International Journal of Computational and Experimental Science and ENgineering (IJCESEN)

Vol. 11-No.3 (2025) pp. 4483-4492
http://www.ijcesen.com

ISSN: 2149-9144

Research Article

Machine Learning Approach for Parkinson's Disease Detection: A Comparative Study of SVM Kernels on DaTSCAN Data

Nandan N[1*], Sanjay Pande M B[2], Raveesh B N[3], Rakesh Pande M S[4], Chethan Raj C[5], Sharmila D[6]

[1]Research Scholar, Department of CSE, GM Institute of Technology, Davanagere, Visvesveraya Technological University, Belagavi, Karnataka, India.
*** Corresponding Author Email:** nandan.mpn@gmail.com – **ORCID**: 0009-0006-2498-769X

[2]Research Supervisor, GM Institute of Technology, Davanagere, Visvesveraya Technological University, Belagavi, Karnataka, India.
Email: rakroop99@gmail.com - **ORCID:** 0000-0001-5995-283X

[3]Department of Psychiatry, Mysore Medical College, Mysuru, Karnataka, India.
Email: raveesh6@gmail.com **-ORCID:** 0000-0003-2766-9997

[4]Dental surgeon, Private Practitioner, Mysuru, Karnataka, India
Email: rakeshpande021@gmail.com - **ORCID:** 0009-0003-0028-666X

[5]Research Scholar, Department of studies in Computer Science, University of Mysore, Mysore
Email: cchethan005@gmail.com – **ORCID**: 0000-0002-8235-5109

[6]Selection Grade Lecturer, Government Polytechnic, Arakere, Karnataka, India
Email: sharmila51076@gmail.com - **ORCID:** 0009-0003-5877-6565

Article Info:

DOI: 10.22399/ijcesen.3090
Received : 05 May 2025
Accepted : 23 June 2025

Keywords

Parkinson's disease
DaTSCAN
Support Vector Machine
RBF kernel
Polynomial kernel
Sigmoid Kernel

Abstract:

Parkinson's disease (PD) detection using machine learning presents significant potential for improving diagnostic accuracy. This study investigates the classification of PD patients and healthy controls (HC) using Striatal Binding Ratio (SBR) values derived from DaTSCAN imaging data. Initial exploratory analysis, including Principal Component Analysis (PCA), revealed a complex, nonlinear data distribution, prompting the use of models adept at handling such patterns. The study primarily evaluates Support Vector Machines (SVM) with different kernel functions—Radial Basis Function (RBF), Polynomial, and Sigmoid—leveraging their ability to model nonlinear relationships. Comparative analysis demonstrated that the SVM-RBF kernel outperformed other kernels, achieving 98.12% accuracy. The Polynomial kernel followed with 94.63% accuracy (C=10, degree=3), while the Sigmoid kernel lagged at 91.68%. The superior performance of the RBF kernel underscores its effectiveness in capturing the intricate nonlinear patterns in DaTSCAN SBR data. Furthermore, when benchmarked against Random Forest, Logistic Regression, K-Nearest Neighbors (KNN), and Convolutional Neural Networks (CNN), the SVM-RBF model consistently exhibited the highest classification accuracy. This study establishes that an optimized SVM with an RBF kernel provides a robust and highly accurate machine learning approach for distinguishing PD patients from healthy controls using DaTSCAN data.

1. Introduction

Parkinson's disease (PD) is a progressive neurodegenerative disorder characterized by the degeneration of dopaminergic neurons in the substantia nigra pars compacta, leading to debilitating motor dysfunction. Clinically, PD manifests with cardinal symptoms, including bradykinesia, resting tremor, rigidity, and postural instability [1]. Traditional diagnosis relies on subjective clinical assessments of these motor features, which often leads to misdiagnosis due to overlapping symptomatology with atypical Parkinsonian syndromes. Moreover, non-motor symptoms—such as olfactory dysfunction, REM sleep behavior disorder, and autonomic disturbances—frequently precede motor onset but

are under-recognized in early-stage detection. To address these diagnostic challenges, machine learning (ML) has emerged as a transformative tool, enabling the data-driven quantification of biomarkers to improve diagnostic accuracy, sensitivity, and specificity.

The striatum, a critical component of the basal ganglia, comprises the caudate nucleus and putamen, which play pivotal roles in motor control and dopaminergic signaling. Embryologically and functionally linked, these structures integrate cortical inputs to modulate movement via the direct and indirect pathways of the basal ganglia-thalamocortical circuit [2]. The caudate nucleus exhibits a C-shaped morphology, anatomically subdivided into the head (adjacent to the lateral ventricle), body (dorsal to the thalamus), and tail (extending into the temporal lobe to synapse with the amygdala) [3]. It is separated from the putamen by the internal capsule, except at its rostral aspect, where the two structures merge. The putamen, the outermost division of the lentiform nucleus, lies lateral to the globus pallidus and serves as the primary input nucleus for motor regulation [4].

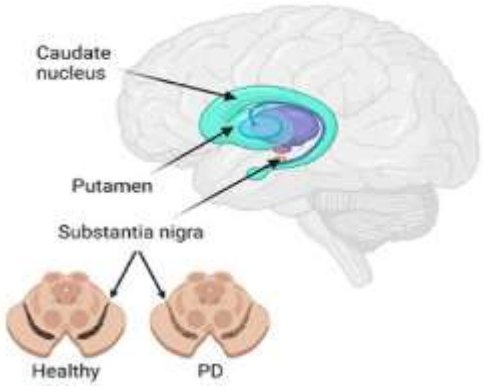


Figure 1. *Substania Nigra region of Brain [25]*

The extrapyramidal motor system, which mostly controls voluntary movements, depends on the caudate nucleus and putamen shown in the Figure 1. These structures prevent undesired or maladaptive motions while facilitating the performance of optimal motor actions. By affecting the timing, scale, and retention of motor programs for well-known tasks, they play a crucial role in the planning and programming of motions [5]. As seen in Parkinson's disease, where dopaminergic depletion impairs motor performance, disorders in the caudate nucleus and putamen can cause severe abnormalities in motor control. This results in recognisable symptoms such as tremors, stiffness, and akinesia, underscoring the vital function of these nuclei in preserving motor integrity [6]. Expressing dopaminergic signalling, the caudate nucleus and putamen have essential roles in the circuitry of the basal ganglia, modulating both motor control and cognitive processes. Dopamine is a critical neurotransmitter involved in processing emotions, motor control, and reinforcement of the brain's reward system. Figure 2 shows the decrease in the concentration of dopamine in Parkinson's patients and Healthy Control in DaTSCAN image.

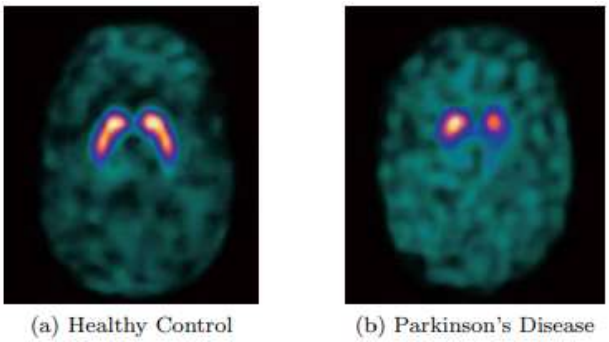


Figure 2. *DaTscan with putamen and caudate regions marked by high contrast [17]*

Dopamine receptors, especially those of the D1 and D2 subtypes, are very abundant in the caudate and putamen, and are critical for interpreting dopaminergic signals originating from the substantia nigra [7]. This dopaminergic interaction facilitates cognitive and motor processes, including learning, habit formation, and reward processing. Dopamine signaling through these structures is crucial in controlling motivational behaviours and movements, allowing adaptive responses to reward-seeking behaviours [8].

Parkinson's disease (PD) is characterized by the degeneration of dopaminergic neurons in the substantia nigra, leading to depleted dopamine levels in the caudate and putamen. This neuronal loss disrupts motor function, causing hallmark symptoms such as rigidity, bradykinesia, and tremors [9]. Beyond motor impairments, PD also affects cognitive functions, impairing movement planning, execution, and reward-based learning.

Recent advances in artificial intelligence (AI) have enabled more accurate PD detection using machine learning (ML) and deep learning (DL) techniques. Studies leveraging datasets from the Parkinson's Progression Markers Initiative (PPMI) have employed learning to achieve high diagnostic accuracy, sensitivity, and specificity [10-11]. AI-driven approaches, particularly those integrating multi-modal data and advanced feature extraction, show significant promise in early PD diagnosis. However, challenges remain, including the need for large, high-quality datasets, model interpretability

issues, and risks of overfitting [12]. Addressing these limitations could further enhance AI's role in revolutionizing PD diagnosis and treatment strategies.

2. Related Works

Parkinson's Disease (PD) diagnosis has been significantly advanced through machine learning (ML) and deep learning (DL) techniques, leveraging diverse data modalities such as neuroimaging (DAT-SPECT, DaTscan SPECT, MRI), clinical assessments, and genetic markers. Recent studies highlight the efficacy of deep neural networks in PD detection, though methodological and generalizability challenges persist. One notable study employed an AlexNet-based artificial neural network (ANN) to analyze DAT-SPECT images, specifically focusing on the putamen region, achieving an accuracy of 86% in PD identification [13]. While this demonstrates the feasibility of ANNs in PD diagnosis, the study's reliance on a limited dataset raises concerns about model generalizability. Additionally, variability in imaging protocols across different medical institutions may further constrain the model's applicability, suggesting a need for larger, multi-center datasets to enhance robustness.Deep CNNs have shown remarkable performance in PD classification using DaTscan SPECT images. A study leveraging the InceptionV3 architecture, pre-trained on ImageNet, achieved an impressive 98.48% accuracy on the Parkinson's Progression Markers Initiative (PPMI) dataset (n=659, 449 PD, 210 non-PD) [14]. Key methodological strengths included rigorous preprocessing to enhance dopaminergic regions, data augmentation, and ten-fold cross-validation to mitigate overfitting. To enhance diagnostic stability, an ensemble approach combining four pre-trained CNNs (VGG16, ResNet50, InceptionV3, Xception) with Fuzzy Rank Level Fusion (FRLF) was proposed, achieving 98.45% accuracy on the PPMI dataset (n=645, 432 PD, 213 non-PD) [15]. The FRLF method effectively aggregated predictions from multiple models, reducing individual network biases.

Classical machine learning (ML) techniques have been extensively applied to Parkinson's disease (PD) diagnosis, though their performance varies significantly depending on methodology and data constraints. A comparative study leveraging PPMI data (n=548) evaluated Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) for feature reduction, with LDA outperforming PCA in discriminative capability. Among clustering methods (DBSCAN, K-means, Hierarchical Clustering), Hierarchical Clustering achieved the highest accuracy (64%), though the study acknowledged limitations in handling high-dimensional data and the necessity for external validation [16].

Recent advances in deep learning have demonstrated superior performance in PD classification. A transfer learning approach using VGG16 on DaTSCAN images (PPMI) achieved 95.2% accuracy, with model interpretability enhanced through the Local Interpretable Model-Agnostic Explainer (LIME) framework. However, the authors highlighted the need for broader validation and further refinement of interpretability mechanisms [17]. Similarly, an automated CAD system employing Haralick texture features and SVM classification on DaTSCAN SPECT images attained 95.9% accuracy and 97.3% sensitivity, though computational complexity and image quality dependencies were noted as key limitations [18]. Further improvements were proposed through advanced ML architectures and larger datasets.

Hybrid deep learning models have shown particular promise. A fusion of VGG-16 and AlexNet for feature extraction, combined with a Multi-Kernel SVM (MSVM), achieved remarkable classification accuracy (98.60%) on PPMI DaTSCAN images. Despite this, the study emphasized the critical need for validation across diverse, larger cohorts to ensure generalizability [19]. Multimodal approaches integrating MRI, SPECT, and CSF biomarkers were explored using both feature-level (deep learning on combined data) and modal-level (MRI feature reduction followed by fusion) frameworks. While CNNs achieved accuracies of 93.33% and 92.38%, respectively, the study's conclusions were constrained by a small, imbalanced dataset (73 PD, 59 healthy) [20].

Beyond imaging, genomic and clinical data have also been leveraged. A GenoML-based model trained on PPMI data and externally validated with PDBP achieved an AUC of 89.72% (85.03% in external validation). The authors suggested further optimization via hyperparameter tuning and multimodal integration [21]. Premotor symptom-based models have also been investigated. A deep learning model outperformed twelve other ML methods in classifying early-stage PD (96.45% accuracy) using premotor features (REM sleep behavior disorder, olfactory loss) from PPMI data (183 healthy, 401 PD). Despite high accuracy, the small sample size restricted broader applicability [22].

A comparison of parametric (logistic regression) and non-parametric (KNN) models on PPMI data (n=919) revealed that optimized KNN (96.8%

accuracy) surpassed logistic regression (94.82%), with ANOVA-driven feature selection. However, the need for larger datasets and cross-validation remains a limitation [23]. Finally, a Random Forest model analyzing Striatal Binding Ratio (SBR) values in DaTSCAN images (PPMI, n=2,071) achieved 97% accuracy, underscoring the potential of early detection and multimodal enhancements [24]. While existing studies demonstrate the efficacy of ML in PD diagnosis, several limitations persist: reliance on single-center datasets (e.g., PPMI) raises concerns about generalizability; small or imbalanced cohorts limit statistical power; interpretability remains a challenge despite tools like LIME; and computational complexity hinders clinical translation. Future work should prioritize analyzing the data patterns (linear or nonlinear), large-scale multimodal validation, robust interpretability frameworks, and efficient model architectures for real-world deployment.

3. Proposed Methodology

3.1 Dataset

The dataset comprises Striatal Binding Ratio (SBR) values derived from DaTSCAN imaging, specifically focusing on the caudate and putamen regions of the brain. These measurements were obtained from the Parkinson's Progression Markers Initiative (PPMI), a leading research database dedicated to Parkinson's disease (PD) studies. The dataset taken is already preprocessed by removing the duplicate values and outliers using the iForest algorithm, which is having 1,862 samples, with 1,449 from Parkinson's disease patients and 413 from healthy controls [24], reflecting a real-world clinical distribution where PD cases are more prevalent than controls.

The key features in this dataset are the DaTSCAN SBR values for six specific brain regions: the right and left caudate, right and left putamen, and right and left anterior putamen. These measurements are critical in assessing dopamine transporter (DAT) availability, which is typically reduced in Parkinson's disease due to the degeneration of dopaminergic neurons.

The PCA visualization of the DaTSCAN dataset (shown in the image) reveals a nonlinear distribution of data points, where healthy controls and Parkinson's disease patients are not separable by a straight line or simple hyperplane, as shown in Figure 3. Instead, the clusters appear intertwined or follow curved patterns, indicating that the relationship between the features (caudate and putamen SBR values) and the target classes (PD vs. healthy) is complex and nonlinear. This nonlinearity arises because neurodegenerative processes like PD affect brain regions in non-uniform ways, leading to intricate interactions between variables that cannot be captured by linear methods alone.

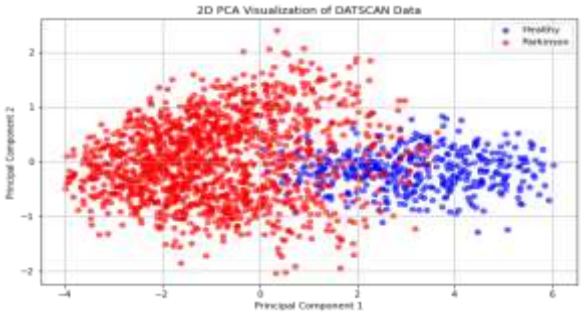


Figure 3. 2D Visualization of DaTSCAN Dataset

In linear datasets, classes can be separated using straight lines (in 2D) or hyperplanes (in higher dimensions), making models like Linear SVM or Logistic Regression effective. However, nonlinear datasets (like the DaTSCAN data) require more sophisticated approaches because of boundaries are curved or irregular – A linear decision boundary would misclassify many points, feature interactions matter – Dopamine loss in PD may not follow a linear trend across brain regions and Higher-dimensional patterns exist – Linear projections (like PCA) may collapse critical nonlinear separability.

The kernel functions in SVM help classify complex, nonlinear data (like the DaTSCAN brain scan results) by cleverly transforming it into a higher-dimensional space where it becomes easier to separate. Imagine trying to draw a straight line to divide two mixed-up groups of dots on paper—it's impossible. But if you could lift some dots "up" into 3D space, as shown in Figure 4, you could slide a flat sheet between them. Kernels do this mathematically without actually moving the data, using tricks like the RBF (Gaussian) or polynomial functions to measure similarities between points. For the DaTSCAN dataset, this means SVM can detect subtle, curved patterns in brain region values that linear methods would miss, helping distinguish Parkinson's patients from healthy controls more accurately.

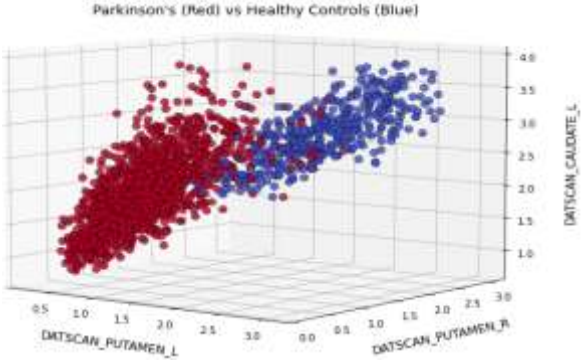


Figure 4. *3D Visualization of DaTSCAN Dataset*

3.2. Proposed Method

Support Vector Machines (SVM) with kernel functions are particularly effective for classifying nonlinear DaTSCAN datasets that distinguish between Parkinson's patients and healthy individuals. By applying the Radial Basis Function (RBF) kernel, the SVM transforms the original 3D DaTSCAN data—typically comprising striatal binding ratios or asymmetry indices—into a higher-dimensional space where the complex, nonlinear patterns of dopamine transporter distribution become linearly separable. The RBF kernel measures the similarity between data points using a Gaussian function, allowing the model to construct a flexible decision boundary that adapts to the intricate spatial relationships in the data. For instance, it captures subtle differences in putamen or caudate nucleus uptake that are characteristic of Parkinson's disease, even when these patterns overlap in the original feature space. The kernel's hyperparameters, such as gamma (γ) and the regularization term (C), are optimized through cross-validation to balance model complexity and generalization, ensuring accurate classification while avoiding overfitting to noise in the dataset.

Algorithm: SVM Kernel Comparison

Input:

- Dataset $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$ where $\mathbf{x}_i \in \mathbb{R}^d$, $y_i \in \{0,1\}$
- Test size ratio $\alpha \in (0,1)$

Procedure:

1. **Feature Engineering:**
   For each subject $i \in \{1, \ldots, n\}$:
   $\text{Asymmetry}_i \leftarrow [x_i^L - x_i^R \ \forall x \in \{\text{Putamen,Caudate,AntPutamen}\}]$
   $\tilde{\mathbf{x}}_i \leftarrow [\mathbf{x}_i \parallel \text{Asymmetry}_i]$

2. **Data Preprocessing:**
   $(\mathbf{X}_{train}, \mathbf{X}_{test}, \mathbf{y}_{train}, \mathbf{y}_{test}) \leftarrow \text{Split}(\mathcal{D}, \alpha)$
   $\mathbf{X}_{train}^{std} \leftarrow \text{StandardScaler}(\mathbf{X}_{train})$
   $\mathbf{X}_{test}^{std} \leftarrow \text{ApplyScaler}(\mathbf{X}_{test})$

3. **Kernel Analysis:**
   For each kernel $k \in \{\text{RBF,Poly,Sigmoid}\}$:
   $\theta^* \leftarrow \arg\max_{\theta \in \Theta_k} \text{CrossValScore}(\text{SVM}(k, \theta), \mathbf{X}_{train}^{std}, \mathbf{y}_{train})$
   Train $f_k \leftarrow \text{SVM}(k, \theta^*)$
   $\mathbf{y}_{pred} \leftarrow f_k(\mathbf{X}_{test}^{std})$
   $\text{Accuracy} \leftarrow \frac{1}{n_{test}} \sum_{i=1}^{n_{test}} \mathbb{I}(y_i = \hat{y}_i)$
   $\text{AUC} \leftarrow \int_0^1 \text{ROC}(f_k) df$

Output: $\{f_k, \text{Accuracy}_k, \text{AUC}_k\}_{k \in K}$

The algorithm is designed to classify DaTSCAN brain imaging data into Parkinson's disease (PD) or healthy control (HC) categories using an ensemble of machine learning models. It processes striatal binding ratio (SBR) values from key brain regions (putamen, caudate, and anterior putamen) through a systematic pipeline of data preparation, feature engineering, model training, and evaluation.

The process begins with the dataset containing brain scan measurements (x_i) and their corresponding labels (y_i = 0 for HC, 1 for PD). The algorithm uses stratified sampling to create multiple subsets of the data, ensuring each subset maintains the original ratio of PD to HC cases. This prevents bias in model training and provides representative samples for each iteration.

For each brain region, the algorithm calculates the difference between left and right hemisphere measurements. These asymmetry features are clinically significant since PD often manifests with uneven dopamine loss between hemispheres. The original measurements and these new asymmetry features are combined into enhanced input vectors. All features are then standardized (mean-centered and scaled) to ensure equal contribution during model training.

The algorithm uses the prepared brain scan data to teach the SVM (Support Vector Machine) model how to distinguish between healthy patients and those with Parkinson's. For each of the three kernels (RBF, Polynomial, and Sigmoid), it first performs a grid search—testing multiple combinations of settings (like sensitivity and complexity) through cross-validation to find the optimal configuration. The selected kernel then learns patterns from the training data by solving an optimization problem: it adjusts a mathematical boundary to maximize the separation between the two groups while minimizing errors. The model's performance is fine-tuned using parameters like the penalty C (for misclassification tolerance) and kernel-specific values (e.g., γ for RBF). Once trained, the model is ready to classify new brain scans based on the learned patterns.

The 3D visualization shows how the SVM with an RBF kernel separates brain scan data from healthy individuals and Parkinson's patients using three key components (likely derived from DaTSCAN imaging). The wavy, non-linear pink surface represents the decision boundary - a complex hyperplane that optimally divides the two classes in three-dimensional space in Figure 5. Points above this surface would be classified as Parkinson's cases, while those below as healthy. The RBF kernel's strength is evident in how the boundary curves and folds to accommodate clustered data points rather than forcing a flat plane.

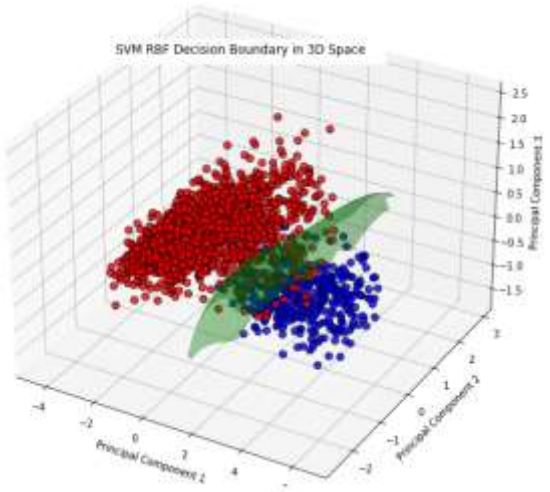


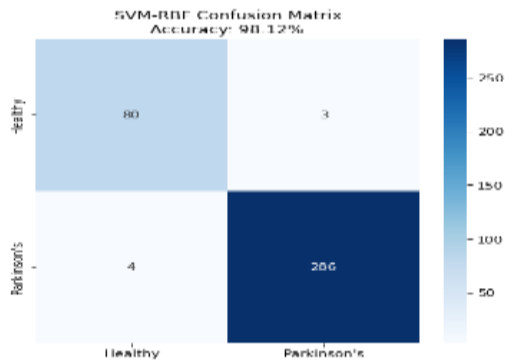
Figure 5. *SVM (RBF) Decision Boundary in 3D Space*

This flexibility allows it to capture subtle patterns in the asymmetrical dopamine uptake that distinguish Parkinson's patients. The varying elevation of the surface reflects how the model weights different combinations of these components - higher regions indicate stronger Parkinson's predictions based on specific biomarker interactions.The visualization demonstrates SVM's ability to handle non-linear relationships in neurological data, where simple straight boundaries would fail. The complex terrain shows areas of high classification confidence (peaks/valleys) versus uncertain zones (gentler slopes), which could correspond to borderline cases or early-stage Parkinson's detection. This matches clinical needs for detecting nuanced neurodegenerative patterns.

3.2.1. Radial Basis Function (RBF) Kernel

The RBF kernel, also known as the Gaussian kernel, is the most commonly used kernel for SVM classification tasks, particularly with medical imaging data like DaTSCAN results. Mathematically, it transforms the input space into an infinite-dimensional feature space using the formula

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2\right) \quad (1)$$



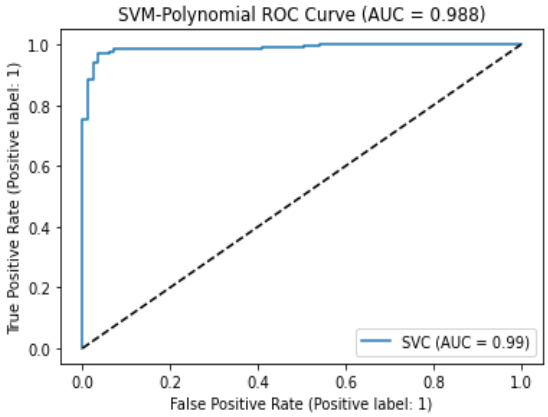
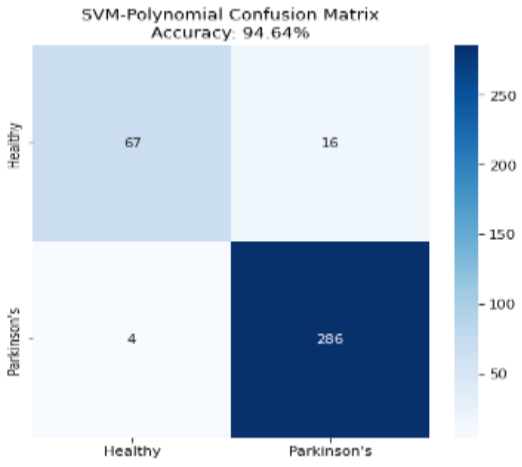


Figure 6. *Confusion matrix and ROC curve of SVM (RBF Kernel)*

where $\gamma>0$ is the kernel parameter that determines how far the influence of a single training example reaches. $\|xi-xj\|^2$ is the squared Euclidean distance. In the results, the RBF kernel achieved outstanding performance with 98.12% accuracy and a near-perfect ROC AUC of 0.994, as shown in Figure 6. The optimal parameters found (C=1, gamma=0.1) indicate that the model benefits from a moderate regularization strength and a reasonably wide decision boundary. This exceptional performance suggests that the patterns differentiating caudate and putamen regions in our DaTSCAN dataset are best captured by smooth, non-linear decision boundaries that don't require extreme complexity. The balanced precision and recall (both around 98.1%) further confirm that the RBF kernel handles both positive and negative cases equally well without significant bias.

3.2.2. Polynomial Kernel

The polynomial kernel introduces non-linearity by computing the polynomial expansion of features up



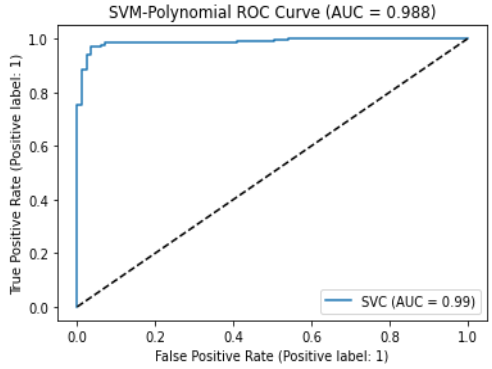


Figure 7. *Confusion matrix and ROC curve of SVM (Polynomial Kernel)*

to a specified degree. Its mathematical form is

$$K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma \mathbf{x}_i^{\intercal} \mathbf{x}_j + r)^d \quad (2)$$

where d is the polynomial degree, γ scales the dot product, and r is a coefficient term. In the results, the polynomial kernel achieved respectable performance with 94.64% accuracy using a 3rd-degree polynomial (degree=3) and C=10 for regularization, as shown in Figure 7. The ROC AUC of 0.988 remains excellent, although it is slightly lower than that of the RBF kernel. The results suggest that while polynomial relationships exist in the DaTSCAN data (likely representing multiplicative interactions between image features), they don't capture the underlying patterns as completely as the RBF transformation. The choice of gamma='scale' (which automatically uses 1/(n_features * X.var())) indicates the data benefits from feature-wise scaling. The marginally lower F1-score (0.945) compared to accuracy shows the polynomial kernel has a slightly less balanced performance than RBF.

3.2.3. Sigmoid Kernel

The sigmoid kernel, mathematically expressed as $K(x_i, x_j) = \tanh(\gamma x_i^{T} x_j + r)$, produces a neural network-like transformation of the data.

$$K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\gamma \mathbf{x}_i^{\intercal} \mathbf{x}_j + r) \quad (3)$$

Where $\tanh(\cdot)$ is the hyperbolic tangent function. γ controls the slope of the sigmoid, and r is the intercept term. In the results, it showed the weakest performance among the three kernels with 91.69% accuracy and ROC AUC of 0.959, as shown in Figure 8. The

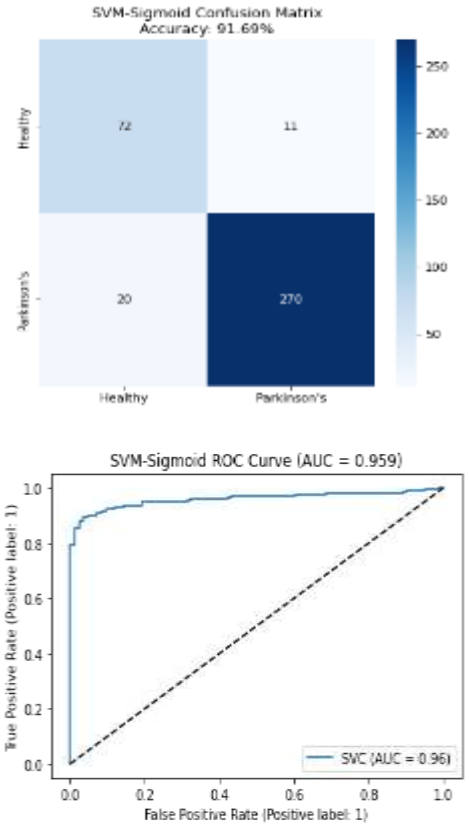


Figure 8. *Confusion matrix and ROC curve of SVM (Sigmoid Kernel)*

optimal parameters (C=0.1, gamma=0.1, coef0=0) suggest the model required stronger regularization (higher C might lead to overfitting) and didn't benefit from an intercept term (coef0=0). The performance gap between the sigmoid and the other kernels indicates that the hyperbolic tangent transformation doesn't align as well with the underlying data structure of DaTSCAN images. Interestingly, the precision (92.12%) is slightly higher than the recall (91.69%), suggesting the sigmoid kernel is somewhat more conservative in making positive predictions. This kernel might be more suitable for data that naturally follows a neural network activation pattern, which doesn't appear to be the case for our caudate/putamen analysis.

The study compared three SVM kernels in Table 1 f or classifying Parkinson's disease using DaTSCAN imaging data and shown in Figure 9, with the RBF kernel achieving superior performance (98.12% acc uracy, 0.993 ROC AUC, and balanced precision/rec all of 0.981) using optimal parameter C=1 and gam ma=0.1. The polynomial kernel followed with 94.6 3% accuracy (degree=3, C=10), while the sigmoid kernel trailed at 91.68% accuracy (C=0.1, gamma= 0.1). The results demonstrate that non-linear RBF tr ansformations best capture the

SVM Kernel Performance Comparison

Table 1. *SVM Kernel Performance Comparison*

Kernel	Accuracy	Parameters	ROC AUC	Precision	Recall	F1
RBF	98.12%	{'C': 1, 'gamma':0.1}	0.993	0.981	0.981	0.981
Polynomial	94.63%	{'C': 10, 'degree': 3, 'gamma': 'scale'}	0.988	0.946	0.946	0.944
Sigmoid	91.68%	{'C': 0.1, 'coef0': 0, 'gamma': 0.1}	0.959	0.921	0.916	0.918

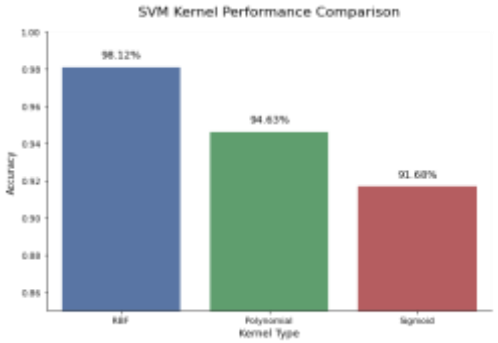


Figure 9. *SVM Kernel Performance Comparison*

complex patterns in caudate-putamen degeneration when compared to the polynomial kernels and sigmoid kernels.

Comparison of the Proposed Model Performance

Several studies have analyzed DaTSCAN striatal binding ratio (SBR) values in the putamen and caudate regions to differentiate between Parkinson's disease (PD) patients and healthy controls (HC), utilizing data from the Parkinson's Progression Markers Initiative (PPMI) database. Table 2 presents a comparative analysis of these prior findings alongside the results of our proposed model.

Table 2. *Comparison Proposed model Performance*

Author & Year	Feature used	Dataset Source & Size	Learning Model	Results
WuWang et al. (2020) [22]	Caudate [Left and Right] and Putamen [Left and Right] of DaTSCAN	584 Sample from PPMI (183 HC and 401 PD)	Deep learning model	96.45% Accuracy
Madhusudhana G K et al. (2021) [23]	Caudate [Left and Right] and Putamen [Left and Right] of DaTSCAN	919 Sample from PPM (290 HC and 629 PD)	Logistic Regression and KNN	94.82% LR and 96.8% KNN
Nandan N et al.(2024) [24]	Caudate [Left and Right] and Putamen [Left and Right] of DaTSCAN	1,862 samples from PPMI (413 HC and 1499 PD)	Random Forest	97% Accuracy
Proposed Method	**Caudate [Left and Right] and Putamen [Left and Right] of DaTSCAN**	**1,862 samples from PPMI (413 HC and 1499 PD)**	**SVM (RBF)**	**98.12% Accuracy**

Additionally, we evaluate each approach using the dataset from the suggested model that computes the difference between measurements from the left and right hemispheres. Since Parkinson's disease (PD) frequently presents with unequal dopamine loss between hemispheres, these asymmetry traits are clinically significant. Improved input vectors are created by combining these new asymmetry features with the original measurements. To guarantee equitable contribution during model training, all features are subsequently standardized (mean-centered and scaled). Table 3 shows the outcomes of the algorithms (CNN, Random Forest, KNN, Logistic Regression, and SVM(RBF)).

Table 3. *Comparison of the performance of different models for the same data*

Models	Accuracy	Precision	Recall	F1	ROC AUC	Best Params
SVM (RBF)	0.9812	0.9813	0.9812	0.9813	0.9938	{'C': 1, 'gamma': 0.1}
Random Forest	0.9759	0.9758	0.9759	0.9758	0.9948	{'max_depth': 5, 'n_estimators': 50}
Logistic Regression	0.9651	0.9649	0.9651	0.9648	0.9928	{'C': 1, 'penalty': 'l2'}
KNN	0.9705	0.9704	0.9705	0.9704	0.9868	{'n_neighbors': 5, 'weights': 'uniform'}
CNN	0.9732	0.9730	0.9732	0.9731	0.9570	{'dropout_rate': 0.3, 'learning_rate': 0.0001}

A comparative analysis of accuracy across five classification models is shown in Figure 10, which presents SVM (RBF), Random Forest, Logistic Regression, KNN, and CNN. The SVM (RBF) model, serving as the performance baseline indicated by the red dashed line, achieved the highest accuracy at 98.12%. all evaluated models demonstrated high predictive accuracy, and the SVM (RBF) has outperformed.

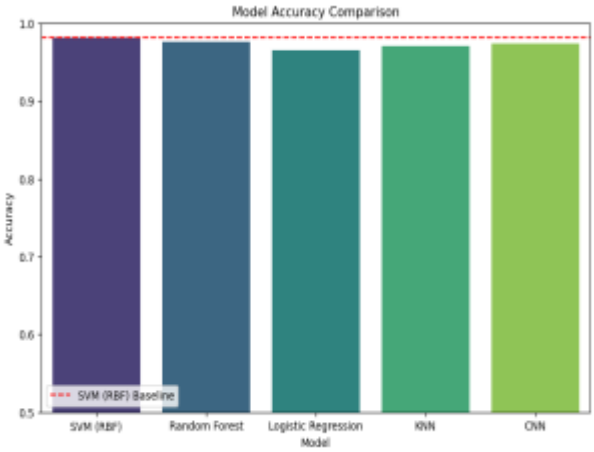


Figure 10. *Models Performance comparison*

Conclusion

This study focused on classifying Parkinson's disease (PD) patients and healthy controls (HC) using Striatal Binding Ratio (SBR) values from DaTSCAN data. Initial analysis, including PCA visualization, revealed a complex, nonlinear distribution of the data, necessitating models capable of handling such patterns. Consequently, the core methodology centered on evaluating Support Vector Machines (SVM) with different kernel functions (RBF, Polynomial, Sigmoid) known for their ability to manage nonlinearity. The approach was further enhanced by engineering asymmetry features, reflecting clinical observations of uneven dopamine loss in PD, and standardizing all features before comparing the optimized SVM model against other common machine learning algorithms.

Based on the comparative analysis of Support Vector Machine (SVM) kernel performance, the Radial Basis Function (RBF) kernel demonstrates significantly superior results across all evaluated metrics. Achieving an accuracy of 98.12%, the RBF kernel, optimized with parameters C=1 and gamma=0.1, stands out as the most effective option. The Polynomial and Sigmoid kernels, while functional, yielded lower performance compared to the RBF kernel. The Polynomial kernel reached an accuracy of 94.63% with parameters C=10, degree=3. The Sigmoid kernel ranked last, achieving 91.68% accuracy using parameters C=0.1 and gamma=0.1. This suggests that the RBF kernel's method of transforming data into a higher-dimensional space effectively captures the intricate, non-linear patterns characteristic of dopaminergic changes in PD as measured by DaTSCAN SBR values. When compared against Random Forest, Logistic Regression, KNN, and CNN using the same comprehensive feature set, the SVM-RBF model consistently demonstrated superior accuracy. In conclusion, the study demonstrates that an optimized SVM with an RBF kernel provides a highly accurate and robust approach for differentiating PD patients from healthy controls based on DaTSCAN imaging data, offering potential value for clinical diagnostic support systems.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1] Jie Mei, Christian Desrosiers, and Johannes Frasnelli (2021). Machine Learning for the Diagnosis of Parkinson's Disease: A Review of Literature *Frontiers in Aging Neuroscience*. 13. https://doi.org/10.3389/fnagi.2021.633752.

[2] De Deurwaerdere P, Gaetani S, Vaughan RA. (2020). Old neurochemical markers, new functional directions? *J Neurochem.* Mar;152(6):623-626. doi: 10.1111/jnc.14929. *Epub* 2020 PMID: 31917872.

[3] Andrea Brovelli, Bruno Nazarian, Martine Meunier, Driss Boussaoud, (2011). Differential roles of caudate nucleus and putamen during instrumental learning, *NeuroImage*, Volume 57(4), 1580-1590, ISSN 1053-8119, https://doi.org/10.1016/j.neuroimage.2011.05.059.

[4] Hörtnagl H, Pifl C, Hörtnagl E, Reiner A, Sperk G. (2020). Distinct gradients of various neurotransmitter markers in the caudate nucleus and

putamen of the human brain. *J Neurochem*. 152:650–62. doi: 10.1111/jnc.14897

[5] Abedelahi A, Hasanzadeh H, Hadizadeh H, Joghataie MT, (2013). Morphometric and volumetric study of caudate and putamen nuclei in normal individuals by MRI: Effect of normal aging, gender, and hemispheric differences. *Pol J Radiol*. 78(3):7-14. doi: 10.12659/PJR.889364. PMID: 24115954; PMCID: PMC3789937.

[6] Dauer W, Przedborski S. (2003). Parkinson's disease: mechanisms and models. *Neuron*. 11;39(6):889-909. doi: 10.1016/s0896-6273(03)00568-3. PMID: 12971891

[7] Nandan, N., Sanjay Pande, M.S., Raveesh, B.N. et al, (2024). Sensitive Two-Dimensional Photonic Crystal Biosensor for The Detection of Parkinson's Disease. *J Opt*. https://doi.org/10.1007/s12596-024-02306-x

[8] Akam T, Walton ME. (2021). What is dopamine doing in model-based reinforcement learning? *Curr Opin Behav Sci*. 38:74-82. doi: 10.1016/j.cobeha.2020.10.010. PMID: 37082448; PMCID: PMC7614453.

[9] Seifert KD, Wiener JI. (2013). The impact of DaTscan on the diagnosis and management of movement disorders: A retrospective study. *Am J Neurodegener Dis*. 2(1):29-34. Epub PMID: 23515233; PMCID: PMC3601468.

[10] Arash Yaghoobi, Homa Seyedmirzaei, Marzie Jamaat, Moein Ala, (2024). The role of AI and machine learning in the diagnosis of Parkinson's disease and atypical parkinsonisms. *Parkinsonism and Related Disorders*, Volume 126, 106986, ISSN

[16] Khachnaoui H, Khlifa N, Mabrouk R. (2022). Machine Learning for Early Parkinson's Disease Identification within SWEDD Group Using Clinical and DaTSCAN SPECT Imaging Features. *J Imaging*. 8(4):97. doi: 10.3390/jimaging8040097. PMID: 35448224; PMCID: PMC9032319.

[17] Pavan Rajkumar Magesh, Richard Delwin Myloth, Rijo Jackson Tom, (2020). An Explainable Machine Learning Model for Early Detection of Parkinson's Disease using LIME on DaTSCAN Imagery, *Computers in Biology and Medicine*, Volume 126, 104041, ISSN 0010-4825, https://doi.org/10.1016/j.compbiomed.2020.104041

[18] Martínez-Murcia, F.J., Górriz, J.M., Ramírez, J., Illán, I.A., Puntonet, C.G. (2013). Texture Features Based Detection of Parkinson's Disease on DaTSCAN Images. In: Ferrández Vicente, J.M., Álvarez Sánchez, J.R., de la Paz López, F., Toledo Moreo, F.J. (eds) Natural and Artificial Computation in Engineering and Medical Applications. IWINAC 2013. Lecture Notes in Computer Science, vol 7931. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-38622-0_28

[19] Kavitha Paranjothi, Fathima Ghouse, Revathi Vaithiyanathan, (2024). Detection of Parkinson's Disease on DaTSCAN Image Using Multi-kernel Support Vector Machine, *International Journal of Intelligent Engineering and Systems*, Vol.17(5), DOI: 10.22266/ijies2024.1031.05

1353-8020, https://doi.org/10.1016/j.parkreldis.2024.106986 .

[11] Bega D, Kuo PH, Chalkidou A, Grzeda MT, Macmillan T, Brand C, Sheikh ZH, Antonini A(2021). Clinical utility of DaTscan in patients with suspected Parkinsonian syndrome: a systematic review and meta-analysis. *NPJ Parkinsons Dis*. Vol. 7(1):43. doi: 10.1038/s41531-021-00185-8. PMID: 34031400; PMCID: PMC8144619.

[12] Sourabarna Roy, Tannistha Pal, Swapan Debbarma, (2024). A Comparative Analysis of Advanced Machine Learning Algorithms to diagnose Parkinson's Disease, *Procedia Computer Science*, Volume 235, 122-131, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2024.04.015 .

[13] Chien CY, Hsu SW, Lee TL, Sung PS, Lin CC, (2020). Using Artificial Neural Network to Discriminate Parkinson's Disease from Other Parkinsonisms by Focusing on Putamen of Dopamine Transporter SPECT Images. *Biomedicines*. 9(1):12. doi: 10.3390/biomedicines9010012. PMID: 33374377; PMCID: PMC7823797.

[14] Justin Quan, Lin Xu, Rene Xu, Tyrael Tong, and Jean Su, (2019). DaTscan SPECT Image Classification for Parkinson's Disease, *ArXiv*, abs/1909.04142

[15] Kurmi A, Biswas S, Sen S, Sinitca A, Kaplun D, Sarkar R. (2022). An Ensemble of CNN Models for Parkinson's Disease Detection Using DaTscan Images. *Diagnostics (Basel)*. 12(5):1173. doi: 10.3390/diagnostics12051173. PMID: 35626328; PMCID: PMC9139649.

[20] Babita Majhi and Bhanu Prasad, (2022). Deep learning architectures for Parkinson's disease detection using multi-modal features. Comput. Biol. Med. 146. https://doi.org/10.1016/j.compbiomed.2022.105610

[21] Makarious, M.B., Leonard, H.L., Vitale, D. et al., (2022). Multi-modality machine learning predicting Parkinson's disease. *npj Parkinsons Dis*. 8, 35. https://doi.org/10.1038/s41531-022-00288-w

[22] Wuwang, Junho Lee, Fouzi Harrou and Ying Sun (2020). Early Detection of Parkinson's Disease Using Deep Learning and Machine Learning, *IEEE Access Digital Object Identifier* 10.1109/ACCESS.2020.3016062.

[23] Madhusudhana G K, Sanjaypande M B, Raveesh B N, (2021). Prediction of Parkinson's disease using the Parametric and Non-Parametric Machine Learning Models, *IJIRT*. Volume 8(7). ISSN: 2349-6002

[24] Nandan N, Sanjay Pande M B, Raveesh B N, Rakesh, (2024). A Machine Learning Approach to Analyzing DATSCAN SBR Values for the Detection of Parkinson's Disease, *J. Electrical Systems* 20-11s. 1119-1131, https://doi.org/10.52783/jes.7383

[25] Klokkaris, Anthony, and Anna Migdalska-Richards. (2024). An Overview of Epigenetic Changes in the Parkinson's Disease Brain *International Journal of Molecular Sciences* 25(11). 6168. https://doi.org/10.3390/ijms25116168