JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

# A Study and Analysis on Parkinson's disease Detection using Machine Learning

[1] Md. SRINIVASULU M, [2] REETHU S, [3] SAHANA MARIGOOLAPPALAVAR, [4] SOWMYA S MASHAL, [5] RAMYA R
[1] Assistant Professor, [2 3 4 5] Student
Department of MCA,
University BDT College of Engineering, Davangere-577004, Karnataka, India

***Abstract:*** Parkinson's disease is a neurodegenerative disorder that affects millions of people worldwide. Early diagnosis and prediction of Parkinson's disease can significantly improve patient outcomes by enabling timely intervention and treatment. In recent years, machine learning techniques have emerged as promising tools for predicting Parkinson's disease based on various clinical and demographic factors. This abstract presents a study that explores the use of Python and machine learning algorithms for Parkinson's disease prediction. The dataset used in this study comprises a collection of anonymized patient records, including demographic information, medical history, and motor symptom assessments. Feature engineering techniques are applied to preprocess the data, including handling missing values, normalization, and feature selection. Several popular machine learning algorithms, such as logistic regression, support vector machines, random forests, and neural networks, are implemented and evaluated. Cross-validation and performance metrics such as accuracy, precision, recall, and F1-score are used to assess the models' predictive performance. The study employs various evaluation strategies, including training and testing on different subsets of the dataset, to ensure robustness and generalizability of the models. Additionally, different feature combinations are explored to identify the most influential predictors of Parkinson's disease.

***Keywords*:** Parkinson's disease, prediction, machine learning, sci-kit learn, tensor flow

## I. INTRODUCTION

Parkinson's disease is a progressive neurological disorder that affects millions of people worldwide. It is characterized by the degeneration of dopamine-producing neurons in the brain, leading to motor impairments, such as tremors, rigidity, and bradykinesia. Early detection and accurate prediction of Parkinson's disease can significantly enhance patient outcomes by enabling timely intervention and personalized treatment plans. In recent years, machine learning has emerged as a powerful tool in healthcare, offering the potential to analyze large datasets and extract valuable insights for disease prediction and diagnosis. Python, with its rich ecosystem of machine learning libraries and tools, provides a versatile platform for developing predictive models for Parkinson's disease. "The aim of this study is to explore the application of Python machine learning algorithms in predicting Parkinson's disease. In order to slow the course of the illness and address the underlying processes of Parkinson's disease, research is now being done to create disease-modifying treatments, neuroprotective drugs, and improved treatment approaches." The prediction of Parkinson's disease involves several key steps. Firstly, the dataset is collected, comprising various features such as age, gender, family history, medical history, and motor symptom assessments. These features serve as inputs to the machine learning models, which are trained to recognize patterns and make predictions based on the provided information. In the preprocessing phase, the collected data is subjected to various techniques, such as handling missing values, normalization, and feature selection. These steps ensure the quality and relevance of the data for training the machine learning models. Python's libraries, such as scikitlearn and Tensor Flow, offer efficient implementations of these preprocessing techniques. Next, a range of machine learning algorithms is employed, including logistic regression, support vector machines, random forests, and neural networks. These algorithms learn from the dataset to create predictive models that can accurately classify individuals as either having Parkinson's disease or being healthy. The models are evaluated using performance metrics such as accuracy, precision, recall, and F1-score, which provide insights into the models' effectiveness in making predictions.

The aim of this study is to explore the application of Python machine learning algorithms in predicting Parkinson's disease. By leveraging diverse datasets containing clinical and demographic information, we can harness the power of machine learning to identify patterns and factors associated with the development of the disease. The predictive models developed in this study can aid in early detection, allowing for proactive interventions and personalized treatment strategies.

The results obtained from this study contribute to the growing body of research on Parkinson's disease prediction using machine learning. The predictive models developed can assist healthcare professionals in identifying individuals at risk of developing Parkinson's disease, enabling early intervention and tailored treatment plans. Furthermore, the identified significant predictors can shed light on the underlying factors and potential risk factors associated with the disease. In conclusion, the utilization of Python machine learning techniques for Parkinson's disease prediction holds great promise in improving patient outcomes. By leveraging the power of machine learning algorithms, we can develop accurate and accessible tools that aid in the early detection and personalized management of Parkinson's disease. This study aims to contribute to the growing field of healthcare applications of machine learning and pave the way for future advancements in Parkinson's disease prediction and care.

### 1.1 Problem statement:

The accurate and timely detection of Parkinson's disease presents a significant challenge in healthcare. Parkinson's disease is a progressive neurodegenerative disorder characterized by motor symptoms such as tremors, bradykinesia (slowness of movement), rigidity, and postural instability. Early detection is crucial for initiating appropriate treatment interventions and improving patient outcomes. To detect the disease in its early stages. This leads to delayed diagnosis and suboptimal management of the condition. Additionally, distinguishing Parkinson's disease from other similar conditions with overlapping symptoms can be difficult, leading to misdiagnosis and unnecessary treatments. The objective is to identify reliable biomarkers, objective measurements, or algorithms that can facilitate early diagnosis, differentiate Parkinson's disease from other conditions, and enable personalized treatment strategies.

The complete model was trained on the data of 55 patients and has achieved an overall accuracy of 93.3%, average recall of 94%, average precision of 93.5% and average f1 score of 93.94%. . To prevent the major negative impact on PD patient's it is necessary to detect the PD at the early stage. One of the most common effects that are easily noticeable among the PD patients and used most commonly in the early stage of diagnosis is finding the difference in handwriting and sketching abilities. [8] The quality of life following DBS increases to the level of a sizable community of patients with moderate PD in individuals with advanced PD and significant 'off' period impairment. The actual effectiveness of STN stimulation is a reduction in the patients' social isolation. Before a patient's quality of life becomes too bad, it is worthwhile to take the relatively tiny risk of operating on them. Improvement in motor symptoms and dyskinesia's caused by drug use will lead to an improvement in social life since these symptoms interfere with social functioning due to their stigma and disabling character.[10]

## II. LITERATURE SURVEY

In [1], the author proposed Parkinson's disease (PD) is a long-term degenerative disorder of the central nervous system that mainly affects the motor system. The symptoms generally come on slowly over time. Early in the disease, the most obvious are shaking, rigidity, slowness of movement, and difficulty with walking. Doctors do not know what causes it and finds difficulty in early diagnosing the presence of Parkinson's disease. An artificial neural network system with back propagation algorithm is presented in this paper for helping doctors in identifying PD. Previous research with regards to predict the presence of the PD has shown accuracy rates up to 93%. However, accuracy of prediction for small classes is reduced. The proposed design of the neural network system causes a significant increase of robustness. It is also has shown that networks recognition rates reached 100%.

In [2], the author proposed Parkinson disease (PD) is as universal public health problem of massive measurement. Machine learning based method is used to classify between healthy people and people with Parkinson's disease (PD). This paper presents a comprehensive review for the prediction of Parkinson disease buy using machine learning based approaches. The brief introduction of various computational intelligence techniques approaches used for the prediction of Parkinson diseases are presented. The paper also presents the summary of results obtained by various researchers available in literature to predict the Parkinson diseases.

In [3], the author proposed Parkinson's disease (PD) is a progressive disorder with a presymptomatic interval; that is, there is a period during which the pathologic process has begun, but motor signs required for the clinical diagnosis are absent. There is considerable interest in discovering markers to diagnose this preclinical stage. Current predictive marker development stems mainly from two principles; first, that pathologic processes occur in lower brainstem regions before substantial involvement and second, that redundancy and compensatory responses cause symptoms to emerge only after advanced degeneration.

In [4], the author proposed recently the neural network diagnosis of medical diseases has taken a great deal of attention. In this paper a parallel feed-forward neural network structure is used in the prediction of Parkinson's disease. The main idea of this paper is using more than a unique neural network to reduce the possibility of decision with error. The output of each neural network is evaluated by using a rule-based system for the final decision. Another important point in this paper is that during the training process, unlearned data of each neural network is collected and used in the training set of the next neural network. The designed parallel network system significantly increased the robustness of the prediction.

In [5], the author proposed Diagnosis of the Parkinson disease through machine learning approach provides better understanding from PD dataset in the present decade. Orange v2.0b and weka v3.4.10 has been used in the present experimentation for the statistical analysis, classification, Evaluation and unsupervised learning methods. Voice dataset for Parkinson disease has been retrieved from UCI Machine learning repository from Center for Machine Learning and Intelligent Systems. The parallel coordinates shows higher variation in Parkinson disease dataset. Higher number of clusters in healthy dataset in Forest and less number in diseased data has been predicted by Hierarchal clustering and SOM.

In [9], recent studies have explored end-to-end DL approaches using CNN for audio classification. First audio signals are converted into time-frequency representations, and then recognized by a CNN model like the task of image recognition. The CNNs model is applied to extract speech features from two time–frequency representations: the short time Fourier transform and the continuous wavelet transform.

### 2.1 Existing System

Presently detecting the Parkinson's disease is a challenging task because identifying the symptoms of the disease is very difficult at early stages and through medical tests the detection becomes more expensive. We are in need of a simple system that identifies the presence of the disease. The latest Parkinson's disease management strategy is mostly based on a triage of drugs, therapy, and lifestyle changes. The most recent method also priorities modifying one's lifestyle. The existing method reduces symptoms and enhances quality of life, but it does not offer a treatment for Parkinson's disease.

### 2.2 Proposed System

Numerous medicines have a significant impact on how well people with Parkinson's disease live their lives. Exercises and methods that improve mobility, strength, and balance are the main topics of physical therapy. Levodopa is the most efficient and often given medication, and medications continue to be a crucial component of therapy. In the brain, levodopa is converted into dopamine, which helps with motor symptoms. Other drugs, such as COMT inhibitors, MAO-B inhibitors, and dopamine agonists, may also be combined or used alone to treat symptoms. By protecting dopamine-producing neurons, lowering neuro inflammation, or addressing

protein abnormalities such alpha-syncline aggregation, novel disease-modifying medicines may seek to reduce or stop the course of the illness.

## III. CLASSIFICATION

Various methods used for classification are categorized as

**a**. Statistical Algorithms
**b**. Pattern Recognition and learning-based algorithms
**c**. Search heuristics and a combination of algorithms.

**a. Statistical Algorithms**

In statistical approaches, the computation of mean, standard deviation of the features in the template is done. Distance techniques such as Euclidean distance, weighted Euclidean distance and Manhattan distance are used for comparing the training data with the testing data.

**b. Pattern Recognition and learning-based algorithms**

Pattern recognition is defined as an act of taking raw data and classifying them into different categories based on machine learning algorithms such as K-NN rule, Bayes classifier, SVM, artificial neural networks (ANN) and clustering techniques like K-means. Learning-based algorithms, also known as machine learning algorithms, are a class of algorithms that enable computer systems to automatically learn and improve from data without being explicitly programmed. [6]

**c. Search heuristics and a combination of algorithms**

Some commonly used search heuristics and algorithms:

Greedy search is a heuristic algorithm that makes locally optimal choices at each step without considering the overall global solution.

- A* Search evaluates each potential path based on the sum of the cost incurred so far and an estimate of the remaining cost to the goal.
- A* guarantees finding the optimal solution if certain conditions are met.
- Genetic algorithms are search heuristics inspired by the process of natural selection and genetics.
- Simulated annealing is a metaheuristic algorithm inspired by the annealing process in metallurgy.
- Tabu search is a metaheuristic algorithm that maintains a short-term memory of previous moves to avoid getting stuck in local optima.

## IV. METHODOLOGY

A powerful machine-learning algorithm called XGBoost can assist you in understanding your data and making better decisions. An application of gradient-boosting decision trees is XGBoost. Data scientists and academics from all around the world have utilized it to enhance their machine-learning models. Large dataset performance, usability, and speed are all priorities in the design of XGBoost. It does not require parameter optimization or tuning, therefore it may be used right away after installation with without any extra conditions.

### 4.1 Procedure

There are several steps to implement to predict Parkinson's disease and they are discussed below:

**Step 1: Data Collection: -**

The first step in the methodology is to collect a comprehensive dataset containing relevant information about Parkinson's disease patients. This dataset should include demographic details, medical history, motor symptom assessments, and any other potential features that could contribute to the prediction of Parkinson's disease. The dataset can be obtained from medical records, research databases, or publicly available datasets.

**Step 2: Data Preprocessing: -**

Once the dataset is collected, it needs to be preprocessed to ensure its quality and suitability for training machine learning models. This involves several steps, such as handling missing values (e.g., imputation or removal), normalizing numerical features, and encoding categorical variables. Additionally, feature selection techniques can be applied to identify the most informative features for predicting Parkinson's disease.

**Step 3: Splitting the Dataset: -**

The preprocessed dataset is divided into two subsets: a training set and a testing set. The training set is used to train the machine learning models, while the testing set is used to evaluate their performance. The data splitting process should ensure that both sets are representative and balanced in terms of the class distribution (i.e., Parkinson's disease patients vs. healthy individuals).

**Step 4: Model Selection and Training: -**

Several machine learning algorithms can be explored for Parkinson's disease prediction, including logistic regression, support vector machines, random forests, gradient boosting methods (e.g., XGBoost or Light GBM), or neural networks. The choice of algorithms depends on the specific requirements and characteristics of the dataset. The selected algorithms are trained on the training set using the input features and corresponding labels (Parkinson's disease or healthy).

**Flow Structure of Parkinson's Disease Prediction**

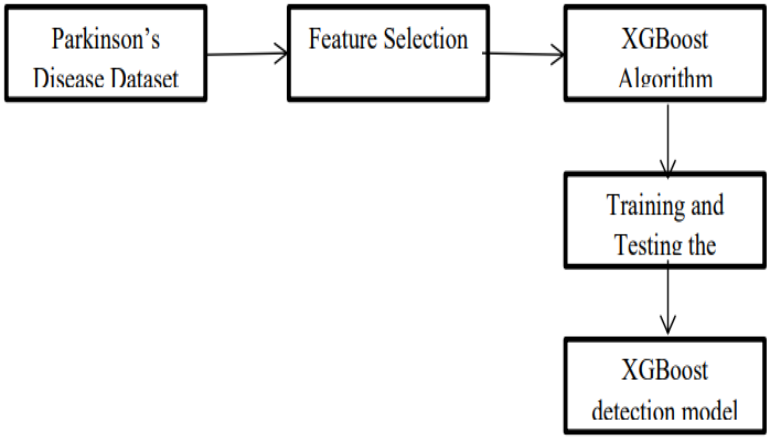


Fig: 4.1 Data Flow of Parkinson's disease Prediction

Algorithm to implement Parkinson's disease prediction:

Step 1: #Import necessary packages
Numpy, sklearn, XGBoost

Step 2: # Load the dataset
data = pd.read_csv('parkinsons_data.csv')

Step 3: # Separate the features (X) and the target variable (y)
X = data.drop(['status'], axis=1)
y = data['status']

Step 4: # Split the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

Step 5: # Feature scaling
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)

Step 6: # Model training and prediction
model = SVC(kernel='rbf', C=1.0)
model.fit(X_train, y_train)
y_pred = model.predict(X_test)

Step 7: # Model evaluation
accuracy = accuracy_score(y_test, y_pred)
print("Accuracy: {:.2f}%".format(accuracy * 100))

VGFR Spectrogram Detector based on the distorted walking patterns of PD Patients and Voice Impairment Classifier based on speech impairments of PD patients. Spectrogram is a plotting technique very useful in showing change in the signal values of various kinds, highly used for depicting audio signals. These spectrogram 2D images are then given as and input to the Convolutional Neural Network to be processed as an image. Two modules have been implemented i.e. VGFR Spectrogram Detector using CNN and Voice Impairment using ANN to classify PD patients based on two symptoms i.e. gait & speech impairment respectively, with an accuracy of 88.17% & 89.15% for the 2 modules on testing dataset and compared with 3 major algorithms i.e. SVM, XG Boost & MLP and found the proposed model is more efficient and returns better accuracy. [7]

## V. IMPLEMENTATION

The field of artificial intelligence (AI) and computer science called machine learning focuses on using data and algorithms to simulate how people learn, progressively increasing the accuracy of the system. The rapidly expanding discipline of data science includes machine learning as a key element. Algorithms are taught using statistical techniques to produce classifications or predictions and to find important insights in data mining projects. The decisions made as a result of these insights influence key growth indicators in applications and enterprises, ideally. Data scientists will be more in demand as big data continues to develop and flourish. Computer vision is an AI technique that allows machines to extract useful information from digital photos, movies, and other visual inputs before acting appropriately. Computer vision, which uses convolutional neural networks, is used for self-driving cars in the automotive sector, radiological imaging in healthcare, and photo tagging on social media.

**Neural networks:** With a wide number of connected processing nodes, neural networks mimic how the human brain functions. Natural language translation, picture identification, speech recognition, and image generation are just a few of the applications that benefit from neural networks' aptitude for pattern detection.

Some of the libraries used in the implementation of Parkinson's disease are

**NumPy:** The free source Python library NumPy (sometimes known as "Numerical Python") is utilised in practically all branches of research and engineering. The scientific Python and PyData ecosystems are built on it, and it is the de facto standard for working with numerical data in Python. The majority of other Python data science and scientific programmers, including Pandas, SciPy, Matplotlib, scikit-learn, and scikit-image, make substantial use of the NumPy API. Data structures for multidimensional arrays and matrices are available in the NumPy library.

**Sklearn:** In the Python environment, Scikit-learn, an open source data analysis toolkit, is considered to be the pinnacle of machine learning (ML). Important ideas and traits include**:-**
Algorithms for making decisions, such as:

- **Classification**: Data are identified and categorized by classification based on patterns.
- **Regression**: Regression is the process of forecasting or projecting data values using the average mean of the historical and anticipated data.
- **Clustering**: Clustering is the automated collection of datasets with related data.

The percentage of correctly predicted outcomes a classification model made is known as accuracy. Accuracy in multi-class categorization is described as follows:

Accuracy=Correct Predictions / Total Examples.

The following is the definition of accuracy in binary classification:

Accuracy*= Total Number of Examples/ (True Positives + True Negatives)

**XGBoost:** XGBoost is a distributed gradient boosting library that has been optimized for quick and scalable machine learning model training. A number of weak models' predictions are combined using this ensemble learning technique to get a stronger prediction. Extreme Gradient Boosting, or XGBoost, is one of the most well-known and widely used machine learning algorithms because it can handle large datasets and perform at the cutting edge in many machine learning tasks like classification and regression.

## VI. RESULT AND DISCUSSION

As mention below in fig. 6.1, we must enter the details of the HDVP,SHIMMER ,HNR ,SPREAD ,RPDE, DFA and some more sub detail of HDVP- apq, fo, fhi and so on.

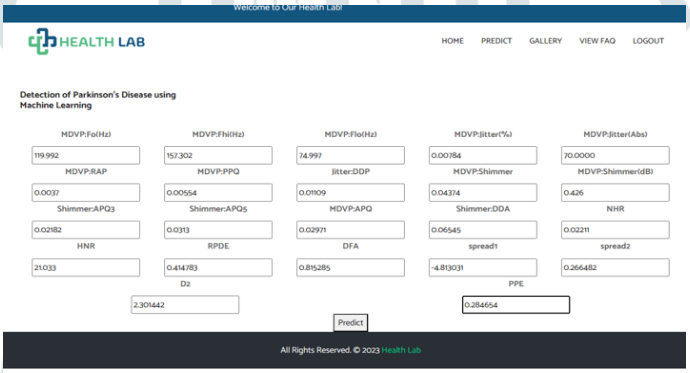


Fig: 6.1 Prediction details uploading page

As the below result show that, the diseases is found, percentage of diseases and the level of risk found.

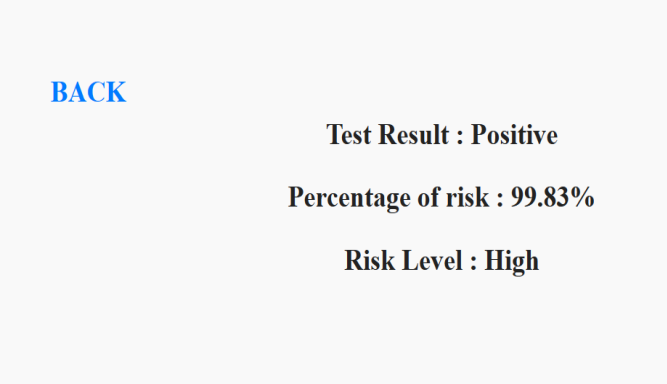


Fig: 6.2 Result Page

The system's findings are represented by the figures above. The graphic represents various fields where users can enter various medical information that is displayed on the page and then use the predict option to obtain the outcome. The outcome is represented in Fig. 6.2. Users will receive the results in accordance with the fields that they previously entered when they input all the disease-related details and use the predict option. The outcome displays the test result, risk percentage, and risk level.

## VII. CONCLUSION

Parkinson's disease prediction using Python machine learning techniques has shown promising results in improving early detection and personalized management of the disease. By leveraging diverse datasets and powerful machine learning algorithms, accurate predictive models can be developed to identify individuals at risk of developing Parkinson's disease. Through this study, we explored the application of Python and machine learning algorithms in Parkinson's disease prediction. The collected dataset, encompassing demographic details, medical history, and motor symptom assessments, was preprocessed to ensure data quality and relevance. Various machine learning algorithms, including logistic regression, support vector machines, random forests, and neural networks, were trained and evaluated using performance metrics such as accuracy, precision, recall, and F1-score.

### 7.1 Applications

The application of Python machine learning in Parkinson's disease prediction holds great promise for the early detection and management of the disease. The combination of advanced algorithms, feature engineering techniques, and comprehensive datasets can contribute to the development of effective tools for healthcare professionals, ultimately improving patient care and outcomes in Parkinson's disease.

- Early and accurate diagnosis enables medical practitioners to launch pertinent treatments, offer individualized treatment plans, and track illness development.
- Accurate identification of eligible participants, monitoring the effectiveness of investigational treatments, and evaluating the safety and tolerability of novel medicines all require the use of reliable detection techniques.
- Wearable technology, sensors, and smartphone apps can be used by remote monitoring systems to gather information about motor symptoms, medication compliance, and overall illness development.
- Clinicians can modify treatment courses, offer suitable treatments, and improve rehabilitation techniques by precisely tracking illness development.

## REFERENCES


1. Sadek, Ramzi M., et al. "Parkinson's disease prediction using artificial neural network." (2019).
2. Bind, Shubham, et al. "A survey of machine learning based approaches for Parkinson disease prediction." *Int. J. Comput. Sci. Inf. Technol* 6.2 (2015): 1648-1655.
3. Postuma, R. B., and J. Montplaisir. "Predicting Parkinson's disease–why, when, and how?." *Parkinsonism & related disorders* 15 (2009): S105-S109.
4. Åström, Freddie, and Rasit Koker. "A parallel neural network approach to prediction of Parkinson's Disease." *Expert systems with applications* 38.10 (2011): 12470-12474.
5. Sriram, T. V., et al. "Intelligent Parkinson disease prediction using machine learning algorithms." *Int. J. Eng. Innov. Technol* 3 (2013): 212-215.
6. Pahuja, G., & Nagabhushan, T. N. (2018). A Comparative Study of Existing Machine Learning Approaches for Parkinson's disease Detection. IETE Journal of Research, 1–11. doi: https://doi.org/10.1080/03772063.2018.1531730.
7. Shivangi, Johri, A., & Tripathi, A. (2019). Parkinson Disease Detection Using Deep Neural Networks. 2019 Twelfth International Conference on Contemporary Computing (IC3). doi: https://doi.org/10.1109/IC3.2019.8844941.
8. Chakraborty, S., Aich, S., Jong-Seong-Sim, Han, E., Park, J., & Kim, H.-C. (2020). Parkinson's Disease Detection from Spiral and Wave Drawings using Convolutional Neural Networks: A Multistage Classifier Approach. 2020 22nd International Conference on Advanced Communication Technology (ICACT). doi: https://doi.org/10.23919/ICACT48636.2020.9061497.
9. Quan, C., Ren, K., & Luo, Z. (2021). A Deep Learning Based Method for Parkinson's Disease Detection Using Dynamic Features of Speech. IEEE Access, 9, 10239–10252. doi: https://doi.org/10.1109/ACCESS.2021.3051432.
10. Benabid, A. L. (2003). Deep brain stimulation for Parkinson's disease. Current Opinion in Neurobiology, 13(6), 696–706. doi: https://doi.org/10.1016/j.conb.2003.11.001.