

Prediction of Parkinson's Disease at Early Stage using Big Data Analytics



Siva Sankara Reddy Donthi Reddy, Udaya Kumar Ramanadham

Abstract: Due to technological improvements in healthcare industry and clinical medicine, it requires to adapt new software techniques and tools to predict, diagnose and analyze disease patterns for making decisions in the early stage of disease. Parkinson's disease is a neurodegenerative disorder. The PD damage the motor skills and may create speech problem and also affect the decision making process. Many people suffers with PD all over the world from many years. Day by day, the PD data has been increased, so the existing data mining predictive methods and tools does not give accurate results early for making decisions by doctors to save and increase the patient life period. Early PD symptoms can be detected by Big Data Analytics and proper medicine will be provided at the right time. In this paper, we are doing survey of predictive methods, Big Data Analytical techniques and also earlier researchers results presented.

Keywords: Healthcare Industry, Clinical Medicine, Parkinson's disease, Neurodegenerative, Big Data Analytics, Prediction.

I. INTRODUCTION

In healthcare industry, Big Data Analytics is used to play good role to extend in the prediction of diseases in the early stage, find difficulties, flaws, better procedures, increase caring patients and decrease the treatment cost [1]. Many people concentrate on healthcare issues and to solve the problems of healthcare industry. It has become too difficult to identify the disease in the right time by using existing conventional tools and methods of data mining because complex disease [2]. This paper describes mainly various Big Data Analytical techniques that may be used in diagnosing of right disease in the right time. Most of the researchers use Big Data Analytical techniques in diagnosing Parkinson's disease such as Decision Tree Induction, Bayesian Classification, Support Vector Machine, Rule-Based Classification, Multilayer Perceptron, Random Forest, Boosted Logistic Regression, principal component analysis (PCA), linear discriminant analysis(LDA), Artificial Neural Network,

K-Nearest Neighbor, M5Rules, M5P, REP Tree and other methods [3].

II. LITERATURE SURVEY

Zhenhao Cai et al. [1] proposed a method a novel approach CBFO-FKNN to differentiate PD people with normal people. In this proposed method, the main two key parameters are determined by using chaos theory. After conducting experiment, the results shows the difference between PD affected and normal people with the key terms such as presbyphonic and dysphonic. The CBFO method has good performance and it provides more convenience in the clinical diagnosing process to take accurate decision.

Srishti Grover et al. [2] developed a method to predict the Parkinson's disease status by using NN on patient's Parkinson's telemonitoring voice data. The TensorFlow deep learning library was used by authors to identify status of PD to implement neural networks. The results of this method are best other than the previous method of earlier research work.

D.Karthiga et al. [3] discussed the study of various data mining methods in diagnosing the Parkinson and atypical Parkinson disease in the early stage to enhance the quality of living.

Muthumanickam S et al. [4] proposed several machine learning and deep learning algorithms were discussed. In this paper, after analyzing many classifiers the Deep Belief Network (DBN) outperforms all other classifiers. The use of artificial intelligence techniques in the field of medicine, especially in the early diagnosis of PD has proved to be very efficient and effective.

Mahalakshmi Senthilarumugam Veilukandammal et al [5] described the various methods to meet the PPMI data challenges. Mainly this work concentrates in grasping of basic identities and discoveries using descriptive tests with different attributes. The significant attributes are identified by exploration of data.

R. Prashanth et al. [6] described MDS-UPDRS patient questionnaires for classifying early stage Parkinson's disease. The authors tested these prediction models to predict PD symptoms with high accuracy. These diagnostic models can have strength to use in clinical process by clinicians when doctors are not available to save patient lives.

Athithiyan G et al. [7] presented about the introduction of machine learning in brain diseases prediction. The decision tree algorithm was used in predicting that which type of disease. The types of disease considered are Parkinson's disease, stress, amnesia, bipolar disorder. The symptoms and physical behavior of the user has been considered. The result has been discussed through the graphical representation.

Revised Manuscript Received on April 27, 2020.

* Correspondence Author

Siva Sankara Reddy Donthi Reddy*, PhD scholar, Department of CSE, Bharath Institute of Higher Education and Research (BIHER), Chennai, Tamilnadu, India, d.sivasankarareddy@gmail.com

Udaya Kumar Ramanadham, Professor, Department of Information Technology, Bharath Institute of Higher Education and Research (BIHER), Chennai, Tamilnadu, India, rsukumar2007@gmail.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Diellza Nagavci et al. [8] described methods for exploring and analyzing earlier research which is disease predicted healthcare data. The authors concentrated mainly big data analytics research. The Big Data Analytics give better prediction of PD than the existing methods.

Athmaja. S et al. [9] described various machine learning algorithms for predicting PD at the earliest period. The authors also presented the more advancement of prediction techniques for achieving accurate and reliable results on big data.

Sakina Imam et al. [10] proposed modified PCA-KNN classifier methodology to diagnose the Parkinson disease. The simulation of proposed methodology has performed on MATLAB2012 a simulator toolbox, which contain various data mining functions and the comparative analysis of the proposed method and existing method was performed using accuracy precision parameter. The simulation results of proposed methodology outperform accuracy 96.92% than the existing method. The modified PCA-KNN classifier method is better for diagnosing the PD than the earlier methods.

Sonu S R et al. [11] discussed data mining algorithms which are used voice recording data sets for predicting Parkinson's disease. The PRAAT script is used to store the voice data and converts as voice attributes. After conducting tests on voice data, the Parkinson's disease status is predicted.

Jesper F et al. [12] discussed clinical and experimental of Parkinson's disease with metabolite profiling. The authors were reviewed various analytical algorithms and methods. The results are obtained after conducting tests which involving metabolic pathways.

Musaed Alhussein et al. [13] proposed a framework for PD monitoring for using in smart cities. The people health is monitored continuously by using this framework and detecting the symptoms of PD in the early stage. Then, proper treatment will be given to the PD patients to save their life. The registered doctors are received the specified decisions which are generated by vector machine-based classifier. Then, the doctors can prescribe certain medicines to the patients. By using this framework, experiments were carried out for demonstrating results. The accuracy of framework is 97.2% to predict Parkinson's disease.

Zhenhao. Cai et al. [14] proposed SVM method which can improve the prediction accuracy than the earlier prediction models. After conducting experiments, the results shows good accuracy with 97.42%.

Siva Sankara Reddy Donthi Reddy et al. [15] presented the factors such as unknown valuable hidden intuition, life style desires, social constraints, financial constraints and clinical aspects which may damage the health of a person. The authors presented the facts that used for improving healthcare analysis. This paper also describes the merits, demerits and pitfalls which can exhibit more impact in the healthcare industry.

N. Shamli et al. [16] designed a predictive structure of Parkinson's Brain Disease. The structure uses various machine learning algorithms such as DT, ANN and SVM. This framework can be used to exhibit the insights from patients. The input is PD voice dataset which is retrieved from UCI Machine Learning Repository. The experimental results will demonstrate that it can help clinical monitoring process to detect the PD in early stage and to give treatment for PD patients to increase the life period to lead peaceful life.

Kamal Nayan Reddy Challa et al. [17] described non-motor features such as RBD and olfactory loss and important biomarkers. In this paper, the authors were used various ML algorithms. The authors were designed new automation diagnosis models using Multilayered Perception algorithm, Bayes Net, Boosted Logistic Regression and Random Forest. In the observation, the Boosted Logistic Regression was provided good performance with expected accuracy of 97.16 %. The area curve was 98.9% under the ROC. Thus, this model was used to predict Parkinson's disease in the early stage.

I.D. Dinov et al. [18] proposed classification models for Parkinson's disease such as model-free and complementary model based for implementing testing and validation. The authors used big data methodology to process demographic, genetics, clinical and complex PPMI imaging data for exploring PD risk.

Arvind Kumar Tiwari [19] discussed feature selection algorithms for selecting the best feature in the existing features for predicting PD. In the observation, the random forest was provided overall accuracy 90.3% by using feature selection algorithms with selection of 20 numbers. The performance of correlation coefficient is best than the other ML algorithms such as bagging, boosting, random subspace, support vector machine, multilayer perceptron, rotation forest, decision tree based methods and random forest.

Tawseef Ayoub Shaikh [20] proposed psychiatric datasets of Parkinson's & primary tumor disease model for prediction of PD. This model can be used for predicting the probable occurrence of PD in the patients. The authors conducted tests on various algorithms and the accuracy of ANN is 90.7692%. Roger A. Barker et al. [21] reviewed the mild cognitive impairment deficits in Parkinson's disease. These are raised from different pathological pathways. In this paper, the authors are presented some important implications for better treatment and modification of future disease therapy trial.

Chandrashekhar Azad et al. [22] presented some important prediction models for the purpose of PD prediction such as ID3, decision stumps, and prediction model tree. The authors used k-fold technique to predict PD accurately. In this case, each record is trained and tested. The prediction model of decision tree exhibits better accuracy among all other prediction models.

Dr. Hari Ganesh S et al. [23] described the comparison of best accurate classification algorithms. In this paper, the authors are presented good models for test cases. By using classification techniques of data mining, the test cases were conducted successfully and the various classifiers are verified for data set of PD. The Random Tree exhibits the accuracy with 84%.

Farhad Soleimanian Gharehchopogh et al. [24] proposed ANN with two types of MLP which can use back propagation algorithms and RBF classifier. For conducting test, the authors were used MATLAB toolbox. The accuracy of MLP, RBF is 93.22% and 86.44% respectively.

Sanjivani Bhande et al. [25] described comparative analysis for illustrating benefits of various PD predictive techniques. The authors were conducted survey for providing guidance for beginner researchers for reference the existing literature on a specific application.

Tarigoppula V.S Sriram et al. [26] proposed PD diagnosis process on voice-data through machine learning algorithms. The authors presented various algorithms such as MDVP, SVM and RF which has shown good performance. The Random Forest has shown 90.26% accuracy.

Peyman Mohammadi et al. [27] described various methods for data mining algorithms such as SLR, MLP and DT for predicting PD. For testing, the 4406 training data set and 1469 data of test set were used. The correlation coefficient was produced 99.85% accuracy. The authors were used WEKA 3.6.9 tool for result observation.

Dr. R. Geetha Ramani et al. [28] described the impact of feature relevance algorithms and classification algorithms for the Parkinson Tele-monitoring dataset. There are two classes in a dataset such as Motor-UPDRS and Total-UPDRS. In this paper, the authors are presented six features selection algorithm which comprises Forward Logistic Regression, Stepwise Discriminates, Runs Filtering, ReliefF, Fisher Filtering, and Backward Logistic Regression. The ReliefF is the best feature selection algorithm. The random tree classification algorithm has shown 100% accuracy.

Siva Sankara Reddy Donthi Reddy, Dr. R. Udaya Kumar [29] described about Big Data Analytics which are used for analyzing data accurately. In this paper, Hadoop framework was designed and discussed to solve the issues.

III. PARKINSON’S DISEASE AND ITS SYMPTOMS

PD is the most dangerous neurological disorder which mainly affects the some cells in the brain and gradually it reduces the secretion of dopamine in Substantia Nigra [4]. This dopamine is a neurotransmitter found in the brain. It is essential one to function the central nervous system. The substance by name, dopamine will be produced by neurons. The dopamine level decreases in the brain, then the person faces problems in movement. If the dopamine secretion is reduced, then all the symptoms will be sub sequentially followed one by one within the life span of 3 to 4 years, and then we make sure that the person affects with PD [5]. PD increases when a person crosses age 60 years. More than 6.3 million people have been living with PD. Day by day, PD patients are increased in developing countries. But, there are no proper identification methods for PD progression in the early stage to save the patient life. The medicine was found in 1967 for PD [6].

Symptoms Of Parkinson’s Disease

The treatment of PD is more successful when it is in initial stage. The PD symptoms are notified slowly and the PD patients live many years. The one million American people were affected by PD and more than sixty thousand people are diagnosed every year. The almost all early signs of PD may be similar to other PSP, so it is too difficult to diagnose. Symptoms of Parkinson’s disease [7]:

- Cramped Hand Writing
- Tremor in finger, hand or foot
- Sleeping Problem
- Limb Hardness & Steady Moving
- Throat Voice Change
- Rigid Facial Expression
- Stooped Posture

IV. BIG DATA-AN OVERVIEW

Big Data

The huge complex data sets is termed as “Big Data” and measured by terabytes and petabytes. It is more demanding domain currently for research. This data may be formed as structured, semi-structured and unstructured. This type of data is not processed efficiently by conventional methods [8].

The big data has widely used in healthcare and medicine. The healthcare cost is increased rapidly due to increasing technology cost. There is more help in healthcare and medicine through Big Data [9].

It helps the physicians to keep the track of history of all patients and these details are fetched by only the owner or doctor [10]. All the records related to patients are stored safely in database forever in specific high-end storage systems. Currently no need to visit the doctor, but treatment is given based on the previous visited reports which are available and accessible to the patient in database systems. The patient heart beat rating and temperature is monitored through the fitted hand watch on patient’s hand when they stay in a remote place [11]. The structured data is handled by traditional database tools because the size of the data is limited for manage and it is centralized.

The conventional tools are not managed and processed the various kinds of complex formats of data [12]. There are so many issues and challenges to store, process and analyze huge sets of data. Hadoop solves the above problems with good functionality and framework. In Hadoop, there are two main sub systems such as DFS and MR. Hadoop works on the basis of distribution feature, with multiple systems. To access the data from various servers, the DFS plays very important role [13].

The operations are performed parallel by Map Reduce framework. It processes huge sets of data in distributing environment [14]. This process of this framework is having two main phases such as Mapper and Reducer. In first phase, the huge data is fragmented into small chunks and input key value is attached to entire the data. Each data unit is sent to next phase. In the second phase, the received data from first phase as input is reduced the size of output and share the same key. The software tools are used to perform analytical process on data. The traditional data warehouses may not provide good service for unstructured data of Big Data analytics and may not handle high processing of Big Data on demand [15]. The Big Data Analytical software framework is core structure for processing the large sets of Big Data across the DFS and clustered systems [16]. The Big Data representation is depicted in Figure 1.

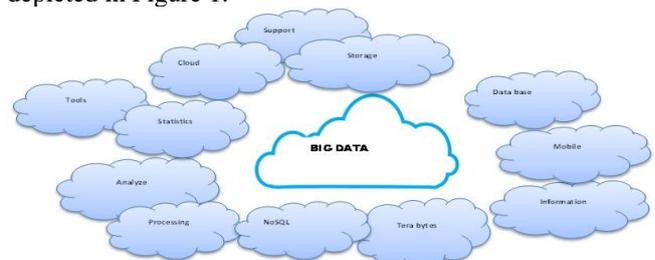


Fig. 1 Big Data [15]

Big Data Characteristics

Big Data was described with five characters that the 5 V's [17]. –Volume, Velocity, Variety, Veracity and Value. We can also consider variability. They are:

Volume- It specifies the size of data with terabytes, petabytes and exabytes. The Big Data creates as huge sets of data from various sources like social media, business processing, computers and networks. We can use Green-Plum and Hadoop to store big data.

Velocity- It specifies the data accessing speed from big data sources. In a broader prospect, it comprises the rate of change, linking of incoming data sets at varying speeds, and activity bursts.

Variety- It refers that Big Data divides into three types such as structured data, Un-structured data and semi-structured data. Structured data consists of Rows and Columns and re-sides in fixed fields in a file. Un-Structured data does not have any pre-defined format data.

Veracity – It refers that it is trustworthiness of the data. Sometimes data may be inconsistent and uncertainty in data. So, the data should be maintained integrity.

Value – It refers that the data itself is of no use or importance but it needs to be converted into something valuable to extract information.

The Big Data Characteristics are depicted in Figure 2.

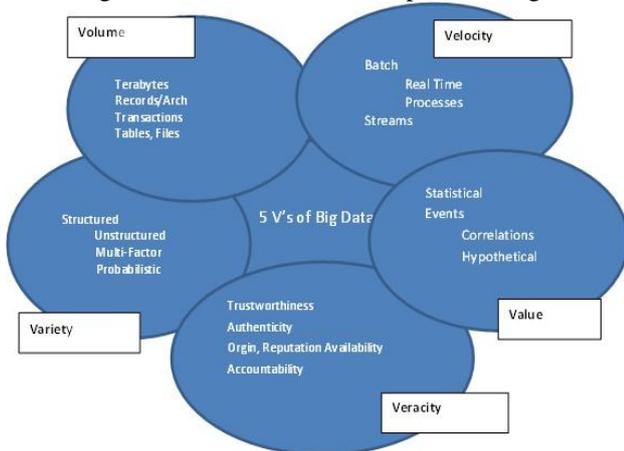


Fig. 2 Big Data Characteristics [15]

V. BIG DATA ANALYTICS

Big Data analytics is the process of applying qualitative and quantitative techniques when examining data, with the goal of providing valuable insights. Big Data Analytics largely involves collection of data from different data sources and mung it in a way that it becomes available to be consumed by analysts and finally deliver data products useful to the organization business.

Big Data Analytics provides more and useful insights that could manage business operations in right way and provide new business opportunities for organizations to generate more revenues [18].

The general process of data analytics is shown in Fig.3.

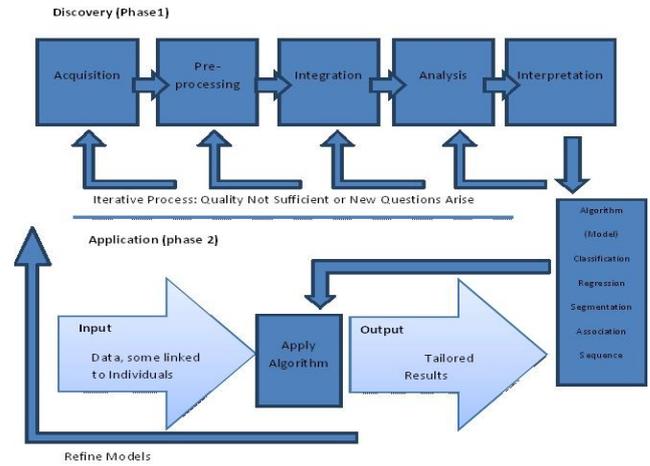


Figure3 Functional Architecture of Big Data Analytics Process [15]

Big Data Analytics In Health Care

The Big Data Analytics applications can extend the services that identifying disease status at the earliest period, getting new ideas of disease mechanism, observe the treatment methods and medicine.

Data mining techniques applied on EHRs, social media and clinical process, web related data for identifying the possible instructions in the hospitals.

In these days, smart phones play very important role in delivering the personal information to patients to participate them in physical treatment to improve their health in well manner. The patients are easily getting medical advices through mobile phones by doctors.

Big Data Tools Used In Disease Prediction

Now, many open source tools are provided for people by companies with free of cost. The tools are Hadoop, Storm, Casandra, HBase, Hive, Pig Latin, and R Programming [19]. These tools are used to analyze healthcare data and finding insights to take decisions by doctors.

The evaluation procedure and requirements are varied for different organizations. The cost and security can be decided based on some platform and tool which used in analytic process of healthcare with availability of data, visualization support, quality assurance and data handling capability. The Big Data analytical tools and platforms are described in the following table.

Table 1 Platforms & Tools for Big Data Analytics in Healthcare [15]

Big Data Analytical Tools	Description
HDFS	HDFS provides more space to store data for all clusters. The data divides as small chunks and transfers all chunks to various nodes in network.
MapReduce	MapReduce distributes sub units of data and collecting the results. After tasks are being executed, MapReduce can track to process every cluster system.

PIG and PIG Latin	Apache <i>Pig</i> is a high-level platform for creating programs that run on Apache Hadoop. Pig Latin is a type of programming language utilized for working in Apache Pig, which is a software resource for creating certain kinds of data analysis programs.
Hive	Apache <i>Hive</i> is a <i>data</i> warehouse system for <i>data</i> summarization, analysis and querying of large <i>data</i> systems in open source Hadoop platform. It is an Open Source <i>data</i> warehousing system. It is exclusively used to query and analyze <i>huge data</i> sets stored in the Hadoop storage.
Jaql	Jaql is a functional data processing and query language most commonly used for JSON query processing on BigData. Jaql perform the MapReduce task operations in parallel manner.
Zookeeper	Apache <i>Zookeeper</i> is a coordination service for distributed application that enables synchronization across a cluster. <i>Zookeeper</i> in Hadoop can be viewed as centralized repository where distributed applications can put <i>data</i> and get <i>data</i> out of it.
HBase	HBase is an open source NoSQL database system which cannot provide SQL approach and it is part of HDFS. It provides functionality for large data sets.
Cassandra	Apache Cassandra is a free and open-source, distributed, wide column store, NoSQL database management system designed to handle large amounts of data across many commodity servers, providing high availability with no single point of failure.
Oozie	Apache Oozie is a server-based workflow scheduling system to manage Hadoop jobs. Workflows in Oozie are defined as a collection of control flow and action nodes in a directed acyclic graph. Control flow nodes define the beginning and the end of a workflow as well as a mechanism to control the workflow execution path.
Avro	Apache Avro is a data serialization and remote procedure call framework. It provides both a serialization format to get persistent data and a wire format for providing communication between Hadoop nodes, as well as connecting client programs to the Hadoop services.
Mahout	Apache Mahout is a project of the Apache Software Foundation to produce free implementations of distributed or otherwise scalable machine learning algorithms focused primarily in the areas of collaborative filtering, clustering and classification.
Tableau	<i>Tableau Big Data Analytics</i> is an analysis and reporting tool from <i>Tableau</i> . With this system, users can experience platform variety and utilize popular frameworks such as Apache Hadoop, Spark and NoSQL to examine <i>data</i> .
R Programming	R is open source tool and it is a statistical analytical tool for analyzing data in the form of numerical analysis, graphical representation and result reporting. It provides more packages which is related to different categories of files that are used to store data in corresponding formats.

VI. RESULTS TABLE

The various Big Data Analytical techniques and algorithms produced the accuracy which is shown in the following table.

Year	Author	Big Data Analytical Technique/ Algorithm	Accuracy
2011	Dr. R. Geetha Ramani et al.	Random Tree	100%
2012	Dr. R. Geetha Ramani et al.	Random Tree Classification Algorithm	100%
2013	Peyman Mohamma di et al.	Correlation Coefficient M5Rules Algorithm	99.85% 99.67%
2013	Farhad Soleimania n Gharehcho pogh et al.	MATLAB neural network Radial Basis Function	93.22% 86.44%
2014	Dr. Hari Ganesh S et al.	Random tree	84%
2014	Chandrashe khar Azad et al.	Decision Tree ID3 Decision Stumps Based Model	85.08% 75.33% 83.55%
2014	Tawseef Ayoub Shaikh	Artificial Neural Network Decision trees Naïve Bayes	90.77% 80.51% 69.23%
2014	M. Hariharan et al.	Classification algorithm	100%
2016	Kamal Nayan Reddy Challa et al.	Boosted Logistic Regression area curve	97.16 % 98.9%
2016	N. Shamli et al.	Decision Tree Artificial Neural Network Support Vector Machine	89% 91% 90.71%
2017	Zhenao Cai et al.	Support Vector Machine	97.42%

2017	Musaed Alhusein et al.	Support Vector Machine	97.2%
2017	Sakina Imam, Prof. Sarwesh Site	PCA-KNN Classifier	96.92%

VII. CONCLUSION

In this paper, the various analytical methods and comparison of different algorithms results are described. The main intention is to verify the accuracy of prediction algorithms. In my literature study, various classification methods and data analytics were discussed which used for diagnosing the Parkinson's disease through symptoms. The survey was conducted from 2011 to 2018 that presents various methods, ideas, analytical and mining techniques. The results are also compared. In this study, most of researchers used data mining and machine learning techniques. This survey covers the overall research possible opportunities in Big Data domain. The future study aims to propose an efficient method to diagnose this type of neurological disorder by some symptoms at the early stage with better accuracy using different Big Data Analytical techniques like Hadoop, Hive, R Programming, MapReduce, PIG, Zookeeper, HBase, Cassandra, Mahout etc. The efficient methods will be played greater role in predicting the Parkinson's disease symptoms in the early stage for increasing the patient life period.

CONFLICT OF INTEREST

There is no any kind of conflict of interest in the publication of article.

ACKNOWLEDGMENT

I would like to thank Prof. R. Udaya Kumar for his valuable guidance, support and sparing valuable time for conducting this research work. I am also thankful for authors of journal papers which I have referred for studying of literature work successfully.

REFERENCES

- Zhennao Cai , Jianhua Gu, Caiyun Wen, Dong Zhao, Chunyu Huang, Hui Huang , Changfei Tong, Jun Li, and Huiling Chen, "An Intelligent Parkinson's Disease Diagnostic System Based on a Chaotic Bacterial Foraging Optimization Enhanced Fuzzy KNN Approach", Hindawi Computational and Mathematical Methods in Medicine Volume 2018, Article ID 2396952.
- Srishti Grover, Saloni Bhartia, Akshama, Abhilasha Yadav, Seeja K. R., "Predicting Severity Of Parkinson's Disease Using Deep Learning", International Conference on Computational Intelligence and Data Science (ICCIDS 2018) , 1877-0509.
- D.Karthiga, Dr.P.Sumitra,"A Survey of Predicting Parkinson's & Atypical Parkinson Disease in the Primordial Stage by using Classification techniques in Data Mining", International Journal of Advance Engineering and Research Development *Volume 5*, Issue 02, February -2018, 2348-4470.
- Muthumanickam S, Gayathri J, Eunice Daphne, "Parkinson's disease Detection And Classification Using Machine Learning And Deep Learning Algorithms–A Survey", International Journal of Engineering Science Invention (IJESI), Volume 7 Issue 5 Ver. 1 || May 2018 || PP 56-63, ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 672.
- Mahalakshmi Senthilarumugam Veilukandammal, Dr.Sree Nilakanta, Dr.Baskar Ganapathysubramanian, Dr.Vellareddy Anantharam, Dr.Anumantha Kanthasamy, Dr. Auriel A Willette, "Big Data and

- Parkinson's Disease: Exploration, Analyses, and Data Challenges". Proceedings of the 51st Hawaii International Conference on System Sciences 2018.
- R. Prashanth, Sumantra Dutta Roy, "Early Detection of Parkinson's Disease through Patient Questionnaire and Predictive Modelling", International Journal of Medical Informatics, 2018.
- Athithiyam G, Anandaraj K, "Brain Disease Prediction by Machine Learning over Bigdata from Healthcare", International Journal of Contemporary Research in Computer Science and Technology (IJCRCSST, Volume 4, Special Issue 3 (March '2018), ISSN: 2395-5325.
- Diellza Nagavci, Mentor Hamiti, Besnik Selimi, " Review of Prediction of Disease Trends using Big Data Analytics", International Journal of Advanced Computer Science and Applications, Vol. 9, No. 8, 2018.
- Athmaja S, Hanumanthappa M, Vasantha Kavitha, "A Survey of Machine Learning Algorithms For Big Data Analytics", International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), 2017.
- Sakina Imam, Prof. Sarwesh Site, "Parkinson's disease Diagnosis using Modified PCA-KNN Classifier", *International Journal of Electrical, Electronics and Computer Engineering* 6(2): 57-63(2017), ISSN No. (Online): 2277-2626.
- Sonu S R, Ravi Ranjan, "Prediction of Parkinson's Disease using Data Mining", International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS-2017), 978-1-5386-1887-5, 2017 IEEE.
- Jesper F. Havelund , Niels H. Heegaard, Nils J. K. Færgeman, Jan Bert Gramsbergen, "Biomarker Research in Parkinson's Disease Using Metabolite Profiling", *Metabolites*, 7, 42, 2017.
- Musaed Alhusein, "Monitoring Parkinson's Disease in Smart Cities", *IEEE. Translations*, Volume 5, 2017, 2169-3536.
- Zhennao Cai, Jianhua Gu, Hui-Ling Chen,"A New Hybrid Intelligent Framework for Predicting Parkinson's Disease", *IEEE. Translations*, VOLUME 5, 2017, 2169-3536.
- Siva Sankara Reddy Donthi Reddy, Udaya Kumar Ramanadham, " Big Data Analytics for Healthcare Organization, BDA Process, Benefits and Challenges of BDA: A Review", *Advances in Science, Technology and Engineering Systems Journal* Vol. 2, No. 4, 189-196 (2017).
- N. Shamli, B. Sathiyabhama, "Parkinson's Brain Disease Prediction Using Big Data Analytics", *IJ. Information Technology and Computer Science*, 2016, 6, 73-84.
- Kamal Nayan Reddy Challa, Venkata Sasank Pagolu, Ganapati Panda, "An Improved Approach for Prediction of Parkinson's Disease using Machine Learning Techniques", International conference on Signal Processing, Communication, Power and Embedded System (SCOPE5)-2016.
- I.D. Dinov, B. Heavner, M. Tang, et al. "Predictive Big Data Analytics: A Study of Parkinson's Disease Using Large, Complex, Heterogeneous, Incongruent, Multi- Source and Incomplete Observations", *PLOS* (2016).
- Arvind Kumar Tiwari, "Machine Learning Based Approaches for Prediction of Parkinson's Disease", *Machine Learning and Applications: An International Journal (MLAIJ)* Vol.3, No.2, June 2016.
- TawseefAyoub Shaikh, "A Prototype of Parkinson's and Primary Tumor Diseases Prediction Using Data Mining Techniques", *International Journal of Engineering Science Invention* April 2014, Vol: 3 Issue 4, pp. 23-28.
- Roger A. Barker, Caroline H. Williams-Gray, "Mild Cognitive Impairment and Parkinson's Disease", *Journal of Parkinson's Disease* 4 (2014) 651–656, ISSN 1877-7171.
- Chandrashekar Azad, Sanjay Jain, Vijay Kumar Jha, "Design and Analysis of Data Mining Based Prediction Model for Parkinson's disease", *International Journal of Computer Science Engineering (IJCSE)*, Vol. 3 No.03 May 2014, ISSN 2319-7323.
- Dr.Hari Ganesh S, Gracy Annamary S, "Comparative study of Data Mining Approaches for Parkinson's Diseases ", *International Journal of Advanced Research in Computer Engineering & Technology(IJARCET)*Volume 3 Issue 9, September 2014, ISSN: 2278 –1323.



24. Farhad Soleimanian Gharehchopogh, Peyman Mohammadi, “A Case Study of Parkinson’s disease Diagnosis using Artificial Neural Networks “,*International Journal of Computer Applications, Volume 73– No.19, July 2013, 0975 – 8887.*
25. Sanjivani Bhande, Dr. Ranjan Raut, “Parkinson Diagnosis using Neural Network: A Survey“,*International Journal of Innovative Research in Science, Engineering and Technology, Vol. 2, Issue 9, September 2013, ISSN: 2319-8753.*
26. Tarigoppula V.S Sriram, M. Venkateswara Rao, G V Satya Narayana , DSVGK Kaladhar, T Pandu Ranga Vital. “Intelligent Parkinson Disease Prediction Using Machine Learning Algorithms”, *IJEIT September 2013, Vol: 3, Issue 3.*
27. Peyman Mohammadi, Abdolreza Hatamlou, Mohammad Masdari, “A Comparative Study on Remote Tracking Of Parkinson’s Disease Progression Using Data Mining Methods“, *International Journal in Foundations of Computer Science & Technology (IJFCST), Vol. 3, No.6, November 2013.*
28. Dr.R.Geetha Ramani G. Sivagami, Shomona Gracia Jacob, “Feature Relevance Analysis and Classification of Parkinson Disease Tele-Monitoring Data Through Data Mining Techniques”,*International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 3, March 2012 ISSN: 2277 128X.*
29. Siva Sankara Reddy Donthi Reddy, Udaya Kumar Ramanadham, “Hadoop Framework for Big Data Analytics to Analyze Attacks on Business Data“, *International Journal of Innovative Technology and Exploring Engineering, Volume-9 Issue-2, December 2019, 3995-4000.*

AUTHORS PROFILE



Siva Sankara Reddy Donthi Reddy received his Master of Engineering degree in CSE from Sathyabama Institute of Technology and Science, Deemed to be University, Chennai in 2007 and he is research scholar at Bharath Institute of Higher Education and Research, Department of Computer Science and Engineering, Chennai, India. His research interests include Big Data for Healthcare Domain, Computer Networks, and

Information Security. He is working as Associate Professor in Annamacharya Institute of Technology and Sciences, Rajampet, AP.



Dr. R. Udaya Kumar is Professor at Department of Information Technology, Bharath Institute of Higher Education and Research, Chennai, India. He received PhD from BIHER, Chennai. He has published more than 600 papers and he is Supervisor for Research Scholars at BIHER. His research interests lie in the areas of Cloud Computing, Data Mining, Big Data Analytics and

Information Security.