ISSN Print: 2518-4245
ISSN Online: 2518-4253
Vol. 60(4), December 2023

PROCEEDINGS
OF THE PAKISTAN ACADEMY OF SCIENCES:
A. Physical and Computational Sciences

PAKISTAN ACADEMY OF SCIENCES
ISLAMABAD, PAKISTAN

Proceedings of the Pakistan Academy of Sciences: Part A
Physical and Computational Sciences

Proceedings of the Pakistan Academy of Sciences A. Physical and Computational Sciences is the official flagship, the peer-reviewed quarterly journal of the Pakistan Academy of Sciences. This open-access journal publishes original research articles and reviews on current advances in the field of Computer Science (all), Materials Science (all), Physics and Astronomy (all), Engineering Sciences (all), Chemistry, Statistics, Mathematics, Geography, Geology in English. Authors are not required to be Fellows or Members of the Pakistan Academy of Sciences or citizens of Pakistan. The journal is covered by Print and Online ISSN, indexed in Scopus, and distributed to scientific organizations, institutes and universities throughout the country, by subscription and on an exchange basis.

Annual Subscription: **Pakistan:** Institutions, Rupees 4000/-; Individuals, Rupees 2000/- (Delivery Charges: Rupees 150/-)
Other Countries: US$ 200.00 (includes air-lifted overseas delivery)

HEC Recognized, Category Y; Scopus Indexed

Published by **Pakistan Academy of Sciences**, 3 Constitution Avenue, G-5/2, Islamabad, Pakistan
Email: editor@paspk.org; **Tel:** 92-51-920 7140 & 921 5478; **Websites:** www.paspk.org/proceedings/; www.ppaspk.org

Printed at **Graphics Point.,** Office 3-A, Wasal Plaza, Fazal-e-Haq Road Blue Area Islamabad.
Ph: 051-2806257, **E-mail:** graphicspoint16@gmail.com

PROCEEDINGS OF THE
PAKISTAN ACADEMY OF SCIENCES: PART A
Physical and Computational Sciences

CONTENTS

Volume 60, No. 4, December 2023 **Page**

Review Article

Research Articles

Submission of Manuscripts: Manuscripts may be submitted as an e-mail attachment at editor@paspk.org or submit online at http://ppaspk.org/index.php/PPASA/about/submissions. Authors must consult the ***Instructions for Authors*** at the end of this issue or at the Website: www.paspk.org/proceedings/ or www.ppaspk.org.

Proceedings of the Pakistan Academy of Sciences: A Physical and Computational Sciences 60(4): 1-12 (2023)

ISSN (Print): 2518-4245; ISSN (Online): 2518-4253
https://doi.org/10.53560/PPASA(60-4)831

Pakistan Academy of Sciences

Review Article

A Supervised Machine Learning Algorithms: Applications, Challenges, and Recommendations

Aqib Ali[1,2*], and Wali Khan Mashwani[3]

[1]Department of Computer Science, Concordia College Bahawalpur, Pakistan
[2]Key Laboratory Measurement and Control of CSE Ministry of Education, School of Automation, Southeast University, Nanjing 210002, China
[3]Institute of Numerical Sciences, Kohat University of Science & Technology, Kohat 26000, Pakistan

Abstract: Machine Learning (ML) is an advanced technology that empowers systems to acquire knowledge autonomously, eliminating the need for explicit programming. The fundamental objective of the machine learning paradigm is to equip computers with the ability to learn independently without human intervention. In the literature, categorization in data mining has received a lot of traction, with applications ranging from health to astronomy and finance to textual classification. The three learning methodologies in machine learning are supervised, unsupervised, and semi-supervised. Humans must give the appropriate input and output and offer feedback on the prediction accuracy throughout the training phase for supervised algorithms. Unsupervised learning methods differ from supervised learning methods because they do not require any training. However, supervised learning methods are more accessible to implement than unsupervised learning methods. This study looks at supervised learning algorithms commonly employed in data classification. The strategies are evaluated based on their objective, methodology, benefits, and drawbacks. It is anticipated that readers will be able to understand the supervised machine learning techniques for data classification.

Keywords: Machine Learning, Supervised Learning, Classification, Supervised Algorithms.

1. INTRODUCTION

Machine learning (ML) is a broad term that encompasses computer science, statistics, probability, artificial intelligence, psychology, neuroscience, and various other fields [1, 2]. Problems may be solved using ML by simply creating a model that is a good representation of a given dataset. ML has enhanced the study of statistics by establishing fundamental computational statistical theories of learning processes [3]. It has also elevated the subject of statistics into a broad discipline by teaching computers to emulate the human brain. ML allows a system to learn and improve without being explicitly designed. ML algorithms are advantageous when explicitly coding for fast-speed performance is impractical. *Sorting numbers* is a significant task that involves providing a set of numbers as input and obtaining a sorted list as the output. We know input values and the appropriate algorithm to achieve the desired outcome [4]. However, specific tasks, like email filtering to differentiate between legitimate and spam messages, can be perplexing. Although we comprehend the required input and anticipated true or false output, the precise instructions guiding the programmer in executing these tasks are yet to be determined [5]. In unusual instances when there is no precise technique to accomplish success, we rely on data and direct the computer to evaluate and make intelligent sense of the data. Consider computer software that learns to identify and forecast cancer based on medical research reports from a patient. Your performance will improve as you experience interpreting medical research papers from a more comprehensive patient group [6]. Correct

Received: July 2023; Accepted: December 2023
*Corresponding Author: Aqib Ali <aqibcsit@gmail.com>

predictions and cancer case detections will be counted and evaluated by an experienced oncologist to determine your performance. ML is used in various fields, including robotics, natural language processing, computer games, pattern recognition, etc. [7]. ML methods involve training models on a dataset to learn patterns and subsequently applying these models to a separate test dataset for tasks such as classification or prediction. Figure 1 depicts the process of supervised ML algorithms.

1.1 ML Algorithm Categories

Supervised learning algorithms, supporting vector machines, decision trees, probabilistic summaries, algebraic functions, and other methods can be used to describe classifiers. Classification is one of the most researched models and possibly relevant in practice, alongside regression and probability estimation. Advances in this category have enormous potential advantages since the approach greatly influences other fields, both inside data mining and its applications [8].

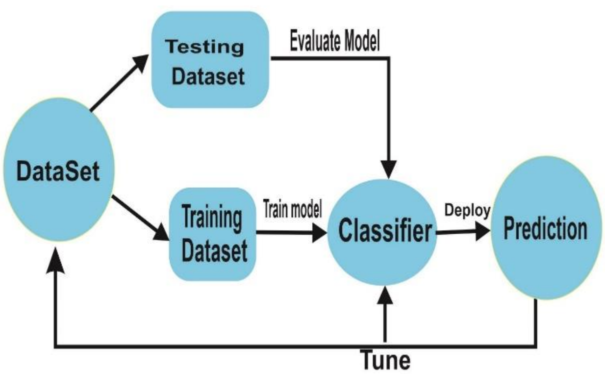

Fig. 1. Supervised Learning Workflow

Unsupervised learning algorithms [9], on the other hand, have no categories. The primary goal of unsupervised learning is to build grading labels automatically. These algorithms seek patterns in the data to see whether they can be categorized and grouped. These groupings represent the whole family of clustered ML approaches, referred to as clusters. The machine does not know how the collections are organized in this unsupervised categorization (cluster analysis). We have a better chance of surprising ourselves if we use cluster analysis. As a result, cluster analysis is a promising method for delving into the correlations between various variables. Figure 2 shows the categories of ML algorithms.

This article depicts many types of supervised ML algorithms and how they may be used to make more efficient judgments and complete tasks more efficiently. This article will demonstrate how different algorithms provide the machine with a varied learning experience and how the machine adopts other things from the environment before making a choice and doing specialized jobs.

2. SUPERVISED LEARNING

In a fundamental machine learning model, the learning process comprises two distinct phases: training and testing. During the training phase, the training data samples serve as inputs for the learning algorithm or learner, allowing it to acquire knowledge and construct the learning model [10]. Subsequently, in the testing phase, the learning model utilizes an execution engine to generate

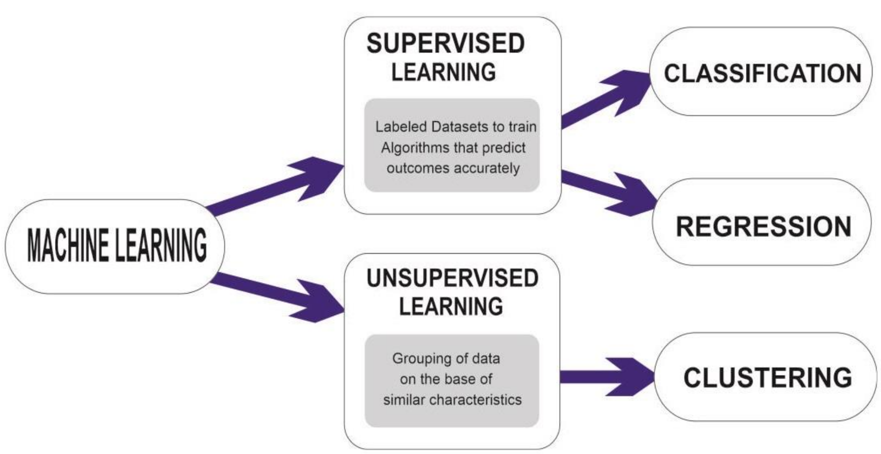

Fig. 2. Machine Learning Algorithm Categories

test or production data predictions. The output of the learning model consists of labeled data, which provides the final forecasts or categorized information. Supervised learning is the most commonly employed approach in classification problems, as it aims to create a classification system based on the constructed machine learning model.

In supervised learning, the input probabilities, such as inputs with known predicted results, are typically unspecified [11]. This process yields a dataset comprising labeled data and corresponding features. The primary objective is to develop an estimator that can determine an item's label based on a set of characteristics. The learning algorithm is supplied with inputs consisting of feature sets and their respective desired outputs. The algorithm learns from errors by comparing its actual output to the correct outputs and adjusts the model accordingly. However, if input values are missing, no reliable conclusions can be drawn about the outputs. Supervised learning is widely employed in training neural networks and decision trees, which rely on labeled data for default categorization [12]. This learning approach is also utilized in programs that predict probable occurrences of functions based on historical data. Its practical applications are vast, including tasks such as inferring the species of an iris flower based on measurements. As mentioned earlier, supervised learning tasks can be categorized into classification and regression. While regression deals with continuous labels, classification involves discrete labels [13].

The approach described in Figure 3 distinguishes between the observed data X, which serves as the training data, and the structured data presented to the model during the training phase. The objective of the supervised learning algorithm is to create a prediction model through this process. Once trained, the fitted model will attempt to predict the most likely labels for a new set of samples X in the test set. Based on the nature of the objective and the type of labels involved, supervised learning can be categorized as follows:

- **Classification:** Classification is termed as such when the objective is to predict the value of y from a predetermined set of classes. The prediction model is trained to assign the most appropriate class label to each input sample.
- **Regression:** The challenge lies in predicting the value y when it consists of floating-point values. The prediction model is designed to estimate continuous numerical values rather than discrete classes.

These two categories, classification, and regression,

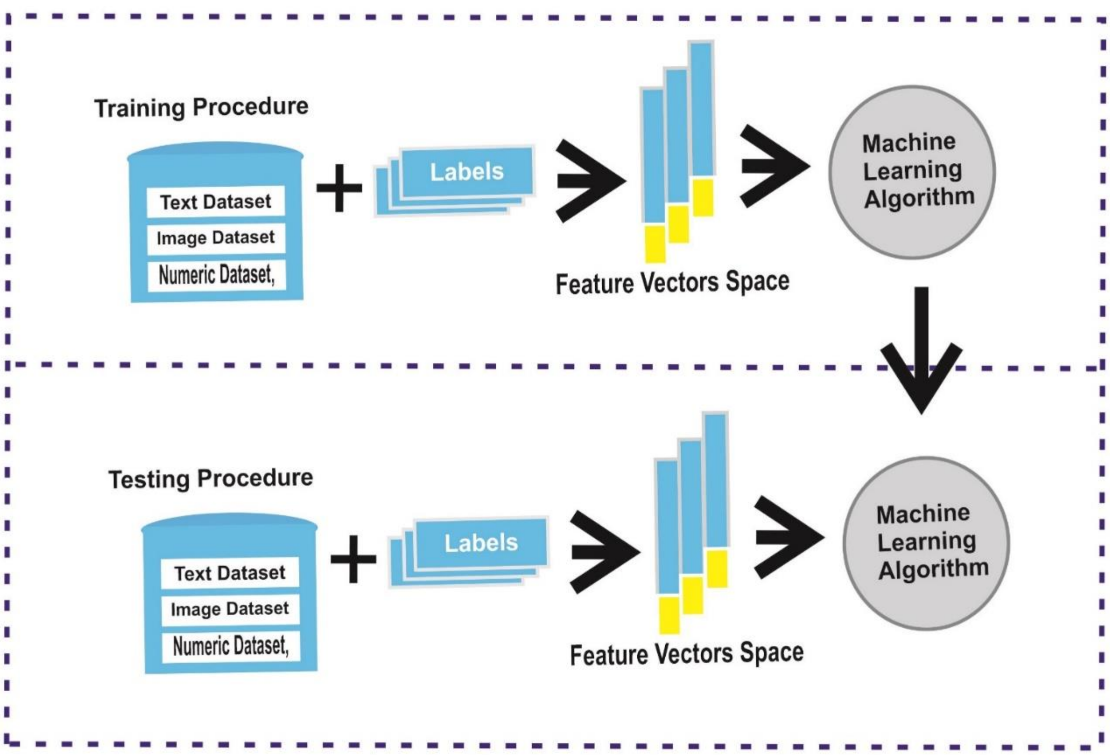

Fig. 3. Supervised Learning Model

capture the primary objectives of supervised learning, depending on the predicted label type.

3. REVIEW ON ML SUPERVISED ALGORITHM

We undertake a thorough analysis of the literature on supervised learning methodologies and algorithms [14-16], performance measures used in supervised learning, and the benefits and drawbacks of eight supervised learning research in this paper, as sown in Figure 4. This research will help academics go in a new direction by identifying new research areas and filling a research vacuum in supervised learning.

3.1 Bayesian Network

The non-dependence on functions of the BN classifier is a crucial property. Another factor to consider is that all operations are interdependent. Returns the BN model as a directed acyclic graph with random variables and conditional dependencies as nodes and edges. It is assumed to be a complete model for the variable and its connection. As a result, a total joint probability distribution (JPD) is defined for all variables in a model, as shown in Figure 5 [17].

3.1.1 Advantages

In Bayesian learning, a priori probability distribution is initially selected and subsequently updated to generate a posteriori distribution. As additional observations become available, the a posteriori distribution can be utilized as a new priori distribution for further updates. The Bayesian network can deal with incomplete datasets. The strategy can prevent data from being over-fitted. It is not essential to eliminate the data's inconsistencies [18].

3.1.2 Disadvantage

The following are some of the drawbacks of Bayesian learning: preselection is a challenging task. The anterior distribution might have an important influence on the posterior distribution. If the preceding option is wrong, the forecasts will be erroneous. It has the potential to be computationally demanding. Text extraction is not suggested since the BN classifier is computationally demanding [19].

3.1.3 Applications

Bayesian learning may be utilized in various areas, including medical diagnosis and disaster victim identification.

3.2 Naive Bayes

Naïve Bayes (NB) classification [20] is a supervised learning method that shares similarities with the

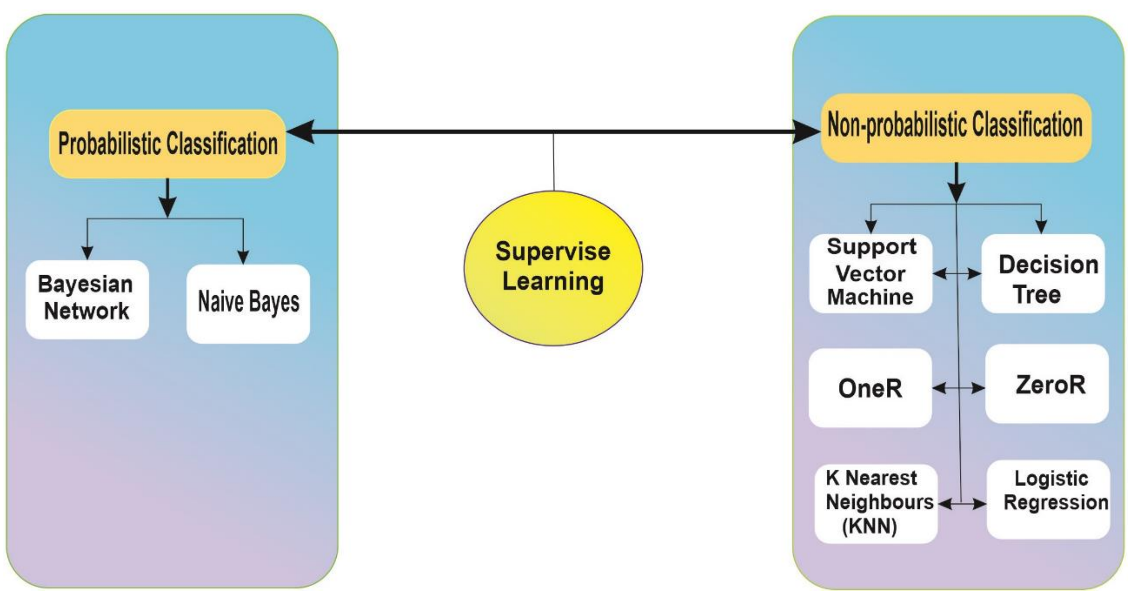

Fig. 4. Supervised Learning Algorithm Review

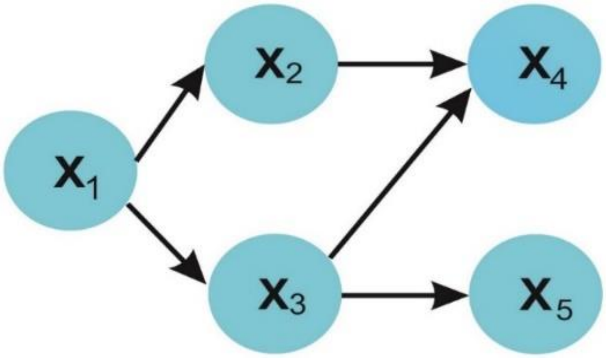

Fig. 5. Visualization and Workflow of a Bayesian Network

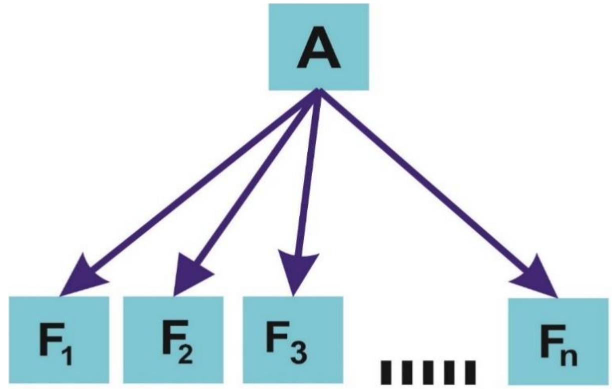

Fig. 6. Visualization and Workflow of a Naive Bayes Algorithm

statistical approach to classification. It is based on a fundamental probabilistic model, which allows for representing uncertainty in the model by computing outcome probabilities. The primary objective of Bayesian classification is to address prediction tasks, and it incorporates observed data while employing effective learning techniques. One of the advantages of Bayesian classification is that it can facilitate the understanding and evaluation of learning algorithms [21]. The graphical representation and primary usage of the NB classifier can be seen in Figure 6. By calculating explicit probabilities for the assumptions and utilizing incoming data to refine the model, the NB classification approach aims to make accurate predictions.

In the context of a two-valued generic probability distribution, without sacrificing generality, we can derive the following equation using Bayes' rule:

$$P(y_1, y_2) = P(y_1 \mid y_2)P(y_2) \quad (1)$$

If another variable of class c is present, we get the following equation:

$$P\ (y_1 \mid y_2 \mid c)\ =\ P\ (y_1 \mid y_2, c)\ (y_2 \mid c)\ P \quad (2)$$

When we extend the scenario to include two variables and assume conditional independence for a collection of variables Y1, . . ., Yn that depends on another variable c, we obtain the following equation:

$$P(y \mid c) = \prod_{i=i}^{N} P(y_i \mid c) \quad (3)$$

3.2.1 Advantages

Naive Bayes (NB) offers several benefits, including its straightforward implementation, efficient performance, and ability to achieve reliable results with smaller training datasets. The growth of predictors and data points remains linear, scaling appropriately with their numbers. NB is versatile, accommodating binary and multi-class classification tasks while generating probabilistic predictions. Furthermore, it handles many data types, encompassing continuous and discrete variables [22].

3.2.2 Disadvantages

The drawbacks of Naive Bayes include its inherent simplicity, which makes it prone to being outperformed by more complex and accurately fitted models. Directly using Naive Bayes becomes challenging when dealing with features like continuous variables (e.g., time). Although approximations like creating "cubes" for continuous variables are possible, they are not entirely accurate. Moreover, the absence of an online Naive Bayes variant necessitates saving all data for retraining, which can be impractical. The scalability of Naive Bayes is limited, especially when the number of classes is substantial, exceeding 100,000. In such cases, resizing becomes problematic. Even for predictions, alternatives like Support Vector Machines (SVM) or basic logistic regression tend to consume more runtime RAM. Additionally, computations can be time-consuming, particularly for models with many variables. [23].

3.2.3 Applications

Naive Bayes may be used in seed classifications, tumor classification, and stock exchange predictions.

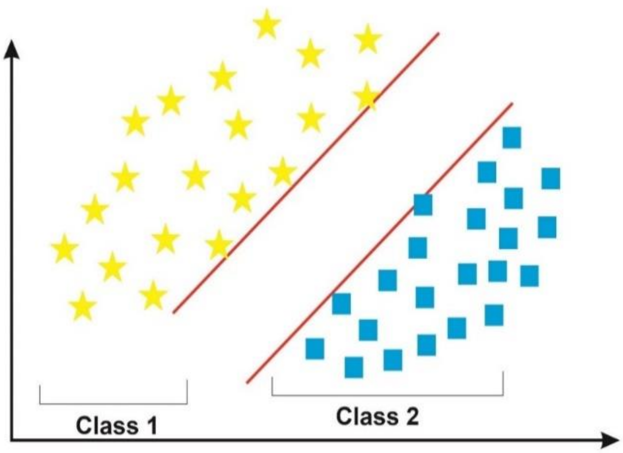

Fig. 7. Visualization and Workflow of a Support Vector Machine

3.3 Support Vector Machine (SVM)

The Supporting Vector Machines (SVM) is a sort of algorithm that can solve classification and regression issues. The hyper-plane, which serves as the decision boundary, must be defined in this technique. When a collection of components from distinct classes has to be separated, judgments must be made. Whether or not items can be divided linearly, complicated mathematical functions known as kernels are necessary to separate objects from distinct classes, as shown in Figure 7. SVM is used to properly categorize things using examples from the training dataset [24].

3.3.1 Advantages

SVM offer several advantages. They are adept at handling both structured and semi-structured data, particularly when a suitable kernel function can be determined. It is scalable with massive datasets and does not get trapped in local optimality.

3.3.2 Disadvantage

The following are some of SVM's drawbacks: with a vast data collection, its performance suffers because of the bigger training time. The proper kernel function will not be easy to find. The SVM does not perform well on noisy datasets. SVM does not offer probability estimates. The final SVM model is hard to understand.

3.3.3 Applications

The final SVM model is challenging to comprehend. Cancer diagnosis, credit card fraud finding, handwriting recognition, face recognition, and text categorization are all support vector machine applications.

3.4 Decision Tree

A decision tree serves as a recursive partitioning of the instance space, functioning as a classifier [25]. It comprises various nodes, with the root node being

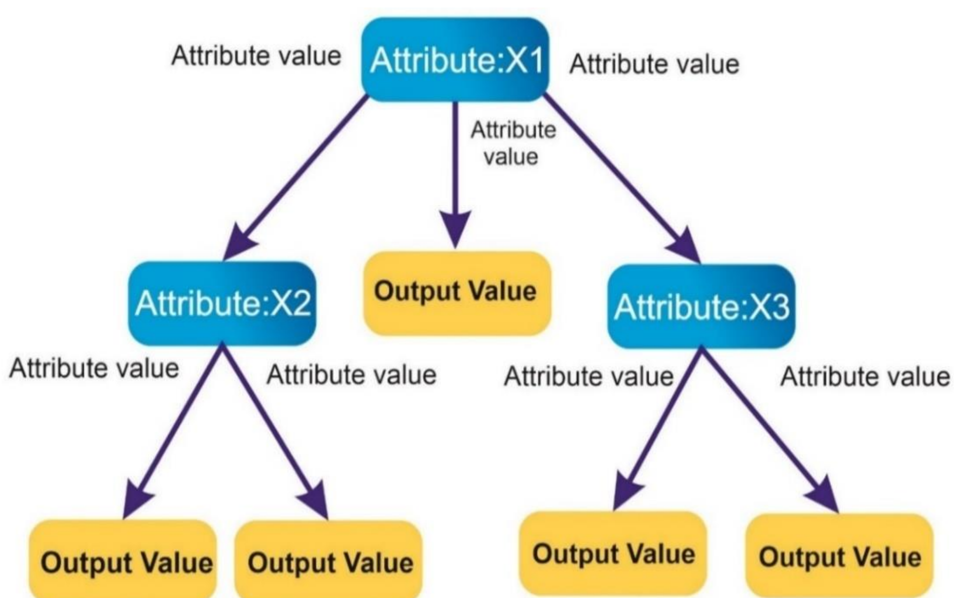

Fig. 8. Visualization and Workflow of a Decision Tree Algorithm

the primary node and the remaining nodes forming the tree structure. Internal or test nodes possess outgoing branches, while leaves represent terminal nodes. In a decision tree, each test node divides the instance space into multiple subspaces based on discrete functions of the input values. Typically, a single attribute is considered in each test, leading to the division of the instance space according to attribute values. For numerical properties, ranges are used as criteria. Each leaf node is assigned to the class with the highest objective value, and the likelihood of a specific target property value can be represented through a probability vector. Based on the test results, instances are categorized along the path from the root to the leaf node. The practical utility of decision trees is demonstrated in Figure 8, where each node is labeled with the checked attribute, and the corresponding attribute values are marked on the branches. By employing this classifier, analysts can predict the responses of elite prospects and analyze the behavioral characteristics of the entire prospect population [26].

Mathematically, decision trees can be understood as collections of hyperplanes orthogonal to the axes, particularly in the case of numeric attributes. Decision-makers often prefer less complex decision trees as they are perceived to be more comprehensive.

3.4.1 Advantages

Here are some advantages of using a Decision Tree:

- Suitable for Classification and Regression: Decision Trees are well-suited for classification and regression problems. They can effectively handle tasks that require predicting categorical outcomes or estimating numerical values.
- Interpretability: Decision Trees offer a straightforward and intuitive interpretation. The tree structure allows for clear visibility into the decision-making process, making it easier to understand and explain the reasoning behind predictions.
- Handling Categorical and Quantitative Values: Decision Trees can handle categorical and quantitative input variables. They can accommodate various data types without extensive data preprocessing or feature engineering.
- Efficient Performance: Decision Trees often demonstrate good performance due to the efficiency of the tree traversal technique. Navigating through the tree to make predictions is computationally efficient, enabling quick decision-making for new instances.

Additionally, when it comes to addressing the overfitting problem, Decision Trees can benefit from ensemble modeling techniques. Random Forest, in particular, is a popular solution that utilizes multiple Decision Trees to mitigate overfitting issues [27].

3.4.2 Disadvantages

The decision tree has the drawbacks of being unbalanced, hard to regulate the scope of the tree, prone to sampling mistakes, and providing a locally optimum answer rather than an overall optimal result [28].

3.4.3 Applications

Decision trees may be utilized in various applications, including forecasting future library book usage.

3.5 OneR

"A One-R rule" is a ranking method that generates a unique rule for each of the predictors in the data before selecting the "unique rule" with the fewest overall error. It's a straightforward and accurate algorithm. Create a frequency table for each predictor against their objectives to develop a rule for the estimator. OneR is virtually as accurate as state-of-the-art classification algorithms and is simple to understand [29].

- The One-R algorithm counts the frequency of each target value first (class).
- Choose the category more frequently.
- The predictor is assigned to that class by the rule.
- The total inaccuracy of the regulations for each estimator is then calculated.
- Choose the estimate with the minor gross error.

3.5.1 Advantages

High performance on non-linear problems, not affected by outliers, and not overfitting sensitive. Effective, unaffected by outliers, work on non-linear problems and takes a probabilistic approach [30].

3.5.2 Disadvantage

The assessment tables (Gain, Lift, K-S, and ROC) are not applicable because OneR does not provide a score or probability [30].

2.5.3 Applications

OneR, known for its simplicity and interpretability, is versatile in educational contexts, teaching fundamental ML concepts and serving as a foundational model for complex classification tasks.

3.6 ZeroR

It is one of the most basic classification approaches since it is goal-oriented and ignores all predictors. In basic terms, the “zero rule” categorization technique asks you to predict which class has the most members. The “zero rule” can be used to analyze the performance of a benchmark as the baseline for specific different ranking techniques, albeit it is not predictive. The Zero-R technique is straightforward: construct a table of frequencies and select the most frequent value [31].

3.6.1 Advantages

ZeroR is the most straightforward goal-based classification algorithm that ignores all predictors. The majority category is anticipated by the ZeroR classifier (class) even though ZeroR has no predictive power [31].

3.6.2 Disadvantage

The assessment tables (Gain, Lift, K-S, and ROC) are not applicable because ZeroR does not provide a score or probability.

3.6.3 Applications

It helps to establish a benchmark’s performance for other ranking systems.

3.7 K Nearest Neighbor

The K Nearest Neighbor (KNN) technique is used to classify data. The method attempts to classify the supplied sample data point as a classification problem utilizing a database with data points sorted into multiple groups, as shown in Figure 9. Because KNN assumes no underlying data distribution, it is called non-parametric [32].

3.7.1 Advantages

Here are several benefits of the KNN algorithm: it offers a straightforward approach that is easy to implement. Constructing the model is cost-effective. KNN is a highly flexible classification system that performs well with multimodal classes. It can handle various class labels within the records. While its error rate can be double that of Bayesian error, KNN is often considered the most effective method. In predicting protein function based on the expression profile, KNN has been found to outperform SVM. [33].

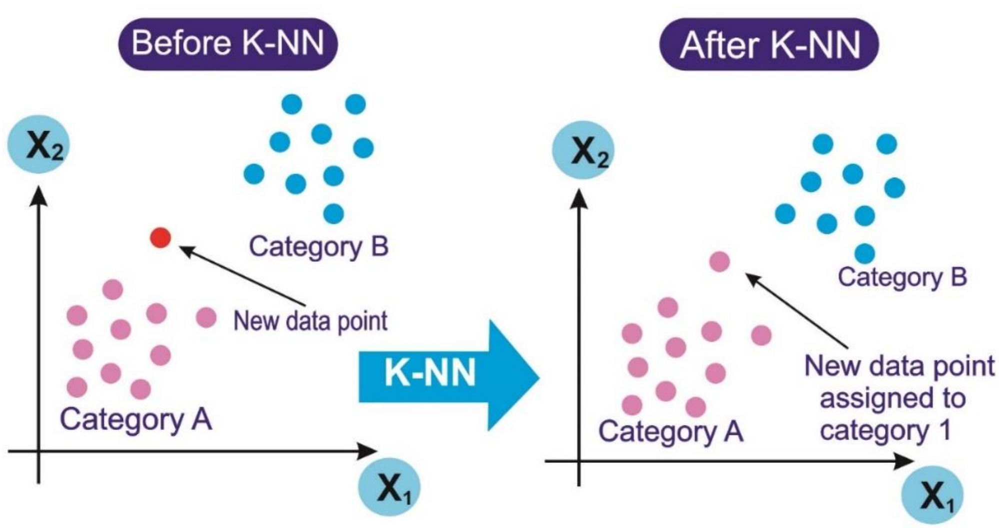

Fig. 9. Visualization and Workflow of a K Nearest Neighbor

3.7.2 Disadvantage

The following are some of KNN's drawbacks: the cost of classifying unknown records is relatively high. Calculation of the nearest k-neighbor distance is required. The method becomes more computationally costly as the training set size grows. Accuracy will suffer as a result of noisy or irrelevant attributes. It is worth noting that KNN can be computationally expensive, particularly for large datasets. Despite these drawbacks, KNN is a method that preserves all training data and makes no assumptions about it. This characteristic allows it to handle enormous datasets, albeit with the trade-off of potentially costly computations.

3.7.3 Applications

The precision of the areas will be reduced as the data becomes more dimensional. KNN has applications in recommendation systems, medical analysis of different diseases with similar symptoms, credit rating based on resemblance of traits, and handwriting recognition [33].

3.8 Logistic Regression

It's a tried-and-true method of resolving binary classification issues. The likelihood that a result has just two values is predicted using logistic regression. The logistic function, an S-shaped curve that takes any number with an accurate weight and transfers it to a value between 0 and 1, but never precisely between 0 and 1, is at the heart of logistic regression. [34].

The logistic regression utilizes data to fit a logistic function and calculate the probability that an event will occur [35]. Logistic regression uses several predictive variables, some of which may be numerical or categorical, like many other forms of regression analysis. As stated by the logistic regression hypothesis:

$$g(y) = g(\theta^T y) \quad (4)$$

where g is a sigmoid function defined as follows:

$$g(x) = \frac{1}{1+e^{-x}} \quad (5)$$

As shown in Figure 10, the sigmoid function has

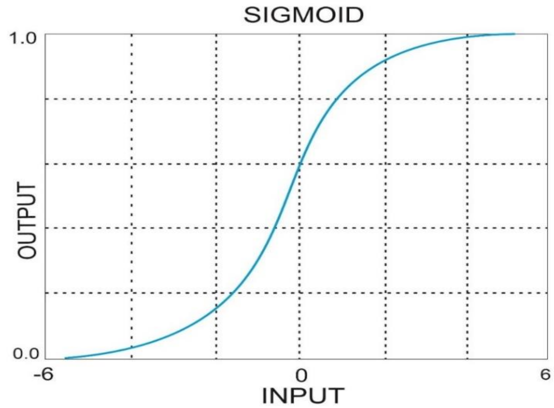

Fig. 10. The logistic function as a visual depiction

specific features that result in the range [0.1] values. The logistic regression cost function is as follows:

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\left[-z^{(i)}\log\left(g_\theta\left(y^{(i)}\right)\right) - \left(1 - z^{(i)}\right)\log\left(1 - g_\theta\left(y^{(i)}\right)\right)\right] \quad (6)$$

3.8.1 Advantages

Implementation ease, computational efficiency, training efficiency, and regularization ease are only a few of the benefits of logistic regression. There is no need to scale the input functions. This strategy is mainly utilized in the industry to address large-scale difficulties. Because the logistic regression output is a probability score, using it to solve a business problem necessitates the creation of unique performance indicators to obtain a breakpoint that can be used to prioritize the objective. Furthermore, modest data noise and multicollinearity have little effect on logistic regression [36].

3.8.2 Disadvantage

The following are some of the drawbacks of logistic regression: because its decision surface is linear and vulnerable to over-fitting, it cannot address nonlinear issues. It will not operate successfully until all independent variables are known.

3.8.3 Applications

Practical uses of Logistic Regression include predicting the risk of acquiring the disease and diagnosing cancer.

4. COMPARATIVE ANALYSIS

We present a comparison of the most popular

Table 1: A Comparison of Eigth Commonly Used Supervised Classification Algorithms

	Naive Bayes	Decision Tree	Logistic Regression	Support Vector Machine	K Nearest Neighbor	Bayesian Network	OneR	ZeroR
Accuracy in General	Satisfactory	Good	Good	Superb	Good	Excellent	Good	Good
Speed of Learning	Superb	Excellent	Satisfactory	Satisfactory	Superb	Superb	Excellent	Satisfactory
Speed of Classification	Superb	Superb	Good	Superb	Satisfactory	Superb	Superb	Good
Tolerance to Missing Values	Superb	Excellent	Superb	Good	Satisfactory	Excellent	Good	Good
Tolerance to Irrelevant Attributes	Good	Excellent	Good	Superb	Good	Good	Excellent	Excellent
Tolerance to Redundant Attributes	Satisfactory	Good	Excellent	Excellent	Good	Good	Satisfactory	Satisfactory
Tolerance to Highly Interdependent Attributes	Satisfactory	Good	Excellent	Excellent	Satisfactory	Good	Good	Satisfactory
Tolerance to Noise	Excellent	Good	Satisfactory	Good	Satisfactory	Good	Satisfactory	Good
Dealing with Danger of Overfitting	Excellent	Good	Superb	Good	Excellent	Superb	Satisfactory	Satisfactory
Attempts for Incremental Learning	Superb	Good	Superb	Good	Superb	Good	Good	Satisfactory
Transparency of Knowledge/ Classification	Superb	Superb	Good	Satisfactory	Good	Superb	Satisfactory	Good
Support Multiclassi-fication	Naturally Extended	Superb	Good	Binary Classifier	Superb	Excellent	Good	Satisfactory

supervised classification algorithms. Several strategies have been created, some of which have been addressed in earlier sections. Based on available facts and theoretical studies, Table 1 compares various regularly used supervised algorithms. This comparison demonstrates that no single learning algorithm beats the others.

5. CONCLUSIONS

This article aimed to look at some of the most often used supervised machine learning methods for regression, classification, and clustering. The benefits and drawbacks of various algorithms were highlighted, as in terms of performance, learning speed, and other factors, several algorithms are compared (where possible). In addition, examples of how these algorithms may be used in the real world were addressed. Machine learning approach such as supervised learning was considered. It is hoped that by recognizing the various machine learning algorithm possibilities, subsequently, in the precise troubleshooting situation, selecting the appropriate machine learning algorithm, the readers would be able to make an educated decision.

6. CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

7. REFERENCES

1. B. Dou, Z. Zhu, E. Merkurjev, L. Ke, L. Chen, J. Jiang, and G.W. Wei. Machine learning methods for small data challenges in molecular science. *Chemical Reviews* 123(13): 8736-8780 (2023).
2. A. Ali, S. Naeem, S. Anam, and M.M. Ahmed. A Supervised Machine Learning Algorithms: Applications, Challenges, and Recommendations. *International Journal of Computing and Digital Systems* 13(1): 1-10 (2023).
3. M. Zubair. Machine Learning Based Biomedical Image Analysis and Feature Extraction Methods. *Journal of Applied and Emerging Sciences* 13(1): 31-39 (2023).
4. J.G. Greener, S. M. Kandathil, L. Moffat, and D.T. Jones. A guide to machine learning for biologists. *Nature Reviews Molecular Cell Biology* 23(1): 40-55 (2022).
5. N. Burkart, and M.F. Huber. A survey on the explainability of supervised machine learning. *Journal of Artificial Intelligence Research* 70: 245-317 (2021).
6. J. Ranjan, and C. Foropon. Big data analytics in building the competitive intelligence of organizations. *International Journal of Information Management* 56: 102231 (2021).
7. X. Xiang and S. Foo. Recent Advances in Deep Reinforcement Learning Applications for Solving Partially Observable Markov Decision Processes (POMDP) Problems: Part 1—Fundamentals and Applications in Games, Robotics and Natural Language Processing. *Machine Learning and Knowledge Extraction* 3(3): 554-581 (2021).
8. T. Jiang, J.L. Gradus, and A.J. Rosellini. Supervised machine learning: a brief primer. *Behavior Therapy* 51(5): 675-687 (2020).
9. A. Roohi, K. Faust, U. Djuric, and P. Diamandis. Unsupervised machine learning in pathology: the next frontier. *Surgical Pathology Clinics* 13(2): 349-358 (2020).
10. Y. Liu, S. Arunachalam, and K. Temme. A rigorous and robust quantum speed-up in supervised machine learning. *Nature Physics* 17(9): 1013-1017 (2021).
11. I.H. Sarker. Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science* 2(3): 1-21 (2021).
12. S. Laghmati, B. Cherradi, A. Tmiri, O. Daanouni, and S. Hamida. Classification of patients with breast cancer using neighbourhood component analysis and supervised machine learning techniques. *3rd International Conference on Advanced Communication Technologies and Networking (CommNet)* pp. 1-6, IEEE (September 2020).
13. L.J. Muhammad, E.A. Algehyne, and S.S. Usman. Predictive supervised machine learning models for diabetes mellitus. *SN Computer Science* 1(5): 1-10 (2020).
14. A. Rawson, and M. Brito. A survey of the opportunities and challenges of supervised machine learning in maritime risk analysis. *Transport Reviews* 43(1): 108-130 (2023).
15. J.A. Pruneski, A. Pareek, K.N. Kunze, K.R. Martin, J. Karlsson, J.F. Oeding, and R.J. Williams III. Supervised machine learning and associated algorithms: applications in orthopedic surgery. *Knee Surgery, Sports Traumatology, Arthroscopy* 31(4): 1196-1202 (2023).
16. A. Dinesh, S.A. Selvasofia, K.S. Datcheen, and D.R. Varshan. Machine learning for strength evaluation of concrete structures–Critical review. *Materials Today: Proceedings* (2023).
17. S. Gil-Begue, C. Bielza, and P. Larrañaga. Multi-dimensional Bayesian network classifiers: A survey. *Artificial Intelligence Review* 54(1): 519-559 (2021).
18. G.A. Ruz, P.A. Henríquez, and A. Mascareño. Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers. *Future Generation Computer Systems* 106: 92-104 (2020).
19. P. Awasthi, N. Frank, and M. Mohri. On the existence of the adversarial bayes classifier. *Advances in Neural Information Processing Systems* 34: 2978-2990 (2021).
20. Z. Xue, J. Wei, and W. Guo. A real-time Naive Bayes classifier accelerator on FPGA. *IEEE Access* 8: 40755-40766 (2020).
21. K.G. Reddy, and P.S. Thilagam. Naïve Bayes classifier to mitigate the DDoS attacks severity in ad-hoc networks. *International Journal of Communication Networks and Information Security* 12(2): 221-226 (2020).
22. I.E. Tiffani. Optimization of naïve bayes classifier by implemented unigram, bigram, trigram for sentiment analysis of hotel review. *Journal of Soft Computing Exploration* 1(1): 1-7 (2020).
23. A. Kurani, P. Doshi, A. Vakharia, and M. Shah. A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting. *Annals of Data Science* 10(1): 183-208 (2023).
24. M. Tanveer, T. Rajani, R. Rastogi, Y.H. Shao, and M.A. Ganaie. Comprehensive review on twin support vector machines. *Annals of Operations Research*: pp. 1-46 (2022).
25. P. Priyanka, and D. Kumar. Decision tree classifier: A detailed survey. *International Journal of Information and Decision Sciences* 12(3): 246-269 (2020).
26. S.H. Yoo, H. Geng, T.L. Chiu, S.K. Yu, D.C. Cho, J. Heo, and H. Lee. Deep learning-based decision-tree classifier for COVID-19 diagnosis from chest X-ray imaging. *Frontiers in Medicine* 7: 427 (2020).
27. A. Ali, S. Naeem, S. Anam, and M.M. Ahmed. Current State of Artificial Intelligence (AI) in Oncology: A Review. *Current Trends in OMICS* **3**(1): 01-17 (2023).
28. O. Mehrpour, C. Hoyte, F. Goss, F.M. Shirazi, and S. Nakhaee. Decision tree algorithm can determine the outcome of repeated supratherapeutic ingestion

(RSTI) exposure to acetaminophen: review of 4500 national poison data system cases. *Drug and Chemical Toxicology* 46(4): 692-698 (2023).
29. Y.T. Chang, and N.H. Fan. A novel approach to market segmentation selection using artificial intelligence techniques. *The Journal of Supercomputing* 79(2): 1235-1262 (2023).
30. M. Farahani, S.V. Razavi-Termeh, A. Sadeghi-Niaraki, and S.M. Choi. A Hybridization of Spatial Modeling and Deep Learning for People's Visual Perception of Urban Landscapes. *Sustainability* 15(13): 10403 (2023).
31. M.U. Abdulazeez, W. Khan, and K.A. Abdullah. Predicting child occupant crash injury severity in the United Arab Emirates using machine learning models for imbalanced dataset. *IATSS Research* 47(2): 134-159 (2023).
32. P. Cunningham, and S.J. Delany. k-Nearest neighbour classifiers-A Tutorial. *ACM Computing Surveys (CSUR)* 54(6): 1-25 (2021).
33. C. Li, S. Ding, X. Xu, H. Hou, and L. Ding. Fast density peaks clustering algorithm based on improved mutual K-nearest-neighbor and sub-cluster merging. *Information Sciences* 647: 119470 (2023).
34. Q. Li, S. Zhao, S. Zhao, and J. Wen. Logistic Regression Matching Pursuit algorithm for text classification. *Knowledge-Based Systems* 277: 110761 (2023).
35. G. Troiano, L. Nibali, H. Petsos, P. Eickholz, M.H. Saleh, P. Santamaria, and A. Ravidà. Development and international validation of logistic regression and machine-learning models for the prediction of 10-year molar loss. *Journal of Clinical Periodontology* 50(3): 348-357 (2023).
36. C. Hercus, and A.R. Hudaib. Delirium misdiagnosis risk in psychiatry: a machine learning-logistic regression predictive algorithm. *BMC Health Services Research* 20(1): 1-7 (2020).

Proceedings of the Pakistan Academy of Sciences: A Physical and Computational Sciences 60(4): 13-28 (2023)

ISSN (Print): 2518-4245; ISSN (Online): 2518-4253
https://doi.org/10.53560/PPASA(60-4)824

Pakistan Academy of Sciences

Research Article

Monitoring Drought Events and Vegetation Conditions in Pakistan: Implications for Drought Management and Food Security

Rabia Tabassum[1,4]*, Imran Ahmed Khan[2], and Mudassar Hassan Arsalan[3]

[1] Department of Computer Science, National University of Computer and Emerging Sciences (FAST), Karachi, Pakistan
[2] Department of Geography, University of Karachi, Pakistan
[3] School of Computing, Engineering and Mathematics, Western Sydney University, Parramatta South Campus, Rydalmere, NSW 2116, Australia
[4]Institute of Space Science and Technology, University of Karachi, Pakistan

Abstract: Drought is an environmental and humanitarian concern, affecting various regions and landscapes with detrimental impacts on the natural environment and human lives. It is crucial to have access to accurate and timely information on vegetation conditions to mitigate its effects, which is possible through remote sensing techniques. The Google Earth Engine (GEE) platform offers powerful tools and a vast collection of geospatial data that can improve drought monitoring efforts. In Pakistan, a country prone to droughts and natural disasters, this study employs GEE to monitor drought events over croplands and determine their severity using indices such as Vegetation Health Index (VHI), Vegetation Condition Index (VCI), and Temperature Condition Index (TCI). Additionally, the study generates mean yearly and monthly VHI maps, allowing for the observation of trends in drought occurrences over time. Analysis of yearly and monthly mean VHI maps reveals that in 2001 and 2002, the Pakistan cropland experienced severe to moderate drought. From 2011 to 2021, overall drought occurrences were relatively low, and the annual VHI reflected healthy vegetation conditions. The study emphasizes the adequacy of vegetation in 2019, 2020, and 2021. Notably, agricultural areas in Punjab demonstrate sufficient soil moisture and healthy vegetation, while in Sindh, cropland is predominantly affected by severe to moderate drought, characterized by a continuous deficiency of soil moisture. Temporal variations highlight an increasing trend in VHI, indicating a promising outlook for healthy vegetation in Pakistan's cropland. This study provides valuable information for drought management, planning, water resource management, food production, and food security.

Keywords: Drought, Remote Sensing, Google Earth Engine, Normalized Difference Vegetation Index, Vegetation Health Index, Food Production and Food Security.

1. INTRODUCTION

Drought is a creeping disaster that processes with accumulative effects and is particularly a slow occurrence phenomenon according to time durations such as months or weeks [1]. Since the 1970s, a global drying trend has been observed, affecting numerous regions, particularly in high northern latitudes [2, 3]. Drought is usually monitored through the vegetation indices derived from satellite images. The "index" determination is a process of developing "value added" information connected to drought with the help of current and past knowledge [4, 5]. Indices, which are drought indicators, are obtained to measure drought and its severity. Drought indices provide more information instead of comparing current and past situations and identifying water scarcity related to a drought period and intensity [6]. Zhao *et al.* [7] assessed drought conditions by rebuilding the Land Surface Temperature (LST) using the Annual Temperature Cycle (ATC) model and drought indices (TCI, VCI, VH, and Temperature-Vegetation Drought Index (TVDI)). Khan and Arsalan [8] assessed drought conditions using Normalized Difference Vegetation Index (NDVI) and found that NDVI is positively

Received: May 2023; Accepted: December 2023
*Corresponding Author: Rabia Tabassum <rabia.tabassum@nu.edu.pk>

correlated with rainfall and soil moisture. Whereas, Kogan [9] used NDVI to derive another index called VCI. Particularly, VCI identifies the spatio-temporal deviations of vegetation situations related to stress due to precipitation shortage. Zambrano *et al.* [10], Kamble *et al.* [11], and Bento *et al.* [12] assessed drought using VCI and compared it with SPI (Standardized Precipitation Index); the result showed a good correlation between VCI and SPI. LST is used to derive TCI and various studies used TCI with VCI for drought monitoring [13]. The composite of TCI and VCI gives VHI, a widely used drought index for drought monitoring applications. Tabassum *et al.* [14] assessed the drought condition in an arid region of Pakistan using VCI, TCI, and Temperature Vegetation Index (TVX). TVX, the ratio of LST and NDVI, provides more spectral information for drought monitoring because it is derived from both reflective and thermal bands [15]. TVX is also used to extract soil moisture [16]. Masitoh and Rusydi [17] and Ma'Rufah *et al.* [18] used VHI to characterize vegetation health.

ArcGIS is a famous software for handling spatial and temporal data, and all of the above studies used it [10-14]. After the popularity of cloud computing, spatiotemporal data processing with downloading is easily possible on the GEE Platform. GEE with cloud computing remote sensing application is an appropriate platform to detect a region with a large amount of data [19]. GEE performs planetary-scale geospatial analysis with the massive computational ability to identify various challenging socio-economic issues like deforestation, drought, hazards, food insecurity, water supply and distribution, climate monitoring, and conservation of the environment [20]. GEE collectively can investigate environmental conditions, agriculture, soil, water, and various climate variables and map observable changes, trends, and variations [20, 21]. Wang *et al.* [22] used GEE for monitoring the long-term surface water changes of lakes [22]. Thilagaraj *et al.* [23] developed the system through the GEE platform that effectively monitors drought events in different time durations using the remote sensing indices such as LST, VCI, NDVI, TCI, and VHI. Many research studies considered the GEE in agricultural mapped and monitor, and used different drought indices for drought conditions [19]. Drought assessment needs long-duration data with good coverage of condition. GEE can fulfill this requirement and provides global coverage with a bulk amount of geospatial data.

Food security and rural development depend on the agriculture sector's continued expansion. The agriculture sector in Pakistan contributes the most to the country's GDP about 22.7 % and it employs approximately 37.4% of the labor force [24]. In the current period of April-October 2023, it is estimated that approximately 10.52 million rural people in Pakistan, accounting for 29 percent of the analyzed 36.7 million rural population, are in IPC Phase 3 (Crisis) and IPC Phase 4 (Emergency) [25]. This population has been impacted by a number of shocks, including drought, livestock diseases, high food prices, and the COVID-19 pandemic's effects [26]. The country's arid and semi-arid regions suffer the most from the effects of drought, which include crop failure and decreased food production [27-29]. A comprehensive drought assessment for agricultural land is required to lessen these effects and increase food security.

Many studies have assessed the drought condition in Pakistan [15]. Ullah et al. [27] used in-situ observations and drought indices (SPI and SPEI) and reanalysis results to evaluate Pakistan's meteorological drought characteristics. Ali et al. [30] observed that the temporal rainfall evaluation in Pakistan indicates an average reducing trend at the country and provisional levels. The identified lowering rainfall should strengthen the probability of droughts, thereby affecting the agricultural region [30]. Ahmed et al. [29] predicted that rising temperatures caused by global warming would exacerbate the intensity and frequency of droughts in Pakistan. Jamro, et al. [28] monitored the spatio-temporal variability of drought in Pakistan and found that the droughts of 1920 and 2000, which affected all zones and persisted for over ten months in three, can be considered as among the most severe dry spells in Pakistan's history. Khan et al. [31] used machine learning for drought prediction with the help of SPEI. Pasha et al. [32] examined the reason for Sindh Drought in 2014 and concluded that it was a famine-like situation brought on by a prolonged lack of precipitation in the Thar Desert. With a large amount of geospatial data and global coverage, GEE can meet this requirement. Khan et al. [33] employed the GEE platform, incorporating diverse drought indices such as SPEI, SPI, VCI,

Table 1. List of the GEE Dataset with Product Name, Temporal Resolution, Spatial Resolution and, Time Duration

Products Name	Data Type	Spatial Resolution	Temporal Resolution	Time Duration
MOD13Q1	NDVI	250m	16-days	2001-2021
MOD11A2	LST	1km	8-days	2001-2021
GFSAD1000	Land-cover	1km		

TCI, PCI (Precipitation Condition Index), and SMCI (Soil Moisture Condition Index), to evaluate drought conditions in the non-irrigated region of Potohar plateau, Punjab, Pakistan. Notably, there is a shortage of studies utilizing GEE for drought monitoring, particularly in Pakistan's croplands. Effective drought monitoring is crucial for adequately preparing Pakistani communities to cope with the repercussions of droughts on food security. The cultivated area observation helps to monitor the crop condition, which will ultimately help in securing food security for Pakistan. By viewing this gap, this study aims to monitor vegetation conditions, particularly in the Cropland of Pakistan, with the help of drought indices like VCI, TCI, and VHI using the MODIS dataset on the GEE platform. Further, a time series analysis has been done to observe the trend of vegetation health conditions.

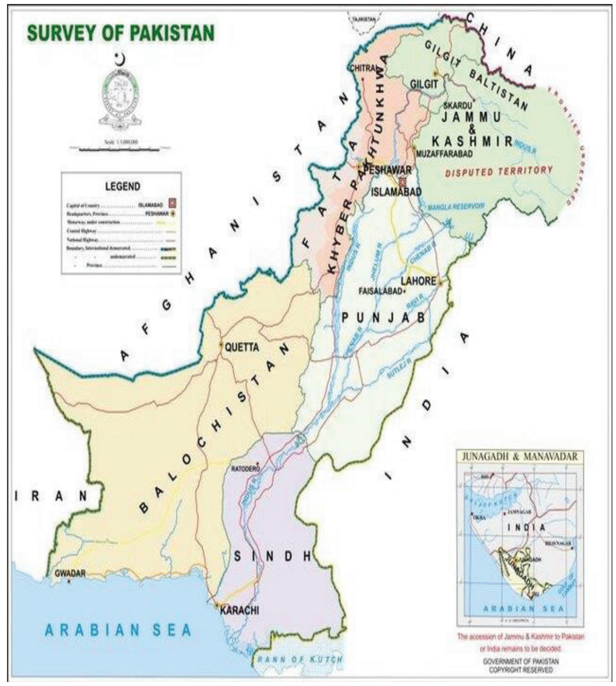

Fig. 1. Pakistan Map (Source: Survey of Pakistan (2015).

2. MATERIALS AND METHODS

2.1. Study Area

Pakistan (latitudes 30.3753°'N and longitudes 69.3451°E), located in South Asia covers an area of 881,913 km^2 (Figure 1). The country experiences four distinct seasons based on temperature because it is in the northern hemisphere; winter that is cool and dry (December to February), a hot and dry spring (March to May), a hot and humid summer (June to August), and a dry fall (Sept. to Nov.) [34]. Pakistan experiences two monsoon seasons; the Indian monsoon (July–Sep) and the western disturbance (Dec-Mar). Pakistan has a population of around 160 million people, with 32% living below the poverty line, and faces significant challenges [34]. Due to diverse topography and altitudes, the country's climate can be characterized as continental with tremendous diversity [34].

2.2. Datasets

GEE is a cloud-based computing web that provides several remotely sensed datasets and powerful geospatial processing capabilities. This paper utilized the GEE repository of freely available datasets. MODIS is an extensive program, and its datasets are extensively used for drought assessment and prediction [35]. Therefore, two MODIS products (LST, NDVI) were used in this study. The GFSAD (Global Food Security-support Analysis Data) is a High-resolution global cropland data provided by NASA, and it contributes to global food security in this era. The GFSAD1000 was used for extracting cropland of the study area. The type, spatial-temporal resolution, and time span of the data products are shown in Table 1.

2.3. Methodology

Drought assessment is crucial for four phases of hazard management (Preparedness, Response, Recovery, and Mitigation) [35]. In this regard, remotely sensed technology presents valuable geo-information as satellite images with spatiotemporal capability. It builds recurring and precise drought

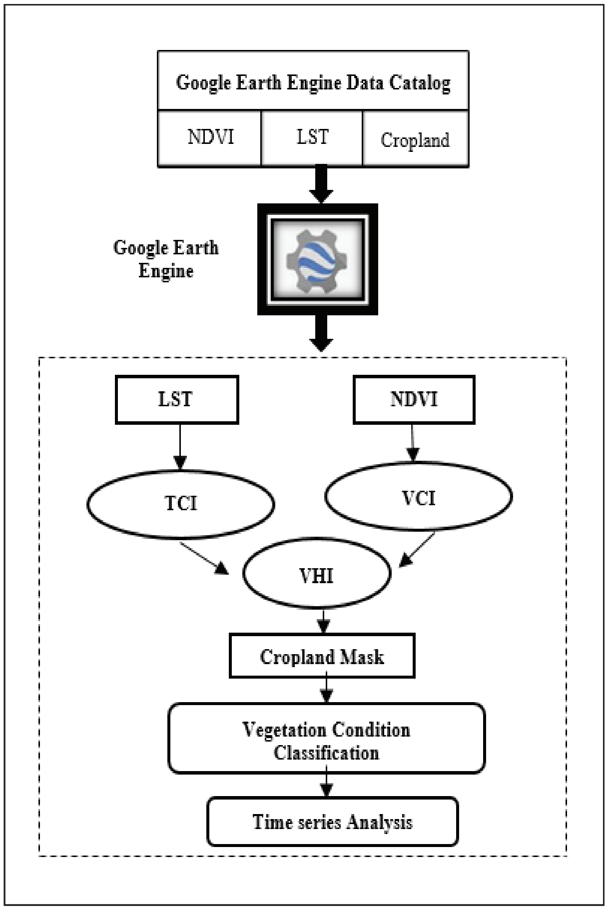

Fig. 2. Methodological Work that shows the GEE software system is utilized to determine the VCI, TCI, and VHI to observe drought conditions and Time-series Analysis.

event maps of particular regions at a specific time. The study presents spatiotemporal maps (mean monthly and yearly) for the Pakistan region spanning from 2001 to 2021, along with the yearly time series analysis and methodology depicted in Figure 1. Data for the study were collected and processed on the GEE platform. Monthly LST and NDVI from 2001 to 2021 were derived from MODIS data products. The High-Resolution Global Cropland Data (GFSAD) was utilized to apply a cropland mask to vegetated areas in Pakistan. LST and NDVI serve as inputs for calculating the indices VCI, TCI, and VHI. Applying a cropland mask provides the Vegetated areas of Pakistan. The yearly mean and monthly mean maps of VHI were produced from 2001 to 2021 to identify the drought severity, particularly in the cropland of Pakistan. Time series were also analyzed to deep monitor the trend of drought conditions over the period.

2.3.1 Satellite Based Drought Indices

The satellite-based drought indices were determined using remote sensing data products. Some of them are given below:

2.3.1.1 Normalized Difference Vegetation Index (NDVI)

NDVI is a well-known vegetation index to measure the development and the amount of vegetation and has a range from -1 to +1. For water bodies and clouds, the value of NDVI is negative; close to negative values represent the soil or build-up area, whereas equal and above 0.6 represents the healthy vegetation [36]. The near-infrared (NIR) and red bands are used to calculate NDVI. Consider that "ρ_{nir}" is the near-infrared band reflectance and "ρ_{red}" is the red band reflectance, then NDVI can be measured as shown in equation (1) [37].

$$NDVI = \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red}} \tag{1}$$

2.3.1.2 Vegetation Condition Index (VCI)

VCI can effectively assess the drought condition by delineating the drought region for
the definite value of NDVI and VCI [38]. VCI can be computed using the following equation (2):

$$VCI = \frac{NDVI_i - NDVI_{min}}{NDVI_{max} - NDVI_{min}} \times 100 \tag{2}$$

Where $NDVI_{max}$ and $NDVI_{min}$ are the maximum and minimum NDVI of each pixel determined monthly, and *i* is the current month's index.

The VCI measure in a percentage ranging from 1 to 100 and above 50% shows the normal condition of vegetation, whereas less than 50% indicates the drought condition [9]. VCI is a better identifier of vegetation response to precipitation impact because it is more sensitive to precipitation dynamics than NDVI [39]. The VCI classification is based on different drought conditions [9] and is shown in Table 2 [14].

2.3.1.3 Temperature Condition Index (TCI)

The satellite image thermal band is translated to brightness temperature (LST) and is used to

Table 2. Five *Criteria of Drought Conditions from No Drought to* **Extreme Drought** *for Different Indices (VCI, TCI, and VHI).*

Drought Condition	VCI (%)	TCI (%)	VHI (%)
Extreme Drought	1-10	1-10	1-10
Severe Drought	10-20	10-20	10-20
Moderate Drought	20-30	20-30	20-30
Mild Drought	30-40	30-40	30-40
No Drought	40>	40>	40>

develop the Temperature Condition Index (TCI). It is based on vegetative stress due to temperature and excessive wetness stress shown in equation (3) [40].

$$TCI = \frac{LST_{max} - LST_i}{LST_{max} - LST_{min}} \times 100 \quad (3)$$

Where, LST_i is the monthly brightness temperature, LST_{max} and LST_{min} represent the maximum and minimum brightness temperature over the total time duration respectively.

2.3.1.4 Vegetation Health Index (VHI)

VHI is one of the most extensively used remotely sensed drought indices [41, 42]. Many researchers used VHI in various applications like crop yield loss assessment, which is a crucial factor for food security [42]. The combination of VCI and TCI is used to assess the vegetation condition and is given in the equation (4):

$$VHI = \alpha VCI + (1 - \alpha)TCI \quad (4)$$

Where α is a weighted factor that is typically set as 0.5 [41, 42].

LST and NDVI are used to verify VHI values using correlation testing. VHI values can be classified according to their severity as shown in Table 2.

2.3.2 Time Series Analysis

2.3.2.1 Man-Kendall Test

The Man-Kendall test was used for the trend analysis of *VHI* in this research. The non-parametric Mann-Kendall test was developed by Mann [43] for trend detection, and Kendall [44] provided the test statistic distribution for testing non-linear trends and turning points [45]. Trends in hydro-meteorological time series have frequently been quantified using the Mann–Kendall statistical test [46]. The Mann–Kendall test statistic (Mann [43]; Kendall [44]) is calculated as:

$$S = \sum_{i=1}^{n-1} . \sum_{j=i+1}^{n} sgn\left(VHI_j - VHI_i\right) \quad (5)$$

A time series VHI_i which is ranked from i = 1, 2, ..., n-1, and a time series VHI_j, which is ranked from $j = i + 1$, 2, ... n, are used for the trend test. In order to compare each data point with the rest of the data points VHI_j, a reference point is used for each of them.

$$sgn\left(VHI_j - VHI_i\right) = \begin{cases} +1 & if\ VHI_j - VHI_i > 0 \\ 0 & if\ VHI_j - VHI_i = 0 \\ -1 & if\ VHI_j - VHI_i < 0 \end{cases} \quad (6)$$

It has been demonstrated that the statistic S has a distribution that is approximately normal with the mean when n > 8. E(S) = 0

The variance statistic is given as:

$$Var(S) = \frac{n(n-1)(2n+5) - \sum_{i=1}^{p} t_i(t_i - 1)(2t_i + 5)}{18} \quad (7)$$

Where t_i is the number of ties up to sample i, and Z_S represents the test statistics.

$$Z_S = \begin{cases} \frac{S-1}{\sqrt{Var(S)}} & if\ S > 0 \\ 0 & if\ S = 0 \\ \frac{S+1}{\sqrt{Var(S)}} & if\ S < 0 \end{cases} \quad (8)$$

Here, Z_S has a normal distribution that is typical. A Z value that is either positive or negative indicates an upward or downward trend. A two-tailed test can also be used to test for either an upward or downward monotone trend using a significance level. The trend is deemed significant if Z_S appears greater than $Z_{\alpha/2}$ where α is the significance level.

2.3.2.2 Sen's Slope estimator Test:

The Sen's estimator predicts the magnitude of the trend. The slope *Ti* of the all data pairs is computed as:

$$Ti = \frac{VHI_j - VHI_k}{j-k} \quad (9)$$

for $i = 1,2, \ldots N$

where, VHI_j and VHI_k are the *VHI* values at time *j* and *k* $(j > k)$. The median of these N values of *Ti* is represented as Sen's estimator of slope which given as:

$$AVHI = \frac{VHI_i - VHI_{avg}}{VHI_{avg}} \quad (10)$$

Sen's estimator is calculated as $T_{N+1/2}$ if N appears odd, and it is regarded as $½(T_{N/2} + T_{(N+2)/2})$ if N appears even. The non-parametric test can then be used to determine a true slope after Q_{med} is calculated using a two-sided test with a confidence interval of 100 (1-α) %. The time series' upward or increasing trend is represented by a positive Q_{med} value, while the downward or decreasing trend is represented by a negative Q_{med} value.

2.3.2.3 Anomaly vegetation health index

Anomaly Vegetation Health Index *(*AVHI*)* is a measure of the variation in healthy vegetation compared to a long-term average. AVHI is estimated by deducting the long-term average value of VHI from the current value of VHI, and then dividing the result by the standard deviation of the long-term VHI [47, 48]. AVHI formula for estimation:

$$AVHI = \frac{VHI_i - VHI_{avg}}{VHI_{avg}} \quad (11)$$

Where, VHI_i is the vegetation health index during a particular time *i*. VHI_{avg} is the average VHI for the total time duration.

Positive values of AVHI indicate above-average vegetation health, while negative values suggest below-average vegetation health. The use of AVHI can help identify anomalies or deviations from the long-term average of vegetation health, which can be indicative of changes in vegetation conditions due to various factors such as drought, floods, or other environmental stressors. By tracking AVHI over time, it is possible to detect changes in vegetation health and monitor the impact of environmental factors on vegetation conditions the use of AVHI provides a useful tool for monitoring and assessing changes in vegetation health, which is important for various applications such as drought management, water resource planning, and food security.

3. RESULTS

This section summarizes and discusses the main findings of the work by observing the mean yearly and monthly spatial and temporal variations of VHI as well as its anomaly.

3.1 Mean Yearly and Monthly Spatial Variation of VHI

GEE platform extracted the study outcomes with the help of the two MODIS data products.

The yearly mean and monthly mean VHI maps were created from 2001 to 2021, as shown in Figure 3 and Figure 4, respectively. Two MODIS data products (LST and NDVI) help to estimate the vegetation indices TCI and VCI. The combination of TCI and VCI gives vegetation health indices, VHI, to monitor drought in cropland. The annual mean of VHI from 2001 to 2021 indicated mild to extreme drought conditions in Pakistan's vegetated areas, while the drought condition is highest in the Sindh region. In 2001 and 2002, VHI maps indicated severe drought in almost all over the study area. The vegetation conditions were healthy in 2019 and 2020.

The monthly mean maps were prepared from 2001 to 2021 and shown in Figure *5* (from January to June) and in Figure 6 (from July to December). Vegetation conditions were good in February, April, September, and October. In November, mostly cropland was affected by severe drought in the Northern parts of Pakistan. Moreover, some of the Punjab regions also showed moderate to severe drought. Most of the Sindh areas faced extreme drought conditions in July.

The affected area percentages during the study period indicate 2001 as the driest year, as shown in Figure 7. Extreme drought to moderate drought occurred during the dry years (57% in 2001, 49%

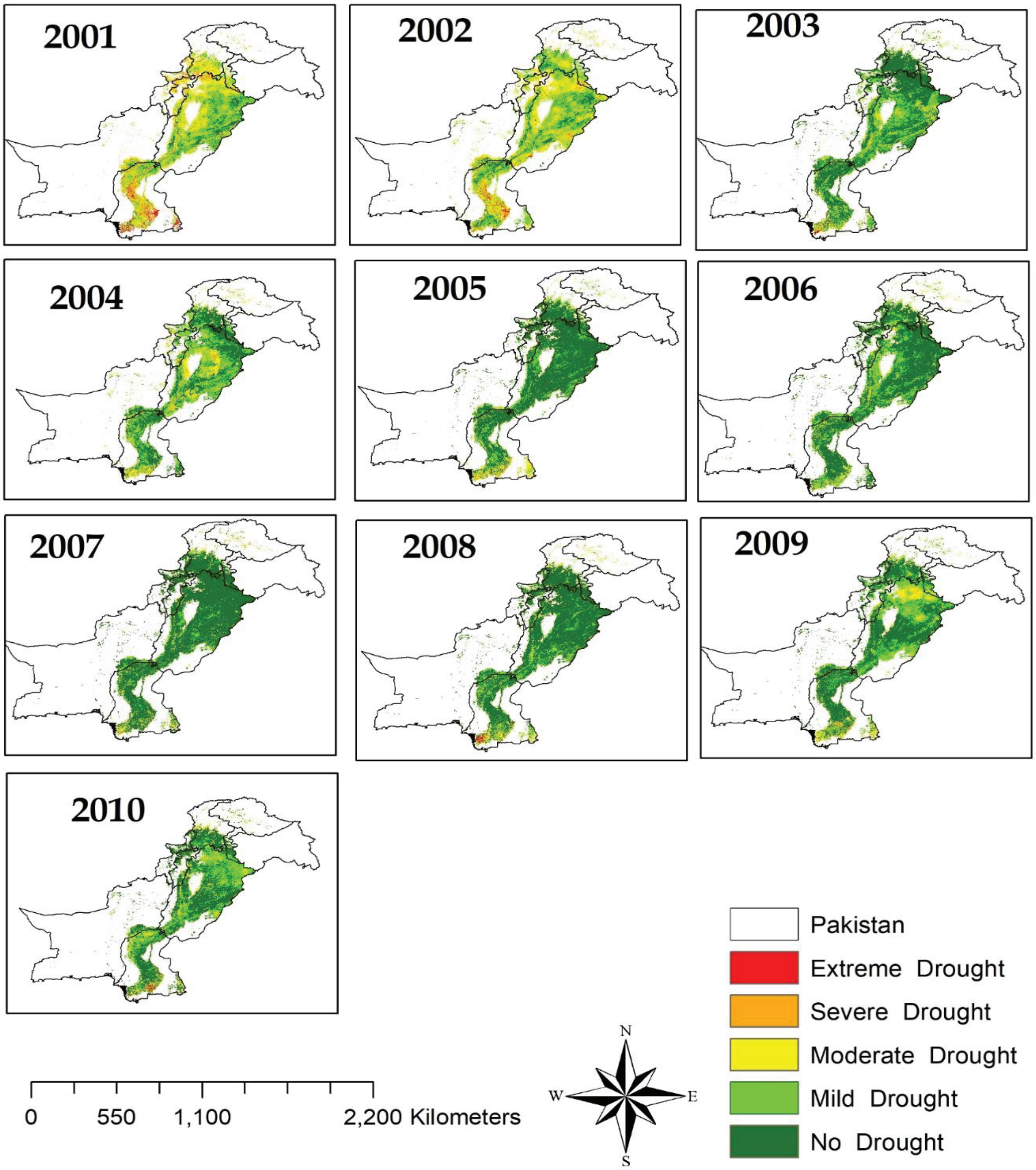

Fig. 3. Yearly mean VHI maps from 2001 to 2010 for five drought conditions: No Drought, Mild drought, Moderate Drought, Severe Drought, and Extreme Drought.

in 2002, and 22% in 2004). In the next decade of the study period (2011 to 2021), large crop areas showed healthy vegetation with some droughts of about 5% to 12%. In contrast, during very humid years, the croplands were less affected by drought (about 6% in 2007, 5% in 2014, 2015, and 2021, and 4% in 2020). Thus, as a general feature, healthy vegetation observation from 2011 to 2021 was decreasing the drought effect in Pakistan, although only some regions have been severely affected (as shown in Figure 3 to Figure 7). The graphical representation (Figure 7) showed an annual mean value of VHI, and seasonal changes in VHI can alter these results.

3.2 Temporal Variation of VHI

The research utilized statistical techniques such as Mann-Kendall and Sen's Slope Estimator to determine the trend in VHI time series data. The

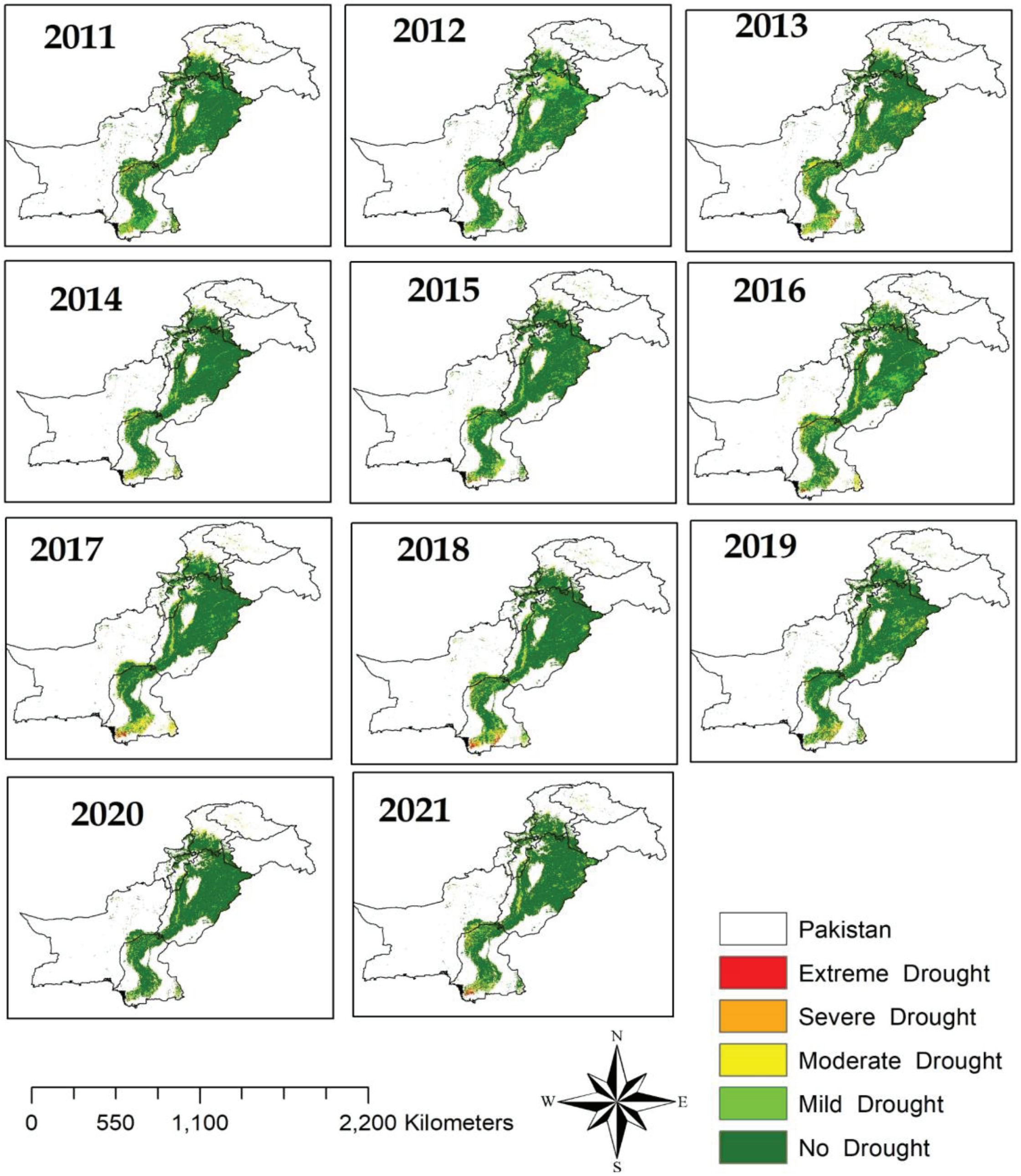

Fig. 4. Yearly mean VHI maps from 2011 to 2021 for five drought conditions: No Drought, Mild drought, Moderate Drought, Severe Drought, and Extreme Drought

Table 3. Statistical Test Results that show Trend, Min, Max, P-value, Z value, Tau, Standard deviation, Variance, and Slope of the time series

Trend	Min	Max	Mean	P.Value	Zs	Tau	S	Var(S)	Slope
Increasing	20.53	57.98	43.1	0.006	2.74	0.4	92	1096.67	0.85

results indicated a positive trend in VHI, as shown by positive values of *Zs* (2.74) and a p-value of 0.006 (Table 3). The Sen's slope estimator was also used to calculate the annual mean trend increase in VHI, which was found to be 0.85 (Table 3), reflecting healthy vegetation. These findings were verified by the annual VHI time series data, as shown in Figure 8, and were attributed to adequate rainfall, as predicted by Ahmed, *et al.* [49].

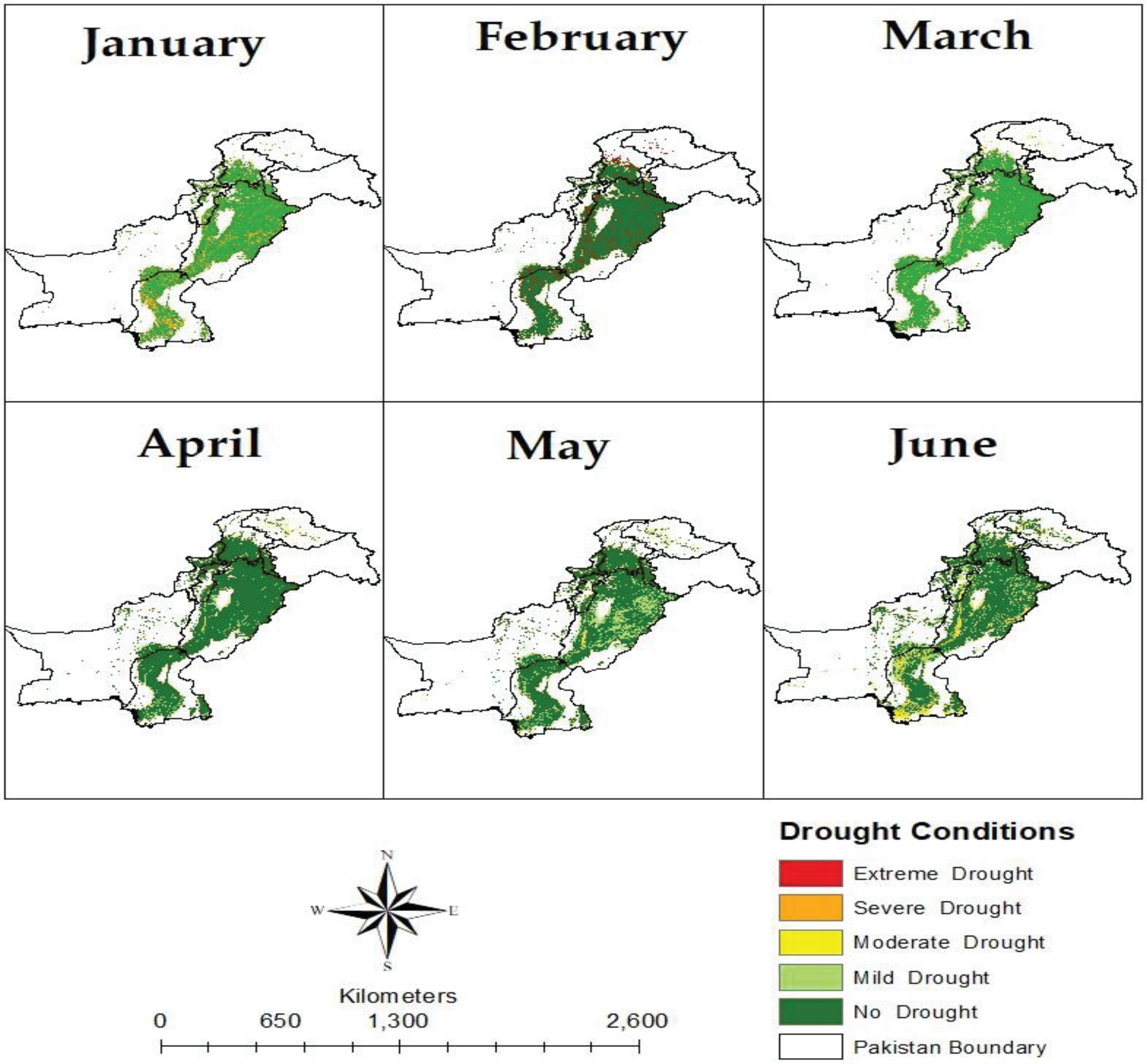

Fig. 5. Monthly mean VHI maps from January to June (2001-2021) for five drought conditions (No Drought, Mild drought, Moderate Drought, Severe Drought, and Extreme Drought)

Overall, the use of statistical techniques employed in the present study provided important insights into the trend of vegetation health in Pakistan, highlighting the positive impact of sufficient rainfall on vegetation conditions. Figure 9 shows the VHI annual mean variations with the help of horizontal lines for three drought conditions (severe, moderate, and no drought). The threshold of severe drought indicates by the red line at 20%, moderate drought with the yellow line (20% to 30%), and healthy vegetation with the green line (above 30%). Figure 9 shows that severe drought was observed in 2001, while moderate in 2002 on the annual time scale. Moreover, crop conditions were healthy in 2007, 2011 to 2015, and 2019 to 2021.

3.3. Anomaly VHI

Anomaly Vegetation Health Index (AVHI) represents the soil moisture content efficiency and deficiency. The anomaly observations from 2001 to 2021 vary from -0.52 to 0.35, as shown in Figure 10. According to AVHI, the length of mild to extreme drought was eleven years (2001 to 2006, 2009 to 2010, and 2016 to 2018), and two years (2001 and 2002) were the driest years. Cropland Soil moisture content was satisfactory in 2007, between 2011 and 2015, and from 2019 to 2021. In the first decade (2001 to 2010), the study area faced a dry spell, whereas, in the next decade (20011 to 2121), vegetation condition was satisfied.

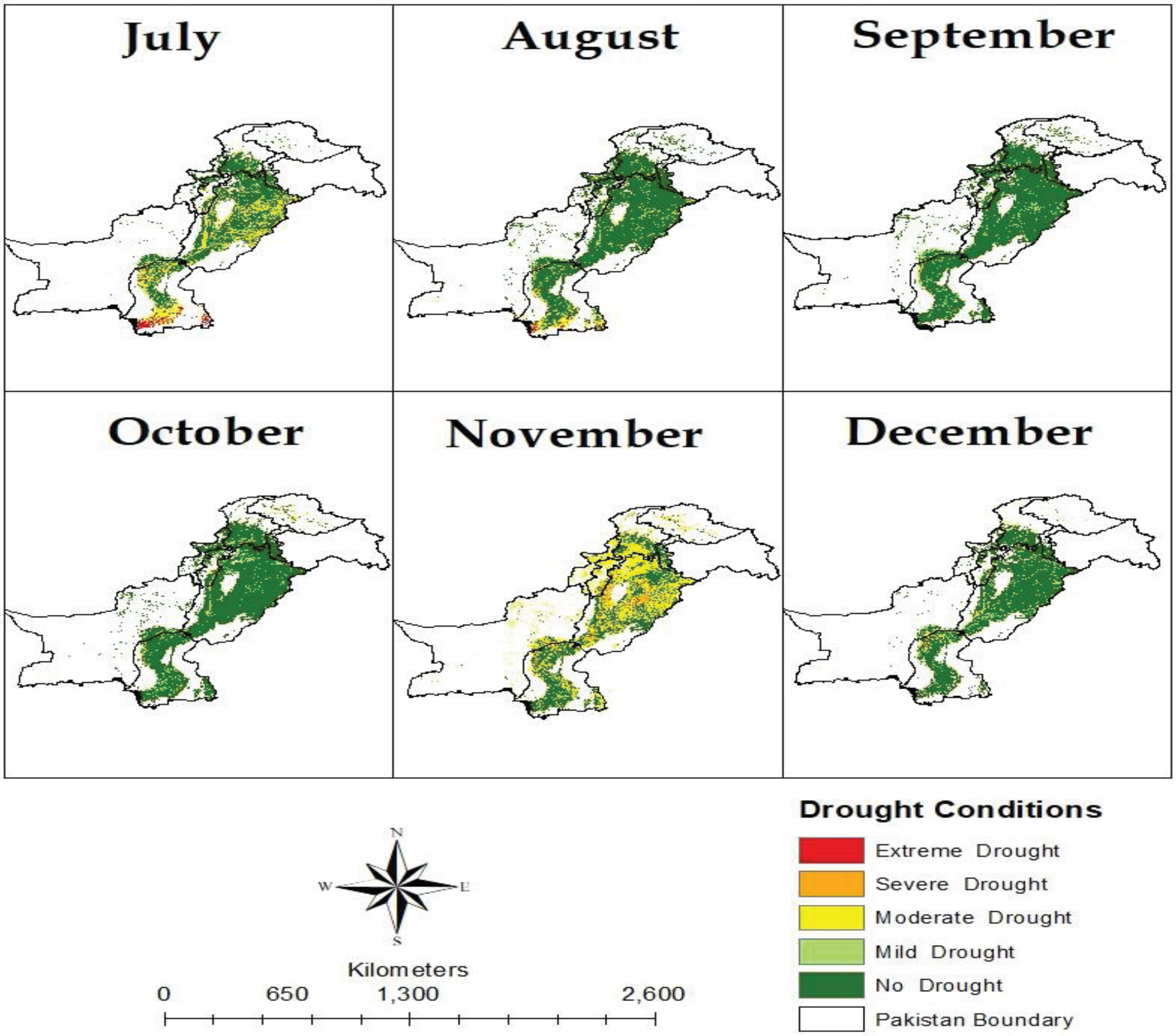

Fig. 6. Monthly mean VHI maps from July to December (2001-2021) for five drought conditions (No Drought, Mild drought, Moderate Drought, Severe Drought, and Extreme Drought).

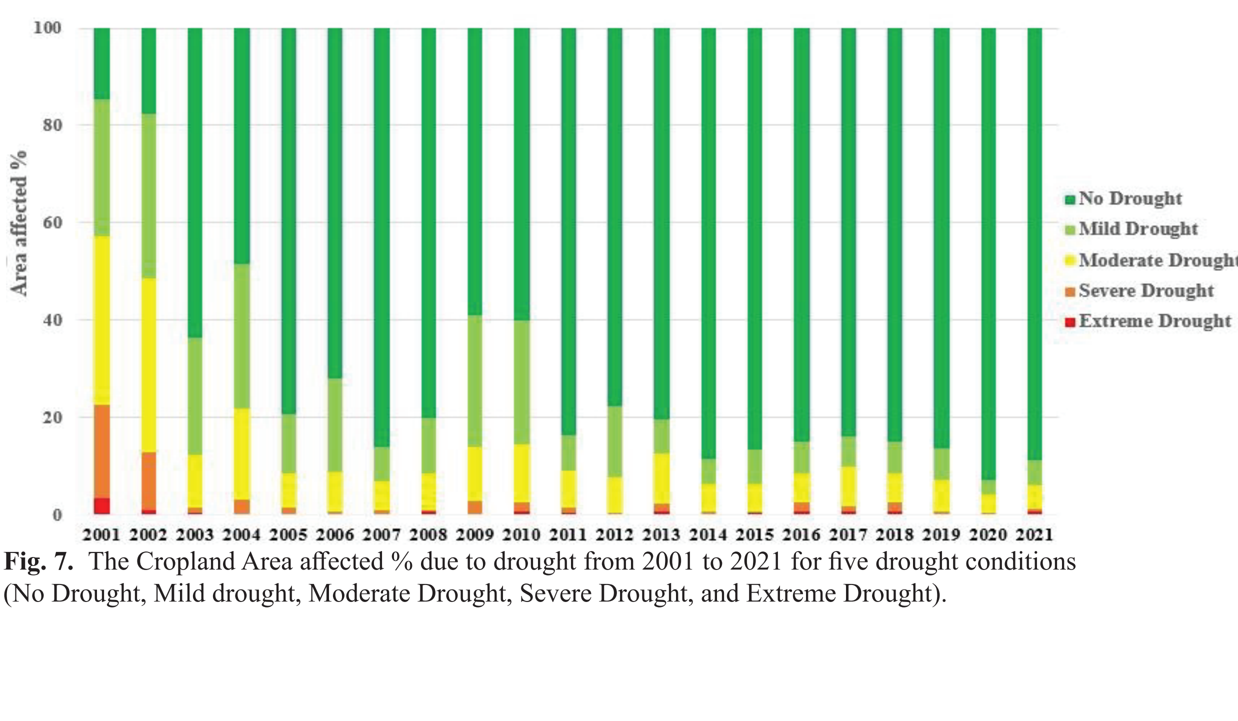

Fig. 7. The Cropland Area affected % due to drought from 2001 to 2021 for five drought conditions (No Drought, Mild drought, Moderate Drought, Severe Drought, and Extreme Drought).

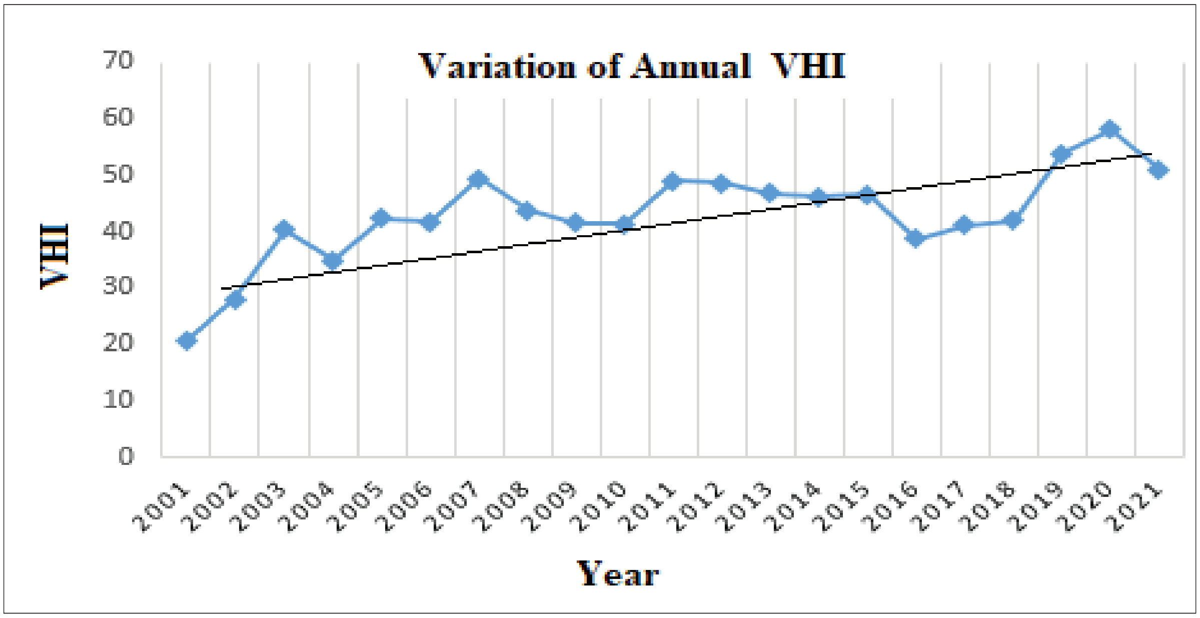

Fig. 8. Mean Annual VHI Variation from 2001 to 2021 with trend line.

By examining the multiannual mean VHI during the analyzed period (2001–2020), about 28 % of the arable land in Pakistan has suffered from mild to extreme drought. It can be seen in Figure 11 that about 14% of the arable land of the entire study area faced mild drought conditions, and 11% faced moderate drought. Collectively 3.5% area of the cropland bore extreme and severe drought.

4. DISCUSSION

The effects of climate warming, such as heat waves, droughts, and increases in salinity, pose a threat to global food security [50]. Pakistan is one of the ten worst-affected nations by climate change. Droughts, in particular, are expected to endure longer and be more severe as a result of climate change [28]. According to NDMC 's report [51], the drought conditions will undoubtedly ensnare the vulnerable regions of Pakistan if subsequent seasons fail to produce significant precipitation. Major droughts have occurred in all of Pakistan's provinces in the past [51]. The effects of drought, which include crop failure and decreased food production, are most severe in the arid and semi-arid regions of the country [27, 28]. According to the Food and Agriculture Organization of the United Nations (FAO), Pakistan is one of the countries, most likely

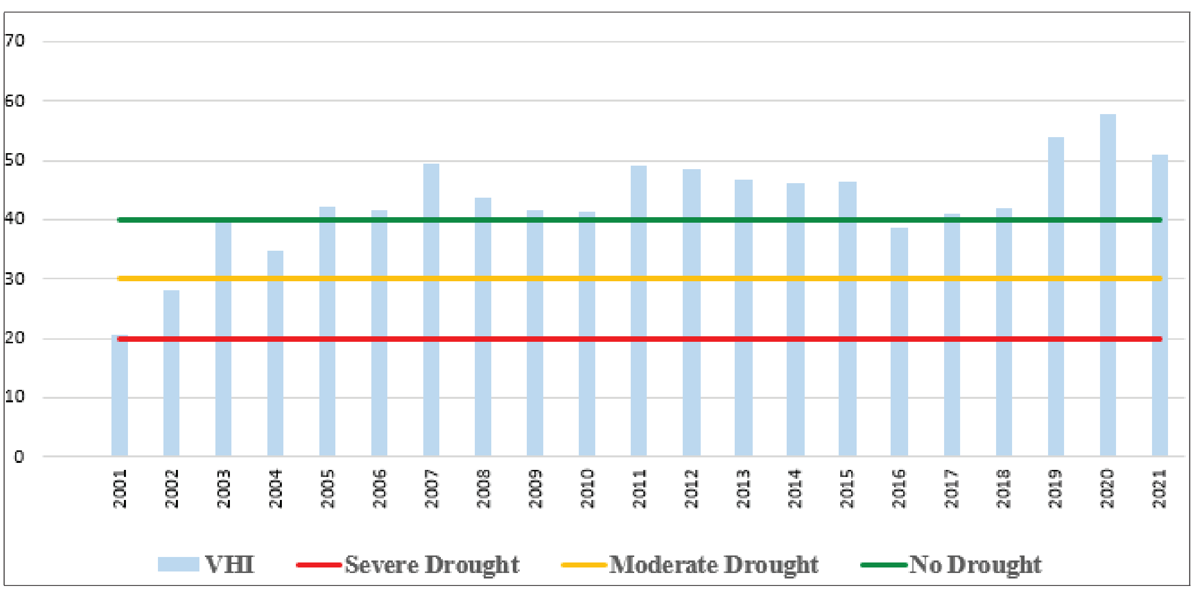

Fig. 9. Temporal variations in VHI from 2001 to 2021 for three drought conditions (No Drought, Moderate Drought, and extreme drought)

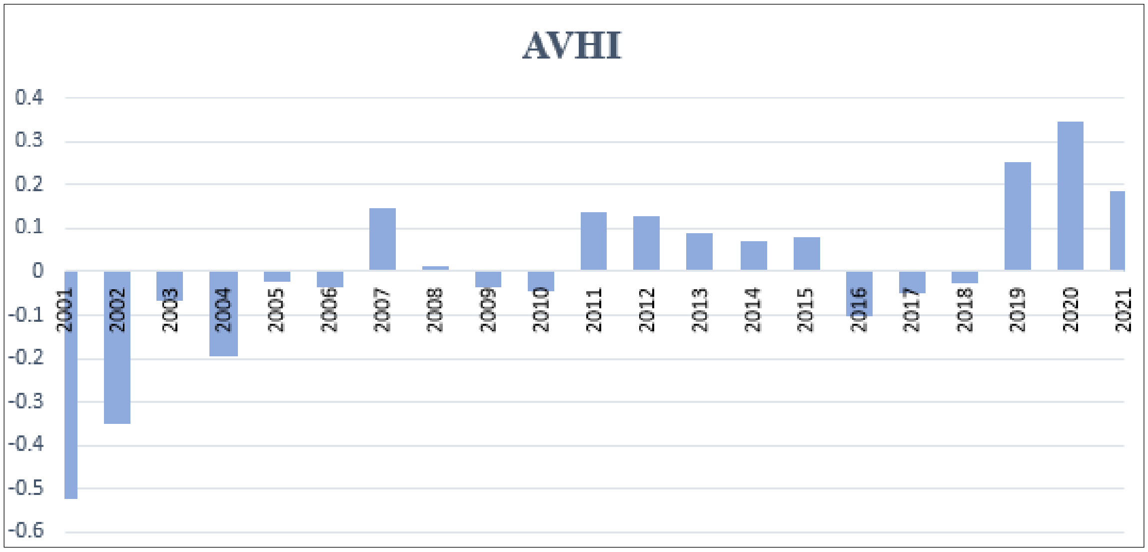

Fig. 10. Anomaly Vegetation Health Index (AVHI) from 2001 to 2021.

to experience food insecurity [52].

For drought mitigation and food security improvement, there is a need for a comprehensive drought early warning system in the cultivated area of Pakistan. One key component of such a system is the remote sensing indices used, such as the Vegetation VHI, to assess vegetation health and vigour and monitor drought conditions [53]. VHI can be calculated using remote sensing data, such as satellite imagery or aerial photographs, to create maps of vegetation health across large areas. These maps can identify the vegetation stress regions and monitor the changes in vegetation health over time. VHI and NDVI indices are the relative values and indicate vegetation health and vigour; but cannot be directly translated into a specific amount of vegetation or chlorophyll content. Additionally, VHI can only indicate moisture stress and can be affected by other factors such as pests, disease, or human-induced changes. Therefore, considering other factors and ground measurements is essential to understand the vegetation conditions.

Remote sensing data and platforms such as GEE can play a crucial role in such a system. GEE provides access to a wealth of satellite imagery and other geospatial data to create NDVI, Standard Vegetation Index (SVI), VHI, and other indices and analyze historical data to track changes over time. However, remote sensing data is only one source of information; ground-based measurements are also necessary to complement the remote sensing data for more accurate and precise findings [54]. Because Pakistan's population is growing at almost 2% per year, there is a greater demand for food supply in the market [55]. The current food production and drought early warning system in Pakistan is not adequate for the country's huge population [51]. A system that integrates modern technology, such as machine learning and the GEE platform, is needed to secure food security in Pakistan.

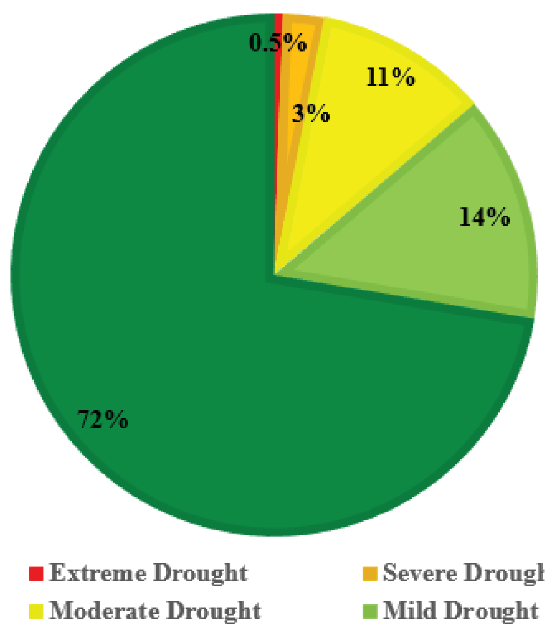

Fig. 11. Overall percentage of Drought Affected Area from 2001 to 2021 for five drought conditions (No Drought, Mild drought, Moderate Drought, Severe Drought, and Extreme Drought).

According to the above result, VHI 's positive trend indicates the healthy vegetation condition due to the rise in rainfall. Ahmed, et al. [49] predicted that the increase in precipitation will be gradual between 2010 and 2099. By the end of this century, Pakistan's total precipitation would rise by 8 to 41% throughout the year, depending on the season. Pakistan Economic Survey Report 2021-2022 also highlighted the healthy vegetation in agricultural land. According to the report, the agriculture industry experienced remarkable growth of 4.4% in 2021 and 2022. High yields, attractive output prices, supportive government practices, and improved selection of certified seeds, pesticides, and agriculture credit are the primary drivers of this expansion [56]. The cultivated area observation facilitates crop condition monitoring, which will ultimately help Pakistan's food security. Pakistan needs to invest in building its capacity to analyse remote sensing data and integrate it with other data to develop an effective and sustainable food production and drought early warning system. In summary, remotely sensed data is beneficial for drought monitoring, food production, and food security, specifically in Pakistan's arid and semi-arid regions. This preliminary study in Pakistan's drought-prone agricultural region has laid the groundwork for the utilization of more comprehensive predictor data products (such as satellite data with reanalysis). The remotely sensed data can be a valuable tool in building a comprehensive system that can help mitigate the effects of drought and enhance food safety in the country.

5. CONCLUSIONS

Spatio-temporal information about agricultural drought is crucial for policy-makers and ecological and farming applications. Drought is an extremist risk in Pakistan, and assessing it can assist determine the best mitigation strategy. GEE is a cloud-based system that evaluates massive data sets rapidly and efficiently. In this study, GEE is used to investigate the spatio-temporal variation of drought-based satellite-derived VHI obtained from TCI and VCI. Yearly and monthly mean VHI maps indicated that in 2001 and 2002, study area cropland faced severe to moderate drought. Moreover, AVHI results verified this condition by showing soil moisture efficiency in the same years. Monthly VHI observations indicate that the most cultivated land identified severe and moderate drought in November. From 2011 to 2021, the overall drought occurrences were not so much, and annual VHI has shown healthy vegetation conditions.

The results demonstrate the adequacy of vegetation in 2019, 2020 and 2021. Overall, all observations represent that the agricultural areas of Punjab have an adequate amount of soil moisture and healthy vegetation, whereas in Sindh majority of cropland has been affected due to the severe to moderate drought with continuous deficiency of soil moisture. Temporal variations indicated the increasing trend in VHI that represents the healthy vegetation future in the cropland of Pakistan. Annual yearly observation shows that percentage of severe and moderate drought occurring is comparatively less (about 3.5%), whereas moderate drought is collectively 25%.

Pakistan's food security and rural development depend on the agriculture sector's continued expansion. It employs approximately 37.4 percent of the workforce, manages rural landscapes, and serves as an environmental shield for climate-resilient production and ecosystems while also contributing 22.7 percent to the GDP [51]. Agricultural production is a crucial contributor to the Pakistani economy, and it should be satisfied to fulfill the food demand for a growing population. The shortage of surface and groundwater is the main reason for drought in the agricultural sector and causes crop failure with pasture losses. There is a need to be aware of temporal and spatial changes in cropland so that government can develop strategies to mitigate the drought impact. This study is a small effort to observe the overall drought situation during the last two decades. The results indicate that instead of mild to extreme drought occurrence, a positive trend in VHI time series shows a good sign of improvement in the agricultural sector in recent years. Pakistan Economic Survey report 2021-22 verified this result and it says that the agriculture industry grew by a remarkable 4.4% in 2022. The increased availability of certified seeds, pesticides, agricultural credit, attractive output prices, beneficial government policies, and excellent yields contribute to this expansion.

6. ADVANTAGES AND LIMITATION OF THE STUDY

The use of Google Earth Engine (GEE) for monitoring drought events and vegetation conditions in Pakistan has several advantages. GEE provides access to a vast collection of geospatial data, which can be used to generate accurate and timely information on vegetation conditions, drought occurrence, and severity. This information is crucial for drought management, planning, water resource management, food production, and food security. The study used GEE to calculate VHI, VCI, and TCI to monitor drought events over croplands and determine their severity. The study also produced mean yearly and monthly VHI maps and observed trends in drought occurrence over time.

However, the study has some limitations. The study's preciseness depends on the quality and consistency of the data utilized in the analysis. The study also has not considered other factors such as soil moisture and rainfall, which can affect vegetation conditions and drought severity. Additionally, the study only mainly focused on croplands, and the results may not be representative of other land cover types. Therefore, further research is needed to address these limitations and provide a more comprehensive understanding of drought events and vegetation conditions in Pakistan.

7. CONFLİCT OF INTEREST

The authors declare no conflict of interest.

8. REFERENCES

1. M.J. Hayes, M.D. Svoboda, B.D. Wardlow, M.C. Anderson, and F. Kogan. Drought Monitoring: Historical and Current Perspectives. In: Remote Sensing of Drought: Innovative Monitoring Approaches. B.D. Wardlow, M.C. Anderson, and J.P. Verdin (Ed.), *CRC Press, Boca Raton, USA,* pp. 1 (2012).
2. L.T. Berner, P.S. Beck, A.G. Bunn, A.H. Lloyd, and S.J.Goetz. High-latitude tree growth and satellite vegetation indices: Correlations and trends in Russia and Canada (1982–2008). *Journal of Geophysical Research: Biogeosciences* 116(G1): (2011).
3. M. Masud, M. Khaliq, and H. Wheater. Analysis of meteorological droughts for the Saskatchewan River Basin using univariate and bivariate approaches. *Journal of Hydrology* 522: 452-466 (2015).
4. S. Eslamian, *Handbook of engineering hydrology: modeling, climate change, and variability*: CRC Press, (2014).
5. M. Azmi, C. Rüdiger, and J.P. Walker. A data fusion-based drought index. *Water Resources Research* 52(3): 2222-2239 (2016).
6. Y. Yihdego, B. Vaheddoost, and R.A. Al-Weshah. Drought indices and indicators revisited. *Arabian Journal of Geosciences* 12(3): 1-12 (2019).
7. X. Zhao, H. Xia, L. Pan, H.Song, W. Niu, R. Wang, R. Li, X. Bian, Y.Guo, and Y.Qin. Drought monitoring over yellow river basin from 2003–2019 using reconstructed MODIS land surface temperature in google earth engine. *Remote Sensing*13(18): (2021).
8. I.A. Khan, and M.H. Arsalan, M.F. Siddiqui, N. Kiran, and M. Ajaib. Short-term drought assessment in Pakistan and adjoining areas by remote sensing MODIS-NDVI data: A potential consequence of climate change. *Pakistan Journal of Botany 48*(5): 1887-1892. (2016).
9. F.N. Kogan. Droughts of the late 1980s in the United States as derived from NOAA polar-orbiting satellite data. *Bulletin of the American Meteorological Society* 76(5): 655-668 (1995).
10. F. Zambrano, M. Lillo-Saavedra, K. Verbist, and O. Lago. Sixteen years of agricultural drought assessment of the BioBío region in Chile using a 250 m resolution Vegetation Condition Index (VCI). *Remote Sensing* 8(6): 530 (2016).
11. D.B. Kamble, S. Gautam, H. Bisht, S. Rawat, and A. Kundu. Drought assessment for kharif rice using standardized precipitation index (SPI) and vegetation condition index (VCI). *Journal of Agrometeorology* 21(2): 182-187 (2019).
12. V.A. Bento, C.M. Gouveia, C.C. DaCamara, and I.F. Trigo. A climatological assessment of drought impact on vegetation health index. *Agricultural and Forest Meteorology* 259: 286-295 (2018).
13. V.A. Bento, C.M. Gouveia, C.C. DaCamara, R. Libonati, and I.F.Trogo. The roles of NDVI and Land Surface Temperature when using the Vegetation Health Index over dry regions. *Global and Planetary Change* 190: 103-198 (2020).
14. R. Tabassum, M. Arsalan, A. Khalid, I. Ahmad, and A.I. Mirza. Assessment of Drought Risk by Using Vegetation Indices from Remotely Sensed Data: A Perspective from Hot and Arid District of Pakistan. *University of Engineering and Technology Taxila. Technical Journal* 20(4): 2-7 (2015).
15. A. Zargar, R. Sadiq, B. Naser, and F.I. Khan.. A review of drought indices. *Environmental Reviews*19: 333-349 (2011).
16. S. Galiano. Assessment of vegetation indexes from remote sensing: Theoretical basis. *Options*

Méditerranéennes 67: 65-75(2012).

17. F. Masitoh, and A. Rusydi. Vegetation Health Index (VHI) analysis during drought season in Brantas Watershed. In *IOP Conference Series: Earth and Environmental Science* 389(1):012-033.(2019).
18. U. Ma'Rufah, R. Hidayat, and I. Prasasti. Analysis of relationship between meteorological and agricultural drought using standardized precipitation index and vegetation health index. *In IOP Conference Series: Earth and Environmental Science* 389(1): 012-033 (2019).
19. O. Mutanga, and L. Kumar. Google Earth Engine Applications. *Multidisciplinary Digital Publishing Institute,* (2019).
20. N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D.Thau, and R.Moore. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote sensing of Environment* 202: 18-27 (2017).
21. L. Kumar, and O. Mutanga. Google Earth Engine applications since inception: Usage, trends, and potential. *Remote Sensing*10(10): 1509 (2018).
22. Y. Wang, J. Ma, X. Xiao, S. Dai, and B.Zhao. Long-term dynamic of Poyang Lake surface water: A mapping work based on the Google Earth Engine cloud platform. *Remote Sensing* 11(3): 313 (2019).
23. P. Thilagaraj, P. Masilamani, R. Venkatesh, and J. Killivalavan. Google earth engine based agricultural drought monitoring in kodavanar watershed, part of amaravathi basin, Tamil Nadu, India. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 43: 43-49 (2021).
24. Finance Division Government of Pakistan. *Pakistan Economic Survey 2021-22, Islamabad, Pakistan* (2022).
25. IPC, Pakistan: IPC Acute Food Insecurity Analysis-April 2023-January 2024. (2023).
26. FAO, Crop Prospects and Food: July 2022, Rome, Italy (2022).
27. I. Ullah, X. Ma, J. Yin, T.G. Asfaw, K. Azam, ,S. Syed, M.Liu, M. Arshad, and M. Shahzaman. Evaluating the meteorological drought characteristics over Pakistan using in situ observations and reanalysis products. *International Journal of Climatology* 41(9): 4437-4459 (2021).
28. S. Jamro, G.H. Dars, K. Ansari, and N.Y. Krakauer. Spatio-temporal variability of drought in Pakistan using standardized precipitation evapotranspiration index. *Applied Sciences* 9(21): 45-88 (2019).
29. K. Ahmed, S. Shahid, and N. Nawaz. Impacts of climate variability and change on seasonal drought characteristics of Pakistan. *Atmospheric research* 214: 364-374 (2018).
30. G. Ali, M. Sajjad, S. Kanwal, T. Xiao, S. Khalid, F. Shoaib, and H.N. Gul. Spatial–temporal characterization of rainfall in Pakistan during the past half-century (1961–2020). *Scientific Reports* 11(1):1-15 (2021).
31. N. Khan, D. Sachindra, S. Shahid, K Ahmed, M.S. Shiru, and N. Nawaz. Prediction of droughts over Pakistan using machine learning algorithms. *Advances in Water Resources* 139: 103562 (2020).
32. M. Pasha, A. Ali, and A. Waheed. Sindh drought 2014-Pakistan: was it a natural or a man-made disaster. *American International Journal of Social Science Research* 1(1): 16-20 (2015).
33. R. Khan, H. Gilani, N. Iqbal, and I. Shahid. Satellite-based (2000–2015) drought hazard assessment with indices, mapping, and monitoring of Potohar plateau, Punjab, Pakistan. *Environmental Earth Sciences* 79(1): 1-18(2020).
34. S. Khan. Climate classification of Pakistan. *International Journal of Economic and Environmental Geology* 10(2): 60-71 (2019).
35. S. Aksoy, O. Gorucu, and E. Sertel. Drought monitoring using MODIS derived indices and google earth engine platform. In *2019 8th International Conference on Agro-Geoinformatics (Agro-Geoinformatic* IEEE: 1-6(2019).
36. R.-V. Dobri, L. Sfîcă, V.A. Amihăesei, L.Apostol, and S.Țîmpu. Drought extent and severity on arable lands in Romania derived from normalized difference drought index (2001–2020). *Remote Sensing* 13(8): 1478 (2021).
37. C.J. Tucker. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment* 8(2): 127-150 (1979).
38. N. Vaani, and P. Porchelvan. Monitoring of agricultural drought using fortnightly variation of vegetation condition index (VCI) for the state of Tamil Nadu, India. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 42: 159-164 (2018).
39. F.N. Kogan. Remote sensing of weather impacts on vegetation in non-homogeneous areas. *International Journal of Remote Sensing* 11(8): 1405-1419, (1990).
40. R.P. Singh, S. Roy, and F. Kogan. Vegetation and temperature condition indices from NOAA AVHRR data for drought monitoring over India. *International Journal of Remote Sensing* 24(22): 4393-4402 (2003).
41. F.N. Kogan. Global drought watch from space. *Bulletin of the American Meteorological Society.* 78(4): 621-636 (1997).
42. F.N. Kogan. Operational space technology for global vegetation assessment. *Bulletin of the American Meteorological Society* 82(9): 1949-1964 (2001).
43. H.B. Mann. Nonparametric tests against trend. *Econometrica: Journal of the Econometric Society*: 245-259(1945).

44. M.G. Kendall. Rank correlation methods. *Griffin.* (1948).
45. A. Mondal, S. Kundu, and A. Mukhopadhyay. Rainfall trend analysis by Mann-Kendall test: A case study of north-eastern part of Cuttack district, Orissa. *International Journal of Geology, Earth and Environmental Sciences* 2(1): 70-78 (2012).
46. R.M. Da Silva, C.A. Santos, M. Moreira, J. Corte-Real, V.C. Silva, and I.C. Medeiros. Rainfall and river flow trends using Mann–Kendall and Sen's slope estimator statistical tests in the Cobres River basin. *Natural Hazards (*77): 1205-1221 (2015).
47. L. Liang, Q. Sun, X. Luo, J. Wang, L. Zhang, M. Deng, L.Di, and Z. Liu. Long-term spatial and temporal variations of vegetative drought based on vegetation condition index in China. *Ecosphere* 8(8): 01919 (2017).
48. B. Baniya, Q. Tang, X. Xu, G.G. Haile, and G. Chhipi-Shrestha. Spatial and temporal variation of drought based on satellite derived vegetation condition index in Nepal from 1982–2015. *Sensors* 19(2): 430 (2019).
49. K. Ahmed, S. Shahid, E.S. Chung, X.J. Wang, and S.B. Harun. Climate change uncertainties in seasonal drought severity-area-frequency curves: Case of arid region of Pakistan. *Journal of Hydrology.* 570: 473-485 (2019).
50. Z. Fatima, M. Ahmed, M. Hussain, G. Abbas, S.Ul-Allah, S.Ahmad, N.Ahmed, M.A. Ali, G. Sarwar, E.U. Haque, and P. Iqbal. The fingerprints of climate warming on cereal crops phenology and adaptation options. *Scientific Reports* 10(1) 1-21 (2020).
51. National Drought Monitoring & Early Warning Centre. Drought Bulletin of Pakistan : October-December 2022. *National Drought Monitoring Centre, Islamabad, Pakistan* (2022).
52. FAO. Crop Prospects and Food: December 2022, *Rome, Italy* (2022).
53. J. Zeng, R. Zhang, Y. Qu, V.A. Bento, T. Zhou, Y. Lin, X. Wu, J. Qi, , W. Shui, and Q. Wang. Improving the drought monitoring capability of VHI at the global scale via ensemble indices for various vegetation types from 2001 to 2018. *Weather and Climate Extremes* 35: 100412 (2022).
54. S. Kocaaslan, N. Musaoğlu, and S. Karamzadeh. Evaluating Drought Events by Time-Frequency Analysis: A Case Study in Aegean Region of Turkey. *IEEE Access* 9: 125032-125041 (2021).
55. M. Ahmad, M. Iqbal, and M. Khan. Climate change, agriculture and food security in Pakistan: adaptation options and strategies. *Pakistan Institute of Development Economics, Islamabad. (Climate Change Brief)* (2013).
56. Finance Devision Government of Pakistan. Highlights - Pakistan Economic Survey 2021-22. *Finance Devision Government of Paksitan* (2022).

Proceedings of the Pakistan Academy of Sciences: A Physical and Computational Sciences 60(4): 29-33 (2023)

ISSN (Print): 2518-4245; ISSN (Online): 2518-4253
https://doi.org/10.53560/PPASA(60-4)670

Pakistan Academy of Sciences

Research Article

New Numerical Approach to Calculate Microstates of Equivalent and Non-Equivalent Electrons

Misha Shafi[1], Saba Javaid[2,3*], Roohi Zafar[2], Ahmed Ali Rajput[1], Muhammad Mustaqeem Zahid[1], and Muhammad Daniyal[1]

[1]Department of Physics, University of Karachi, Karachi, Pakistan
[2]Department of Physics, NED University of Engineering & Technology, Karachi, Pakistan
[3]Department of Physics, Federal Urdu University of Arts, Science & Technology, Karachi, Pakistan

Abstract: A term symbol is used to describe atomic microstate states, which give the multiplicity and total angular momentum of the atomic state. Russel Sauder coupling scheme is used to generate terms of equivalent and non-equivalent electronic configurations. For equivalent electrons, the terms are calculated using Pauli's principle, and the number of terms is limited and is calculated by the combination rule. The total possible electrons and total available electrons are used in the combination formula. In case of non-equivalent electrons, the number of terms are found by the permutation rule. The number of terms for equivalent electrons is less than the terms for non-equivalent electrons. The number of possible microstates for p^2 and d^5 configurations are 15 and 252 respectively. While the number of final microstates for 1p2p and 3d4d configurations are 36 and 100. In the proposed study, a Python programme was developed that generates the microstate according to filled and half-filled subshell electronic configurations for equivalent, non-equivalent, and combinations of both. Examples of microstates for non-equivalent electrons of configuration *1s2s, sp, sd, ss, 2p3p, pd, pf, 3d4d, df, 4f5f* and for equivalent electrons of configuration s^u, p^v, d^x, and f^y are presented.

Keywords: Microstates, Term Symbols, Equivalent and Non-Equivalent Electrons, Pauli Exclusion Principle, L-S Coupling.

1. INTRODUCTION

Atomic microstates are identified by a spectral term that specifies their multiplicity and overall angular momentum. Term symbols provide information on the spectral and magnetic properties of various elements [1]. These term values are helpful to identify coefficients of fractional parentage of wavefunctions of elements having complex structures like Praseodymium, Tantalum etc. Henry Russell and Frederick Saunders introduced the Russell-Saunders (which is abbreviated as R-S) scheme for the first time in 1923. Initially, it was used to electrons in half-filled orbits of atoms with lower atomic numbers since the spin-orbit coupling is less effective than the electrostatic effect [2]. However, spin-orbit coupling due to higher nuclear charge seems to be more important for those elements that have a higher atomic number. The R-S technique is still useful for rare Earth elements and heavier transition elements. The complex spectra of structures containing valence electrons in distinct sub shells were successfully interpreted by using a vector model for terms that were established before the quantum mechanical approach [3]. The three vectors, $\vec{\mathbf{S}}, \vec{\mathbf{L}}$ and $\vec{\mathbf{J}}$ are produced as a result of the R-S coupling, which is the basis for the present nomenclature for the specific energy level (2S+1) LJ, where 2S+1 is the multiplicity or spin multiplicity of a term, the orbital angular momentum vectors of the valence electrons are vectorially added together to form L and spin angular momentum vectors of valence electrons are vectorially summed to generate S, and vectorial sum of *L* and *S* is *J*. A given term produces the number of microstates which are simply (2S + 1) ×

Received: September 2022; Accepted: December 2023
*Corresponding Author: Saba Javaid <hars4@yahoo.com>

(2L + 1) and by applying Hund's rule, ground state term and the order of stability can be determined. A state of definite *J* can be achieved when *L* and *S* coupled together and J's allowed values range from |L + S| to |L – S| [4, 5].

$$\mathbf{L} = \sum_{\mathbf{i}} \mathbf{l_i} \quad (1)$$

$$\mathbf{S} = \sum_{\mathbf{i}} \mathbf{s_i} \quad (2)$$

Each term spit into (**2J** + **1**) terms with energy difference proportional to the applied field intensity (Zeeman Effect) and a quantum number $\mathbf{M_J}$ that can have the values **J, J-1, J-2, …., -J** is used to define states. As a result, **J** can take **2S** + **1** values when **L ≥ S** but **2L + 1** values when **L** < **S** and **J** can only take one value when **L** = **0**. The valence electrons' energy levels may be defined if the possible L, S and J values are known [6]. Energy associated with the state of an atom taking part in a transition is described by an R-S spectral term, and energy levels in an atom with many electrons are briefly described by term symbols.

When an ion or atom is placed into a lattice, electronic repulsion splits the degenerate state into two or more states. Equivalent electrons are ones whose l and n values are the same such as $\mathbf{np^2}$, $\mathbf{nd^6}$, $\mathbf{nf^4}$, or $\mathbf{nf^6}$ configuration [7]. As a result, identical terms are produced for $\mathbf{nf^4}$ and $\mathbf{nf^{10}}$ configurations. The number of microstates for the sub shell increases when the number of electrons in orbital of the incomplete sub shell increases, but the non-equivalent electronic system produces a number of microstates that is much greater than that of the similar equivalent electronic system [8-13]. In 2019, Javaid *et al.* [14] evaluated 187 spectral terms with 457 *J* values and 106 wavefunctions for *4f2 5d2* configuration of Praseodymium II by using Russell – Saunders technique. Zafar *et al.* [15] determined 46 orthonormal wave functions for $4f^3$ $6s^2$ ground state configuration of Praseodymium I using spectral terms in 2020.

The main purpose of our study is to develop a machine algorithm using the Python language to evaluate term symbols of equivalent and non-equivalent electronic configurations. This program provides a user input interface for saving datasets into the program directory. It asks either to enter '0' for equivalent electron configuration or to hit '1' for non-equivalent electron configuration.

2. METHODOLOGY

In this article, the term symbols and the microstates of filled and half-filled subshell electronic configurations are calculated. Microstates are the numerous ways in which electrons can be arranged in a set of orbitals, each of which has a unique energy. The total number of microstates (W) of a system is the total number of definite arrangements for "e" number of electrons to be placed in "n" number of possible orbital positions. The number of microstates for equivalent electrons can be calculated by using a simple expression:

$$\mathbf{W} = \frac{\mathbf{n}!}{\mathbf{e}!\,(\mathbf{n}-\mathbf{e})!} \ldots\ldots\ldots\ldots (\mathbf{3})$$

where e denotes the number of electrons, and n is the total available orbitals [16].

Spectral Term corresponds to energy states and provides knowledge of angular momenta. For the non-equivalent electrons, there are more available microstates than for the equivalent electrons. Some of the available microstates for non-equivalent electrons are forbidden for equivalent electrons because of Pauli's principle [17, 18]. In Russell Saunders Coupling Scheme, term symbols are provided by (2S+1) LJ, where S shows the total spin angular momentum, L denotes the orbital angular momentum and J symbolizes the total angular momentum [19-21]. L can have the following values in a term symbol: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9,…, and the English capital letters "S, P, D, F, G, H, I, K, K" are used to represent each value of *L*. 2S+1 denotes the spin multiplicity of the spectral terms like singlet, doublet, triplet and so on. The Russell-Saunders technique makes the assumption that spin–orbit coupling < orbit-orbit coupling < spin-spin coupling [22, 23].

A computer algorithm was designed that generates terms of equivalent electrons and non equivalent electrons of any configuration. A user inputs either '0 and 1' for generation of microstates of equivalent and non-equivalent electrons. An algorithm flowchart has been given in Figure 1. It further requires orbital numbers (e.g., 0, 1, 2, 3 for *s, p, d, f* respectively) as input. These inputs must be given for the task's completion, or it may lead to the failure of the program. Permitted terms for the configuration of s^u, p^v, d^x, **and** f^y where **u=1 and 2, v = 1 to 5, x = 1 to 9 and y =1 to 13** are found by considering these equivalent configuration. The configurations ***1s2s, sp, sd, ss, 2p3p, pd, pf, 3d4d, df, 4f5f*** have non-equivalent electrons. Term symbols for these non equivalent electronic system are generated by calculating total orbital angular momentum and total spin angular momentum. For non equivalent electrons, Pauli's Exclusion Principle is not taken into account [24, 25].

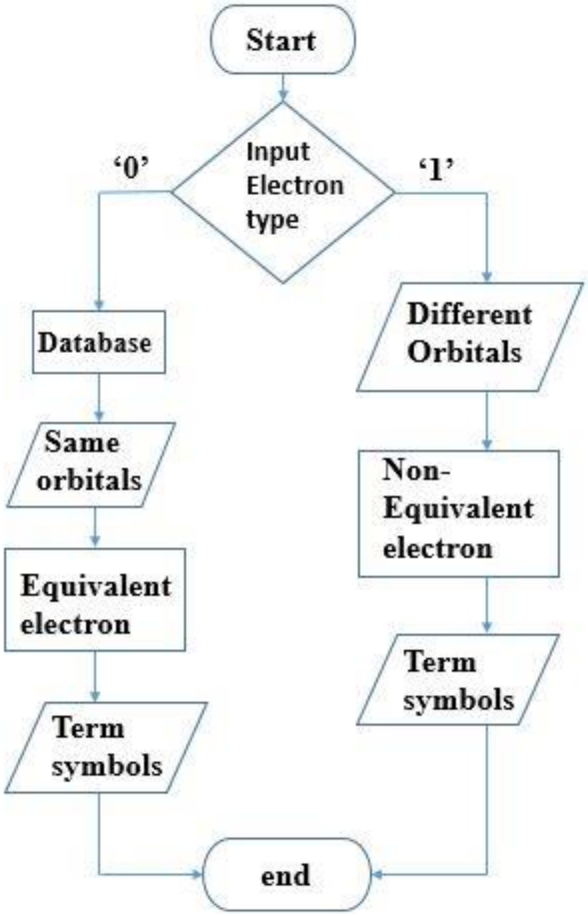

Fig. 1. Algorithm flowchart

3. RESULTS AND DISCUSSION

The filling of electrons in a shell depends on the nature of electrons, the electrons having same principal quantum number and same orbital quantum number face the limitations in filling the energy level slots, As they have to follow the Pauli's principle, no two electrons having same four set of quantum number can occupy the same energy level. Hence the number of microstates in case of equivalent electrons are less than that are available for non-equivalent electrons.

We used the theory of the coupling of orbital and spin angular momenta. Using a Python programme, we evaluated the term symbols for both equivalent and non-equivalent electrons.These configurations ***s^u, p^v, d^x*, and *f^y*** have equivalent electrons (where **u=1 and 2, v = 1 to 5, x = 1 to 9 and y =1 to 13**) and the microstates for these configurations are calculated by our program and are given in Table 1. Moreover, the configurations ***1s2s, sp, sd, ss, 2p3p, pd, pf, 3d4d, df, 4f5f***have non-equivalent electrons, the microstates for these configurations are found by using combination rule for two non-equivalent orbitals. Table 2. shows the some of the microstates for two non-equivalent orbitals.

For example $l_1 = 1, l_2 = 1$ then L= 2, 1, 0 using $L = |l_1 + l_2|,\ldots, |l_1 - l_2|$ so the states *D, P*, and *S* will be generated. To find the multiplicity, we find the

Table 1. Term symbols for equivalent electrons

Orbitals	No. of Electrons	Total No. of Microstates	Final No. of Microstates	Term Symbols
S	1	2	(0, 0.5)	^{2}S
	2	1	(0, 0.0)	^{1}S
P	1	6	(1, 0.5)	^{2}P
	2	15	(2, 0.0), (1, 1.0), (0, 0.0)	^{1}D, ^{3}P, ^{1}S
	3	20	(2, 0.5), (1, 0.5), (0, 1.5)	^{2}D, ^{2}P, ^{4}S
D	1	10	(2, 0.5)]	^{2}D
	2	45	(4, 0.0), (3, 1.0), (2, 0.0), (1, 1.0), (0, 0.0)	^{1}G, ^{3}F, ^{1}D, ^{3}P, ^{1}S
	3	120	(5, 0.5), (4, 0.5), (3, 1.5), (3, 0.5), (2, 0.5), (2, 0.5), (1, 1.5), (1, 0.5)	^{2}H, ^{2}G, ^{4}F, ^{2}F, ^{2}D, ^{2}D, ^{4}P, ^{2}P
F	1	14	(3, 0.5)	^{2}F
	2	91	(6, 0.0), (5, 1.0), (4, 0.0), (3, 1.0), (2, 0.0), (1, 1.0), (0, 0.0)	^{1}I, ^{3}H, ^{1}G, ^{3}F, ^{1}D, ^{3}P, ^{1}S
	3	364	(8, 0.5), (7, 0.5), (6, 1.5), (6, 0.5), (5, 0.5), (5, 0.5), (4, 1.5), (4, 0.5), (4, 0.5), (3, 1.5), (3, 0.5), (3, 0.5), (2, 1.5), (2, 0.5), (2, 0.5), (1, 0.5), (0, 1.5)	^{2}L, ^{2}K, ^{4}I, ^{2}I, ^{2}H, ^{2}H, ^{4}G, ^{2}G, ^{2}G, ^{4}F, ^{2}F, ^{2}F, ^{4}D, ^{2}D, ^{2}D, ^{2}P, ^{4}S

Table 2. Term symbols for non-equivalent electrons

1st Orbital	1st Orbital Number	2nd Orbital	2nd Orbital Number	Total Orbital Quantum Number	Total Spin Quantum Number	Multiplicity	Final Microstates
s	0	s	0	[0]	[0. 1.]	[1, 3]	['1S', '3S']
s	0	p	1	[1]	[0. 1.]	[1, 3]	['1P', '3P']
s	0	d	2	[2]	[0. 1.]	[1, 3]	['1D', '3D']
s	0	f	3	[3]	[0. 1.]	[1, 3]	['1F', '3F']
p	1	p	1	[0 1 2]	[0. 1.]	[1, 3]	['1S', '1P', '1D', '3S', '3P', '3D']
p	1	d	2	[1 2 3]	[0. 1.]	[1, 3]	['1P', '1D', '1F', '3P', '3D', '3F']
p	1	f	3	[2 3 4]	[0. 1.]	[1, 3]	['1D', '1F', '1G', '3D', '3F', '3G']
d	2	d	2	[0 1 2 3 4]	[0. 1.]	[1, 3]	['1S', '1P', '1D', '1F', '1G', '3S', '3P', '3D', '3F', '3G']

total spin generated by two electrons each having spin half. The total spin will be 1 and 0, that would lead to a multiplicity of 3 and 1. Therefore, the terms generated by two non-equivalent p-electrons are ^3D, ^3P, ^3S, ^1D, ^1P, and ^1S. The GitHub program repository can be accessed at https://github.com/AhmedAliRajput/Term-Symbol-Calculation.git.

4. CONCLUSIONS

A computer program is developed in Python to generate term values of various electronic configurations of atoms and ions. Russell Saunder coupling scheme is implemented for equivalent, non-equivalent, and combinations of both electrons in open and close shells of atoms and ions. It is very complicated to calculate terms using pen and paper, even for three equivalent electrons in a *d* shell. With the help of this program the task can be completed in a few seconds. This term calculator is useful for students working in the field of spectroscopy or quantum chemistry. For example, if students require to determine the microstates of np^2 configuration that is the case of equivalent electrons, generates 3 microstates that accommodates 18 electrons, wherease for np-(n+1)p configuration of non-equivalent electrons, program will generate 6 microstates that accommodate 36 electrons. If microstates for higher orbitals is to determine, it would cost a lot of time, therefore, this programs would be handy tool to deal with such lengthy calculations. The code can be downloaded from https://github.com/AhmedAliRajput/Term-Symbol-Calculation.git.

5. CONFLICT OF INTEREST

The authors declare no conflict of interest.

6. ACKNOWLEDGEMENTS

We would like to express our sincere gratitude to our research supervisor Professor Dr. Zaheer Uddin, Department of Physics, University of Karachi, Pakistan for his invaluable guidance throughout this research work.

7. REFERENCES

1. P.L. Meena, P.K. Jain, N. Kumar, and K.S. Meena. A comparative study of the atomic term symbols of f^3 and f^{11} configuration. *Acta Chimica and Pharmaceutica Indica* 2(1): 32-45 (2012).
2. P.L. Meena, N. Kumar, A.S. Meena, and K.S. Meena. Comparative studies on Russell-Saunders atomic term symbols (terms) for equivalent electrons of nf^4 and nf^{10} configurations. *Bulgarian Chemical Communications* 46: 141-149 (2014).
3. P. Armentrout. Chemistry of excited electronic states. *Science* 251(4990): 175-179 (1991).
4. E.S.J. Gauerke, and M.L. Campbell. A simple, systematic method for determining j levels for jj

coupling. *Journal of Chemical Education* 71(6): 457 (1994).

5. J.A. Olson. Atomic term symbols via partitioning techniques. *International Journal of Quantum Chemistry* 111(12): 2844-2850 (2011).
6. W. Demtröder. Atoms, molecules and photons. *Springer Heidelberg Dordrecht London New York* (2010).
7. A.K Bose, and A. Bose. Combinatorial computation of atomic terms for equivalent electrons. *Computer Physics Communications* 13(4): 241-246 (1977).
8. R.F. Curl Jr, and J.E. Kilpatrick. Atomic term symbols by group theory. *American Journal of Physics* 28(4): 357-365 (1960).
9. N. Karayianis. Atomic terms for equivalent electrons. *Journal of Mathematical Physics* 6(8): 1204-1209 (1965).
10. E. Tuttle. Terms obtained from configurations of equivalent electrons. *American Journal of Physics* 35(1): 26-29 (1967).
11. B. Judd. Atomic shell theory recast. *Physical Review* 162(1): 28 (1967).
12. Y. Liu, Y. Liu, and B. Liu. A New Method for Obtaining Russell– Saunders Terms. *Journal of Chemical Education* 88(3): 295-298 (2011).
13. C.J. Foot. Atomic physics. *Oxford University Press, Oxford, USA* (2004).
14. S. Javaid, R. Zafar, and Z. Uddin. Theoretical analysis of $4f^2 5d^2$ configuration of singly ionized praseodymium. *Journal of Physics Communications* 3(9): 095012 (2019).
15. R. Zafar, S. Javaid, and Z. Uddin. Wavefunctions for ground state $4f^3\ 6s^2$ configuration of praseodymium to calculate energy of fine levels and other spectroscopic quantities. *Journal of Physics Communications* 4(3): 035003 (2020).
16. P.L. Meena. Terms for Non-equivalent Electrons in $d^x p^y s^z$ Configurations. *Journal of Science Education* 19: 52 (2018).
17. H.A. Murthy, and V.M. Rao. A Study on Atomic Spectroscopic Term Symbols for Non-equivalent Electrons of (n-1) d s p Configuration Using Russell-Saunders Coupling Scheme. *International Journal of Engineering & Research* 7: 27-33 (2017).
18. M. Suhail, and I. Ali. A magic trick for determination of ground state term of s, p & d-orbital electrons. *Chemistry International* 6(4): 218-223 (2020).
19. P.L. Meena, P.K. Jain., N. Kumar, K.S. Meena, R. Goyal. Study on the Atomic Term Symbols for f^4 (M^{+3} free ion) Configuration. *Journal of Chemical Biological and Physical Science* 1(2): 188-203 (2011).
20. J.A. Odutola. Review of Term Symbols Computation from Multi-Electron Systems for Chemistry Students. *Contemporary Chemistry* (2019).
21. E. Kiremire. A numerical algorithm technique for deriving Russell-Saunders (RS) terms. *Journal of Chemical Education* 64(11): 951 (1987).
22. L. Guofan, and M. Ellzey. Finding the terms of configurations of equivalent electrons by partitioning total spins. *Journal of Chemical Education* 64(9): 771 (1987).
23. Y. Ge. On Teaching Molecular Term Symbols: From the Atomic Term Symbols in a Two-Dimensional World to the Molecular Term Symbols in a Three-Dimensional World. *Journal of Chemical Education* 95(9): 1682-1683 (2018).
24. M. Ivanova, and L. Dospatliev. Use of Vectors and Factorials for Calculation of Microstates from p^2, d^2, f^2 Configurations and Description of the p^2 Electron Configuration. *Journal of Chemistry. Natural Sciences in Education*, 26(3): 677-685 (2016).
25. R.B. Jha, W. Rehman, S. Kumar, and S. Lamba. Assigning Atomic Term Symbols for d-block Elements-A. *International Conference on Chemistry Frontiers & Challenges* 2:(3) (2013).

Proceedings of the Pakistan Academy of Sciences: A Physical and Computational Sciences 60(4): 35-44 (2023)

ISSN (Print): 2518-4245; ISSN (Online): 2518-4253
https://doi.org/10.53560/PPASA(60-4)820

Pakistan Academy of Sciences

Research Article

Measuring the Performance of Supervised Machine Learning Algorithms for Optimizing Wheat Productivity Prediction Models: A Comparative Study

Malik Muhammad Hussain[1], Farrukh Shehzad[2], Muhammad Islam*[3], Ashique Ali Chohan[4], Rashid Ahmed[2], and H.M. Muddasar Jamil Shera[3]

[1]Department of Statistics, Emerson University Multan, Pakistan
[2]Department of Statistics, the Islamia University of Bahawalpur, Punjab, Pakistan
[3]Crop Reporting Service, Agriculture Department, Punjab, Pakistan
[4]Department of Energy and Environment, Faculty of Agricultural Engineering, Sindh Agriculture University Tando Jam, Pakistan

Abstract: The issue of precise crop prediction gained worldwide attention in the midst of food security concerns. In this study, the efficacies of different machine learning (ML) algorithms, i.e., multiple linear regression (MLR), decision tree regression (DTR), random forest regression (RFR), and support vector regression (SVR) are integrated to predict wheat productivity. The performances of ML algorithms are then measured to get the optimized model. The updated dataset is collected from the Crop Reporting Service for various agronomical constraints. Randomized data partitions, hyper-parametric tuning, complexity analysis, cross-validation measures, learning curves, evaluation metrics and prediction errors are used to get the optimized model. ML model is applied using 75% training dataset and 25% testing datasets. RFR achieved the highest R^2 value of 0.90 for the training model, followed by DTR, MLR, and SVR. In the testing model, RFR also achieved an R^2 value of 0.74, followed by MLR, DTR, and SVR. The lowest prediction error (P.E) is found for the RFR, followed by DTR, MLR, and SVR. K-Fold cross-validation measures also depict that RFR is an optimized model when compared with DTR, MLR and SVR.

Keywords: Model Optimizations, Machine Learning Algorithms, Prediction Models, Performance Measurement.

1. INTRODUCTION

Data science is an interdisciplinary field that comprises scientific computing, processes, algorithms and systems to extract meaningful insights and knowledge from data [1, 2]. Machine learning (ML) is viewed as an advanced tool of data science and artificial intelligence and focuses on developing algorithms and statistical models that enable computers to perform tasks without explicit programming [3, 4]. Arthur Samuel, a pioneer in ML, defined ML as "field of study that gives computers the ability to learn without being explicitly programmed", and used to learn the machines, how to handle the data in more efficient way [5, 6]. ML algorithms created new dimensions and possibilities for data intensive research in the applied agricultural science [7-9]. Various supervise machine learning models, i.e., linear regression, decision tree, random forest and artificial neural networks, etc., are used for the prediction of parameters using agricultural constrains [10, 11].

Global food supply system is facing challenges due to rapid population growth and unpredictable changes in climate [12, 13]. Agriculture being core component, got key significance in the midst of food security situation in the world, and the focus of the agricultural research is diverted towards the improvement and true prediction of crop productivity [14-16]. The high growth rate of population, increases the demand of staple food crop [17]. Wheat, maize and rice are the world's significant cereals crops in terms of area and food supply channels [18]. Wheat is important food crop

Received: May 2023; Accepted: December 2023
* Corresponding Author: Muhammad Islam <mislam6667@gmail.com>

in the developing world after rice [19, 20]. Wheat crop, ranks first in acreage and production among all others food crops [21].

Chlingaryan *et al.*, [22] conducted a study using various machine learning techniques to predict the crop yield, and they concluded that ML approaches provide comprehensive and effective solutions for better crop projection. Bondre and Mahagaonkar [23] studied the efficacies of support vector and random forest machine learning algorithms to predict the crop yield using agricultural dataset of fertilizer application. They demonstrated that random forest is better choice to predict the crop yield.

Cedric *et al.*, [24] proposed prediction system using machine learning algorithms to predict the crops yield in African countries using weather, climatic and chemical datasets. They applied decision tree, multivariate regression, k-nearest neighbor models. They performed hyper-parameter tuning and cross-validation to optimize the model performance to avoid over fitted model. The results indicated that decision tree model is better fitted model. Mohapatra and Chaudhary [25] applied machine learning models, i.e., multiple linear regression, decision tree regression, random forest regression, k-nearest neighbor approach and support vector regression. They applied evaluation metrics to measure the model performances and reported that random forest regression performed well.

Currently, Food supply system in the world is facing major stress due to food insecurity. The advancement in agricultural science and technology leaded immense volume of data. Machine learning algorithms have been emerged as advanced tools of artificial intelligence and applied as an alternative of traditional statistical modeling, to extract meaningful information inside from data using algorithms. The optimization and validations of machine learning algorithms is one of the key concerns for the world, to predict the true agricultural productions. This research will focus, to apply and to integrate different machine learning algorithms such as multiple regression, decision tree, random forest and support vector regression with the aim to search out, which one is best, and able to validate and to optimize the learning algorithms, to predict the wheat productivity in Pakistan using various agronomical constrains.

2. MATERIALS AND METHODS

2.1 Data Collection and Variables Features

The secondary cross-sectional data of 15000 crop cut experiments (CCE), conducted by the Crop Reporting Service is collected along with different agronomical constrains comprise from year 2019 to 2021. This data is further pre-processed and centralized using the centroid clustering scheme introduced by Islam and Shehzad [14] for the 136 wheat area sown tehsils zones in Punjab. The experiment is performed using the python' key library called Scikit Learn (https://scikit-learn.org/stable/supervised_learning.html). The wheat average yield in mds/acre is used as depended (labeled) variable along with agronomical constrains, i.e., urea, DAP and other fertilizers used in kg/acre, water, spray pest, soil type (chikny loom or not), adoption of advanced varieties trend (yes/no), wheat harvested from April 1 to 20 (yes/no), and wheat sown in November (yes/no).

2.2 Hierarchy of Machine Learning Algorithms

The layout hierarchy of supervised machine learning model deployment fallows under the following seven prominent steps.

2.2.1 Steps-1: Data Collection

Collection of the quality data relevant to the problem.

2.2.2 Steps-2: Data Preparation

Accurate data format is essential to apply machine learning algorithms. Sometimes, collected dataset is found incomplete and inadequate. Data preparation is a machine learning tool used to gather, combine, clean, complete, transform and feature selection of raw data in well shape to make true prediction [14, 26, 27].

2.2.3 Steps-3: Choosing of Machine Learning Model

There are three types of machine learning algorithms, i.e., supervised machine learning, unsupervised machine learning and reinforcement machine learning [5, 28]. The following condition is applied to choose the ML model. When class label (predicted variable) is categorical, the well-known algorithms applied as decision tree, random forest, support vector machines and artificial neural networks, etc. When class label (predicted variable) is real values or continuous, the well-known algorithms applied are linear regression, decision tree regression, random forest regression and support vector regression, etc.

2.2.4 Steps-4: Mutually Exclusive Randomized Data Partition

The whole dataset *(d)* is split into two mutually exclusive randomized datasets called training and testing datasets [29, 30]. The training dataset (d_{train}) used to train the ML model. For the current study 75% of datasets is used to train the model. The remaining 25% used as test dataset ($d_{test} = (d) - d_{train}$) and used to evaluate the ML model performance for unseen datasets.

2.2.5 Steps-5: ML Model Evaluation, Model Complexity, Over and Under-Fit Model

Evaluation of ML model is assessed for training and testing algorithm. The ML model error found in training datasets called bias or in sample error or training error. The ML model error found in testing dataset called testing error or out-of-sample error or variance. The model with lowest values of root mean square error (RMSE), mean square error (MSE), mean absolute error (MAE) and highest performance score (R^2), viewed best model. A model depicts low bias and high variance called over fit, while a model depicts high bias and high variance called under fit. Bias and variance tradeoff is common property of ML model [31, 32]. Optimum model complexity means a condition, where model depict optimum lowest range of bias and variance (Figure 1). The following equation is used to determine the prediction error (PE).

$$y = h(x) + e \quad (1)$$

Where *"e"* is error term. Let the estimate of *"h(x)"* is *"$h^*(x)$"* and the expected prediction squared error at *"x"* is determine as under:

$$\text{PE}(x) = E[(y - h^*(x))^2] \quad (2)$$

The prediction error further decomposed into bias and variance components as:

$$\text{P.E}(x) = \text{Var}[h^*(x)] + [\text{Bias}\{h^*(x)\}]^2 + \text{Var(error)} \quad (3)$$

$$\text{P.E}(x) = \text{variance} + \text{bias}^2 + \text{error} \quad (4)$$

2.2.6 Step-6: Hyper Parametric Tuning of ML Model

The ML models involve different parameters learning rates, etc. The process to extract the optimum values of these parameters learning rates called hyper-parameters tuning of ML problem. Hyper parameter tuning is a way to produce optimized ML model to avoid under and over fit conditions [33, 34]. The prominent ML function called GridSearchCV is used to tune the ML model.

2.2.7 Step-7: Model Deployment for Final Predictions

ML algorithms are applied after data preparation and hyper parametric tuning, and used to test/unseen data with the aim to make the true prediction for the dependent variable (wheat productivity). For the current study the following ML models are applied.

- Multiple linear regression
- Decision tree regression
- Random forest regression
- Support vector regression

2.3 Multiple Linear Regression Model (MLR)

The entire dataset is split into two mutually exclusive datasets, i.e., training and testing datasets. Multiple linear regression model is deployed for the dependent variable (wheat productivity) using the features (independents variables).

$$y_i = \beta_i z_i + \varepsilon \quad (5)$$

Where *"y_i"* stands wheat productivity in mds/acre, *"z_i"* stands for the features, *"β_i"* stands for regression coefficients of features.

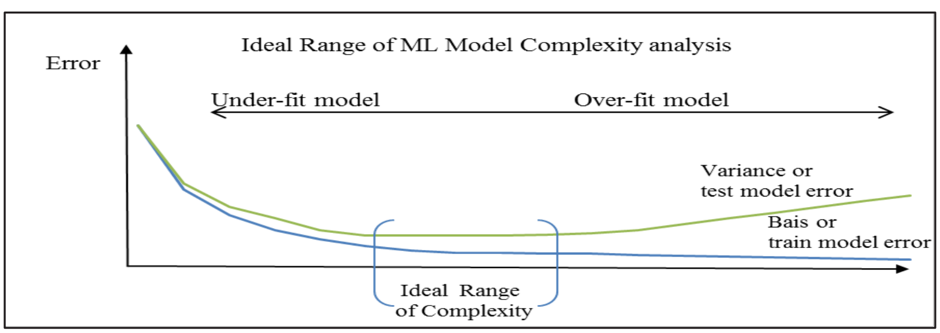

Fig. 1. Ideal range of model complexity for ML model

2.4 Decision Tree Regression Algorithm (DTR)

Decision tree is supervised machine learning algorithms used to breaks down the dataset into smaller homogenous subset using flowchart structure of root-to-leaf direction [35, 36]. Figure 2 shows the root-to-leaf direction flowchart of decision tree algorithms. The each homogenous subset is associated with specific decision about the whole data. The roots node is split into intermediate nods based on specifics characteristic of root node, and signifies a test. The intermediate node is further split into more intermediate or leaf nodes based on specifics characteristic of previous intermediate nodes, and each leaf node signifies the decision or outcomes. For the current study, the root nodes are constructed for the wheat yield using the different agronomical constrains. The mean square error (MSE) is used to build the structure of root-to-leaf nodes.

2.5 Random Forest Regression

Random forest regression (RFR) is supervised machine learning algorithm that combined the large set of decision trees [37-39]. RFR search has the best features and uses precision techniques to make the forest random. Figure 3 shows the flowchart structure of RFR algorithm. For RFR, decision tree is constructed for a random set of variable dataset. In RFR, various trees are merged together by taking the average of the prediction trees. RFR algorithm is based on three hyper parameters (no. of trees forest, no. of features and maximum depth of the tree). Machine learning function named GridSearchCV is used to tune the RFR hyper parameters.

2.6 Support Vector Regression (SVR)

Support vector is a non-linear generalization of generalized portrait algorithm developed by Vladimir Vapnik and his colleagues [40, 41]. SVR is based on the structural risk minimization principle for minimizing a bound on the generalization prediction error and to fit the optimum hyper plane with the aim to have maximum number of points within the decision boundary line. SVRs are limiting the prediction error margin (ε) and search to fit maximum points between the decision boundary

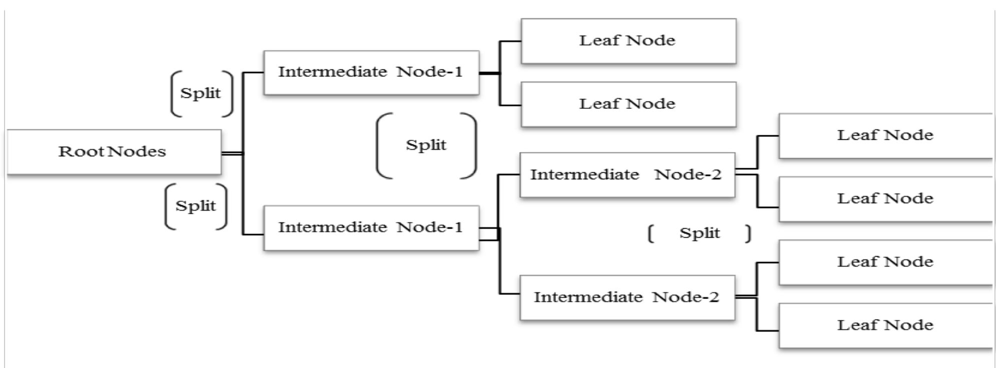

Fig. 2. Root-to-leaf direction flowchart of decision tree algorithms

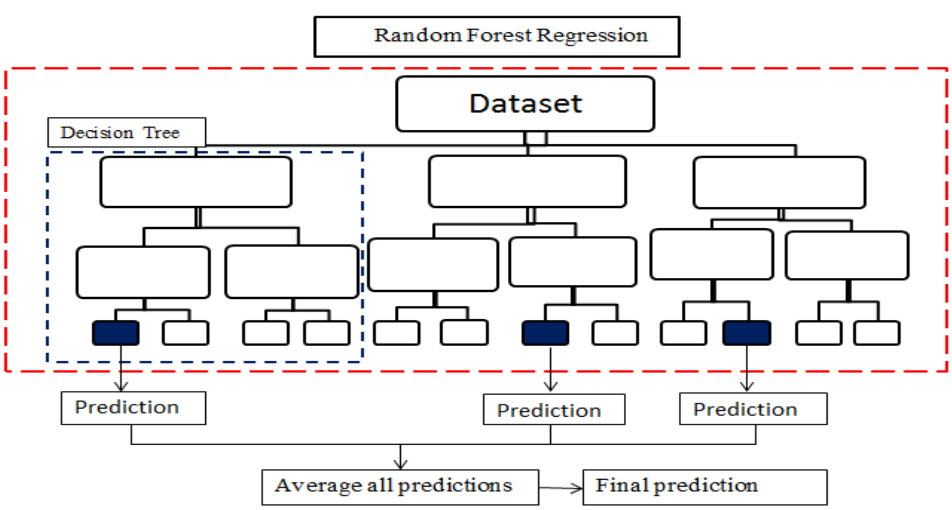

Fig. 3. Flowchart structure of RFR algorithm

lines. The SVRs are dependent on different kernel function such as linear, radial basis function (RBF) and polynomial. RBF kernel can better fit and handle the situation in case of nonlinear relation. Smola and Scholkopf [41] defined for a given train set *[(x_1, y_1),…,(x_ℓ, y_ℓ)]* ⊂ X × R, where *"X"* is the space of the input patterns *($X = R^d$)*. The basic aim of this technique is to obtain *"f(x)"* has at most one *"ε"* deviation obtained from the current targets *"y_i"* at the time and that is as flat as possible. The flatness is determine by the small value of *"w"* and the problem can be written as:

$$f(x) = (w, x) + b \ \ with \ \ w\epsilon x, b\epsilon R \qquad (6)$$

A linear function that approximate all pairs (x_i, y_i) with precision (*ε*).

$$\text{Min } \frac{1}{2}||w||^2 \qquad (7)$$

$$\text{Subject to} \begin{cases} y_i - \langle w, x_i \rangle - b \le \epsilon \\ \langle w, x_i \rangle + b - y_i \le \epsilon \end{cases}$$

Figure 4 and Figure 5 show the structure for SVR. The key hyper parameters in SVR are noted as:

- ***Hyperplane***: Hyper plane means a line used to separate two datasets
- ***Support vector***: Datasets points lies on either side of the hyper plane as closest as, called support vectors.
- ***Kernel***: Kernel means a set of mathematical functions used to transform input data into required forms and it is generally used to search optimum hyper plane.
- ***Decision boundary lines:*** The two lines drawn around the hyper plane at a distance of "ε" and used to create a margin between the dataset points.

2.7 K-Fold Cross-Validations Measures

Cross-validation is resampling, out of sample testing or rotation estimations, that partitioning the data into complementary subsets. In K-Fold cross validations various rounds of train test split are applied and average of all rounds of partitions is taken to assess the model performance [42-44]. For K-Fold cross- validation, various learning disjoint subsets of equal size *"k"* is made and model is trained for *"k-1"* subsets, and test for complementary subset. For the current study "10" fold cross- validation is applied. Figure 6 shows the flowchart for the application of K-Fold cross-validation.

3. RESULTS AND DISCUSSION

3.1 Randomized Data Partitions

For the current study collected datasets is complete and free from any inadequacy. Data partition is applied using the randomized data splitting. The 75% datasets is used to train the model and 25% is used to test the model. The centroid clustering scheme is applied to prepare the dataset of 136 tehsils zones in Punjab for wheat crop CCE. The wheat crop CCE of 102 tehsils is split to train the

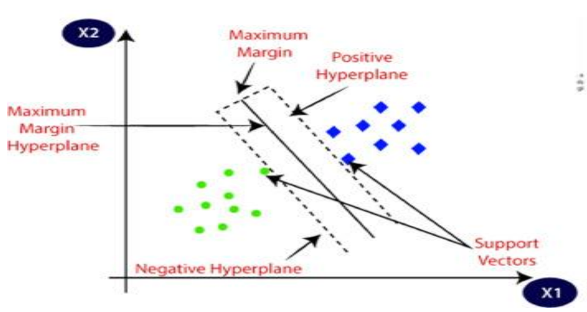

Fig. 4. Structure of hyper plane for SVR

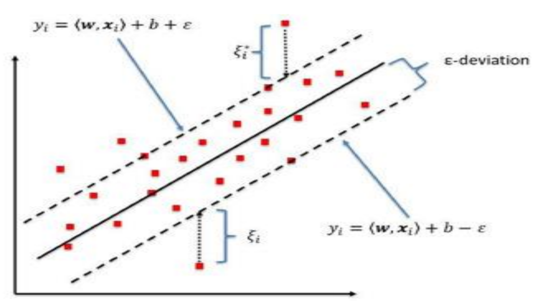

Fig. 5. Prediction error margin of SVR

K-Fold cross-validation using randomized data partitions							
Average of Fold	Fold-1	Test	Train	Train	Train	Train	Test score -1
	Fold-2	Train	Test	Train	Train	Train	Test score -2
	Fold-3	Train	Train	Test	Train	Train	Test score -3
	Fold-4	Train	Train	Train	Test	Train	Test score -4
	Fold-5	Train	Train	Train	Train	Test	Test score -5

Fig. 6. Flowchart structure of K-Fold cross-validation

model, and 34 tehsils is split to test the model performance for unseen datasets.

3.2 Hyper Parametric Tuning of Machine Learning Models

Figure 7 shows the machine learning model complexity for DTR and RFR. Learning curves show high variance and high bias at lower model complexity (under fit). The learning curves reached their lowest point at the ideal range of model complexity with an optimum depth value of 4. However, the test curve rises again (over fitting) after reaching the lowest point in the optimum range of model complexity. GridSearchCV, a Python's key hyper parameter tuning library, produced the optimal value for the minimum sample split is 4, while the number of trees in the forest is found optimum at 10 for RFR.

3.3 Integrating the Performance of MLR, DTR, RFR and SVR

Table 1 shows the performance of multiple regression, decision tree regression, random forest regression and support vector regression using the machine learning approach. The highest value of R^2 found is 0.90 for RFR, followed by DTR, MLR and SVR for train model ($R^2_{RFR} > R^2_{DTR} > R^2_{MLR} > R^2_{SVR}$), while for the test model, it is 0.74, followed by MLR, DTR and SVR ($R^2_{RFR} > R^2_{MLR} > R^2_{DTR} > R^2_{SVR}$). Comparing the performance of MSE and MAE, the lowest MSE and MAE are reported for the RFR as 4.58 and 1.66, followed by DTR, MLR and SVR for the train model (MSE $(MAE)_{RFR}$ < MSE $(MAE)_{DTR}$ < MSE $(MAE)_{MLR}$ < MSE $(MAE)_{SVR}$). In the test model, RFR produced an MSE of 8.98 and an MAE of 2.36, followed by MLR, DTR, and SVR (MSE $(MAE)_{RFR}$ < MSE $(MAE)_{MLR}$ < MSE $(MAE)_{DTR}$ < MSE $(MAE)_{SVR}$). Integrating the performance of prediction error (P.E) for the MSE and MAE, lowest value of P.E found for the RFR is 13.56 for MSE and 4.02 for MAE, and followed by DTR, MLR and SVR. The relationship among the prediction error (P.E) values for RFR, DTR, MLR, and SVR is represented as follows $P.E_{RFR} < P.E_{DTR} < P.E_{MLR} < P.E_{SVR}$. Figure 8 shows the learning curves for the MSE and Figure 9 shows the learning curves for the MAE. It is found that RFR is better and optimized fitted model, as the lowest error is found for the RFR with highest performance score (R^2).

3.4 K-fold Cross-Validations Measures

Table 2 shows the K-Fold cross-validation measures for the MLR, DTR, RFR and SVR using the MAE and MSE criterion. For the MLR, the smallest values of MAE and MSE are found for K-Fold-5, with 2.25 and 7.65, respectively, while the largest values are found for K-Fold-7, with 3.74 for MAE

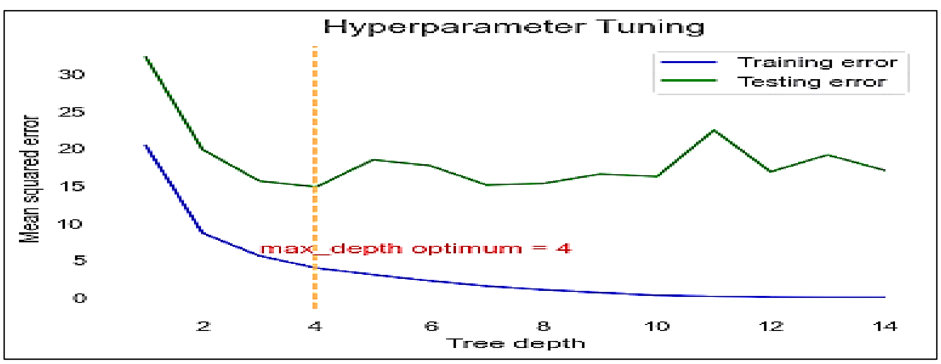

Fig. 7. Machine learning model complexity for DTR and RFR

Table 1. Performance of machine learning models

Evaluation Metrics	Train set				Test set			
	MLR	DTR	RFR	SVR	MLR	DTR	RFR	SVR
R^2	0.75	0.89	0.90	0.48	0.68	0.508	0.74	0.31
MSE	11.09	4.65	4.58	22.9	11.23	17.36	8.98	24.17
MAE	2.76	1.67	1.66	3.55	2.59	3.06	2.36	3.67
P.E (MSE)	22.32	22.01	13.56	47.07	--	--	--	--
P.E (MAE)	5.35	4.72	4.02	7.22	--	--	--	--

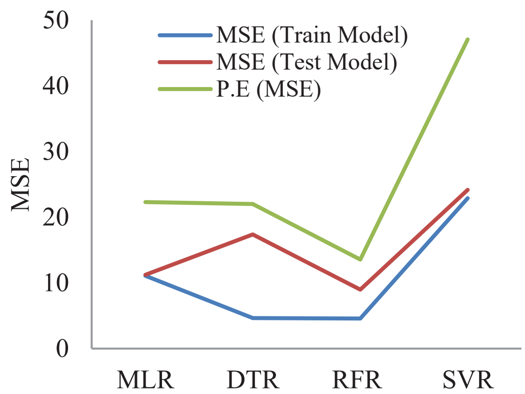

Fig. 8. Learning curves for the MSE

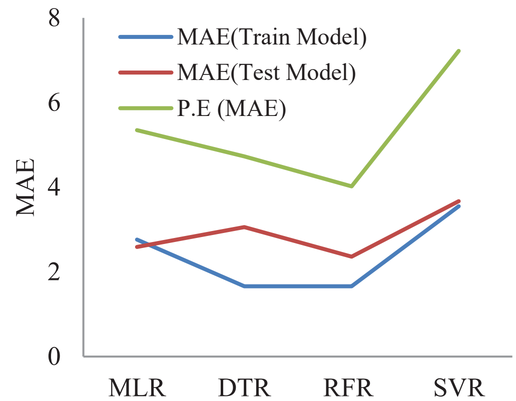

Fig. 9. Learning curves for the MAE

Table 2. K-Fold cross-validation measures for the MLR, DTR, RFR and SVR

ML models	Criterions	K-Fold average	K-Fold 1	K-Fold 2	K-Fold 3	K-Fold 4	K-Fold 5	K - Fold 6	K-Fold 7	K-Fold 8	K-Fold 9	K-Fold 10
MLR	MAE	2.97	2.65	3.22	2.30	2.45	2.25	3.14	3.74	3.16	3.10	3.74
	MSE	13.10	9.27	16.04	8.19	10.15	7.65	15.78	19.96	13.29	14.59	16.71
DTR	MAE	2.50	2.23	2.02	2.11	1.90	2.86	3.20	3.17	2.78	2.06	2.08
	MSE	10.56	7.55	12.12	7.38	6.96	12.05	16.16	15.97	9.96	7.82	9.00
RFR	MAE	2.29	2.08	2.24	2.18	1.55	1.99	2.24	3.49	2.79	2.04	2.45
	MSE	8.41	5.84	6.68	6.20	4.50	7.53	6.92	20.04	9.83	8.16	10.77
SVR	MAE	3.57	2.93	2.94	3.04	2.36	3.62	3.43	4.42	5.08	3.69	4.18
	MSE	21.80	14.55	12.72	20.93	7.17	21.73	21.22	29.26	35.69	21.89	32.93

and 19.96 for MSE. The lowest values of MAE and MSE are found at K-Fold-4, with 1.90 and 6.96 for DTR, 1.55 and 4.50 for RFR, and 2.36 and 7.17 for SVR, respectively. The largest values of MAE and MSE are found to be 3.20 and 16.16 for the DTR at K-Fold-6, 3.49 and 20.04 for RFR at K-Fold-7, and 5.08 and 35.69 for SVR at K-Fold-8. Figure 10 shows the learning curves of MAE and MSE for comparing RFR, DTR, MLR and SVR. RFR are found to perform better as it predicted the lowest learning curve compared to other machine learning models.

4. CONCLUSIONS

Machine learning (ML) emerged as an advanced tools of data science and artificial intelligence. ML is used to extract reliable information from data using various algorithms. The issue of precise crop prediction gained worldwide attention in the midst of food security concerns. ML models, i.e., linear regression, decision tree, random forest and support vector regression are applied as an alternative of statistical model in almost every field. True crop prediction got key significance in the midst of

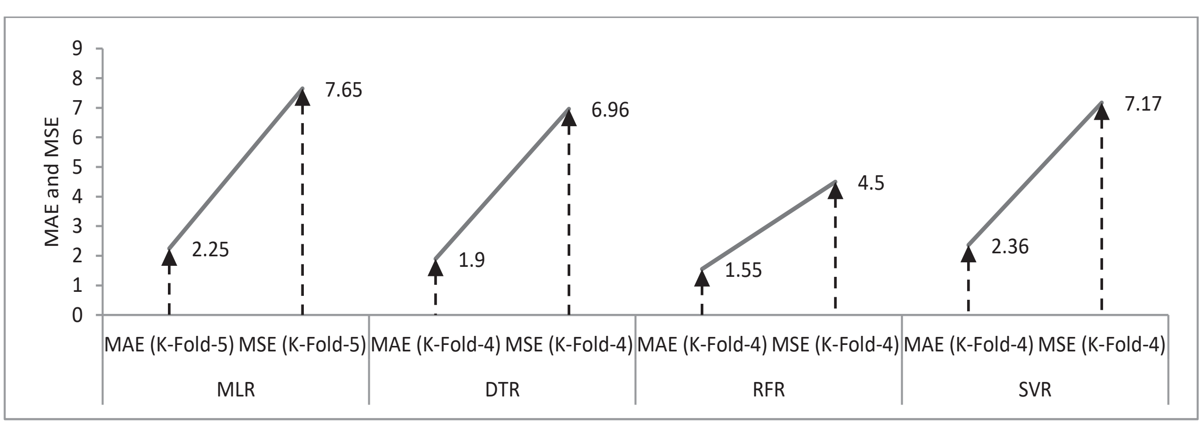

Fig. 10. Learning curves of MAE and MSE for K-Fold cross-validations measures

food security situation in the world. Validations and optimization of ML algorithms have become the key concerns for the world to predict the true agricultural productions. This study integrated the efficacies of different machine learning algorithms, i.e., multiple linear regression (MLR), decision tree regression (DTR), random forest regression (RFR), and support vector regression (SVR) with the aim to get optimized model to predict the wheat productivity in Pakistan using various agronomical constrains. The updated centroid clustering based dataset is collected from the Crop Reporting Service. The python's key library called ScikitLearn is used to analyze the results. Different machine learning tools such as randomized data partitions, hyper parametric tuning, complexity analysis and cross-validation measures are also performed to get the optimized model.

Highest value of R^2 is reported from RFR, followed by DTR, MLR and SVR for train model ($R^2_{RFR} > R^2_{DTR} > R^2_{MLR} > R^2_{SVR}$), and for test model, it is followed by MLR, DTR and SVR ($R^2_{RFR} > R^2_{MLR} > R^2_{DTR} > R^2_{SVR}$). Lowest MSE and MAE are reported for the RFR, followed by DTR, MLR and SVR for the train model (MSE $(MAE)_{RFR}$ < MSE $(MAE)_{DTR}$ < MSE $(MAE)_{MLR}$ < MSE $(MAE)_{SVR}$), and for the test model, it is followed by MLR, DTR and SVR (MSE $(MAE)_{RFR}$ < MSE $(MAE)_{MLR}$ < MSE $(MAE)_{DTR}$ < MSE $(MAE)_{SVR}$). Lowest value of prediction error (P.E) is found from RFR, followed by DTR, MLR and SVR both for the MSE and MAE ($P.E_{RFR} < P.E_{DTR} < P.E_{MLR} < P.E_{SVR}$). Learning curves of MAE and MSE depict that RFR is an optimized machine learning model compared to DTR, MLR, and SVR. K-Fold cross-validation results indicate that the lowest error is found for RFR when compared with DTR, MLR, and SVR. RFR found a validated and optimized model when compared with other machine learning models. This research can also be extended by applying other clustering approaches and machine learning algorithms.

5. CONFLICT OF INTEREST

The authors declare no conflict of interest.

6. ACKNOWLEDGEMENT

All authors acknowledge the strong data collection mechanism of Crop Reporting Service, Punjab being a sole source of agricultural crop statistics in Punjab, Pakistan.

7. REFERENCES

1. J.A. Polonsky, A. Baidjoe, Z.N. Kamvar, A. Cori, K. Durski, W.J. Edmunds, R.M. Eggo, S. Funk, L. Kaiser, P. Keating, O.P. Waroux, M. Marks, P. Moraga, O. Morgan, P. Nouvellet, R. Ratnayake, C.H. Roberts, J. Whitworth and T. Jombart. Outbreak analytics: a developing data science for informing the response to emerging pathogens. *Philosophical Transactions of the Royal Society B* 374: 1-11 (2019).
2. H. Yang. Building an Agile and Scalable Data Science Organization; (Chapter 3) In: Data Science, AI, and Machine Learning in Drug Development. *Chapman and Hall/CRC* (2022).
3. T.H. Davenport. From analytics to artificial intelligence. *Journal of Business Analytics* 1(2): 73-80 (2018).
4. Y, Nazarathy, and H. Klok. Statistics with Julia: Fundamentals for data science, machine learning and artificial intelligence. *Springer Nature* (2021).
5. B. Mahesh. Machine learning algorithms-a review. *International Journal of Science and Research (IJSR)* 9: 381-386 (2020).
6. C. Park, C.C. Took, and J.K. Seong. Machine learning in biomedical engineering. *Biomedical Engineering Letters* 8: 1-3 (2018).
7. N. Yadav, S.M. Alfayeed, and A. Wadhawan. Machine learning in agriculture: techniques and applications. *International Journal of Engineering Applied Sciences and Technology* 5(7): 118-122 (2020).
8. K. Patel, and H.B. Patel. A comparative analysis of supervised machine learning algorithm for agriculture crop prediction. *Paper presented at the 2021 Fourth International Conference on Electrical, Computer and Communication Technologies (ICECCT)* (2021).
9. A.M. Lad, K.M. Bharathi, B.A. Saravanan, and R. Karthik. Factors affecting agriculture and estimation of crop yield using supervised learning algorithms. *Materials Today: Proceedings* 62: 4629-4634 (2022).
10. K.G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis. Machine learning in agriculture: A review. *Sensors* 18(8): 1-29 (2018).
11. Y. Mekonnen, S. Namuduri, L. Burton, A. Sarwat, and S. Bhansali. Machine learning techniques in wireless sensor network based precision agriculture. *Journal of the Electrochemical Society* 167(3): 1-10 (2019).
12. E. Kamir, F. Waldner, and Z. Hochman. Estimating wheat yields in Australia using climate records, satellite image time series and machine learning methods. *ISPRS Journal of Photogrammetry and Remote Sensing* 160:124-135 (2020).
13. M.S. Ud-Din, M. Mubeen, S. Hussain, A. Ahmad, N. Hussain, M.A. Ali, A.E. Sabagh, M. Elsabagh, G.M Shah, S.A. Qaisrani, M. Tahir, H.M.R. Javeed, M.A. H. Ali and W. Nasim. World nations priorities on climate change and food security. *Springer* pp. 365-384 (2022).
14. M. Islam, and F. Shehzad. A prediction model optimization critiques through centroid clustering by reducing the sample size, integrating statistical and machine learning techniques for wheat productivity. *Scientifica* 1-11 (2022).

15. D. Elavarasan, and P.D.R. Vincent. A reinforced random forest model for enhanced crop yield prediction by integrating agrarian parameters. *Journal of Ambient Intelligence and Humanized Computing* 12: 10009–10022 (2021).
16. M. Islam, F. Shehzad, A. Qayyum, M.W. Abbas, and R. Siddiqui. Growth analysis of production of food crops and population growth for Food Security in Pakistan. *Proceedings of the Pakistan Academy of Sciences: B. Life and Environmental Sciences* 60(1): 83-90 (2023).
17. P. Feng, B. Wang, D.L. Liu, C. Waters, and Q. Yu. Incorporating machine learning with biophysical model can improve the evaluation of climate extremes impacts on wheat yield in south-eastern Australia. *Agricultural and Forest Meteorology* 275: 100-113 (2019).
18. O. Erenstein, M. Jaleta, K. Sonder, K. Mottaleb, and B. Prasanna. Global maize production, consumption and trade: trends and R&D implications. *Food Security* 14(5): 1295-1319 (2022).
19. P. Giraldo, E. Benavente, F. Manzano-Agugliaro, and E. Gimenez. Worldwide research trends on wheat and barley: A bibliometric comparative analysis. *Agronomy* 9(7): 1-18 (2019).
20. R. Ghimire, H. Wen-Chi, and R.B. Shrestha. Factors affecting adoption of improved rice varieties among rural farm households in Central Nepal. *Rice Science* 22(1): 35-43 (2015).
21. S. Ali, Y. Liu, M. Ishaq, T. Shah, A. Ilyas, and I.U. Din. Climate change and its impact on the yield of major food crops: Evidence from Pakistan. *Foods* 6(6): 1-19 (2017).
22. A. Chlingaryan, S. Sukkarieh, and B. Whelan. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Computers and Electronics in Agriculture* 151: 61-69 (2018).
23. D.A. Bondre, and S. Mahagaonkar. Prediction of crop yield and fertilizer recommendation using machine learning algorithms. *International Journal of Engineering Applied Sciences and Technology* 4(5): 371-376 (2019).
24. L.S. Cedric, W.Y.H. Adoni, R. Aworka, J.T. Zoueu, F.K. Mutombo, M. Krichen, and C.L.M. Kimpolo. Crops yield prediction based on machine learning models: case of west African countries. *Smart Agricultural Technology* 2: 1-14 (2022).
25. S. Mohapatra, and N. Chaudhary. Statistical analysis and evaluation of feature selection techniques and implementing machine learning algorithms to predict the crop yield using accuracy metrics. *Engineered Science* 21: 1-11 (2022).
26. V. Gudivada, A. Apon, and J. Ding. Data quality considerations for big data and machine learning: Going beyond data cleaning and transformations. *International Journal on Advances in Software* 10(1): 1-20 (2017).
27. A. Dogan, and D. Birant. Machine learning and data mining in manufacturing. *Expert Systems with Applications* 166(2): 114060 (2021).
28. M. Alloghani, D. Al-Jumeily, J. Mustafina, A. Hussain, and A.J Aljaaf. A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science. In: Supervised and Unsupervised Learning for Data Science. *Springer* pp. 3-21 (2020).
29. J.J. Salazar, L. Garland, J. Ochoa, and M.J. Pyrcz. Fair train-test split in machine learning: Mitigating spatial autocorrelation for improved prediction accuracy. *Journal of Petroleum Science and Engineering* 209: 109885 (2022).
30. J. Tan, J. Yang, S. Wu, G. Chen, and J. Zhao, A critical look at the current train/test split in machine learning. *arXiv preprint arXiv*:2106-04525 (2021).
31. M. Islam. Integrating Statistical and Machine Learning Techniques to Predict Wheat Production in Pakistan. Ph.D. Thesis, Department of Statistics, *The Islamia University of Bahawalpur, Pakistan* (2022).
32. P. Chakraborty, S.S. Rafiammal, C. Tharini, and D.N. Jamal. Influence of bias and variance in selection of machine learning classifiers for biomedical applications smart data intelligence. *Proceedings of ICSMDI, Springer:* 459-472 (2022).
33. A. Vabalas, E. Gowen, E. Poliakoff, and A.J. Casson. Machine learning algorithm validation with a limited sample size. *PLoS One* 14(11): e0224365 (2019).
34. L. Li, K. Jamieson, A. Rostamizadeh, E. Gonina, J. Ben-Tzur, M. Hardt, B. Recht and A Talwalkar. A system for massively parallel hyper parameter tuning. *Proceedings of Machine Learning and Systems* 2: 230-246 (2020).
35. F. Osisanwo, J. Akinsola, O. Awodele, J. Hinmikaiye, O. Olakanmi, and J. Akinjobi. Supervised machine learning algorithms: classification and comparison. *International Journal of Computer Trends and Technology (IJCTT)* 48(3): 128-138 (2017).
36. B.T. Jijo, and A.M. Abdulazeez. Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends* 2(1): 20-28 (2021).
37. M. Maduranga, and R. Abeysekera. Treeloc: An ensemble learning-based approach for range based indoor localization. *International Journal of Wireless and Microwave Technologies (IJWMT)* 11(5): 18-25 (2021).
38. K. Rawal, and A. Ahmad. A comparative analysis of supervised machine learning algorithms for electricity demand forecasting. *Paper presented at the 2022 Second International Conference on Power, Control and Computing Technologies (ICPC2T):* (2022).
39. F. Farooq, A.M. Nasir-Amin, K. Khan, M. Rehan-Sadiq, M.F. Javed, F. Aslam, and R. Alyousef. A comparative study of random forest and genetic engineering programming for the prediction of compressive strength of high strength concrete (HSC). *Applied Sciences* 10(20): 7330 (2020).
40. H. Drucker, C.J. Burges, L. Kaufman, A. Smola, and V. Vapnik. Support vector regression machines. *Advances in Neural Information Processing Systems* 9: 155-161 (1997).
41. A.J. Smola, and B. Scholkopf. A tutorial on support vector regression. *Statistics and Computing* 14(3):

199-222 (2004).
42. R.T. Nakatsu. An evaluation of four resampling methods used in machine learning classification. *IEEE Intelligent Systems* 36(3): 51-57 (2020).
43. B. Bischl, O. Mersmann, H. Trautmann, and C. Weihs. Resampling methods for meta-model validation with recommendations for evolutionary computation. *Evolutionary Computation* 20(2): 249-275 (2012).
44. G. Afendras, and M, Markatou. Optimality of training/test size and resampling effectiveness in cross-validation. *Journal of Statistical Planning and Inference* 199: 286-301 (2019).

Proceedings of the Pakistan Academy of Sciences: A Physical and Computational Sciences 60(4): 45-54 (2023)

ISSN (Print): 2518-4245; ISSN (Online): 2518-4253
https://doi.org/10.53560/PPASA(60-4)811

Pakistan Academy of Sciences

Research Article

Enhancing Efficiency of Solar Heater Box with Linear Actuator for Maximizing Solarization

Agha Mushtaque Ahmed[1*], Imran Khatri[1], Irfan Ahmed Halepoto[2], and Masood Nazir Khoso[3]

[1]Department of Entomology, Faculty of Crop Protection, Sindh Agriculture University Tando Jam, Sindh, Pakistan
[2]Department of Electronic Engineering, Mehran University of Engineering and Technology, Jamshoro, Pakistan
[3]Department of Mechatronic Engineering, Mehran University of Engineering and Technology, Jamshoro, Pakistan

Abstract: Worldwide, stored grain pests are massively infesting most stored crops and their by-products. The main losses are due to infestation of these pests occurring on various carriers just prior to harvest, during storage or shipping. Methyl bromide and phosphine fumigation have been widely used for phytosanitary treatment of stored grains but are recognized as highly effective in depleting ozone, and similarly, residue-free grains are important for thermal disinfection. Solarization is one of the best ways to manage and disinfect crops, the traditional solarization methods are already practised by farmers but are inefficient to kill all stages of pests and require additional land exposed to the sun. In this work, thermal disinfection systems using solar heaters are proposed, designed, and developed to offset the actual lethal heat window. For the experiment work, the solar heater boxes were constructed in an octagon shape with 135^{o} at the base for trapping maximum heat inside the solar the heater box. The characteristics of the proposed system proved simpler, faster, and inexpensive but equally effective to achieve the desired results in terms of heat generation and seed moisture.

Keywords: Solar Heater Boxes, Chickpea Grains, Temperature, Moisture, Seed Volume.

1. INTRODUCTION

Due to increasing population, food consumption is always on the rise globally, where the cost of food has reached new high scales, resulting in food insecurity, particularly in underdeveloped countries [1]. In developing countries, grain loss is significant because of inadequate post-harvest management knowledge and technology, improper storage facilities, and other factors. Around 20-30% losses of grains are recorded from most of the African countries [2]; and further 10 to 30% to other parts of the world [3]. Before storage of seeds, drying is a widely explored method for longer shelf life of agricultural products [4]. Mainly, it involves removing moisture from food materials to a safe level, slowing down the action of enzymes, bacteria, yeasts, and molds, and preventing insect damage. Grain needs to be dried 13% for long-term storage and 15% for short-term storage to be stored safely to keep the quality and quantity of grains [5]. Quantity losses happen when the grain is consumed by various arthropod insect pests, whereas quality losses show up as a reduced economic value of the crops or a decreased market worth of the product [6]. This results in decreased nutritional value, decreased seed germination, increased moisture, free fatty acid levels, decreased pH, loss of protein contents in food grain and existence of various protozoa that may cause dysentery and typhoid [7]. Approximately, 20,000 species are identified as pests of stored products [8] and estimated to destroy

Received: March 2023; Accepted: December 2023
*Corresponding Author: Agha Mushtaque Ahmed <maagha@sau.edu.pk>

about one-third of the global production of food that is worthy of more than 100 billion dollars [9]. However, the performance of insecticides over the past few decades has made them widely accepted for use against a variety of pests that affect agriculture and public health. Using pesticides frequently has several negative effects, including the recurrence of pests, toxicity to mammals and non-target animals, notably beneficial for insects, environmental degradation, and the emergence of resistance [6]. Methyl bromide fumigation has been widely employed to treat stored grains for phytosanitary reasons, however the US Environmental Protection Agency (USEPA) has acknowledged that it has significant consequences on ozone depletion [10]. Therefore, alternatives to methyl bromide are required for all postharvest applications [11]. Insect resistance to phosphine and market demands for residue-free grain has increased the importance of heat disinfestations [12]. By adopting modern techniques in storage systems, grain losses can generally be low in the intermediate stage of the supply chain. A sustainable way to enhance grain supply, reduce the strain on natural resources, and help farmers on a wide scale is by reducing post-harvest losses of grains, especially in developing nations. It will surely resolve hunger and improve farmers' livelihoods.

Traditionally, agricultural commodities like grains, fruits, and vegetables are left to dry out in the sun on surfaces like compacted dirt, mats, concrete floors, and roadways. Solarization is one of the most reliable practices for seed disinfection for various noxious store grain insect pests. Such management of store grain pests is more successful, particularly in tropical and sub-tropical countries where there is a lot of sunshine throughout the year. But, to accommodate the world's expanding population, agricultural fields are being turned into residential and commercial sectors, which reduces the amount of land available for crop cultivation. Furthermore, applying this technique results in successful drying but exposes the agricultural products to contamination from dirt, dust, insect infestation, and loss from birds and other animals [4]. In addition, there is a risk that cereal grains and other products may be contaminated with aflatoxin. In this regard, numerous studies have been undertaken on the natural convection of solar drying for agricultural products [13-15]. For successful pest management particularly for store grain pests, use of solar energy with specified materials, tools and techniques achieved a certain lethal dose of heat without affecting the seed quality [16] but still require further improvement for maximum seed volume and efficiency. However, the prime aim of this proposing work is to trap and retain the solar thermal in a new constructed solar heater box installed with motorized machine to acquire lethal heat level of 50-62^{o}C for less than 1 hour [17] to disinfect the store grains from insect pests without damaging the grains quality.

2. MATERIALS AND METHODS

2.1 Design and Construction of Octagonal Solar Heater Boxes

Previously, solar heater boxes were constructed with an inverted pyramid shape and an obtuse base angle of 118^{o} which had proved to be more effective in capturing more solar energy [18 -20]. In the present study, a slight adjustment was made to the overall design of the solar heater box composed of an acrylic sheet of 2 mm thickness. The design of the solar box was an octagon in shape with 135^{o} at the base for trapping much better heat inside. According to the longitude and latitude of an experiment for capturing the most solar radiation, the angle N-S for the upper side of the boxes was 45^{o}. The total height of solar heater box was 92 cm, and the top face and base measure 40×40 cm^2 in size. For maximum heat absorption, the walls of the boxes were painted with black paint on both the inside and the outside. To save heat losses from solar boxes by conduction and convection, the Styrofoam sheet of 10mm was attached outside the boxes because it is an insulator that contains bubbles filled with air. Later, aluminum foil was adhered to the outside of solar boxes with glue to maximize solar radiation reflection, as previously reported by Masood et al., [21] for the seed thickness of 2 kg. The present experiment consists of different volumes of seeds for testing the similar or better efficiency of the designed solar heater box.

For heat penetration and retention inside the solar box, the face and half of the arms of the boxes were covered in glass. The inner and outer shapes of solar heater boxes are shown in Figure 1 (b1 and b2). In the first experiment, a total of 5 kg grains were placed on the tray (50×50 cm^2) containing holes that allow temperature penetration from all directions. To allow heat to enter from all sides, the tray is fixed in the center of each box at a perpendicular height.

1.2 Recording of Temperature and Seed Moisture in Solar Heater Box

Temperature and seed moisture in solar heater box were recorded by mini portable digital data logger (16000 points recoding capacity, 40°C~+85°C Model: Elitech RC-4HC; China) assembled with external thermal sensor (thermocouples) as shown in Figure 2. The data after recording was exported in MS-excel via data management software 5.16000 data recording capacity (record interval=10s~24hrs adjustable) in hp laptop for further review and analyzation. A sharp pointed sensor needle was inserted inside the seeds to know the penetration of temperature and moisture in solar heater box as shown in Figure 3. The thermal sensor was placed between the seeds at random depths to record the temperature and seed moisture with data loggers [21]. The different seed volumes such as 5, 10, 15, 22, 30 and 40 kg were placed inside the solar heater box. The temperature readings were set auto to take at every 15 minutes of each time interval. The ambient temperature was also recorded with digital thermometer (°C). The exposure time for experiment was 6 hours starting from 9.00 am to 3.00 pm. To increase the quantity of grains inside solar heater box, the efficiency of solar heater box was better enhanced through installation of a 12v linear actuator assembled with a wooden log (saw tooth like comb) fitted in Solar Heater Box as shown in Figure 4. Through a 12v linear actuator, the temperature in the solar heater box was kept constant at various grains volume by rotating the seeds inside box. The H-bridge motor driver regulates the linear actuator's direction and speed. Using an Arduino UNO micro-controller and C++ programming, a linear actuator's stroke automatically ceases after extending and retracting continuously for 10 minutes [21].

2.3 Data Analysis

The temperatures recorded in the solar thermal box for each 60 minutes exposure time, for seven days, were evaluated with the (RCBD) method using analysis of variance (ANOVA) in statistical analysis software (Statistic 8.1). The data loggers and thermocouples were configured to collect temperature data, and the relationship between temperature measurements taken at 15 minutes intervals within the solar thermal box and exposure time was examined using regression and coefficient correlation analysis. Temperature trend data were transformed prior to regression analysis. ANOVA

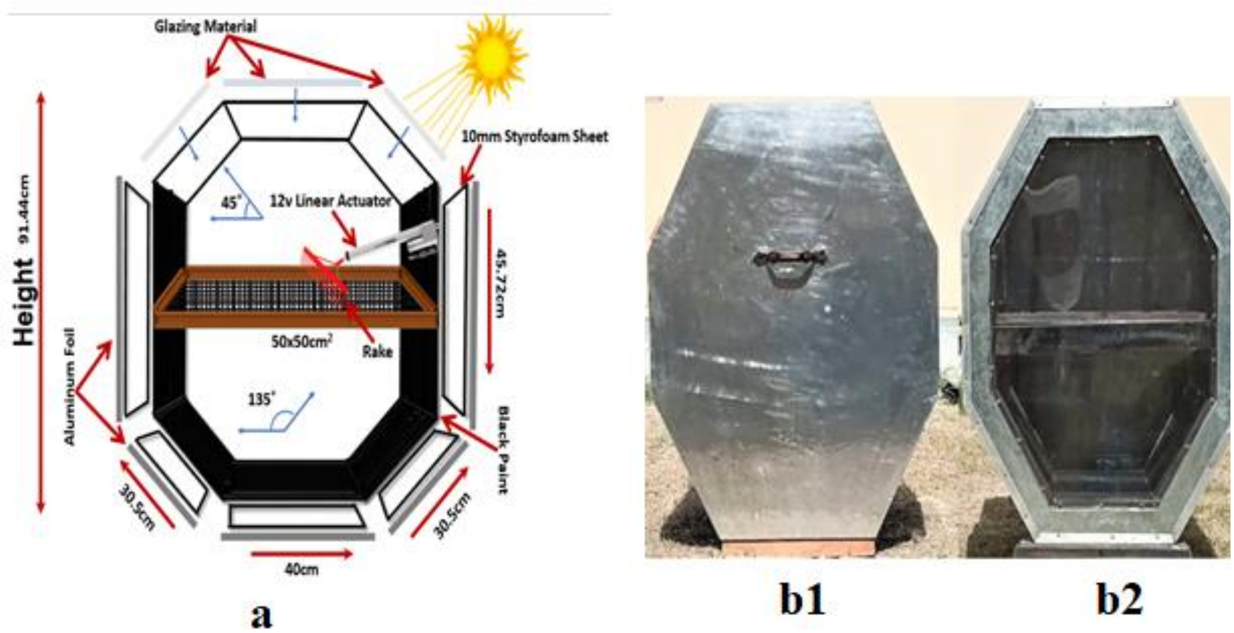

Fig.1. Design and construction of Octagonal solar heater box (a) animated picture of box showing all the measurement, design, and structure (b1 and b2) original pictures inside and outside of box used for experiment [21]

Fig. 2. Data logger and digital thermometer for recording temperature and Humidity [21]

Fig. 3. Recording temperature and seed moisture at different depths using pointed sensor needle [21]

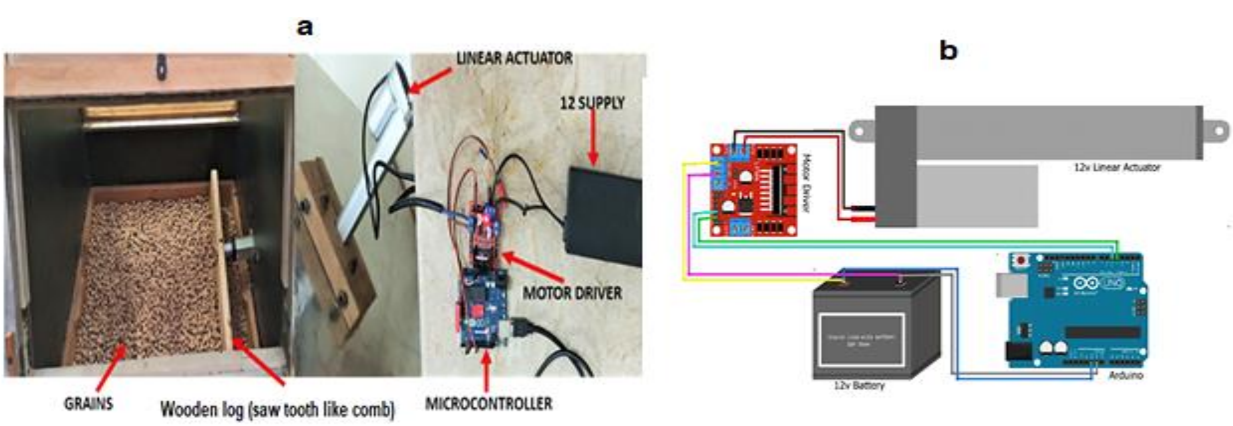

Fig. 4. Prototype (a) and circuit diagram (b) showing linear actuator assembled with wooden log for betterer retention of temperature and reduce moisture between grains inside solar heater box [21]

with RCBD was used to analyze data on the effect of seeding rate on optimal solar heating boxes, meanwhile, student *t* test for comparing seed moisture inside solar heater box with and without linear actuator was used. The least significant difference LSD (P=0.05) was used to demonstrate the heat trapping capacity of the solar heater boxes individually.

3. RESULTS AND DISCUSSION

3.1 Data Analysis of Different Solar Heater Boxes using 5kg Seed

The results regarding the temperature trapped inside the solar heater box was significantly different (P<0.05) with and without application of linear actuator as shown in Table 1. The maximum mean temperature of 43.35±0.89 °C was recorded in the first hour of observation in linear actuator as compared to the solar heater box without linear actuator. Meanwhile, the ambient temperature was quite lesser 29.48±0.32 °C. With time, the intensity of ambient temperature moderately raised up to 36.33± 0.68 °C in the last hour of observation (360 minutes), but the intensity of seed temperature recorded in solar heater box with linear actuator was the highest (78.11±0.84 °C) in comparison to seed temperature (63.21±6.02 °C) in solar heater box without linear actuator. Thus, the solar heater box with and without linear actuator successfully trapped the enough seed temperature inside the box as compared to ambient temperature. The present study was conducted after designing an octagonal solar heater box to obtain optimal results for capturing the maximum amount of solar radiation that can be detrimental to pest control of stored grains as (above 60 °C) with retention of 1 hour which is quite lethal for all stages of store grain pests [22]. Thus, the solar heater box with and without linear actuator successfully trapped the enough seed temperature inside the box as compared to ambient temperature. It has been previously well studied that for an effective structural heat treatment, the temperature must be maintained above 50°C for at least 24 hours [23]. Controlling arthropod pests by using high temperatures (above 50 °C) to disinfect products and/or structures is an environmentally friendly approach to protect stored products [19, 22, 24]. Thus, these arguments are clear evidence for the success of this study as with or without linear actuator, the designed solar heater box trapped enough heat energy that would be environmentally friendly for keeping grains after their harvest.

The seed moisture with temperature was also found varied ($P<0.05$) in solar heater box with and without linear actuator at different intervals as given in Table 2 in comparison to control seed moisture. In very first hour of experimental period (60 minutes), seed moisture (%) was maximum (62.32±0.6 %) in control condition whereas seed moisture recorded in the solar heater box without linear actuator was 45.89±1.56 % and with 39.74±1.97% with significant difference. With time, the seed moisture

Table 1: Record of temperature with and without linear actuator in solar heater box

Time (minutes)	Temperature (°C) of solar heater box		Ambient temperature (°C)
	with Linear Actuator	without Linear Actuator	
60	43.35±0.89[a]	39.95±1.66[b]	29.48±0.32[c]
120	49.13±1.11[a]	43.23±2.52[b]	31.65± 0.44[c]
180	52.00±1.30[a]	45.52±3.49[b]	33.32± 0.48[c]
240	61.41±1.53[a]	51.69±6.55[b]	34.48± 0.49[c]
300	69.51±1.67[a]	56.99±7.10[b]	35.40± 0.55[c]
360	78.11±0.84[a]	63.21±6.02[b]	36.33± 0.68[c]

The mean temperatures (Mean ± SE) were separated with least significant difference (LSD) at $p<0.05$ and represented by different letters within the same line.

Table 2. Record of moisture with and without linear actuator in solar heater box

Time (minutes)	Moisture (%) of grains in solar heater box		Seed moisture (%) in control
	With Linear Actuator	without Linear Actuator	
60	39.74±1.97[a]	45.89±1.56[b]	50.42±1.11[c]
120	30.34±1.90[a]	38.75±2.10[b]	46.53±1.95[c]
180	22.55±1.10[a]	36.18±2.00[b]	42.00±1.81[c]
240	17.20±1.95[a]	29.64±2.09[b]	38.14±1.62[c]
300	14.15±1.00[a]	24.73±3.08[b]	35.67±1.2[2c]
360	12.01±1.8[2]a	19.14±3.1[8]b	29.69±2.21[c]

The mean moistures (Mean ± SE) were separated with least significant difference (LSD) at $p <0.05$ and represented by different letters within the same line.

reduced in all treatments, but the reduction of seed moisture was the lowest 12.01±1.82 % in solar heater box with linear actuator and lower 19.14± 3.18 % without linear actuator at the end of experimental hours (360 minutes). Thus, these results show that the application of linear actuator successfully reduced the maximum percentage of seed moisture as compared to other treatments. Furthermore, the regression between temperature of different solar heater boxes and ambient temperature in respect of time and model was also recorded and is shown in Figure 5.

According to the regression analysis, there were no significant differences in temperature or seed moisture over time. There was a linear relationship between the temperature of the solar heater box and the surrounding air, with the significant coefficient of determination (r^2) lying between 0.93 and 0.99. Additionally, it resulted that the seed moisture in a solar heater box without a linear actuator had a strong linear relationship (r^2=0.99) with non-significant effect (P>0.05) regarding time.

With temperature, seed moisture was also found to be influenced with the designed study at different intervals in comparison to seed moisture kept under open sunshine. At beginning, the seed moisture (%) was maximum (almost 60 %) but after 1 hour of solarization it dropped and recorded around 50% (in control condition), whereas the seed moisture recorded in the solar heater box without linear actuator was 45.89±1.56 % and with linear actuator was 39.74±1.97% that shows the exact difference of open solarization and solarization with and without linear actuator. At the end of day, the seed moisture reduced in all treatments, but the reduction of seed moisture was the lowest 12.01±1.82 % in solar heater box with linear actuator and lower 19.14± 3.18 % without linear actuator. The application of linear actuator successfully reduced the maximum percentage of seed moisture as compared to other treatments.

3.2 Data Analysis of Different Solar Heater Boxes using Different Seed Volume

The overall mean temperature and the seed

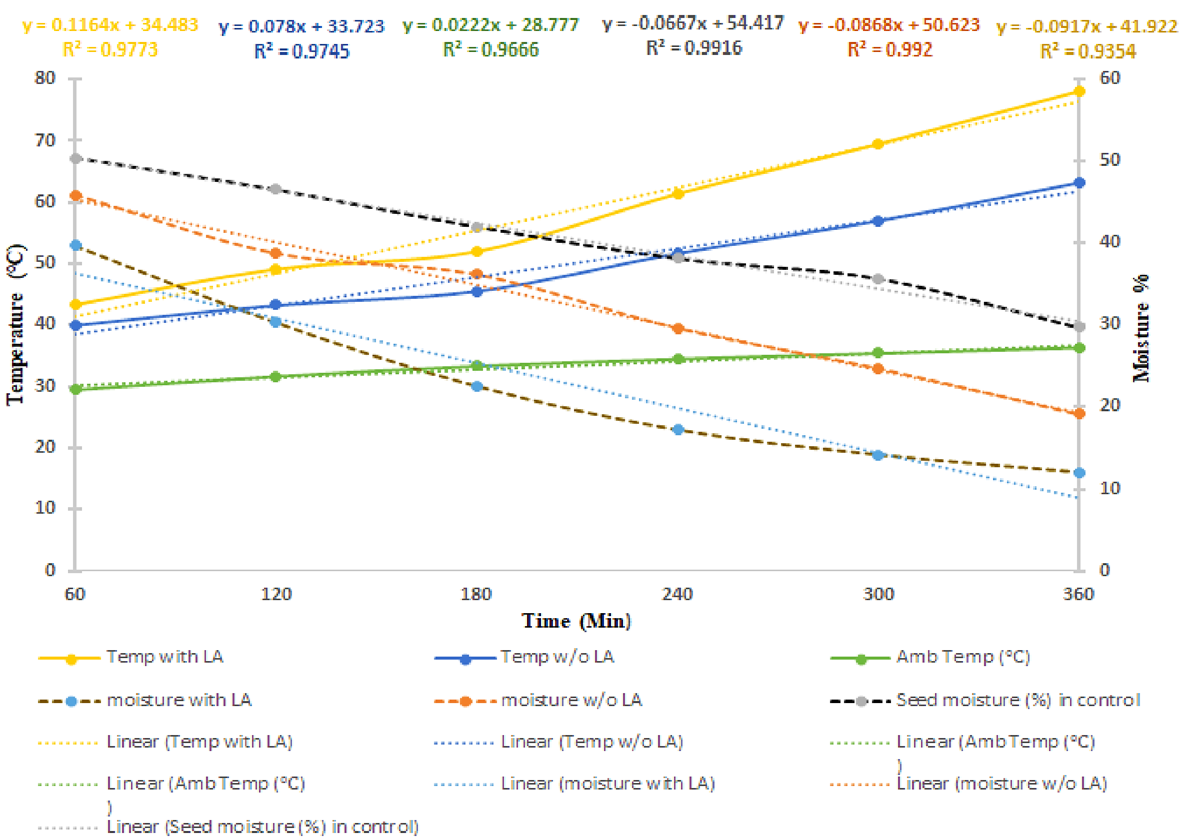

Fig. 5. Regression line showing relationship between temperature and moisture of ambient, with and without linear actuator in solar heater box.

moisture with different quantity of chickpea seeds at different intervals were recorded and are shown in Table 3. These results indicate variability in mean temperature recorded in solar heater box with or without linear actuator for different volume. It statistically ($P<0.05$) effected the penetration of heat between seeds in solar heater box with linear actuator but still it was enough in comparison to the solar heater box without linear actuator and ambient temperature. The temperature was successfully trapped above 60 ℃ in linear actuator solar heater box with different quantity of seeds (5-40kg). The maximum temperature with minimum volume of seeds was found 76.4±0.99 ℃ in initial hour of experiment with 5 kg seeds. The increased volume of seeds literally decreased the temperature inside both solar heater boxes. However, the increased volume of seeds inside linear actuator solar heater box at maximum limits of 40 kg still got better results (61.35±0.89 ℃) to trap the temperature as compared to the solar heater box without linear actuator (44.95 ±9.6 ℃) and ambient temperature (36.33±0.68 ℃) at the end of experimental hours (360 minutes).

Overall, these results show that the temperature in linear actuator solar heater box is much better as compared to temperature trapped without linear actuator solar heater box. Similarly, the seed moisture tested in a solar heater box with linear actuator after a period of 6 hours was ideal (11.09±1.26%) for 5kg seed bed (Table 4). Later, it started to increase with the decrease of temperature. The results regarding seed moisture were quite interesting as the volume of seeds was increased it did not allow seed moisture to dry efficiently but still the seed moisture in solar heater box applied with linear actuator produced better results and dried maximum seed moisture 24.74±1.17 % as compared to seed moisture (45.14±1.29 %) without linear actuator at the end of experiment. Meanwhile, the best results came with 5 kg with linear actuator and thus it showed that an application of further powerful motor for linear actuators will work better to get more desired results regarding seed moisture and heat trapping for solar heater box.

The regression estimated that there were significant differences in temperature and seed moisture with respect to seed weight (kg). The temperature in the solar heater box and ambient temperature was linearly related with an essential coefficient of determination (r^2) falling between 0.93 and 0.99. It is also demonstrated that there is not a significantly different ($P<0.05$) but moderate ($r^2=0.67$) linear relationship between seed moisture in solar heater box without linear actuator with respect to the seed weight (kg). Only 67% of the seed moisture variation depended on the weight of seed, according to the moderate coefficient of determination (r^2) value for seed moisture in solar heater box without linear actuator as shown in Figure 6.

In the second experiment, the volume of seeds was increased as to observe the efficiency of linear actuator. The overall mean temperature and the seed moisture with different quantity of chickpea seeds at different intervals were recorded. The results indicated variability in mean temperature recorded

Table 3. Record of temperature in terms of different seed volume with and without linear actuator in solar heater

Seed weight (kg)	Temperature (℃) of solar heater box		Ambient temperature (℃)
	with linear actuator	without linear actuator	
5	76.40±0.99^a	62.69±2.02^b	29.48±0.32^c
10	73.32±1.05^a	59.21±2.71^b	31.65± 0.44^c
15	69.21±1.21^a	56.78±4.65^b	33.32± 0.48^c
22	67.02±1.11^a	53.52±6.49^b	34.48± 0.49^c
30	64.13±1.11^a	49.23±8.62^b	35.40± 0.55^b
40	61.35±0.89^a	44.95±9.60^b	36.33±0.68b

The mean temperatures (Mean ± SE) were separated with least significant difference (LSD) at p <0.05 and represented by different letters within the same line.

Table 4. Record of seed moisture in terms of different seed volume with and without linear actuator in solar heater

Seed weight (kg)	Moisture (%) of grains in solar heater box	
	with linear actuator	**without linear actuator**
5	11.09±1.26[a]	20.89±1.21[b]
10	15.15±1.01[b]	35.75±1.97[b]
15	17.20±1.15 [b]	39.18±1.83 [b]
22	19.55±1.20 [a]	43.64±1.19 [b]
30	21.34±1.30 [a]	44.73±1.08 [b]
40	24.74±1.17 [a]	45.14±1.29 [b]

The mean moistures (Mean ± SE) were separated with least significant difference (LSD) at $p < 0.05$ and represented by different letters within the same line.

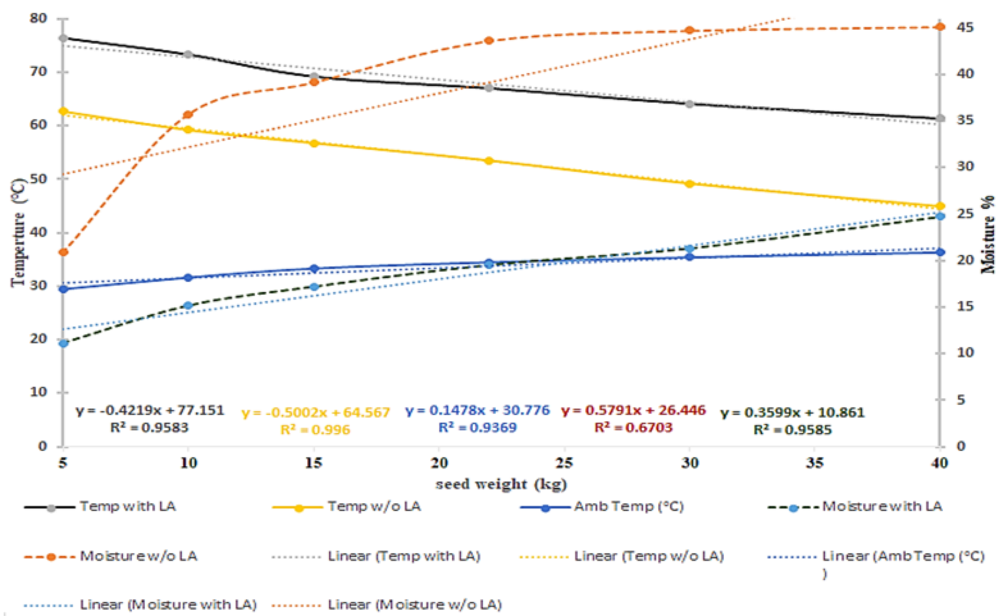

Fig. 6. Regression line showing relationship between temperature and moisture of ambient, with and without linear actuator in solar heater box

in solar heater box with and without linear actuator for different volume. The enough temperature in comparison to the solar heater box without linear actuator and ambient temperature in linear actuator solar heater trapped above 60 °C with different quantity of seeds (5-40kg). The increased volume of seeds literally decreased the temperature inside both solar heater boxes. However, the increased volume of seeds inside linear actuator solar heater box at maximum limits of 40 kg still got better results. Likewise seed moisture was ideal (13.09±1.26%) at 5kg seed at the end of experiment. But the moisture was higher with maximum volume of seed but still the seed moisture in solar heater box applied with linear actuator produced better results and dried maximum seed moisture 24.74±1.17 % as compared to seed moisture (45.14±1.29 %) without linear actuator at the end of experiment. Meanwhile, the best results came with 5 kg with linear actuator and thus it shows that an application of further powerful motor for linear actuators will work better to get more desired results regarding seed moisture and heat trapping for solar heater box. Temperature (>45°C) is the heat lethal zone for most storage pests [23], nevertheless may exist at certain point in the product disinfection system with temperature as high as 50°C [25], unless the heat is maintained for a longer time. The large amount of grain in the solar heating box required longer time to raise the temperature. However, temperature maintenance was maintained for a longer period. A small amount of grain heats up quickly and so as heat losses instantly [16]. The resulting temperature in a solar heater box with linear actuator was suitable for disinfesting the grains from store grain insect pests. Meanwhile, the most studies continue to define 45–50℃ as the lowest effective treatment temperature that must be achieved in a capability or between and in many heat-treated products [22, 23]. Others detailed a temperature range of 50-60°C for product disinfection on preserved arthropods is well enough [26, 27]. The results from previous studies supported the hypothesis that solar heater box is an ideal tool for solarization and can disinfect more products when used for pest control of stored product and further additional application of linear actuator enhanced its efficiency.

4. CONCLUSIONS

An acrylic solar heater box with a linear actuator generated incredibly the highest thermal performance compared to an ambient temperature. The highest mean temperature 61.35°C was found in linear actuator solar heater box which is comparatively the best in performance compared to the solar heater box without linear actuator and an ambient temperature when increasing seed volume up to 40kg in solar heater box. Based on these results, it can be concluded that the solar heater box with a linear actuator is quite safe, cheaper, and environmentally friendly control method that produces enough heat energy to disinfest the store grain pests. Furthermore, the application of linear actuator made the efficiency of solar heater box much better to increase volume of seed inside it. Therefore, such types of solar heater boxes with more powerful motors are recommended in future work.

5. ACKNOWLEDGEMENTS

Authors highly acknowledge the research facilities provided by Sindh Agriculture University Tando Jam, Pakistan under "SAU Small Research Grant Program (SRGP)" for research project title "A novel method of seed solarization through solar heater box for disinfecting the store grains pests", (Grant No. 848. SAU.ORIC)

6. CONFLICT OF INTEREST

The authors declare no conflict of interest.

7. REFERENCES

1. F. Hussain, M. Ahmad, A. Ghafoor, and A. Tanvir. A developmental study of a solar-assisted maize dryer for quality drying, *The Journal of Animal Plant Science* 30(3): 742–748 (2020).
2. N. Zhang, Wu, Wenfu. Y. Wang, and S. Li. Hazard Analysis of Traditional Post-Harvest Operation Methods and the Loss Reduction Effect Based on Five Time (5T) Management: The Case of Rice in Jilin Province, China. *Agriculture* 11(9): 1-21 (2021).
3. K. Neme, A. Nafady, S. Uddin, and Y.B. Tola. Application of nanotechnology in agriculture, postharvest loss reduction and food processing: Food security implication and challenges. *Heliyon* 7(12): 1-12 (2021).
4. O.K. Oni, A.M. Oyinloye, A.B. Adepeju, C.I. Okoro, and F. Idowu-Adebayo. Effect of Drying Methods and Sulphiting on the Physico-Chemical and Microbial Quality of Tomato Powder. *FUOYE Journal of Pure and Applied Sciences* 6(2): 109-117 (2021).
5. M.A. Lewis, S. Trabelsi, and S.O. Nelson. Development of an eighth-scale grain drying system with real-time microwave monitoring of moisture content. *Applied Engineering in Agriculture* 35(5): 767-774 (2019).
6. W.R. Morrison, A. Bruce, R.V. Wilkins, C.E. Albin, and F. H. Arthur. Sanitation improves stored product insect pest management. *Insects* 10(3): 1-20 (2019).
7. R.I. Sanchez-Mariñez, M.O. Cortez-Rocha, F. Ortega-Dorame, M. Morales-Valdes, and M. I. Silveira. End-use quality of flour from Rhyzoperthadominica infested wheat. *Cereal Chemistry* 74(4): 481-483 (1997).
8. L. Sidauruk, E. Panjaitan, P. Sipayung, and P.S. Hutauruk. Botanical pesticides, a potential ethnobotany Karo Regency to support food safety of the horticultural product. *In IOP Conference Series: Earth and Environmental Science* 1005(1): 1-9 (2022).
9. K.D. Singh, A.J. Mobolade, R. Bharali, D. Sahoo, and Y. Rajashekar. Main plant volatiles as stored grain pest management approach: *A review. Journal of Agriculture and Food Research* 4(1): 1-12 (2021).
10. W. Thomas. Methyl bromide: effective

pest management tool and environmental threat. *Journal of Nematology 28*(4S): 586 (1996).

11. S. Kumar, D. Mohapatra, N. Kotwaliwale, and K.K. Singh. Vacuum hermetic fumigation: *A review. Journal of Stored Products Research* 71(1): 47-56 (2017).
12. R. Qaisrani, and J. Banks. The prospects for heat disinfestation of grain. *In: Australian Postharvest Technical Conference, Adelaide, Australia*. 1-4 August, (2000).
13. D. Kumar, and P. Kalita. Reducing postharvest losses during storage of grain crops to strengthen food security in developing countries. *Foods* 6(1): 1-22 (2017).
14. A.G. Watson, S. Aleckovic, and R. Nallamothu. A novel and improved solar drying system appropriate for smallholder farmers. *Drying Technology* 40(11): 2274-2282 (2021).
15. R.B. Pinto, S. Oduro-Kwarteng, J.A. Hamidu, and H.M.K. Essandoh. Sensitivity of nutritional and microbial content of food wastes to drying technologies. *Scientific African* 16(1): 1-13 (2022).
16. R.M.E. Elhadaa, R.M. Awang, D. Omar, U.R. Sinniah and I. V. Grozescu. Selection and evaluation of materials for solar heater boxes and their capacity in trapping solar energy. *Journal of Agriculture and Veterinary Science* 10(1): 25-32 (2017).
17. P.G. Fields. The control of stored-product insects and mites with extreme temperatures. *Journal of Stored Products Research* 28(2): 89-118 (1992).
18. C. Mekasha, O. Dzolkifli, S. Yusuf, M. Rita and O. Noorma. Effect of heat treatment on developmental stages of *Callosobruchus maculatus* (Coleoptera: Bruchidae) in stored adzuki bean Vigna angularis. *International Journal of Tropical Insect Science* 26(4): 273-279 (2006).
19. G. Abdullahi, R. Muhamad, O. Dzolkhifli, and U.R. Sinniah, Analysis of quality retentions in cocoa beans exposed to solar heat treatment in cardboard solar heater box. *Cogent Food & Agriculture* 4(1): 1-12 (2018).
20. G. Abdullahi, R. Muhamad, O. Dzolkhifli, and U.R. Sinniah. Efficiency of cardboard solar heater boxes for disinfestations of stored grains against arthropod pest. *Agricultural Science and Technology* 11(3): 247–256 (2019).
21. M.N. Khoso, I.A. Halepoto, and A.M. Agha. Impact Analysis of Solar Heater Box with Linear Actuator on Characteristics of Chickpea Seeds. *Journal of Applied Research in Plant Sciences* 5(1): 1-11(2024).
22. P.G. Fields, and N.D. White. Alternatives to methyl bromide treatments for stored product and quarantine insects. *Annual Review of Entomology* 47: 331–359(2002).
23. S.J. Beckett, P.G. Fields, and B. Subramanyam. Disinfestation of stored products and associated structures using heat. *In: Heat treatments for postharvest pest control: theory and practice. CAB International, Oxon, United Kingdom*: pp. 182-237 (2007).
24. J.D. Hansen, J.A. Johnson, and D.A. Winter. History and use of heat in pest control: a review. *International Journal of Pest Management* 57(4): 267-289 (2011).
25. S.J. Beckett. Insect and mite control by manipulating temperature and moisture before and during chemical-free storage. *Journal of Stored Products Research* 47(4): 284-292 (2011).
26. R. Mahroof, B. Subramanyam, and D. Eustace. Temperature and relative humidity profiles during heat treatment of mills and its efficacy against *Tribolium castaneum* (Herbst) life stages. *Journal of Stored Products Research* 39(5): 555-569 (2003).
27. O. Dosland, B. Subramanyam, K. Sheppard, and R. Mahroof. Temperature modification for insect control. In: Insect management for food storage and processing, 2nd ed. J.W. Heaps (Ed.), *Elsevier Publishers, St. Paul. MN, USA,* pp. 89-103 (2006).

Proceedings of the Pakistan Academy of Sciences: A Physical and Computational Sciences 60(4): 55-67 (2023)

ISSN (Print): 2518-4245; ISSN (Online): 2518-4253
https://doi.org/10.53560/PPASA(60-4)826

Pakistan Academy of Sciences

Research Article

A Statistical Survey on the Socioeconomic and Demographic Livelihood of Brick Kiln Workers: A Case Study of Bahawalpur District, Punjab, Pakistan

Fakhra Anwar[1], Muhammad Mohsin[2*], and Sana Arshad[1]

[1]Department of Geography, The Islamia University of Bahawalpur, Pakistan
[2]Department of Geography, Govt. Sadiq Egerton Graduate College, Bahawalpur, Pakistan

Abstract: The working population at the brick kilns is considered as the people living under poverty line. This study aims to investigate the socioeconomic and demographic characteristics and housing and living conditions of the brick kiln workers in Bahawalpur district. Total 20 kilns were visited in 5 tehsils of the district as sample sites. A questionnaire was designed and a field survey was conducted by using simple random sampling techniques to collect the data of 400 workers from 20 brick kilns. The quantitative analysis was performed in Statistical Package for Social Sciences (SPSS) v. 21 by applying descriptive and inferential statistics. Findings showed that majority of the workers and heads of working households were male (85.75% and 97.5% in rural and urban kilns respectively) and Saraiki speaking (70%) belonged to Muhajir, Ranagr and Khokhar castes. The early age (14-18 years) marriages were common among the workers especially among the females. Dependency ratio of the kiln workers was 37.6% including mostly the children and aged. Dominant share of the kiln workers (70-80%) were illiterate and very few were literate from primary to intermediate level. The workers' income varied between PKR 11,000 to 15,000 which was lower than the average wage (PKR 25,000) as per the government rules. Brick moulders and soil suppliers were the main occupants at kilns. Most of the workers were resided in their own houses (1 or 2 rooms) made with *kacha* (mud made) material. Although various facilities were available to the workers at brick kilns but they have limited or no proper access to safe drinking water (20%), first aid and medicare (18%) and accident risk prevention (0%) crucial for their health. Thus, these facts demonstrated that kiln workers led a meagre life style. This work will serve as a reminder to authorities, planners and Non-Governmental Organizations (NGOs) to take action to improve the living conditions of kiln workers. Lastly, few suggestions were proposed to uplift the lives of the kiln workers.

Keywords: Socioeconomic factors, Dependency ratio, Literacy rate, Kiln workers, Bahawalpur.

1. INTRODUCTION

The brick kiln industry produces 21% of the world bricks. Globally, Asia alone produces 1,300 billion bricks, which accounts for 86.67% of the world's brick production [1]. It is estimated that over 16 million people are the working labour in this industry. China and India are the leading producers of bricks in the world with over 200 billion bricks annually [2]. It is one of the most challenging works for workers. The most problematic task in brick making workers for improvement of working method and working place is to save the environment from pollution [3]. At world level, building bricks are the part of the ceramic industry that produces most bricks. Bricks are made from vide range of clay and shale with excellent particle sizes. This clay contains many different minerals that influence in final colour of the fire bricks. South Asia is producing over 21% of the entire world's bricks [4]. Bricks construction in south Asian countries is the major source of building material in construction sector. It is one of the oldest unorganized labour intensive informal sectors in South and South East Asia [5]. Therefore, the brick kiln production has been dominated by few Asian countries such as China, India, Pakistan, Bangladesh, Nepal and Vietnam. China has occupied the first place with

Received: June 2023; Accepted: December 2023
* Corresponding Author: Muhammad Mohsin <mohsinshahzad10@yahoo.com>

54% of the total production, India is the second with the 11%, Pakistan is placed in third place with 8% and Bangladesh is placed at the fourth place in total production. The brick industry is an expletive industry on the world level and bricks are used as main construction material all over the world in every kind of building purpose. Yet the developing countries populations are still struggling to come up with sustainable solution to combat the problem of kilns impact and poverty [6]. Especially, in Pakistan, the environmental and health impacts of brick kilns have been researched by conducting various surveys from field professionals and investigators [7]. In brick industry, huge amount of wood and waste are used as a fuel which is highly inefficient and tends to air pollution and cause damage to vegetation and human health [8]. The kiln industry has been perceived as the main stationary source of environmental pollution as various pollutants like CO, SO, Nitrogen, Nitric Acid are also affecting the ozone layer as well [9].

Brick is one of the main components of construction industry, and mud brick kiln in particular, have made significant contribution to human development throughout history. However, the brick kiln industry products were not entirely absorbed by the real estate industry fast growth which caused shift in the conventional bricks kiln products network chain [10]. The process of brick kiln production is based on manual labour and brick kiln units are expected to give employment to around 10 million workers. But majority of them are hand to mouth and thus take advance loan from the kiln owners for their perusals. Hence, they are bound to work at kilns for an extended period of time longer than the time limit worldwide [11]. It is estimated that about 4.5 million people (including 1 million children), are working in slave-like conditions at around 20,000 brick kilns in Pakistan where almost 50% are women [12]. Mitchell et al., [13] argued that most brick kiln workers migrated with their families from the backward and poor areas of country. Gosal [14] stated that families, including young children, work in harsh condition and low wages situations. Brick production is season based work, as brick kiln units do not operate during the period of the rainy season, hence brick kiln workers returned to their localities and get employed in other works like agriculture for livelihood or remains unemployed. Kumari [15] highlighted the economic sustainability of the seasonal migration of the brick workers. The brick kiln system is mainly based on a manual labour. Kainth [16] studied the phenomena of kiln labour migration which is a permanent feature by pull and push factors that constitute a large population engaged in this industry in Punjab (India). Atangana [17] explained the typically push factors that initiate migration. Due to these factors, the workers in rural and urban areas did work to earn their livelihood. If they earn less, they migrate from one place to another in kiln areas.

Although, brick kiln sector all over the world is in a very unregulated and uncoordinated state, but it is responsible for 1.5% of gross domestic production (GDP) to meet the requirements of population of Pakistan [11]. Pakistan is 3rd largest brick producer (after China and India) with the production of over 45 billion bricks annually [7]. There are more than 16,000 brick kilns with over 0.5 million workers making 6 billion USD revenue annually [18]. Above 99% of the brick kilns in Pakistan are nearly a millennium old Fixed Chimney Bull's Trench Kiln (FCBTK) and many are of these are located in rural and peri-urban areas [19, 20]. These kilns contribute significantly in backing bricks to fulfil the basic need of shelter for humans. Basically, in Pakistan, the most widely used method of bricks is handmade, which are made with the help of moulders by workers and then dried in sun light and backed in a FCBTK. The use of heating material did not cause a wide spread pollution because these people use kuttal, wood, bora and coal for heating. The kiln trench chimneys are used to reduce the smoke and thermally insulated chamber is used to produce temperature like oven they provide heat to bricks for backing like a cake. But the sufficient temperature is required to complete this process.
Brick kiln workers are considered among the most susceptible and excluded in society. Therefore, there is a dire need for social concern to delineate measures to improve the working and living conditions of brick kiln workers overall. In November 2020, there were around 450 brick kilns in Bahawalpur [21]. Additionally, the brick industry is an unorganized sector where most of the workers are illiterate and working in meagre and vulnerable conditions. The health of brick kiln workers is highly affected due to working in unhealthy environment for an extended period and whole of family gives their skills yet they scarcely have enough to eat. Nevertheless,

this informal sector generates a large number of employments. Thus, this research investigates the socioeconomic characteristics and disparities of the working population in the brick kiln industry in five tehsils of Bahawalpur district. A clear strategy or action plan to solve their issues and eliminate the anticipated risk with alternative chances their slandered of living and welfare will be suggested to the concerned authorities and community for their awareness and action.

2. METHODOLOGY

Present study was an exploratory study in which socioeconomic and demographic conditions of working population at brick kilns of Bahawalpur district was investigated. The nature of this study is quantitative because it quantified the data variables.

2.1 Study Area

The study area Bahawalpur district is situated in southern Punjab between 27°48′ to 29°50′ North latitudes and 70°54′ to 72°50′ East longitudes. The district has covered a land area of 24,830 km^2 with a population of 4,284,964 souls in 2023 [22]. Administratively, there are six tehsils, i.e., Ahmadpur East, Hasilpur, Bahawalpur City, Bahawalpur Saddar, (both City and Saddar considered Bahawalpur tehsil as a whole in this research), Khairpur Tamewali and Yazman (Figure 1).

2.2 Data Collection

A self-administrative questionnaire was developed

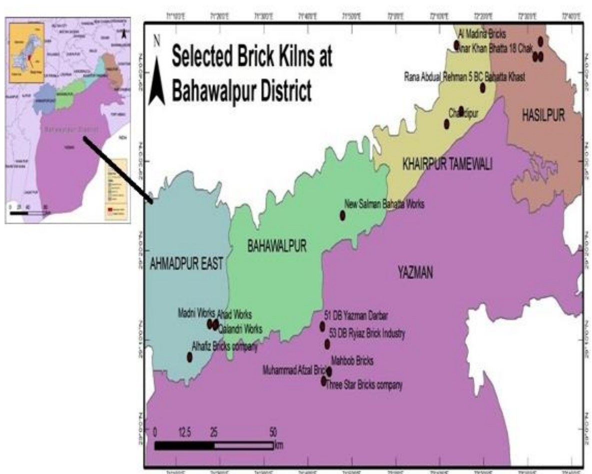

Fig. 1. Location of brick kilns in study area

to get data from the working labour to measure the variables of proposed study. Total 20 brick kilns were selected in overall district as study sites and visited. The questionnaire was divided into four parts viz. information about the head of the kiln, the socioeconomic conditions of the workers, the demographic characteristics of the interviewed labour and staff and the living facilities available for the labour and staff. The independent variables in this study were socioeconomic and demographic conditions, while the dependent variables were the workers and heads of the kiln.

2.3 Sampling Procedure

A random sampling technique was selected, where the individuals are chosen randomly, giving each member of the population an equal chance of being selected as the sample. While random stratified sampling method was used for the selection of the brick kilns in the district. The selected kilns were chosen from all over the district with the division of four kilns (two urban and two rural clusters) each from Ahmadpur East, Hasilpur, Khairpur Tamewali, Yazman and Bahawalpur (City and Saddar) tehsils. The sample size (n) calculated for this study was total of 400 (20×20 = 400). The population of this study was working people at brick kilns including labour, manager, supervisor and heads.

2.4 Data Analysis

The reliability index (Cronbach's alpha value) of questionnaire was 0.846, which shows a high score of reliability [23]. Data analysis was carried out using SPSS version 21. Both descriptive (frequency and percentages) and inferential statistics (chi-square) were used in this analysis. For the primary findings, followings formulae were used;

Proportion of Male (%) = Male population/total population

Proportion of female (%) = Female population/total population

Literacy ratio (%) = Total literate population/total population (above 10 years)

Marital status/Married population (%) = married population/total population

Study area map was developed in ArcGIS 10.5 software by using GPS to point out the kilns on map. Results were displayed in the form of charts, text, tables, and figures. These statistical results were then compared and discussed with supporting theoretical literature. Lastly, few preventive suggestions were proposed to facilitate the working labour and save the environment.

3. RESULTS AND DISCUSSION

3.1 Gender of Working Households

Table 1 reveals that in Bahawalpur tehsil, 54.4% were active male kiln workers and 45.6% were female. In Khairpur Tamewali tehsil 65.5% kiln workers were male and 34.4% were female working in kilns. In Yazman tehsil, 78.7% kiln workers were male and 21.3% were female workers active in work place. In rural areas, females were working more than in urban areas. In Ahmadpur East tehsil, among the active working kiln workers, 61% were male and 38.8% were female. The female population ratio was less because the workers were not living in the surrounding of kilns. Mostly population was illiterate and they were hesitant to give proper information. Majority of them (70%) were Saraiki speaking and belonged to Muhajir, Rangar and Khokhar castes, which is more than half of the kiln working population. Atangana [17] explained the long term relocation of an individual, household or group of new location outside the community of origin for finding employment opportunities.

3.2 Gender of Households Heads

In the current study, gender differential on the basis of household head has been studied. In a typical Pakistani society, primarily male people are head of the household because of social customs and traditions developed to accept male in this role. Since the human race discovered fire a million years ago, they have continually sought to control and utilized it to their advantage to create diverse construction materials that would satisfy their fundamental need for shelter [24].

Results of the current study indicate that 85.75% and 97.5% of males were the heads of the households in rural and urban areas, respectively. Only 14.25% and 2.5% of females were playing the role of household heads in rural and urban areas, respectively (Table 2).

Hence, notable inequalities of household headships were observed in rural and urban localities in Bahawalpur district.

3.3 Marital Status of Brick Kiln Workers

Marital status has practically nothing to do with biological characteristics, but even then, the study is quite helpful in studying imputation structure. Marriage is an essential factor in fertility because marriage increases population structure. In the study area, analysis of marital status suggests that women have lower status than men because the age of marriage of females can be argued to be an indicator of relative status. Generally speaking, women who marry early, soon after puberty, are likely to belong to more conservative families, have less chance to acquire education, to under employed and are more likely to depend on spouses than the women who marry at more mature age. In kiln sector, women marry generally at the early twenties, and they are not educated. In rural areas mostly females marry at an early age of 15-19. These results depict that 75% females and only 7% males were married; this shows the apparent gender differences. These differences are not only shown in rural area, but are found in all societies.

Table 1. Gender of households working in brick kilns (n=400)

Tehsil	**Bahawalpur**		**Khairpur Tam.**		**Yazman**		**Hasilpur**		**Ahmadpur East**	
Gender	M	F	M	F	M	F	M	F	M	F
Percentage	54.40	45.60	65.80	34.20	78.70	21.30	52.50	37.50	61	39

Table 2. Gender of households heads working at brick kilns (n=400)

Area	**Total Housing Units**	**Housing Units (Male)**	**Percentage**	**Housing Units (Female)**	**Percentage**
Rural	200	143	85.75	57	14.25
Urban	200	195	97.5	5	2.5

Table 3 shows the age at marriage of brick kilns population in district Bahawalpur. The 80 workers of age group of 14-18 were married in Bahawalpur, 40 working population in Khairpur Tamewali, 30 in Yazman tehsil, 40 in Hasilpur and 32 in Ahmadpur East. In the age group of 18-22, the married working population of the kilns was 22 in Bahawalpur, 27 in Khairpur Tamewali, 19 in Yazman, 20 in Hasilpur and 14 in Ahmadpur East tehsils. In the age group of 22-26, the married working population of the kilns was 7 in Bahawalpur, 15 in Khairpur Tamewali, 7 in Yazman, 15 in Hasilpur and 9 in Ahmadpur East tehsils. In the age group 26-30, the married working population of the kilns was 1 person only in Bahawalpur, 5 in Khairpur Tamewali, 5 in Yazman, 5 in Hasilpur and 7 in Ahmadpur East tehsils. Susan Golombok [25] examined the process involved women survival and behaves differently in different ways they bear in kilns for their survival. Likewise, Sadiq [26] studied the kiln population differences by social class, age at marriage in Pakistan and found the noteworthy differences in ages particularly the marriage at early age. Khan and More [27] studied the socioeconomic status of female workers in some selected kilns of India that is ranked second in brick manufacturing in the world. It is found that half of the workers at brick kilns were women working to earn money and as well as do the job of bearing and rearing of children. Due to poverty most of them were uneducated.

3.4 Dependency Ratio of Brick Kilns Workers

An active population of any country or city may comprise the working population, disabled persons, retirees and students. In Pakistan the dependency ratio is very high; however, in the kiln population dependency ratio is low because every child or adult in kiln family members works at the work place, i.e., men, women and children. Dessy and Pallage [28] concluded that child labour is common at kilns, which is the employment of children energy and potential at work. Generally, workers in brick industry are from deprived sectors of life and whole family works as a unit. Many recessive, illiterate and poor populations of backward communities are involved in this sector. Even the head of the kiln or labour in Asian countries is usually illiterate.

Table 4 shows the dependency ratio of the kiln population at kilns. According to the results 22% of kiln population in tehsil Bahawalpur was dependent and 78% was independent. In Ahmadpur East tehsil 29% of the population was dependent and 71% population was independent. In Hasilpur tehsil 29% population was dependent and 71% was independent. In Khairpur Tamewali tehsil, the majority of the 67% population was dependent (children and aged) and 33% population was independent, and in Yazman tehsil, 36% population was dependent and 64% was independent. Generally at kilns, the working population is thought to be ever working population and they do their work from childhood to old age. Ali [29] conducted a study in district Swabi (Pakistan) and asserted that existing resources exert pressure on the people and they need to improve their income, so they are forced to ask their children to work.

Table 3. Age at marriage of working population in brick kilns (n=400)

Age Group	Bahawalpur	Khairpur Tam.	Yazman	Hasilpur	Ahmadpur East
14-18	80	40	30	40	32
18-22	22	27	19	20	14
22-26	7	15	7	15	9
26-30	1	5	5	5	7

Table 4. Dependency ratio of working population of brick kilns (n-400)

Tehsil	Working Population	Dependent (%)	Independent (%)
Bahawalpur	110	20 (22)	90 (78)
Ahmadpur	90	20 (29)	70 (71)
Hasilpur	80	18 (29)	62 (71)
Khairpur Tamewali	67	27 (67)	40 (33)
Yazman	53	14 (36)	39 (64)
Total	400	109 (37.4)	291 (62.6)

3.5 Literacy Ratio of Brick Kilns Workers

A person is treated as literate if he could read a newspaper or any journal of the same standard and could write a simple letter. But in kiln area majority of the working population is ranked as illiterate. Figure 2 shows that the education division or levels of grades divided into in five groups' primary, middle, matric, intermediate and above. The percentage shown in graph displayed that dominant share of population was illiterate i.e. 62% in Bahawalpur, 50% in Ahmadpur East, 95% in Khairpur Tamewali, 93% in Hasilpur and 91% in Yazman. The working population at kiln attained Primary education was 25% in Bahawalpur, 17% in Ahmadpur East, 4% in Khairpur Tamewali, 6% in Hasilpur and 5% in Yazman. The share of Middle passed population among working population at kilns was 13% each in Bahawalpur and Ahmadpur East, and 1% each in Khairpur Tamewali, Hasilpur and Yazman. The working population share having Matric was highest (10%) in Ahmadpur East and 1% each in Khairpur Tamewali and Yazman. While in Bahawalpur and Hasilpur the share of Matric passed working population was zero. Likewise, Intermediate level in education attainment was very low (10% in Ahmadpur East and 1% in Yazman) and no single worker in whole working population attained above Intermediate education. Though, there is no gender discrimination in the attainment of education. These results show that the workers living nearby villages and towns of Bahawalpur and Ahmadpur East were primary passed due to the availability of schooling facilities. These children were treated as literate because they were primary passed but in three tehsils those were not more than 15%. Khan and Shahzadi [11] also found that lack of education is a major cause of kiln workers backwardness.

It is estimated that in Pakistan, there were more than 12 million child labourers working in various sectors of life including brick kilns [30]. Gosal [14] in his work on spatial pattern and progress of literacy at kilns in India contends that children are a work force for their parents no matter whether they are educated or not. The increase in children means increase in the workforce for the parents. They only know to work on kilns and learnt that their requirement is not an education. They make and mould the bricks in the kilns.

3.6 Occupational Structure of Brick Kilns Workers

Both male and female workers in these kilns were either too old or too young to do this work equally. Women and children worked to make the clay lumps, brick mouldings, bricks makings and transport bricks to other places using donkeys. This is done by all the members of the family of kiln workers. They are classified as Supervisor, Manager, Mud Maker, Brick Moulder, Soil Supplier, Sand Spreader, Rubbish Man and Fire Man having different functions they perform (Table 5). It is noted that 60% supervisors lived in the work place and other 40% lived in nearby village or town. While Managers, Mud Makers, Brick Moulders, Soil Suppliers, Sand Spreaders, Rubbish Men and Fire Men all belonged to one family. They work

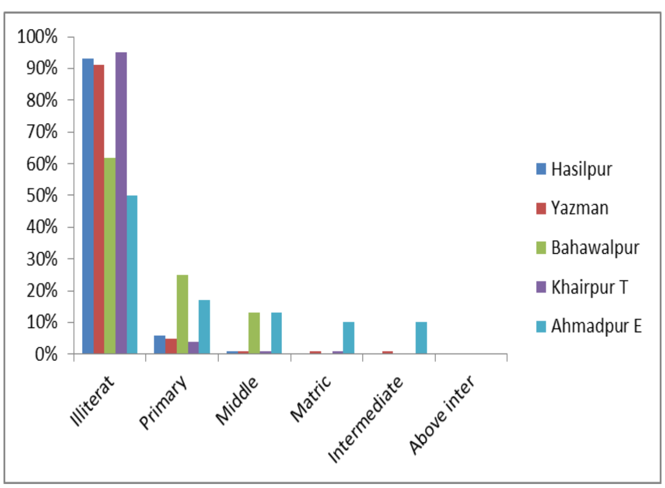

Fig. 2. Literacy rate of workers at brick kilns in Bahawalpur district (n=400)

Table 5. Classification of brick kiln workers occupation at kiln

Occupation	Definition
Supervisor	A person who directs and oversees the work of kiln workers.
Manager	A person responsible for controlling or administrating of staff members.
Mud Maker	A person who mixes mud with water.
Brick Moulder	A person who uses a structured frame (*sanchi*) to mould mud in the shape of bricks.
Soil Supplier	A person who takes the supply the mud in a hand cart or donkey cart.
Sand Spreader	A person who spreads sand.
Rubbish Man	A person who collects waste material.
Fire Man	A person who makes the fire in underground oven or furnace.

together with their family members. Mostly every family has separate work from the other family or household.

Table 6 showed that there were different kinds of workers working at the selected kilns. In tehsil Bahawalpur there were 400 soil suppliers, 200 brick moulders, 72 mud makers, 20 fire men, 11 rubbish men and 39 sand spreaders, 5 mangers and 4 supervisors. In Ahmadpur East tehsil, there were 400 soil suppliers, 150 mud makers, 100 brick moulders, 10 fire men, 10 rubbish men, 10 sand spreaders, 6 mangers and 4 supervisors. In tehsil Hasilpur, there were 400 brick moulders, 385 soil suppliers, 200 mud makers, 10 fire men, 8 rubbish men, 10 sand spreaders, 4 mangers and 4 supervisors. In Khairpur Tamewali tehsil, there were 300 soil suppliers, 110 brick moulders, 100 mud makers, 10 fire men, 2 rubbish men, 11 sand spreaders, 4 mangers and 4 supervisors. In tehsil Yazman there were 200 brick moulders, 100 each soil suppliers and mud makers, 30 fire men who have backed bricks in the underground baking oven, 10 rubbish men and 19 sand spreaders were working at the sampled kilns. Sain and Meena [31] analyzed the occupational structure of kilns in Rajasthan (India) and found that kiln workers work in adverse working conditions, having poor pastures with substantial loadings and traditionally designed hand tool results in occupational health problems. Owing to lack of proper training and absence of personal protection equipment have a potential of causing adverse impacts on the workers' health as 79.23% were found suffering with respiratory and skin diseases.

3.7 Income Level of Brick Kiln Workers

The income of kiln workers depends upon the job of their resources as they perform at the kilns so far. Income is an economic term that is generally defined in terms of consumption and earning of wealth. In the preparation of bricks, the whole family of men, women and children are involved. However, the wages are given to a single person only. The brick kiln generally is situated in a desolate place away from the main cities and towns in a few tehsils. The kiln working days are hardly 240-260 in a year. On rainy days, there is no work. As such, during the monsoon months in summer, their work is stopped and if rain comes suddenly then the loss of many unbaked bricks might be faced. In a few kilns on the rainy days, the loss of bricks is born by the kiln head. They depend upon the kiln owners entirely.

Table 6. Occupational structure of workers at brick kilns in Bahawalpur district

Occupation	Bahawalpur	Ahmadpur East	Hasilpur	Khairpur Tamewali	Yazman
Supervisor	4	4	4	4	6
Manager	5	6	4	4	6
Mud Maker	72	150	200	100	100
Brick Moulder	200	100	400	110	200
Soil Supplier	400	400	385	300	100
Sand Spreader	39	10	10	11	19
Rubbish Man	11	10	8	2	10
Fire Man	20	10	10	10	30

Table 7 shows the monthly average income of the households' kiln workers in Bahawalpur district. Kiln workers earn the highest income of PKR 21,000-25,000 per month of Khairpur Tamewali (10) followed by 8 in Hasilpur, 6 in Ahmadpur East, 5 in Bahawalpur and 2 in Yazman. The per month income of PKR 16,000-20,000 is earned by 39 kiln workers of Ahmadpur East, followed by 30 in Bahawalpur, 10 in Khairpur Tamewali, 8 in Yazman and 4 in Hasilpur. The per month income of PKR 11,000-15,000 is earned by the 50 kiln workers in Khairpur Tamewali followed by 48 in Hasilpur, 40 in Yazman, 30 in Bahawalpur and 20 kiln workers in Ahmadpur East. The kiln workers earned the per month income of PKR 6,000-10,000 were earned by 20 in Yazman, 17 in Hasilpur, 10 in Bahawalpur and 5 each in Khairpur Tamewali and Ahmadpur East. The least income per month of PKR 1,000-5,000 is earned by the 10 kiln workers each in Yazman and Ahmadpur East, 5 each in Bahawalpur and Khairpur Tamewali and 3 in Hasilpur. These results justify that income levels of the kiln workers households were not very much high as compared to their needs and inflation rates. Niaz et al., [20] conducted a study at the brick kilns of Faisalabad district (Pakistan) and found that majority of the workers (52.72%) have a monthly income of PKR 6,000-9,000 and most of the workers (87%) haven't any kind of agricultural or residential property. Thaheem et al. [30] also noted that the incomes of the brick kiln workers in Pakistan are less than the international levels. Thirupathi and Anthonisamy [32] conducted a study in Salem district (India) and found a significant difference between the workers satisfaction and level of wages.

3.8 Expenditure of Brick Kiln Workers

Expenditure depends on the income level of kiln population. Therefore, they are compromising the inequity by working hard but didn't earn enough to be economically strong. They need more income to meet their daily based necessities and improve the socioeconomic conditions on a local level. Table 8 presents the average monthly expenses of kiln workers family in all tehsils of district Bahawalpur. Kiln workers households of Khairpur Tamewali (20) made the highest expenses of PKR 16,000-20,000 per month followed by 15 in Hasilpur, 10 each in Yazman and Bahawalpur and 5 in Ahmadpur East. The expenses of PKR 11,000-15,000 were made by the 50 kiln workers households of Bahawalpur, followed by 45 each in Ahmadpur East and Hasilpur, 40 in Khairpur Tamewali and 20 in Yazman. Likewise, the highest expenses of PKR 6,000-10,000 were made by the 40 kiln workers households of Yazman followed by 22 in Ahmadpur East, 17 in Bahawalpur, 15 in Hasilpur and 12 in Khairpur Tamewali. The least expenses of PKR 1,000-5,000 were made by the 12 kiln workers households of Ahmadpur East, 10 in Yazman, 8 in Khairpur Tamewali, 5 in Hasilpur and 3 in Bahawalpur. These results suggest that the expenditures of the kiln workers households were almost even to their income levels and they hardly meet their daily based necessities.

3.9 Housing Condition and Living Facilities

The condition of the houses is a significant factor in analyzing the social and economic state of an area.

Table 7. Average monthly income of households working at brick kilns in Bahawalpur district

Income (PKR)	Bahawalpur	Ahmadpur East	Hasilpur	Khairpur Tamewali	Yazman
1,000-5,000	5	10	3	5	10
6,000-10,000	10	5	17	5	20
11,000-15,000	30	20	48	50	40
16,000-20,000	30	39	4	10	8
21,000-25,000	5	6	8	10	2

Table 8. Average monthly expenditure of households working at brick kilns in Bahawalpur district

Expenditure (PKR)	Bahawalpur	Ahmadpur East	Hasilpur	Khairpur Tamewali	Yazman
1,000-5,000	3	12	5	8	10
6,000-10,000	17	22	15	12	40
11,000-15,000	50	45	45	40	20
16,000-20,000	10	5	15	20	10

The shelter is the basic need of every human being. In the kiln, all family members work to make bricks for others, but they do not have their own house. In a civilized society, all persons should have their own house or shelter. The housing condition depends on constructional material. In the study area, three types of houses were present depending on the construction materials, *Kacha* (mud made), *Semi-kacha* (mud and brick made) and *Pakka* (brick and cemented).

Table 9 portrays the ownership and condition of houses at the brick kilns. In Bahawalpur, 800 houses were owned by the kiln workers while 300 were on rent, the housing material used was *Kacha* of 300 households, *Semi-kacha* of 405 households and *Pakka* of 395 households. The number of rooms of in a house was one or two throughout the brick kilns in the district. In Ahmadpur East, 335 houses were owned by the kiln workers while 280 were on rent, the housing material used was *Kacha* of 400 households, *Semi-kacha* of 115 households and *Pakka* of 100 households. In Hasilpur, 115 houses were owned by the kiln workers while 85 were on rent, the housing material used was *Kacha* of 115 households, *Semi-kacha* of 35 households and *Pakka* of 50 households. In Khairpur Tamewali, 400 houses were owned by the kiln workers while 147 were on rent, the housing material used was *Kacha* of 300 households, *Semi-kacha* of 200 households and *Pakka* of 47 households. In Yazman, 300 houses were owned by the kiln workers while 85 were on rent, the housing material used was *Kacha* of 35 households, *Semi-kacha* of 300 households and *Pakka* of 50 households. These results show that workers at the kilns resided in the houses constructed by different building materials made mostly by *Kacha*, *Semi-kacha* and *Pakka*. A study conducted in Jhang (Pakistan) also found that majority of the respondents (67%) resided in *Kacha* houses. They belonged to low income class and engaged in minor jobs, i.e., kiln workers [33]. A recent study undertaken in the Punjab, (Pakistan) also asserted that 62.5% workers at kilns were offered housing facility by the brick kiln owners in the surroundings of the kilns; of which, 23.8% were *Pakka* houses and 38.6% were *Kacha* houses [21].

Table 10 shows that the living facility check list at kilns workers of Bahawalpur does not have essential life needs and they are spending a poor and vulnerable life. The electricity facility is crucial in the modern day life as well as all kiln sites of Bahawalpur district therefore 89% workers were availing this facility and remaining 11% were not availing this facility. These results show that 78% workers had radio facility and remaining 22% workers do not have any radio facility. They were hand to mouth and cannot avail the luxuries of life. The 85% workers had the television facility and 15% did not have a TV. In the study area, the telephone facilities were shared. Cell phone facility was availed by 91% working population like simple Nokia mobiles and other 15% have no cell phone because they cannot afford it. The kiln workers

Table 10. Living facilities check list at kilns of Bahawalpur district

Living Facility Check List	Yes (%)	No (%)
Electricity Facility	89	11
Radio Facility	78	22
Telephone Facility	85	15
Cell Phone	91	9
Television Facility	76	34
Tape Recorder	90	10
Loan Availability from the Kiln Owner	80	20
Clean Drinking Water	20	80
Sanitation Facility	30	70
First Aid & Medicare Facility	18	82
Accident Risk Prevention	0	100

Table 9. Ownership status and housing condition of households working in brick kiln in Bahawalpur district

Tehsil	House Ownership		Housing Material			
	Owned	Rented	Kacha	Semi-kacha	Pakka	No. of Rooms
Bahawalpur	800	300	300	405	395	1-2
Ahmadpur East	335	280	400	115	100	1-2
Hasilpur	115	85	115	35	50	1-2
Khairpur Tamewali	400	147	300	200	47	1-2
Yazman	300	85	35	300	50	1-2

mostly enjoy by listening the songs while making the bricks as 76% had a tape recorder. The loan (*peshgi*) facility was also availed by the majority of the kiln workers (80%) in advance by agreeing a contract, and resultantly they are bound to work for specific time period agreed in contract usually giving over time of work. Normally, the loan facility for working household at kilns started with the amount of 500 USD or equal to nearly PKR 150,000 [34]. It is also found in previous studies that majority of the workers (70%) have taken the loan facility and were bound to work at kilns and there is extensive bonded labour at the kilns throughout Pakistan [20, 35]. Often this loan lasts for a long period of time and passes from one generation to another and hence this loan slavery is causing the breakdown of the poor workers in many developing countries (e.g., Pakistan) and ultimately increasing their clashes and casualties in such contexts [36, 37]. Alarmingly, the majority of the kiln population (80%) had a lack of access to safe drinking water and use hand pump water while 20% had access to nearby water canals and water channels for potable water. About 70% working population had no sanitation facility while remaining 30% had the sanitation facility. Similarly, in Bahawalpur the majority of the kiln population (82%) can't afford the first aid kit because of poorness and even chance to be suffered in accidents being present every time in the kiln area while baking bricks at a higher risk. A study conducted at the kilns of Faisalabad district (Pakistan) also found that majority of the kiln workers are facing lack of proper drinking water source and majority (94%) have no idea about water purity. Likewise, majority of them (76%) didn't have sanitation facility [20]. The brick kiln workers did not have access to the proper medical treatment because of unaffordability and the inaccessibility of health and Medicare facility. Moreover, the accident risk prevention facility was totally absent during the work of making and baking bricks at the kilns, as kiln workers have no awareness about this. . Hence, the lives of kiln workers were at serious risk in the case of an accident, especially the rubbish men and fire men.

Table 11. Chi-square results of selected variables

Variable	Value	Df	Sig.
Monthly Income and Expenditures	11.4	3	0.029*
Available Facilities and Workers Satisfaction	7.81	1	0.000**

Note: *significant at 95% level of confidence
**significant at 99% level of confidence

3.10 Chi-square Results

Table 11 shows the results of Chi-square to test the association between the selected variables viz. Monthly Income (Independent) and Expenditures (Dependent) of the kiln workers and Available Facilities (Independent) and Workers Satisfaction (Dependent). The value of the chi-square of monthly income and expenditures is 11.4 with alpha value of 0.029 that is significant at 95% level of confidence. It exhibits a strong association between the monthly income and expenditures of the kiln workers. Moreover, there is also a strong positive association between available facilities and workers' satisfaction. The value of chi-square of available facilities and workers' satisfaction is 7.81 with degree of freedom (Df) value of 1 and shows a strong association between the workers' satisfaction and available facilities. The alpha value of 0.000 is also highly significant at 99% level of confidence. Hence, these results also justify the brick kiln workers miserable and deprived livelihood. Khan and Shahzadi [11] also highlighted the kiln workers clashes and limited livelihood opportunities due to the low wages.

4. CONCLUSIONS

The workers at brick kiln represent a class of individuals who live below the poverty line. The present study aimed to investigate the socioeconomic and demographic inequalities in Bahawalpur district. Based on quantitative analysis of socioeconomic and demographic livelihood of the workers, the study found wide disparities in various aspects of kiln workers' life. The study concludes that the majority of the household heads of workers at kilns were male. Early young age (14-18 years) marriages were common in working population at the kilns. Therefore, the dependency ratio was also considerable among kiln workers due to large family size. Most of the workers were Saraiki speakers, and belonged to Muhajir, Rangar and Khokhar castes. The education level of the kiln workers showed that majority of the workers (above 90%) were illiterate in overall district except Ahmadpur East and Bahawalpur tehsils having comparatively lower number of illiterate than other tehsils. The socioeconomic indicators also showed the deteriorated conditions of the kiln workers. The income level of the workers per month varied between PKR 11,000 to 15,000, which is lower than the average wage of labourer as per the government rules. Similarly, the expenditures of the majority of the workers were also equal to their monthly income. Workers mostly involved in mud making, mud moulding and soil supplying works at the kilns. The housing conditions and living facilities showed that workers' houses were small and consisted of 1 or 2 rooms only. There were a lot of people living in rented houses and most of the houses were built with *kacha* material of unbaked mud bricks. Most of the workers were migrated and they had no proper facilities of daily life and most importantly lack of access to safe and clean drinking water and Medicare facility.

It is also found that due to the time constraints on the brick kiln production, the owners of kilns were having aggressive and harsh behavior with workers. Thus, Child labour, harassment, long working hours, violation of environmental law, labour law, human rights are common features of the kilns. The loan is also given to the majority of the workers by the owners which trapped the workers of kiln due to high interest rates and low wages during the working period. Therefore, the whole family is virtual prisoner of the kiln owner and need special permission to leave the premises. So the brick kiln workers in district Bahawalpur and its tehsils are living a meager life. This work will serve as a reminder for authorities, planners and Non-Governmental Organizations (NGOs) to take action to improve the living conditions of kiln workers. So, based on the findings the current study suggests;

1. The kiln workers must be aware of alternate works apart from brick making as hand making craft, stitching and embroidery into monsoon to fulfill their needs.

2. If they are not given the formal education than they must be given informal and islamic education so they can spend their lives as civilized persons.
3. Kiln workers must be given first aid facility as they get injured like using sharp instruments to carve out mud or from the hot oven hard to sustain the heat and work pressure.
4. The living conditions of kiln workers should be improved.
5. The clean drinking water with hand pumps should be made available.
6. Financial and medical advice should be provided to the kiln workers from NGOs, and healthcare professionals to sustain a better healthy life.
7. The owners should provide a safe and healthy environment for the workers because brick kilns are perceived as dirty, unsafe with a backing environment and poor infrastructure.
8. Development of the veterinary services should be initiated to support animals' projects for local community.
9. No interest based policies should be made by kiln heads/ owners for workers.
10. Provincial government must take tangible steps in the implementation of 1992 Bounded Labour System (Abolition) Act, regarding the elimination of bonded labour system and ensuring to stop this in practice.

5. CONFLICT OF INTEREST

The authors declare no conflict of interest.

6. REFERENCES

1. ILO. Environment, Human Labour, and Animal Welfare - Unveiling the Full Picture of South Asia's Brick Kilns and Building the Blocks for Change/ International Labour Office; The Brooke Hospital for Animals; The Donkey Sanctuary volume 1. (Geneva) (2017).
2. A.J. Nath, R. Lal, and A.K. Das. Fired bricks: CO_2 emission and food insecurity. *Global Challenges* 2(4): 1700115–1700115 (2018).
3. K. Pandey, and A. Vats. An Owas-based analysis of worker engaged in the brick making factories, Faizabad district of Uttar Pradesh, India. *Journal of Ergonomics* 2(2): 1000104 (2012).
4. A. Eil, J. Li, P. Baral, and E. Saikawa. Dirty Stacks, High Stakes: An Overview of Brick Sector in South Asia, (2020). https://documents1.worldbank.org/curated/en/685751588227715919/pdf/Dirty-Stacks-High-Stakes-An-Overview-of-Brick-Sector-in-South-Asia.pdf
5. S. Halder, and U.K. Patra. Status of brick kiln workers in south-east Asia. *Journal of Natural Remedies* 21(1): 6–16 (2021).

6. S. Jaffee, and J.S. Hyde. Gender differences in moral orientation: A meta-analysis. *Psychological Bulletin* 126(5): 703–26 (2000).
7. M.W. Khan, Y. Ali, F. De Felice, A. Salman, and A. Petrillo. Impact of brick kilns industry on environment and human health in Pakistan. *Science of Total Environment* 678: 383–389 (2019).
8. R. Khan, and H. Vyas. A study of impact of brick industries on environment and human health in Ujjain city (India). *Journal of Environmental Research and Development,* 2(3): 421–425 (2008).
9. A. Rauf, S. Shakir, A. Ncube, H.M. Abd-ur Rehman, A.K. Janjua, S. Khanum, and A.H. Khoja. Prospects towards sustainability: A comparative study to evaluate the environmental performance of brick kiln in Pakistan. *Environmental Impact Assessment Review* 94: 106746 (2022).
10. A. Ercelawn, and M. Nauman. Unfree labour in South Asia: Debt bondage at brick kilns in Pakistan. *Economic and Political Weekly* 39(22): 2235–2242 (2004).
11. K.K. Khan, and A. Shahzadi. Socio-economic Determinants & Dynamics of Debt Bondage: A descriptive Analysis of Brick Kiln Workers in Punjab, Pakistan. *Technium* 3(7): 64–86 (2021).
12. K. Chaudhry. The brick kiln slaves of Pakistan (2021). Available from https://www.ucanews.com/news/the-brick-kiln-slaves-of-pakistan/92287 (6-5-2023).
13. R.B. Mitchell, N. Choucri, and L. Gasser. Intentional Oil Pollution at Sea: Environmental Policy and Treaty Compliance. *MIT Press* (1994).
14. G.S. Gosal. Regional aspects of rural literacy in India. *Rural Sociology* 9(3): 1–15 (1967).
15. S. Kumari. Occupational Health of brick workers of India. *International Journal of Health Sciences & Research* 8(9): 183–187 (2018).
16. G.S. Kainth. Push and pull factors of migration: A case study of brick kiln industry of Punjab state. *Asia-Pacific Journal of Social Sciences* 1(1): 82–116 (2009).
17. C.A. Atangana. Textbook of World Social Geography. *New Delhi, India* (2013).
18. M. Sohail, and A.K. Mahmood. Extension of Labor Laws in the Informal Economy of Pakistan: Brick Kiln Industry as a Case Study. *Pakistan Journal of Social Sciences* 40(2): 591–602 (2020).
19. S. Nasim, and F. Sharif. To adopt, or not to adopt, ‘why’ is the question: A case for clean kiln technologies in developing countries. *Journal of Cleaner Production* 257: 120553 (2020).
20. U. Niaz, Z. Jamil, M.S. Chaudhry, and S. Nasim. Socio-Economic and cultural problems of brick kiln working families (A case study of district Faisalabad. *Journal of Applied Environmental and Biological Sciences* 8(1): 106–111 (2018).
21. F. Ahmad, G.M. Shah, B.B. Pradhan, and R. Dixit. Towards an environment-and-worker-friendly brick kiln sector in Punjab, Pakistan. Working Paper, International Centre for Integrated Mountain Development (ICIMOD), Kathmandu, Nepal (2022). Available from https://lib.icimod.org/record/36001 (10-11-2023).
22. GOP. Province Wise Results of Census-2023. Government of Pakistan, Pakistan Bureau of Statistics (PBS), Islamabad (2023). Available from www.pbs.gov.pk (6-11-2023).
23. L.J. Cronbach. Coefficient alpha and internal structure of the tests. *Psychometrika* 16: 297–334 (1951).
24. S. Kumari. Neo-bondage in the brick kiln industry: A case study of Bihar. *Social Change* 48(3): 183–187 (2018a).
25. R.F. Susan Golombok. Gender Development. *Cambridge University Press* (1994).
26. M.M. Sadiq. The economic effects of postponement of marriage for Pakistan. PhD Thesis, *The American University* (1964).
27. I. Khan, and R. More. Study of socio economic status of female workers in some selected bricks kiln in Karad, Maharashtra, India. *International Journal of Multidisciplinary Research and Development* 2(10): 299–304 (2015).
28. S.E. Dessy, and S. Pallage. Child labor and coordination failures. *Journal of Development Economics* 65(2): 469–476 (2001).
29. G. Ali. Economic Factors responsible for child labor (A case study of District Swabi). *Journal of Managerial Sciences* V(1): 85–96 (2011).
30. D. Thaheem, M. Sohail, and A. Mahmood. Extension of Labor Laws in the Informal Economy of Pakistan: Brick-Kiln Industry as a Case Study. *Pakistan Journal of Social Science*s 40(2): 591–602 (2020).
31. M.K. Sain, and M. Meena. Analysing the occupational health of workers in brick kilns of Rajasthan. *Interdisciplinary Environmental Review* 18(2): 101–107 (2017).
32. T. Thirupathi, and M. Anthonisamy. Health problems faced by brick kiln workers in Salem district. *International Journal of Applied Research* 1(15): 266–269 (2015).
33. Z. Iqbal, and M. Mohsin. Evaluating the level of household poverty: A case study of Jhang, Pakistan. *Journal of Environmental and Agricultural Sciences* 22(4): 16–24 (2020).
34. M.I. Shah, A. Alam, and M. Shabbir. Problems of Bonded Child Labor in Brick Kilns Industry at Peshawar, Pakistan. *Pakistan Social Sciences Review* 4(III): 209–217 (2020).
35. N. Malik. Bonded Labour in Pakistan. *Advances in Anthropology* 6(4): 127–136 (2016).

36. D. A. Usman, J.R. Kataria, H. Awan, S. Hussain, and M. Usman. Elimination of Bonded Labour in South Asia: The Process Patronage and Labourers' Right to Access Micro Credit in Brick Kiln Industry of Pakistan. *South Asian Studies* 30(1): 71–80 (2015).
37. K. Shaikh, U. Imran, A. Khan, W.A. Khokhar, and H. Bakhsh. Health risk assessment of emissions from brick kilns in Tando Hyder, Sindh, Pakistan using the AERMOD dispersion model. *SN Applied Sciences* 2(7): 1–11 (2020).

Proceedings of the Pakistan Academy of Sciences: A Physical and Computational Sciences 60(4): 69-72 (2023)

ISSN (Print): 2518-4245; ISSN (Online): 2518-4253
https://doi.org/10.53560/PPASA(60-4)669

Pakistan Academy of Sciences

Research Article

Analytical Solution for Cylindrical Shell of Permeable Material in Fractional Dimensional Space

Muhammad Akbar[1], Saeed Ahmed[2*], Muhammad Imran Shahzad[3], Muhammad Ahmad Raza[4], and Sania Shaheen[3]

[1]Department of Physics, Air University, Islamabad, Pakistan
[2]Department of Physics, Quaid-i-Azam University, Islamabad, Pakistan
[3]Department of Applied Physics, Federal Urdu University of Arts, Science and Technology, Islamabad, Pakistan
[4]Department of Statistics, Federal Urdu University of Arts, Science and Technology, Islamabad, Pakistan

Abstract: We have investigated the Laplacian equation in fractional dimensional space (FDS) that is widely used in physics to describe many complex phenomena. Using this concept, we have applied it on a cylindrical shell of permeable material to find the analytical solution of electric potential in FDS. The derivation of this problem is performed by applying Gegenbauer polynomials. The general solution has been obtained in a closed form in the FDS and can be applied to the cylindrical shell for different materials inside the cylinder core and outside the shell**.** By setting the fractional parameter $\alpha = 3$, the derived solution is retrieved for the integer order.

Keywords: General Close Form Solution, Gegenbauer Polynomials, Laplace Equation, Method of Separation Variable.

1. INTRODUCTION

The concept of fractional-dimensional space (FDS) has been applied in many areas of physics for the last many years. It has been investigated and cited by many theoretical physicists [1-19] Wilson [3] has utilized FD space to explain the quantum field theory. Stillinger [4] has formulated Gibbsian statistical mechanics and Schrodinger wave by using this novel idea of the FD space. In the Ising domain [6], the quantum field theory FD space can be utilized as a parameter. Zeilinger and Svozil [10] have elaboratedthe meaning of the dimension of space time by which space-time dimension may be predicted. By using this concept, it has also been estimated that the FD of space-time is approximately less than 4 Gauss law [11] has been derived in α-dimensional fractional space. Various electrostatic issues have been resolved [13-18] in the fractional-dimensional space for $(2 < \alpha \leq 3)$.
This problem has been studied from Jackson [13] for permeable material for the FDS. The main objective of this paper is to evaluate the electric field and potential due to a permeable cylindrical shell by applying the Laplacian equation in FD space. The summary of this paper is described as follows. First of all, we have formulated the boundary value problem. Then in a host medium of uniform electrical field, a permeable cylindrical shell is placed Then, we have evaluated a general solution for the given problem by applying the separation variable method, next we have constructed the solution for the three regions accordingly. Here**,** we apply boundary conditions, and finally**,** we have calculated unknown constants, by using the boundary conditions as well as known values. Now we are able to find the complete solution in form of electric potential due to the cylindrical shell of highly permeable material for fractional dimensional space.

2. MATHEMATICAL MODEL

A cylindrical shell of infinite length filled with permeable material $\mu 1/\mu 0$ having outer and inner radii '*b*' and '*a*' respectively has been considered

Received: January 2023; Accepted: December 2023
*Corresponding Author: Saeed Ahmed <saeedqau@gmail.com>

cylinder. We have investigated the potential in NID space ($2 < \alpha \leq 3$), in the three regions. For an appropriate solution the cylindrical coordinates (r, θ) are employed.

$$\nabla^2\Psi(r,\theta) = 0 \quad (1)$$

This is Laplace equation in cylindrical coordinates.

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\Psi}{\partial}\right) + \frac{1}{r^2}\frac{\partial^2\Psi}{\partial\theta^2} + \frac{\partial^2\Psi}{\partial z^2} + k^2\Psi = 0 \quad (2)$$

As this problem is considered in magneto statics, where $\omega = 0$ it means $k = 0$ and for the symmetry of the problem about the z-axis, Ψ is independent of z, therefore, the above wave equation reduces to:

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\Psi}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2\Psi}{\partial\theta^2} = 0 \quad (3)$$

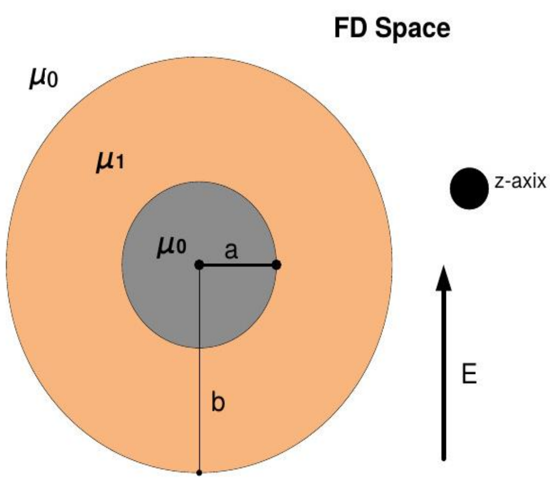

Fig. 1. Cylindrical Shell of Permeable Material Placed in Fractional Space.

Eq. (3) is solved by separable method

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\Psi}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2\Psi}{\partial\theta^2} = 0 \quad (4)$$

Suppose the solution is:

$$\Psi(r,\theta) = R(r)\Theta(\theta) \quad (5)$$

The obtained angular and radial part of the differential equations (4) in NID Space by following the previous research work [18, 19], we find:

$$\left[\frac{d^2}{d\theta^2} + (\alpha-2)cot\theta\frac{d}{d\theta} + l(l+\alpha-2)\right]\Theta(\theta) = 0 \quad (6)$$

$$\left[\frac{d^2}{dr^2} + \frac{\alpha-2}{r}\frac{d}{dr} + \frac{l(l+\alpha-3)}{r^2}\right]R(r) = 0 \quad (7)$$

Here, the generalized solution of the scalar potential of cylindrical shell in fractional space followed by Morse and Feshbach [17] can be expressed as:

$$\Psi(r,\theta) = \sum_{l=0}^{\infty}\left[a_l r^l + b_l r^{-(l+\alpha-3)}\right]P_l^{\alpha/2-1}(\cos\theta) \quad (8)$$

For the physical solution, we find a close form solution inside and outside of the cylindrical region. But, we are interested only in the solution for $l = 1$, that is $P^{(\alpha/2-1)}{}_l(\cos\theta = (\alpha-2)\cos(\theta))$ As there is uniform symmetry with respect to the external field for each region, we find the potential outside region.

$$\Psi(r,\theta) = \left(-E_0 r + A_1 r^{-(\alpha-2)}\right)(\alpha-2)cos\theta, r > b \quad (9)$$

In between the cylinders:

$$\Psi(r,\theta) = \left(B_1 r + C_1 r^{-(\alpha-2)}\right)(\alpha-2)\cos\theta,\ a < r < b \quad (10)$$

and inside the cylinder:

$$\Psi(r,\theta) = D_1 r(\alpha-2)cos\theta, 0 < r < a \quad (11)$$

Here, we have imposed boundary conditions at r = a and r = b These four boundary conditions lead to the following equations in the simplified form such that:

$$A = E_0 b^{\alpha-1} + Bb^{\alpha-1} + C \quad (12)$$

$$a_1 A = -E_0 b^{\alpha-1} - \kappa_1 Bb^{\alpha-1} + a_1\kappa_1 C \quad (13)$$

Where $\boldsymbol{\alpha}_1 = (\alpha-2)$ and $\kappa_1 = \mu/\mu_0$

$$D = B + a^{-(\alpha-1)}C \quad (14)$$

$$D = \kappa_1\left(B - a_1 a^{-(\alpha-1)}C\right) \quad (15)$$

By solving these equations simultaneously, we find the required unknown constants:

$$B = \frac{-(\alpha-1)(\kappa_1 a_1+1)E_0 b^{(\alpha-1)}}{(\kappa_1+a_1)(\kappa_1 a_1+1)b^{\alpha-1}-a_1} \quad (16)$$

$$C = \frac{(\alpha-1)(1-\kappa_1)E_0 a^{(\alpha-1)}b^{(\alpha-1)}}{(\kappa_1+a_1)(\kappa_1 a_1+1)b^{\alpha-1}-a_1} \quad (17)$$

$$D = \frac{-(\alpha-1)(\alpha-1)E_0 b^{(\alpha-1)}}{(\kappa_1+a_1)(\kappa_1 a_1+1)b^{\alpha-1}-a_1} \quad (18)$$

And

$$A = E_0 b^{(\alpha-1)} + \frac{(1-\kappa_1)a^{(\alpha-1)}-(1+\kappa_1 a_1)b^{(\alpha-1)}}{(\kappa_1+a_1)(\kappa_1 a_1+1)b^{\alpha-1}-a_1} \quad (19)$$

Then we recover the exact solution [13] by setting α = 3 and $\alpha_1 = (\alpha\text{-}2)$ that is given below:

$$B = \frac{-2E_0 b^2(1+\kappa_1)}{b^2} \quad (20)$$

$$C = \frac{2E_0 a^2 b^2(1-\kappa_1)}{b^2} \quad (21)$$

$$D = \frac{-4E_0 b^2}{b^2} \quad (22)$$

Similarly, we find constant A for α=3

$$A = E_0 b^2 + 2E_0 b^2 \frac{a^2(1-\kappa_1)-b^2(1+\kappa_1)}{b^2} \quad (23)$$

Special Cases

For a magnetic cylinder, if we allow a → 0, we obtained the solution such that for the fractional-dimensional space.

$$A = \frac{\kappa_1-1}{\kappa_1+1} E_0 b^{\alpha-1} \quad (24)$$

For Integer order

$$A = \frac{\kappa_1-1}{\kappa_1+1} E_0 b^2, for\ \alpha = 3 \quad (25)$$

In FD Space

$$B = \frac{-(\alpha-1)E_0}{\kappa_1+1},\ for\ \alpha = 3 \quad (26)$$

In Integer order

$$B = \frac{-2E_0}{\kappa_1+1},\ for\ \alpha = 3 \quad (27)$$

$$C = 0 \quad (28)$$

In FD Space

$$D = \frac{-(\alpha-1)(\alpha-1)E_0}{(\kappa_1+1)^2} \quad (29)$$

$$D = \frac{-4E_0}{(\kappa_1+1)^2},\ for\ \alpha = 3 \quad (30)$$

For the cylindrical cavity, if we allow b → ∞, we find

$$B = \frac{-(\alpha-1)E_0}{\kappa_1+1} \quad (31)$$

$$B = \frac{-2E_0}{\kappa_1+1}, for\ \ \alpha = 3 \quad (32)$$

In FD Space

$$C = (\alpha-1)E_0 a^{\alpha-1} \frac{1-\kappa_1}{(1+\kappa_1)^2} \quad (33)$$

$$C = 2E_0 a^2 \frac{1-\kappa_1}{(1+\kappa_1)^2},\ for\ \alpha = 3 \quad (34)$$

In FD Space

$$D = \frac{-(\alpha-1)(\alpha-1)E_0}{(1+\kappa_1)^2} \quad (35)$$

$$D = \frac{-4E_0}{(1+\kappa_1)^2},\ for\ \alpha = 3 \quad (36)$$

3. RESULTS AND DISCUSSION

In this article, we have solved a boundary value problem analytically by using the Laplace equation for NID Space. Here, we have applied four boundary conditions on a cylindrical shell that is made of magnetic material. Then we have calculated an electric potential due to the permeable cylindrical shell. Using the electric potential, we can find electric field. In this article we have extended the classical solution of electric potential in FDS due to the spherical shell of highly permeable material from Jackson [13]. Further, this solution can be applied for various materials. For an example, we can take plasma material as a host medium, also we can fill the cylindrical shell with plasma media like magnetized plasma, anisotropic plasma, isotropic plasma, uniaxial plasma, bi-axial plasma, cold plasma, hot plasma, un magnetized plasma. Similarly, we can also use meta materials as host medium as well as material between the cylindrical shell. The shielding effect due to the highly permeable material causes enough reduction in the field inside cylinder, even if the cylindrical shell is thin. Further, we have also discussed its special cases, (i) If its radius α approaches to zero, we find a solution for a uniform magnetic cylinder, (ii) as we allowed radius *b* approach to infinity, we find a solution for a cylindrical cavity.

4. CONCLUSIONS

In this paper, the analytical solution has been calculated for the permeable cylindrical shell in non-integer dimensional (NID) space. Considering **α** = 3, the classical results are retrieved. The shielding effect due to the highly permeable material causes enough reduction in the field inside cylinder, even if the cylindrical shell is thin. We may apply it further for multiple materials like meta-materials, plasma etc. as the host medium and core medium.

5. CONFLICT OF INTEREST

Authors have no conflict of interest.

6. REFERENCES

1. C.G. Bollini, and J.J. Giambiagi. Dimensional renormalization: The number of dimensions as a regularizing parameter. *Nuovo Cimento B* 12: 20-26 (1972).
2. J.F. Ashmore, On renormalization and complex space time dimensions. *Communications in Mathematical Physics* 29: 177-187 (1973).
3. K.G. Wilson. Quantum Field - Theory Models in Less Than 4 dimensions. *Physical Review D* 7(10): 2911-2926 (1973).
4. F.H. Stillinger. Axiomatic basis for spaces with non integer dimension. *Journal of Mathematical Physics* 18 (6): 1224-1234 (1977).
5. X.F. He. Excitons in anisotropic solids: The model of fractional-dimensional space. *Physical Review B* 43(3): 2063-2069 (1991).
6. C.M. Bender, and S. Boettcher. Dimensional expansion for the Ising limit of quantum field theory. *Physical Review D* 48 (10): 4919-4923 (1993).
7. C.M. Bender, and K.A. Milton. Scalar Casimir effect for a D-dimension sphere. *Physical Review D* 50(10): 6547-6555 (1994).
8. V.E. Tarasov. Fractional generalization of Liouville equations. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 14: 123-127 (2004).
9. V.E. Tarasov. Electromagnetic fields on fractals. *Modern Physics Letters A* 21(20): 1587-1600 (2006).
10. A. Zeilinger, and K. Svozil. Measuring the dimension of space time. *Physical Review Letters*. 54: 2553-2555 (1985).
11. S. Muslih, and D. Baleanu. Fractional multipoles in fractional space. *Nonlinear Analysis: Real World Applications* 8: 198-203 (2007).
12. C. Palmer, and P.N. Stavrinou. Equations of motion in a non-integer-dimensional space. *Journal of Physics A* 37: 6987-7003 (2004).
13. J. D. Jackson. Classical Electrodynamics (3rd Ed.), *John Wiley, USA, pp. 201-202.*
14. J. Adams. Stratton Electromagnetic Theory. *Wiley IEEE Press, pp. 640*
15. T. Myint-U, and L. Debnath, Linear Partial Differential Equations for Scientists and Engineers, (4th ed.). *Springer Science & Business Media,* Birkhauser, *Boston, USA* (2007).
16. V.E. Tarasov. Gravitational field of fractals distribution of particles. *Celestial Mechanics and Dynamical Astronomy* 94: 1-115 (2006).
17. P.M. Morse, and H. Feshbach. Methods of Theoretical Physics, Part II. *McGraw-Hill*, *New York,* pp. 1184-1185 (1953).
18. D. Baleanu, A.K. Golmankhaneh, A.K. Golmankhaneh, On electromagnetic field in fractional space. *Nonlinear Analysis: Real World Applications* 11(1): 288-292 (2010).
19. V.E. Tarasov. Vector Calculus in Non-Integer Dimensional Space and its Applications to Fractal Media. *Communications Nonlinear Science and Numerical Simulations* 20(2): 360-374 (2015).

Instructions for Authors

Manuscript Format

The manuscript may contain Abstract, Keywords, INTRODUCTION, MATERIALS AND METHODS, RESULTS, DISCUSSION (or RESULTS AND DISCUSSION), CONCLUSIONS, ACKNOWLEDGEMENTS, CONFLICT OF INTEREST and REFERENCES, *and any other information that the author(s) may consider necessary*.

Abstract (font size 10; max 250 words)**:** Must be self-explanatory**,** stating the rationale, objective(s), methodology, main results, and conclusions of the study. Abbreviations, if used, must be defined on the first mention in the Abstract as well as in the main text. Abstract of review articles may have a variable format.

Keywords (font size 10)**:** Three to eight keywords**,** depicting the article.

INTRODUCTION: Provide a clear and concise statement of the problem, citing relevant recent literature, and objectives of the investigation.

MATERIALS AND METHODS: Provide an adequate account of the procedures or experimental details, including statistical tests (if any), concisely but sufficient enough to replicate the study.

RESULTS: Be clear and concise with the help of appropriate Tables, Figures, and other illustrations. Data should not be repeated in Tables and Figures, but must be supported with statistics.

DISCUSSION: Provide interpretation of the RESULTS in the light of previous relevant studies, citing published references.

ACKNOWLEDGEMENTS: (font size 10)**:** In a brief statement, acknowledge the financial support and other assistance.

CONFLICT OF INTEREST: State if there is any conflict of interest.

REFERENCES (font size 10)**:** Cite references in the text **by number only** in **square brackets**, e.g. "Brown et al [2] reported ..." or "... as previously described [3, 6–8]", and list them in the REFERENCES section, in the order of citation in the text, Tables and Figures (not alphabetically). Only published (and accepted for publication) journal articles, books, and book chapters qualify for REFERENCES.

Declaration: Provide a declaration that: (i) the results are original; (ii) the same material is neither published nor under consideration elsewhere; (iii) approval of all authors have been obtained; and (iv) in case the article is accepted for publication, its copyright will be assigned to *Pakistan Academy of Sciences*. Authors must obtain permission to reproduce, where needed, copyrighted material from other sources and ensure that no copyrights are infringed upon.

Manuscript Formatting

Manuscripts must be submitted in Microsoft Word (2007 Version .doc or .docx format); **pdf** files not acceptable. Figures can be submitted in Word format, TIFF, GIF, JPEG, EPS, PPT. Manuscripts, in *Times New Roman*, 1.15spaced (but use single-space for Tables, long headings, and long captions of tables & figures). The text must be typed in a double-column across the paper width. The Manuscript sections must be numbered, i.e., **1. INTRODUCTION, 2. MATERIALS AND METHODS,** and so on… (a) **Title** of the article **(**Capitalize initial letter of each main word; font-size 16; **bold**), max 160 characters (no abbreviations or acronyms), depicting article's contents; (b) Author' first name, middle initial, and last name (font size 12, **bold**), and professional affiliation (i.e., each author's Department, Institution, Mailing address and Email; but no position titles) (font size 12); (c) Indicate the corresponding author with *; (d) **Short running title**, max 50 characters (font size 10).

Headings and Subheadings (font size 11): All flush left

LEVEL-1: **ALL CAPITAL LETTERS; Bold**

Level-2: **Capitalize Each Main Word (**Except prepositions**); Bold**

Level-3: Capitalize each main word (Except prepositions)***; Bold, Italic***

Level-4: *Run-in head*; *Italics*, *in the normal paragraph position. Capitalize the initial word only and end in a colon (i.e.,* ***:****)*

List of REFERENCES must be prepared as under:

a. Journal Articles *(Name of journals must be stated in full)*

1. I. Golding, J. Paulsson, S.M. Zawilski, and E.C. Cox. Real time kinetics of gene activity in individual bacteria. *Cell* 123: 1025–1036 (2005).
2. W. Bialek, and S. Setayeshgar. Cooperative sensitivity and noise in biochemical signaling. *Physical Review Letters* 100: 258–263 (2008).
3. R.K. Robert, and C.R.L.Thompson. Forming patterns in development without morphogen gradients: differentiation and sorting. *Cold Spring Harbor Perspectives in Biology* 1(6) (2009).
4. D. Fravel. Commercialization and implementation of biocontrol. *Annual Reviews of Phytopathology* 43: 337359 (2005).

b. Books

5. W.R. Luellen. Fine-Tuning Your Writing. *Wise Owl Publishing Company, Madison, WI, USA* (2001).
6. U. Alon, and D.N. Wegner (Ed.). An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman & Hall/CRC, Boca Raton, FL, USA* (2006).

c. Book Chapters

7. M.S. Sarnthein, and J.D. Stanford. Basal sauropodomorpha: historical and recent phylogenetic developments. In: The Northern North Atlantic: A Changing Environment. P.R. Schafer, & W. Schluter (Ed.), *Springer, Berlin, Germany*, pp. 365–410 (2000).
8. J.E. Smolen, and L.A. Boxer. Functions of Europhiles. In: Hematology, 4th ed. W.J. Williams., E. Butler and M.A. Litchman (Ed.), *McGraw Hill, New York, USA*, pp. 103–101 (1991).

d. Reports

9. M.D. Sobsey, and F.K. Pfaender. Evaluation of the H2S method for Detection of Fecal Contamination of Drinking Water, Report WHO/SDE/WSH/02.08, *Water Sanitation and Health Programme, WHO, Geneva, Switzerland* (2002).

e. Online references

These should specify the full URL for reference and give the date on which it was consulted. Please check again to confirm that the work you are citing is still accessible:

10. L. Branston. SENSPOL: Sensors for Monitoring Water Pollution from Contaminated Land, Landfills and Sediment (2000). http://www.cranfield.ac.uk/biotech/senspol/ (accessed 22 July 2005)

Tables and Figures

Insert all tables as editable text, not as images. Number tables consecutively following their appearance in the text, Figures should appear in numerical order, be described in the body of the text, and be positioned close to where they are first cited. Each figure should have a caption that describes the illustration, and that can be understood independently of the main text (Caption Table 1. and Fig 1. font size 10; Bold; Captions should be in sentence case; left-aligned). All Figures should have sufficiently high resolution (minimum 1000 pixels width/height, or a resolution of 300 dpi or higher) to enhance the readability. Figures may be printed in two sizes: column width of 8.0 cm or page width of 16.5 cm; number them as **Fig. 1**, **Fig. 2**, … in the order of citation in the text. Parts in a figure can be identified by A, B, C, D, ... and cited as Figure 2A, Figure 2B, Figure 2C. Captions to Figures must be concise but self-explanatory. Laser printed line drawings are acceptable. Do not use lettering smaller than 9 points or unnecessarily large. Photographs must be of high quality. A scale bar should be provided on all photomicrographs.

Tables: with concise but self-explanatory headings must be numbered according to the order of citation (like **Table 1.**, **Table 2.**). *Do not* abbreviate the word "Table" to "Tab.". Round off data to the nearest three significant digits. Provide essential explanatory footnotes, with superscript letters or symbols keyed to the data. Do not use vertical or horizontal lines, except for separating column heads from the data and at end of the Table.

Figures: Figures may be printed in two sizes: column width of 8.0 cm or page width of 16.5 cm; number them as **Fig. 1**, **Fig. 2**, … in the order of citation in the text. Captions to Figures must be concise but self-explanatory. Laser printed line drawings are acceptable. Do not use lettering smaller than 9 points or unnecessarily large. Photographs must be of high quality. A scale bar should be provided on all photomicrographs.

Note: The template of the manuscript is available at http://www.paspk.org/proceedings/; http://ppaspk.org/

Reviewers: Authors may suggest four relevant reviewers, two National and two International (with their **institutional E-mail** addresses).

SUBMISSION CHECKLIST

The following list will be useful during the final checking of an article before sending it to the journal for review.

Ensure that the following items are present:

One author has been designated as the corresponding author with contact details:

- E-mail address (Correct and valid)
- Full address of Institute/organization
- Keywords
- All figure captions
- All tables (including title, description, footnotes)

Further considerations

- Manuscript has been 'spell-checked' and 'grammar checked'
- References are in the correct format for this journal
- All references mentioned in the Reference list are cited in the text, and vice versa
- Permission has been obtained for the use of copyrighted material from other sources (including the Internet)

In case of any difficulty while submitting your manuscript, please get in touch with:

Editor
Pakistan Academy of Sciences
3-Constitution Avenue, Sector G-5/2
Islamabad, Pakistan
Email: editor@paspk.org
Tel: +92-51-920 7140
Websites: http://www.paspk.org/proceedings/; http://ppaspk.org/

PROCEEDINGS OF THE
PAKISTAN ACADEMY OF SCIENCES: PART A
Physical and Computational Sciences

CONTENTS

Volume 60, No. 4, December 2023 **Page**

PAKISTAN ACADEMY OF SCIENCES, ISLAMABAD, PAKISTAN

HEC Recognized, Category Y; Scopus Indexed

Websites: http://www.paspk.org/proceedings/; http://ppaspk.org