

AN EFFECTIVE PARKINSON'S DISEASE PREDICTION USING LOGISTIC DECISION REGRESSION AND MACHINE LEARNING WITH BIG DATA

Kranthi Kumar Singamaneni¹, Dr.G.Puthilibai², D.Saravanan³, Dr.P,Sagaya Aurelia⁴,
P Gopala Krishna⁵, Dr.D.StalinDavid⁶

¹Associate Professor, Department of CSE, Gokaraju Rangaraju Institute of Engineering & Technology,
Kukatpally, Hyderabad,

²Professor, Department of Chemistry, Sri Sairam Engineering College, Chennai,

³Associate Professor, Department of CSE, IFET College of Engineering, Villupuram,

⁴Assistant Professor, Department of CSE, CHRIST (Deemed to be University), Bangalore,

⁵Associate Professor, Department of IT, Gokaraju Rangaraju Institute of Engineering & Technology,
Kukatpally, Hyderabad,

⁶Assistant Professor, Department of CSE, IFET College of Engineering, Villupuram,

kkranthicse@gmail.com, vgputhili@gmail.com, saranmds@gmail.com,
sagaya.aurelia@christuniversity.in, gopalakrishna@griet.ac.in, sdstalindavid707@gmail.com

ABSTRACT

Background: Medical data is conducive to early identification of diseases, patient treatment, and community service. Parkinson's disease prediction by Machine Learning (ML) in large data and reliable study of biomedical and healthcare community develop big data; medical data is conducive to early detection of diseases, patient care, and community service. The machine learning algorithm is being used to successfully forecast the prevalence of chronic illness populations.

Proposed Methodology: Parkinson's disease is a serious neurodegenerative disease that affects people when they become older (mostly past the age of 50). It is the most serious and harmful of the non-curable neurodegenerative diseases. Parkinson's disease is challenging to diagnose at an early level and the origin of subtle early signs is difficult to identify. Due to the heavy responsibility of the condition on the Parkinson disease patient, a clinical care scheme has been developed. To classify Parkinson's disease and overcome this complicated challenge, the suggested Machine Learning (ML) induced Logistic Decision Regression (LDR) algorithm is used. Early diagnosis of Parkinson's disease will contribute to improved care and disease control, thereby enhancing the quality of life of patients. To create such successful decision support, an automatic prediction system focused on machine learning was developed and presented.

Result and Findings: This data collection would be used to classify possible biomarkers of Parkinson's disease using Machine Learning (ML) and Big Data (BD) technologies. The organization's disease prediction technology, which is focused on machine learning and large data, enhances human wellbeing while further promoting the big data industry of disease prediction. In comparison to other current approaches, the simulation findings indicate a strong reliability.

Keywords: Parkinson's disease, Machine Learning (ML), Big Data (BD), Logistic Decision Regression (LDR), Prediction, Classify.

I. INTRODUCTION

Parkinson's disease is a neurodegenerative disorder that affects millions of people around the globe. Muscle fatigue, tremor (upper and lower arms, as well as jaw vibration), voice disturbances, deadpan, slow motion, apathy, postural dysfunction (depression and mood changes), repetitive gestures, dementia (loss of memory), what not to do, sleep disorders, and thinking are also typical symptoms of Parkinson's disease. The batch size, all the properties of velocity, veracity, length, meaning, and variety are all present in Parkinson's disease data, which

is referred to as Big Data (BD). These five have been identified in greater depth in the context of Parkinson's data in the following manner. Velocity is provided by the pace of the data contained within, as well as the processing speed of the storage indication, which aids in real-time decision making and modern real-time prediction processing. Veracity is a software company that specialises in data and precision analysis. A key aspect of consistency is issues and knowledge concerning the reliability of data quality. The information on Parkinson's disease is heterogeneous, multi-source, unreliable, contradictory, and scarce. In disorders like Parkinson's disease, medical and biological predictions are extremely significant. Stable people with Parkinson's disease treatment judgments boundary evidence from their alienation and Parkinson's disease for the diagnosis efficacy and computational performance class boundaries mechanisms

Parkinson's disease is a chronic brain condition marked by non-motor symptoms as well as two motor symptoms (motor). The personal introduction shown in the conditions, in addition to all of the typical symptoms, is something that everyone will witness. Stiffness or rigidity may occur in people with Parkinson's disease. People with Parkinson's disease may also become frozen or unable to shift for a brief amount of time. Parkinson's disease is a neurodegenerative disease characterised by the destruction of dopamine-producing cells. Parkinson's syndrome may be isolated from other disorders that have identical clinical manifestations, but accurate testing is difficult to come by. The scientific practise of patient background and analysis is used to make a diagnosis. Movement conditions, commonly involving neurological illnesses such as depression and schizophrenia, are the most common causes, although they may often lead to the progression of other diseases. It's possible that you'll lose your autonomy and be in agony as a result. People with a deteriorating standard of life and major obstacles and challenges were impacted as the disease advanced. There is a chance that family members and caregivers may have an indirect impact.

The physical activity data obtained on the wearable smart terminal is also pre-processed on the computer, after which it is fixed and stored by the wireless relay base station to the online tracking management network, and all data is summarised as decided on the cloud online management platform for web pages and smartphone apps, called visual. Both physical activity details, on the other hand, must be in a particular data format based on Hadoop-bigdata, which implements the format of data files from the online monitoring portal, utilising high data analysis and corresponding data analysis technologies, with high data integration advantages, and highly effective data analysis.

To maintain the integrity of the data files, all of them will be copied at the same time. After that, I'll run the abbreviated research software for the particular diagram in order to do the parallel analysis efficiently and in accordance with the calculation's analysis specifications. Following the completion of the calculation, different analyses are performed, and the calculation's result is output, with many of the resulting files being generated and extracted. Big Data (BD) and Machine Learning (ML). Following that, the company will notify the online administration in compliance with the agreed-upon procedure and apply the site's sports data processing results. Finally, partner members can access related analytics data using the online management tool for spoken smartphone apps.

II. RELATED WORK

An organisation, which is generally described in depth as a structured collection of processes, protocols, and the concept of designing procedures to carry out specific activities or solve specific problems, successfully applies to public health from a variety of mechanisms, including mechanical systems. The data representation framework relevant to continuous surveillance, review, and device control, management, and strategic preparation is referred to as System Health Monitoring and Management (SHMM) [1]. A dynamic engineering system's requirement for brain health and quality pre-sales testing has become critical. The integration of big data, intelligent flight coding, and the introduction of popular classification algorithms, the secret mining of historical data knowledge, and the capacity to achieve quality forecasts of brain health are all dependent on intelligent decision-making [2]. Stuff for measuring brain health content are developed on an internet-based architecture, and classification approaches are used to accomplish real-time data collection and analysis. Obstructive Sleep Apnoea (OSA) is a serious sleep condition that has a strong detrimental effect on one's quality of life. Reduced mood changes are a symptom of depression, and certain OSAs are caused by behavioural and attitude problems [3]. As a result, there is a pressing need for treatment solutions for this disease to be monitored in real time. There are many OSA detection systems available. The pervasive usage of medical / health-related data is the at an exponential pace, thanks to the growth of maturity and the Internet, cloud infrastructure, and technology across the Internet, medical / health information technology. Simultaneously, the usage of mainstream and genomic technologies, as well as the exponential

growth of mobile medical technology and mobile medical technology, have resulted in big data programmes in the medical and healthcare fields [4].

Wireless sensor networks for essential applications such as self-organizing autonomous cities, digital medicine, precision farming, and industrial control systems are often developed and equipped with resource-controlled sensing equipment. Many of these instruments are conscious of their surroundings and the data they collect, and they submit it to the data collection flow server. The real-time monitoring and decision-making capabilities of Data Stream Manager (DSM) gathers server data flow (commonly referred to as big data) for these critical applications, which can be exploited during malicious attack traffic or to destroy data [5]. As a consequence, ensuring the reliability of the gathered data such that any reports are reliable is a difficult challenge in these applications.

[illegible]

function error feature identification and operational error clever detection action feature to obtain a system instability error prediction approach and function error feature detection and operational error clever detection action feature to obtain a hierarchical correlation between big data and function error [14]. Diabetes has become more common in the world. To meet this aim, diabetics must be monitored on a regular basis and must be actively engaged with their health care. Mobile Health (MH) is a growing trend in information and telecommunications technology that can help chronic patients in a smart world. It will review the existing status of MH technology in order to overcome its shortcomings [15]. Current study papers are evaluated in terms of validity and degree of applicability by patients and medical care professionals in the MH literature. The Patient will remotely track patients using Internet of Things (IoT) big data processing sensors. The literature, on the other hand, supports experiments that acknowledge the relatively large number of virtual instruments and the common use of a sleep predictor. A simulation-based smart bed load sensor is provided by the Design sluggish motion agent-based simulation platform [16]. The arrangement specifies the sleep predictor identification indicators and enables the outcomes of the dull taken indicator to be compared. The effects of the graph star map, chart evolution, and the final visual output of the bed sensor state are presented in this novel befit dummy, which helps users to discover the results of the graph star chart, chart evolution, and the final visual performance of the bed sensor state. While health networks for Parkinson's disease (PD) exist and have been discussed in the literature, the majority of them lack the capacity to analyse vast volumes of data generated and gathered from medical exams and organised in a pre-defined manner. Centered on the study of massive numbers, proposes a modern model of health network [17]. The planned architecture's key goal is to assist physicians in conducting purposeful assessments of common Parkinson's disease motor issues and improvements. Machine health control devices are commonly used in the modern sector to accomplish proactive maintenance, with applications such as fault monitoring, unemployment reduction, and property safety. Since predicting the appearance of certain faults (up to and including forecasting potential operating conditions and remaining service life [18], data-based health management has produced significant results in the big machine data age. The data source sensor is comprised of numerous effective trapezoids of digital data expression. The conventional approach is labor-intensive since it typically necessitates specialised knowledge and relies on manual labour. Wearable Health Monitoring Systems (WHMS) that are disabled will allow for continuous monitoring of possible patient physiological parameters from either place. Furthermore, utilising multiple biosensors, such a device will be unable to achieve a comprehensive measure of the user's wellbeing. Java-based simulation platform [19] that is multi-sensor supported or equivalent to the petri net model. This work has been extended to accommodate for synchronisation issues as well as time-dependent variables.

III. MATERIALS AND METHODS

Parkinson's disease is caused by the death of dopamine-producing neurons in the brain. Early signs of infections may be tough to detect since they are mild at first. Owing to the high pressure placed on patients, there is a health-care system that delays detection. The diagnosis of Parkinson's disease using Machine Learning (ML) and Big Data (BD) methods leads to a greater explanation of the disease decade.

Figure 1 shows the pre-processing of a Parkinson's disease dataset, which eliminates noise and unnecessary data while using feature selection to find the right attributes and eventually classifying the Parkinson's disease prediction result.

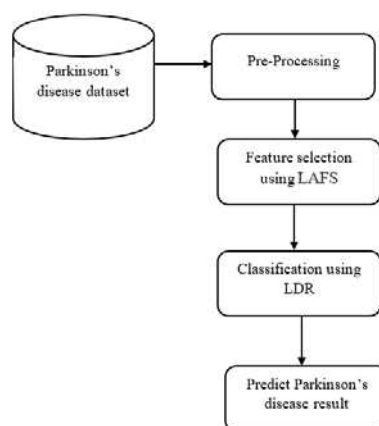


Figure 1: Block diagram

3.1 Pre-processing

It is a typical data pre-processing approach that is defined by the use of a data processor. Filtering, normalisation, and target recognition are also part of the pre-processing procedure. The performance at this stage is a set of important parts and components. Light varies often, and there is less comparison, according to the data. As a result, the platform must process the data immediately. To protect random lighting and noise medical records, improvements are needed. It's a technique for filtering out unnecessary information from the background noise. Clinical evidence were often altered prior to improvement. This is a phase in the data pre-processing method that will increase performance.

3.2 Feature selection using Least Absolute Shrinkage and Feature Selection (LAFS) algorithm

In machine learning, feature selection refers to the process of choosing the most appropriate features for the model disease results. There is a propensity to increase the tempo, precision, or both training by limiting the number of features you want to use (not only the data that hasn't modified the model supplied), but there is also a tendency to improve the speed, accuracy, or both training by limiting the number of features you want to use (not just the data that hasn't changed the model supplied). Using the LAFS algorithm, it picks shivering, voice problems, postural instability, and movement problems.

Algorithm Steps

Input: Parkinson's disease dataset

Output: Feature selection disease Dataset

Start

Step 1: Initialize the disease dataset

Step 2. Read the dataset

Step 3. Split the Dataset

Step 4. For (Calculate each intensity data)

Feature selection disease dataset

End of

Step 5. Select best features (shivering, speech problem, Movements)

Exit

Any Parkinson's disease data is gathered during the phases of this algorithm, making the data a very delicate operation. The LAFS algorithm is used to select the Parkinson's disease functions.

3.3 Classification using Logistic Decision Regression

The large-scale estimation issue of Parkinson's disease is found in the general public health. Machine learning-based approaches were used to differentiate between stable individuals and those with Parkinson's disease. Via a network of space exploration lectures, LDR is able to achieve its goals. It is reliant on a number of inputs, which are usually used to estimate the estimated function.

Algorithm steps

Input: Feature selection disease dataset

Output: Classification of Parkinson's disease

Start

Step 1: Import the feature selection disease dataset

Step 2: Read the feature selection disease dataset

Step 3: Remove unnecessary data

Step 3: Calculate the classification using LDR

Stop

IV. RESULT AND DISCUSSION

The outcomes and consequences of the new development process would be evaluated using the Anaconda tool and a medical dataset for Parkinson's disease. The results evaluation During the re-run process of test data,

sensitivity, accuracy, and classification efficiency are determined. Support Vector Machine (SVM) and Convolutional Neural Network (CNN) are two known approaches that are compared to the proposed Logistic Decision Regression (LDR) algorithm (CNN).

Table 1: Simulation parameters

Parameters	Values used
Tool	Anaconda
Language	Python
Input data	Parkinson's disease data set
Training dataset	120
Testing dataset	30

Table 1 shows the simulation parameters for proposed implementation process using python language.

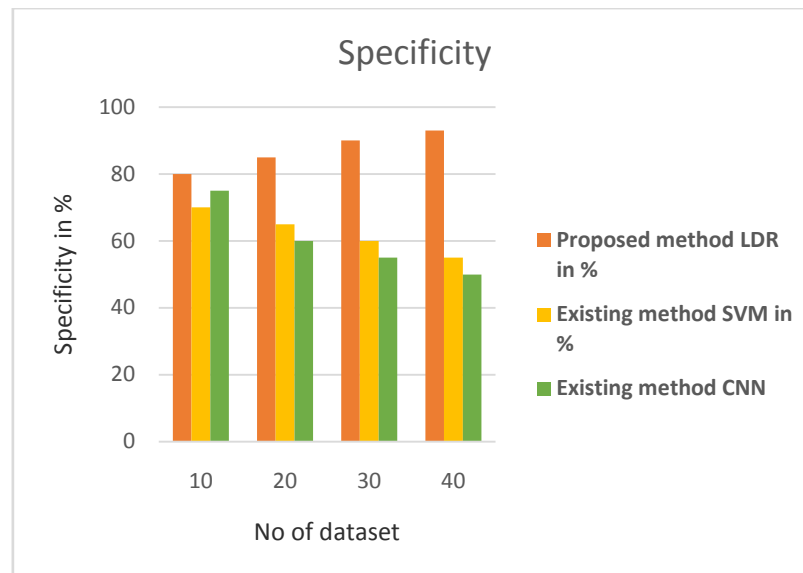


Figure 2: Analysis of specificity

The study of accuracy in percentages is depicted in Figure 2. The sensitivity to Parkinson's disease tests refers to the number of times they are correctly identified. The suggested sensitivity result from Logistic Decision Regression (LDR) is 93 percent. The results of the current approaches are 55 percent for the Support Vector Machine (SVM) and 50 percent for the Convolutional Neural Network (CNN).

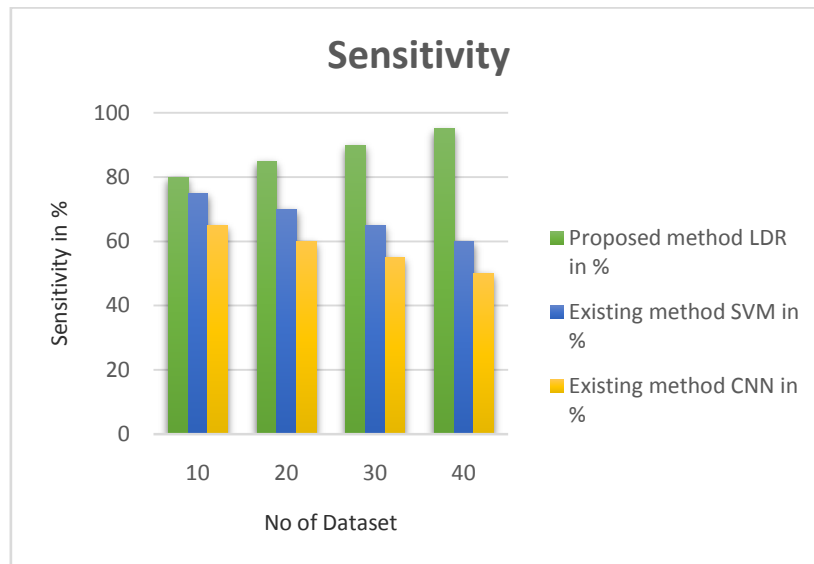


Figure 3: Analysis of sensitivity

The sensitivity study in percentage as seen in Figure 3. As a proportion of correctly negative for Parkinson's disease, specificity tests have been established. The specificity result of the suggested Logistic Decision Regression (LDR) is 95%. Support Vector Machine (SVM) results are 60% and Convolutional Neural Network (CNN) results are 50%, respectively, with the latest processes.

Table 2: Analysis of classification performance

No of dataset	LDR in %	SVM in %	CNN in %
14	81	76	72
25	87	61	66
33	92	66	64
42	98	58	49

Table 2 shows the analysis of classification performance the proposed algorithm classify the Parkinson disease and it provide high performance compared to other methods.

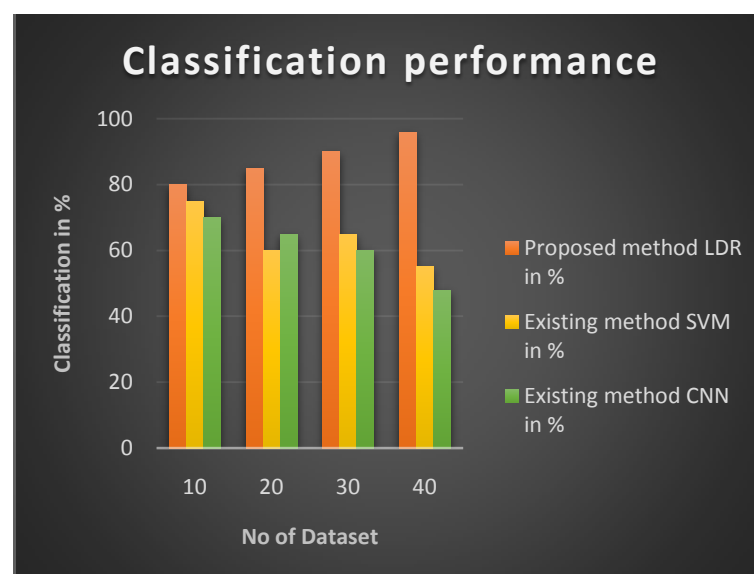


Figure 4: Classification performance

Figure 4 depicts the classification success rate as a percentage. The suggested specificity result for Logistic Decision Regression (LDR) is 96 percent. The results of the latest approaches are 55 percent for Support Vector Machines (SVM) and 48 percent for Convolutional Neural Networks (CNN).

V. CONCLUSION

People may not have a clear way of determining whether or not they are suffering from Parkinson's disease through a particular examination, such as a blood test or an ECG. Diagnosis of Parkinson's disease requires a specific test, such as a blood test or an ECG. To classify Parkinson's disease and overcome this complicated challenge, the suggested Machine Learning (ML) induced Logistic Decision Regression (LDR) algorithm is used. Early diagnosis of Parkinson's disease will contribute to improved care and disease control, thereby enhancing the quality of life of patients. The organization's disease prediction technology, which is focused on machine learning and large data, enhances human wellbeing while further promoting the big data industry of disease prediction. The proposed LDR algorithm has a sensitivity of 95 percent, a precision of 93 percent, and a classification efficiency of 97 percent.

REFERENCE

1. Kwok Leung Tsui ; Yang Zhao, "Big Data Opportunities: System Health Monitoring and Management", IEEE Access (Volume: 7) 2019.
2. Yuan Huang ; Qiang Zhao, "Air Quality Forecast Monitoring and Its Impact on Brain Health Based on Big Data and the Internet of Things", IEEE Access (Volume: 6) 2018.
3. Diana C. Yacchirema ; David Sarabia-JáCome, "A Smart System for Sleep Monitoring by Integrating IoT With Big Data Analytics", IEEE Access (Volume: 6) 2018.
4. Jingyi Zhang ; Tong Zhao, "Analysis Method of Motion Information Driven by Medical Big Data", IEEE Access (Volume: 7), 2019.
5. Deepak Puthal ; Xindong Wu, "SEEN: A Selective Encryption Method to Ensure Confidentiality for Big Sensing Data Streams", IEEE Transactions on Big Data (Volume: 5 , Issue: 3 , Sept. 1 2019).
6. Yin Zhang ; Meikang Qiu, "Health-CPS: Healthcare Cyber-Physical System Assisted by Cloud and Big Data", IEEE Systems Journal (Volume: 11 , Issue: 1 , March 2017).
7. Abdulsalam Yassine ; Shailendra Singh, "Mining Human Activity Patterns From Smart Home Big Data for Health Care Applications", IEEE Access (Volume: 5), 2017.
8. Antonino Galletta ; Lorenzo Carnevale, "An Innovative Methodology for Big Data Visualization for Telemedicine", IEEE Transactions on Industrial Informatics (Volume: 15 , Issue: 1 , Jan. 2019).
9. Abdur Rahim Mohammad Forkan ; Ibrahim Khalil, "BDCaM: Big Data for Context-Aware Monitoring—A Personalized Knowledge Discovery Framework for Assisted Healthcare", IEEE Transactions on Cloud Computing (Volume: 5 , Issue: 4 , Oct.-Dec. 1 2017).
10. Jong Wook Kim ; Jong Hyun Lim, "Collecting Health Lifelog Data From Smartwatch Users in a Privacy-Preserving Manner", IEEE Transactions on Consumer Electronics (Volume: 65 , Issue: 3 , Aug. 2019).
11. Ying Yang, "Medical Multimedia Big Data Analysis Modeling Based on DBN Algorithm", IEEE Access (Volume: 8) 2020.
12. Tianshu Wu ; Shuyi Chen, "Intelligent fault diagnosis system based on big data", The Journal of Engineering (Volume: 2019 , Issue: 23 , 12 2019).
13. Shaker El-Sappagh ; Farman Ali, "Mobile Health Technologies for Diabetes Mellitus: Current State and Future Challenges", IEEE Access (Volume: 7), 2018.
14. Iván García-Magariño ; Raquel Lacuesta, "Agent-Based Simulation of Smart Beds With Internet-of-Things for Exploring Big Data Analytics", IEEE Access (Volume: 6), 2017.
15. Abdulsalam Yassine ; Shailendra Singh "Mining Human Activity Patterns From Smart Home Big Data for Health Care Applications", IEEE Access (Volume: 5), 2017.
16. Leonarda Carnimeo ; Gianpaolo Francesco Trotta, "Proposal of a health care network based on big data analytics for PDs", The Journal of Engineering (Volume: 2019 , Issue: 6 , 6 2019).
17. Rui Zhao ; Dongzhe Wang, "Machine Health Monitoring Using Local Feature-Based Gated Recurrent Unit Networks", IEEE Transactions on Industrial Electronics (Volume: 65 , Issue: 2 , Feb. 2018).
18. Alexandros Pantelopoulou, "SPN-model based simulation of a wearable health monitoring system" IEEE Access, 2017.
19. Mary M. Rodgers ; Vinay M. Pai, "Recent Advances in Wearable Sensors for Health Monitoring", IEEE Sensors Journal (Volume: 15 , Issue: 6 , June 2017).
20. In cheol Jeong ; David Bychkov, "Wearable Devices for Precision Medicine and Health State Monitoring", IEEE Transactions on Biomedical Engineering (Volume: 66 , Issue: 5 , May 2019).
21. Stalin David D, Saravanan D, "Enhanced Glaucoma Detection Using Ensemble based CNN and Spatially Based Ellipse Fitting Curve Model", Solid State Technology, Volume 63, Issue 6, PP.3581-3598.
22. Stalin David D, Saravanan M, Jayachandran A, "Deep Convolutional Neural Network based Early Diagnosis of multi class brain tumour classification", Solid State Technology, Volume 63, Issue 6, PP.3599-3623.
23. R.Parthiban, Dr.K.Santhosh Kumar, Dr.R.Sathya, D.Saravanan, "A Secure Data Transmission And Effective Heart Disease Monitoring Scheme Using Mecc And Dlmnn In The Cloud With The Help Of Iot", International Journal of Grid and Distributed Computing, ISSN: 2005 – 4262, Vol. 13, No. 2, (2020), pp. 834 – 856.
24. R.Bhavya, G.I.Archanaa, D.Karthika, D.Saravanan, "Reflex Recognition of Tb Via Shade Duplicate Separation Built on Geometric Routine", International Journal of Pure and Applied Mathematics 119 (14), 831-836.
25. D Saravanan, R Bhavya, GI Archanaa, D Karthika, R Subban, "Research on Detection of Mycobacterium Tuberculosis from Microscopic Sputum Smear Images Using Image Segmentation", 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC).
26. D Saravanan, R Parthiban, "Automatic Detection of Tuberculosis Using Color Image Segmentation and Statistical Methods", International Journal of Advance Research in Science and Engineering, Volume 6, Issue 10.
27. D. Stalin David, 2019, "Parasagittal Meningioma Brain Tumor Classification System based on MRI Images and Multi Phase level set Formulation", Biomedical and Pharmacology Journal, Vol.12, issue 2, pp.939-946.
28. D. S. David and A. Jeyachandran, "A comprehensive survey of security mechanisms in healthcare applications," 2016 International Conference on Communication and Electronics Systems (ICCES), Coimbatore, 2016, pp. 1-6, doi: 10.1109/CESYS.2016.7889823.

29. D Stalin David, A Jayachandran, 2018, Robust Classification of Brain Tumor in MRI Images using Salient Structure Descriptor and RBF Kernel-SVM, TAGA Journal of Graphic Technology, Volume 14, Issue 64, pp.718-737.
30. D Stalin David, 2016, Robust Middleware based Framework for the Classification of Cardiac Arrhythmia Diseases by Analyzing Big Data, International Journal on Recent Researches In Science, Engineering & Technology, 2018, Volume 4, Issue 9, pp.118-127.
31. Stalin David D, Saravanan D, 2020, 'Multi-perspective DOS Attack Detection Framework for Reliable Data Transmission in Wireless Sensor Networks based on Trust', International Journal of Future Generation Communication and Networking , Volume 13, Issue 4, PP.1522–1539.
32. Dr. D. Stalin David, Mr. D. Saravanan, "Certain Investigation On Iot Therapeutic Image Recognition And Rivaroxabanpreclude Thrombosis In Patients", 2021, pg.no:51-66, ISBN: 978-81-948555-1-4.
33. D Saravanan, R Parthiban," Automatic Detection of Tuberculosis Using Color Image Segmentation and Statistical Methods", International Journal of Advance Research in Science and Engineering, Volume 6, Issue 10.
34. D.Saravanan, S.Rajasekaran, Dr. D.Stalin David, P.Hemalatha, Dr.U.Palani. (2021). Detection of Sickle Cell Anemia from Microscopic Blood Images Using Different Local Adaptive Thresholding Techniques. Annals of the Romanian Society for Cell Biology, 6549 –. Retrieved from <http://annalsofscb.ro/index.php/journal/article/view/3254>.
35. Dr.BrahmadesamViswanathan Krishna, Dr.G.Amuthavalli, Dr.D.StalinDavid, E. FantinIrudaya Raj, D.Saravanan. (2021). Certain Investigation of SARS-COVID-2-Induced Kawasaki-Like Disease in Indian Youngsters. Annals of the Romanian Society for Cell Biology, 1167–1182. Retrieved from <http://annalsofscb.ro/index.php/journal/article/view/4469>.