Bulletin of Environment, Pharmacology and Life Sciences
Bull. Env. Pharmacol. Life Sci., Vol 11 [5] April 2022 : 124-129

Online ISSN 2277-1808
Journal's URL:http://www.bepls.com
CODEN: BEPLAD

ORIGINAL ARTICLE **OPEN ACCESS**

Develop the machine learning model to support physicians in early diabetes forecasting

Biswajeet Champaty[1],Aiswarya Dash[2], Mrutyunjaya S Yalawar[3].
1.Associate Professor, School of Engineering, Ajeenkya DY Patil University, Lohegaon, Pune, Maharashtra-412105,India.
2.Associate Professor, School of Engineering, Ajeenkya DY Patil University, Lohegaon, Pune, Maharashtra-412105,India.
3.Assistant Professor, Dept of CSE, CMR ENGINEERING COLLEGE, Hyderabad, Telangana, India..501401
Correspondence Email: biswajeet.champaty@gmail.com

ABSTRACT
In various industries, machine learning methods are applied to perform predictive analytics using massive amounts of data. Predictive analytics in health care seems to be a hard process, but that can adjacent community clinicians in making timely choices concerning a participant's health service based on a large volume of data. This study examines data analysis in healthcare and uses 6 traditional machine learning methods.A collection in a specific format was assembled for the study and 6 different machine learning algorithms were used for the information.The effectiveness and appropriateness of the methods employed will be evaluated and compared.The original study evaluating different machine learning techniques shows which model would be more suitable for the diagnosis of diabetes.Using machine learning techniques, this research is intended to help physicians and clinicians identify diabetics early.
Keywords: Predictive analytics; healthcare; diabetes; machine learning

Received: 18.02.2022 Revised :12.03.2022 Accepted: 24.03.2022

INTRODUCTION

As technology advances, gadgets generate vast amounts of information daily. There has been an explosion in the data available for researchers throughout the world. To effectively manage, evaluate, & visualize data, one must look for, discover, & implement new technology tools and techniques owing to the difficulty, large amount, & heterogeneity of the information.From 2008 to 2015, Google Scholar results for the phrase "Big data" [1]. These findings demonstrate how this discipline has progressed over time, as well as the growing number of papers in the field of big data. This explosive expansion in the field of big data began in 2012, and this research area continues to draw an increasing number of academics.Many scholars from all around the world have been working on big data analytics & prescriptive modeling in medicine and also other fields in recent decades [2]. This study adopts & expands on this taxonomy. There were various big data resources from which information was analyzed, and also different pieces & large advanced analytics. We would concentrate on machine learning for predictive in this work [3]. The research project of several authors was examined as a foundation for our investigation & comprehension. A few research publications were presented below in this connection.Hydrocephalus diagnosis using imaging techniques & deep learning. Researchers retrieved 77 imaging characteristics from brain ventriculomegaly [4]. Support vector machines, a deep learning method, were used to the ventricle characteristics of 25 children. Whom required shunts & who didn't was the question. [5] The information was tabulated & analyzed. According to the findings, shunts were required in three out of every 4 kids, with a sensitivity of 75% & specificity of 95%. A novel fuzzy regulation classification has been developed.For analysis & group construction, algorithms based on function optimization & fuzzy-rule basis classifiers have been used. The suggested scheme was compared to modern strategies, with the outcomes evaluated in terms of reliability, refresh rate, false alarm rate, & computing price [6]. The suggested technique outperforms Bayes networks, multi-layers, & decision tables, according to the outcomes.Predicted the complications of diseases, symptoms, & treatment options to individuals [7]. For the predictions & therapy kinds, a predictive analytic method & Hadoop map reduction were employed. Huge data collection was obtained from several labs & hospitals, EHR & PHR were analyzed in Hadoop, & end findings were shared across different computers based on the geographical areas [8].

A thorough review of the literature on big data & analytics. The authors focused on using machine learning algorithms to forecast failures & power demand in manufacturing power applications & systems. The Naive Bayes method has been used to create a healthcare predictive model [9]. The proposed system searches a health dataset for data hiding linked to various illnesses & retrieves it. Customers can discuss their wellness issues with this platform, which subsequently uses Naive Bayes to forecast the proper disease [10]. Researchers optimize machine learning methods for the precise estimation of cardiac illnesses in areas with regular chronic disease outbreaks [11].The testing was performed on a patient with a persistent case of brain infarction. The results from the experiment reveal that Naive Bayes works better for data structure, and when organized & text information was mixed, the suggested method performed best. The authors used septic mortality as the predictive application instance, because of the clinical relevance of septic [12]. For 12 months, data were collected from four emergency departments. K-mean grouping was employed for information processing & grouping, while the randomized forests method has been used for predictions [13]. In emergency treatment, the traditional models of prognosis have been the logistic regression models and CART. In comparison to other algorithms, the findings indicate that random forest predicts better reliable data.The Randomized Forests method surpassed other methods & provided 88 percent accuracy, based on the 26 factors & 8 classifications of female infertility. An automated recommendation system that aids doctors & patients in determining the short-term danger of heart problems. The authors suggested a heart disease prediction system, & depending on the outcomes, the software also advises clients on the necessity for certain tests & visits to a specialist [14].The time Series data management method seems to be the fundamental component of the recommender systems. Real-world data had been used to evolve the proposed solution. The researchers carried out a pilot study on a group of patients with heart failure, gathering information through regular medical readings. There have been 7147 patient information entries collected [15]. According to the findings, the suggested program's suggestion accuracy varies between 75% & 100%.There have only been a few individuals in the collection, and also information & measurements recorded were only ever numeric numbers.

The literature focuses on diabetes, a disorder in which the body's ability to retain malted grain levels in the blood was impaired. In a Hadoop/Map reduction context, the process utilizes many computers to predict predictions. The suggested technique can predict which form of diabetes a patient would have. Learning algorithms classification algorithms have been used to determine if a person has non-CKD or CKD.The UCI repository [16] provided a collection of 400 data entries & 25 characteristics. The authors used a smaller sample with 14 CKD-related parameters. Various machine learning methods were applied to the image utilizing Microsoft Azure Machine Learning Studios. The outcome indicates that the Multiclass Decision Forest Method works better & has a 99.1% average accuracy.The capacity to forecast the prognosis of breast cancer survivors. The UCI machine learning collection [17] provided a collection, including information from over 683 instances. Every database entry had 26 variables related to the illness. On the information, the authors utilized five machine learning techniques. According to the findings, gradient boosting machines outperformed other methods, achieving a 97 percent average accuracy. Utilizing machine learning techniques, telemonitoring information could be used to forecast asthma symptoms before they happen.

MATERIAL AND METHODS

Data mining was among the most essential & widely utilized technologies in the industry today for analyzing the result & obtaining insights from the data. Data mining algorithms, like as artificial intelligence, deep learning, & statistics, have been used in information retrieval. In this research, a machine learning model was employed to predict illness. Machine learning provides a variety of tools & techniques that allows a computer to transform raw data into usable, useful knowledge.Machine learning algorithms were already applied in four ways. It must be primarily used for predictive because it creates a model from information, which contains events or answers. Labeled deep learning uses the algorithm. Whenever the outcomes or answers were uncertain, a learning algorithm was utilized, and the classifier is designed utilizing unlabeled data. Patterns recognition & description, modeling were two of the major applications of this type of learning.Grouping issues arise in unsupervised classification. Semi-supervised training would be a hybrid of supervised and unsupervised machine learning techniques. Finally, the learning algorithm tries to use the information gained through interaction with the environment to adopt behaviors that improve the benefit or decrease the risk. Because this study assesses the effectiveness of deep learning systems for predictive within medicine, it employs supervised learning.The basic mechanism of learning algorithms was depicted in Figure 1. In reinforcement methods, information was transformed into a system, also known as a model, for the aim of learning. This information includes the input data, also known as predictive variables, & also the accurate output value

systems. The model can learn the connections, trends, & correlations among 's an attribute and also the output value using this information. Researchers could use models that predict reactions given new information once they have learned those tendencies.

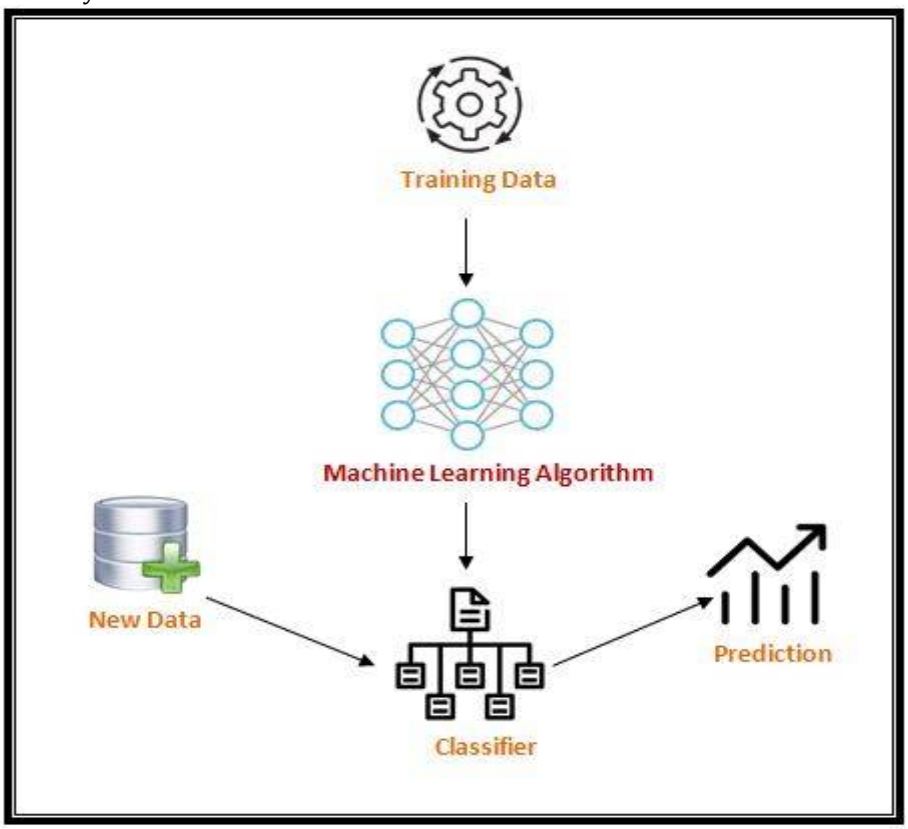

Figure 1: The Method of Supervised Machine learning

Figure 2 depicts the steps involved in constructing & assessing prediction models. Enthought Canopy has been used to code in the Python programming language. Enthought Canopy seems to be a Python package release providing core integrated capabilities for business applications, continuous analytical techniques, & information visualization. Following the acquisition of the information from the UCI machine learning repository.Information was done before being done with the information in the first stage. The information must be formatted for good functioning & evaluation. Information was collected for missing values & diabetic cases were converted to a numeric value, such as 1 or 0. The number of cases with a zero value was found to be extremely large during the data gathering.To deal with missing or zero values in the collection, information imputing has been used. Following that, eight characteristics were chosen from a total of nine. The data were then separated into two groups: learning information and test data. Then, using the training data, a deep learning model has been trained to make accurate predictions. After the training has been completed using data for training, test data set were used to predict answers & assess correctness, and also the system was finally assessed.

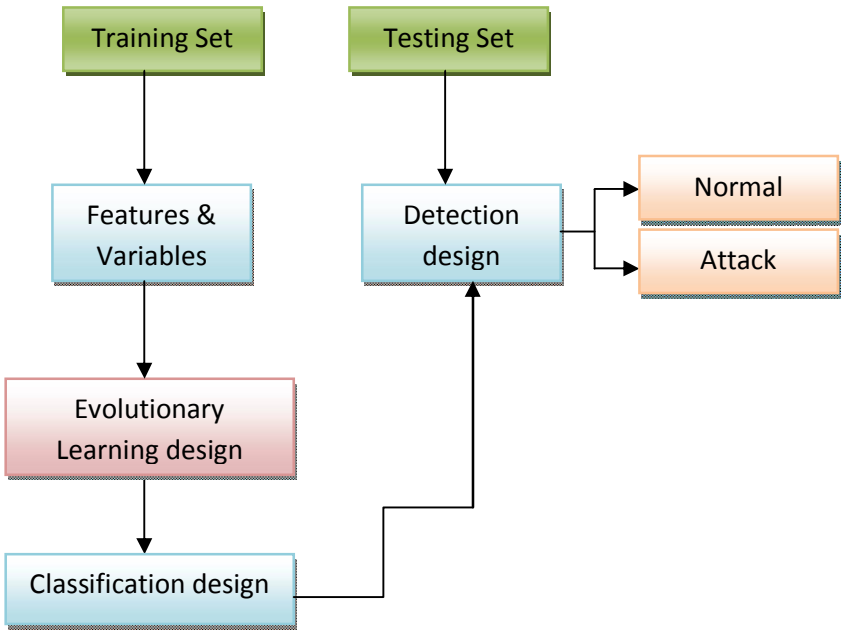

Figure 2: Evaluating methods.

RESULTS AND DISCUSSIONS

Six machine learning algorithms were employed in this study. NB, KNN, SVM, LR, DT, & RF seem to be the algorithms. The PIMA Indian database was used to test all of these methods. The information was split into two categories: train information & testing set, with each section containing 70% & 30% of the entire dataset. Utilizing Enthought Canopy, all six algorithms have been applied to the same data & conclusions were produced. The key form of assessment we used here study was prediction accuracy. Equation 1 could be used to defy precision. The individual's ultimate success rate was considered accuracy.

Accuracy = (TP+TN) / (P + N) …….(1)

All genuine positive & genuine negative predictions were split by all positively and negatively predictions. Table 1 shows the projected True Positive (TP), True Negative (TN), False Negative (FN), & False Positive (FP) for all methods. In this situation, TP stands for both existing and projected diabetes. FN, diabetes that isn't expected to become diabetic. FP anticipated being diabetic but did not get it. Real diabetes has not been a disease, & projected diabetic was never diabetes.

Table 1:Confusion matrix

Algorithm	True Positive	False Negative	False Positive	True Negative
LR	45	25	37	131
DT	62	49	21	105
RF	45	32	38	122
NB	53	35	29	119
KNN	42	21	26	139
SVM	38	17	38	142

Figure 3 depicts the significance of the characteristics. Among several characteristics, it has been shown that plasma glucose level has been the most important. The BMI & gender seem to be the 2nd and 3rd most essential traits, respectively in Figure 4. It may be deduced that these critical characteristics play a vital part in diabetes prognosis also are predictive of whether or not a person would develop diabetes.Algorithm efficiency was evaluated and can be seen in Picture 4. LR has a 74 percent accuracy, SVM has a 77 percent accuracy, NB has a 74 percent accuracy, DT & RF have a 71 percent accuracy, & KNN has a 77 percent accuracy. As a result, SVM & KNN scored the highest accuracy of 77 percent. The Support vector machine & K-nearest neighbor algorithms were suitable for forecasting the diabetic state of individuals, according to the experimental data.

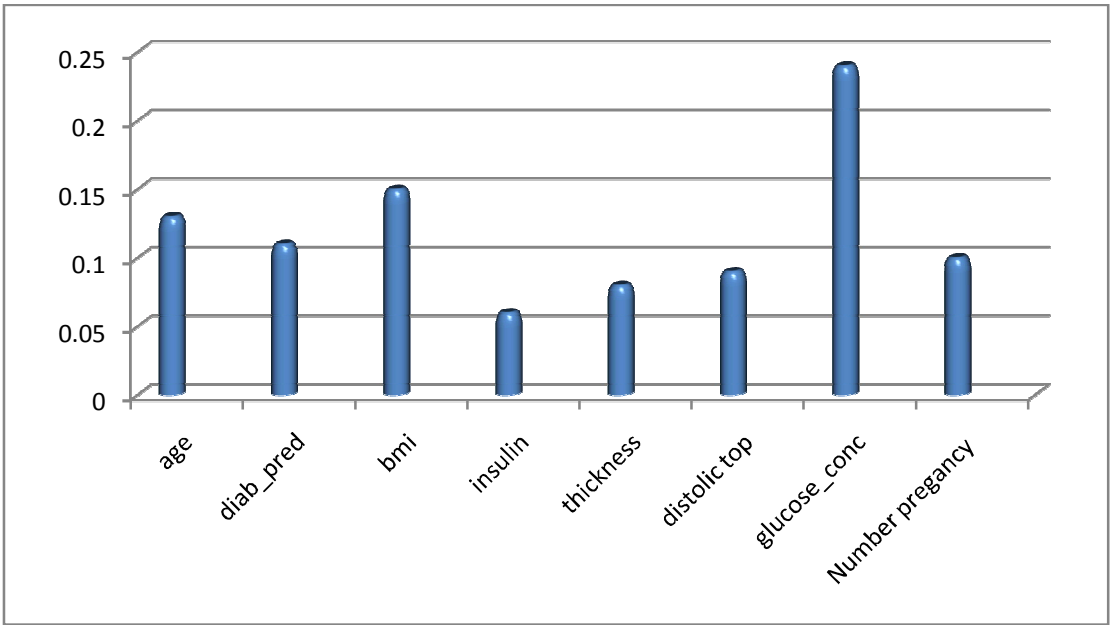

Figure 3. Features characteristics

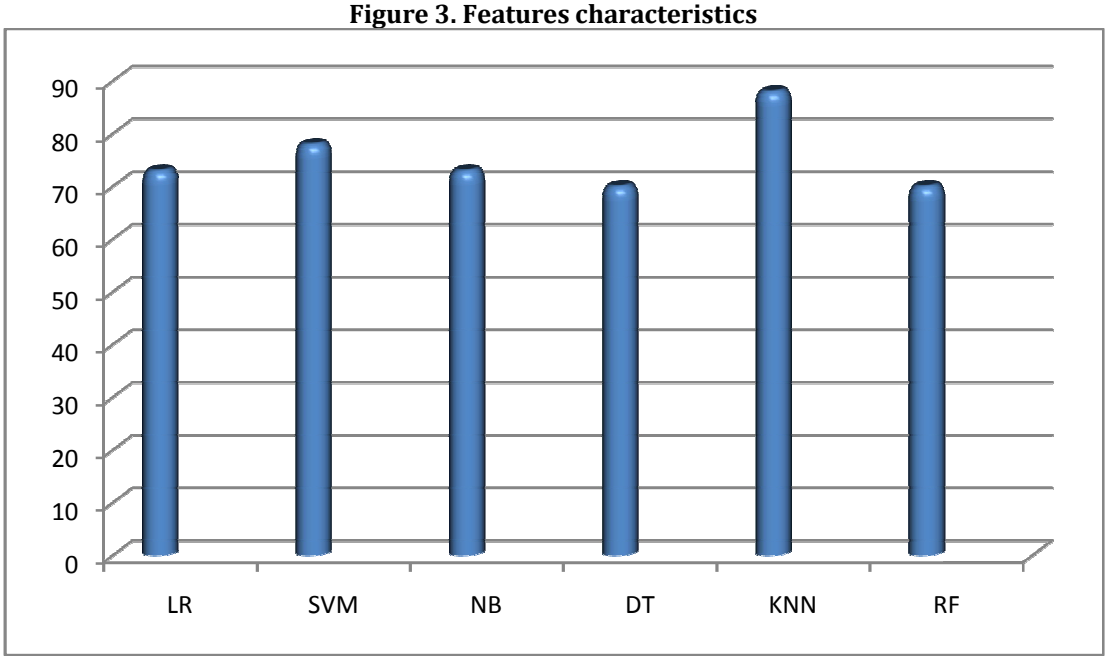

Figure 4. Accuracy of algorithms

CONCLUSION

Prescriptive modeling in healthcare has the potential to revolutionize the way healthcare practitioners & researchers analyze information and make choices. Researchers employed six common machine learning techniques for predictive in this work. SVM, KNN, and other methods were examples of these. Diabetic forecasts were developed using the PIMA Indian dataset, which had 768 individuals. The prediction classifier was tested & tested using 8 variables.From the results of the research, it would be clear that SVM & KNN have the highest precision for diagnosing diabetes. In comparison to the other four methods employed in this research, both of these methods achieve 77 percent accuracy. As a result, it could be inferred that SVM & KNN are useful for predicting diabetes.

ACKNOWLEGEMENT

The authors acknowledge the subjects who were involved in the study.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest for this study

REFERENCES

1. Sarwar, M. A., Kamal, N., Hamid, W., & Shah, M. A. (2018, September). Prediction of diabetes using machine learning algorithms in healthcare. In *2018 24th international conference on automation and computing (ICAC)* (pp. 1-6). IEEE.
2. Yahyaoui, A., Jamil, A., Rasheed, J., &Yesiltepe, M. (2019, November). A decision support system for diabetes prediction using machine learning and deep learning techniques. In *2019 1st International Informatics and Software Engineering Conference (UBMYK)* (pp. 1-4). IEEE.
3. Ezhilarasi, T. P., Sudheer Kumar, N., Latchoumi, T. P., &Balayesu, N. (2021). A secure data sharing using IDSS CP-ABE in cloud storage. In Advances in Industrial Automation and Smart Manufacturing (pp. 1073-1085). Springer, Singapore.
4. Latchoumi, T. P., & Parthiban, L. (2021). Quasi oppositional dragonfly algorithm for load balancing in cloud computing environment. Wireless Personal Communications, 1-18.
5. Garikapati, P., Balamurugan, K., Latchoumi, T. P., &Malkapuram, R. (2021). A Cluster-Profile Comparative Study on Machining AlSi7/63% of SiC Hybrid Composite Using Agglomerative Hierarchical Clustering and K-Means. Silicon, 13(4), 961-972.
6. Pavan, V. M., Balamurugan, K., &Latchoumi, T. P. (2021). PLA-Cu reinforced composite filament: Preparation and flexural property printed at different machining conditions. Advanced composite materials.
7. Spänig, S., Emberger-Klein, A., Sowa, J. P., Canbay, A., Menrad, K., & Heider, D. (2019). The virtual doctor: An interactive clinical-decision-support system based on deep learning for non-invasive prediction of diabetes. *Artificial intelligence in medicine*, *100*, 101706.
8. Balamurugan, K. (2020). Compressive Property Examination on Poly Lactic Acid-Copper Composite Filament in Fused Deposition Model–A Green Manufacturing Process. *Journal of Green Engineering*, *10*, 843-852.
9. Katarya, R., & Srinivas, P. (2020, July). Predicting heart disease at early stages using machine learning: A survey. In *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)* (pp. 302-305). IEEE.
10. Armstrong, G. W., & Lorch, A. C. (2020). An (eye): a review of current applications of artificial intelligence and machine learning in ophthalmology. *International ophthalmology clinics*, *60*(1), 57-71.
11. Yarlagaddaa, J., Malkapuram, R., & Balamurugan, K. (2021). Machining studies on various ply orientations of glass fiber composite. In *Advances in Industrial Automation and Smart Manufacturing* (pp. 753-769). Springer, Singapore.
12. Le, T. M., Vo, T. M., Pham, T. N., & Dao, S. V. T. (2020). A novel wrapper–based feature selection for early diabetes prediction enhanced with a metaheuristic. *IEEE Access*, *9*, 7869-7884.
13. Chinnamahammad Bhasha, A., & Balamurugan, K. (2019). Fabrication and property evaluation of Al 6061+ x%(RHA+ TiC) hybrid metal matrix composite. *SN Applied Sciences*, *1*(9), 1-9.
14. Sonar, P., &JayaMalini, K. (2019). Diabetes prediction using different machine learning approaches. In *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 367-371). IEEE.
15. Choudhury, A., & Gupta, D. (2019). A survey on the medical diagnosis of diabetes using machine learning techniques. In *Recent developments in machine learning and data analytics* (pp. 67-78). Springer, Singapore.
16. Latchoumi, T. P., Kalusuraman, G., Banu, J. F., Yookesh, T. L., Ezhilarasi, T. P., & Balamurugan, K. (2021, November). Enhancement in manufacturing systems using Grey-Fuzzy and LK-SVM approach. In 2021 IEEE International Conference on Intelligent Systems, Smart and Green Technologies (ICISSGT) (pp. 72-78). IEEE.
17. Latchoumi, T. P., Swathi, R., Vidyasri, P., & Balamurugan, K. (2022, March). Develop New Algorithm To Improve Safety On WMSN In Health Disease Monitoring. In 2022 International Mobile and Embedded Technology Conference (MECON) (pp. 357-362). IEEE.

CITATION OF THIS ARTICLE

B Champaty, Aiswarya Dash, Mrutyunjaya S Yalawar. Develop the machine learning model to support physicians in early diabetes forecasting.Bull. Env. Pharmacol. Life Sci., Vol 11[5] April 2022:124-129