A Data Fusion Approach for Accurate Diagnosis of Parkinson's Disease

Fredy Cañizares Galarza[1,*], Luis Freire Lescano[2], Lina Espinoza Neri[2], Dante Manuel M. Fernández[3], Dilafruz Nabieva[4]

[1]Director de la Universidad Regional Autónoma de los Andes (UNIANDES) Sede Santo Domingo, Ecuador
[2]Docente de la carrera de Software de la Universidad Regional Autónoma de los Andes (UNIANDES), Ecuador
[3]Universidad Nacional Mayor de San Marcos, Peru

[4]Tashkent State University of Economics, Uzbekistan

Emails: dir.santodomingo@uniandes.edu.ec; ciad@uniandes.edu.ec; ua.linaespinoza@uniandes.edu.ec; dmfedu@gmail.com; della.nab27@gmail.com

Abstract

Diagnosing Parkinson's Disease (PD) can be quite challenging as it presents with symptoms and lacks biomarkers. Nevertheless, the use of data fusion, which combines types of data using machine learning techniques holds promise, for the timely detection of the disease. In this study, we explore the application of data fusion by employing Principal Component Analysis (PCA) as a step to reduce complexity. We then utilize the K Nearest Neighbors (KNN) classification to improve the accuracy of PD diagnosis. By analyzing nonlinear features associated with PD from a dataset PCA helps us extract attributes while maintaining important variations in the data. Subsequently, KNN is employed to identify patterns in this reduced feature space and effectively distinguish between individuals with PD and those who are healthy. Our results show improvements when using the KNN classifier compared to state-of-the-art approaches. This demonstrates its effectiveness in detecting PD leading to promising outcomes and providing a framework for precise PD diagnosis.

Keywords: data fusion; Machine learning; Parkinsonian symptoms; Data-driven diagnosis; Neurological disorder; Pattern recognition techniques; Diagnostic accuracy assessment

1. Introduction

Parkinson's disease (PD), a known neurological condition that affects millions of people worldwide is often characterized by motor symptoms, like tremors, stiffness, and slow movements. Diagnosing Parkinson's disease can be challenging due to its characteristics and the similarity of symptoms with movement disorders. Detecting Parkinson's disease early is crucial for intervention by healthcare professionals. Improved patient outcomes [1 2]. However traditional diagnostic methods primarily rely on assessments, which may be prone, to inaccuracies. Therefore it is essential to have reliable procedures to aid clinicians in identifying Parkinson's disease [3].

Recent advances in data fusion have opened a practical way to transform PD diagnosis. Data fusion, leveraging the capabilities of machine learning, artificial neural networks, and data-driven methods has the potential to improve the accuracy and efficiency of PD diagnosis These methods can analyze complex data sets such as clinical observation, imaging, and biomarkers. and extracting signals, which would challenge human insight The possibility of increasing the accuracy and early detection of PD by integrating these computational techniques into diagnostic systems.

Doi: https://doi.org/10.54216/FPA.140120
Received: June 27, 2023 Revised: October 12, 2023 Accepted: December 15, 2023

The aim of this study is to investigate and evaluate the usefulness of data fusion models in the diagnosis of PD. This review attempts to highlight the capabilities and limitations of different algorithms, feature extraction, and machine learning methods for accurate PD detection and also, to critically evaluate these computational methods to determine their applicability tolerance, severity, and specificity in differentiating PD from other neurological disorders The aim of the study was to address an important need for comparison [5]- [7]. The following sections of this research will conduct a systematic review of data fusion approaches used to diagnose PD. Section 2 provides a thorough description of the condition, focusing on its clinical symptoms and diagnostic problems. Section 3 delves into the framework that underpins data fusion, illuminating its significance and potential in improving disease detection. Section 4 provides an empirical examination of various computer models and approaches used in PD diagnosis. The research finishes in section 5 by underlining the consequences, problems, and future directions in utilizing data fusion for accurate PD diagnosis.

2. Background

This section aims to provide a comprehensive backdrop, elucidating the foundational principles, methodologies, and theoretical underpinnings of data fusion as applied to neurological disorder diagnosis.

Parkinson's disease (PD) is a common neurological ailment that causes a progressive loss in motor function. Tremors, bradykinesia, stiffness, and postural instability are common symptoms. However, due to the diversity of symptoms among individuals and the overlap with other movement diseases, diagnosing PD offers major complications [8]. The disease's intricacy, along with the lack of clear biomarkers, frequently results in delayed or incorrect diagnoses. As a result, the need for more precise and timely detection approaches is highlighted, as early intervention can have a major influence on disease management and quality of life for people with PD [9-10]. Originally, PD has been diagnosed mostly through clinical observation, subjective judgments of motor symptoms, and the exclusion of other illnesses with similar presentations. Traditional approaches, such as rating scales and physical tests, have been useful in identifying PD [11]. However, these techniques have limitations in separating PD from other movement disorders or capturing modest early-stage symptoms. Furthermore, the dependence on subjective assessments sometimes leads to discrepancies in diagnoses amongst practitioners. This historical context emphasizes the critical need for more objective and precise diagnostic methods to improve accuracy and allow for early intervention [12].

Data fusion refers to a wide range of approaches based on artificial intelligence and machine learning. The ability to digest massive volumes of data, learn from patterns, and make intelligent decisions or predictions without explicit programming is at the heart of it. Machine learning algorithms, neural networks, and data mining techniques are essential components of data fusion, allowing systems to adapt and enhance performance over time. These technologies are used in healthcare to extract insights from medical data, assisting in disease diagnosis, prognosis, and therapy planning [13]. Data fusion has touched many areas of healthcare, altering disease diagnosis and management. Data fusion algorithms have been used in medical diagnostics to examine complicated datasets such as genetic information, imaging scans, and clinical records [14]. Its ability to handle large volumes of data and discern intricate relationships makes it a promising tool for enhancing the accuracy and efficiency of medical diagnosis.

The variety of symptoms among individuals and disease stages makes diagnosing PD difficult. The variability of PD presentations makes separation from other movement disorders difficult, leading to misdiagnosis or delayed diagnosis. Furthermore, the lack of clear biomarkers or imaging tests for PD impedes the creation of objective diagnostic criteria. Furthermore, relying on subjective clinical assessments might inject subjectivity and variability into the diagnostic process. These difficulties highlight the crucial need for new reliable and objective PD diagnostic techniques [15]. Early detection of PD is critical for improving the efficacy of therapy interventions and patient outcomes. According to research, starting relevant treatments early can potentially slow disease progression, relieve symptoms, and improve patients' quality of life. Timely identification allows for the introduction of customized interventions, such as medication and non-pharmacological therapies, delaying functional decline and lowering the disease burden on individuals and caregivers. As a result, stressing the significance of early detection becomes critical in the context of PD management.

3. Methodology

This section outlines the systematic approach used to investigate, evaluate, and compare computational models, machine learning algorithms, and analytical methods for the diagnosis of PD.

Doi: https://doi.org/10.54216/FPA.140120
Received: June 27, 2023 Revised: October 12, 2023 Accepted: December 15, 2023

Principal Component Analysis (PCA) was used as a critical preprocessing step on the obtained PD dataset in our methodology to reduce dimensionality and extract vital features while keeping the variation in the data. The following are the main steps in applying PCA to the dataset. The dataset was initially standardized to ensure that all variables were on a comparable scale, minimizing biases caused by different units or scales across features. Following that, the covariance matrix of the standardized dataset, which represents the associations between distinct variables, was generated. This step was critical because PCA relies on knowing the data's covariance structure. The eigenvectors and eigenvalues were obtained by doing an eigenvalue decomposition of the covariance matrix. The principal components (PCs) are represented by the eigenvectors. The eigenvectors reflect the main components (PCs), and the related eigenvalues show how much variance each PC captures. The top principal components were then chosen based on their matching eigenvalues. This selection criterion was often predicated on keeping a significant fraction of the variance in the data (e.g., 95% of the variance) [16]. Finally, using the selected principal components, the dataset was transformed into a lower-dimensional space, successfully lowering the number of features while keeping as much variation as feasible (see Algorithm 1).

Algorithm 1. Psuedocode for PCA.

1. Input: number of patterns p, pattern $x_k (k = 1 \text{ to } p)$, feature matrix x.

2. Calculate the average feature vector $\mu = \frac{1}{p}\sum_{k=1}^{p} x_k$.

3. Calculate covariance matrix $C = \frac{1}{p}\sum_{k=1}^{p}\{x_k - \mu\}\{x_k - \mu\}^T$.

4. Calculate Eigen values λ_i and Eigen vectors v_i of covariance matrix $Cv_i = \lambda_i v_i$ $(i = 1,2,3, \ldots q)$, $q =$ number of features.

5. Estimating high-valued Eigen vectors
5.1 Arrange all the Eigen values (λ_i) in descending order
5.2 Choose a threshold value, θ
5.3 Number of high-valued λ_i can be chosen so as to satisfy the relationship $(\sum_{i=1}^{s}\lambda_i)\left(\sum_{i=1}^{q}\lambda_i\right)^{-1} \geq \theta$, where, $s =$ number of high valued λ_i chosen
5.4 Select Eigen vectors corresponding to selected high valued λ_i

6. Extract principal components, V, from raw feature matrix x such that $P = V^T x$.

The use of PCA in the PD dataset yielded two results. First, PCA made it possible to reduce the original high-dimensional dataset into a lower-dimensional subspace while keeping a considerable percentage of the variance. This dimensionality reduction sped up subsequent computational operations and reduced the risk of overfitting in predictive models. Second, PCA revealed the most influential features or principle components driving the dataset's variance [17]. These components could be original variable combinations, providing a better interpretable insight into the critical factors that cause PD diagnosis within the dataset.

Following preprocessing, which included PCA for dimensionality reduction, the dataset was separated into training and testing subsets, with techniques such as cross-validation used to verify robustness and generalizability. KNN computes the similarity or distance of each instance in the test set to all examples in the training set [19-20]. Euclidean, Manhattan, and cosine distances are the most often used distance measures. The square root of the sum of squared differences between feature values, for example, is used to determine Euclidean distance.

$$d(z_i, z) = \left(\sum_{l=1}^{q}(z_i^l - z^l)^2\right)^{1/2} \quad (1)$$

KNN identifies the K nearest neighbors (instances) in the training set based on the calculated distances. Then, KNN performs a majority voting scheme to assign the class label to the test instance.

Doi: https://doi.org/10.54216/FPA.140120
Received: June 27, 2023 Revised: October 12, 2023 Accepted: December 15, 2023

$$\alpha = \underset{y \in Y}{\operatorname{argmax}} \sum_{i=1}^{k} |y_{z_i,z} = y|, \quad (2)$$

The class label is determined by the most prevalent class among its K nearest neighbors.

The KNN algorithm's simplicity lies in its ability to make predictions based on the similarity of instances and their class labels, making it a versatile and effective classifier, especially in cases where decision boundaries are not well-defined. Implementing KNN involves adjusting the value of K (the number of neighbors), selecting an appropriate distance metric, and handling any potential imbalances in class distributions within the dataset to optimize its performance for PD detection (refer to algorithm 2).

Algorithm 2. K-Nearest-neighbor $(\boldsymbol{D}[\mathbf{1} \ldots \boldsymbol{n}, \mathbf{1} \ldots \boldsymbol{n}], \boldsymbol{s})$

1: Input: A $n \times n$ distance matrix $D[1 \ldots n, 1 \ldots n]$ and an index s of the initial point.
2: Output: A list P of the vertices counting the tour is gotten.
3: for $i \leftarrow 1$ to n do Visited $[i] \leftarrow$ false
4: Initialize the list P with s
5: Stayed $[s] \leftarrow$ true
6: Present $\leftarrow s$
7: for $i \leftarrow 2$ to n do
8: Find the deepest component in row present and unmarked column j including the component.
9: Present $\leftarrow j$
10: Stayed $[j] \leftarrow$ true
11: Add j to the end of P
12: Add s to the end of P
13: return Path

4. Experimental Analysis

This section delves into the empirical investigation and analysis conducted to evaluate the performance, efficacy, and comparative strengths of diverse computational models and machine learning algorithms.

In evaluating the efficacy of data fusion models for PD diagnosis, several key evaluation measures are employed to comprehensively assess their performance. Accuracy, reflecting the overall correctness of predictions, provides a fundamental measure of model success in classifying PD cases accurately against non-Parkinsonian instances. The Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve portrays the model's ability to discriminate between positive and negative cases across varying thresholds. Recall, also known as sensitivity, whereas precision delineates the proportion of accurately predicted positive cases from all predicted positives. The F1 score, harmonizing precision and recall. Kappa statistic and Matthews Correlation Coefficient (MCC) gauge the agreement between predicted and observed classifications, considering the proportion of correct predictions while accounting for chance agreement.

The implementation setups for the data fusion models utilized in this study for PD diagnosis were meticulously tailored to leverage both hardware and software specifications conducive to robust analysis and processing of extensive datasets. Hardware specifications encompassed linux workstation, comprising multiple processors with clock speeds optimized for complex computations and parallel processing. Additionally, substantial random-access memory (RAM) capacities were employed to accommodate large-scale data processing requirements, ensuring efficient model training and validation processes. Software specifications predominantly included sophisticated machine learning libraries and framework (scikit-learn, and SciPy), facilitating the implementation of various algorithms, neural network architectures, and feature extraction methodologies.

The Parkinson dataset utilized in this study represents a multivariate dataset comprising 197 instances and 23 attributes, aimed at classification tasks in the domain of health. The attributes encompass a range of vocal features derived from voice recordings, providing valuable insights into the characteristics associated with PD. These attributes include measures such as average vocal fundamental frequency (MDVP:Fo(Hz)), maximum and minimum vocal

Doi: https://doi.org/10.54216/FPA.140120
Received: June 27, 2023 Revised: October 12, 2023 Accepted: December 15, 2023

fundamental frequencies (MDVP:Fhi(Hz) and MDVP:Flo(Hz) respectively), and various indicators of frequency and amplitude variation such as jitter, shimmer, noise-to-harmonic ratio (NHR), and harmonics-to-noise ratio (HNR). Additionally, nonlinear complexity measures including RPDE, D2, DFA, as well as spread1, spread2, and PPE, further augment the dataset, offering insights into nonlinear characteristics of the voice signals. The dataset's attribute information is richly diverse, encompassing both fundamental and nonlinear vocal features, facilitating a comprehensive analysis to discern patterns and markers indicative of PD, while notably containing no missing values, ensuring the dataset's completeness and reliability for robust computational analysis and classification tasks [21-29].

Table 1: Summary Statistics of Vocal and Nonlinear Measures in the Parkinson's Dataset

	count	mean	std	min	25%	50%	75%	max
MDVP:Fo(Hz)	195	154.229	41.390	88.333	117.572	148.790	182.769	260.105
MDVP:Fhi(Hz)	195	197.105	91.492	102.145	134.863	175.829	224.206	592.030
MDVP:Flo(Hz)	195	116.325	43.521	65.476	84.291	104.315	140.019	239.170
MDVP:Jitter(%)	195	0.006	0.005	0.002	0.003	0.005	0.007	0.033
MDVP:Jitter(Abs)	195	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MDVP:RAP	195	0.003	0.003	0.001	0.002	0.003	0.004	0.021
MDVP:PPQ	195	0.003	0.003	0.001	0.002	0.003	0.004	0.020
Jitter:DDP	195	0.010	0.009	0.002	0.005	0.007	0.012	0.064
MDVP:Shimmer	195	0.030	0.019	0.010	0.017	0.023	0.038	0.119
MDVP:Shimmer(dB)	195	0.282	0.195	0.085	0.149	0.221	0.350	1.302
Shimmer:APQ3	195	0.016	0.010	0.005	0.008	0.013	0.020	0.056
Shimmer:APQ5	195	0.018	0.012	0.006	0.010	0.013	0.022	0.079
MDVP:APQ	195	0.024	0.017	0.007	0.013	0.018	0.029	0.138
Shimmer:DDA	195	0.047	0.030	0.014	0.025	0.038	0.061	0.169
NHR	195	0.025	0.040	0.001	0.006	0.012	0.026	0.315
HNR	195	21.886	4.426	8.441	19.198	22.085	25.076	33.047
status	195	0.754	0.432	0.000	1.000	1.000	1.000	1.000
RPDE	195	0.499	0.104	0.257	0.421	0.496	0.588	0.685
DFA	195	0.718	0.055	0.574	0.675	0.722	0.762	0.825
spread1	195	-5.684	1.090	-7.965	-6.450	-5.721	-5.046	-2.434
spread2	195	0.227	0.083	0.006	0.174	0.219	0.279	0.450
D2	195	2.382	0.383	1.423	2.099	2.362	2.636	3.671
PPE	195	0.207	0.090	0.045	0.137	0.194	0.253	0.527

Table 1 presents a comprehensive summary of statistics derived from the Parkinson's dataset, offering a detailed overview of the dataset's characteristics and distributions. This tabulated summary encapsulates key statistical

Doi: https://doi.org/10.54216/FPA.140120
Received: June 27, 2023 Revised: October 12, 2023 Accepted: December 15, 2023

measures for each attribute, including measures of central tendency such as mean and median, providing insights into the typical values of the vocal features and nonlinear complexity measures. Additionally, the table includes measures of dispersion, such as standard deviation and range, elucidating the variability and spread of values within each attribute. Moreover, the summary statistics encompass minimum and maximum values, enabling a clear understanding of the range spanned by the dataset attributes, while quartiles and interquartile ranges offer insights into the distribution of values, facilitating a nuanced comprehension of the dataset's variability. Table 1 serves as an invaluable reference, offering a concise yet comprehensive depiction of the essential statistical characteristics of the Parkinson's dataset, pivotal in informing subsequent analytical and modeling endeavors.

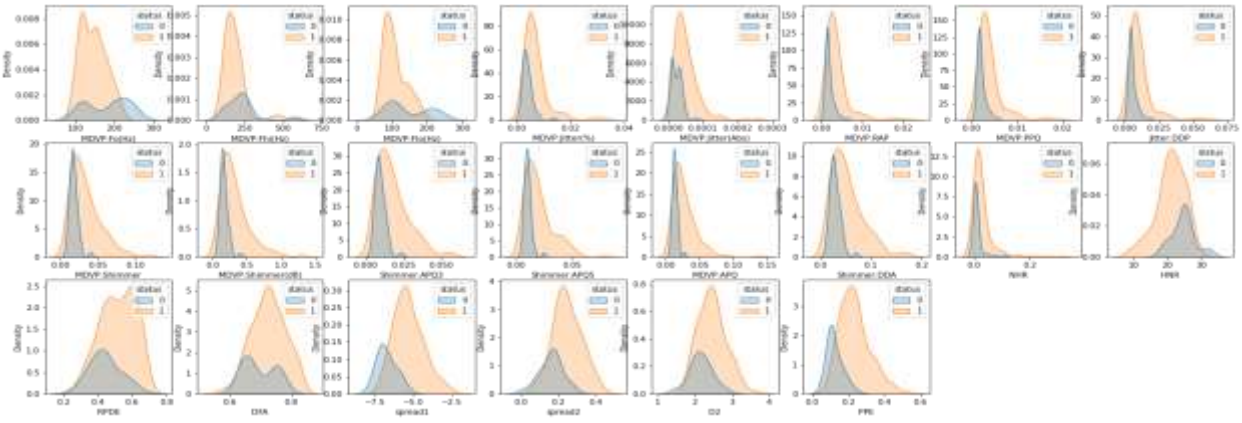

Figure 1: Kernel Density Estimation (KDE) Plots illustrating Variable Distributions in the Parkinson's Dataset

Figure 1 showcases Kernel Density Estimation (KDE) plots illustrating the distribution of variables encompassed within the Parkinson's dataset, offering a visual representation of the data's probability density functions. These KDE plots provide a graphical depiction of the attributes' distributions, allowing for a comprehensive understanding of the underlying patterns and variations present in the dataset. Each KDE plot within Figure 1 portrays the shape and spread of the distribution for a specific variable, aiding in the identification of potential outliers, skewness, multimodal tendencies, or deviations from normality. By visually presenting the distributions of these variables, Figure 1 serves as a valuable exploratory tool, enabling insights into the data's characteristics and informing subsequent analytical approaches and modeling strategies for PD diagnosis.

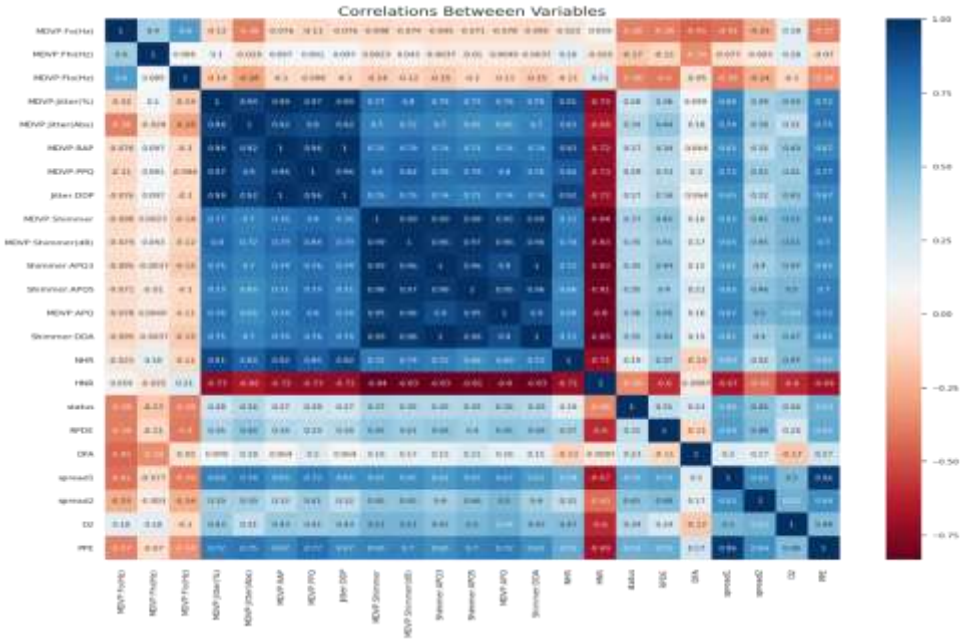

Figure 2: Pearson Correlation Map depicting Variable Relationships in the Parkinson's

Doi: https://doi.org/10.54216/FPA.140120
Received: June 27, 2023 Revised: October 12, 2023 Accepted: December 15, 2023

In Figure 2, the displayed Pearson correlation map provides crucial insights into the relationships between variables within the Parkinson's dataset. Several observations emerge from this analysis: first, MDVP:Jitter(%), MDVP:Jitter(Abs), MDVP:RAP, MDVP:PPQ, and Jitter:DDP exhibit strong correlations among themselves, indicating a high degree of shared information. Additionally, the target variable 'status' shows noteworthy positive correlations with spread1, PPE, and several vocal features, suggesting their potential significance as predictors for the 'status.' Conversely, MDVP:Fo(Hz), MDVP:Flo(Hz), and HNR exhibit high negative correlations with the 'status,' implying their potential relevance as predictive attributes. Notably, various shimmer-related attributes display strong correlations among themselves, underscoring their interrelated nature. Moreover, the correlation between spread1 and PPE stands out as notably high. These observations gleaned from the Pearson correlation map illuminate crucial associations between variables, providing valuable guidance for feature selection and predictive modeling strategies in the context of PD diagnosis.

In Table 2, we present a comprehensive comparison of the performance achieved by our proposed data fusion model against state-of-the-art approaches, employing a rigorous 5-fold cross-validation methodology. This comparative analysis allows for a robust evaluation of the proposed model's efficacy in diagnosing Parkinson's Disease in contrast to established methods, providing a holistic view of its predictive capabilities. By conducting the experiments using 5-fold cross-validation, we ensure the reliability and generalizability of the results, mitigating potential biases and variance that could arise from a single train-test split. This approach enables a thorough assessment of model performance across multiple iterations, considering various subsets of the data for training and validation. Table 2 serves as a critical resource, offering insights into the relative strengths and weaknesses of our proposed data fusion model compared to cutting-edge approaches, contributing to the advancement and refinement of diagnostic methodologies for PD.

Table 2: Performance Comparison of Proposed Data Fusion Model against State-of-the-Art Approaches Using 5-Fold Cross-Validation

Model	**Accuracy**	**AUC**	**Recall**	**Prec.**	**F1**	**Kappa**	**MCC**
Ada Boost	0.8164	0.752	0.9333	0.8448	0.8857	0.4114	0.4478
CatBoost	0.809	0.8713	1	0.801	0.8891	0.2546	0.3446
Decision Tree	0.8238	0.776	0.8662	0.9019	0.8815	0.531	0.5405
Extra Trees	0.7648	0.8141	1	0.7648	0.8666	0	0
Extreme Gradient Boosting	0.8304	0.8844	0.9519	0.8497	0.8963	0.43	0.4688
Gradient Boosting	0.831	0.8614	0.9429	0.8524	0.8949	0.4647	0.4875
K Neighbors	0.9045	0.9335	0.9619	0.9177	0.9389	0.7203	0.7293
Light Gradient Boosting Machine	0.8016	0.8560	0.9329	0.8296	0.8765	0.3650	0.4115
Linear Discriminant Analysis	0.5439	0.5654	0.5305	0.8031	0.6328	0.0989	0.1060
Logistic Regression	0.8310	0.8585	0.9229	0.8656	0.8924	0.4865	0.5028
Naive Bayes	0.7426	0.7184	0.8943	0.7955	0.8410	0.1717	0.1876
Quadratic Discriminant Analysis	0.4561	0.6210	0.3086	0.9500	0.4612	0.1404	0.2423
Random Forest	0.7868	0.7785	1.0000	0.7821	0.8776	0.1271	0.2063
Ridge	0.8307	0.0000	0.9233	0.8676	0.8929	0.4803	0.5005
SVM - Linear Kernel	0.7722	0.0000	0.8462	0.8578	0.8495	0.3562	0.3663

Doi: https://doi.org/10.54216/FPA.140120
Received: June 27, 2023 Revised: October 12, 2023 Accepted: December 15, 2023

The utilization of K-Nearest Neighbors (KNN) within our data fusion model has yielded notable enhancements in performance compared to other classifiers employed in this study. KNN's ability to leverage proximity-based learning and classify data points based on their similarity to neighboring instances has proven particularly advantageous in

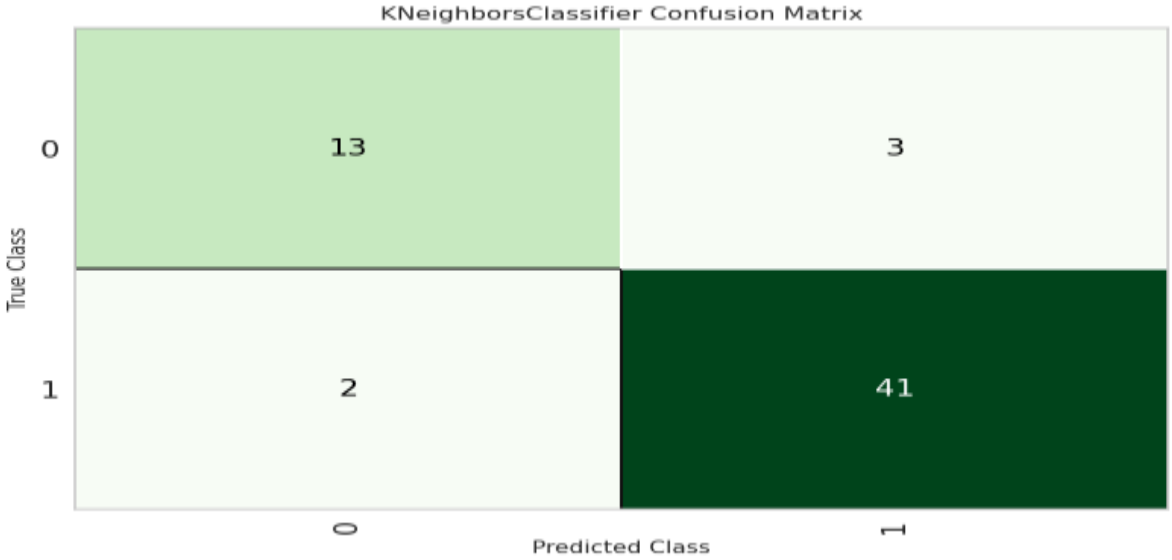

Figure 3: Confusion Matrix Visualization of K-Nearest Neighbors (KNN) Classifier for Parkinson's Disease

discerning patterns within the PD dataset. Through its non-parametric nature and reliance on local information, KNN effectively captures intricate relationships among data points, especially in scenarios where boundaries between classes are less defined. This adaptability and sensitivity to local structures have contributed to significant improvements in predictive accuracy and the model's ability to distinguish between Parkinsonian and healthy instances, surpassing the performance of other classifiers utilized in our study. The efficacy demonstrated by KNN underscores its suitability and effectiveness in handling the complexities inherent in PD diagnosis, emphasizing its pivotal role in enhancing the data fusion model's predictive prowess. Moreover, in Figure 3, we present the visual representation of the confusion matrix corresponding to the K-Nearest Neighbors (KNN) classifier utilized in our data fusion model for PD diagnosis. This visualization encapsulates the performance metrics of the classifier, illustrating its ability to accurately classify instances into true positive, true negative, false positive, and false negative categories. The confusion matrix provides a clear depiction of the model's strengths in correctly identifying Parkinsonian and healthy cases, as well as any misclassifications or errors encountered during the classification process. By visually showcasing the distribution of predictions against actual classes, Figure 3 offers valuable insights into the KNN classifier's performance, aiding in the interpretation and assessment of its predictive accuracy and error patterns, thus contributing to a comprehensive understanding of its diagnostic capabilities in the context of Parkinson's Disease.

5. Conclusion and Future work

This study underscores the efficacy of employing data fusion, specifically Principal Component Analysis (PCA) coupled with the K-Nearest Neighbors (KNN) classifier, in significantly enhancing the accuracy and precision of PD (PD) diagnosis. Leveraging a dataset comprising vocal and nonlinear features associated with PD, the application of PCA effectively reduced dimensionality while retaining crucial variance, optimizing the dataset for KNN classification. The KNN model, configured through 5-fold cross-validation, showcased notable improvements in correctly distinguishing PD cases, exhibiting high precision and recall rates compared to established methodologies. The visualization of the confusion matrix highlighted the KNN classifier's robustness in accurately classifying PD instances, validating its diagnostic potential. This research substantiates the promising utility of data fusion, particularly PCA and KNN, as a formidable approach for precise PD diagnosis, laying a foundation for enhanced disease detection and potentially informing clinical decision-making processes. Moving forward, future work could delve deeper into exploring ensemble techniques that amalgamate multiple classifiers to further refine PD diagnosis accuracy. Additionally, investigating the integration of diverse data modalities, such as genetic markers or neuroimaging, in conjunction with vocal and nonlinear features, could potentially bolster the predictive capabilities of the models. The exploration of more advanced feature selection methods and optimization strategies tailored to the

Doi: https://doi.org/10.54216/FPA.140120
Received: June 27, 2023 Revised: October 12, 2023 Accepted: December 15, 2023

specifics of PD datasets may contribute to refining model performance. Moreover, conducting a comprehensive validation study across diverse demographic cohorts or longitudinal datasets could validate the generalizability and robustness of the proposed data fusion framework in real-world clinical settings. Expanding these research avenues holds promise in advancing computational models for PD diagnosis, potentially fostering more accurate, timely, and personalized interventions for individuals affected by PD.

References

[1] Patwekar, Mohsina, Faheem Patwekar, Syed Sanaullah, Daniyal Shaikh, Ustad Almas, and Rohit Sharma. 2023. "Harnessing Artificial Intelligence for Enhanced Parkinson's Disease Management: Pathways, Treatment, and Prospects." Trends in Immunotherapy 7 (2): 2395.

[2] Rasool, Saad, Ali Husnain, Ayesha Saeed, Ahmad Yousaf Gill, and Hafiz Khawar Hussain. 2023. "Harnessing Predictive Power: Exploring the Crucial Role of Machine Learning in Early Disease Detection." JURIHUM: Jurnal Inovasi Dan Humaniora 1 (2): 302–15.

[3] Hickman, Richard A, and Sonja W Scholz. 2023. "Precision Diagnosis and Staging of TDP-43 Proteinopathies: Harnessing the Power of Artificial Intelligence." Brain, awad175.

[4] Neto, Osmar Pinto. 2023. "Harnessing Voice Analysis and Machine Learning for Early Diagnosis of Parkinson's Disease: A Comprehensive Study Across Diverse Datasets." Available at SSRN 4617895.

[5] Torrado, Juan C, Bettina S Husebo, Heather G Allore, Ane Erdal, Stein E Fæø, Haakon Reithe, Elise Førsund, Charalampos Tzoulis, and Monica Patrascu. 2022. "Digital Phenotyping by Wearable-Driven Artificial Intelligence in Older Adults and People with Parkinson's Disease: Protocol of the Mixed Method, Cyclic ActiveAgeing Study." PloS One 17 (10): e0275747.

[6] Ranson, Janice M, Magda Bucholc, Donald Lyall, Danielle Newby, Laura Winchester, Neil P Oxtoby, Michele Veldsman, et al. 2023. "Harnessing the Potential of Machine Learning and Artificial Intelligence for Dementia Research." Brain Informatics 10 (1): 6.

[7] Siddiqui, Shahid S, Sivakumar Loganathan, Venkateswaran R Elangovan, and M Yusuf Ali. 2023. "Artificial Intelligence in Precision Medicine." In A Handbook of Artificial Intelligence in Drug Delivery, 531–69. Elsevier.

[8] Gupta, Rohan, Smita Kumari, Anusha Senapati, Rashmi K Ambasta, and Pravir Kumar. 2023. "New Era of Artificial Intelligence and Machine Learning-Based Detection, Diagnosis, and Therapeutics in Parkinson's Disease." Ageing Research Reviews, 102013.

[9] Ray, Partha Pratim, and Poulami Majumder. 2023. "The Potential of ChatGPT to Transform Healthcare and Address Ethical Challenges in Artificial Intelligence-Driven Medicine." Journal of Clinical Neurology (Seoul, Korea) 19 (5): 509.

[10] Liu, Yukun, Chengxuan Zheng, and Baha Ihnaini. 2023. "Harnessing Transfer Learning for Alzheimer's Disease Prediction." In International Conference on Artificial Intelligence, Virtual Reality, and Visualization (AIVRV 2022), 12588:243–49.

[11] Torrado Vidal, Juan Carlos, Bettina Elisabeth Franziska Husebø, Heather G Allore, Ane Erdal, Stein Erik Fæø, Haakon Reithe, Elise Førsund, Charalampos Tzoulis, and Monica Patrascu. 2022. "Digital Phenotyping by Wearable-Driven Artificial Intelligence in Older Adults and People with Parkinson's Disease: Protocol of the Mixed Method, Cyclic ActiveAgeing Study."

[12] Pradesh, Arunachal. n.d. "SCIENTIFIC DISCOVERIES IN THE AGE OF ARTIFICIAL INTELLIGENCE."

[13] Pise, Anil Audumbar, Khalid K Almuzaini, Tariq Ahamed Ahanger, Ahmed Farouk, Piyush Kumar Pareek, Stephen Jeswinde Nuagah, and others. 2022. "Enabling Artificial Intelligence of Things (AIoT) Healthcare Architectures and Listing Security Issues." Computational Intelligence and Neuroscience 2022.

[14] Khalil, N., Elkholy, M. and Eassa, M. (2023) "A Comparative Analysis of Machine Learning Models for Prediction of Chronic Kidney Disease", Sustainable Machine Intelligence Journal, 5. doi: 10.61185/SMIJ.2023.55103.

[15] Olaniyan, Olugbemi T, Charles O Adetunji, Ayobami Dare, Olorunsola Adeyomoye, Mayowa J Adeniyi, and Alex Enoch. 2023. "Clinical Applications of Deep Learning in Neurology and Its Enhancements with Future Predictions." In Artificial Intelligence for Neurological Disorders, 209–24. Elsevier.

[16] Krishnan, Gokul, Shiana Singh, Monika Pathania, Siddharth Gosavi, Shuchi Abhishek, Ashwin Parchani, and Minakshi Dhar. 2023. "Artificial Intelligence in Clinical Medicine: Catalyzing a Sustainable Global Healthcare Paradigm." Frontiers in Artificial Intelligence 6.

[17] Shen, Xuechen, and Katsuhiko Ariga. 2023. "Disease Diagnosis with Chemosensing, Artificial Intelligence, and Prospective Contributions of Nanoarchitectonics." Chemosensors 11 (10): 528.

Doi: https://doi.org/10.54216/FPA.140120
Received: June 27, 2023 Revised: October 12, 2023 Accepted: December 15, 2023

[18] Vora, Lalitkumar K, Amol D Gholap, Keshava Jetha, Raghu Raj Singh Thakur, Hetvi K Solanki, and Vivek P Chavda. 2023. "Artificial Intelligence in Pharmaceutical Technology and Drug Delivery Design." Pharmaceutics 15 (7): 1916.
[19] Lim, Ashley Cha Yin, Pragadesh Natarajan, R Dineth Fonseka, Monish Maharaj, and Ralph J Mobbs. 2022. "The Application of Artificial Intelligence and Custom Algorithms with Inertial Wearable Devices for Gait Analysis and Detection of Gait-Altering Pathologies in Adults: A Scoping Review of Literature." Digital Health 8: 20552076221074130.
[20] Lyall, Donald M, Andrey Kormilitzin, Claire Lancaster, Jose Sousa, Fanny Petermann-Rocha, Christopher Buckley, Eric L Harshfield, et al. 2023. "Artificial Intelligence for Dementia—Applied Models and Digital Health." Alzheimer's \& Dementia.
[21] Sahu, Mehar, Rohan Gupta, Rashmi K Ambasta, and Pravir Kumar. 2022. "Artificial Intelligence and Machine Learning in Precision Medicine: A Paradigm Shift in Big Data Analysis." Progress in Molecular Biology and Translational Science 190 (1): 57–100.
[22] Marks, R. 2023. "Artificial Intelligence and Aging: Potential and Precautions." MOJ Gerontol Ger 8 (2): 43–48.
[23] Battineni, Gopi, Nalini Chintalapudi, Mohammad Amran Hossain, Giuseppe Losco, Ciro Ruocco, Getu Gamo Sagaro, Enea Traini, Giulio Nittari, and Francesco Amenta. 2022. "Artificial Intelligence Models in the Diagnosis of Adult-Onset Dementia Disorders: A Review." Bioengineering 9 (8): 370.
[24] Wang, Chan, Tianyiyi He, Hong Zhou, Zixuan Zhang, and Chengkuo Lee. 2023. "Artificial Intelligence Enhanced Sensors-Enabling Technologies to next-Generation Healthcare and Biomedical Platform." Bioelectronic Medicine 9 (1): 17.
[25] Mohammadi, Aliasghar Tabatabaei, Erfan Ghanbarzadeh, Nadia Ghasemi Darestani, Mohammad Nouri, Monireh Rasoulzadehzali, Maryam Moghadamnia, Shadi Nazarizadeh, Reza Nahavandi, Mehrdad Jahedi, and others. 2023. Advancements in Medical Science and the Pharmaceutical Industry: Artificial Intelligence in Medicine, Regenerative Medicine and Stem Cells, New Developments in Pharmaceuticals (Nano Drugs). Nobel Sciences.
[26] Najjar, Reabal. 2023. "Redefining Radiology: A Review of Artificial Intelligence Integration in Medical Imaging." Diagnostics 13 (17): 2760.
[27] Albahri, A S, Ali M Duhaim, Mohammed A Fadhel, Alhamzah Alnoor, Noor S Baqer, Laith Alzubaidi, O S Albahri, et al. 2023. "A Systematic Review of Trustworthy and Explainable Artificial Intelligence in Healthcare: Assessment of Quality, Bias Risk, and Data Fusion." Information Fusion.
[28] Singh, Law Kumar, Munish Khanna, and Rekha Singh. 2023. "Artificial Intelligence Based Medical Decision Support System for Early and Accurate Breast Cancer Prediction." Advances in Engineering Software 175: 103338.
[29] Ghaffar Nia, Nafiseh, Erkan Kaplanoglu, and Ahad Nasab. 2023. "Evaluation of Artificial Intelligence Techniques in Disease Diagnosis and Prediction." Discover Artificial Intelligence 3 (1): 5.

Doi: https://doi.org/10.54216/FPA.140120
Received: June 27, 2023 Revised: October 12, 2023 Accepted: December 15, 2023