An Analysis of Parkinson Disease Prediction Using Machine Learning Approaches

Supervised by

Dr. Md. Asraf Ali

Associate Professor

Department of Software Engineering

Daffodil International University

Submitted by

Ekramul Kabir Biplab

ID: 151-35-1038

Surovi Akter Trishna

ID: 151-35-857

Department of Software Engineering

Daffodil International University

This Thesis report has been submitted in fulfillment of the requirements for the Degree of Bachelor of Science in Software Engineering.

APPROVAL

This Thesis titled "An Analysis of Parkinson Disease Prediction UsingMachine Learning Approaches", submitted by Ekramul Kabir Biplab, ID: 151-35-1038 and Surovi Akter Trishna, ID: 151-35-857 to the Department of Software Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Software Engineering and approved as to its style and contents.

BOARD OF EXAMINERS

Dr. Touhid Bhuiyan
Professor and Head
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Chairman

Dr. Md. Asraf Ali
Associate Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 1

Mohammad Khaled Sohel
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 2

Prof Dr. Mohammad Abul Kashem
Professor
Department of Computer Science and Engineering
Faculty of Electrical and Electronic Engineering
Dhaka University of Engineering & Technology, Gazipur

External Examiner

ACKNOWLEDGEMENT

Firstly, we express our heartiest thanks and gratefulness to almighty Allah for His divine blessing makes us possible to complete this study successfully.

We sincerely and heartily grateful to our advisor, **Dr. Md. Asraf Ali**, **Associate Professor,** Department of Software Engineering, Daffodil International University, Dhaka. For the support and guidance, he showed us throughout the study. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to other faculty members of Software Engineering department of Daffodil International University.

Finally, we must acknowledgement with due respect the constant support and patients of our parents.

DECLARATION

We hereby declare that, we have taken this thesis under the supervision of **Dr. Md. Asraf Ali,** Associate Professor, Department of Software Engineering, Daffodil International University. We also declare that neither this thesis nor any part of this thesis has been submitted elsewhere for award of any degree or diploma.

Ekramul Kabir Biplab
ID: 151-35-1038
Program: B.Sc.
Department of Software Engineering
Daffodil International University

Surovi Akter Trishna
ID: 151-35-857
Program: B.Sc.
Department of Software Engineering
Daffodil International University

Certified By

11/05/2019

Dr. Md. Asraf Ali
Associate Professor
Department of Software Engineering
Daffodil International University

iii © Daffodil International University

ABSTRACT

Objective: The main goal of the study is to inspect the performance of three Supervised algorithms for improving the Parkinson disease diagnosis by detection.

Methods: We used three machine learning techniques for the detection of Parkinson disease datasets. SVM, KNN, and LR were used for prediction of Parkinson Disease. The performance of the classifiers was evaluated via recall, precision, f 1 measure and accuracy.

Results: SVM shows the accuracy level 100% for Parkinson disease prediction. LR achieved the second highest classification accuracy of 97%. Moreover, in the terms of accuracy for analyzing Parkinson disease datasets, KNN achieved the worst performance (i.e. 60%).

Conclusion: Our finding showed that the SVM obtained the highest performance for analyzing the Parkinson datasets. This study has emphasized the current Parkinson research trends and scope in relation to clinical research fields by machine learning techniques. That will be an effective impact in the field of Parkinson disease.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

TABLES **PAGE NO**

CHAPTER ONE
INTRODUCTION

1.1 Background

In the present era, disease prediction is the most important task for medical institutions and physicians in order to decide the finest possible physicians' decisions. Possible incorrect decisions are opine of the major cause to delays in medical care or even loss of life. On the other hands, there is another viewpoint about medical services is a big commercial in every time. The business marketplace in this fields always running and rapidly growing than other fields. Most of the patients are always searching for good treatment care for better medical services. In reality, they cannot afford the treatment cost most of the time. Possibly, it's very crucial situation for this patient. Therefore, researchers have always searching in this area how can it is less possible expense for every patient and there is need of one umbrella platform for solving these problems in medical fields. Here is our main objective for improving Parkinson disease treatment is to more importance on early detection of Parkinson disease with less expenses and live for healthier life.

1.2 Motivation of the Research

Parkinson diseases are the most critical causes of death and disability worldwide. According to the Parkinson disease foundation, The affected peoples in the worldwide of Parkinson disease is projected that the 1 million people are Living by 2020 in USA (Marras et al., 2018). The medical treatment of Parkinson disease can be endorsed on Neuropathologic and Histopathologic (Gelb, Oliver, neurology, & 1999, n.d.). Medical diagnostic detection of Parkinson Disease can be done on widespread selection basing on the sensitivity and specificity of the characteristic Parkinson disease features. Therefore, the Parkinson Disease are needed to explore the clinical, pathologic, and nosology studies grounded on frequency of occurrence, characteristics, and including risk factors of samples (Aarsland, Andersen, neurology, & 2003, n.d.). Parkinson usually affects a large part of worldwide patients over the age of 50, which has affected up to now ("Parkinson's Disease Information Page | National Institute of Neurological Disorders and Stroke," n.d.). Still now there is no known cause of Parkinson disease, however, it is very likely possible to assuage symptoms knowingly in the early stage of the subjective patients (Singh, Pillay, neurobiology, & 2007, n.d.). A study claimed that around 90% of the patients affected with vocal damage ("Speech impairment in a large sample of patients with Parkinson's disease," n.d.). The Parkinson treatment is likely very costly. This causes most of the patients cannot afford the cost of the Parkinson

disease. Because if the disease is detected in the initial stage, then the cost will decrease and it will also be possible to save the patient's life. Nowadays, Parkinson disease prediction is most critical matter for clinical practitioners to take accurate decision of such disease. It's a great exercise at present time, machine learning based extensive platform can detect Parkinson disease.

1.3 Problem Statement

Medical data has growing a vast scale of volume from different clinical areas including health care services. To handle this data and attaining insights from this data there is need of Big Data analysis through Machine learning that aim to solve diverse medical and clinical problem (Hossain, Mahmud, Hossin, Haider Noori, & Jahan, 2018). Already, many of the studies show that machine learning algorithms has gained meaningfully high performance in classification-based medical problems. However, supervised learning-based methods are one of the most effective method for the research community and real-life applications on clinical fields. (Dwivedi, 2017). This works main objective is to improve the detection and diagnosis techniques of Parkinson disease treatment. Parkinson's disease can't be cured, but medications can help control your symptoms, often dramatically. So if it detects in the early stage, the cost of medication will reduce. Therefore, our study can be playing an important role for the detecting Parkinson disease with machine learning algorithms.

1.4 Research Question

In the following research questions associated to the Parkinson detection is also addressed:

(1) What is the best machine learning techniques within popular supervised learning for Parkinson detection?

1.5 Research Objectives

To study different machine learning algorithm. To evaluate the performance of different machine learning algorithms for Parkinson disease prediction. As well as compare the algorithms result and find out which algorithm is giving the best results.

1.6 Research Scope

In recent, machine learning algorithms have generated a significant influence and commitment in the Parkinson research community for detection of Parkinson disease. Moreover, machine learning

techniques are specified more precise results in disease prediction as compared to others data taxonomy techniques (Dwivedi, 2017)(Mahmud & Ahmed, 2018) . Motivated by this, the authors have used three prominent machine learning algorithms for detection and proper diagnosis of Parkinson patients. The main goal of this study is to examine the performance measurement of various prominent classification methods for this study we used three supervised learning techniques were used including k-Nearest Neighbors, Support Vector Machine and Logistics Regression. Moreover, the performance of the three classifiers was evaluated using different methods.

1.7 Thesis Organization

The rest of the study is ordered as follows, chapter 1 illustrates the objectives of this thesis, inspiration behind this thesis, research possibility and thesis organization. Chapter 2 portrayed the literature review and related works in these Parkinson disease areas. And the materials and methodology are designated with the evaluation criteria of different classifiers in Section 3. Moreover, the performance results and discussion are illustrated in Section 4. Finally, conclusions and further study are deliberated in Section 5

CHAPTER TWO
LITERATURE REVIEW

In this study, this work was designed by the qualitative research process and implements the strategies of Kitchenham and Charters (Kitchenham, Brereton, …, & 2009, n.d.). Our searching process of collecting paper was systematic mapping study (Petersen, Feldt, Mujtaba, Ease, & 2008, n.d.) for searching publication. Basically, cumulative use for paper selection. We have used a few keywords in our searching process. These studies are searched in Parkinson Disease, Parkinson Disease data sets, Machine Learning Techniques, Prediction, detection. We have used and/or/not. By using those keywords, we have created a search sequence which used in different online databases such as –

- IEEE Xplore
- Springer Link
- ACM Digital Library
- Science Direct
- Google Scholar
- Hindawi

Ensuing this search string, to find out journal articles we have used a condition in the searching process, and it was "English language between the years 2010 to 2019". Then 17 articles we have finally selected that have published in the above-mentioned journals or conferences. We have

applied inclusion and exclusion standards which is proposed by the Kitchenham. Our partner researchers reviewed the search verdicts from the systematic process, which helped to moderate the validity extortions. Figure 2.1 shows that the searching process of the present study.

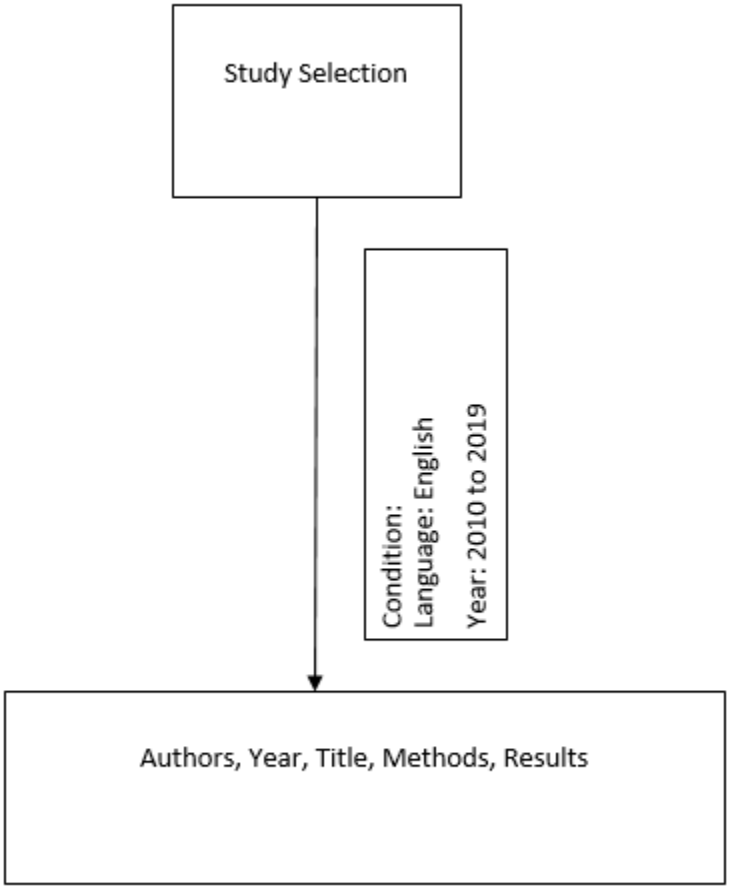

Figure 2.1. Study Selection Process

Through related work, 17 studies were done on applying and using different machine learning approaches to determine detection of Parkinson Disease. Previous work also introduces a set of studies-based detection of Parkinson diseases using machine learning algorithms. However, the outcomes of the 17 articles on machine learning used in disease prediction as follows:

Tarigoppula et al. (Sriram, Rao, Narayana, Kaladhar, & Vital, 2013) presented a comparative study between Naïve Bayes, Random Forest, Logistics Regression, Support Vector Machine to detect

Parkinson disease. SVM (i.e. 88.9%) has shown the good performance to compared NB (i.e. 69.23%), and RF (90.26%) shown the compared to SVM for the Parkinson detection. Moreover, LR (i.e. 83.66%) shown the quiet good performance. 86%). And the SVM and LDA have superior sensitivity in comparison to other classifiers. The contribution of this study is to analysis of voice data to understand presence of Parkinson diseases. In order to additional improve the diagnosis accuracy for detection of Parkinson Disease, the study (Chen et al., 2013) proposed a fuzzy based KNN model to predict Parkinson. Their study shown to the best accuracy (96.07%) obtained by the proposed algorithm including a 10-fold cross validation. Another study (Chen et al., 2016) also considers a hybrid model of detection Parkinson with compared to the existing methods and their proposed model has achieved the brilliant accuracy through 10-fold cross-validation analysis, the topmost accuracy of 96.47% and quite good accuracy of 95.97%. Moreover, The experimental (Hariharan, Polat, & Sindhu, 2014) results show that the maximum classification accuracy of 100% for the Parkinson's dataset via feature pre-processing . Hanzel et al.(Hazan, Hilu, Manevitz, Ramig, & Sapir, 2012) presented a new prediction system that can detect of Parkinson from voice data seems to be possible and precise with results approaching (90%) in two different data sets. Another hybrid method (Ma, Ouyang, Chen, & Zhao, 2014) named SCFW-KELM has been presented for the diagnosis of Parkinson disease. The result of proposed method is effective for Parkinson detection by MAE for the Total-UPDRS and Motor-UPDRS were achieved respectively MAE = 0.4656 and MAE = 0.4967 (Nilashi, Ibrahim, Ahmadi, Shahmoradi, & Farahmand, 2018). Moreover, A study (Ozcift, 2012) uses kernel Support Vector Machine for their classification and Neural Network classification scheme. Thus, the prediction performances of the 2 classifiers respectively are 91.4% and 92.9%. Hence, one study (Geetha, Professor, Head, & Sivagami, 2011) found they showed into their study that the Random Forest obtained the highest performance. But, other study showed SVM reaches upright accuracy of 83.33% (Shetty & Rao, 2017). Ferdous et al. (Wahid, Begg, Hass, Halgamuge, & Ackland, 2015) presented a comparative study between different classifiers. Their analysis shown that the RF attained the accuracy of 92.6% after standardizing gait data using the multiple regression method, competed to 80.4% (Support Vector Machine) and 86.2% (Kernel Fisher Discriminant). Hence, the study (Yadav, Kumar, & Sahoo, 2012) compared to different classifiers and showed into the results LR obtained the highest performance than others.

CHAPTER THREE
MATERIALS AND METHODS

3.1 Experimental Setup

In this study, this section represents the experimental process (figure 3.1) of the experiment including machine learning techniques. Parkinson Disease data sets have been considered in this work. Firstly, we focused on preparing and combined data from the main datasets. Moreover, we extracted 30 features from the Parkinson datasets. Then, we checked the missing values and co related values. Secondly, Data set splitting is an important task of this machine learning based fields. In this dataset, we have not found split and test datasets. Figure 3.1 shows the Parkinson data set has split into trainset and test sets. After that, 3 supervised based classifiers performed the operation. After successfully executed these algorithms SVM obtained the highest performance.

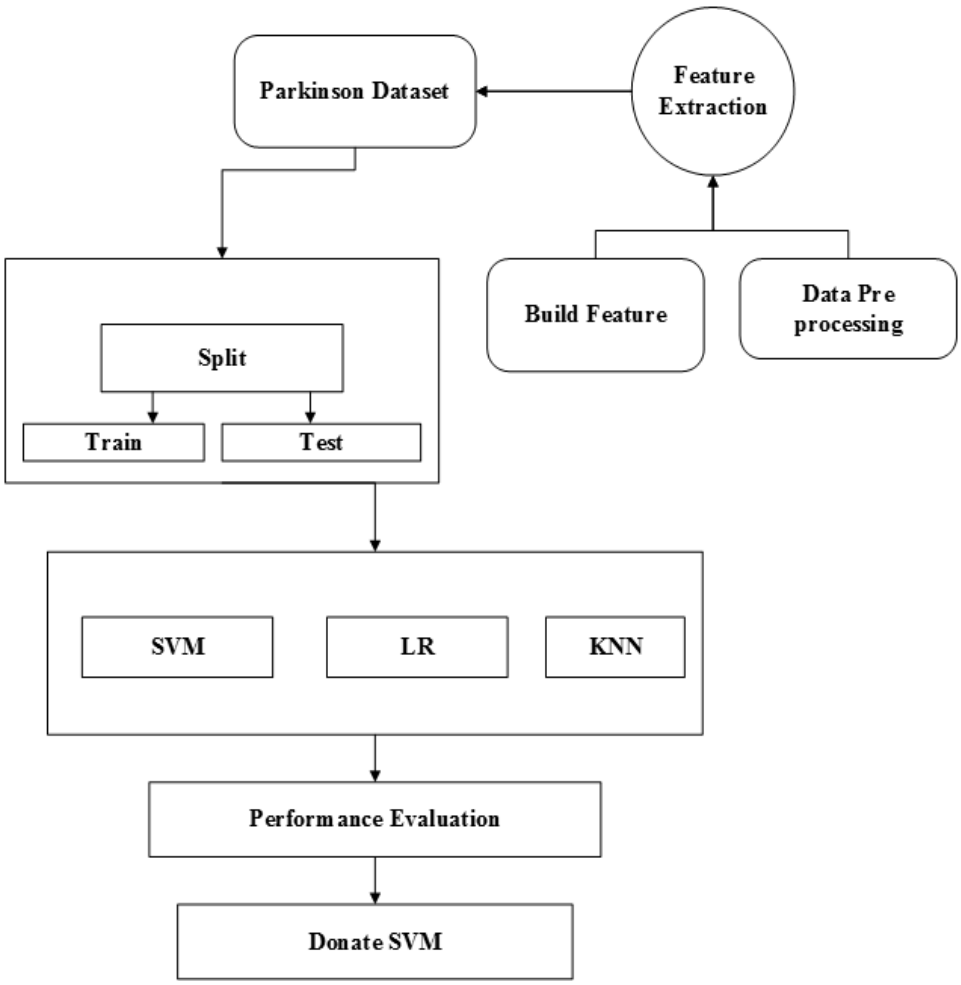

Figure 3.1 The experimental setup

3.2 Data Collection

3.2.1 Parkinson Disease Datasets

In this study, we used the Parkinson disease data from provided by the UCI Machine Learning Repository. In addition, this dataset is consisting of 62 people with Parkinson disease and 15 peoples were healthy. The authors used three types of recording are taken such as static spiral test, dynamic spiral test and stability test score. However, we chosen the particular features for data analysis which are below presented,

I. No of strokes
II. Stroke speed

III. Velocity
IV. Acceleration
V. Jerk
VI. Horizontal velocity/acceleration/jerk
VII. Vertical velocity/acceleration/jerk
VIII. Number of changes in velocity direction
IX. Number of changes in acceleration direction
X. Relative NCV
XI. Relative NCA
XII. In air time
XIII. On surface time
XIV. Normalized in-air time
XV. Normalized on-surface time
XVI. In air/on surface ratio

3.3 Data Preprocessing

In this section, firstly we extracted features from the Parkinson disease datasets. We picked the 30 columns and 77 entries of data. Then, we conducted several experiments to checking missing values, redundant values. Figure 3.2 has shown that the 30 features from the dataset which were we collected.

```
In [40]: data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 77 entries, 0 to 76
Data columns (total 30 columns):
no_strokes_st              77 non-null float64
no_strokes_dy              77 non-null float64
speed_st                   77 non-null float64
speed_dy                   77 non-null float64
magnitude_vel_st           77 non-null float64
magnitude_horz_vel_st      77 non-null float64
magnitude_vert_vel_st      77 non-null float64
magnitude_vel_dy           77 non-null float64
magnitude_horz_vel_dy      77 non-null float64
magnitude_vert_vel_dy      77 non-null float64
magnitude_acc_st           77 non-null float64
magnitude_horz_acc_st      77 non-null float64
magnitude_vert_acc_st      77 non-null float64
magnitude_acc_dy           77 non-null float64
magnitude_horz_acc_dy      77 non-null float64
magnitude_vert_acc_dy      77 non-null float64
magnitude_jerk_st          77 non-null float64
magnitude_horz_jerk_st     77 non-null float64
magnitude_vert_jerk_st     77 non-null float64
magnitude_jerk_dy          77 non-null float64
magnitude_horz_jerk_dy     77 non-null float64
magnitude_vert_jerk_dy     77 non-null float64
ncv_st                     77 non-null float64
ncv_dy                     77 non-null float64
nca_st                     77 non-null float64
nca_dy                     77 non-null float64
in_air_stcp                77 non-null float64
on_surface_st              77 non-null float64
on_surface_dy              77 non-null float64
target                     77 non-null float64
dtypes: float64(30)
memory usage: 18.1 KB
```

Figure 3.2 Features extraction from Parkinson Data

Therefore, analyzing the attributes of the selected Parkinson's datasets, some of them presented a very few values whereas others appeared not correlated with the specific medical event. There were no missing values exist in this dataset. Figure 3.3 shows the number of missing values is empty. Moreover, the Parkinson's datasets were also checked to verify the correlation of parameters. The heatmap, which is a two-dimensional graphical representation of data where the individual values that are contained in a matrix are represented as colors that is shown in figure 3.4 appear to have some correlated parameters.

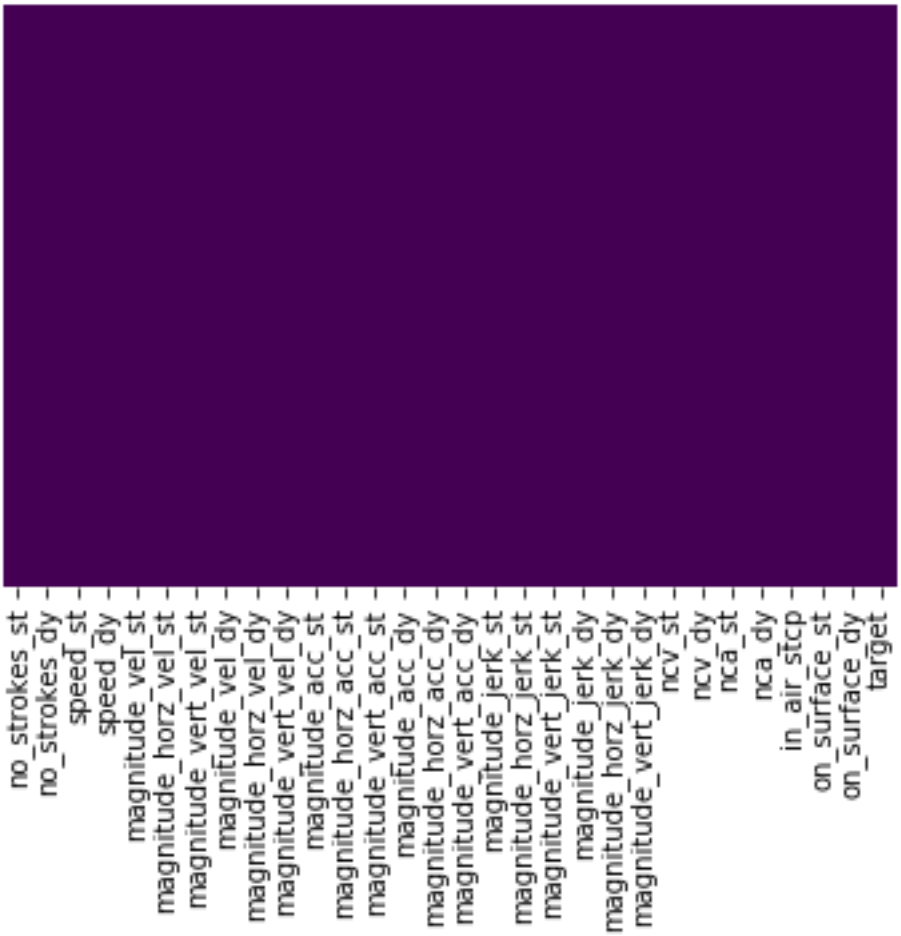

Figure 3.3. No missing values in Parkinson Data sets

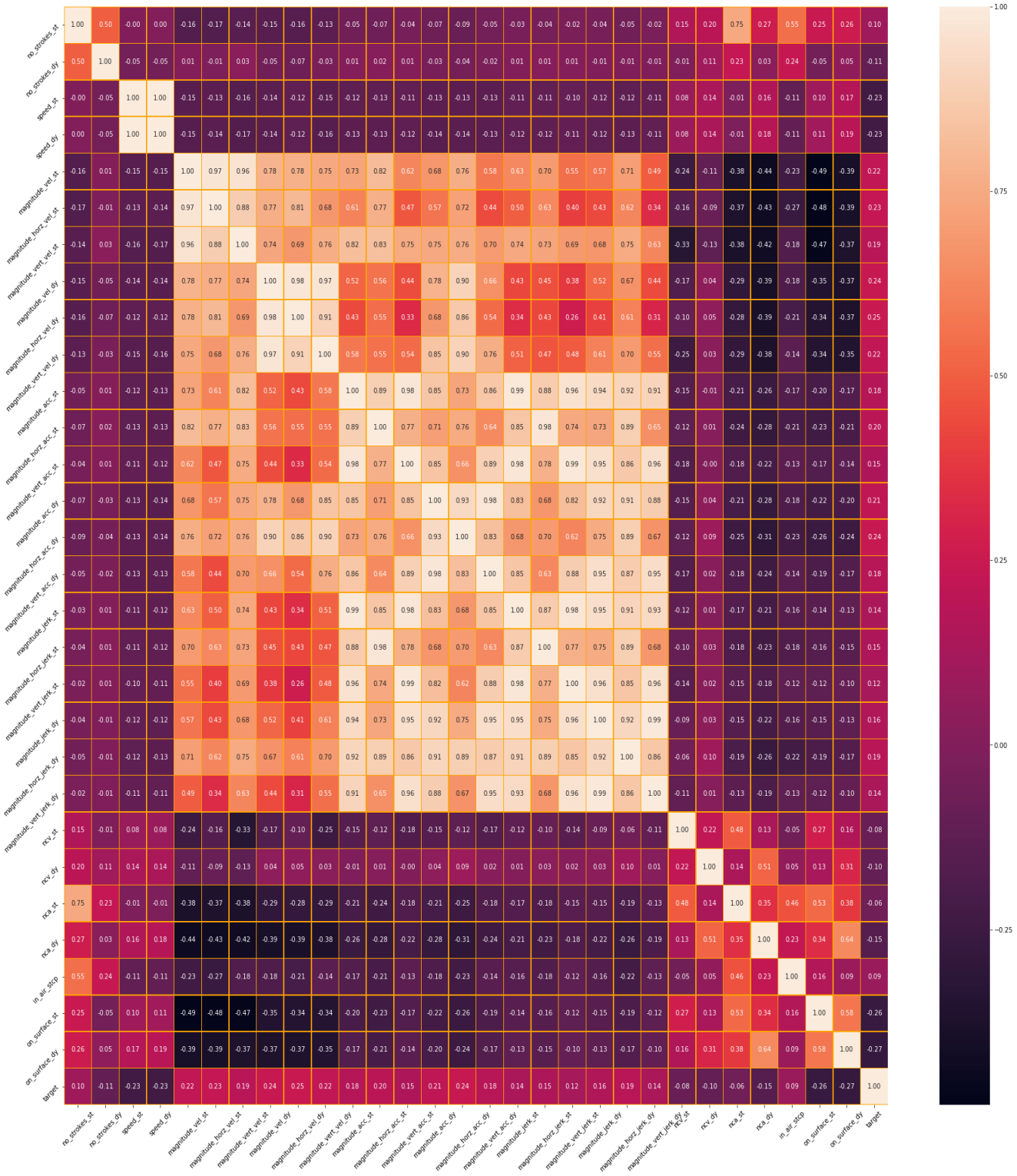

Figure 3.4. Heat map for checking correlated columns in Parkinson data sets

3.4 Classification Techniques

3.4.1 Logistics Regression

Logistic Regression was mostly used in the biological research and applications in the early 20th century (Jr, Lemeshow, & Sturdivant, 2013). Logistic Regression (LR) is one of the most used machine learning algorithms that is used where the target variable is categorical. Recently, LR is a popular method for binary classification problems. Moreover, it presents a discrete binary product between 0 and 1. Logistic Regression computes the relationship between the feature variables by assessing probabilities (p) using underlying logistic function.

3.4.2 Support Vector Machine (SVM)

Support vector machine has been first introduced by Vladimir Vapnik and Alexey Chervonenkis (Chervonenkis, 2013)(Vapnik, Guyon, Learn, & 1995, n.d.). SVM is a method of machine learning that can solve both linear and nonlinear problems. It provides good performance to solve both regression and classification problem. The SVM classification technique inspects for the optimal separable hyperplane in order to classify the dataset between two classes (Smola & Schölkopf, 2004). Finally, the model can estimate noisy data problems for new cases.

3.4.6 k- Nearest Neighbors (KNN)

The K-Nearest Neighbors is one of the most basic instance-based classification algorithms in Machine Learning. However, the KNN works on the concept that samples are close to fit in the same samples class (Zhang & Zhou, 2007). A KNN categorizes a sample to the class that is most determined among K neighboring. K is constraint for fine-tuning the classification algorithms (Guo, Wang, Bell, Bi, & Greer, 2003).

3.5 Evaluation Criteria

In this work, we used three supervised learning techniques for the detection of Parkinson disease. Therefore, the performance measurements of the classifiers are evaluated by different statistical procedures. Such as Recall, Precision, f1- measure etc. Hence, the computation method of the measurement considerations are as follows,

Accuracy = $(TP + TN) / (TP + FP + TN + FN)$ (1)

Recall or sensitivity = $TP / (TP + FN)$ (2)

Precision = $TP / (TP + FP)$ (3)

f1 = $2 * (Recall * Precision) / (Recall + Precision)$ (4)

3.6 Software and Tools

In the current study all analysis was applied in Python version 3.7.0 using Anaconda Distribution including Jupyter Notebook. The version of the Jupyter notebook server is: 5.6.0-3badce9.

CHAPTER FOUR
RESULTS & DISCUSSION

4.1 Analysis of the Results

In this section, we conducted various experiment to evaluate the three-machine learning supervised algorithms for detection of Parkinson Disease. The analysis of three classification techniques were evaluated for the detection of Parkinson disease data. Figure 4.1 shows the accuracy of three supervised techniques. Here, SVM outperformed than LR and KNN, by obtaining the highest accuracy and it is 100%.

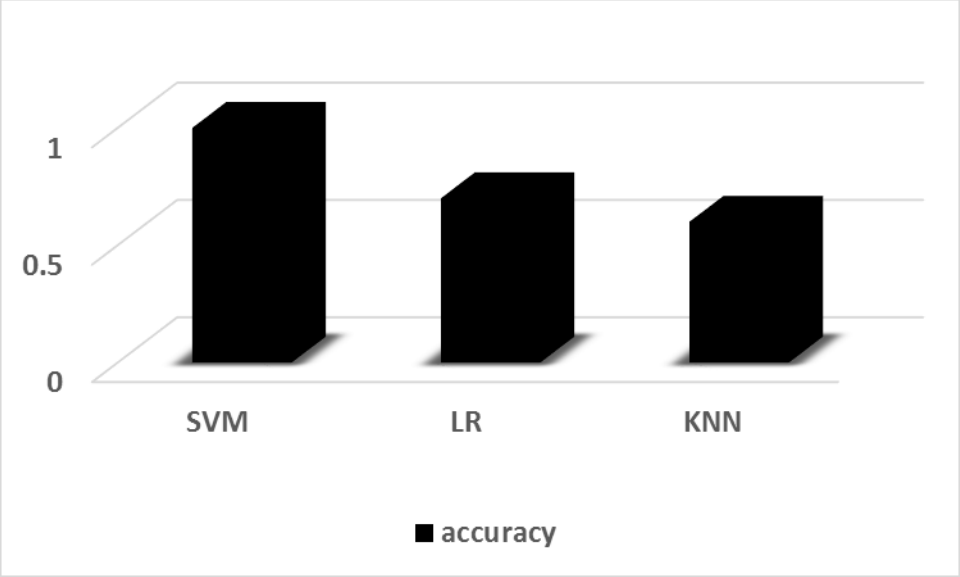

Figure 4.1. Performance of three supervised classification techniques

However, the LR achieved 70% accuracy and KNN obtained 60% accuracy. Table 4.1 and shows the classification performance measurements of three classification techniques.

Table 4.1 Classification performance measurements

	Recall	F1	Precision
SVM	1	0.67	0.5
LR	0.625	0.55	0.5
KNN	0.57	0.47	0.4

According to the performance measurements of three classification algorithms are presented is figure 4.2. The results clearly show that the SVM reached to the highest recall (100%). LR achieved the highest F1, it's 67%. KNN obtained the worst performance in terms of f1 measure (i.e. 0.47) and LR achieved the second highest score (i.e. 0.550. And KNN also achieved the worst precision (40%). Considering precision, SVM and LR shows the same performance, it's around 50%, respectively. Finally, SVM is the highest performer by overall performance.

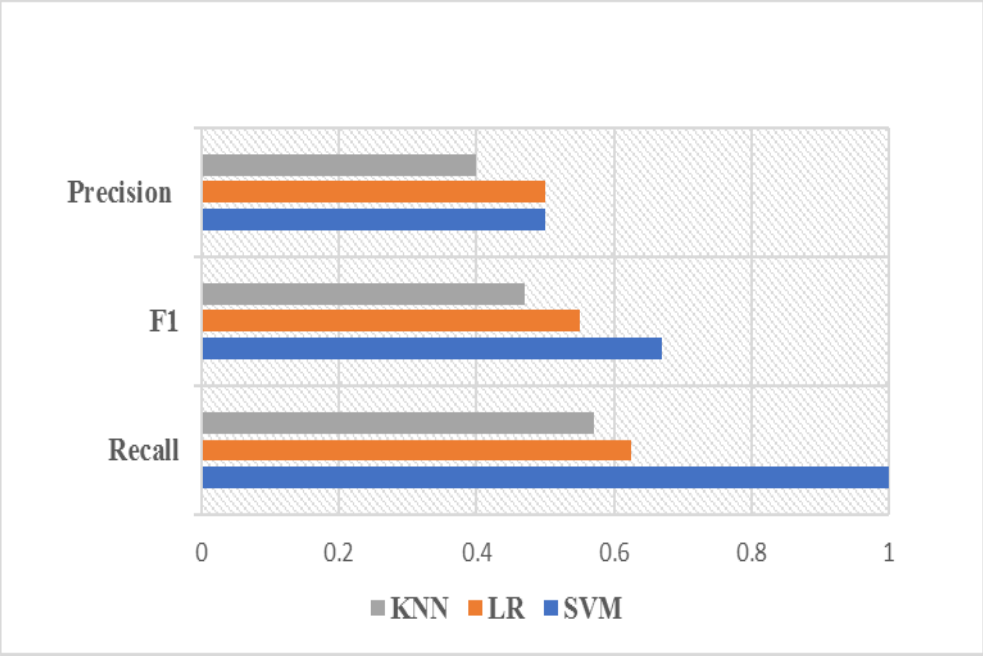

Figure 4.2. Classification performance of three classifiers

CHAPTER FIVE
CONCLUSION & RECOMMENDATION

5.1 Findings and Contribution

In this analysis, we have illustrated three supervised learning machine learning approaches. Afterwards, we evaluated the performance of the three classifiers which are used in the prediction of Parkinson disease and assessed their performance using different statistical methods. The tentative performance shows that the SVM have achieved the highest performance than the other two classifiers within the Parkinson datasets. It is 100%. This analysis has used three machine learning techniques for the detection of Parkinson disease based on several parameters. In addition, this work is part of a project that has the aim to cultivate an automated application to give more accurate action to normal occurrences and make a greater decision to multifaceted situations. The application will be able to detect in Parkinson disease in very few minutes and notify dangerous probability of having disease. This application can be outstandingly helpful in low-income peoples where is lack of medical institutes and as well as particular physicians.

5.2 Recommendation for Future Work

In our experiments, each classification algorithms were prepared and assessed on a training set that includes both positive and negative samples. Moreover, the work can be supportive for Parkinson disease detection by collecting data from different clinical and medical center and can provide more accurate results for disease prediction and diagnosis. In our research goal, there are several directions for the future work in this area of research. We will develop an application using different type of classification techniques for predicting and monitoring new and old patients. We have only investigated to three popular supervised algorithms; it can be preferring more algorithm for develop the precise model of these Parkinson disease prediction and performance can be more improved. In summary, we have painted the research objective and opportunity in relation to Parkinson disease fields by machine learning approaches, which has arising impression in health fields.

References

Aarsland, D., Andersen, K., neurology, J. L.-A. of, & 2003, undefined. (n.d.). Prevalence and characteristics of dementia in Parkinson disease: an 8-year prospective study. *Jamanetwork.Com*. Retrieved from https://jamanetwork.com/journals/jamaneurology/article-abstract/783882

Chen, H. L., Huang, C. C., Yu, X. G., Xu, X., Sun, X., Wang, G., & Wang, S. J. (2013). An efficient diagnosis system for detection of Parkinson's disease using fuzzy k-nearest neighbor approach. *Expert Systems with Applications*, *40*(1), 263–271. https://doi.org/10.1016/j.eswa.2012.07.014

Chen, H. L., Wang, G., Ma, C., Cai, Z. N., Liu, W. Bin, & Wang, S. J. (2016). An efficient hybrid kernel extreme learning machine approach for early diagnosis of Parkinson's disease. *Neurocomputing*, *184*, 131–144. https://doi.org/10.1016/j.neucom.2015.07.138

Chervonenkis, A. Y. (2013). Early History of Support Vector Machines. In *Empirical Inference* (pp. 13–20). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-41136-6_3

Dwivedi, A. K. (2017). Analysis of computational intelligence techniques for diabetes mellitus prediction. *Neural Computing and Applications*, 1–9. https://doi.org/10.1007/s00521-017-2969-9

Geetha, R., Professor, R., Head, &, & Sivagami, G. (2011). Parkinson Disease Classification using Data Mining Algorithms. *International Journal of Computer Applications*, *32*(9), 975–8887.

Gelb, D., Oliver, E., neurology, S. G.-A. of, & 1999, undefined. (n.d.). Diagnostic criteria for Parkinson disease. *Jamanetwork.Com*. Retrieved from https://jamanetwork.com/journals/jamaneurology/article-abstract/774675

Guo, G., Wang, H., Bell, D., Bi, Y., & Greer, K. (2003). KNN Model-Based Approach in Classification (pp. 986–996). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-

540-39964-3_62

Hariharan, M., Polat, K., & Sindhu, R. (2014). A new hybrid intelligent system for accurate detection of Parkinson's disease. *Computer Methods and Programs in Biomedicine*, *113*(3), 904–913. https://doi.org/10.1016/j.cmpb.2014.01.004

Hazan, H., Hilu, D., Manevitz, L., Ramig, L. O., & Sapir, S. (2012). Early diagnosis of Parkinson's disease via machine learning on speech data. *2012 IEEE 27th Convention of Electrical and Electronics Engineers in Israel, IEEEI 2012*, 29–32. https://doi.org/10.1109/EEEI.2012.6377065

Hossain, R., Mahmud, S. M. H., Hossin, M. A., Haider Noori, S. R., & Jahan, H. (2018). PRMT: Predicting Risk Factor of Obesity among Middle-Aged People Using Data Mining Techniques. *Procedia Computer Science*, *132*, 1068–1076. https://doi.org/10.1016/J.PROCS.2018.05.022

Jr, D. H., Lemeshow, S., & Sturdivant, R. (2013). *Applied logistic regression*. Retrieved from https://books.google.com/books?hl=en&lr=&id=64JYAwAAQBAJ&oi=fnd&pg=PR13&dq=applied+logistics+regression+wiley&ots=DsfO4Z4rkK&sig=qrAYpO4bqZMuaADsHhl_mjXWZFE

Kitchenham, B., Brereton, O., … D. B.-I. and software, & 2009, undefined. (n.d.). Systematic literature reviews in software engineering–a systematic literature review. *Elsevier*. Retrieved from https://www.sciencedirect.com/science/article/pii/S0950584908001390

Ma, C., Ouyang, J., Chen, H.-L., & Zhao, X.-H. (2014). An Efficient Diagnosis System for Parkinson's Disease Using Kernel-Based Extreme Learning Machine with Subtractive Clustering Features Weighting Approach. *Computational and Mathematical Methods in Medicine*, *2014*, 1–14. https://doi.org/10.1155/2014/985789

Mahmud, S. M. H., & Ahmed, R. (2018). Machine Learning Based Unified Framework for Diabetes Prediction.

Marras, C., Beck, J. C., Bower, J. H., Roberts, E., Ritz, B., Ross, G. W., … Tanner, C. (2018). Prevalence of Parkinson's disease across North America. *Npj Parkinson's Disease*, *4*(1), 21. https://doi.org/10.1038/s41531-018-0058-0

Nilashi, M., Ibrahim, O., Ahmadi, H., Shahmoradi, L., & Farahmand, M. (2018). A hybrid intelligent system for the prediction of Parkinson's Disease progression using machine learning techniques. *Biocybernetics and Biomedical Engineering*, *38*(1), 1–15. https://doi.org/10.1016/j.bbe.2017.09.002

Ozcift, A. (2012). SVM feature selection based rotation forest ensemble classifiers to improve computer-Aided diagnosis of Parkinson disease. *Journal of Medical Systems*, *36*(4), 2141–2147. https://doi.org/10.1007/s10916-011-9678-1

Parkinson's Disease Information Page | National Institute of Neurological Disorders and Stroke. (n.d.). Retrieved April 8, 2019, from https://www.ninds.nih.gov/disorders/all-disorders/parkinsons-disease-information-page

Petersen, K., Feldt, R., Mujtaba, S., Ease, M. M.-, & 2008, undefined. (n.d.). Systematic mapping studies in software engineering. *Researchgate.Net*. Retrieved from https://www.researchgate.net/profile/Michael_Mattsson/publication/228350426_Systematic_Mapping_Studies_in_Software_Engineering/links/54d0a8e90cf20323c218713d/Systematic-Mapping-Studies-in-Software-Engineering.pdf

Shetty, S., & Rao, Y. S. (2017). SVM based machine learning approach to identify Parkinson's disease using gait analysis. *Proceedings of the International Conference on Inventive Computation Technologies, ICICT 2016*, *2*. https://doi.org/10.1109/INVENTIVE.2016.7824836

Singh, N., Pillay, V., neurobiology, Y. C.-P. in, & 2007, undefined. (n.d.). Advances in the treatment of Parkinson's disease. *Elsevier*. Retrieved from https://www.sciencedirect.com/science/article/pii/S0301008206001511

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, *14*(3), 199–222. https://doi.org/10.1023/B:STCO.0000035301.49549.88

Speech impairment in a large sample of patients with Parkinson's disease. (n.d.). *Hindawi.Com*. Retrieved from https://www.hindawi.com/journals/bn/1999/327643/abs/

Sriram, T. V. ., Rao, M. V., Narayana, G. V. S., Kaladhar, D., & Vital, T. P. R. (2013). Intelligent Parkinson Disease Prediction Using Machine Learning Algorithms. *International*

Journal of Engineering and Innovative Technology (IJEIT), 212–215. Retrieved from http://www.ijeit.com/Vol 3/Issue 3/IJEIT1412201309_33.pdf

Vapnik, V., Guyon, I., Learn, T. H.-M., & 1995, undefined. (n.d.). Support vector machines. *Statweb.Stanford.Edu*. Retrieved from http://statweb.stanford.edu/~tibs/sta306bfiles/svmtalk.pdf

Wahid, F., Begg, R. K., Hass, C. J., Halgamuge, S., & Ackland, D. C. (2015). Classification of Parkinson's disease gait using spatial-temporal gait features. *IEEE Journal of Biomedical and Health Informatics*, *19*(6), 1794–1802. https://doi.org/10.1109/JBHI.2015.2450232

Yadav, G., Kumar, Y., & Sahoo, G. (2012). Predication of Parkinson's disease using data mining methods: A comparative analysis of tree, statistical and support vector machine classifiers. *2012 National Conference on Computing and Communication Systems, NCCCS 2012 - Proceeding*, 21–28. https://doi.org/10.1109/NCCCS.2012.6413034

Zhang, M.-L., & Zhou, Z.-H. (2007). ML-KNN: A lazy learning approach to multi-label learning. *Pattern Recognition*, *40*(7), 2038–2048. https://doi.org/10.1016/J.PATCOG.2006.12.019

APPENDIX

Appendix A: Parkinson Disease Dataset

Sample of Parkinson Disease Data

o_stroke	no_stroke	speed_st	speed_dy	magnitud	magnitud	magnitud	magnitud	magnitud	magnitud	magnitud	magnitud	magnitud	magnitud	magnitud	magnitud	magnitud	magnitud	magnitud	magnitud	magnitud	ncv_st	ncv_dy
12	2	0.000293	0.000431	0.061342	0.038319	0.03905	0.084891	0.053885	0.052829	0.000375	0.000251	0.000212	0.000456	0.000305	0.000282	8.90E-06	6.15E-06	5.03E-06	9.91E-06	6.48E-06	6.10E-06	185.25	412.8571
4	6	0.000286	0.000281	0.119159	0.077012	0.07426	0.164457	0.11136	0.099211	0.000564	0.000331	0.000327	0.000728	0.00051	0.000409	1.25E-05	8.69E-06	7.15E-06	1.48E-05	9.99E-06	8.63E-06	192.7778	173.875
4	4	0.000278	0.000277	0.113859	0.071547	0.07336	0.191965	0.117724	0.127062	0.000713	0.000428	0.000459	0.002179	0.001012	0.001782	1.64E-05	9.69E-06	1.08E-05	5.10E-05	2.31E-05	4.20E-05	263.5714	136.3333
4	4	0.000286	0.000275	0.154621	0.101146	0.0963	0.152413	0.101235	0.091241	0.000501	0.000452	0.000237	0.000539	0.000446	0.000289	1.17E-05	8.92E-06	5.50E-06	1.10E-05	8.68E-06	5.10E-06	225.2857	308.4
2	2	0.000279	0.00027	0.060058	0.039449	0.03573	0.081577	0.052451	0.050634	0.000466	0.000341	0.000236	0.00058	0.000426	0.000298	1.12E-05	8.52E-06	5.65E-06	1.32E-05	9.87E-06	6.90E-06	225.7	168
2	2	0.000251	0.000238	0.159831	0.128719	0.12483	0.275863	0.171905	0.178293	0.00068	0.000432	0.000444	0.001032	0.00069	0.00071	1.23E-05	8.51E-06	7.30E-06	1.54E-05	1.04E-05	9.49E-06	135.8889	208.25
4	2	0.005701	0.005063	0.173355	0.138019	0.07513	0.344131	0.277091	0.140489	0.001002	0.000756	0.000514	0.00179	0.001394	0.000839	2.47E-05	1.82E-05	1.34E-05	3.24E-05	2.47E-05	1.67E-05	403.6	204.75
2	2	0.004495	0.004138	0.282933	0.232454	0.12439	0.233773	0.191026	0.102773	0.001433	0.001273	0.000421	0.001157	0.00101	0.000403	2.78E-05	2.55E-05	6.61E-06	2.41E-05	2.17E-05	7.08E-06	643	193.4286
4	2	0.003696	0.00344	0.111742	0.090968	0.04685	0.143555	0.117886	0.059437	0.001086	0.000834	0.000474	0.001255	0.001077	0.000489	2.63E-05	2.15E-05	1.15E-05	2.98E-05	2.59E-05	1.16E-05	392.8571	379.6667
2	2	0.002756	0.003645	0.312659	0.249655	0.14067	0.424141	0.333874	0.198243	0.001113	0.000928	0.000455	0.001653	0.001319	0.000685	1.75E-05	1.50E-05	6.62E-06	1.89E-05	1.63E-05	7.00E-06	196.1667	296.75
4	2	0.002363	0.002181	0.2423	0.194081	0.10587	0.401756	0.321012	0.179775	0.000789	0.000646	0.000322	0.001657	0.001409	0.000704	1.36E-05	1.12E-05	5.08E-06	2.37E-05	2.07E-05	8.12E-06	302.3333	452
4	2	0.002149	0.002059	0.133674	0.110583	0.05413	0.15661	0.137781	0.067226	0.000723	0.000632	0.000236	0.00059	0.000588	0.000256	1.49E-05	1.30E-05	5.01E-06	1.17E-05	9.94E-06	4.41E-06	333.4286	263.8333
2	4	0.002	0.001918	0.364356	0.303828	0.15167	0.403164	0.340265	0.177291	0.001574	0.001277	0.000758	0.002431	0.00196	0.001069	2.71E-05	2.04E-05	1.38E-05	3.22E-05	2.59E-05	1.51E-05	261.75	279.6667
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	1.77E-05	0	0.06711	0.045253	0.03374	0	0	0	0.000466	0.000321	0.000257	0	0	0	1.12E-05	7.88E-06	6.52E-06	0	0	0	257.375	0
6	0	2.14E-05	0	0.059142	0.064157	0.05246	0	0	0	0.000449	0.000302	0.00026	0	0	0	1.03E-05	7.35E-06	5.75E-06	0	0	0	280.4444	0
0	0	1.00E-05	8.55E-06	0.208928	0.130968	0.13327	0.199471	0.122773	0.133141	0.000773	0.000553	0.00041	0.00034	0.00053	0.000546	1.28E-05	1.03E-05	6.00E-06	1.12E-05	8.70E-06	5.35E-06	174.5	264.3333
20	0	1.62E-05	0	0.142423	0.090339	0.09156	0	0	0	0.00049	0.000292	0.000379	0	0	0	9.95E-06	6.54E-06	6.26E-06	0	0	0	177.5556	0
8	0	1.80E-05	0	0.049118	0.031	0.03085	0	0	0	0.00029	0.000133	0.000171	0	0	0	6.75E-06	4.44E-06	4.11E-06	0	0	0	307.75	0
2	2	1.76E-05	1.64E-05	0.220058	0.142634	0.13517	0.331228	0.214605	0.206542	0.00097	0.00056	0.000658	0.001457	0.000915	0.000918	1.42E-05	8.53E-06	9.59E-06	1.83E-05	1.15E-05	1.19E-05	249.2	156.8
0	0	1.61E-05	1.57E-05	0.128255	0.085838	0.07564	0.1963	0.13132	0.120119	0.000341	0.000228	0.000214	0.000551	0.000328	0.000364	6.92E-06	4.55E-06	4.01E-06	8.50E-06	5.01E-06	5.69E-06	311.8333	297.5
30	0	2.02E-05	1.57E-05	0.027254	0.017237	0.01612	0.1963	0.13132	0.120119	0.000324	0.000202	0.000136	0.000551	0.000328	0.000364	8.09E-06	5.23E-06	4.95E-06	8.50E-06	5.01E-06	5.69E-06	308.4444	297.5
0	0	1.13E-05	1.12E-05	0.218554	0.144268	0.13713	0.351252	0.233852	0.221706	0.000725	0.000448	0.000473	0.001396	0.000856	0.000924	1.22E-05	8.32E-06	7.17E-06	1.57E-05	9.65E-06	1.01E-05	205.1667	256
2	2	1.11E-05	9.74E-06	0.158574	0.106667	0.09356	0.247058	0.154667	0.157234	0.000526	0.000404	0.000335	0.000951	0.000552	0.000665	1.11E-05	7.30E-06	6.65E-06	1.17E-05	6.82E-06	8.00E-06	237.1667	210.25
2	2	1.05E-05	2.40E-05	0.233776	0.143445	0.14319	0.3974	0.547406	0.587756	0.000324	0.000478	0.000558	0.006636	0.004045	0.00454	1.35E-05	8.03E-06	9.13E-06	6.01E-05	3.70E-05	4.03E-05	177.5	318
16	2	1.05E-05	1.48E-05	0.158056	0.099742	0.10182	0.291125	0.184873	0.19029	0.000469	0.000251	0.000329	0.001137	0.000691	0.000761	8.98E-06	4.87E-06	6.52E-06	1.54E-05	8.71E-06	1.07E-05	261.1667	230.2