VISVESVARAYA TECHNOLOGICAL UNIVERSITY

Jnana Sangama, Santhibastawad Road, Machhe

Belagavi - 590018, Karnataka, India

PROJECT WORK PHASE- 1 (18CSP77) REPORT
ON

"Anomaly detection in Parkinson's disease using unsupervised machine learning Approach"

Submitted in the partial fulfillment of the requirements for the award of the degree of

BACHELOR OF ENGINEERING
IN

INFORMATION SCIENCE AND ENGINEERING
For the Academic Year 2021-2022
Submitted by

Abhay PJ	(1JS18IS001)
Hemanth Kumar A	(1JS19IS406)
Jeevan KV	(1JS18IS039)
Manoj GH	(1JS18IS048)

Under the Guidance of
Dr Malini M Patil
Associate Professor, Department of Information Science and Engineering

2021-2022

DEPARTMENT OF INFORMATION SCIENCE AND ENGINEERING
JSS ACADEMY OF TECHNICAL EDUCATION
JSS Campus, Dr.Vishnuvardhan Road, Bengaluru-560060
JSS MAHAVIDYAPEETHA, MYSURU

JSS ACADEMY OF TECHNICAL EDUCATION

JSS Campus, Dr.Vishnuvardhan Road, Bengaluru-560060

DEPARTMENT OF INFORMATION SCIENCE & ENGINEERING

CERTIFICATE

This is to certify that Project Work Phase -1(18CSP77) Report entitled "**Anomaly detection in Parkinson's disease using unsupervised machine learning Approach**" is a bonafide work carried out by Abhay PJ [1JS18IS001], Hemanth Kumar A [1JS19IS406], Jeevan KV [1JS18IS039], Manoj GH [1JS18IS048] in partial fulfillment for the award of degree of Bachelor of Engineering in Information Science and Engineering of Visvesvaraya Technological University Belagavi during the year 2021- 2022.

Signature of the Guide

Dr. Malini M Patil

Associate Professor
Dept. of ISE
JSSATE, Bengaluru

Signature of the HOD

Dr. Rekha P M

Associate Professor & Head
Dept. of ISE
JSSATE, Bengaluru

ACKNOWLEDGEMENT

The satisfaction and euphoria that accompany the successful completion of any task would be incomplete without the mention of the people who made it possible. So with gratitude, we acknowledge all those whose guidance and encouragement crowned my effort with success.

First and foremost we would like to express our heartfelt deep sense of gratitude to **His Holiness. Jagadguru Sri Shivarathri Deshikendra Mahaswamiji**, whose divine blessings have motivated and helped us in completing this project work.

Dr. Mrityunjaya V Latte, Principal, JSSATE, Bangalore for providing an opportunity to carry out the Project Work Phase – 1 (18CSP77) as a part of our curriculum in the partial fulfillment of the degree course.

We express our sincere gratitude for our beloved Head of the department, Dr**. Rekha P M**, for her co-operation and encouragement at all the moments of our approach.

It is our pleasant duty to place on record our deepest sense of gratitude to our respected guide **Dr. Malini M Patil, Associate Professor,** for the constant encouragement, valuable and timely guidance in every possible way in completing this work.

We are thankful to the Project Coordinators **Dr. Nagamani N P,** Asst. Professor and **Mrs. Sahana V** Asst. Professor, for their continuous co-operation and support.

We would like to thank all **ISE department teachers** and **non teaching staff** for providing us with their valuable guidance and for being there at all stages of our work.

Abhay PJ (1JS18IS001)
Hemanth Kumar A (1JS19IS406)
Jeevan KV (1JS18IS039)
Manoj GH (1JS18IS048)

TABLE OF CONTENTS

ABSTRACT:

This Project focuses on anomaly detection techniques for Parkinson’s disease, the insight generated by the algorithm is useful for doctors to diagnose the patient efficiently. It can help the patient with symptoms to diagnose whether he/she is positive or negative of Parkinson’s disease. Parkinson’s Outlier analysis is very useful in real life as it is very difficult to identify specific symptoms as every individual goes through a different type of symptoms and by going through all undesired symptoms we can get a brief knowledge of the disease.

The Project uses Google collab as the AI/ML platform and Tableau as the visualization tool.

The Front end of the project is developed using Flask/Django with AWS/Azure as the service provider. The possibility of getting an accurate prediction is low therefore we study the undesirable symptoms leading to Parkinson’s disease by this manner we get a bigger picture of Parkinson’s symptoms.The main intention of this project is to increase the chances of predicting the parkinson's disease.

INTRODUCTION:

What is artificial intelligence?

Artificial intelligence means training a computer in such a way that it attains intelligence through experience and prior knowledge which helps the machine to have the intelligence like the human mind. Many works have been done in the field of AI from 2000-2010,In the year of 2004 John McCarthy stated AI is used to have human intelligence but not their wisdom.As biological feelings and methods are not understood by AI.

Application of AI?

AI is used in many sectors like space exploration, social media analysis, gaming industry, agriculture and so on. It has become important in modern times as it can solve difficult problems appropriately in many industries, such as health care, entertainment, finance, education, etc.AI is used so the work is completed with intelligent decisions.

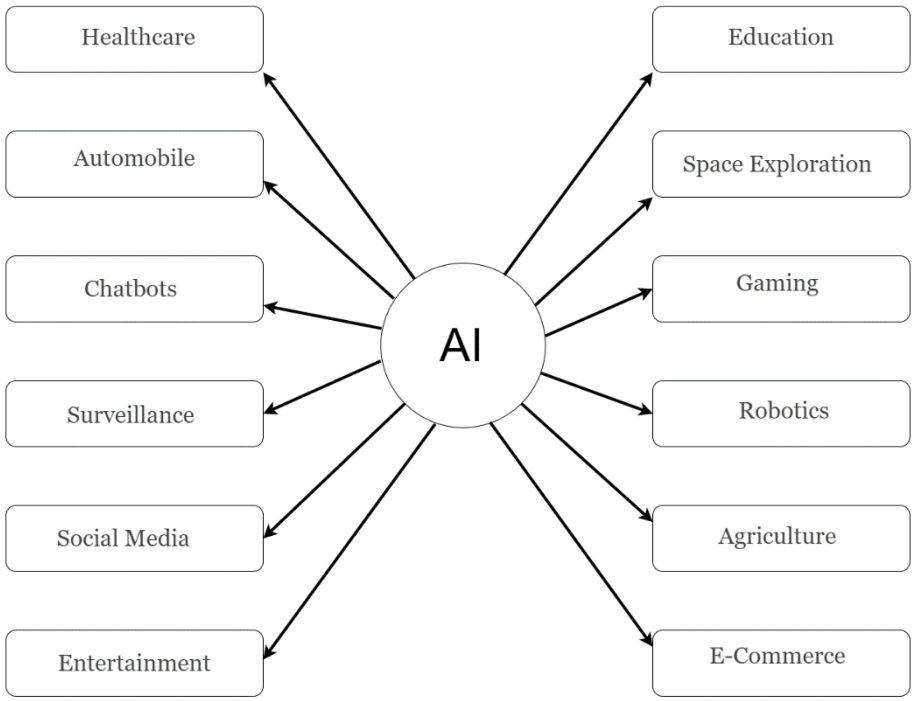

Figure 1. Application of Artificial Intelligence

What is Machine learning?

ML consists of many data structures and algorithms which are used so that the problem can be solved faster and accurately. ML is a part of AI which is used as a part of the thinking and logical sector of the model.

Data science is a field related to data and data analysis so it requires data structures and algorithms to handle and manipulate the model so ML plays a significant role.Likewise, many formula and prediction techniques are used to discover the knowledge or to predict the future instinct with respect to the input condition . This information later conducts decision-making within applications and businesses, which has a positive impact on key growth metrics.As the size is data will keep on increasing through time or through the number of the user increases so the demand for machine learning and data science will increase.

Methods of Machine learning

Machine learning classifiers fall into three primary categories.

- **Supervised machine learning**

  Supervised machine learning is ML model in which their is an interfere of the human algorithm which helps the machine to understand the data and problem more efficiently which help to predict the future more accuracy, the label is used to train the model with data classification with can give a positive and negative feedback of a condition with helps the supervised machine learning model to understand the data model more easily.

- **Unsupervised machine learning**

Unsupervised machine learning is a method in which the model tries to figure out the surrounding and the condition in which it is fitted.It tries out all the possible ways to reach the desired result by constructing its own model to solve the problem without any outworld intelligence.

Semi-supervised learning

This learning technique consists of both learning techniques which means it uses both unlabeled and labeled classifiers to analyse the situation. It uses its prior knowledge to understand the situation and tries to figure out the way to reach the desired result.

Reinforcement Learning:

RL is a machine learning model in which a model is supervised and interacts with the environment.Rl learns by understanding its surroundings and trying to adapt to it .

The RL model is a recursive model in which there is continuous learning .Basically how this RL model work is it takes a note of all the positive and negative output, it remember for which

condition of input we get positive and negative feedback and in future it will try to avoid negative output.

Data Mining

Data mining is basically extracting big data from the real world and try to analyse the complexity of the data, its main objective is to fetch insight or meaning out of this mass collection of data. This field consists of computer science concepts, mathematics and basic logics which helps in dealing with huge complexity.

Data mining is used to give a meaningful knowledge which is given by going through the data step by step. In addition to the non-critical analysis step, it also includes data retention features and data management features, data pre-processing, model and speculative thinking, interesting metrics, complex considerations, post-processing analysis of acquired properties, visibility, and online reviews. .

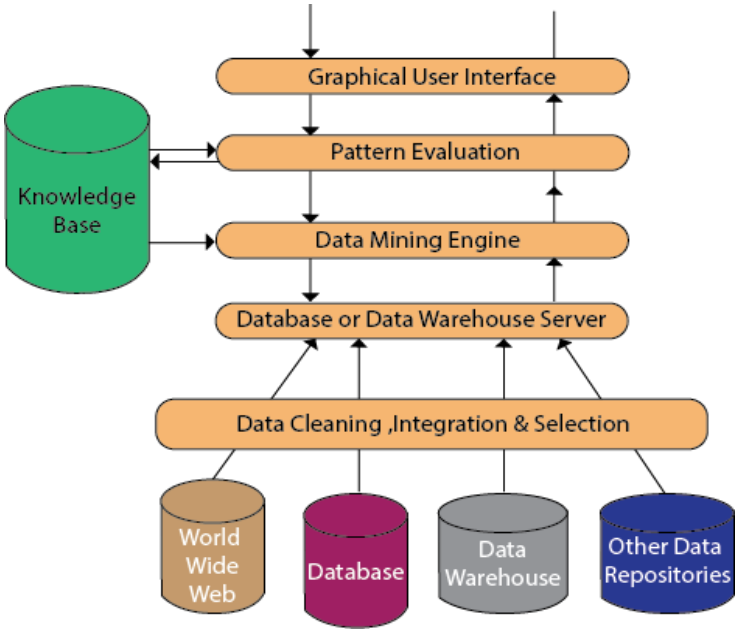

Figure 2. Data mining process

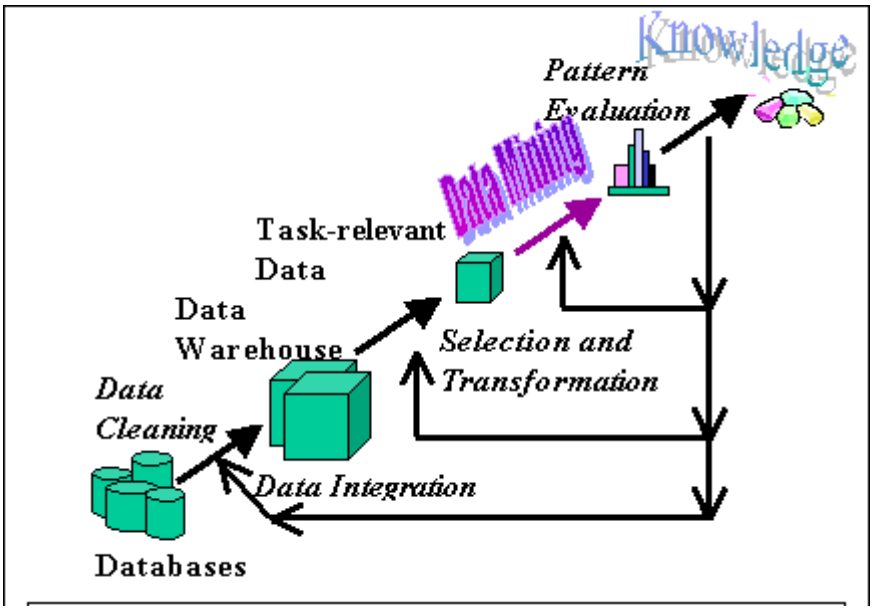

Figure 3. Data mining as core in knowledge

Data mining techniques

The data mining technique consist of algorithm,mathematical model,data structure which helps in understanding the data and helps in gaining information.So we can basically classify,understand the pattern,get knowledge on the pattern of occurrence of a state and many more.

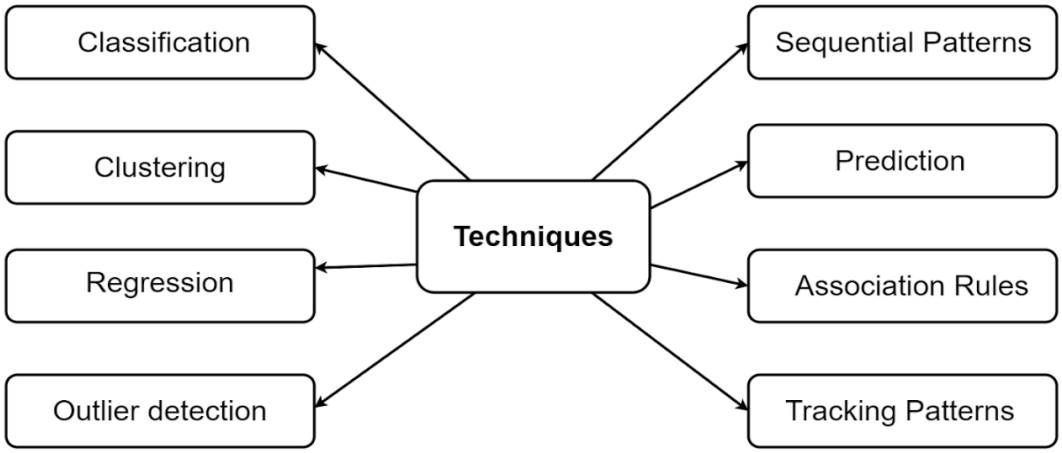

Figure 4. Data mining technique

Outlier Mining

Excess value outliers deviate from other observations in the data, they may indicate variability in measurement, test errors, or something new. In other words, the outlier is the recognition of the differences in the overall pattern in the sample.

External acquisition in high-density databases is an important and challenging issue in all fields and has practical implications, when we remove the data points that are present outside the cluster of data points we understand the model more properly and help in predicting the future more easily . Figure 5 shows Outlier in Scatter Plot

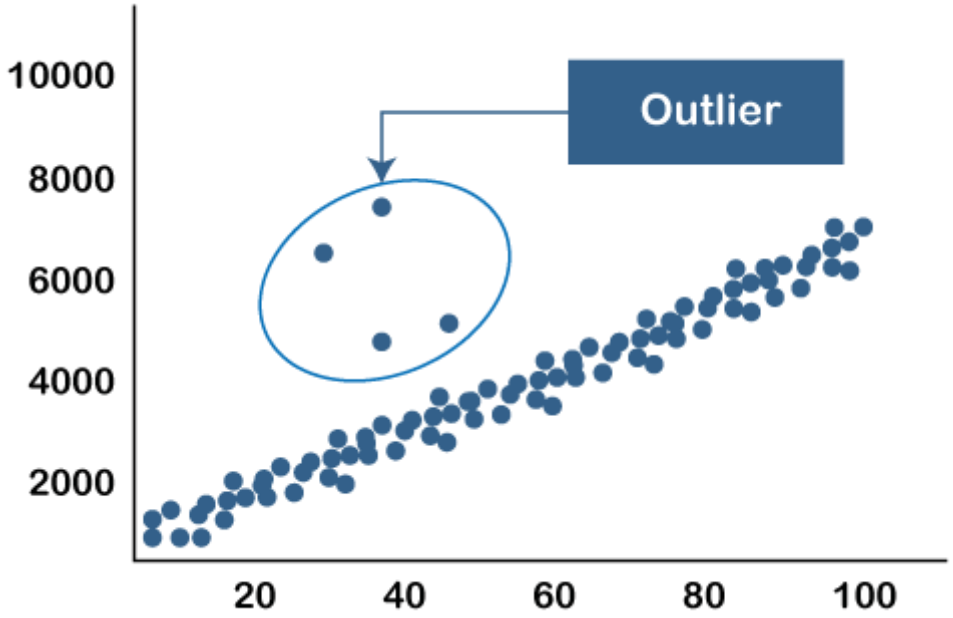

Figure 5. The outlier in Scatter Plot

Types of outliers

Outliers can be of two kinds: univariate and multivariate.

- **Univariate**: This type of data contains only one variable. Stable data analysis is thus an easy way to analyze as the information is related to only one variable value.It consists of one label which effects in change of other labels and we study the outlier data point of the major label. An example of a single outlier label can be length.
- **Multivariate**:When a data model with two or more labels is affecting the relation of rest labels then it is considered as multivariate. An example of multivariate data is that an advertiser wants to compare the popularity of four ads on a website, and then rate their click-through rates for men and women and re-evaluate the relationships between mutations. It resembles double but has more than one dependent variant.The data is always related to an object in real life so we use various analyses techniques to gain the knowledge of the changes of one object parameter and achieve the accuracy to predict the result with respect to the changes in the object.

Parkinson's disease

Parkinson's disease is caused due to many unpredictable symptoms but leads to brain disorder. The neuron present in the brain is affected and leads to malfunction of the nervous system leading to shaking, stiffness and changes in movement of the body.

In the start the disease is unnoticed and later on we observe slowness of the body moment and as time progresses it gets worse. They may also have psychological and behavioral changes, sleep problems, depression, memory problems and fatigue.

Symptoms of Parkinson's disease

Parkinson's disease has four main symptoms:

- trembling of body parts
- Stiffness of the limbs and trunk
- Slowness of movement
- Impaired balance and coordination, sometimes leading to falls

Treatment of Parkinson's disease

As we know there exist no cure to the disease but there exist some medical, surgical treatment and methods used to decrease the symptoms of the disease and which can slow down or restrict the condition being worse.

Medicines prescribed for Parkinson's Disease include:

- By increasing the level of dopamine helps the brain to function properly and increases the reaction time
- By doping chemical into the body which attacks the chemical present in the human body
- Drugs that help control no motor symptoms

LITERATURE SURVEY:

1. Mei J, Desrosiers C, Frasnelli J. Machine Learning for the Diagnosis of Parkinson's Disease: A Review of Literature. Front Aging Neurosci. 2021 May 6;13:633752. DOI: 10.3389/fnagi.2021.633752. PMID: 34025389; PMCID: PMC8134676.

diagnosis of Parkinson's disorder (PD) is frequently primarily based on medical exams and clinical signs and symptoms, which include the manifestation of numerous motor signs. but, conventional diagnostic techniques may also suffer as a result of subjugation as they depend on checks that are on occasion subtle in human eyes and therefore tough to distinguish, main to possible misdiagnosis. within the intervening time, the early signs and symptoms of non-PD vehicles may be moderated and may be due to many different situations. therefore, those symptoms are frequently not noted, making early prognosis of PD a challenge. To deal with those troubles and refine the analysis and evaluation of PD strategies, gadget mastering techniques have been used to differentiate PD from healthy controls or patients with comparable scientific presentations (e.g., movement issues or other Parkinsonian diseases). offer a comprehensive evaluation of the records strategies and machine studying strategies that have been used within the diagnostic and differential analysis of PD.

2. Cheng, Zhang Y, Zou, Chengming, Dong, Jianwei, 2019/09/24, 161, 168, 978-1-4503-6843-8, Outlier detection using isolation forest and local outlier factor, 10.1145/3338840.3355641, RACS '19: Proceedings of the Conference on Research in Adaptive and Convergent Systems

Outlier discovery, additionally known as confusing discovery, is one of the hottest troubles inside the facts mining industry. out off purchase algorithms, IF (wooded area) & nearby OF (LOF) highly used. however, iForest is best touchy to global outliers and is susceptible to managing neighborhood outsiders. although LOF does properly on outside acquisition, it has a excessive time constraint. to conquer the weaknesses of Isolation forest and LOF, a -layer integration approach is proposed for outside publicity. it can appropriately come across outliers on complex databases of a low period. This method first makes use of the low-density iForest to quick test a database, trim normal records, and produce an external candidate set. continuing to enhance the accuracy of the pruning, an external coefficient turned into brought to layout a threshold dedication method, based totally at the outside level of information. Then LOF is used to improve the separation of the external candidate set and to locate extra accurate outsiders. The proposed combination approach uses algorithms and focuses on critical pc sources in a vital phase. in the end, a big variety of assessments are accomplished to verify the integration method. The outcomes show that compared with present techniques, the integration method can extensively enhance the extent of outside acquisition and substantially reduce the duration of time.

3. **McDonnell MN, Rischbieth B, Schammer TT, Seaforth C, Shaw AJ, Phillips AC. Lee Silverman Voice Treatment (LSVT)-BIG to improve motor function in people with Parkinson's disease: a systematic review and meta-analysis. Clin Rehabil. 2018 May;32(5):607-618. DOI: 10.1177/0269215517734385. Epub 2017 Oct 5. PMID: 28980476**.

(LSVT) -LOUD This technique is used to improved the voice quality in human having PD. this was completly taken away the positive of another intervention, LSVT-BIG, to increase method.

4. W. Wang, J. Lee, F. Harrow and Y. Sun, "Early Detection of Parkinson's Disease Using Deep Learning and Machine Learning," in *IEEE Access*, vol. 8, pp. 147635-147646, 2020, DOI: 10.1109/ACCESS.2020.3016062.

Accurately detecting Parkinson's disease (PD) is very important in slowing down its progression and giving patients access to alternative therapies. A new in-depth learning proceed is introduced to determine in advance whether a person is affected by PD or not based on premotor factors. To detect PD early, several indicators have been considered in this study, including Rapid Eye Movement and hearing loss, Cerebrospinal fluid data, and dopaminergic cognitive symptoms. Comparisons between the proposed in-depth learning model and twelve machine learning and integrated data-based learning methods comprising 183 healthy individuals and 401 early PD patients show high efficacy of the designed model, achieving the highest accuracy, 96.45% on average. In addition to the PD detection, we also provide an important feature of the PD detection process based on the Boosting method.

5. T. J. Wrote, Y. Özkanca, C. Demiroglu, D. Si, D. C. Atkins, and R. H. Ghomi, "Parkinson's Disease Diagnosis Using Machine Learning and Voice," ***2018 IEEE Signal Processing in Medicine and Biology Symposium (SPMB),*** **2018, pp. 1-7, DOI: 10.1109/SPMB.2018.8615607.**

This paper examines the effectiveness of using classified classification algorithms, such as deep neural networks, to accurately identify people with the disease. Our high 85% accuracy given by machine learning models exceeds the accuracy of non-specialist clinical diagnostics (73.8%) and the accuracy among movement therapists (79.6% without follow-up, 83.9% after follow-up) with pathology. post-mortem examination as a basic fact [3]

6. JOUR, Maitin, Ana, García-Tejedor, Alvaro, Romero Muñoz, Juan Pablo, 2020/12/03, 8662, T1 - Machine Learning Approaches for Detecting Parkinson's Disease from EEG Analysis: A Systematic Review, 10, 0.3390/app10238662, Applied Sciences

The review processed is proposed in series with the well-known value of systematic analysis & meta-analysis (PRISMA). All the announcements before May 2020 are included and your main features and results are tested and documented.
Results: Nine subjects were included. Seven used EEG downtime and EEG unlocking two engines. Sample samples were used in 83.3% of the studies. This value of the PD grouping is 62-99.62%. There was no standard EEG cleaning procedure and significant variability in features extracted from the EEG. However, spectral features were more common. Conclusions: The features presented in the model and its structure are important for good performance in predicting classification. In contrast, the EEG cleaning protocol differed significantly between different studies and did not influence outcomes.

7. Wang, C., Liu, Z., Gao, H., & Fu, Y. (2019). Applying anomaly pattern score for outlier detection. IEEE Access, 7, 16008-16020

Outlier discovery is an important part of data mining and has been studied in depth by researchers over the past few decades. For neighbor-based acquisition methods such as KNN and LOF, settings that differ in neighbor value (indicated by parameter k) may significantly affect model performance. Thus, some recent research has focused on identifying the total value of k by analyzing global or local databases.

8. Wang, H., Bah, M. J., & Hammad, M. (2019). Progress in outlier detection techniques: A survey. Ieee Access, 7, 107964-108000.

External access is a major problem studied in various areas of research and applications. Researchers continue to develop robust strategies to provide effective outreach solutions. In this survey, they presented a comprehensive and systematic review of the progress of foreign diagnostic methods from 2000 to 2019. First, they have provided basic concepts of external acquisition and categorized them into different strategies from different external acquisition strategies, such as distance, cluster congestion, and learning-based methods.

9. Singh, K., & Upadhyaya, S. (2012). Outlier detection: applications and techniques. International Journal of Computer Science Issues (IJCSI), 9(1), 307.

Outliers are regarded as sound data in mathematics and have become an important problem that is still being researched in various research fields and domains used. Many external diagnostic methods are designed for specific system domains, but other methods are standard. Other application areas are investigated for strict confidentiality such as criminal investigations and terrorist activities. The tricks and results of such tactics are not readily available. A few surveys, research articles, reviews, and books cover outsourcing techniques for machine learning and individual mathematical domains in more detail. This paper attempted to integrate external identification methods, with a structured and standard definition.

10. Djenouri, Y., Belhadi, A., Lin, J. C. W., Djenouri, D., & Cano, A. (2019). A survey on urban traffic anomalies detection algorithms. IEEE Access, 7, 12192-12205.

This paper examines the use of external acquisition methods in urban traffic analysis. Divide existing solutions into two main categories: flow output detection and trajectory outlier detection. The first section includes solutions for detecting flow outflows and incorporating statistics, comparisons, and drilling methods. The second phase contains solutions for trajectory outliers, which include offline trajectory outliers and online processing of sub-trajectory outliers. Solutions for each of these categories are explained, drawn up, and discussed, and then open up ideas and research styles.

11. Park, C. H. (2019). Outlier and anomaly pattern detection on data streams. The Journal of Supercomputing, 75(9), 6118-6128

Data transfer is a series of data that is continuously generated over time. too large to be stored in memory and its original data distribution may change over time.Finding Outlier aims to find the most deviant data in the distribution of basic data. Although many external detection methods work in a bulk mode where all data samples are simultaneously available, the need for external detection methods and methods for obtaining unusual patterns in data transmission has grown. External detection is performed at a certain level of occurrence, and abnormal pattern detection involves obtaining Sometimes when data behavior is abnormal and different from normal behaviour.

12. JOUR, Singh, Karanjit, Upadhyaya, Shuchita, 2012/01/01, Outlier Detection: Applications And Techniques, 9, International Journal of Computer Science Issues

In this paper, we attempt to integrate external identification methods, with a structured and standard definition. With this application, we hope to gain a better understanding of the various

research analytical guidelines outside of us and beginners in this field of research who can find links in different application areas in detail.

13. H. Wang, M. J. Bah and M. Hammad, "Progress in Outlier Detection Techniques: A Survey," in IEEE Access, vol. 7, pp. 107964-108000, 2019, DOI: 10.1109/ACCESS.2019.2932769

This paper provides current developments in outsourcing strategies and provides a better understanding of external procurement methods. Research problems and open challenges will eventually provide researchers with a clear future for these external diagnostic methods.

14. BOOK, Liu, Fei Tony, Ting, Kai, Zhou, Zhi-Hua, 2009/01/19, 413, 422, Isolation Forest, 10.1109/ICDM.2008.17

This paper proposes a method based on a very different model that clearly distinguishes confusing rather than standard profiling points. To our knowledge the concept of separation has not been explored in modern literature. The benifits of isolation qualify the put forward, IF, to utalize a some model than has never seen in the present model, to create a complex algorithm of line time with a low constant need and low memory. Our robust experiment shows that iForest works well in ORCA, a near-time-based complexity-based approach, LOF, and AUC randomized jungle and time processing, especially on large data sets. Isolation Forest also works well for high-volume issues with a large number of non-essential features, and in situations where the training set does not have any confusion.

15. Evgeniou, Theodoros, Pontil, Massimiliano, 2001/01/01, 249, 257, Support Vector Machines: Theory and Applications, 2049, 10.1007/3-540-44673-7_12

The purpose of this chapter is twofold: The concepts of the past theory & present knowing of SVM was presented, as well as presentations & problem that emerged during the workshop. Vector Support Equipment (SVM) has presently incremented in the framework of math knowing theory and has been pass proposed in many practical request, from time-series predictions to observations, to biological data analysis for treatment identification. Their theoretical foundations and their experimental success encourage further research on their features, as well as their continued use.

16. Y. Ma and X. Zhao, "POD: A Parallel Outlier Detection Algorithm Using Weighted kNN," in *IEEE Access*, vol. 9, pp. 81765-81777, 2021, DOI: 10.1109/ACCESS.2021.3085605

In I technique uses info entropy to calculate the weight of every attribute & then sums the maximum size of the Z curve data into the Z value.The heavy kNN of each item is searched for the Z value. On this basis, we propose a similar external acquisition algorithm called POD to improve the efficiency of external acquisition.

17. Anila M, Dr. G. Pradeepini, 2020, A Review on Parkinson's Disease Diagnosis using Machine Learning Techniques, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 09, Issue 06 (June 2020),

This paper is a study of Parkinson's disease prediction using machine learning algorithms, using various new technologies, and their achieved accuracy.

18. T. M. Thang and J. Kim, "The Anomaly Detection by Using DBSCAN Clustering with Multiple Parameters," *2011 International Conference on Information Science and Applications*, 2011, pp. 1-5, DOI: 10.1109/ICISA.2011.5772437

In this article, we propose a new way to get DBSCAN parameters and use DBSCAN within those limits. Each collection may have different epsilon values and concepts in our algorithm. The algorithm is called DBSCAN-MP. We also suggest how to update common behavior by updating the size or creating new clusters as network location changes over time. KDD Cup 1999. We test the proposed algorithm using a data set. The result tells how good integration algorithms.

19. Fourure, Damien, Javaid, Muhammad, Posocco, Nicolas, Tihon, Simon, 2021/09/10, 3, 18, 978-3-030-86513-9, Anomaly Detection: How to Artificially Increase Your F1-Score with a Biased Evaluation Protocol, 10.1007/978-3-030-86514-6_1

In this essay, we display that F1-score and AVPR are highly tactful to pollution levels. One result is that it is possible to increase their prices hypocritically by adjusting the train-segregation process. This leads to misleading comparisons between textbook algorithms, especially if the test protocol is not well defined.Furthermore, we show that F1-score and AVPR cannot be used to compare performance across all different sites because they do not reflect the internal difficulties of modeling such data. We recommend a standard testing process for uncomplicated confusion detection, which includes the use of other metrics such as the AUC, It is very powerful in random selection of test protocols.

20. Kaur, Parmeet, 2016/12/01, 693, 696, Outlier Detection Using Kmeans and Fuzzy Min-Max Neural Network in Network Data, 10.1109/CICN.2016.142

In this paper, we propose kmean clustering and neural network as a novel to find an outsider in network analysis. Particularly in social networking, k means integration, and the neural network is used to locate a scattered social user in a network and acquires an additional clique that describes a strong data connection. In this article, we suggest that this approach works to find outsiders in social media analysis. upcomming, we show the profit of this new method by test data.

PROBLEM IDENTIFICATION :

Data mining is performed using methods like classification, clustering, and association rule. Most of the time, knowledge discovery is performed through these methods but, in the last few years, the use of outlier mining has emerged as a promising technique. There has been a limited amount of research related to the use of these methods in detecting anomalous data . In this feature,e we study the outlier data points in Parkinson's Disease. We check the accuracy of the bench data set bypassing the real-time dataset into the model with the bench dataset as reference Real-Time data stream analysis is performed to check the OSE model plot on the real-time dataset.

OBJECTIVE :

1. **Data Pre-Processing**

The input data is most likely to be imperfect, contains errors, and is redundant for starting a data mining process. If the data size is large, we need a well-built model to analyze it. Data mining is required in data mining as many data mining algorithms need feasible data to mined and give meaningful insight.

2. Data Modelling:

To detect an outlier is by graphing the features or the data points. Visualization is one of the best and easiest ways to have an inference about the overall data and the outliers. We use different visualization tools to represent outlier detection data points. The outlier can also be detected by using methods like the Isolation technique, SVM, and NN.

3. Outlier Analysis:

The outlier analysis is done by studying the data without outlier data points. Data Visualization tool is variouso compare varies outlier mining algorithms. The outlier data points are studied to give a meaningful insight which helps in diagnose osis of the disease, Similarly the,e data without an outlier is studied how accurately the target can be predicted.

4. Performance Analysis:

The outlier model performance is measured using accuracy_score which gives the accuracy of the model in percent, a confusion matrix that classifies all the columns in the model, Real-time Performance analysis is done by plotting the curve of some right prediction with respective time.

METHODOLOGY:

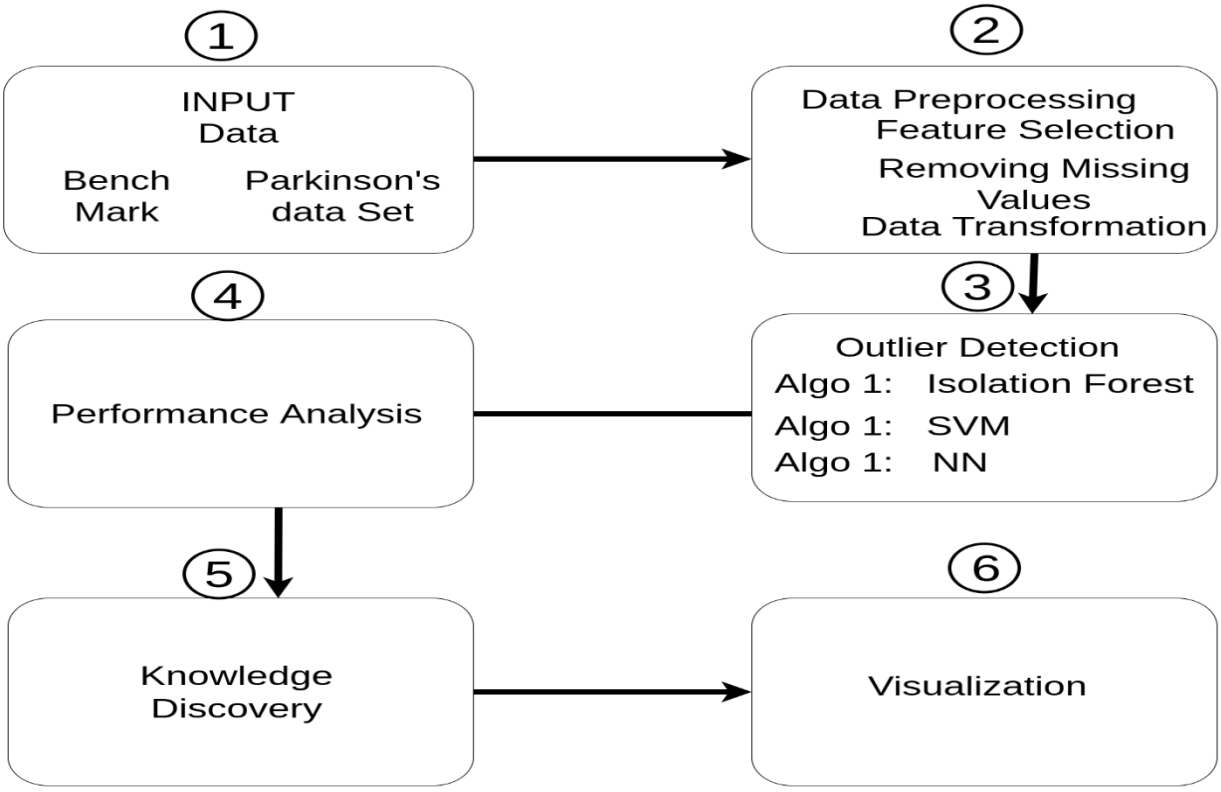

Figure 6. System Design

Input Data: Data is given to the algorithmic model in two ways. Bench dataset: The Data or information is gather from various streams like Kaggle, UCI repository etc. The data is used as a bench data set where this static dataset is preprocessed and used to get an insight of the data creating the standards Real-time DataStream: The DataStream generated from real-time API is sent to the model. The model is tested by keeping the bench dataset as standards and generating the output by undergoing the artificial intelligence methods

Data Preprocessing: Because data is partial and also conflicting and may have huge number of repeated information, it cannot be directly fed to data mining process. If the Data is huge then advances techniques are used to analyse it, to get a proper insight with high Accuracy. Pre-processing of data helps data, adapt to the requirements of the model. which helps in faster analysis, otherwise it may become very difficult to the algorithm to provide accurate insight. Transformation of data for model requirement, data integration, Removing missing or null or

exceptional values present in the dataset and Normalising the data, helps for better accuracy of the analysis and prediction. The processed data can be said as more attest and appropriate form of data that may be used for better analysis using various data mining process.

Outlier Detection: To detect an outlier is by graphing the features or the data points. Visualization is one of the best and easiest ways to have an inference about the overall data and the outliers. Outlier that are present in the dataset can be identified visually by using some of visualization techniques such as Box plotting and Scatter plotting. The outlier can also be detected by using methods like the Isolation technique, SVM and NN

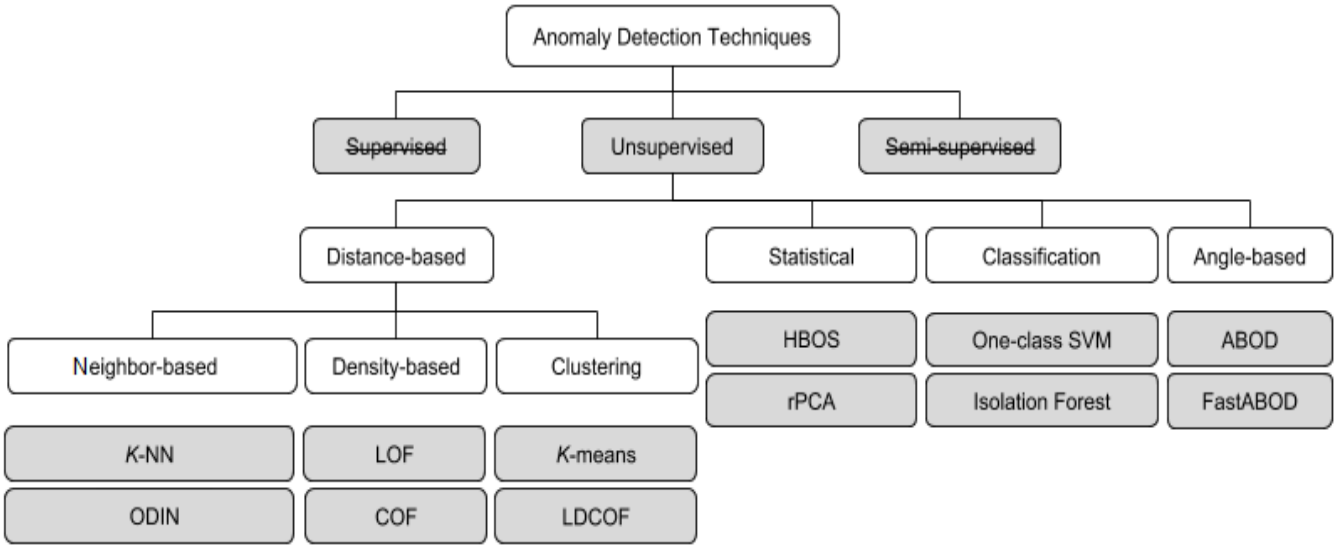

Figure 7 .Outlier Detection Methods

Performance Analysis: In the initial stage of Software Development process, Analysing performance of the data may become very difficult. Until now, many researchers have developed statistical techniques. Subjects having High stress can be classified using many ML techniques. The percentage is recorded for all subjects after they are introduced to stress-inducing stimuli. High Stress or Average Stress of Subject are calculated based on PPG signals, PRV parameters. Logistic Regression, Support Vector Machine (SVM), Decision Tree, and Random Forest which are different ML algorithms and PRV parameters are used to form the dataset, and these dataset are tested trained for classification of subjects as average stress or high stress. Performance Analysis of the trained model is analysed by various performance parameters like confusion matrix, accuracy score , f1-score.

Knowledge Discovery: Extracting a new insight or explicit information from the data based on the previous knowledge, can be said as Knowledge discovery. Knowledge discovery can be done through conveyance, unification and scheme of various sources of proper knowledge.

Visualization: Visualization of Data is nothing but presenting knowledge and statistical data in the form of insight. This can be accomplished by using various Visualization software like Tableau, Matplotlib which is an python library, Seaborn, Power BI , Microsoft Azure analysis etc. Many knowledgeable insight of data can be obtained through various graphs, maps and also various charts. By the help visualisation, insight can be understood more efficient.

EXPECTED OUTCOME OF THE PROPOSED RESEARCH:

- The outcome of the model will contain a list of graphical diagrams and knowledge which is gained by using outlier mining. We will perform accuracy testing on the outcome to validate if the input is given results proper status of Parkinson's disease.
- We will be using various visualization tools to evaluate the algorithm used and ensure that the meaning of the visual graph is insightful. We will validate the bench dataset by passing a real-time dataset to the model keeping the bench dataset as a reference.
- Our Anomaly detection model will come across as a very useful application, to a wide range of doctors studying Parkinson's disease to understand the outliers in the patient and help them to diagnose properly.

CONCLUSION:

- The main strength of this project is the identification of potential outliers as it is important for increasing the accuracy of the model. An outlier may indicate bad data. Outlier mining can give meaningful insight into the dataset and helps in the analysis of the dataset more accurately. By understanding the errors we can analyze the reason behind the outlier obtained from the data source
- By analyzing the outlier data points of Parkinson's disease, the doctors can analyze Parkinson's patients more efficiently.
- By the interactive process, the data inputted from the sensors can be analyzed quickly and no input sensed is neglected as outlier uses every bad data sensed to give more accurate results.

REFERENCES:

1. https://www.ibm.com/search?lang=en&cc=in&q=ai%20ml
2. http://www.aiml.foundation/
3. https://trends.google.com/trends/?geo=IN
4. https://www.mindmeister.com/?utm_source=google&utm_medium=cpc&utm_campaign=ind_en_search&utm_content=mm&gclid=
5. Kaur, Parmeet, 2016/12/01, 693, 696, Outlier Detection Using Kmeans and Fuzzy Min-Max Neural Network in Network Data, 10.1109/CICN.2016.142
6. Mathew, Melvin John, and Jomon Baiju. "MACHINE LEARNING TECHNIQUE BASED PARKINSON'S DISEASE DETECTION FROM SPIRAL AND VOICE INPUTS." European Journal of Molecular & Clinical Medicine 7.4: 2020.
7. T. M. Thang and J. Kim, "The Anomaly Detection by Using DBSCAN Clustering with Multiple Parameters," 2011 International Conference on Information Science and Applications, 2011, pp. 1-5, DOI: 10.1109/ICISA.2011.5772437
8. Anila M, Dr. G. Pradeepini, 2020, A Review on Parkinson's Disease Diagnosis using Machine Learning Techniques, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 09, Issue 06 (June 2020),
9. Y. Ma and X. Zhao, "POD: A Parallel Outlier Detection Algorithm Using Weighted kNN," in IEEE Access, vol. 9, pp. 81765-81777, 2021, DOI: 10.1109/ACCESS.2021.3085605.
10. Evgeniou, Theodoros, Pontil, Massimiliano, 2001/01/01, 249, 257, Support Vector Machines: Theory and Applications, 2049, 10.1007/3-540-44673-7_12
11. BOOK, Liu, Fei Tony, Ting, Kai, Zhou, Zhi-Hua, 2009/01/19, 413, 422, Isolation Forest, 10.1109/ICDM.2008.17
12. H. Wang, M. J. Bah and M. Hammad, "Progress in Outlier Detection Techniques: A Survey," in IEEE Access, vol. 7, pp. 107964-108000, 2019, DOI: 10.1109/ACCESS.2019.2932769.
13. JOUR, Singh, Karanjit, Upadhyaya, Shuchita, 2012/01/01, Outlier Detection: Applications And Techniques, 9, International Journal of Computer Science Issues
14. Park, C. H. (2019). Outlier and anomaly pattern detection on data streams. The Journal of Supercomputing, 75(9), 6118-6128.
15. ***Djenouri, Y., Belhadi, A., Lin, J. C. W., Djenouri, D., & Cano, A. (2019). A survey on urban traffic anomalies detection algorithms. IEEE Access, 7, 12192-12205.***
16. *Singh, K., & Upadhyaya, S. (2012). Outlier detection: applications and techniques. International Journal of Computer Science Issues (IJCSI), 9(1), 307*
17. ***Wang, H., Bah, M. J., & Hammad, M. (2019). Progress in outlier detection techniques: A survey. Ieee Access, 7, 107964-108000.***
18. ***Wang, C., Liu, Z., Gao, H., & Fu, Y. (2019). Applying anomaly pattern score for outlier detection. IEEE Access, 7, 16008-16020.***
19. *JOUR, Maitin, Ana, García-Tejedor, Alvaro, Romero Muñoz, Juan Pablo, 2020/12/03, 8662,*
20. *T1 - Machine Learning Approaches for Detecting Parkinson's Disease from EEG Analysis: A Systematic Review, 10, 0.3390/app10238662, Applied Sciences*
21. *W. Wang, J. Lee, F. Harrow and Y. Sun, "Early Detection of Parkinson's Disease Using Deep Learning and Machine Learning," in IEEE Access, vol. 8, pp. 147635-147646, 2020, DOI: 10.1109/ACCESS.2020.3016062.*
22. *McDonnell MN, Rischbieth B, Schammer TT, Seaforth C, Shaw AJ, Phillips AC. Lee Silverman Voice Treatment (LSVT)-BIG to improve motor function in people with Parkinson's disease: a systematic review and meta-analysis. Clin Rehabil. 2018 May;32(5):607-618. DOI: 10.1177/0269215517734385. Epub 2017 Oct 5. PMID: 28980476.*
23. *Mei J, Desrosiers C, Frasnelli J. Machine Learning for the Diagnosis of Parkinson's Disease: A Review of Literature. Front Aging Neurosci. 2021 May 6;13:633752. DOI: 10.3389/fnagi.2021.633752. PMID: 34025389; PMCID: PMC8134676.*
24. *T. J. Wrote, Y. Özkanca, C. Demiroglu, D. Si, D. C. Atkins, and R. H. Ghomi, "Parkinson's Disease Diagnosis Using Machine Learning and Voice," 2018 IEEE Signal Processing in Medicine and Biology Symposium (SPMB), 2018, pp. 1-7, DOI: 10.1109/SPMB.2018.8615607.*

Anomaly detection in Parkinson's disease using unsupervised machine learning Approach

ORIGINALITY REPORT

12%	6%	5%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Wu Wang, Junho Lee, Fouzi Harrou, Ying Sun. "Early Detection of Parkinson's Disease Using Deep Learning and Machine Learning", IEEE Access, 2020 Publication	2%
2	medium.com Internet Source	2%
3	uwe-repository.worktribe.com Internet Source	1%
4	Submitted to University of East London Student Paper	1%
5	www.nia.nih.gov Internet Source	1%
6	Tran Manh Thang, Juntae Kim. "The Anomaly Detection by Using DBSCAN Clustering with Multiple Parameters", 2011 International Conference on Information Science and Applications, 2011 Publication	1%

7	www.ncbi.nlm.nih.gov Internet Source	1 %
8	www.journaltocs.ac.uk Internet Source	1 %
9	Submitted to South Asian University Student Paper	1 %
10	Submitted to University of Hertfordshire Student Paper	<1 %
11	Submitted to Acknowledge Education Pty Ltd Student Paper	<1 %
12	Vindya Wijesinghe, Manoj Perera, Chamod Peiris, Praveen Vidyaratne, Dasuni Nawinna, Janaka Wijekoon. "Power Profiling:Assessment of Household Energy Footprints", 2021 5th International Conference on Green Energy and Applications (ICGEA), 2021 Publication	<1 %
13	Submitted to University of Nebraska at Omaha Student Paper	<1 %
14	Submitted to Baze University Student Paper	<1 %
15	Damien Marquis, Eric Guillaume, Damien Lesenechal. "Accuracy (Trueness and Precision) of Cone Calorimeter Tests with and	<1 %

Without a Vitiated Air Enclosure", Procedia Engineering, 2013
Publication

16 Kushagra Awasthi, Pranav Nanda, K V Suma. "Performance analysis of Machine Learning techniques for classification of stress levels using PPG signals", 2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), 2020 <1 %
Publication

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		