinguish between classes and is calculated as the area under the receiver operating characteristic curve. It is a scalar value between 0 and 1, where a higher AUC indicates better model performance. AUC provides an aggregate measure of performance across all classification thresholds. It is particularly useful for comparing the performance of different models.

These metrics together provide a comprehensive evaluation of the model's performance, highlighting its strengths and areas for improvement. Using a combination of these metrics ensures a thorough and balanced assessment of the model's effectiveness in diagnosing PD based on the augmented hand-drawing images.

### 3.4 Training details

This section provides training details. Initially, the dataset consisting of 204 images was insufficient, so augmentation was applied to create a larger dataset. Augmentation included rotations (90°, 180°, 270°), horizontal and vertical flips, and color transformations applied to the hand-drawn images. These operations increased the dataset size to 1,635 images. After augmentation, the images were resized to 224 × 224 pixels to ensure consistency with the input requirements of CNN models. The dataset was then split into 80% training and 20% test sets. A series of CNN models pre-trained on the ImageNet dataset were initially used for feature extraction. This approach enhances the model's generalization ability when the dataset is limited, yielding better performance with less data. Additionally,

models trained on a large dataset like ImageNet learn general features, enabling effective feature extraction for different tasks. Subsequently, fine-tuning was performed using the augmented dataset. During the training phase, the categorical cross-entropy loss function was used, and the Adam optimization algorithm was employed with a batch size of 32. The models were trained for 20 epochs using the default parameters of the Adam optimizer.

## 4 Experiment and results

This section provides a general overview of experimental studies conducted to diagnose PD. Initially, nine different transfer learning models widely recognized in the literature were employed to extract meaningful features from the images in the dataset. These models are: InceptionV3, DenseNet201, EfficientNetB0, ResNet50, MobileNetV2, VGG16, Xception, NASNetMobile, and InceptionResNetV2. Each model used pre-trained weights from the ImageNet dataset and was then fine-tuned with the Parkinson Hand Drawings dataset. This process is crucial for the models to learn features specific to PD. To ensure consistent evaluation of these transfer learning models, the feature maps produced by the final layer of each model were summarized using a GAP layer. The GAP layer compresses the feature maps into a single vector, facilitating the classification task. Subsequently, a dense layer with a Softmax activation function was added for the classification task. This layer produces an output vector corresponding to the number of classes, providing probability distributions for the predicted classes. This uniform strategy was applied to each model, and the experimental results are presented in Table 2.

Table 2 presents a comprehensive performance comparison of the fine-tuned models used in this study. Each model's performance is measured using several key metrics: accuracy, specificity, sensitivity, precision, F1 score, training time, inference time, the number of parameters (Params), and computational load (FLOPs). InceptionV3 achieved the highest performance with an accuracy of 97.55% along with efficient training (3.22 minutes) and inference times (0.04 minutes). Xception followed closely with an accuracy of 97.25%. InceptionResNetV2 also performed well, with an accuracy of 96.64%, but had the highest parameter count (54.3M) and computational load (13B FLOPs). DenseNet201 and NASNetMobile demonstrated moderate performance, with accuracies of 81.35% and 80.73%, respectively. ResNet50 and MobileNetV2 had varying degrees of success, with ResNet50 achieving an accuracy of 65.14% and MobileNetV2 achieving an accuracy of 77.06%. EfficientNetB0 and VGG16 performed less effectively, with EfficientNetB0 showing an

accuracy of 52.60% and VGG16 demonstrating an accuracy of 48.62%. In summary, InceptionV3 and Xception were the top performers, while EfficientNetB0 and VGG16 had the lowest performance in PD detection. Figure 3 shows the loss and accuracy graphs for the best-performing model, InceptionV3, and the lowest-performing model, VGG16, across epochs.

In Fig. 3, InceptionV3 shows a consistent decrease in training loss, indicating effective learning, while validation loss generally decreases with minor fluctuations, suggesting occasional overfitting. The best epoch is 16. Training accuracy steadily increases, and validation accuracy trends upward despite fluctuations. For VGG16, training loss decreases steadily, but validation loss has more pronounced fluctuations, indicating overfitting issues. The best epoch is 18. Training accuracy increases, but validation accuracy shows significant fluctuations, reflecting unstable performance. In summary, InceptionV3 demonstrates a more stable performance with fewer fluctuations and higher validation accuracy compared to VGG16. Both models exhibit some overfitting; however, InceptionV3 is more reliable and stable, making it a potentially better model for this task.

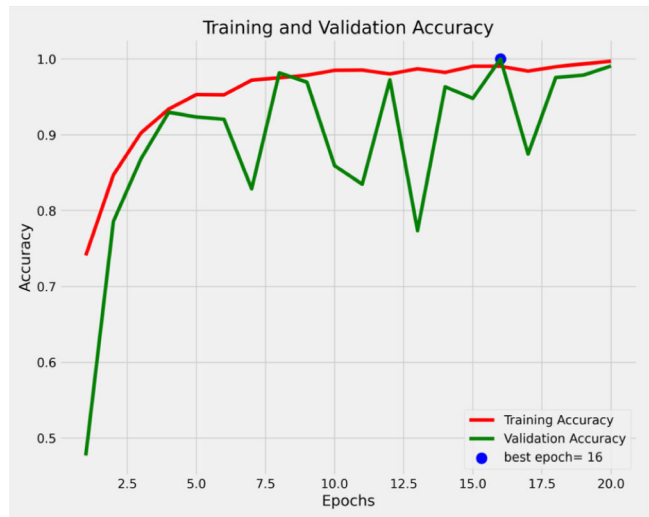
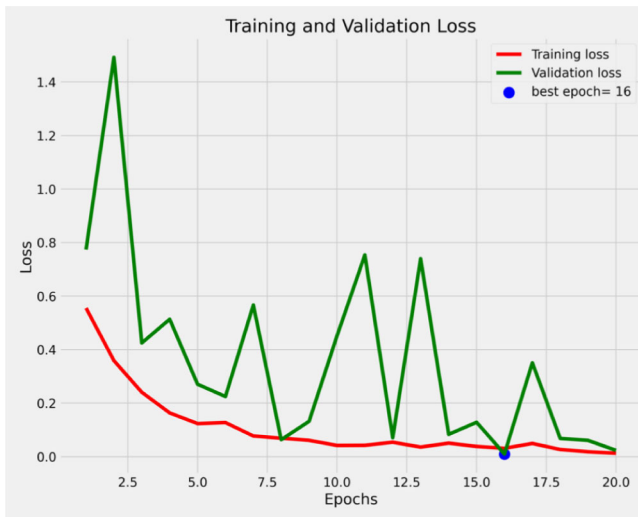
In the next phase of the experiments, the pre-trained models were used as feature extractors, utilizing the feature vectors from the GAP layers of each model. Subsequently, the NCA method was applied for feature selection and dimensionality reduction. NCA is a learning strategy designed to preserve the dataset's structure by clustering relevant formations in the feature space and classifying them by class. This is achieved by learning a transformation matrix that better represents the feature space. The primary goal of NCA is to group similar examples and use the learned transformation matrix to separate them from other classes in the dataset. As a result, NCA retained 17% of the features obtained from the GAP layers of the respective models and discarded the remaining 83%. The features preserved by the NCA method were classified using the SVM machine learning algorithm. Table 3 shows the results of using the features obtained from the respective model, applying the NCA dimensionality reduction approach, and using the SVM classification algorithm. Since the transfer learning models were not retrained at this stage, the variables training time, inference Time, params, and FLOPs remained almost the same. Therefore, these variables were no longer included in the tables.

In the initial Table 2 (without NCA and SVM), InceptionV3, Xception, and InceptionResNetV2 demonstrated high accuracies of 97.55%, 97.25%, and 96.64%, respectively. In Table 3 (with NCA and SVM), these models either maintained or improved their performance, with InceptionV3 and Xception both achieving 97.55%, and InceptionResNetV2 reaching 98.47%. DenseNet201

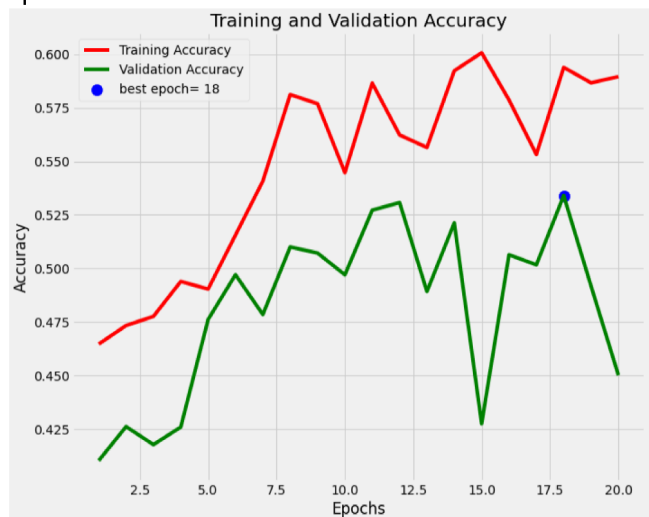
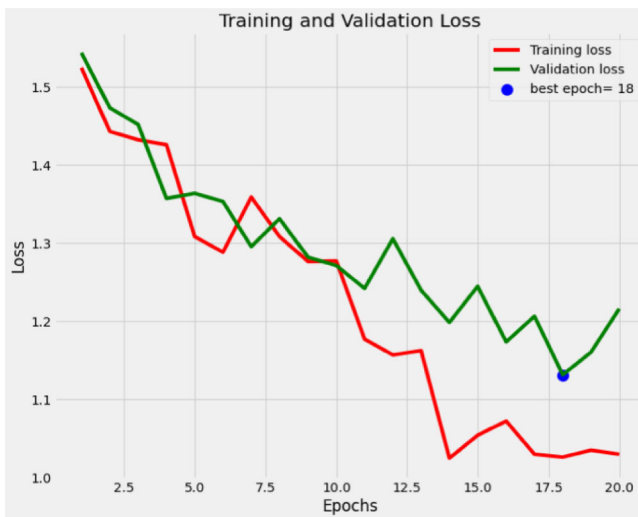


**Table 2** Performance of fine-tuned models

Models	Accuracy (%)	Precision (%)	Specificity (%)	Sensitivity (%)	F1 Score (%)	AUC (%)	Training Time (Min)	Inference Time (Min)	Params (M)	FLOPs
InceptionV3	97.55	97.55	97.45	97.55	97.55	97.58	3.22	0.04	21.8	5.7B
DenseNet201	81.35	79.83	81.2	81.25	80.54	80.48	6.87	0.08	18.3	8.6B
EfficientNetB0	52.60	28.01	53.5	52.70	36.59	50.32	4.16	0.04	4.1	0.8B
ResNet50	65.14	64.63	67.4	65.14	64.89	65.57	3.55	0.04	23.6	7.7B
MobileNetV2	77.06	74.61	78.8	77.06	75.81	76.07	1.98	0.03	2.3	0.6B
VGG16	48.62	24.86	38.6	48.62	33.81	50.82	3.57	0.02	14.7	30.7B
Xception	97.25	97.25	96.5	97.25	97.25	97.17	4.75	0.04	20.9	9.1B
NASNetMobile	80.73	79.74	79.7	80.73	80.23	80.06	5.69	0.09	4.3	1.1B
InceptionResNetV2	96.64	96.64	97.1	96.64	96.64	96.73	7.03	0.08	54.3	13B



Model: InceptionV3



Model: VGG16

**Fig. 3** The top graphs display the training and validation loss and accuracy of the best-performing model, InceptionV3, while the bottom graphs show these for the lowest-performing model, VGG16

**Table 3** Performance of fine-tuned models with NCA and SVM

Models	Accuracy (%)	Precision (%)	Specificity (%)	Sensitivity (%)	F1 Score (%)	AUC (%)
InceptionV3	97.55	98.15	98.6	97.55	97.55	97.58
DenseNet201	96.02	96.53	97.6	96.02	96.02	96.00
EfficientNetB0	87.16	88.02	88.4	87.16	87.16	87.16
ResNet50	70.03	70.86	71.0	70.03	69.80	69.66
MobileNetV2	93.88	94.34	94.08	93.88	93.89	93.98
VGG16	47.71	48.52	48.63	47.71	30.82	50.00
Xception	97.55	98.15	98.6	97.55	97.53	97.52
NASNetMobile	91.44	92.13	92.32	91.44	91.43	91.33
InceptionResNetV2	98.47	98.86	98.7	98.47	98.47	98.46

improved significantly from 81.35% to 96.02%, and MobileNetV2 increased from 77.06% to 93.88%. EfficientNetB0 also showed substantial improvement, rising from 52.60% to 87.16%. ResNet50's accuracy improved from 65.14% to 70.03%, while VGG16 remained low at 47.71%, indicating minimal impact from the NCA-SVM method. NASNetMobile's accuracy increased significantly from 80.73% to 91.44%, showing a notable improvement in classification performance. Overall, experimental studies indicate that the combined use of CNN, NCA, and SVM approaches generally enhances results.

The study's next step focused on evaluating the impact of combining features from different CNN models. In this case, features derived from binary combinations of nine different transfer learning models were first concatenated. Specifically, features taken independently from each pair of transfer learning models were concatenated, and the resulting feature sets went through feature selection and dimensionality reduction using the NCA approach. The newly obtained features were then trained using the SVM method. This method allows each combination to highlight distinct structures inside the feature space. A total of 36 distinct experimental studies were conducted, considering pairwise combinations of the nine transfer learning models. Table 4 presents the top and bottom 10 experimental results based on the performances obtained from these model combinations.

Table 4 shows that features derived from pairwise combinations of models significantly improve model performance. Notably, the combination of InceptionV3 and EfficientNetB0 yielded an impressive accuracy of 99.1%. In addition, the top ten model combinations achieved accuracy rates ranging from 98.5% to 99.1%. Examining the worst-performing model combinations, the pairing of the VGG16 model, which previously achieved a single accuracy rate of 47.71%, and the ResNet50 model, which had an accuracy rating of 70.03%, resulted in the lowest

accuracy rate of 71.6%. Despite being the least favorable combination, it still demonstrates a significant improvement compared to the individual performance of VGG16. Overall, the combination of models such as InceptionV3 + EfficientNetB0, InceptionV3 + Xception, and Xception + NASNetMobile proves to be the most promising, with consistently high accuracy rates. This indicates that leveraging complementary strengths from different models can lead to superior performance. Comparing Tables 3 and 4, it is evident that paired combinations of models produce very high accuracy rates. The effectiveness of model combinations is particularly evident in the substantial performance gains seen in previously lower-performing models when paired with stronger models.

The next phase of the study evaluated the impact of combining features from different CNN models in triple combinations. Features from each trio of the nine transfer learning models were concatenated, followed by feature selection and dimensionality reduction using NCA. The refined features were then trained using the SVM method, allowing each combination to highlight distinct structures within the feature space. A total of 84 experimental studies were conducted. Table 5 presents the top and bottom 10 experimental results based on the performance of these model combinations.

Table 5 summarizes the performance results of triple feature combinations derived from fine-tuned transfer learning models integrated with NCA and SVM. Upon examining the experimental results, it is clear that six distinct combinations, e.g., InceptionV3 + DenseNet201 + Xception, achieved the best performance, each with an accuracy of 99.39%. To determine the most efficient model among these top-performing combinations, it is crucial to evaluate the total number of parameters, training time, inference time, and FLOPs, in addition to accuracy. For instance, the combination of InceptionV3, DenseNet201, and Xception has 61.0M parameters, with medium training

**Table 4** Performance of pairwise model combinations with NCA and SVM

Model 1	Model 2	Accuracy (%)	Precision (%)	Specificity (%)	Sensitivity (%)	F1 Score (%)	AUC (%)
InceptionV3	EfficientNetB0	99.1	99.6	99.2	99.1	99.1	99.1
InceptionV3	Xception	99.1	99.5	99.2	99.1	99.1	99.0
InceptionV3	InceptionResNetV2	99.1	99.6	99.2	99.1	99.1	99.1
Xception	NASNetMobile	99.1	99.6	99.2	99.1	99.1	99.1
Xception	InceptionResNetV2	99.1	99.6	99.2	99.1	99.1	99.1
InceptionV3	MobileNetV2	98.8	99.1	98.8	98.8	98.8	98.8
MobileNetV2	InceptionResNetV2	98.8	99.1	98.8	98.8	98.8	98.8
NASNetMobile	InceptionResNetV2	98.8	99.1	98.8	98.8	98.8	98.8
InceptionV3	NASNetMobile	98.5	98.9	98.5	98.5	98.5	98.5
EfficientNetB0	InceptionResNetV2	98.5	98.9	98.5	98.5	98.5	98.5
–	–	–	–	–	–	–	–
–	–	–	–	–	–	–	–
DenseNet201	VGG16	95.7	96.3	95.7	95.7	95.7	95.7
MobileNetV2	NASNetMobile	95.7	96.3	95.7	95.7	95.7	95.7
EfficientNetB0	NASNetMobile	95.4	96.1	95.4	95.4	95.4	95.4
ResNet50	MobileNetV2	95.4	96.1	95.4	95.4	95.4	95.4
MobileNetV2	VGG16	93.9	94.5	93.9	93.9	93.9	94.0
VGG16	NASNetMobile	93.6	94.5	93.6	93.6	93.6	93.5
ResNet50	NASNetMobile	91.7	92.3	91.7	91.7	91.7	91.6
EfficientNetB0	VGG16	87.5	87.9	87.5	87.5	87.5	87.4
EfficientNetB0	ResNet50	86.2	86.8	86.2	86.2	86.2	86.3
ResNet50	VGG16	71.6	72.2	71.4	71.6	71.4	71.2

and inference times, and high FLOPs. The combination of InceptionV3, EfficientNetB0, and Xception has 46.8M parameters, with faster training and inference times, and medium FLOPs. InceptionV3, MobileNetV2, and Xception have 45.0M parameters, with fast training and inference times, and low FLOPs. InceptionV3, VGG16, and Xception have 57.4M parameters, with slower training and inference times, and high FLOPs. InceptionV3, Xception, and NASNetMobile have 47.0M parameters, with fast training and inference times, and medium FLOPs. DenseNet201, Xception, and NASNetMobile have 43.5M parameters, with medium training and inference times, and low FLOPs. Considering these factors, DenseNet201, Xception, and NASNetMobile emerge as the most parameter-efficient and computationally efficient combination while still maintaining the highest performance level. On the other hand, the combination of VGG16, NASNetMobile, and InceptionResNetV2 performed poorly, achieving the lowest accuracy of 83.18%. Despite having high computational resources, this combination did not translate into better performance, highlighting the importance of model selection and feature combination. Table 5 shows that combinations involving InceptionV3 and Xception consistently outperformed other models among the most

successful combinations. In contrast, VGG16 frequently appeared in the least effective combinations, indicating its lower effectiveness in these triple-model setups. Figure 4 displays a confusion matrix plot demonstrating the classification performance of the successful model combination DenseNet201, Xception, and NASNetMobile.

The confusion matrix in Fig. 4 illustrates the performance of the DenseNet201, Xception, and NASNetMobile model combination with NCA and SVM on test images. The matrix provides a clear representation of the model's classification accuracy for two classes: Healthy and PD. The model correctly classified 170 images as Healthy (true negatives) and 155 images as PD (true positives). There were minimal misclassifications, with only 1 image incorrectly classified as Healthy when it was PD (false positive) and 1 image incorrectly classified as PD when it was Healthy (false negative). This high level of accuracy, where the model correctly classified 325 out of 327 images, highlights the robustness and reliability of this model combination. The near-perfect classification performance aligns with the previously noted accuracy rate of 99.39%, demonstrating that the combination of DenseNet201, Xception, and NASNetMobile is both parameter-efficient and highly effective in accurately classifying the dataset.

**Table 5** Performance of triple model combinations with NCA and SVM

Model 1	Model 2	Model 3	Accuracy (%)	Precision (%)	Specificity (%)	Sensitivity (%)	F1 Score (%)	AUC (%)
InceptionV3	DenseNet201	Xception	99.39	99.36	99.42	99.36	99.36	99.39
InceptionV3	EfficientNetB0	Xception	99.39	99.36	99.42	99.36	99.36	99.39
InceptionV3	MobileNetV2	Xception	99.39	99.36	99.42	99.36	99.36	99.39
InceptionV3	VGG16	Xception	99.39	99.36	99.42	99.46	99.36	99.39
InceptionV3	Xception	NASNetMobile	99.39	99.36	99.42	99.36	99.36	99.39
DenseNet201	Xception	NASNetMobile	99.39	99.36	99.42	99.36	99.36	99.39
InceptionV3	DenseNet201	EfficientNetB0	99.08	99.08	98.88	99.08	99.08	99.08
InceptionV3	DenseNet201	MobileNetV2	99.08	99.08	98.88	99.08	99.08	99.08
InceptionV3	DenseNet201	InceptionResNetV2	99.08	99.08	98.88	99.08	99.08	99.08
InceptionV3	ResNet50	Xception	99.08	99.08	98.88	99.08	99.08	99.08
–	–	–	–	–	–	–	–	–
–	–	–	–	–	–	–	–	–
MobileNetV2	NASNetMobile	InceptionResNetV2	90.52	90.52	89.82	90.52	90.53	90.52
ResNet50	NASNetMobile	InceptionResNetV2	90.21	90.21	89.68	90.21	90.29	90.21
ResNet50	MobileNetV2	VGG16	88.99	88.99	88.15	88.99	88.99	88.99
ResNet50	VGG16	NASNetMobile	88.99	88.99	89.35	88.99	89.03	88.99
MobileNetV2	VGG16	NASNetMobile	88.38	88.38	87.75	88.38	88.39	88.38
EfficientNetB0	NASNetMobile	InceptionResNetV2	88.07	88.07	88.07	88.07	88.07	88.07
EfficientNetB0	VGG16	InceptionResNetV2	87.77	87.77	87.52	87.77	88.17	87.77
EfficientNetB0	VGG16	NASNetMobile	86.85	86.85	87.4	86.85	86.85	86.85
ResNet50	VGG16	InceptionResNetV2	86.24	86.24	87.3	86.24	86.24	86.24
VGG16	NASNetMobile	InceptionResNetV2	83.18	83.18	82.23	83.18	83.28	83.18

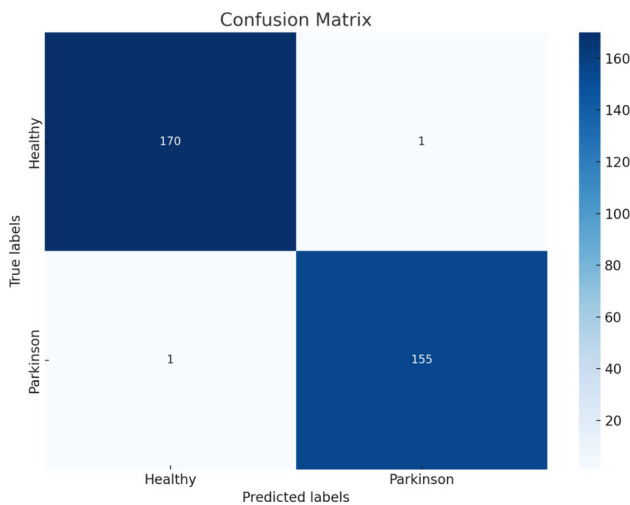
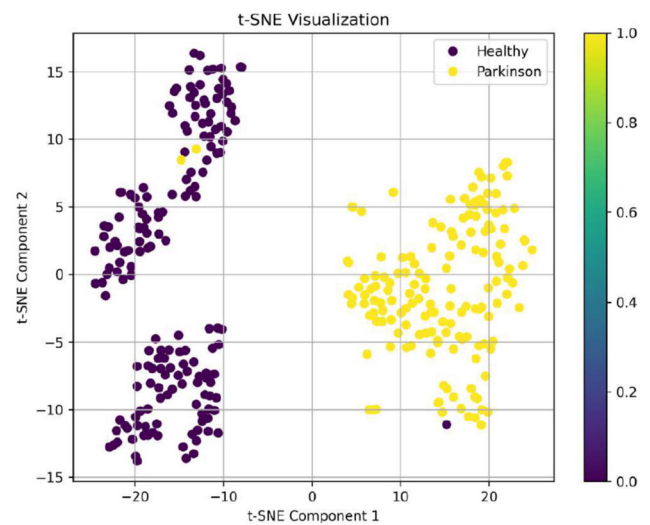
**Fig. 4** Confusion matrix for DenseNet201 + Xception + NASNet-Mobile with NCA and SVM

Figure 5 visually evaluates the performance of the model, composed of the InceptionV3 + DenseNet201 + Xception combination, using the t-SNE approach.

Figure 5 visualizes the distinctiveness of the features obtained in the final stage for the test samples using the

**Fig. 5** Feature space distribution visualization of DenseNet201 + Xception + NASNet-Mobile model using t-SNE

t-SNE approach. t-SNE is a visualization technique that transforms high-dimensional features of a dataset into a lower-dimensional space, clustering similar samples together while separating them from other classes. t-SNE



operates in two stages: first, it estimates similarity values and then performs a low-dimensional embedding to preserve the similarity structures between the samples in the dataset. t-SNE is excellent for visualizing complex data structures and patterns, revealing clustering trends and relationships between dataset samples. Figure 5 shows the distribution of features for the test samples using t-SNE. The plot displays two distinct clusters representing the healthy and PD classes. Purple dots represent the healthy class, while yellow dots represent the PD class. The clear separation between these clusters indicates that the model effectively distinguishes between the two classes. The tight clustering of similar points suggests that the model has learned meaningful features, using these features to distinguish healthy individuals from those with PD. This t-SNE visualization confirms the model's effectiveness in classifying the dataset. The proposed hybrid model's features are highly distinctive, and the model's success is evident. This visualization shows that the model has effectively learned the complex data structures and clustering trends, using this knowledge successfully in classification tasks. Table 6 presents a comparison of the proposed approach with state-of-the-art studies conducted on the Parkinson Hand Drawings dataset, evaluating the models based on the metrics of accuracy, precision, sensitivity, and F1-score.

The Parkinson Hand Drawing Dataset, consisting of spiral and wave drawings, serves as a crucial resource for examining motor impairments associated with PD. Table 6 presents state-of-the-art studies utilizing this dataset. While some of these studies focus solely on spiral images, others analyze both types of drawings or develop separate models for each. MORALES-CASTRO et al. [34] found that employing ResNet50 as a feature extractor and SVM as a classifier resulted in the best performance for PD identification, with an accuracy of 89%. Kamran et al. [33] proposed an end-to-end deep transfer learning method for the early diagnosis of PD using handwriting. By using various deep transfer CNN architectures, combining different PD handwriting datasets, and applying data augmentation techniques, they achieved a 90% accuracy score with GoogleNet. Das et al. [42] introduced an approach aiming to expedite the costly process of PD diagnosis. They conducted a comparative analysis evaluating the effectiveness of manually crafted features versus deep-level features, achieving a notable accuracy of 93%. Chakraborty et al. [32] developed a system for detecting PD using spiral and wave drawings with two distinct CNNs. They employed logistic regression and random forest meta-classifiers with ensemble voting, achieving 93.3% accuracy. This multi-stage approach enhances PD detection precision and highlights the importance of combined decision-making for better clinical interventions. Hossain et al. [43] introduced

the MetaParkinson model, an advanced health framework combining industrial cyber-physical systems with a meta-learning approach. This model achieves a diagnostic accuracy of 95.0% for spiral images and 90.0% for wave images in a 10-shot training setup, training separate models for spiral and wave data, using a CNN encoder and Siamese network for PD classification. Saleh et al. [2] proposed a hybrid method for PD diagnosis by combining spiral and wave CNN-KNN architectures, achieving 96.67% accuracy. This approach leverages CNN's feature extraction and KNN's classification through ensemble voting, capturing nuanced data relationships and preventing overfitting. Jahan and Nesa [39] utilized the ResNet50 model with transfer learning to diagnose PD using spiral and wave images. By employing data thinning and augmentation, they enhanced model performance, achieving an accuracy of 96.67%, thereby improving the reliability of PD detection. Kumar and Bansal [35] introduced a modified MobileNetV2 model for real-time PD detection on mobile devices, achieving 97.70% accuracy using spiral and wave hand drawings. Their lightweight, efficient model highlights the diagnostic value of hand drawings for Parkinson's, offering a non-invasive and accurate method for detection. Fiza et al. [38] integrated CNN and ANN with GridSearchCV to effectively utilize drawing and acoustic features for Parkinson's detection, leading to a high classification accuracy of 98.0%. Zhou et al. [37] proposed the Diplin model, which achieves outstanding performance in image classification by incorporating WGAN and migration learning, exhibiting a remarkable accuracy rate of 98% on the validation dataset. Varalakshmi et al. [1] conducted a study focusing on predicting PD using only spiral images. They found that hybrid models like ResNet50 + SVM outperform other machine learning, deep learning, and hybrid models, achieving an accuracy rate of 98.45%. Krishnasmorthy et al. [36] developed the Lf-HWFRNet system for early PD detection using handwritten samples. The system integrates FRCNN and BiGRU with a weighted average ensemble technique and optimized hyperparameters using LFDO, achieving a remarkable PD classification accuracy of 98.82%. Many of these studies emphasize the combination of deep learning and traditional machine learning techniques in detecting PD using hand drawings. The proposed hybrid model successfully identified Parkinson's patients with a significant accuracy rate of 99.39% across six different combinations utilizing various feature extraction models. This research highlights the combination of features obtained from binary and ternary combinations of transfer learning models, followed by feature selection and dimensionality reduction using the NCA method, and ultimately classification using the SVM algorithm, significantly enhancing performance. This performance is notably remarkable

**Table 6** Comparison between the proposed model with previous literature (only Parkinson Hand Drawing Dataset)

Study	Dataset	Year	Features	Classifier	Accuracy (%)	Precision (%)	Sensitivity (%)	F1-Score (%)
[32]	Spiral-Wave	2020	CNN-based features	LR, RFC	93.3	93.5	94	93.94
[39]	Spiral-Wave	2021	CNN-based features	FC layers	96.67	–	–	–
[33]	Spiral-Wave	2021	CNN-based features	MLP	90	–	–	–
[42]	Spiral-Wave	2021	HOG	MLP	93	–	–	–
[38]	Spiral-Wave	2022	CNN-based features	Decision Tree	98	100	100	98
[34]	Spiral-Wave	2022	CNN-based features	SVM	89	90	89	87
[1]	Spiral	2022	CNN-based features	SVM	Spiral: 98.45	Spiral: 99	Spiral: 99	Spiral: 99
[43]	Spiral-Wave	2023	Siamese Networks features	Meta-learning framework	Spiral: 95, Wave: 90	Spiral: 97.4, Wave: 90.3	Spiral: 96.1, Wave: 93.2	–
[36]	Spiral-Wave	2023	Faster RCNN, BiGRU	WAE	98.82	97.75	94.13	92.64
[35]	Spiral-Wave	2023	CNN-based features	Dense layer	97.70	98	98	98
[37]	Spiral-Wave	2023	CNN-based features	FC layer	98	97	96	97
[3]	Spiral-Wave	2024	CNN-based features	MLP	94	95	94	94
[2]	Spiral-Wave	2024	CNN-based features	Ensemble voting classifier	96.67	100	93.33	96.55
Proposed approach	Spiral-Wave	–	CNN Fusion features	NCA + SVM	99.39	99.36	99.36	99.36

when compared to the accuracy rates ranging from 89% to 98.82% in other research. The strong performance of the proposed approach in identifying PD suggests the potential utility of such models in clinical applications.

## 5 Conclusion

This study proposes a new hybrid model consisting of CNN, NCA, and SVM algorithms to diagnose PD based on handwritten drawings. The primary aim is to develop a feature extraction methodology that captures the geometric dynamics of handwriting associated with symptoms of PD. In this context, the performance of fine-tuned transfer learning models on the Parkinson Hand Drawings dataset was compared using several key metrics, i.e., accuracy, specificity, sensitivity, precision, F1 score, training time, inference time, number of parameters, and computational load (Table 2). The analyses revealed significant disparities in the relevant metrics among the models examined.

InceptionV3 achieved the highest performance with an accuracy of 97.55%, efficient training (3.22 min), and inference times (0.04 min). It also excelled in other metrics like specificity, sensitivity, precision, and F1 score. Xception followed closely with 97.25% accuracy and strong performance across all metrics. InceptionResNetV2 also performed well, achieving 96.64% accuracy, but it had the highest parameter count (54.3M) and computational load (13B FLOPs). DenseNet201 and NASNetMobile demonstrated moderate performance, with accuracies of 81.35% and 80.73%, and balanced results in other metrics. ResNet50 and MobileNetV2 had accuracies of 65.14% and 77.06%, respectively, with varied performance in other metrics. EfficientNetB0 and VGG16 were the least effective, with accuracies of 52.60% and 48.62%, and lower scores across all metrics. Overall, InceptionV3 and Xception were top performers across all evaluated metrics, while EfficientNetB0 and VGG16 showed the lowest performance.

Previous research has demonstrated that using CNNs as feature extractors and SVM as the classifier outperforms using CNN classification layers. In this study, the NCA approach was implemented before SVM to filter out irrelevant features, discarding 83% of the features. Initially, without NCA and SVM (Table 2), InceptionV3, Xception, and InceptionResNetV2 showed high accuracies of 97.55%, 97.25%, and 96.64%, respectively. Upon applying NCA and SVM (Table 3), these models either maintained or improved their performance, with InceptionV3 and Xception both achieving 97.55%, and InceptionResNetV2 increasing to 98.47%. Other models also saw significant improvements: DenseNet201 from 81.35% to 96.02%, MobileNetV2 from 77.06% to 93.88%, and EfficientNetB0 from 52.60% to 87.16%. ResNet50 saw a modest increase in accuracy from 65.14% to 70.03%, while VGG16's accuracy remained low at 47.71%. Overall, the combination of CNN, NCA, and SVM enhanced the models' performance.

An important contribution of this study is to demonstrate that utilizing various combinations of transfer learning models yields an effective approach to capturing handwriting geometric dynamics related to disease symptoms. In this context, initially, binary combinations of these transfer learning models were examined, and it was observed that the features generated through these combinations enhance model performance (Table 4). In this regard, 5 different binary combinations, such as InceptionV3 + EfficientNetB0, achieved an impressive accuracy of 99.1%. When the model combinations with the lowest performance were investigated, pairing the VGG16 model, which previously achieved a standalone accuracy rate of 47.71%, with the ResNet50 model, which has an accuracy rate of 70.03%, resulted in the lowest accuracy rate of 71.6%. This suggests that utilizing the complementary strengths of different models can result in improved performance. To identify the best combination, it is essential to consider the number of parameters, training time, inference time, and FLOPs. Among the combinations, InceptionV3 + EfficientNetB0 and Xception + NASNet-Mobile emerge as the most efficient. The InceptionV3 + EfficientNetB0 combination has approximately 29.2M parameters. EfficientNetB0's design allows for faster training and inference times, and its lower FLOPs enhance computational efficiency. Similarly, the Xception + NASNet-Mobile combination has around 28.2M parameters, offering even better parameter efficiency. This combination also benefits from faster training and inference times due to NASNet-Mobile's efficient architecture and low FLOPs, contributing to its computational efficiency. When comparing these combinations, Xception + NASNet-Mobile stands out as the most efficient in terms of parameters, computational demands, and speed. This

combination provides a balanced approach with high accuracy and practical efficiency, making it the preferred choice for applications that require optimal performance with limited computational resources. In this context, the effect of combining features obtained from different transfer learning models in triple combinations was also evaluated (Table 5). Among the experimental results, six distinct combinations achieved the highest performance, each with an accuracy of 99.39%. However, the combination of DenseNet201, Xception, and NASNet-Mobile emerged as the most efficient model when considering additional metrics. This combination has 43.5M parameters, medium training and inference times, and low FLOPs, maintaining lower computational costs while achieving the same high accuracy as the other top-performing models. In these experimental trials, similar to previous experimental studies, the NCA approach was used after feature fusion, and SVM was used for classification. The findings provide valuable insight into the relationship between various combinations of transfer learning models.

This study also contributes to identifying transfer learning models that are compatible with one another and lead to better feature extraction when utilized together. Experimental results show that the Inception and Xception models are compatible with other models and improve performance. On the other hand, the VGG16 model exhibits low performance for PD and adversely affects the overall performance. Furthermore, it was discovered that the ResNet50 model made no substantial contributions to other models after VGG16 and had a minor impact on overall performance. These findings give an important insight for understanding transfer learning models' behaviors and determining suitable combinations.

In addition, the t-SNE approach was used to assess the proposed hybrid model's performance qualitatively. The t-SNE method maps the obtained feature vectors to a low-dimensional space, positioning similar instances close to each other and dissimilar instances far apart. In this context, visualization using the feature set obtained from the combination of DenseNet201, Xception, and NASNet-Mobile, which achieved the highest accuracy, revealed the feature set to be highly discriminative.

Finally, the proposed approach was compared with state-of-the-art studies in the literature using the Parkinson Hand Drawings dataset, as detailed in Table 6. This comparison evaluated models based on accuracy, precision, sensitivity, and F1-score. Some of these studies focused on spiral images, while others concentrated on both types or developed separate models for each. The proposed hybrid model achieved an outstanding accuracy of 99.39%, surpassing other studies with accuracy rates between 89% and 98.82%, highlighting its potential for clinical applications in early and accurate PD diagnosis. The success of the

proposed approach in this study lies in presenting an innovative hybrid method that combines deep learning and machine learning algorithms. Due to the limited number of samples in the dataset, not all features obtained from transfer learning models may be meaningful, and the fully connected layers in these models can lead to overfitting. To address this issue, the study extracts a combination of features from different models and uses NCA to eliminate less informative features. This enhances the model's discriminative power by selecting only the most relevant features. In the classification stage, SVM is used to process the enriched feature vectors. SVMs are effective in maximizing the margin between classes, making them less prone to overfitting with small datasets. By integrating various transfer learning models, selecting features efficiently with NCA, and classifying robustly with SVM, this approach ensures superior performance.

## 6 Future work

Future research will focus on applying the proposed hybrid model to different health datasets to investigate its performance across various medical conditions. This will involve obtaining and analyzing datasets related to other neurodegenerative disorders, cardiovascular diseases, and various types of cancers to evaluate the model's adaptability and effectiveness in diverse medical contexts.

Furthermore, to enhance the model's performance on the new datasets to be utilized in future research, various methods for both feature selection and classification will be investigated. Specifically, methods such as principal component analysis, linear discriminant analysis, and uniform manifold approximation and projection will be considered replacements for NCA. These methods will be evaluated for their ability to retain the most informative features while reducing dimensionality, ultimately aiming to enhance classification accuracy and model efficiency. Similarly, other classification models like decision trees, random forests, k-nearest neighbors, logistic regression, and gradient boosting models will be evaluated in place of SVM. These models will be assessed for their performance in terms of accuracy, robustness, and computational efficiency. Additionally, ensemble methods that combine multiple classifiers could be investigated to further improve diagnostic performance.

Another promising direction for future research is the application of generative adversarial networks (GANs) for data augmentation and synthetic data generation. GANs can significantly enhance the size and diversity of training datasets, which is crucial for medical data that are often scarce. By producing synthetic samples that mimic real data, GANs provide a broader range of examples,

improving the training process. This is especially vital for small datasets where traditional methods fall short. The use of GANs helps mitigate overfitting, ensuring the model remains robust and generalizable.

Finally, collaboration with healthcare professionals will be sought to ensure the practical applicability of the model in real-world clinical settings. This includes conducting clinical trials to validate the model's effectiveness and user studies to gather feedback from medical practitioners. The ultimate goal is to develop a user-friendly diagnostic tool that seamlessly integrates into existing healthcare workflows, thereby aiding in the early detection and treatment of various diseases.

**Funding** No Funding

**Data availability** Dataset Public

## Declarations

**Conflict of interest** The authors declare no Conflict of interest.

## References

1. Varalakshmi P, Priya BT, Rithiga BA, Bhuvaneaswari R, Sundar RSJ (2022) Diagnosis of Parkinson's disease from hand drawing utilizing hybrid models. *Parkinsonism Relat Disord* 105:24–31
2. Saleh S, Ouhmida A, Cherradi B, Al-Sarem M, Hamida S, Alblwi A, Mahyoob M, Bouattane O (2024) A novel hybrid cnn-knn ensemble voting classifier for parkinson's disease prediction from hand sketching images. *Multimed Tools Appl* 1–33
3. Aldhyani TH, Al-Nefaie AH, Koundal D (2024) Modeling and diagnosis Parkinson disease by using hand drawing: deep learning model. *AIMS Math* 9(3):6850–6877
4. Rajinikanth V, Yassine S, Bukhari SA (2024) Hand-sketches based Parkinson's disease screening using lightweight deep-learning with two-fold training and fused optimal features. *Int J Math Statist Comput Sci* 2:9–18
5. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 2818–2826
6. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 4700–4708
7. Tan M, Le Q (2019) Efficientnet: rethinking model scaling for convolutional neural networks. In: *International Conference on Machine Learning*, pp 6105–6114. PMLR
8. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 770–778
9. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen L-C (2018) Mobilenetv2: inverted residuals and linear bottlenecks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 4510–4520



10. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556)
11. Chollet F (2017) Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 1251–1258
12. Zoph B, Vasudevan V, Shlens J, Le QV (2018) Learning transferable architectures for scalable image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 8697–8710
13. Szegedy C, Ioffe S, Vanhoucke V, Alemi A (2017) Inception-v4, inception-resnet and the impact of residual connections on learning. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol 31
14. Goldberger J, Hinton GE, Roweis S, Salakhutdinov RR (2004) Neighbourhood components analysis. *Adv Neural Inf Process Syst* **17**
15. Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B (1998) Support vector machines. *IEEE Intell Syst Appl* **13**(4):18–28
16. Saleh S, Cherradi B, El Gannour O, Hamida S, Bouattane O (2024) Predicting patients with Parkinson's disease using machine learning and ensemble voting technique. *Multimed Tools Appl* **83**(11):33207–33234
17. Wang X, Huang J, Chatzakou M, Medijainen K, Toomela A, Nõmm S, Ruzhansky M (2024) LSTM-CNN: an efficient diagnostic network for Parkinson's disease utilizing dynamic handwriting analysis. *Comput Methods Prog Biomed* **247**:108066
18. Bdaqli M, Shoeibi A, Moridian P, Sadeghi D, Pouyani MF, Shalhaf A, Gorriz JM (2024) Diagnosis of Parkinson disease from EEG signals using a CNN-LSTM model and explainable AI. In: International Work-Conference on the Interplay Between Natural and Artificial Computation, pp 128–138. Springer
19. Sun M, Jung W, Koltermann K, Zhou G, Watson A, Blackwell G, Helm N, Cloud L, Pretzer-Aboff I (2023) Parkinson's disease action tremor detection with supervised-learning models. In: Proceedings of the 8th ACM/IEEE International Conference on Connected Health: Applications, Systems and Engineering Technologies, pp 1–10
20. Yang Y, Yuan Y, Zhang G, Wang H, Chen Y, Liu Y, Tarolli CG, Crepeau D, Bukartyk J, Junna MR, Videnovic A, Ellis TD, Lipford MC, Dorsey R, Katabi D (2022) Artificial intelligence-enabled detection and assessment of Parkinson's disease using nocturnal breathing signals. *Nat Med* **28**(10):2207–2215
21. George N, Kulkarni S, James J, Parsewar S, Shriram R (2021) Detection of Parkinson's disease through speech and smell signatures. In: Advances in Electrical and Computer Technologies: Select Proceedings of ICAECT 2020, pp 659–670. Springer
22. Bhan A, Kapoor S, Gulati M, Goyal A (2021) Early diagnosis of Parkinson's disease in brain mri using deep learning algorithm. In: 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), pp 1467–1470. IEEE
23. Wang X, Hao X, Yan J, Xu J, Hu D, Ji F, Zeng T, Wang F, Wang B, Fang J et al (2023) Urine biomarkers discovery by metabolomics and machine learning for Parkinson's disease diagnoses. *Chin Chem Lett* **34**(10):108230
24. Folador JP, Santos MCS, Luiz LMD, Souza LAPS, Vieira MF, Pereira AA, Oliveira Andrade A (2021) On the use of histograms of oriented gradients for tremor detection from sinusoidal and spiral handwritten drawings of people with parkinson's disease. *Med Biol Eng Comput* **59**:195–214
25. Drotár P, Mekyska J, Smékal Z, Rektorová I, Masarová L, Faundez-Zanuy M (2015) Contribution of different handwriting modalities to differential diagnosis of Parkinson's disease. In: 2015 IEEE International Symposium on Medical Measurements and Applications (MeMeA) Proceedings, pp 344–348. IEEE
26. Drotár P, Mekyska J, Rektorová I, Masarová L, Smékal Z, Faundez-Zanuy M (2016) Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease. *Artif Intell Med* **67**:39–46
27. Impedovo D (2019) Velocity-based signal features for the assessment of Parkinsonian handwriting. *IEEE Signal Process Lett* **26**(4):632–636
28. Naseer A, Rani M, Naz S, Razzak MI, Imran M, Xu G (2020) Refining Parkinson's neurological disorder identification through deep transfer learning. *Neural Comput Appl* **32**:839–854
29. Pereira CR, Pereira DR, Silva FA, Masieiro JP, Weber SA, Hook C, Papa JP (2016) A new computer vision-based approach to aid the diagnosis of Parkinson's disease. *Comput Methods Prog Biomed* **136**:79–88
30. Pereira CR, Weber SA, Hook C, Rosa GH, Papa JP (2016) Deep learning-aided Parkinson's disease diagnosis from handwritten dynamics. In: 2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), pp 340–346. IEEE
31. Pereira CR, Pereira DR, Rosa GH, Albuquerque VH, Weber SA, Hook C, Papa JP (2018) Handwritten dynamics assessment through convolutional neural networks: an application to Parkinson's disease identification. *Artif Intell Med* **87**:67–77
32. Chakraborty S, Aich S, Han E, Park J, Kim H-C, et al. (2020) Parkinson's disease detection from spiral and wave drawings using convolutional neural networks: a multistage classifier approach. In: 2020 22nd International Conference on Advanced Communication Technology (ICACT), pp 298–303. IEEE
33. Kamran I, Naz S, Razzak I, Imran M (2021) Handwriting dynamics assessment using deep neural network for early identification of parkinson's disease. *Futur Gener Comput Syst* **117**:234–244
34. Morales-Castro JC, Ruiz-Pinales J, Lozano-García JM, Guzmán-Cabrera R (2022) Use of image processing for the detection of Parkinson's disease. *J Physiother Med Technol* **6**:27–32
35. Kumar BA, Bansal M (2023) A transfer learning approach with mobilenetv2 for parkinson's disease detection using hand-drawings. In: 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), pp 1–7. IEEE
36. Krishnsmoorthy S, Zhang Y, Kadry S, Mustafa N, Sumitha C, Sundararaj V, MR R. Decoding handwriting patterns with the If-hwfrnet classifier: A novel method for early Parkinson's disease detection and improved patient care. Available at SSRN 4404864
37. Zhou F, Hu S, Wan X, Lu Z, Wu J (2023) Diplin: a disease risk prediction model based on efficientnetv2 and transfer learning applied to nursing homes. *Electronics* **12**(12):2581
38. Fiza S, Kumar A, Yadav J, Reddy S, Sidda TK, Sai K (2022) Classification of parkinson's disease using CNN and ANN with the aid of drawing and acoustic feature. In: 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), pp 513–517. IEEE
39. Jahan N, Nesa A, Layek MA (2021) Parkinson's disease detection using resnet50 with transfer learning. *Int J Comput Vis Signal Process* **11**(1):17–23
40. Mader K (2017) Parkinson's drawings. Accessed: 2024-06-01. <https://www.kaggle.com/datasets/kmader/parkinsons-drawings>
41. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L (2009) Imagenet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp 248–255. IEEE
42. Das A, Das HS, Neog A, Bharat Reddy B, Choudhury A, Swargiary M (2021) Detection of Parkinson's disease from hand-drawn images using machine learning algorithms. In: Soft Computing and Signal Processing: Proceedings of 3rd ICSCSP 2020, Vol 1, pp 241–252. Springer

43. Hossain MS, Shorfuzzaman M (2023) Metaparkinson: a cyber-physical deep meta-learning framework for n-shot diagnosis and monitoring of Parkinson's patients. *IEEE Syst J* 17:5251

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.