INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNICAL DEVELOPMENT

International Journal of Scientific & Technical Development

Table of Contents

International Journal of Business Management & Research
(A Peer-Reviewed Bi-Annual Journal) ISSN: 2249-2143

Medical Data Security and ECG Steganography:
Issues and Challenges

Butta Singh, Himali Sarangal and Manjit Singh
Guru Nanak Dev University, Regional Campus, Jalandhar
Corresponding Author Email ID: bsl.khana@gmail.com

Abstract
It becomes utmost important that the patient confidentiality is ensured while medical data is being sent over the public networks as well as when it is stored in the healthcare repository used by a remote monitoring system. ECG monitoring facilitates in providing accelerated health status of the concerned patient to the healthcare centre in case of the hostile cardiac behaviour. Transmission of this compressed ECG via a communication channel introduces various security and privacy issues [5]. To counteract with these concerns, there is a need for implementation of efficient security protocols. To this effect, several algorithms have been developed in the past two to three decades. The patient's confidential data hiding through watermarking and the ECG encryption techniques are the emerging biometric security mechanisms. This paper attempts to explore the state of the are approaches developed for ECG encryption and steganography.
Keywords: ECG; Steganography; Encryption; Medical Data

Introduction

In the modern era, cardiovascular diseases had emerged as one of the essential causes of mortality in both urban and rural areas [1], [2]. ECG monitoring facilitates in providing accelerated health status of the concerned patient to the healthcare centre in case of the hostile cardiac behaviour [4]. Cardiac monitoring using ECG signal is the best representative of heart's electrical functionality and had proven to be useful in the diagnosis of most of the heart diseases. Transmission of this compressed ECG via a communication channel introduces various security and privacy issues [3]. To counteract with these concerns, there is a need for implementation of efficient security protocols. To this effect, several algorithms have been developed in the past two to three decades [4].

It becomes utmost important that the patient confidentiality is ensured while data is being sent over the public networks as well as when it is stored in the healthcare repository used by a remote monitoring systems [5]. The information sent over the public network or the internet should be highly protected and secured [5], [6]. The techniques used for data ciphering are based on encryption and cryptographic algorithms. After, the process of encryption healthcare centre receives the encrypted ECG data which can be de-encrypted using either the symmetric or asymmetric keys which ensures the higher end-to-end security, when applied to e-health monitoring system [7].

To simultaneously protect the biometric credentials of the patient and to ensure reliable diagnosis of ECG, hybrid approaches that incorporates both the confidential patient data hiding as well as encryption continue to advance in literature. The confidential patient data includes information like Patient Reference ID, Name, Sex, Age, Case History, Temperature, Blood Pressure, Concern Doctor, Prescription etc.

2. ECG DATA SECURITY TECHNIQUES

The patient's confidential data hiding through watermarking and the ECG encryption techniques are the emerging biometric security mechanisms. The ECG encapsulates the sensitive information about the cardiovascular state of patient's ECG. Transmitting it without encryption makes it vulnerable to spoof attack and also is against the HIPAA regulations. So to prevent the imposters from capturing the ECG segments and gaining the unauthorized access to the secured facilities, the segments needed to be efficiently encrypted.

Ayman and Ibrahim [5] proposed a wavelet-based steganography technique which combines encryption and scrambling technique to defend confidential patient data. The proposed method allows ECG signal to hide its corresponding patient private data and other physiological information thus guaranteeing the integration between ECG and the rest.

Lu et al. [8] presented an effective scheme to protect patients personal privacy for a medical information system. In the scheme, personal data was encrypted before being stored in the database of the server of a medical

information system, so that even if server information was disclosed, data be difficult to be decrypted and interpreted. Han et al. [39] proposed the use of multi scroll chaos to encrypt the ECG packets. The encryption was achieved by XORing the ECG packet with the chaos key generated from chaos generator server. The devised approach proved to be 18 times faster than permutation based ECG encoding, 25 times faster than wavelet-based ECG and 31 times faster than noise based ECG data ciphering technique.

Khalil et al. [9] devised a specialized permutation based ECG encryption technique. The permutation key is only known to the authorized personnel, who can decrypt the ECG from encrypted ECG. The original ECG can be transformed into fully encoded ECG (in the form of scrambled ASCII letters). When combined with existing encryption schemes, the strength can be further raised thus providing unmatched protection against spoof attacks.

Sufi et al. [10] proposed an efficient ECG obfuscation method which involved detection of P, QRS complex and T wave feature from ECG and then replacing them with their noisy versions. The advantage of the proposed approach over existing encryption method was that the corrupted ECG appears as regular ECG though it is encrypted ECG. The limitation of using this method was that the overall ECG data size increased by 0.9% by the addition of the key for decryption.

Zhou et al. [11] used quantization-based digital watermark encryption technology on the ECG to protect patient rights and information. The patient's confidential data, e.g., name, age, and ID, etc. are collected and treated as a watermark for medical data. They concluded that after testing with ten selected data sets from the MIT-BIH arrhythmia database, the difference between the watermarked ECG and the original one is very small and negligible for physiological diagnostics.

Murillo-Escobar et al. [12] proposed a symmetric encryption algorithm based on logistic map with double chaotic layer encryption in diffusion process and just one round of confusion-diffusion for the confidentiality and privacy of clinical information such as ECG, Electroencephalograms (EEG), and Blood Pressure (BP) for applications in telemedicine. The achieved information entropy is 7.96 for 8-bit quantization.

Raeiatibanadkooki et al. [13] designed a mechanism for encrypting the ECG signal. The sequence of steps followed for preprocessing were the removal of the baseline noise, peak detection, and determination of heart rate. Huffman coding with chaos accomplishes the ECG signal encryption.

A steganography approach based on Discrete Wavelet Transform (DWT) and SVD was developed by Jero et al. [14]. DWT was used to decompose the signal and SVD was employed to embed the secret information into the signal. Experimenting with 76,800 ECG samples from MIT-BIH arrhythmia database with secret data size equal to 350 bytes, the reported values of Peak Signal to Noise Ratio (PSNR), PRD, and Kulback-Leibler Divergence (KL-Div) were equal to 69.13, 0.0687, and 6.84×10^{-5}. The designed method resulted in an overall signal degradation of 0.6%. The approach is very sensitive to the selection of mother wavelet and provides low embedding rate. Another Transformed-domain quantization based scheme for ECG steganography was proposed in [11]. 4096 ECG samples from MIT-BIH arrhythmia database has been selected for experimentation. The achieved amplitude similarity, Relative Root Mean Square Error (rRMSE), and amplitude Root Mean Square Error (RMS) were equal to 99.96, 0.153, and 22.847 for ECG record 103 for a secret data size of 32 bits. Later in the year 2015, authors in [14] explored fast discrete curvelet transform with adaptive thresholding for ECG Steganography. After experimenting on 128 trains of ECG from MIT-BIH Normal Sinus Rhythm (NSR), the obtained values were 43.44, 0.0132, and 0.1448 for the PSNR, PRD, and KL Distance at the payload equal to 502 bytes. Every 1.5 times increase in the patient confidential data, the ECG signal deterioration increases by 10 %. Same research group proposed a Continuous Ant Colony Optimization (CACO) algorithm based ECG steganography scheme using DWT and SVD [48]. The PSNR, PRD, and KL-Div values for a secret data size of 21 Kb were 34.46, 0.06, and 2.04 for Normal Sinus Rhythm (NSR) dataset. The parameter selection for the CACO was computationally expensive while extending the proposed method to other signals. Yang and Wang [49] validated their lossy and the reversible ECG steganography technique on MIT-BIH arrhythmia database. Investigations on Lead II ECG signals resulted in average values of SNR, and mean absolute error equal to 56.34, and 0.90 at a payload equal to 7500 bits.

3. PERFORMANCE EVALUATION METRICS

The performance evaluation metrics are broadly divided into the categories of ECG compression, encryption, and the transmission. The brief description is given as under:

3.1 ECG Data Compression

The performance evaluation metrics for ECG compression include CR, PRD, Percentage Root Mean Square Difference with Base Removed (PRD1024), Percentage Root Mean Square Difference Normalized (PRDN), RMS Error, SNR, Quality Scores (QS, QS1024, QSN), Wavelet-based Weighted Percentage Root Mean Square Difference (WWPRD) and Wavelet Energy Based Diagnostic Distortion (WEDD).

3.2 ECG Data Security

The performance evaluation analysis for ECG stenography includes reconstructed ECG security and sensitivity analysis, Bit Error Rate for Secret data (BERs), Embedding Score (ES). Evaluation analysis for ECG encryption includes Histogram analysis, Entropy analysis, Correlation analysis.

3.3 ECG Transmission

The performance evaluation analysis for processed ECG transmission and reception includes Channel Signal to Noise Ratio: SNRc and Bit Error Rate due to Channel and receiver: BERc.

4. CONCLUSION

Most of the approaches mentioned above lack a suitable trade-off between the attributes of higher secret data security, reversibility of secret data, lower degradation of stego-ECG signal and higher Embedding Capacity (EC). Increasing EC often leads to higher degradation of stego-ECG signal. ECG steganography causes irreversible degradation to stego-ECG as compared to the cover ECG signal. Many approaches have been devised for ECG data encryption, yet the previously reported works did not focus on the issues of simultaneously processing the patient confidential data in ciphered version and ECG signal in the encrypted version.

REFERENCES

[1] M. Nichols, N. Townsend, P. Scarborough, and M. Rayner, "Cardiovascular disease in Europe: Epidemiological update," Eur. Heart J., vol. 35, no. 42, pp. 3028–3034, 2014.

[2] S. Gupta, R. Gudapati, K. Gaurav, and M. Bhise, "Emerging risk factors for cardiovascular diseases: Indian context," Indian J. Endocrinol. Metab., vol. 17, no. 5, pp. 806–814, 2013.

[3] S. S. Mahmoud, Q. Fang, Z. M. Hussain, and I. Cosic, "A blind equalization algorithm for biological signals transmission," Digit. Signal Process. A Rev. J., vol. 22, no. 1, pp. 114–123, 2012.

[4] A. Ibaida, D. Al-Shammary, and I. Khalil, "Cloud enabled fractal based ECG compression in wireless body sensor networks," Futur. Gener. Comput. Syst., vol. 35, pp. 91–101, 2014.

[5] A. Ibaida and I. Khalil, "Wavelet-based ECG steganography for protecting patient confidential information in point-of-care systems," IEEE Trans. Biomed. Eng., vol. 60, no. 12, pp. 3322–3330, 2013.

[6] W. B. Lee and C. D. Lee, "A cryptographic key management solution for HIPAA privacy/security regulations," IEEE Trans. Inf. Technol. Biomed., vol. 12, no. 1, pp. 34–41, 2008.

[7] S. K. Chen, T. Kao, C. T. Chan, C. N. Huang, C. Y. Chiang, C. Y. Lai, T. H. Tung, and P. C. Wang, "A reliable transmission protocol for zigbee-based wireless patient monitoring," IEEE Trans. Inf. Technol. Biomed., vol. 16, no. 1, pp. 6–16, 2012.

[8] C. Lu, Z. Wu, M. Liu, W. Chen, and J. Guo, "A patient privacy protection scheme for medical information system," J. Med. Syst., vol. 37, no. 6, 2013.

[9] F. Sufi and I. Khalil, "Enforcing secured ECG transmission for realtime telemonitoring : A joint encoding , compression , encryption mechanism," Secur. Comm. Networks, vol. 1, no. 5, pp. 389–405, 2008.

[10] F. Sufi and I. Khalil, "A new feature detection mechanism and its application in secured ECG transmission with noise masking," J. Med. Syst., vol. 33, no. 2, pp. 121–132, 2009.

[11] S.-T. T. Chen, Y.-J. J. Guo, H.-N. N. Huang, W.-M. M. Kung, K.-K. K. Tseng, and S.-Y. Y. Tu, "Hiding patients confidential datainthe ECG signal viaa transform-domain quantization scheme," J. Med. Syst., vol. 38, no. 6, p. 54, 2014.

[12] M. A. Murillo-Escobar, L. Cardoza-Avendaño, R. M. López-Gutiérrez, and C. Cruz-Hernández, "A double chaotic layer encryption algorithm for clinical signals in telemedicine," J. Med. Syst., vol. 41, no. 4, pp. 1–17, 2017.

[13] M. Raeiatibanadkooki, S. R. ahati Quchani, M. M. KhalilZade, and K. Bahaadinbeigy, "Compression and encryption of ECG signal using wavelet and chaotically huffman code in telemedicine application," J. Med. Syst., vol. 40, no. 3, p. 73, 2016.

ASSESSING AND FORECASTING ACADEMIC ACHIEVEMENT OF STUDENTS

Maninderjit kaur

CT University, Ludhiana

Abstract

Research in the fields of educational data mining (EDM) and learning analytics (LA) has become more intriguing since it reveals practical information from educational databases for a variety of uses, including forecasting students' progress. Predicting a student's performance can be useful for decisions in contemporary educational systems. While family expenditures and student personal information are frequently disregarded, existing techniques have employed aspects that are mostly connected to academic success and family financial assets. Learning analytics, discriminative classification models, and generative classification models are used to forecast a student's likelihood of successfully completing his degree.

Keywords: Educational data mining (EDM), Learning analytics (LA)

Introduction

Research in the fields of educational data mining (EDM) and learning analytics (LA) has become more intriguing since it reveals practical information from educational databases for a variety of uses, including forecasting students' progress. Predicting a student's performance can be useful for decisions in contemporary educational systems. While family expenditures and student personal information are frequently disregarded, existing techniques have employed aspects that are mostly connected to academic success and family financial assets.

It is essential to properly implement the appropriate pedagogical interventions to guarantee that students graduate on time and satisfactorily. This requires making accurate predictions about students' forthcoming performance founded on their continuing academic records. Although there is a wealth of research on forecasting student achievement when tackling issues or preparing for classes using data-driven techniques, predicting student performance when finishing degrees (such as college programmes) is considerably less explored and confronts unique challenges: (1) Backgrounds and course choices vary greatly across students; (2) Courses are not all equally useful for creating precise forecasts; (3) Student growth must be taken into account in the prediction.

The proposed system will comprise the tracking of detailed info of a student concerning his academic and curricular activity and would predict the right learning courses using an algorithm over the information tracked meeting the ambition or the goal for a student. In the last decade, school conducts examination manually. It has so many problems. The existing systems are very time consuming. It is difficult to analyse the exam manually. Results are not precise as calculation and evaluations are done manually. Result processing after summation of exam takes more time as it is done manually. So we introduce a Pre-school examination Portal system, which is fully computerized. Existing system is a large man power process and is difficult to implement. It provides an easy to use environment for both Test Conductors and Students appearing for Examination. In this study, a model for predicting student success in an academic organisation is provided. The method used is a machine learning approach known as a neural network. Furthermore, the significance of various attributes or qualities is studied in order to establish which of these are associated to student success.

It is essential to properly implement the appropriate pedagogical interventions to guarantee that students graduate on time and satisfactorily. This requires making accurate predictions about students' upcoming performance based on their continuing academic records. Predicting student success in finishing degrees (such as college programmes) is far less researched and presents additional obstacles, despite the fact that there is a wealth of literature on doing so:

1. Backgrounds and course choices vary greatly across students;
2. Courses are not all equally useful for creating precise forecasts;
3. Student growth must be taken into account in the prediction.

In this study, we create a cutting-edge machine learning technique for forecasting student success in degree programmes that can deal with these important issues. There are two key components to the suggested tech-

nique. For producing predictions based on students' changing performance states, first a bi-layered structure made up of several base predictors and a cascade of ensemble predictors is built. Second, a data-driven strategy based on probabilistic matrix factorization and latent component models is suggested to find course relevance, which is crucial for developing effective base predictors.

According to the provided studies, academic success of this is mostly based on their prior performance. Our analysis reveals that previous performance has a considerable impact on student achievement. Further, we established that the enactment of SVM increases with increases in dataset size. System will comprise the tracking of detailed information of a student regarding his academic and curricular activity and would predict the right learning courses using an algorithm over the information tracked meeting the ambition or the goal for a student. In the last decade, school conducts examination manually. It has so many problems. The existing systems are very time consuming. It is difficult to analyse the exam manually. Results are not precise as calculation and evaluations are done manually. Result processing after summation of exam takes more time as it is done manually. So we introduce a Pre-school examination Portal system, which is fully computerized. Existing system is a large man power process and is difficult to implement. It provides an easy to use environment for both Test Conductors and Students appearing for Examination. In this study, a model for predicting student success in an academic organisation is provided. The method used is a machine learning approach known as a neural network. Furthermore, the significance of various attributes or qualities is studied in order to establish which of these are associated to student success.

In view of the aforementioned obstacles, we present a unique strategy for forecasting student success in a degree programme in this work. We focus on predicting students' GPAs but the general framework can be used for other student performance prediction tasks. We make three major contributions.

1. We create a unique prediction system based on students' progressing performance phases. It has a bi-layered structure that includes a base predictor layer and an ensemble predictor layer. Multiple base predictors create local predictions in the base layer based on a snapshot of the student's current performance level in each academic term. An ensemble predictor in the ensemble layer forecasts the student's future performance by combining the outcomes of the local predictions and the previous-term ensemble prediction. The cascading of ensemble predictors over academic terms allows the prediction to incorporate students' developing progress while keeping complexity low. We also extract a guarantee of performance for our suggested method.

2. We develop a data-driven course clustering method based on probabilistic matrix factorization, which automatically outputs course clusters based on large, heterogeneous and sparse student course grade data. Base predictors are skilled using a variety of state-of-the-art machine learning techniques based on the discovered course clustering results. Specifically, only relevant courses in the same cluster are used as input to the base predictors. This not only reduces the training complexity but also removes irrelevant information and reduces noise in making the prediction.

3. e perform extensive simulation studies on an undergraduate student dataset collected over three years across 1169 students at the Mechanical and Aerospace Engineering department at UCLA. The results show that our proposed method is able to significantly outperform benchmark methods while preserving educational interpretability.

ARTIFICIAL NEURAL NETWORKS

Artificial Neural Network represents a set of input unites and output unites that are connected to each other by weighted connections. The ANN learns by changing the weights of the connections in a way so it is able to predict the right target label for some input data instances. One of the famous learning algorithms used to train the ANN is Backpropagation Algorithm. ANN has many advantages such as its high resistance to noisy datasets and its well performance on classifying patterns that has not been trained on so it's used in situations when there is a little knowledge of the relation between the class label and the features in the dataset. There are many real world applications of the ANNs such as image and handwritten recognition, speech recognition, laboratory medicine and pathology. There are many types of the ANNs which can be classified based on their architecture and design. One type is a fully connected multilayer feed forward ANN in which the network has an input layer, one or more hidden layers, and the output layer. Furthermore, its connections never return to an input unit or an output unit in the preceding tier. Furthermore,

each unit in layer L supplies input to each unit in layer L+1. A three layer fully connected feed forward ANN has been used in this research. The network consists of an input layer, two hidden layers, and the output layer. The input layer has twenty input unites, neurons, while the first hidden layer has six hidden unites. The second hidden layer has three hidden unites. The fourth layer is the output layer which has only one output unite. The Rectifier Linear Unit has been used as the hidden unites' activation function.

RELATED WORK

In recent years, there has been a lot of interest in machine learning for education. A considerable portion of work focuses on predicting student performance in problem solving or course completion [10]. Many machine learning approaches have been used to construct prediction algorithms, including decision trees [11], artificial neural networks [12], matrix factorization [13], collaborative filters [14], and probabilistic graphical models [15][6]. Most of this work ignores the temporal/sequential effect that students improve their knowledge over time and treats the prediction as a one-time task. To account for the temporal/sequential effect, a three-mode tensor factorization (on student/problem/time) technique for predicting student performance in solving problems in ITSs was developed [16], and a similarity-based algorithm was proposed to issue predictions of student grades in courses only when a certain confidence level is reached [17]. These techniques, however, are not relevant in our environment due to the aforementioned significant discrepancies in forecasting student success in degree programmes. The ensemble learning approach, namely the Exponentially Weighted Average Forecaster (EWAF) [18], is used as a building component in our progressive prediction algorithm to enable progressive prediction of student performance and live update of the predictor as new student data is received. The main difference between an ensemble predictor and the traditional EWAF algorithm is that an ensemble predictor has access to multiple base predictors (experts) as well as the previous-term ensemble predictor, whose output summarises the outputs of all previous-term base predictors (experts), whereas the conventional EWAF algorithm has direct access to all experts. To the best of our knowledge, this is an unique architecture for developing predictors for gradually growing input spaces that decreases design and implementation complexity and scales readily with the number of academic words. In this setting, we prove that each ensemble predictor still performs asymptotically no worse than the best base predictor in hindsight among all previous-term base predictors in the worst case, thereby providing strong performance guarantee. More importantly, when the best base predictor is biased towards current-term base predictors, our algorithm is able to achieve better expected regret than the conventional method that has access to all experts directly and treats them equally.

COURSE RELEVANCE DISCOVERY

Our course relevance discovery technique is based on the latent component model [19] and performs course clustering using the probabilistic matrix factorization algorithm [20], both of which are widely used in recommender systems [21][22][23]. The problem that recommender systems face is similar to that of student performance prediction in that the dataset is sparse in the sense that each user has rated only a small set of items in the entire item space, whereas in our case, each student has taken only a small set of courses in the entire course space. The latent factor model is therefore used to discover the hidden latent factor that resolves the sparsity problem. Unlike recommender systems, which employ the identified latent factor to facilitate user similarity matching and item suggestions, our system leverages the discovered latent factor to cluster relevant courses. It is worth mentioning that sparse factor analysis is utilised in the learning context to evaluate a learner's understanding of the ideas underlying a domain and the linkages between a collection of questions and those concepts [24]. The authors of [25] draw a separate link between recommender systems and student performance prediction. They develop a collaborative filtering algorithm, which is used in recommender systems to recommend items to users based on user similarity. In this paper, we do not develop collaborative filtering prediction algorithms, although they can be adopted as base predictors in our method. More broadly, there is a rich literature on recommending relevant courses or problems to students based on their associated knowledge level, learning styles, and feedbacks [26] [27][28]. Course sequence recommendation, which considers the specific course constraints, was studied in [29]. To utilize logged data for course sequence recommendations and curriculum design, an off-policy estimator was developed to estimate how an unobserved policy performs given an observed policy [30]. A rank aggregation framework is adapted for the discovery of optimal course sequences at the university level [31]. However, whereas this literature aims to

recommend courses/course sequences based on student backgrounds and past performance, the purpose of the current work is to predict future performance based on student backgrounds and past performance for a given curriculum.

METHODOLOGY

We present a unique strategy for developing predictors based on students' growing progress. The important point is that because predictor f_t's input x_{ti} for term t is a superset of predictor f_t's input x_{ti1} for term $t1$, f_t may capitalise on f_{t1}'s prediction output y_{i1} by adding the progressively additional information x_{ti}. This decreases the difficulty of building f_t and makes the prediction algorithm scalable. Our approach to enable such progressive predictions is based on the ensemble learning technique and integrates offline learning and online learning. The proposed architecture consists of two layers — a base prediction layer and an ensemble prediction layer.

1. In the base prediction layer, we construct a set of base predictors H implemented using different prediction algorithms. For each base predictor h ∈H, let $z_{h,it} = h(\theta_i, x_{ti})$ denote the prediction result of h for student i given the student's static feature and the current performance state x_{ti}. The base predictors are trained using a dataset consisting of all student data in the department without differentiating areas to maximally utilize the data. In fact, predictor h may even be trained differently for each term t's prediction task. Therefore, we write $h_t(\theta_i, x_{ti})$ rather than $h(\theta_i, x_{ti})$. Learning the base predictors is done offline.

2. (2) In the ensemble prediction layer, we construct an ensemble predictor for each term. The ensemble predictor f_t for term t synthesizes the previous ensemble output $\hat{y}_{it-1}$ and output of the base predictors $z_{h,it}, \forall h \in H_t$ and makes a final prediction $\hat{y}_{it}$ based on $\hat{y}_{it-1}$ and $z_{h,it}, \forall h \in H_t$. Because students from various places take different courses and in different sequences, the temporal correlation is likely to be variable, the ensemble predictor is trained using student data from the same area. The ensemble predictors are learned online.

3. A system block diagram of the proposed bi-layered architecture for the term t ensemble learning is illustrated in Figure. Although the proposed architecture is easy to understand, a couple of challenges must be addressed in order to achieve good prediction performance. The first challenge is how

to construct the base predictors. Although existing off-the-shelf machine learning algorithms can be used to perform the prediction task, we would like to construct a base predictor that is customized to the considered course grade prediction problem to improve its prediction performance. A specific consideration in this regard is what information should be included in the training of the predictor as well as making the prediction. The second challenge is how to construct the ensemble predictors and take the temporal correlation into account. Specifically, this is to answer how to synthesize the prediction results of the multiple base predictors as well as the prediction from the previous term.
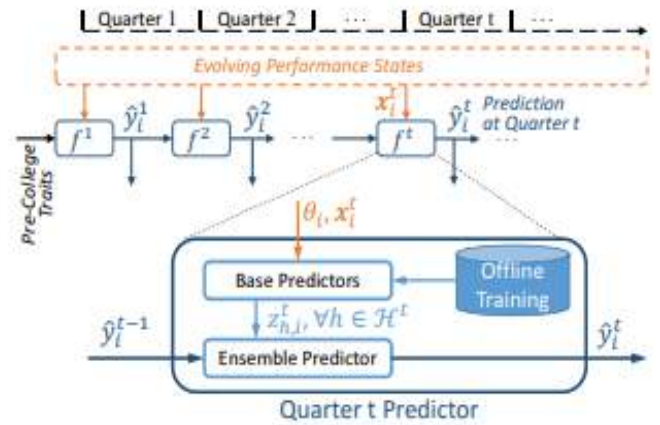

Figure 1: System Block Diagram.

Method Overview

An important question when training h is how to construct the (input) feature vector given the student performance states x . Because students come from different areas as well as have different interests, the courses in the performance states can be very different. A straightforward way is to construct a large feature vector that contains the grade of courses that have appeared in D_t. Entries corresponding to courses that a student did not take in this vector are filled with null values. In this way, students have the same feature vector format. However, there are two major drawbacks for this method. First, the feature vector can be very large, especially in the later terms of the program when students have taken more courses. The problem is more severe since even though students in different areas may have many courses in common, they also have considerably many distinct courses. In addition to the increased complexity, the second drawback is the possible degraded prediction accuracy due to added noise since not all courses, even the courses within the same area, have

predictive power for predicting the grade of the targeted course. Therefore, we will learn the set of courses that are more relevant to the targeted course. Notice that for different targeted courses, the relevant courses will also be different. Once the relevant courses are found, the feature vector is constructed using only elements in x_t that corresponds to the relevant courses. Then our method will utilize various state-of-the-art supervised learning algorithms to train the base predictors. In this paper, we do not invent new supervised learning algorithms but only focus on learning the relevant courses.
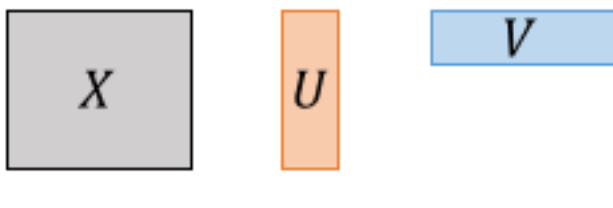
Figure 2: Illustration of matrix factorization.

RESULTS

User interface of the performance Prediction System

Dataset used for the work
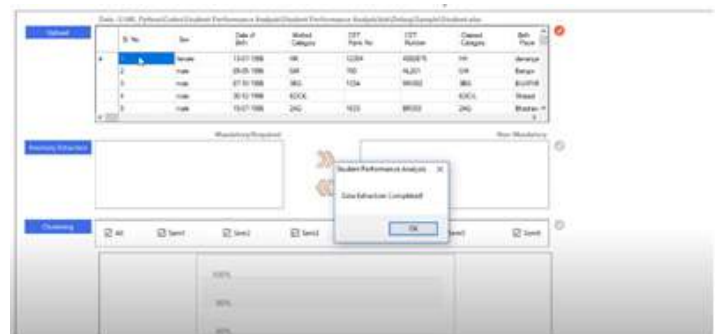
Data Extraction Process

Data extraction
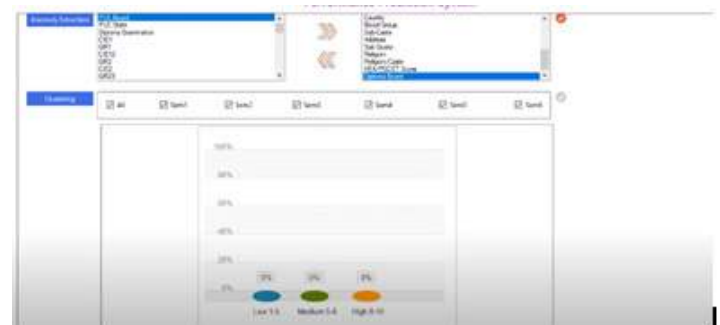
Data Clustering Process

Performance Clustering

Data clustering Completed

Machine Learning Classification
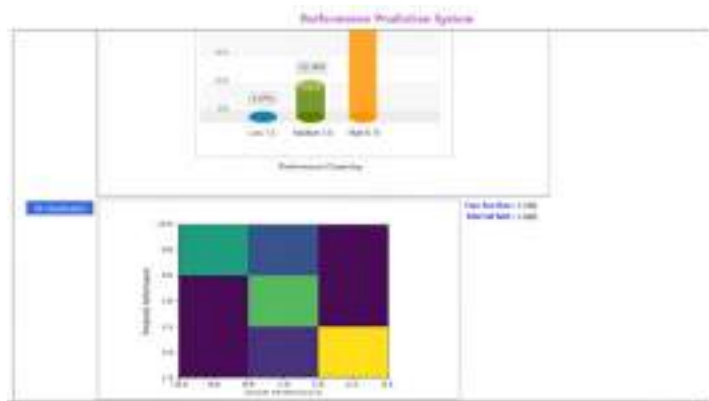
CONCLUSION

According to current research, pupils' academic achievement is mostly determined by their prior performance. Our analysis reveals that previous performance has a considerable impact on pupils' performance. Furthermore, we verified that the performance of neural networks improves as dataset size grows. According to current studies, earlier performance has the greatest influence on students' academic attainment. According to our findings, prior performance has a significant influence on students' performance. Furthermore, we demonstrated that the performance of neural networks increases with increasing dataset size. In the future, applications similar to the one developed, as well as any improvements thereof may become an integrated part of every academic institution. This project can be used in any organization, college as analysis purpose. A latent factor model-based course clustering method was developed to discover relevant courses for constructing base predictors. An ensemble-based progressive prediction architecture was developed to incorporate students' evolving performance into the prediction. These data-driven methods can be used in conjunction with other pedagogical methods for evaluating students' performance and provide valuable information for academic advisors to recommend subsequent courses to students and carry out pedagogical intervention measures if necessary. Additionally, this work will also impact curriculum design in degree programs and education policy design in general. Future work includes extending the performance prediction to elective courses and using the prediction results to recommend courses to students.

REFERENCES

[1] H. Cen, K. Koedinger, and B. Junker, Learning factors analysis a general method for cognitive model evaluation and improvement, in International Conference on Intelligent Tutoring Systems, Springer, 2006, pp. 164175

[2] Ravi TejaYarlagadda, "THE RPA AND AI AUTOMATION", International Journal of Creative Research Thoughts (IJCRT), ISSN:2320-2882, Volume.6, Issue 3, pp.365-373, September 2018,
Available at: http://www.ijcrt.org/papers/IJCRT1133933.pdf

[3] IshaqAzhar Mohammed. (2013). Intelligent authentication for identity and access management: a review paper. International Journal of Managment, IT and Engineering (IJMIE), 3(1), 696–705. Retrieved from https://www.ijmra.us/project doc/IJMIE_JANUARY2013/IJMIEJan13Ishaq.pdf

[4] M. Feng, N. Heffernan, and K. Koedinger, Addressing the assessment challenge with an online system that tutors as it assesses, User Modelling and User-Adapted Interaction, vol. 19, no. 3, pp. 243266, 2009.

[5] J. Xu, K. H. Moon and M. van der Schaar, "A Machine Learning Approach for Tracking and Predicting Student Performance in Degree Programs," in IEEE Journal of Selected Topics in Signal Processing, vol. 11, no. 5, pp. 742-753, Aug. 2017, doi: 10.1109/JSTSP.2017.2692560.

[6] Altabrawee, Hussein & Ali, Osama &Qaisar, Samir. (2019). Predicting Students' Performance Using Machine Learning Techniques. JOURNAL OF UNIVERSITY OF BABYLON for pure and applied sciences. 27. 194-205. 10.29196/jubpas.v27i1.2108.

[7] Rupali, Bagal&Matele, Ramdas&Machhindra, Bhavika& Kailas, Wable& Mohammad, Sameer. (2019). Tracking and Predicting Student Performance Using Machine Learning. SSRN Electronic Journal. 6. 439-442.

[8] HashmiaHamsa, Simi Indiradevi, Jubilant J. Kizhakkethottam, Student Academic Performance Prediction Model Using Decision Tree and Fuzzy Genetic Algorithm, Procedia Technology, Volume 25, 2016, Pages 326-332, ISSN 2212-0173, https://doi.org/10.1016/j.protcy.2016.08.114.

[9] Mustafa Agaoglu, "Predicting Instructor Performance Using Data Mining Techniques in HigherEducation," IEEE Access , Volume: 4 ,2016.

[10] [2] TriptiMishra,Dr. DharminderKumar,Dr. SangeetaGupta,"Mining Students' Data for Performance Prediction," in fourth International Conference on Advanced Computing & Communication Technologies,2014.

[11] Keno C. Piad, Menchita Dumlao, Melvin A. Ballera, Shaneth C. Ambat," Predicting IT Employability

Using Data Mining Techniques," in third International Conference on Digital Information Processing, Data Mining, and Wireless Communications (DIPDMWC), 2016.

[12] BipinBihariJayasingh,"A Data Mining Approach to Inquiry Based Inductive Learning Practice In Engineering Education," in IEEE 6th International Conference on Advanced Computing, 2016.

[13] S. M. Merchán,"Analysis of Data Mining Techniques for Constructing a Predictive Model for Academic,"IEEE Latin America Transactions, vol. 14, no. 6, June 2016.

[14] KonstantinaChrysafiadi and Maria Virvou," Fuzzy Logic for adaptive instruction in an e-learning environment for computer programming," IEEE Transactions on Fuzzy Systems, Volume: 23, Issue: 1, Feb. 2015.

[15] M. Mayilvaganan,D. Kalpanadevi ," Comparison of Classification Techniques for predicting the performance of Students Academic Environment," in International Conference on Communication and Network Technologies (ICCNT), 2014.

[16] Crist´obal Romero," Educational Data Mining: A Review of the State of the Art," IEEE Transactions On Systems, Man, And Cybernetics—Part C: Applications And Reviews, Vol. 40, No. 6, November 2010.

[17] Kuzilek, J.; Hlosta, M.; Herrmannova, D.; Zdrahal, Z.; Wolff, A. OU Analyse: Analysing at-risk students at The Open University. Learn. Anal. Rev. 2015, 2015, 1–16

[18] Kovacic, Z. Early Prediction of Student Success: Mining Students' Enrolment Data. In Proceedings of the Informing Science and Information Technology Education Joint Conference, Cassino, Italy, 19–24 June 2010.

[19] Watson, C.; Li, F.W.; Godwin, J.L. Predicting performance in an introductory programming course by logging and analyzing student programming behavior. In Proceedings of the IEEE 13th International Conference on Advanced Learning Technologies, Beijing, China, 15–18 July 2013; pp. 319–323.

[20] The White House, "Making college affordable," https:// www.whitehouse.gov/issues/education/higher-education/ making-college-affordable, 2016.

A Deep Learning Approach for Effective Predictor of COVID-19 and ICU Requirements

*Irshad Ahmad Palla **Dr. Gurinder Kaur Sodhi,

*Research Scholar, Electronics and Communication Engineering, DeshBhagat University, Mandi Gobindgarh-147301

**Assistant Professor, Electronics and Communication Engineering, DeshBhagat University, Mandi Gobindgarh-147301

Corresponding author Email: eirshadahmad@gmail.com

Abstract

COVID-19 also known as Coronavirus is a global pandemic which has already affected millions of people around the world and is continuing to do so even now. Due to the high positivity rate among patients, the health and medical facilities faced lot of issues such as, lack of medical staff, beds, and intensive care units (ICUs). Therefore, it is important to identify and detect COVID-19 as earliest as possible. Over the years, a large number of methods were proposed for predicting covid-19, but the problem with those methods was that they were high complex and took huge time for training. In this paper, a bidirectional Long Short-Term memory (Bi-LSTM) based model is proposed that performs two classification tasks i.e. detection of COVID-19 and predicting the need for regular wards and ICU and semi-ICUs. The main objective of the proposed work is not only to decrease the complexity and training time of the model but also to enhance the accuracy of the system. For this, a publicly available dataset taken from the Kaggle.com is taken and pre-processing is applied to it, for making the data balanced and normalized. In addition to this, the complexity of the system is decreased by using the Eigenvector centrality feature selection (ECFS), in which only important and crucial features are selected. Furthermore, the effectiveness of the suggested system is enhanced by using the DL based classifier, named as, Bi-LSTM. Finally, the efficacy of the suggested approach is analyzed and compared with several traditional approaches in terms of various dependency factors like, accuracy, specificity, precision, recall and Fscore. The simulated outcomes determine the supremacy of the proposed approach over traditional approaches.

Keywords: COVID-19, ICU requirements, Disease detection, Predicting models, Artificial intelligence, etc.

Introduction

The year 2020 has proven to be a disaster for humanity. In the month of December 2019, a group of strange pneumonia cases was discovered in the Wuhan region of china, which subsequently spread to the rest of the globe. Corona-virus infection, also referred as COVID-19, is a respiration disorder that develops as a result of the acute respiratory disease (sars syndrome corona-virus-2 (abbreviated as SARS-COV-2) [1-2]. The disease spread rapidly over the globe, causing a pandemic and bring the medical systems of several countries to their knees. As India is the second populous country in the world and is one of the top five nations that got severely affected with COVID-19 [3]. The transfer of COVID-19 virus from one individual to another was quickly realized as an additional route of disease. According to recent research, it was discovered that COVID-19 transmission arises when humans are in close vicinity to each other and are infected with the virus. Folks and the general public with lower immune mechanism, like diabetes, heart disease, respiratory illness, leukemia, urinary, and liver failure are at a significantly greater chances for getting infected with COVID-19 infectious disease rather than kids that are less prone to infections or, if they do, have mild illness or even asymptomatic infection [4]. COVID-19 has a broad range of health symptoms, including asymptomatic and symptomatic types which include serious breathing issues requiring ventilatory and ICUs (intensive care unit) support, multi-organ failure and various other systemic manifestations such as septic shock, sepsis, and MODS or multiple organ dysfunction syndromes [5]. Illnesses with zero indications were also reported, although their origin is uncertain. Temperature, coughing, breathlessness, dry mouth, and other indications are very common in COVID19 positive patients. In addition to this, Pneumonia is also one of the serious symptoms found in COVID19 patients [6]. Headaches, sore throats, and rhinorrhea are some of the relatively prevalent symptoms. Other than this, various gastrointestinal indications like vomiting and diarrhea are also prevalent in some cases. The primary route of transmission are droplets that are disbursed during coughing or sneezing, and asymptomatic person-to-person encounters.

1.1 TRANSMISSION OF COVID-19

COVID-19 can be transmitted from one person to another in number of ways, some of the most common ways of transmitting the disease are explained below;

• The predominant route of transmission of SARS-CoV-2 is by close touch and potentially infectious

respiratory secretions from pre-symptomatic, asymptomatic and symptomatic patients.

- Fomite transmission of COVID-19 results from the contaminated surfaces and it was also observed that the COVID-19 virus is more effective on plastic and stainless-steel materials rather than copper and paper items. Also, infection tends to be significantly greater in ICUs, with COVID-19 virus accumulated on rugs, computer mice, waste bins, and hospital bed railings, and also in atmosphere up to 4 meters from victims, implying the fomite spread.

- According to epidemiologic information from many investigations, individuals with covid-19 illness had flu virus in their feces, suggesting fecal-oral spread.

- A meta-analysis of 936 infants from COVID-19-positive moms found that vertical transmission can occur, but only in a small percentage of cases.

This transmission can be reduced by following the guidelines issued by health experts, like, covering mouth while coughing and sneezing, washing hands regularly or using sanitizers, wear mask, maintain distance and avoid public gatherings. However, apart from enacting global health and disease prevention steps to avoid COVID-19 spread, the most important step is vaccination in societies. The vaccination results in producing neutralizing antibodies in the body against COVID-19 that empower the immunity in humans and reduce the probability of getting effected. India developed Covaxin, Spitnik V was developed by Russia and CoronaVAc was developed by China. These vaccines have already been used in different countries in order to prevent the transmission of COVID-19 further [7]. Due to the widespread number of persons getting infected of this deadly virus, several nations' hospitals and health-care systems are overburdened. As a result, early detection of these diseases is critical. In this regard, Machine learning and deep learning methods have been used for detecting the presence of COVID-19 diseases in humans by various researchers.

1.2 ROLE OF AI IN DETECTING COVID-19

Artificial Intelligence (AI), a rapidly emerging software tool in the field of medical image processing, has also aided in the fight against the new coronavirus [8-9] by quickly delivering superior detection rates while drastically decreasing or removing the need for human intervention. Over the years, different ML based algorithms which include, SVM, KNN, Naïve Bayes, Decision Tree etc. are used. However, the problem with the ML methods is that they cannot handle the large datasets generated by the medical field. Therefore, the researchers are moving towards the implementation of DL based methods. Deep learning (DL) approaches can handle enormous datasets, making them effective for detecting covid-19 that has a massive dataset. Numerous researchers trained their networks using publicly available datasets. The research in [10] found that by using the support vector machine (SVM) method, it was possible to correctly categorize COVID-19 individuals in 85 percent of instances. In this paper, an effective DL based approach is proposed that can efficiently predict Covid-19 as well as the need for regular wards and ICU/Semi-ICU requirements in the Hospitals.

2. LITERATURE REVIEW

In order to have a better understanding of the various ML and DL based COVID-19 detection methods, a significant number of existing techniques are studied and reviewed in this section. Over the years, various researchers proposed models by using different ML and DL classifiers in their work, some of them are discussed here; M. Sevi and İ. Aydin [11], proposed a covid-19 detection model in which data augmentation and classification though multi-class DL based methods was done. Similarly, S. Lafraxo and M. el Ansari, [12], suggested a DL based Covid-19 detection approach, named as, CoviNet in which adaptive median filter, histogram equalization was implemented. In addition to this, a CNN technique was used for classifying the disease. The results demonstrated that the suggested model achieved an accuracy of 98.62% and 95.77% for binary and multi-class classifications. Xiaoshuo Li et al. [13], suggested a DL based covid-19 approach that was based on the hybridization of stacked Generalization ensemble learning along with the VGG16 classifier so that a cascade classifier is developed. The suggested model was able to achieve a accuracy rate of 93.57% and a precision of 89.40%. L. Brunese et al. [14], proposed an effective mechanism for detecting the covid-19 in humans in which they utilized the supervised ML techniques and used a publicly available dataset containing a total of 85 chest x-ray of patients. Through extensive experimentation, the efficacy of the suggested scheme is depicted. SoumyajitPodder et al [15], proposed a Mask R-CNN based covid-19 detection approach that was trained on the publicly available dataset comprising of 668 x-ray frontal chest images. The results showcased that current approach was able to achieve an accuracy of 96.98% and a precision rate of 96.60% respectively. S. Sakib et al. [16], presented an appropriate and effective DL based Chest Radiograph Classification (DL-CRC) paradigm for separating

covid-19 positive samples from the normal or pneumonic cases with an accuracy of 93.94%. A. Channa et al. [17], proposed an extremely effective method for detecting covid-19 by utilizing DeepNets to assemble images of patients. The suggested approach was able to yield an accuracy of 91.67% while as, the survival ratio of accuracy was attained to be 100%. E. F. Ohata et al. [18], proposed a covid-19 detection approach that was based on various structures of CNN that was trained on ImageNet. Moreover, the authors incorporated the CNN with traditional ML algorithms like KNN, NB, RF, MLP and SVM for identifying the best performing model. Sethy, P.K et al. [19], utilized a DL based technique, named as, ResNet-50 along with the ML SVM for identifying and classifying the covid-19 in x-ray images. The results showcased that the proposed strategy was able to achieve a classification accuracy rate of 95.38%. S K T Hwa et al. [20], proposed an effective and improved approach in which local morphological contrast enhancement and Canny edge detection techniques were merged for classifying pneumonia and covid-19 diseases was proposed. From the literature survey conducted, it is observed that over the years a significant number of AI based approaches were proposed by any researchers in order to identify and detect the COVID-19 disease effectively and efficiently in humans. Majority of the researchers have implemented the ML based algorithms in their work for predicting covid-19. Although, the ML based methods were providing god classification and quantification results, they undergo through some issues that degrade their overall performance considerably. One of the main issues faced in such approaches is that ML algorithms are not able to handle the large datasets which directly affect the efficiency of the detection system. Therefore, in order to combat this issue, a DL method are recommended. In addition to this, the complexity in current ML based systems was very high because not much work has been done on reducing the dimensionality of the datasets. Another drawback in these systems was that most of the ML algorithms used by researchers either get stuck in the local minima or have high computational cost. Moreover, feature selection is considered as one of the important factors in enhancing the accuracy of the systems, and no focus was given to it. Inspired from these findings, a new and improved model will be proposed that will not only enhance the accuracy of the system but will also reduce its intricacy and computational cost.

3. PROPOSED WORK

In order to overcome the limitations of traditional Covid-19 detection models, a new and enhanced detection model that is based on DL method is proposed in this research. The suggested method works for two classification phases, the first phase is intended for identifying covid-19 in patients and appropriately the necessity for ICU/semi-ICU requirement if predicted in the second phase. The main objective of the proposed DL method is to reduce the complexity of the system as well as enhance the accuracy of the system. To accomplish this task, firstly a dataset is needed upon which more advanced techniques will be applied to generate the final covid-19 and ICU requirement predictions. However, the problem with the available datasets is that they are unbalanced in nature and contain a lot of empty cells, null and NAN values, which enhances the complexity of the system. Therefore, it becomes necessary to apply pre-processing and other advanced techniques to it so that its complexity id reduced and only informative and useful data is present in it. Here, we propose an efficient and effective method where, Eigenvector centrality Feature Selection (ECFS) technique is applied along with the advanced version of LSTM, named as, Bi-LSTM (bidirectional Long Short-Term Memory). The main motive for using the Bi-LSTM is that it can handle large datasets effectively and also it remembers the information of the pasta as well as the future. Along with this, the feature selection technique used helps in reducing the dimensionality of the dataset which in return reduces the overall complexity and increases the accuracy of the system.

The dataset taken from the Israelita Albert Einstein hospital at Sao Paulo, Brazil is utilized in the proposed work, whose details are mentioned briefly in the methodology section. As previously stated, the used database is unbalanced and has a large number of empty and NAN entries which must be refilled or eliminated. To do so, an average Insertion Approach is applied on the selected dataset where a particular column is selected and analyzed so that the missing values are filled by calculating the average values and empty cells are removed from it. This helps in reducing the dimensionality of the dataset. In addition, Eigenvector Centrality Feature Selection (ECFS) approach is also implemented on the given dataset to ensure that only important and significant characteristics are chosen from it, thereby reducing the complexity of the system. Moreover, in conventional detection approaches ML classifiers were used that were not able to handle big datasets, therefore, an DL based Bi-LSTM model is used in the proposed work that handles the large datasets quite effectively and enhances the classification accuracy results. The Bi-LSTM classifier is used in the proposed

work for predicting the covid-19 and the need for normal wards and ICU/Semi-ICU requirements. Another major reason for using the Bi-LSTM in the proposed work is that their prediction rate is quite higher than the regular LSTMs and doesn't get stuck in the local minima. Therefore, an effective and highly accurate model is developed by utilizing the ECFS and Bi-LSTM approaches that not only predicts the covid-19 but also predicts the requirement for ICUs. The step by step process of the proposed model is mentioned in the following section of this paper.

3.1 METHODOLOGY

The proposed bi-LSTM model undergoes through a series of steps like, data collection, pre-processing, feature selection and classification. At each stage the data is modified and refined so that effective results can be obtained at the end. The flowchart of the proposed system is shown in figure 1.2 and is explained below;

• Data acquisition: The first step opted in proposed model is of data collection. In this stage, the necessary information is collected from the available datasets. The dataset used in the proposed work is taken from the Kaggle.com and is utilized for performing two classification tasks (predicting covid-19 and ICU requirement) in the proposed work.

DATASET USED

The dataset used in this research is taken from the Kaggle.com, and it contains information about various patients at Israelita Albert Einstein hospital located in Sao Paulo, Brazil. A sum of 5644 patient samples are included in the collection, together with their 111 characteristics. The difficulty with the existing database, unfortunately, is that it is unbalanced because 90% of the information is for negative instances. Furthermore, the collection contains empty and missing NAN values, which must be eliminated.

•Data pre-processing: once the data is collected. The next step to eb followed is pre-processing. For this, an average insertion approach is used which removes the empty cells from the dataset and fills the missing and NAN values by calculating their average sum. The data is processed and normalized so that only informative and useful information is present in the system.

•Feature selection: In order to reduce the dimensionality and complexity of the dataset, an Eigenvector Centrality Feature Selection (ECFS) is applied to the processed data. The ECFS selects only those features that are important and crucial for predicting the cobid-19 and ICU requirements in proposed work. The two important pa-

rameters of ECFS are alpha and mean threshold whose values are given in table 1.1. The final dataset obtained is then divided into two categories for training and testing.

TABLE 1.1 DIFFERENT PARAMETERS OF PROPOSED MODEL

Sr. No.	Factors	Values
1	Alpha	0.85
2	Mean Threshold	0.7*W1 and W2
3	HiddenUnits	25
4	InputLayer (Type)	Sequential
5	Network Type	BILSTM
6	Epochs	100
7	InitialLearning Rate	0.045
8	GradientThreshold	0.02
9	BatchSize	10
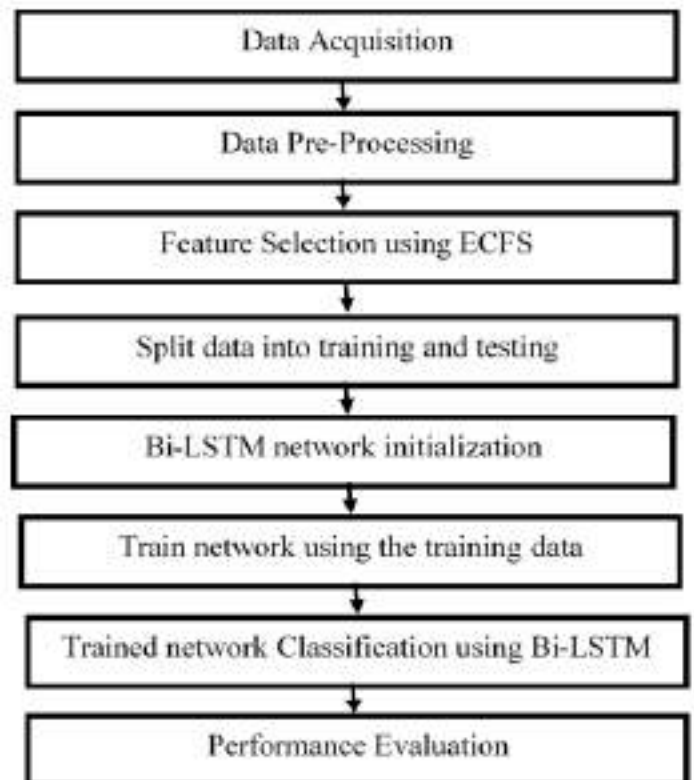
Fig.1.1Proposed Bi-LSTM model

• Network initialization: After this, the proposed Bi-LSTM network initialized wherein, a number of Bi-LSTM parameters are defined. Some of the parameters used in the proposed work are input layers, network type, hidden units and max epochs, whose values are mentioned in table 1.1.

• Training of model: Once the model is initialized, it starts training itself by passing the training data to it. The model trains itself as per this data.

• Classification: In the last phase, the performance of

the proposed model is analyzed by passing the testing data to it. The proposed Bi-LSTM model predicts the covid-19 and ICU requirements as per the training provided to it. On the basis of this classification, its performance is evaluated in terms of number of parameters like accuracy, precision, recall and Fscore that are mentioned briefly in next section.

4. RESULTS AND DISCUSSION

The results of the suggested Bi-LSTM framework are analyzed and explained in this section. Because the findings were acquired for two classification tasks, they were also analyzed separately for the two categorization groups. The suggested Bi-LSTM model is simulated using the MATLAB software. the results were obtained in terms of various dependency factors and is alter on compared with traditional methods to evaluate its supremacy.

4.1 PERFORMANCE EVALUATION FOR COVID-19 PREDICTION

The performance of the proposed Bi-LSTM model is analyzed and compared with various state of art approaches in terms of accuracy, precision, specificity, recall and Fscore. In addition to this, the learning rate of the proposed Bi-LSTM model is 0.0450 throughout all iterations. Moreover, the min. batch loss is also reduced from 0.6920 to 0.1916 and training accuracy is enhanced from 93.896 to 95.250% in proposed Bi-LSTM model. To validate the effectiveness of the suggested approach, its performance is compared with the traditional RF, MLP, LGBM, NB, ETC, S1-RF-XGB-LR, S2-NB-LGBM-LR, V1-ET-RF-LGBM-Hard, V1-RF-LR-SVM-Soft, V2-MLP-NB-LGBM- hard and V2-MLP-NB-LGBM-Soft approaches in terms of their accuracy value. The graph obtained for the same is represented in figure 1.2. The x-axis and the y-axis of the given graph calibrate the different approaches and tehri accuracy values respectively. After examining the graph closely, it is observed that the value of accuracy was lowest achieved in conventional MLP model with 90.55%, followed up by NB with 91.95, followed up by V2-MLP-NB-LGBM-hard, V2-MLP-NB-LGBM-Soft, V1-RF-LR-SVM-Soft, ETC, LGBM and RF with 91.95%, 91.65%, 91.56%, 91.56%, 91.29% and 91.1% respectively. while, as the accuracy values is considerably enhanced by other traditional models like, S1-RF-XGB-LR, S2-NB-LGBM-LR and V1-ET-RF-LGBM-Hard whose accuracy rate was mounted at 94.39%, 92.78% and 92.47% respectively. On the other hand, when the value of accuracy is analyzed for the proposed Bi-LSTM model, it is significantly higher than all the traditional models with value of 95.26129%. This means that the proposed Bi-LSTM model is able to predict Covid-19 and ICU requirement more accurately and effectively, when compared with traditional ML based methods.
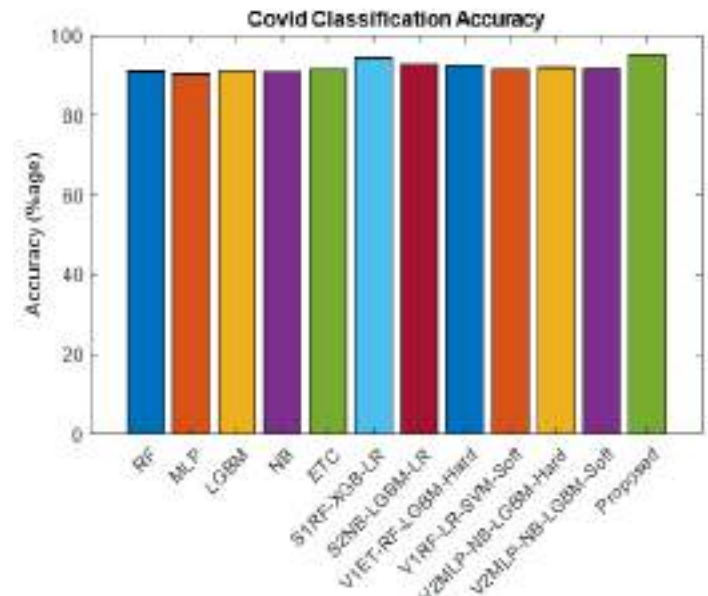
Fig.1.2Comparison graph for Accuracy (COVID)

In addition to the accuracy, the performance of the suggested Bi-LSTM model is evaluated in terms of precision, specificity, recall and Fscore as well. The values attained for each parameter in proposed Bi-LSTM model and traditional models is given in table 1.2.

After analyzing the above table (see table 1.2), it is observed that the proposed Bi-LSTM approach is outperforming the conventional RF, MLP, LGBM, NB, ETC, S1-RF-XGB-LR, S2-NB-LGBM-LR, V1-ET-RF-LGBM-Hard, V1-RF-LR-SVM-Soft, V2-MLP-NB-LGBM- hard and V2-MLP-NB-LGBM-Soft methods in terms of precision, recall, specificity and FScore as well. The precision value in proposed Bi-LSTM model is 100% while as, it was only 90% in RF, MLP, ETC, V1-ET-RF-LGBM-Hard, V1-RF-LR-SVM-Soft and V2-MLP-NB-LGBM-Soft models and 91% in LGBM, NB, S1-RF-XGB-LR, S2-NB-LGBM-LR and V2-MLP-NB-LGBM- hard. Similarly, the values achieved for Recall, specificity and Fscore in proposed Bi-LSTM model came out to be 95.26129%, 95.26129318% and 97.57315 respectively. From the given graph and table, it is proved that the proposed Bi-LSTM scheme for predicting covid-19 is more effective and efficient.

4.2 PERFORMANCE EVALUATION FOR ICU REQUIREMENT

As previously done for the first classification phase, the performance of the proposed Bi-LSTM model is also analyzed and compared with several conventional methods in terms of their accuracy. Figure 1.3 represents the comparison graph of the proposed Bi-LSTM model with traditional models.
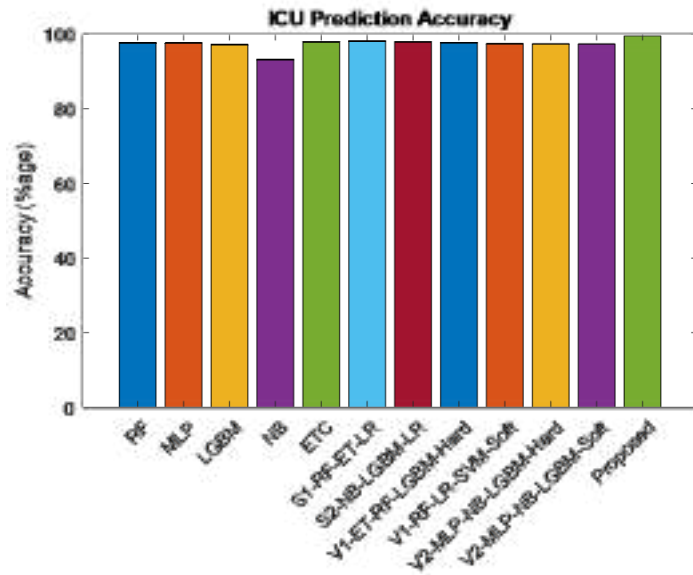
Fig.1.3Comparison graph for accuracy (ICU)

Figure 1.3 illustrates the comparison graph of the suggested Bi-LSTM model along with the traditional RF, MLP, LGBM, NB, ETC, S1-RF-XGB-LR, S2-NB-LGBM-LR, V1-ET-RF-LGBM-Hard, V1-RF-LR-SVM-Soft, V2-MLP-NB-LGBM- hard and V2-MLP-NB-LGBM-Soft methods in terms of their accuracy. The accuracy reached in standard RF, MLP, LGBM, NB, and ETC approaches was only 97.66%, 97.59%, 97.18%, 93.16%, and 97.94%, correspondingly, as shown in the graph. Meanwhile, hybrid approaches such as S1-RF-ET-LR, S2-NB-LGBM-LR, V1-ET-RF-LGBM-Hard, V1-RF-LR-SVM-Soft, V2-MLP-NB-LGBM-Hard, and V2-MLP-NB-LGBM-Soft enhance accuracy by a little percentage with 98.13%, 97.95%, 97.71%, 97.41% and 97.39% respectively. The accuracy achieved in the suggested Bi-LSTM framework, on the other hand, is about 99.555%, which is significantly higher than traditional models, therefore, demonstrating its usefulness and efficacy. Moreover, the efficiency of the proposed Bi-LSTM model is depicted in terms of other performance factors like Precision, recall and fscore, whose exact values are given in table 1.3.

TABLE 1.2
Performance Values attained by traditional and Proposed model (COVID Prediction)

Techniques	Precision	Recall	Specificity	Fscore	Accuracy
RF	90	89	95.9677	89	91.1
MLP	90	91	86.2903	91	90.55
LGBM	91	90	93.5484	90	91.29
NB	91	90	95.9677	90	90.95
ETC	90	89	95.9677	89	91.56
S1-RF-ET-LR	91	92	95.1613	91	94.39
S2-NB-LGBM-LR	91	91	95.1613	91	92.78
V1-ET-RF-LGBM-Hard	90	90	95.9677	89	92.47
V1-RF-LR-SVM-Soft	90	89	94.3548	89	91.56
V2-MLP-NB-LGBM-Hard	91	90	95.1613	90	91.95
V2-MLP-NB-LGBM-Soft	90	90	94.3548	90	91.65
Proposed	100	95.26129	95.26129318	97.57315	95.26129

TABLE 1.3
Comparison table for predicting ICU requirements

Techniques	Precision	Recall	Fscore	Accuracy
RF	98	98	98	97.66
MLP	97	98	97	97.59
LGBM	98	98	98	97.18
NB	98	94	96	93.16
ETC	98	97	98	97.94
S1-RF-ET-LR	99	99	98	98.13
S2-NB-LGBM-LR	98	99	98	97.95
V1-ET-RF-LGBM-Hard	98	98	98	97.71
V1-RF-LR-SVM-Soft	98	97	96	97.41
V2-MLP-NB-LGBM-Hard	98	98	98	97.39
V2-MLP-NB-LGBM-Soft	97	98	97	97.34
Proposed	100	99.11426	99.11426	99.55516

After analyzing the table 1.3, it is observed that the value of precision is highest among all in proposed Bi-LSTM model with 100% results. While as, the precision values were lowest among all in traditional MLP and V2-MLP-NB-LGBM-Soft models with just 97% value, 98% in conventional RF, LGBM, NB, ETC, S2-NB-LGBM-LR, V1-ET-RF-LGBM-Hard, V1-RF-LR-SVM-Soft and V2-MLP-NB-LGBM-Hard models and 99% in standard S1-RF-ET-LR model. In addition to this, the performance was evaluated in terms of recall and Fscore as well, whose values were mounted at 99.11426% for each. These values are enough to prove the superiority of the proposed Bi-LSTM model over traditional models in all factors including, accuracy, precision, recall and Fscore.

5. CONCLUSIONS

In this paper, an effective and efficient Bi-LSTM based approach is proposed for predicting the covid-19 and ICU requirement in hospitals. Through extensive simulations in the MATLAB software, the efficacy and accuracy of the proposed Bi-LSTM system is analyzed in terms of accuracy, precision, recall and Fscore. After examining the results, it is observed that the value of accuracy in proposed Bi-LSTM model for predicting covid-19 were mounted at 95.26129% while as for predicting the ICU requirement the accuracy value was 99.55516% respectively. However, when the accuracy values were analyzed for traditional models, it came out to be 91.1% in RF, 90.55% in MLP, 91.29% in LGBM, 90.95% in NB, 91.56% in ETC, 94.39% in S1-RF-XGB-LR, 92.78% in S2-NB-LGBM-LR, 92.47% and 91.56%

in V1-ET-RF-LGBM-Hard and V1RF-LR-SVM-Soft, 91.95% in V2-MLP-NB-LGBM- hard and 91.65% in V2-MLP-NB-LGBM-Soft approaches for predicting covid-19. On the other hand, the value of accuracy for predicting ICU requirement in traditional approaches were 97.66%, 97.59%, 97.18%, 93.16%, and 97.94% in RF, MLP, LGBM, NB, and ETC and 98.13%, 97.95%, 97.71%, 97.41% and 97.39% in S1-RF-ET-LR, S2-NB-LGBM-LR, V1-ET-RF-LGBM-Hard, V1-RF-LR-SVM-Soft, V2-MLP-NB-LGBM-Hard, and V2-MLP-NB-LGBM-Soft approaches. In addition to this, the values were attained in terms of precision, recall and fscore whose values came out to be 100%, 95.26129% and 97.57315 for predicting covid-19 and 100%, 99.11426% and 99.11426% for predicting ICU requirements. From these results, it is concluded that the proposed Bi-LSTM model for predicting Covid-19 and ICU requirements is more accurate and precise with low complexity

REFERENCES

[1] S. Shaikh, J. Gala, A. Jain, S. Advani, S. Jaidhara and M. Roja Edinburgh, "Analysis and Prediction of COVID-19 using Regression Models and Time Series Forecasting," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2021, pp. 989-995, doi: 10.1109/Confluence51648.2021.9377137.

[2] R. Punia, L. Kumar, M. Mujahid and R. Rohilla, "Computer Vision and Radiology for COVID-19 Detection," 2020 International Conference for Emerging Technology (INCET), 2020, pp. 1-5.

[3] M. Rohini, K. R. Naveena, G. Jothipriya, S. Kameshwaran and M. Jagadeeswari, "A Comparative Approach To Predict Corona Virus Using Machine Learning," 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), 2021, pp. 331-337, doi: 10.1109/ICAIS50930.2021.9395827.

[4] Wikipedia contributors. "COVID-19." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 23 Sep. 2021. Web. 25 Sep. 2021.

[5] Lupia, T.; Scabini, S.; Mornese Pinna, S.; Di Perri, G.; De Rosa, F.G.; Corcione, S. 2019 novel coronavirus (2019-nCoV) outbreak: A new challenge. J. Glob. Antimicrob. Resist. 2020, 21, 22-27.

[6] Yang, Y.; Peng, F.; Wang, R.; Guan, K.; Jiang, T.; Xu, G.; Sun, J.; Chang, C. The deadly coronaviruses: The 2003 SARS pandemic and the 2020 novel coronavirus epidemic in China. J. Autoimmun. 2020.

[7] Shinde V, Bhikha S, Hoosain Z, Archary M, Bhorat Q, Fairlie L, Lalloo U, Masilela MSL, Moodley D, Hanley S, Fouche L, Louw C, Tameris M, Singh N, Goga A, Dheda K, Grobbelaar C, Kruger G, Carrim-Ganey N, Baillie V, de Oliveira T, Lombard Koen A, Lombaard JJ, Mngqibisa R, Bhorat AE, Benadé G, Lalloo N, Pitsi A, Vollgraaff PL, Luabeya A, Esmail A, Petrick FG, Oommen-Jose A, Foulkes S, Ahmed K, Thombrayil A, Fries L, Cloney-Clark S, Zhu M, Bennett C, Albert G, Faust E, Plested JS, Robertson A, Neal S, Cho I, Glenn GM, Dubovsky F, Madhi SA., 2019nCoV-501 Study Group. Efficacy of NVX-CoV2373 Covid-19 Vaccine against the B.1.351 Variant. N Engl J Med. 2021 May 20;384(20):1899-1909.

[8] F. Shi et al., ''Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for COVID-19,'' IEEE Rev. Biomed. Eng., vol. 14, pp. 4-15, Jan. 2021.

[9] R. Vaishya, M. Javaid, I. H. Khan, and A. Haleem, ''Artificial intelligence (AI) applications for COVID-19 pandemic,'' Diabetes Metabolic Syndrome, Clin. Res. Rev., vol. 14, no. 4, pp. 337-339, Jul. 2020

[10] A. F. M. Batista, J. L. Miraglia, T. H. R. Donato, A. D. P. C. Filho, "COVID-19 diagnosis prediction in emergency care patients: a machine learning approach", medRxiv, 2020.

[11] M. Seviand ?. AYDIN, "COVID-19 Detection Using Deep Learning Methods," 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI), 2020, pp. 1-6, doi: 10.1109/ICDABI51230.2020.9325626.

[12] S. Lafraxo and M. el Ansari, "CoviNet: Automated COVID-19 Detection from X-rays using Deep Learning Techniques," 2020 6th IEEE Congress on Information Science and Technology (CiSt), 2020, pp. 489-494, doi: 10.1109/CiSt49399.2021.9357250.

[13] Xiaoshuo Li, Wenjun Tan, Pan Liu, Qinghua Zhou, Jinzhu Yang, "Classification of COVID-19 Chest CT Images Based on Ensemble Deep Learning", Journal of Healthcare Engineering, vol. 2021, Article ID 5528441, 7 pages, 2021.

[14] Luca Brunese, Fabio Martinelli, Francesco Mercaldo, AntonellaSantone, Machine learning for coronavirus covid-19 detection from chest x-rays, Procedia Computer Science, Volume 176, 2020, Pages 2212-2221.

[15] SoumyajitPodder, SomnathBhattacharjee, Arijit Roy. An efficient method of detection of COVID-19 using Mask R-CNN on chest X-Ray images[J]. AIMS Biophysics, 2021, 8(3): 281-290.

[16] S. Sakib, T. Tazrin, M. M. Fouda, Z. M. Fadlullah and M. Guizani, "DL-CRC: Deep Learning-Based Chest Radiograph Classification for COVID-19 Detection: A Novel Approach," in IEEE Access, vol. 8, pp. 171575-171589, 2020

[17] A. Channa, N. Popescu and N. u. R. Malik, "Robust Technique to Detect COVID-19 using Chest X-ray Images," 2020 International Conference on e-Health and Bioengineering (EHB), 2020, pp. 1-6, doi: 10.1109/EHB50910.2020.9280216.

[18] E. F. Ohata et al., "Automatic detection of COVID-19 infection using chest X-ray images through transfer learning," in IEEE/CAA Journal of AutomaticaSinica, vol. 8, no. 1, pp. 239-248, January 2021, doi: 10.1109/JAS.2020.1003393.

[19] Sethy, P.K.; Behera, S.K. Detection of Coronavirus Disease (COVID-19) Based on Deep Features. Preprints 2020, 2020030300 (doi: 10.20944/preprints202003.0300.v1).

[20] S K T Hwa1, A Bade1 and M H Ahmad Hijazi, Enhanced Canny edge detection for Covid-19 and pneumonia X-Ray images, IOP Conference Series: Materials Science and Engineering, Volume 979, International Conference on Virtual and Mixed Reality Interfaces 2020, 16-17 November 2020 16-17

Review on Effect of Replacement Lime and Sodium Silicate on The Stabilization OF Soil

*SaddamHussain **ShivaniThaku
*M. Tech Scholar,
**Assistant Professor, Civil Engineering Department,
Desh Bhagat University MandiGobindgarh-147301,Punjab

Abstract

Soil is the basic foundation for any civil engineering structures. It is required to bear the loads without failure. In some places, soil may be weak which cannot resist the oncoming loads. In such case, soil stabilization is needed. Soil stabilization can be explained as the alteration of the soil Properties by chemical or physical means in order to enhance the engineering quality of the soi. This paper deals with the complete analysis of the improvement of soil properties and its stabilization using lime and sodium silicate. The soil is treated with lime and sodium silicate with different percentages. For this, the percentage of lime was varied as 2.5%, 4.5% and 6.5% (by weight) and sodium silicate percentage was as 1.5%, 2.5%, and 3.5% (by weight). The maximum dry density decreased from 1.70 gm/cc to 1.63 gm/cc. 1.57 gm/cc and 3.5% (by weight). The maximum dry density decreased from 1.70gm/cc to 1.55gm/cc, 1.48gm/cc and 1.50gm/cc with addition of 1.5%, 2.5% and 3.5% sodium silicate to virgin soil. The OMC of soil increased from 16.50% to 18.88%,20.40% and 22.44% with addition of 2.5%, 4.5% and 6.5% lime (by weight). The OMC of soil and sodium silicate from 16.50% to 17.10%, 17.70% and 17.90% with addition 1.5%, 2.5% and 3.5%Sodium silicate (by weight). The unconfined compressive strength increases from 4.10 kg/cm2 to 5.20 kg/cm2 at 1.5% sodium silicate, the unconfined compressive strength increases from 4.10 kg/cm2 to 5.20kg/cm2 at 1.5% sodium silicate (by weight) and to 7.50kg/cm2 at 2.5% sodium silicate (by weight) addition. The value tends to all at 3.5% sodium silicate addition. For optimum mix i.e., soil + 2.5% sodium silicate, the unconfined compressive strength came out to be 9.73 kg/cm2 for uncured sample and 10.27kg/cm2 for cured samples. The C.B.R. Value for unsoaked and soaked virgin soil was found to be 6.2% and 3.2% respectively. When 4.5% lime was added to soil the C.B.R> Value for unsoaked and soaked condition increased to 15.6% and 5.4% respectively. When 2.5% sodium silicate was added to this soil + 4.5% mix, C.B.R. (unsoaked and soaked condition) was observed to be 18.7% and 8.6% respectively.

Key words: Soil -lime Stabilization, Soil Sodium Silicate Stabilization, compressive strength, C.B.R. (unsoaked and soaked condition).

Introduction

In this study of stabilization or stabilization of soil is one of the best or easily available method to improve the properties of soil. There are stabilizers like fly ash, rice-husk ash, lime, sodium silicate, jute, gypsum tyres etc. are used to strength the properties of soil. Mainly used stabilizers are cement and fly these two are costly ones that means by using these construction costs get high. So, in the present study, We add lime and sodium silicate as stabilizer to increase the engineering properties of clayey soil.The purpose of this study is to increase the strength of the of the clayey soil by making soil-lime sodium silicate mixture.

For this study we are going to performed test on the soil sample and determines its properties, and nature of soil.Also the properties and characteristic of stabilizing material is firstly determined. So that we can easily use the quantity and quality of stabilizer. After this test are to be performed to obtain the optimum moister content (OMC) and maximum dry density (MDD)of soil and the unconfined compressive strength of soi sample by adding constant content of lime as 2.5, 4.5, 6.5% and sodium silicate as 1.5, 2.5, 3.5%. The test is performed for obtaining optimum moisture content (OMC) and maximum dry density (MDD) of soil and the unconfined compressive strength of soil samples. On several number of samples, the tests are performed one by one by parent soil with lime and Sodium silicate to sample to provide the bearing capacity of soil little bit but as the percentage of lime and sodium silicate is increased 2.5 and1.5% the weight of soil. The mixture of clay soil and lime, sodium silicate as the dosage 2.5% and 1.5% and conduct the test standard proctor test and unconfined compression test. Due to this test, we get the new bearing capacity, strength, durability of soil. As similarly the percentage of lime and sodium silicate increase in 4.5%, 6.5 and 2.5%, 3.5%. so that we get the bearing capacity of soil to improve the shear strength and durability hardness and increase the life of structure. Because the lime and sodium silicate having good engineering properties to hold the soil particles and bonding the particle of soil together so that the permeability of soil is decreased that's why the bearing capacity and stability of soil is increased. On several number of samples, the tests are performed one by one by parent soil with lime and Sodium silicate to sample to provide the

bearing capacity of soil little bit but as the percentage of lime and sodium silicate is increased 2.5 and1.5% the weight of soil. The mixture of clay soil and lime, sodium silicate as the dosage 2.5% and 1.5% and conduct the test standard proctor test and unconfined compression test. Due to this test, we get the new bearing capacity, strength, durability of soil. As similarly the percentage of lime and sodium silicate increase in 4.5%, 6.5 and 2.5%, 3.5%. so that we get the bearing capacity of soil to improve the shear strength and durability hardness and increase the life of structure. Because the lime and sodium silicate having good engineering properties to hold the soil particles and bonding the particle of soil together so that the permeability of soil is decreased that's why the bearing capacity and stability of soil is increased.

A) Soil -lime Stabilization

Lime stabilization improves the strength, stiffness and durability of fine-grained material. In addition, lime is sometimes used to improve the properties of the fine-grained fraction of granular, under concrete foundations, on embankment and canal linings. Adding lime to soil product of maximum density under higher optimum moisture content than in the untreated Soil, Moreover, Lime produces a decrease in plasticity index.

Which produce a cemented material that increase in strengths with lime. Lime -treated soils, in general, have greater strength and a higher module of elasticity than untreated soils.

Recommended percentages of lime for soil stabilization vary 2 to 10 percent. For coarse soils such as clay gravels, sandy soils with less than 50% silt- clay fraction, the percent of lime varies from 2 to 5, whereas for soils with more than 50 percent silt clay fraction, the percent of lime lies between 5 and 10.

- Lime is very cheap and easily available in India.
- Lime treatment is especially effective for heavy clays or for silt and clayey granular soils, which are likely to lose strength because of high water affinity in their silt clay fraction.
- Where the lime treatment is aimed at "modifying" the soil properties, (by reducing plasticity, improving workability, increasing grain size etc.) the criteria for mechanically stabilized mixture are applied.

B) Soil Sodium Silicate Stabilization

Sodium silicate is easily is easily available and cheap material. Soil stabilized with sodium silicate possess high strength than virgin soil. It is very effective in re-ducing the swelling potential and the swelling pressure of clayey soils. Sodium silicate reacts with soil particle to from colloid which polymerizes further to from a gel that binds soil or sediment particles together and fills voids.

Besides, sodium silicate is a white powder or colourless solution that is readily soluble in water, it has also been considered for use as a peptizing agent to improve the mix ability of the in situ and in this way increase the homogeneity and strength of stabilized soil.

- Sodium silicate is easily available and cheap material.
- It increases the bearing strength of the soil.
- It decreases the swelling potential and swelling pressure.

2. LITERATURE REVIEW

(Mitchell et al. (1961)) Researchers have illustrated the impact of lime addition on the strength of clay soils depends on several factors. These include, soil type, curing time and method, moisture, soil unit weight and time elapsed between mixing and compaction.

(Winterkorn 19705) Based on us experience lime requirements for stabilized base and sub base course of pavement are as follows:

1. Five to ten percent of lime by weight of oil, for heavy clay soils to serve as bases, or one to three percent in sub-bases.
2. Two to four percent of lime for clay-gravel materials to serve as bases in India, the amount of lime needed for stabilization is recommended to be decided by the criteria of ph. value, CBR and an unconfined compressive strength (IRC 1983).

Ai-Rawi et al. (1981) comparatively fewer studies have focused on the impact of lime stabilization on the compressibility as much as by the shear strength. In addition, sodium silicate as a glass material can be used efficiently in soil stabilization because it is cheap and available. This work show that the sodium silicate can improve the geotechnical properties by increasing strength of soil and reducing its volume change.

The aim of this work is to investigate the effect of lime and sodium silicate powder on different geotechnical engineering properties of soil. In order to achieve, an experimental program was designed into three steps.

Firstly, to estimate the optimum amount of lime and sodium silicate that required to improve the soil properties. This optimum value (fixation point) can be estimated by short term reaction using different indepen-

dent method like plasticity index.

Dal Hunter et al. (1988) for effective stabilization ,a soil must have not less than 15% fraction passing a 425-micron sieve and its plasticity index should be at 10. The organic content should not be more than 20% and the sulphate content should be more than 0.2% impurities such as sulphates of calcium and sodium induces considerable have in lime soil. Hence it is desirable that the hydrated lime should be free of such impurities. For proper mixing the soil should pulverized to about 25mm and smaller size, about 50 to 60% passing a 4.75. pulverization and mixing of soil and lime in the field can be done manually or mechanically. The mix is contacted at OMC making allowance for moister loses. It is recommended that all compaction should be completed within four after mixing of soil, lime, and water the based should be cured for 7 to 28 days under moist conditions.

Sivapullaiah et al. (1996) has reported the effect of fly ash, on the index properties namely liquid and plastic of Indian black cotton soil. The effect of lime when added have also and studied. They used black cotton soil collected from Davanagere, Karnataka state in India and two and two fly ashes, such as Nayeli fly ash (NFA) and Vijayawada fly ash '(VFA)' collected from Nayeli thermal power plant, Tamil Nadu and Vijayawada thermal power plant, Andhra Pradesh, respectively. From the experiment they concluded that the liquid limit, plastic limit a plasticity index properties of black cotton soil were significantly altered by the addition of fly ash. The plastic limit of black cotton soil increased with the increase of fly ash content. The addition of 50% VFA to the black cotton soil increased the plastic limit to 62% whereas the same percentage of NFA increased the plastic limit to 116%.

Cokca et al. (2001) studied the effect of fly ash on the swelling characteristics of expansive soil. In his study , high-calcium and low-calcium class C fly ashes from the soma and Tuncbilek thermal power plants, respectively, in Turkey< were used for stabilization of expansive soil. Lime and cement were mixed to the expansive soil at 0 – 8% to obtained optimum values. Soma fly ash and Tuncbilek fly ash were mixed to the expansive soil at 0-25%. The "free swell" testing method was used to determine the swelling potential of the specimens (ASTM D4546-90).

Goswami et al. (2007) in their study of the leaching characteristics of residual lateritic soil stabilized with fly ash and lime for geotechnical applications. With the help of singlebatch leaching test and col-umn leaching test for different soil-fly ash-lime mixes found that the high PH induced by lime treatment of the mixes helps in keeping most of the metals within the stabilized soil matrix.

Sharma et al. (2008) studied the strength characteristics (UCS) of stabilizes expansive clay mixed with lime, calcium chloride, and rice-husk ash (RHA). The changes of unconfined compressive strength by the addition of addition of additives were reported. It was observed that the UCS of expansive clay improvement was observed by the addition of lime up to 5% or up to 1% calcium chloride. A maximum improvement was observed at 4% lime and 1% calcium chloride, respectively. An RHA content of 12% was found to be the optimum with regard to UCS in the presence of either lime (4%) or calcium chloride (1%).

Solanki et al. (2009) Observed the behaviour of UCS of silty clay stabilized with lime, class C fly ash, and cement kiln dust. They observed an increase in unconfined compressive strength with increase in the amount of the stabilizing agent. The increase in UCS value was found to vary with the type of stabilizing agents. The CKD-stabilized specimens exhibited a higher UCS value than the corresponding value of lime and CFA stabilized specimens.

Sahoo et al. (2010) Studied the compaction characteristics of soil mixed fly ash and lime using fly ash (class c category) from Penki thermal power station , Kanpur and soil from mesa road, Allahabad ,and observed that the optimum moisture content (OMC) and maximum dry unit weight (Yd) of the stabilized samples were increased and decreased gradually with the increase of fly ash and lime content respectively.

3) OBJECTIVE OF WORK

- The objective of this work is to study the suitability of sodium silicate for expansive soil stabilization, unsuitable for pavement sub grades, by increasing their bearing capacity and decreasing the swelling pressure and heave.
- In addition to this, it aims to investigate suitability of the chemical to decrease pavement thickness by increasing the bearing capacity of substandard sub base materials.
- Lime is well known sub base and sub grade stabilizers. This study aims to assess the potential improvement in strength and decrease in plasticity of soils though the application of the chemical additives at various curing durations.

4) CONCLUSIONS

- Lime acts immediately and improves various of soil such as resistance to shrinkage during moist conditions, reduction in plasticity, increase in CBR value and Subsequent increase in the compression resistance with the increase in time.
- Lime is used as an excellent soil stabilizing material for highly active soils which undergo through frequent expansion and shrinkage.
- The reaction is very quick and stabilization of soil starts within few hours.
- The maximum dry density decreased by the addition of lime and sodium silicate to the soil.
- The optimum moisture content increased by the addition of lime and sodium silicate to the soil.
- The unconfined compression strength increased by 137% for uncured sample of soil + 4.5% lime + 2.5% sodium silicate mix as compared to virgin soil. The increased in unconfined compression strength for cured sample of soil + 4.5% lime + 2.5% sodium silicate was 152% as compared to virgin soil.
- The California bearing ratio of soil + lime 4.5% + sodium silicate 2.5% increased by 211% (soaked) at 2.5 mm penetration, and 115.9% (unsoaked) and 66.9% (soaked) at 5.0 mm penetration.
- The California bearing ratio of soil + lime 4.5% + sodium silicate 2.5% increased by 211% (unsoaked) and 186% (soaked) at 2.5 mm penetration, and 172.4% (unsoaked) and 161.2% at 5.0 mm penetration.
- Hence, there is an overall gain strength parameter of clayey soil due to the addition of lime and sodium silicate.

5) REFERENCE

[1] AI.Rawi (1981), "Behaviour of Granular Soils Reinforced with Discrete Randomly Oriented Inclusions," Geotextiles and Geomembranes , Geotextile Society, Elsevier Science Publishers,England,10(4),pp.319- 333

[2] Bell, F.G. (1988),"Randomly Distributed Fibre Reinforcement Soil ."The state of the art, "Indian Geotechnical Journals, Vol. 70.

[3] Cokca (2001), "Jute Geotextile : An Ideal Eco-friendly ,Cost-Effective Geo-textile."

[4] Dall Hunter (1988), "Lime induced heave in sulphate bearing clay soils," Journal of Geotechnical Engg.Div.ASCE, Vol. 114,No. 2

[5] Goswami (2007), "Standardization of Geosynthetics materials and test methods in India an approach,"

[6] Mitcher (1981), "Performance studies on California Bearing ratio Value Geosynthetics."

[7] Anders land and Khattak, (1979), "Shear strength of KaolinitieFibre Soil Mixtures ," Proceedings, International Conference on soil reinforcement , Vol. I, Paris .France,pp.11-16

[8] Arora ,K.R., "Soil Mechanics and Foundation Engineering,"

[9] Bansal ,R.K . and Jain A. (2000), "Advance in Ground Improvement Techniques,"

AN ANALYTICAL STUDY FOR RECOGNISING SENTIMENTUSING MACHINE LEARNING MODELS

*Shanun Randev   **Dr. Deepak Gupta
*Dept. of electronics and communication and engineering,
Maharaja Agrasen institute of Technology New Delhi
*Dept. of Computer Science, Maharaja Agrasen institute of Technology New Delhi
Shanunrandev21@gmail.comdrdeepakgupta.cse@gmail.com

Abstract
Sentiment analysis (SA) is a branch of opinion mining that focuses on obtaining people's thoughts and feelings about a specific subject from systematic, semi-structured, or unorganized text data. In this paper, the efficacy of four ML classifiers i.e. Decision Tree (DT), Random Forest (RF), Logistic Regression (LR) and Multinomial NB is analyzed on IMDB dataset. The main objective of the proposed work is to analyze which classifier shows best results on the given dataset. To achieve this objective, necessary movie review or comment data is taken from IMDB dataset that is available on Kaggle. However, this dataset is not balanced and contains a lot of unnecessary and redundant data that needs to be eliminated, therefore, pre-processing is must. During the pre-processing phase, tokenization, stemming, stop words removal and segregation like techniques are implemented to make the data balanced and normalized. After this, the given dataset is divided into subsets by using k-fold cross validation approach. The main motive for doing so is to train the ML classifiers effectively on various combinations of data so that its accuracy can be enhanced. Finally, the classification is performed by DT, RF, LR and Multinomial NB classifiers as per the training provided to them. The efficacy of the system is analyzed usingMATLAB on IMDB dataset for every fold. Simulation results revealed that LR classifiers is outperforming DT, RF and Multinomial NB in terms of accuracy, precision, recall and F1-Score as well, to prove its supremacy.
Keywords: Sentiment Analysis, ML classifiers, Movie review etc.

Introduction

One of the most emerging and in-demand study areas in text mining and natural language processing (NLP) is sentiment analysis (SA). NLP is basically a part of Artificial Intelligence (AI)and can defined as the system that has capabilityto comprehend the human language (spoken or written) that is often known as natural language. On the other hand, SA can be defined as the NLP technique used to assess the polarity of the data which can be either be positive, negative or neutral. Nonetheless, SA can also be defined as the technique that not only determines the polarity of data but also concentrates on sentiments like angry, pleased, sad etc., urgency and intents as well [1].Because of this, SA is often referred as opinion mining or extraction and review or attitude analysis, and is responsible for extracting, recognizing and categorizing opinions on various domains usually written in text. There are numerous online business sites available where people may discover about various product difficulties.The customers can share their opinions or sentiments regarding any product to general audience by using websites like Amazon, IMDB, yelp or e-commerce. Everyday millions of users write reviews about products, movies or organization which results in huge data that is stored in the form of e-documents. The information stored in these documents can be divided into two categories of facts and opinions. Whilst the facts focus on communication of objective data, opinion purely communicate sentiments [2].Sentiment analysis allows one to gauge customer sentiment toward various aspects of your company without having to read through a massive amount of customer reviews at once.

SA can broadly be classified into two types, one is Lexicon based SA and other is ML based SA. The lexicon technique relies on tokenizing text, calculating the amount of times every word appears, and actually looking every word's significance in a preexisting lexicon.As part of the machine learning methodology, the system becomes increasingly sophisticated by using a training set of data to train several classifiers [3, 4 & 5].Moreover, while determining opinions, SAusually goes under following sub tasks of-- sentiment classification (SC), Sentiment Lexicon Generation (SLG), Sentiment Quantification (SQ), Opinion Extraction (OE), Feature-Based Summary (FBS), and Opinion Spam (OS). The job of the SC is to classify any part of the given text into sentiments. These sentiments can be expressed on any three levels of document, sentence and feature levels. While as, the sentiment lexicon is generated by the SLG by marking words with sentiment polarity. Determining the frequency of various

sentiments across a group of texts is the work of SQ. Furthermore, the purpose of OE is to categorize and retrieve every opinion from the components in customer reviews.FBS is concerned with creating a feature description that determine the characteristics, components and other facets of product [6]. Finally, OS identifies the false or bogus content in data like false and untruthful reviews or comments.

Among all the categories, watching movies are considered as most convenient way of entertainment. However, only a small number of films are appreciated and profitable. As mentioned earlier, reviews are brief texts which typically give a viewpoint on films. Therefore, the success of the film is greatly influenced by reviews [7-8].The movie buffs may choose which movies to see and which to skip using one of the various rating websites like IMDB, Rotten Tomatoes etc. The users rate movie on these websites by giving a score out of 10 stars and on the basis of these stars the success of failure rate of the movie is determined. Hence, Reviews play a significant part in bringing audiences to the cinemas in addition to word-of-mouth advertising.To put it another way, SA on movie reviews facilitates Opinion Summarization by capturing the reviewer's emotion. However, with the advancement in technology in the last few years, the researchers are still facing a number of issues in this domain that needs to be resolved. The two main drawbacks are the keywords having different meanings as per their content that leads to ambiguity, and incapability of categorizing reviews that doesn't depict clear emotional keywords. Therefore, it is important to keep these facts in mind while designing a new SA system.

Research Motivation

The impetus for this study originates from the fact that today there are billions of online users around the world today, and it has been seen that the amount of content produced by these people on the web is growing quickly. All kind of textual data, including photographs, video, and videos, can be included in this information. Nevertheless, the majority of consumers even now express their thoughts about brands, movies, and services using sentences. As a result, a substantial amount of information is produced in the form of textual reviews. It is crucial that we analyze and anticipate the reviews because failing to do so will make the data inoperable. In this paper, we are going to analyze the performance of some widely used classifiers specifically for understanding the sentiments of reviews for movies.

The remaining section of the paper is categorized as; Section 2, reviews recent publications for determining sentiments for movies, followed up by the problem statement. The section 3 of this paper describes the proposed work and results obtained are discussed in section 4. Finally, a conclusion of the analytical study is given at the end of paper.

Literature Survey

Over the past few years, a significant number of researchers are paying their attention towards sentiment classification by using various ML and DL algorithms. In this section of the paper, we are going to discuss and review some related papers that particularly use different techniques on IMDB dataset and also what outcomes were obtained. Qaisar et al. [9],proposed a DL based sentiment analysis model wherein they used LSTM classifier for analyzingand categorizing movie reviews. The model preprocesses data and then divided for enhancing the post classification performance. the efficacy of the suggested model was analyzed on IMDB dataset upon which an accuracy of 89.9% was attained.S. Sabba et al. [10],proposed an effective SA system that was based on NLP and Deep CNN models for resolving the various issues faced in user SA. The model was again tested on the IMDB dataset that contained a total of 50,000 reviews and achieved an accuracy of 99 and 89% in training and testing phase respectively. K. Amulya, et al. [11],reviewed the performance of various ML and DL classifiers for identifying sentiments on IMDB dataset. The authors stated that ML only work in single layer that decrease their output value whereas, DL algorithms work on multiple layers to give better results. The given paper aimed at helping the new researchers and scholars to select the best technique for SA. The contrast between machine learning and deep learning methodologies demonstrates that DL algorithms produce precise and effective outcomes.G Donia, et al. [12],analyzed the efficacy of three ML algorithms like ANN, SVM and NB for detecting opinions of users on renowned review datasets including movie, product and smart gadgets of last five years. Through extensive experimentation, it was revealed that ANN produces an accuracy of 90.3% when feature is extracted through Unigram technique. Haque, et al. [13],evaluated and compared the effectiveness of three DL classifiers i.e. CNN and LSTM and hybrid of CNN-LSTM for extracting sentiments from texts. The efficiency of the suggested model was analyzed on IMDB dataset to determine which framework generates more

accurate results. Results simulated that CNN outperforms LSTM and CNN-LSTM models and other standard models with an F-Score of 91% on IMDB movie review database. Similarly, R. Bandana,[14], integrated ML and lexicon-based features along with supervised algorithms (NB and linear SVM) to develop a new sentiment analysis model. The outcomes obtained from the work stated that proposed heterogenous feature and hybrid SA system outperforms all other similar models in terms of various performance dependency factors. Furthermore, they also stated that for analyzing large datasets more accurate and effective SA models can be developed by using heterogenous features and DL classifiers. Again Kumar, H. M., [15], showcased the impact of hybrid features by integrating ML features like TF, TF-IDF along with lexicon features on accuracy of SA. The supremacy of the suggested approach was validated over SVM, NB, KNN and maximum entropy in terms of accuracy and complexity.Shaukat, Z., et al. [16],utilized NN that was trained on IMDB dataset in order to extract emotions or sentiments from reviews. The suggested model achieved an accuracy of 91% on the given dataset. A. Yenter et al. [17],proposed another DL based SA model wherein they used CNN along with the LSTM model for extracting opinions from movie reviews. The results obtained highest accuracy on IMDB dataset to prove its supremacy. T İlhan et al. [18],proposed a SA model in which vector space was developed in KNIME analytics platform, upon which classification was implemented by using DT, SVM and NB. The results were analyzed on IMDB and twitter datasets respectively. On IMDB dataset, the suggested model achieved accuracy of 94, 73.20 and 85.50%, whereas, it was only 82.76, 75.44 and 72.50% for DT, NB and SVM classifiers.

From the above literatures, it can be summarized that a number of ML and DL approaches have already been proposed by various researchers for extracting the opinions from movie reviews datasets. However, a number of limitations are faced by scholars while extracting sentiments from IMDB datasets which degraded its performance. Generally, it has been seen accuracy of the SA model decreases when users use keywords with ambiguity meaning. In such cases, the current SA approaches are unable to detect the polarity of the text which in turn results in decreased accuracy. Moreover, the current systems are unable to extract the meaning of slang words or short forms like "LOL", "ROFL", etc., in reviews that also is a major challenge in SA. In addition to this, we have seen that majority of research-ers are using ML classifiers in their work for analyzing opinions from text, however, each classifier performs differently on different datasets. Moreover, selecting suitable classifier for extracting the polarity of text is still one of the major issues that needs to be resolved. In this regard, an analytical study must be undertaken wherein efficacy of various classifiers should be analyzed on specific dataset.

Present Work

In this paper, an analytical study is conducted specifically for analyzing the sentiments in movie comments or reviews. The main aim of the proposed work is to analyze and understand which classifier shows best accuracy results for a particular movie dataset. To achieve this objective, the proposed model undergoes through various steps of data collection, pre-processing (tokenization, stop word removals and segregation), data segregation, training & testing and finally classification. Initially, the data has been taken from IMDB dataset that comprises a total of 50000 movie reviews for NLP. The detailed description of the given dataset is given in methodology section of this paper. Nonetheless, the data present in this dataset is not normalized and balanced which leads to biased solutions and hence deteriorated accuracy. To overcome this, we have applied a number of techniques like tokenization, stop word removal, stemming and segregating in the pre-processing phase. After this, k-fold cross validation is applied to the given dataset for dividing it into k subsets that are also called as folds. The main reason for doing so is to train classifiers on different various datasets combinations so that accuracy of the system is enhanced. The value of k determines number of subsets that a dataset is divided into. It must be noted here that the classifier is trained on all these subsets but one i.e. (k-1) of the subsets. The one remaining data subset that is not sed in training the classifiers is used for testing its efficacy. Here, we have analyzed the performance of four ML classifiers i.e. Decision Tree (DT), Random Forest (RF), Logistic Regression (LR) and Multinomial NB on the IMDB dataset on 6-fold cross validation technique. The 6-fold cross validation means that classifiers will be trained on 5 folds and the its efficacy is tested on 1 dataset. The results for the given study is obtained in terms of various performance dependency factors like accuracy, precision, recall and Fscorefor 5 folds on IMDB dataset. The general working methodology of the proposed work is given in the next section of this paper.

Methodology

In order to analyze the efficacy of four ML classifiers on the IMDB dataset, the proposed model undergoes through various stages that are described thoroughly here. As mentioned earlier, the proposed work is particularly conducted for analyzing best ML algorithm on IMDB movie dataset.

Data Acquisition: The very first step that is opted in the proposed work is collecting necessary data about movie reviews or comments. Here, we have used IMDB dataset which is one of the popular movie datasets available online on Kaggle.com. The dataset comprises a total of 50K reviews about various movies that is divided into two categories of training and testing. The detailed information of the dataset is given below.

IMDB Dataset

IMDB is a popular movie review dataset that has been utilized by number of researchers in their work for trainingand testing their classifiers. The dataset comprises a total of 50,000 movie review entries that can be used for text analysis and NLP. Basically, the given dataset is binary sentiment classification repository wherein considerably more information is provided than other previous standard datasets. Moreover, the dataset is divided into two categories of training dataset and testing dataset with 25k movie reviews for each. This training data is used for training the classifier and finally the testing data validated its efficacy.

Pre-processing: Nonetheless, the data collected in the previous step is not normalized and balanced which decreases classification accuracy rate. Therefore, it is important to pre-process data before passing it to classifiers for training and testing purpose. During the pre-processing phase, following steps are implemented on the IDB dataset

- Tokenization: This is the first step towards making the dataset balanced and normalized. In this step, the movie reviews or comments present in the dataset are divided into various phrases also called as tokens, symbols and words.
- Stop Word Removal: In this step, pronouns like "I, We, they, she etc." and other words like "and, the, for, should and so on" are eliminated from the movie reviews. By doing so, only important words that depict the polarity of review are retained which helps in enhancing the model's efficiency.
- Segregation: it can be defined as the process wherein any special character that's is present in movie reviews like <, @, &, %, / etc. are removed for mak-

ing data more informative and effective.

- Stemming: During stemming process, the redundant words are minimized to their basic forms. For an example, the word "dancer" and "dancing" are reduced to their basic root word "dance".

Data splitting: Once the data is processed, we have divided it into k subsets or folds. The value of k represents the total number of data subsets formed and is 6 in the proposed work. The classifiers used in the proposed model are trained on all subsets of data except one (k-1). It is basically a resampling technique that determines the efficacy of ML algorithms of limited datasets. The process starts by shuffling the given dataset randomly which is then divided into k groups. Each group represents the different combination of data upon which the classifiers are trained and tested for validating their efficacy.

Classification: In this phase, we have analyzed the performance of four ML classifiers, those are—DT, RF, LR and Multinomial NB on IMDB dataset for 6-fold cross validation.The main reason for doing so is to analyze which classifiers shows best results on which subset of data. The four classifiers are trained on various folds and accordingly their performance is analyzed under various performance parameters. The next section of the paper discusses results that are obtained for proposed methodology.

Analytical Results

The usefulness of the proposed study is performed in MATLAB software.The simulated outcomes were obtained in terms of various metrices like Accuracy, precision, recall and Fscore under various folds and overall results. The detailed description of the results is discussed in this section of the paper.

Performance Evaluation

In order to analyze the efficacy of ML classifiers, we firstly analyzed their overall performance in terms of accuracy, precision, recall and F1-score on IMDB dataset. The values obtained for each ML classifier are recorded in tabular form and is shown in table 1. After analyzing the given table, it can be concluded that the value of accuracy came out to be highest in LR with 90.18%, followed up by Multinomial NB with 86.43% and then RF and DT with 84.68% and 72% respectively. Similarly, the value of precision in DT, RF, LR and Multinomial NB was accounted at 71.42%. 84.19%, 89.36% and 88.60% respectively. Moreover, we have analyzed the performance of four classifiers (DT, RF, LR and Multinomial NB) in terms of their recall score

and F1-measure as well. The recall outcomes were mounted to 71.80%, 83.64%, 91.22% and 82.90% in DT, RF, LR and Multinomial NB respectively, while as, it was 71.61%, 83.91%, 90.28% and 85.60% for Fscore in four DT, RF, LR and Multinomial NB respectively on IMDB dataset.

Table 1: Specific value of ML classifiers for different parameters

ML model	Accuracy	precision_score	recall_score	f1_score
DecisionTree	72%	71.42%	71.80%	71.61%
RandomForest	84.68%	84.19%	83.64%	83.91%
LogisticRegression	90.18%	89.36%	91.22%	90.28%
MultinomialNB	86.43%	88.60%	82.90%	85.60%

Furthermore, we have also evaluated the efficacy of DT, RF, LR and Multinomial NB on IMDB dataseton every fold in terms of accuracy. From the given results, it has been analyzed that LR classifiers is showing better results for accuracy with 0.89, 0.90, 0.89, 0.896 and 0.892 on fold 0, 1, 2 , 3 and 4 respectively. On the other hand, the value of accuracy was mounted at 0.864, 0.863, 0.862, 0.869 and 0.861 in multinomial NB for 0, 1, 2, 3 and 4 folds. While as, the value of accuracy in DT and RF for IMDB dataset on 0, 1, 2, 3 and 4 folds were 0.719, 0.711, 0.7.31, 0.719 & 0.7154 and 0.842, 0.84, 0.847, 0.841 and 0.844 respectively. The specific value for accuracy obtained in DT, RF, LR and Multinomial NB on IMDB dataset on every fold is given in table 2.

Table 2: Specific value for accuracy on every fold

ML model	fold-0	fold-1	fold-2	fold-3	fold-4
MultinomialNB	0.8643	0.8637	0.8625	0.8697	0.8612
decision Tree	0.719	0.711	0.7031	0.719	0.7154
Logistic Regression	0.8924	0.901	0.8995	0.8968	0.8927
RandomForest	0.8428	0.84	0.8479	0.8411	0.8444

Similarly, the efficacy of Multinomial NB, RF, DT and LR were also analyzed on IMDB dataset for every fold in terms of their Fscore value. The results attained showcased that again LR is showing better results than other ML classifiers in terms of Fscore for every fold. The recall values were 0.859, 0.857, 0.864, 0.858 % 0.859 in multinomial NB, 0.713, 0.716, 0.701, 0.715 and 0.716 in DT and 0.84, 0.837, 0.846, 0.840 and 0.843 in RF. On the other hand, the value of F1-score was mounted at 0.893, 0.902, 0.9, 0.897 and 0.893 in LR classifier on 0, 1,2 3 and 4 folds respectively. The specific values for F1-score attained in each ML classifier for every fold is given in table 3.

Table 3: Specific value for F1-Score on every fold

Classifier	fold-0	fold-1	fold-2	fold-3	fold-4
MultinomialNB	0.859	0.8578	0.8642	0.8581	0.8592
Decision Tree	0.7137	0.7168	0.7012	0.7157	0.7164
Logistic Regression	0.893	0.902	0.9	0.8978	0.8937
Random forest	0.84	0.8379	0.8465	0.8403	0.8436

Likewise, the efficacy of four ML classifiers was also analyzed and verified on every fold of IMDB dataset in terms of their recall values. The specific values obtained for the recall for each classifier is given in table 4. After

analyzing the values of table, it can be seen that out of all classifiers LR shows best results for recall with a value of 0.906 for fold-0, 0.911 for fold-1, 0.907 for fold-2 & 3 and 0.9024 for fold-4. While as the recall values were 0.835, 0.831, 0.844, 0.827 and 0.829 in multinomial NB, 0.708, 0.713, 0.709, 0.718 and 0.7226 in DT and 0.84, 0.827, 0.839, 0.8366 and 0.8394 in RF for every fold.

Table 4: Specific value for recall on every fold

Classifier	fold-0	fold-1	fold-2	fold-3	fold-4
MultinomialNB	0.835	0.8316	0.8444	0.8274	0.829
decision Tree	0.7088	0.7132	0.7092	0.7188	0.7226
Logistic Regression	0.9062	0.9116	0.9072	0.9072	0.9024
RandomForest	0.841	0.8272	0.839	0.8366	0.8394

In addition to this, the efficiency of the four ML classifiers (Multinomial NB, DT, LR and RF) was analyzed and determined on every fold in terms of precision. The value of precision was recorded 0.884, 0.885, 0.885, 0.8912 and 0.8917 in Multinomial NB for every fold, 0.7094, 0.72, 0.698, 0.718 and 0.715 for DT on every fold; 0.881, 0.8926, 0.8933, 0.888 and 0.885 for LR on every fold and 0.844, 0.8489, 0.8542, 0.84419 and 0.8478 for RF on every fold. These values specify that LR is showing optimum results for precision as well to prove its supremacy over other classifiers. Table 5 shows the exact values of precision attained in each classifier for very fold.

Table 5: Specific value for Precision on every fold

ML Model	fold-0	fold-1	fold-2	fold-3	fold-4
MultinomialNB	0.8845	0.885811	0.885115	0.8912	0.8917
Decision Tree	0.7094	0.72	0.698	0.718	0.715
Logistic Regression	0.881	0.8926	0.89334	0.888	0.8852
Random forest	0.844	0.8489	0.8542	0.84419	0.8478

After analyzing the results attained in the form of tables, it can be concluded that out of four ML classifiers, LR is showing best results than other three classifiers (Multinomial NB, RF and DT) for all parameters.

Conclusion

In this paper, we have analyzed the performance of four ML classifiers which included Multinomial NB, RF, DT and LR on IMDB dataset by varying folds to see which classifier gives best results. The analysis of the proposed model is performed in MATLAB software under various metrices for every fold. The simulated outcomes revealed that LR classifiers was showing best results than other three ML classifiers. With an accuracy of 90.18% LR showcased its performance over DT, RF and Multinomial NB models whose accuracy rate was only 72%, 84.68% and 86.43% respectively. Moreover, when we analyzed the performance of four classifiers in terms of their precision as well, it came out to be 71.42% in DT, 84.19% in RF, 89.36% in LR and 88.60% in Multinomial NB models respectively. Similarly, the efficacy of four classifiers was also analyzed and validated in terms of their Recall score that acme out to be highest in LR with 91.22%, followed up by RF with 83.64%, followed up by multinomial NB with 88.90% and finally DT with 71.80% respectively. Furthermore, the F1-score values were 71.61%, 83.91%, 90.28% and 85.60% in DT, RF, LR and Multinomial NB classifiers respectively. In addition to this, we also analyzed the efficacy of given classifiers on every fold and results simulated revealed that LR is outperforming DT, RF and Multinomial NB on every fold in terms of accuracy, precision, recall and F1-score respectively to prove its efficiency.

References

1. Sudhir, Prajval, and Varun Deshakulkarni Suresh. "Comparative study of various approaches, applications and classifiers for sentiment analysis." Global Transitions Proceedings 2.2 (2021): 205-211.
2. Raghuvanshi, N., J. M. Patil, A naive bays classifier to classify movie, International Journal of Advanced Research in Computer Engineering & Technology

(IJARCET), vol.5 no.4, pp.1014–1017, 2016.

3. Sailunaz, K., Dhaliwal, M., Rokne, J., & Alhajj, R. (2018). Emotion detection from text and speech: a survey. Social Network Analysis and Mining, 8(1), 28.

4. Bandhakavi, A., Wiratunga, N., Padmanabhan, D., & Massie, S. (2017). Lexicon based feature extraction for emotion text classification. Pattern recognition letters, 93, 133-142.

5. Medhat, W., A. Hassan, H. Korashy, Sentiment analysis algorithms and applications: A survey, International Journal of Ain Shams Engineering Journal, vol.5, no.4, pp. 1093–1113, 2014.

6. Buche, A., D. Chandak, A. Zadgaonkar, Opinion mining and analysis: a survey, International Journal of Natural Language Computing (IJNLC), vol.2, no.3, pp.39–48, 2013.

7. Ghosh, Ayanabha. "Sentiment Analysis of IMDb Movie Reviews: A comparative study on Performance of Hyperparameter-tuned Classification Algorithms." 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS). Vol. 1. IEEE, 2022.

8. Karak, Gahina, et al. "Sentiment Analysis of IMDb Movie Reviews: A Comparative Analysis of Feature Selection and Feature Extraction Techniques." International Conference on Hybrid Intelligent Systems. Springer, Cham, 2021.

9. Qaisar, Saeed Mian. "Sentiment analysis of IMDb movie reviews using long short-term memory." 2020 2nd International Conference on Computer and Information Sciences (ICCIS). IEEE, 2020.

10. S. Sabba, N. Chekired, H. Katab, N. Chekkai and M. Chalbi, "Sentiment Analysis for IMDb Reviews Using Deep Learning Classifier," 2022 7th International Conference on Image and Signal Processing and their Applications (ISPA), 2022, pp. 1-6, doi: 10.1109/ISPA54004.2022.9786284.

11. K. Amulya, S. B. Swathi, P. Kamakshi and Y. Bhavani, "Sentiment Analysis on IMDB Movie Reviews using Machine Learning and Deep Learning Algorithms," 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT), 2022, pp. 814-819.

12. Gamal, Donia, et al. "An Evaluation of Sentiment Analysis on Smart Entertainment and Devices Reviews." International Journal Information Theories and Applications 26.2 (2019): 147-164.

13. Haque, Md Rakibul, Salma Akter Lima, and Sadia Zaman Mishu. "Performance analysis of different neural networks for sentiment analysis on IMDb movie reviews." 2019 3rd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE). IEEE, 2019.

14. R. Bandana, "Sentiment Analysis of Movie Reviews Using Heterogeneous Features," 2018 2nd International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech), 2018, pp. 1-4, doi: 10.1109/IEMENTECH.2018.8465346.

15. Kumar, H. M., B. S. Harish, and H. K. Darshan. "Sentiment Analysis on IMDb Movie Reviews Using Hybrid Feature Extraction Method." International Journal of Interactive Multimedia & Artificial Intelligence 5.5 (2019).

16. Shaukat, Z., Zulfiqar, A.A., Xiao, C. et al. Sentiment analysis on IMDB using lexicon and neural networks. SN Appl. Sci. 2, 148 (2020).

17. A. Yenter and A. Verma, "Deep CNN-LSTM with combined kernels from multiple branches for IMDb review sentiment analysis," 2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON), 2017, pp. 540-546, doi: 10.1109/UEMCON.2017.8249013.

18. Tarımer, İlhan, Adil Çoban, and Arif Emre Kocaman. "Sentiment analysis on IMDB movie comments and Twitter data by machine learning and vector space techniques." arXiv preprint arXiv:1903.11983 (2019).

Comparison of the performance of sintered abrasives to glued abrasives using a rotating magnetic field finishing setup

*Jagdeep Singh Gill **Lakhvir Singh

*Research Scholar, Mechanical Engineering Department, IKGPTU, Kapurthala, Punjab, India
**Mechanical Engineering Department, Baba Banda Singh Bahadur Engineering College,
Fatehgarh Sahib, Punjab, India
Corresponding Author Email ID: erjsgill@gmaol.com

Abstract

Magnetic abrasive finishing is a method of creating high-quality finishes by moving magnetic abrasive particles across the workpiece. The magnetic abrasive particles are composite powder containing hard abrasive grains in a ferromagnetic matrix. Magnetic abrasive finishing relies heavily on a cutting tool formed from magnetic abrasive particles. In the present work, a comparison has been made between the performance of sintered abrasives and glued abrasives. A rotating magnetic field finishing setup is used to conduct experimentation. Experiments were carried out to determine the effect of rotational speed and abrasive quantity on surface roughness. Improvement in surface roughness is taken as the performance parameter. The best result for the sintered magnetic abrasives was at 425 rpm and 6 g abrasive, with an improvement in the surface roughness value of 84.4%. In the case of the glued magnetic abrasives, the best result was at 575 rpm and 6 g of abrasive, with a surface roughness improvement of 65.65%. Sintered magnetic abrasives offer a more significant improvement in surface roughness for the same parameters than glued magnetic abrasives.
Keywords: Magnetic abrasive finishing, Sintering, Surface roughness, Abrasive.

Introduction

High-quality finishes by moving magnetic abrasive particles across the workpiece. The magnetic abrasive particles are composite powder containing hard abrasive grains in a ferromagnetic matrix. Magnetic abrasive finishing relies heavily on a cutting tool formed from magnetic abrasive particles. Surfaces of different shape, size, and material can be finished using this process. It produces very fine surfaces. Shinmura et al. [1] investigated the influence of the grain diameter of the abrasive particles on the material removal and the surface roughness during magnetic abrasive finishing. They discovered that the finishing pressure is affected by the magnetic flux density, the relative permeability of the ferromagnetic substance, and the volume ratio of a ferromagnetic substance contained in the magnetic abrasive particles, rather than the diameter of the MAP or the abrasive grains. Kreman et al. [2] used magnetic abrasive finishing estimating machining time to minimize roundness errors on cylindrical carbon steel specimens. They found that the greatest reduction in runout (OOR) occurs at the beginning of the machining process and that MAF is not affected by runout. The influence of the working gap and the peripheral speed of the workpiece on the performance of the finishing process with magnetic abrasives have been studied by Jain et al. [3]. They used loosely bound magnetic abrasive particles (made from a homogeneous mixture of iron and aluminum oxide particles) and Servospin-12 lubricating fluid to conduct experiments on a cylindrical stainless steel workpiece. Based on their findings, they found that increasing the working distance or reducing the peripheral speed of the workpiece generally reduced the amount of material removed. As the peripheral speed of the workpiece increases, so does the surface quality. To predict the surface e roughness, Raghuram and Joshi [4] performed a parametric study and analytical modeling of the MAP process of stainless steel sheets (SUS304). They found that the size ratio, the surface clearance of the tool work surface, the polishing speed, the diameter of the magnetic abrasive and the polishing time have a significant influence on the value of the surface roughness. Mulik and Pandey [5] examined the effects of voltage, mesh count, revolutions per minute (RPM) of the electromagnet, and weight percent of the abrasive. The response was measured as the percentage change in surface roughness. The mesh count has been shown to have the greatest influence on the percentage change in surface roughness, followed by the% weight of the abrasive, the speed of the electromagnet and the voltage. Givi et al. [6] investigated the influence of the speed of the permanent magnet poles, the working gap between the permanent magnet and the workpiece, the number of cycles and the weight of the abrasive particles on the MAF of aluminum alloy sheets using the experimental technique. According to the researchers, the number of cycles, the working gap and the rotational speed are the elements that have a great influence on the surface roughness. Kang et al. [7] investigated the use of sintered diamond-based magnetic abrasives for inter-

nal magnetic grinding of SUS304 stainless steel. They investigated how the distance between the workpiece and the magnet affected the percentage improvement in surface polish as well as the rotation speed of the magnetic poles, the percentage of abrasives in the iron matrix and the processing time. Accordingly, machining time, diamond abrasive%, and pole rotation speed all have a significant impact on PISF. The SEM images of the samples before and after MAF show that the surface has improved significantly after the finishing process. Kadhum et al. [8] developed a magnetic inductor for polishing flat surfaces produced by a vertical milling machine. They used a Taguchi experiment to see how coil current, working gap, powder component volume, and feed rate affected the surface quality of non-ferromagnetic (7020 aluminum alloy) and ferromagnetic (410 stainless steel) workpieces. They found that with non-ferromagnetic materials, the amount of powder and the working gap are more important properties than current and feed. In contrast to the amount of powder and the feed speed, the current and working gap are important parameters for ferromagnetic materials. Verma et al. [9] developed a new tool for polishing holes, grooves and vertical surfaces based on the MAF method. They used the CCD approach to assess tool performance and examine the influence of variables such as speed, magnetic flux density, abrasive size, and abrasive weight percentage on the PISF of stainless steel (SS304) tubes. In completing an SS304 pipe, they found that the magnetic flux density was the most effective parameter, followed by the speed of rotation. Hang et al. [10] developed a new ultra-precise magnetic abrasive process for wire material that uses a rotating magnetic field. They tested the influence of variables such as the magnetic field speed, the vibration frequency of the wire workpiece and the unlimited magnetic abrasive grain size on the change in surface roughness and the removed diameter of AISI 1085 steel wire material. With a magnetic abrasive grain size of 0.5 m and a vibration frequency of 10 Hz at 800 rpm for 60 seconds, there are the best conditions for finishing workpieces made of wire material. The roughness of the original surface has been reduced from 0.25 m to 0.02 μm. Wu et al. [11] used a low-frequency alternating magnetic field to remove groove edge burrs and to improve the surface quality of the group through a magnetic abrasive finishing technique. They investigated the influence of the magnetic pole shape on the finishing qualities, and found that arc groove poles have higher magnetic flux density than flat, conical, concave, convex and flat groove

poles. During high speed magnetic grinding machining of alumina ceramic rods, Song et al. [12] examined the influence of input parameters such as diamond grain size, vibration frequency and diamond paste weight on output reactions such as surface roughness, diameter variation, roundness and removed weight. Under ideal conditions, they achieved a surface roughness of 0.01 μm and a roundness of 0.14 μm. To treat beta titanium wire with a magnetic abrasive finishing process, Nam et al. [13] used a multiple transfer movement approach. They found that a surface roughness Ra of 2000 rpm, a particle size of 1 m and a machining time of 300 seconds gave the best surface roughness Ra. The roughness of beta surface titans improved from 0.32 μm to 0.05 μm. With a processing time of 300 seconds, the effect of the finishing gap shows that a gap of 3 mm has a higher processing power than a gap of 5 mm. For the magnetic abrasive finishing of aluminum samples, Lee and Chang [14] used a horizontal magnetic field structure of the transverse type. During their research, they found that white aluminum oxide is superior to green silicon carbide when it comes to aluminum samples. Within 2-3 minutes after the final examination, they reach a fine surface of Ra 0.06 μm. During the magnetic abrasive finishing process, Xie and Zou [15] studied the effect of changing the current mode on magnetic flux and finishing force. During their research, they discovered that pulsed current can achieve a better material removal rate compared to static magnetic fields and sinusoidal alternating magnetic field values. To increase the flatness of flat surfaces, Zang and Zou [16] proposed a variable speed magnetic loop finishing process. They changed the speed of the finishing tool to manage the machining time on the workpiece surface and make it flatter. In order to achieve homogeneity of the magnetic flux over the magnetic pole surface, they constructed a small magnetic pole with a diameter of 1 mm.

It is seen from literature review that most of the studies related to magnetic abrasive finishing process are concerned with evaluating the affect of process parameters on the surface finish. So, the present work is undertaken to compare the performance of sintered abrasives and glued abrasives. Aluminium oxide powder is mixed with iron powder in a ratio of 20:80 to produce abrasives. A rotating magnetic field finishing setup is used to conduct experimentation. Experiments were carried out to determine the effect of rotational speed and abrasive quantity on surface roughness. Improvement in surface roughness is taken as the performance parameter. Aluminium tubes are taken as workpiece material

in the present study.

2. MATERIAL AND METHOD
2.1 EXPERIMENTAL SETUP

The photograph in Figure 1 shows a machine used for finishing workpieces with magnetic abrasive. The experimental setup for finishing the aluminum tube in the MAF process consists of four cylindrical permanent magnets mounted on a stainless steel chuck. The steel chuck serves two purposes: on the one hand as a carrier for the magnets and on the other hand as an isolator for separating. The magnets can move radially up and down to vary the gap between the workpiece and the magnet surface. This arrangement offers flexibility to process workpieces of different sizes. The magnetic chuck can be rotated at the desired speed with a DC motor.
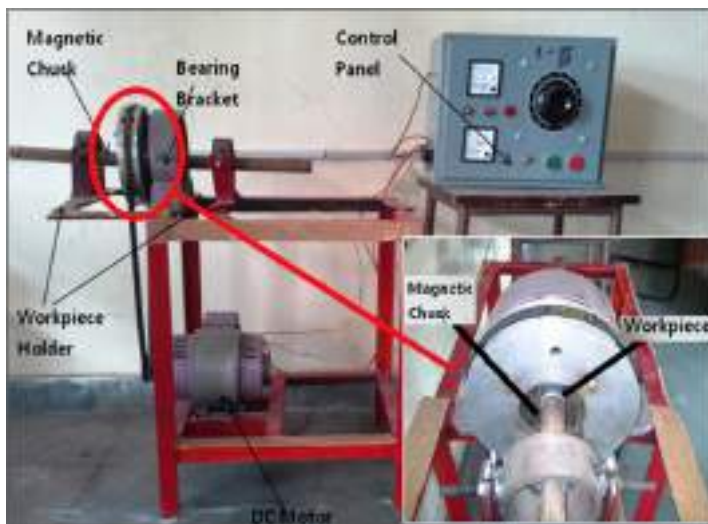
Fig.1 Photographic view of the MAF setup

2.2 MAGNETIC ABRASIVES

The magnetic abrasive particles are composite powder containing hard abrasive grains in a ferromagnetic matrix Two types of magnetic abrasives were prepared for experimentation one is sintered abrasives and other is glued abrasives. Aluminium oxide with mesh size 100-300 is used as abrasive material and iron powder with mesh size 300 is used as ferromagnetic material for preparing magnetic abrasives for the experimentation. Aluminium oxide powder is mixed with iron powder in a ratio of 20:80 to produce abrasives. For preparing sintered magnetic abrasives, the powder mixture was compacted in a cylindrical shape die and then sintered in an H2 environment for 2 hours at 1150°C. Sintered compacts were crushed into small particles for preparing abrasive powder. Figure 2 shows a photograph of the sintering setup. To make glued magnetic abrasives, Fevitite bonding tubes are properly mixed with the mix-ture of aluminium oxide and iron powder so that the mixture soaks up the bonding glue. This mixture takes a day to dry properly. A large pellet was created. Then this compact was mechanically crushed into fine powder to form magnetic abrasives.

Fig.2 Photograph of the sintering setup

Fig.3 Photograph of Fevitite bonding tubes

Fig.4 Photograph of workpieces

2.3 MAGNETIC ABRASIVES

Table 1 shows the experimental conditions. The magnetic abrasive particles are introduced into the workpiece. These particles combine to form a flexible magnetic abrasive brush in the workpiece. The rotation imparted to the chuck causes the magnetic field to rotate, causing the magnetic abrasives to rotate with a tangential force which, along with the normal force, develops pressure on the inner surface of the workpiece. This pressure is responsible for the abrasion of the inner surface of the pipe by magnetic abrasive particles. Since the improvement in surface roughness is considered a performance parameter, it is necessary to measure the surface roughness before and after the finishing process. The surface roughness is measured by using Mitutoyo surface roughness tester (Model SJ 210) at four locations inside the pipe before and after finishing, and its mean value is used as the final value for calculating the improvement in surface roughness. The improvement in surface roughness is calculated as follows:

$$ISR = \frac{(Initial\ roughness - Final\ roughness)}{Initial\ roughness} X100$$

TABLE 1: EXPERIMENTAL CONDITIONS

Workpiece Material	Aluminium tube (ϕ 50 x 2mm)
Machining time	60 min
Lubricant	Light Oil(10% of the quantity of abrasives)
Type of the abrasives	Sintered and Glued
Abrasive percentage	20%
Response	Improvement in Surface roughness

3. RESULTS AND DISCUSSION

The main aim of this experiment was to compare the performance of sintered abrasives with glued abrasives and to determine the effect of rotational speed and abrasive quantity on surface roughness.

3.1 Effect of Rotational speed on surface roughness with 6g of abrasives

Table 2 shows the effect of changing the speed of rotation of magnetic poles on the ISF using 6 g of magnetic abrasive. The results of Table 2 are shown in the form of a graph. As shown in Figure 5, the ISF of sintered magnetic abrasives is higher than that of glued abrasives. Each case shows a similar trend with an increase in ISR

followed by a decrease, but the values are different in both situations. The ISR grows up to 425 rpm on sintered magnetic abrasives before decreasing. For glued abrasives, the ISR increases to 575 rpm before falling off.

TABLE 2: Experimental Conditions

Speed (r.p.m)	Size (μm)	Gap (mm)	Qty of MA (g)	ISR with SA (%)	ISR with GA (%)
350	163	3	6	65.6	47.99
425	163	3	6	84.03	55.37
500	163	3	6	78.98	58.27
575	163	3	6	76.1	65.65
650	163	3	6	71.4	59.7
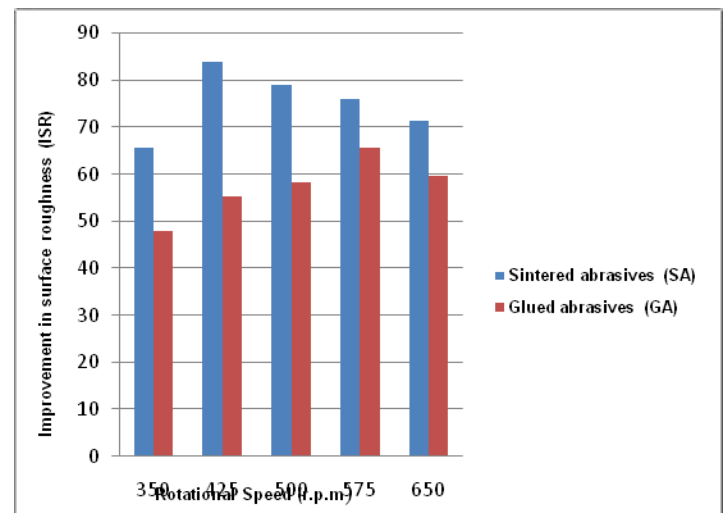
Fig.5 Effect of rotational speed (r.p.m) on ISR with 6g of abrasives

3.2 Effect of Rotational speed on surface roughness with 8g of abrasives

Table 3 shows the effect of changing the speed of rotation of the magnetic poles on the ISF using 8 g of magnetic abrasive. The results of Table 3 are shown in the Figure 6 in the form of a graph. The results show a similar trend as observed in Section 3.1, but the value of the ISR is different. The ISR grows up to 500 RPM on sintered magnetic abrasives before decreasing. For glued abrasives, the ISR increases to 425 RPM before falling off. This difference is caused by a change in the permeability of the work area due to a change in the amount of abrasive. This change leads to a change in the cutting force acting in the area.

TABLE 3: Experimental Conditions

Speed (r.p.m)	Size (μm)	Gap (mm)	Qty of MA (g)	ISR with SA (%)	ISR with GA(%)
350	163	3	8	53.50	51.9
425	163	3	8	57.2	55.8
500	163	3	8	72.02	46.87
575	163	3	8	51.1	42.97
650	163	3	8	41.11	37.78
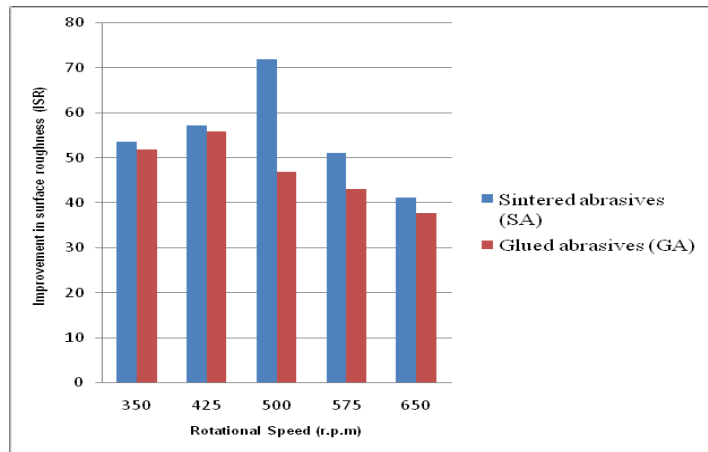
Fig.6 Effect of rotational speed (r.p.m) on ISR with 8g of abrasives

3.3 Effect of Quantity of abrasives on surface roughness with 425 rpm speed of poles

Table 4 shows the effect of changing the amount of abrasive on the ISR at a pole rotation speed of 425 rpm. The results of Table 4 are shown in Figure 7 in the form of a graph. The ISR grows up to 6g on sintered magnetic abrasives before decreasing. For glued abrasives, the ISR increases to 8 g before it falls off. This is because after a certain value the abrasives start to fall or get mixed up instead of moving with the rotating magnetic field.

TABLE 4: Experimental Conditions

Qty of MA (g)	Size (μm)	Gap (mm)	Speed (r.p.m)	ISR with SA (%)	ISR with GA(%)
4	163	3	425	64.7	50.3
6	163	3	425	84.4	55.37
8	163	3	425	57.2	55.8
10	163	3	425	46.39	45.95
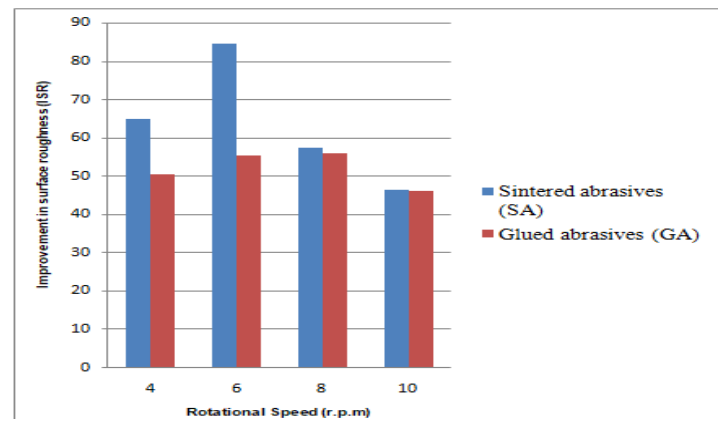
Fig.7 Effect of quantity of abrasive on ISR at 425 rpm speed of poles

3.4 Effect of Quantity of abrasives on surface roughness with 575 rpm speed of poles

Table 5 shows the effect of changing the amount of abrasive on the ISR at a pole rotation speed of 425 rpm. The results of Table 5 are shown in Figure 8 in the form of a graph. The results show that the ISR grows up to 6g on sintered magnetic abrasives and glued abrasives before it falls off.

Table 5: Experimental Conditions

Qty of MA (g)	Size (μm)	Gap (mm)	Speed (r.p.m)	ISR with SA (%)	ISR with GA (%)
4	163	3	575	75.1	64.47
6	163	3	575	76.1	65.65
8	163	3	575	51.1	2.97
10	163	3	575	49.4	36.93

Fig. 8 Effect of quantity of abrasive on ISR at 425 rpm speed of poles

3.5 Surface morphology

In order to get to know the resulting surface better, microscopic photographs were taken of samples before and

after the magnetic abrasive finishing process. Figure 9 (a) shows an untreated surface, while Figures (b) and (c) show surfaces that have been processed with sintered and glued abrasives, respectively. The results show that both abrasives removed scratches and sanding marks, although the results obtained with the surfaces obtained with sintered abrasives (SA) are superior to those obtained with bonded abrasives (GA).

Fig. 9 Microscopic images of unfinished and finished surface

4. CONCLUSIONS

Following is a summary of conclusions from this study:

1. The best result for the sintered magnetic abrasives was at 425 rpm and 6 g abrasive, with an improvement in the surface roughness value of 84.4%.

2. In the case of the glued magnetic abrasives, the best result was at 575 rpm and 6 g of abrasive, with a surface roughness improvement of 65.65%.

3. Sintered magnetic abrasives offer a more significant improvement in surface roughness for the same parameters than glued magnetic abrasives.

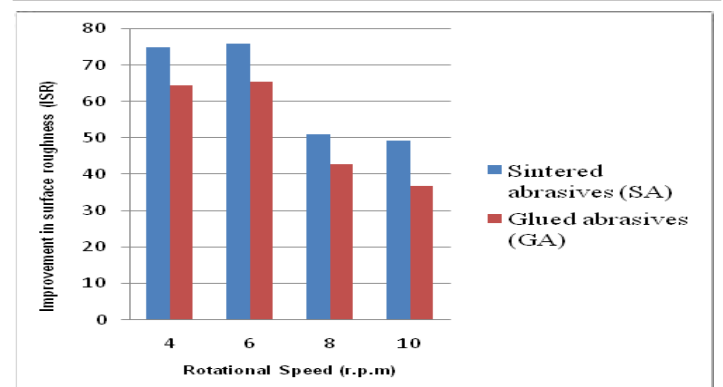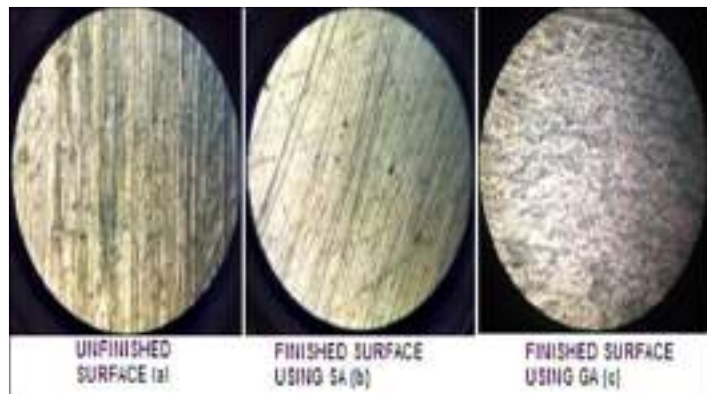4. Microscopic photographs of the final surfaces also confirmed the experimental results that sintered magnetic abrasives gave better surface results than glued magnetic abrasives.

REFERENCES

[1] Shinmura, T., Takazawa, K., & Hatano, E., "Study on Magnetic Abrasive Finishing- Effects of Various Types of Magnetic Abrasives on Finishing Characteristics", Bull. Japan Society of Precision Engineering, 21(2), 1987, 139-141.

[2] KREMAN, G. Z., ELSAYED, E. A., & REBEIRO, J. L., "Machining time estimation for magnetic abrasive processes", INT. J. PROD. RES., 32(12), 1994, 2817-2825.

[3] Jain, V., Kumar, P., Behera, P., & Jayswal, S.,"Effect of working gap and circumferential speed on the performance of magnetic abrasive finishing process", Wear, 250(1-12), 2001, 384-390.

[4] Raghuram, M. V., & Joshi, S. S., "Modeling of Polishing Mechanism in Magnetic Abrasive Polishing", The 12th International Conference of International Association for Computer Methods and Advances in Geomechanics(IACMAG) 1-6 October 2008 ,Goa .344-352.

[5] Mulik, Rahul S., and Pulak M. Pandey. "Magnetic abrasive finishing of hardened AISI 52100 steel." The International Journal of Advanced Manufacturing Technology, vol. 55, no. 5-8, 2010, pp. 501–515.

[6] Givi, Mehrdad, Alireza Fadaei Tehrani, and Aminollah Mohammadi. "Polishing of the aluminum sheets with magnetic abrasive finishing method." The International Journal of Advanced Manufacturing Technology 61.9-12 2011, 989-98.

[7] Kang, Parmvir Singh, et al. "FINISHING OF SUS 304 STAINLESS STEEL BENT TUBES USING MAGNETIC ABRASIVE FINISHING." Mechanica Confab, vol. 2, no. 3, 2013, pp. 123–130.

[8] Kadhum, Ali H, et al. "The Effect of Magnetic Abrasive Finishing on the Flat Surface for Ferromagnetic and non-Ferromagnetic materials." Al-Nahrain University, College of Engineering Journal (NUCEJ), vol. 18, no. 1, 2015, pp. 66–75.

[9] Verma, Girish Chandra, et al. "Experimental investigations into internal magnetic abrasive finishing of pipes." The International Journal of Advanced Manufacturing Technology, vol. 88, no. 5-8, 2016, pp. 1657–1668.

[10] Heng, L., Yin, C., Han, S., Song, J. and Mun, S., "Development of a New Ultra-High-Precision Magnetic Abrasive Finishing for Wire Material Using a Rotating Magnetic Field", Materials, 12(2),2019, p.312.

[11] Wu, J., Yin, S., Xing, B. and Zou, Y., "Effect of magnetic pole on finishing characteristics in low-frequency alternating magnetic field for micro-groove surface" The International Journal of Advanced Manufacturing Technology, 104(9-12), 2019, pp.4745-4755.

[12] Song, J. et al., "Micro-Machining Characteristics in High-Speed Magnetic Abrasive Finishing for Fine Ceramic Bar", Metals, 10(4), 2020, p.464.

[13] Nam S, Kim J, Mun S., "Magnetic Abrasive Finishing of Beta-Titanium Wire Using Multiple Transfer Movement Method." Applied Sciences. 10(19), 2020, 6729.

[14] Lee P, Chang J, "A magnetic abrasive finishing (MAF) platform utilizing horizontal transverse magnetic field magnetized by permanent magnets." Microsystem Technologies, 27(6), 2020, pp.2499-2506.

[15] Xie H, Zou Y., "Study on the magnetic abrasive

finishing process using alternating magnetic field—discussion on the influence of current waveform variation." The International Journal of Advanced Manufacturing Technology, 2021; 114(7-8), pp. 2471-2483.

[16]   Zhang, Y., & Zou, Y., "Study of corrective abrasive finishing for plane surfaces using magnetic abrasive finishing processes." Nanotechnology and Precision Engineering. 4(3), 2021, 033001.

To analyse the behaviour of AISI 4340 steel under lubricated condition using Pin On Disc apparatus

*Amrinder Pal Singh **Krishan Kumar
*Department of Mechanical Engineering, RIMT University (Punjab) India 140406
**Department of Mechanical Engineering, RIMT University (Punjab) India 140406

Abstract
In the sliding pairs, friction and wear are the core reasons for energy losses that reduce the efficiency of the mechanical systems. This paper presents an experimental study to predictthe frictional and wear behaviour of AISI 4340 steel under lubricated conditions. The experiment was conducted on a pin on disc machine in which EN-31 steel disc was used as a counter surface. The tribological characteristics of AISI 4340 material were observed under different loads varied as 30N and 90N at a fixed sliding distance of 1500m at a continuing speed of 1.0, 1.5, 2.0, 2.5 m/s. It was found that variation of load and sliding distance plays a crucial role in frictional force and wear behaviour of the gear material. The wear and tear of AISI 4340 steel was found to be increased with sliding distance. Load is the significant factor which liable for wear loss. The result showed that all loads the COF was minimum at the sliding speed of 2.0 m/s and COF increases with the rise in sliding speed.
Keywords: Sliding Speed , Sliding Distance, Pin on Disk, Coefficient of Friction , Normal Load

Introduction
When the surfaces of two or more components of a machine interact, various types of wear occur [1]. Under severe pressures, abrasion, sliding, and contact fatigue can cause more frequent and complex surface and sub-surface damage to wear behaviour [2–4]. Scuffing occurs when the load and sliding speed cause an impulsive impact on the material's surface during operation. Better gear life is obtained by surface modification methods for several years [5–7]. Friction is the resistance to motion that occurs when a solid slides across another body. Wear, on the other hand, is a difficult concept to define clearly and completely [8]. There's also no way to totally eliminate friction from sliding pairs. Oiling is the most effective way to reduce friction between moving parts. An oil film is placed between the contact surfaces of moving parts to reduce wear [9]. Wear and friction can be easily decreased by using lubricants. It's a really effective way to lessen the effects of wear and tear. When executing an operation, friction occurs. We must use extreme caution while selecting the kind of lubricant to be utilised in the operation [10]. Wearing behaviour using tribometers required to minimize friction and wear of materials. The wear rate of the material can be assisted by using digital balances to calculate weight loss.

A tribometer is a device that allows pin-on-disk or ball-on-disk tests to be performed with precision. The best application of the tribometer device was used to conduct dry or lubrication type wear tests. The tribometer allows researchers to investigate tribological behaviour while altering pressure of contact, rubbing speed, time, and lubricants [11,12].

Materials and methods
AISI 4340 alloy steel was utilised as specimens for the experimental study. The major components of this alloy steel are Mn and Cr, which provide it with excellent strength and toughness. Specimen with a dia(d) of 10 mm and a length of 40 mm was used. Such pins were used because round heads of pins have proper contact with revolving disc. To achieve a hardness of 46 HRC, heat treatment was conducted. The Rockwell hardness machine was used to where the hardness of pins was tested using a diamond ball intender with a load of 150 kg. Chemical composition of work piece AISI 4340 is displayed in the following Table 1.

The hardness of the disc utilized in this wear test was 62 HRC . The disc was made of EN 31 material and the lubricant used for this wear test was SAE 80W-90 gear oil. The process parameters used in this investigation were sliding distance, sliding speed and varied loads. The setup of pin on disc is depicted in Figure 1.
Eightsample of pieces are used in order to determine the weight loss and COF values. For experimental work, sliding speeds of 1, 1.5, 2.0 and 2.5 m/s were employed with two different normal loads of 30 N and 90 N and sliding distances was fixed to 1500 m. Denver electronic machine was used to weigh the specimen. The specimens are washed in acetone before being cleaned. The difference between the original and final

weight loss values, as well as friction, was calculated to determine wear weight loss.

Fig.1. Pin on Disc Apparatus

TABLE 1. Chemical composition of the pin and the disc.

| Material	C	Mn	P	S	Cr	Mo	Si	Al	Cu
AISI-4340	0.420	0.750	0.006	0.03	0.79	0.25	0.270	0.030	0.210
EN-31	1.03	0.4562	0.007	0.04	1.50	0.02	0.328	0.03	

Experimental design

Under lubricated conditions, research was conducted as per ASTM G-99-17 standard. The results that was seen are shown in Table 2.

2-COMBINITION OF PARAMETER AND LEVELS

LOAD(N)	SPEED (m/s)	SLIDING DISTANCE (m)	C.O.F (µ).
90	0.5	1500	0.05476
90	1.0	1500	0.055876
90	1.5	1500	0.047651
90	2.0	1500	0.054423

Results
1-Graph between COF and sliding speed when applied load is 30N and sliding distance fixed 1500m
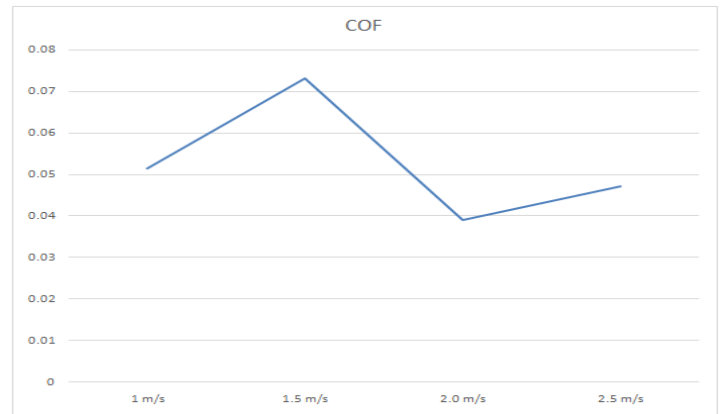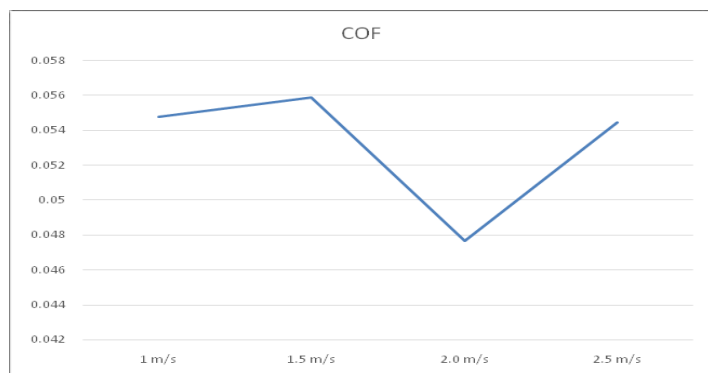
LOAD(N)	SPEED (m/s)	SLIDING DISTANCE (m)	C.O.F (µ).
30	1.5	1500	0.073214
30	2.0	1500	0.039138
30	2.5	1500	0.047303
30	2.5	1500	0.047303

1. COMBINITION OF PARAMETER AND LEVELS

Conclusions

1. At lower loads such as 30N the value of COF increases with the increase in the sliding speed from 1.0m/s to 1.5m/s and COF is minimum at the 2.0 m/s and then COF gradually rises with increase in the sliding speed.
2. At the higher load such as 90 N the value of the COF firstly slightly rises from when speed changes from 1m/s to 1.5 m/s and then COF value fall down at 2.0 m/s but after that COF abruptly increases. The result show that irrespective of the loads the COF is minimum at the 2.0 m/s sliding speed and then COF increases at the 30 N and 90 N loads
3. The weight loss rised with the increase in normal load as well as sliding distance. The amount of wear loss is high when load increases gradually as compared to sliding distance. Thus the load is the most significant parameter in this study

REFERENCES

1. A. Kahraman, H. Ding, Wear in gears, in: Q.J. Wang, Y.-W. Chung (Eds.), Encycl.Tribol., Springer US, Boston, MA, 2013, pp. 3993–4001.
2. S.J. Bull, J.T. Evans, B.A. Shaw, D.A. Hofmann, The effect of the white layer onmicro-pitting and surface contact fatigue failure of nitrided gears, Proc. Inst.Mech. Eng. 213 (1999) 305–313.

3. M. Boniardi, F.D. Errico, C. Tagliabue, Influence of carburizing and nitriding onfailure of gears–a case study, Eng. Fail. Anal. 13 (2006) 312–339.

4. C. Prakash, M.S. Uddin, Surface modification of b-phase Ti implant byhydroaxyapatite mixed electric discharge machining to enhance thecorrosion resistance and in-vitro bioactivity, Surf. Coat. Technol. 326 (2017)134–145.

5. C. Santus, M. Beghini, I. Bartilotta, M. Facchini, Surface and subsurface rollingcontact fatigue characteristic depths and proposal of stress indexes, Int. J.Fatigue 45 (2012) 71–81.

6. C. Prakash, S. Singh, B.S. Pabla, M.S. Uddin, Synthesis, characterization,corrosion and bioactivity investigation of nano-HA coating deposited onbiodegradable Mg-Zn-Mn alloy, Surf. Coat. Technol. 346 (2018) 9–18.

7. T. Bell, Source Book on Nitriding, ASM, Met. Park. OH, pp. 266–278, 1977.

8. 8. E.J. Mittemeijer, M.A.J. Somers (Eds.), Thermochemical Surface Engineering ofSteels: Improving Materials Performance, Elsevier, 2014, pp. 3–5.

9. M.A.J. Somers, E.J. Mittemeijer, Layer-growth kinetics on gaseous nitriding ofpure iron: evaluation of diffusion coefficients for nitrogen in iron nitrides, Metall. Mater. Trans. A 26 (1995) 57–74.

10. J. Halling, Principles of Tribology, The Contributors (1978).

11. C. Prakash, S. Singh, On the characterization of functionally graded biomaterialprimed through a novel plaster mold casting process, Mater. Sci. Eng., C 110(2020) 110654.

12. C. Prakash, H.K. Kansal, B.S. Pabla, S. Puri, Potential of powder mixed electricdischarge machining to enhance the wear and tribological performance of b-Tiimplant for orthopedic applications, J. Nanoeng. Nanomanuf. 5 (4) (2015) 261–269.

13. L. Yang, Pin-on-disc wear testing of tungsten carbide with a new moving pintechnique, Wear 225–229 (1999) 557–562.

14. B. Bhushan, B.K. Gupta, Handbook of Tribology: Materials Coatings and SurfaceTreatments, Krieger Publishing Company, Florida, 1997.

15. H. Singh, A.K. Singh, Y.K. Singla, K. Chattopadhyay, Design & development of alow cost tribometer for nano particulate Lubricants, Mater. Today:.Proc. (2020).

16. H. Singh, A.K. Singh, Y.K. Singla, K. Chattopadhyay, Effect of nanofly ash aslubricant additive on the tribological properties of SAE 10W–30 oil: a novelfinding, Trans. Indian Inst. Met. 73 (9) (2020) 2371–2375.

17. H. Singh, A.K. Singh, Y. Singla, K. Chattopadhyay, Tribological study of nano flyash as lubrication oil additive for AISI 4140 steel for automotive engineapplications, Int. J. Mech. Prod. Eng. Res. Develop. (2020.

IMPLEMENTING LEAN TOOL: POKA-YOKE IN TRACTOR COMPONENT MANUFACTURING INDUSTRY

*Deepak Bubber **Gulshan Babber ***Rakesh Kumar Jain
*Faculty of Engineering & Applied Science, Himalayan University, (Arunachal Pradesh) India 791111.
**Chitkara Business School, Chitkara University, (Punjab) India 14040.
***Faculty of Engineering & Applied Science, Himalayan University, (Arunachal Pradesh) India 791111.

Abstract

Several changes have occurred in the manufacturing industry since the post-liberalization after 1991. Original equipmentmanufacturers (OEMs) dominate the organized sectors. Moreover, companies of all sizes including micro, small and medium (MSME) act as mediators by supporting as Tier-1 and Tier-2 suppliers for these OEMs. The lifting of trade restrictions has been blessing in disguise for both Indian manufacturing industry as well as consumers. In recent times, the Indian subcontinent has seen a paradigm shift in manufacturing culture with the advent of many reputed global brands establishing their manufacturing facilities in India. As a result, India's manufacturing culture has adopted modern technological tools and techniques in place of traditional manufacturing methods. Component manufacturers have been forced to modify their working cultures to adapt and adjust to the dynamic requirements such as change in material specifications, tolerances and surface treatments like heat treatment, plating etc. The purpose of this case study is to gauge the outcome after the implementation of one of the lean tool: Poka-Yoke. This tool was invented by a Japanese industrial engineer in 1961, who specialized in quality control. He is also known for introducing various quality techniques, such as Just-in-time production (JIT), Single Minute Exchange of Die (SMED), Jidoka (building quality into the process), Kaizen (continues improvement) etc. Poka-yoke technique was applied in a tractor parts manufacturing company that is a Tier-1 supplier for an OEM. Although the company has achieved quality due to sophisticated technology and superior manufacturing capabilities, inadvertent errors due to human mistakes can still occur. This was resolved by changing fixture design, where human mistake is negligible. However, applyingpoka-yoke helped to eradicate these issues related due to human errors there by leading to reduction in overall customer complaint
Keywords- Poka-Yoke, Lean Manufacturing, Fixture Design, Customer Complaints

Introduction

AIndian manufacturing industry has seen a sea change after the post economic liberalization. Government of India (GOI) took many initiativesin the nineteen eighties, but it was not before 1991 that GOI laid a roadmap to facilitate a systematic shift towards more open economic systemat par with global practices. Furthermore GOI acted as a facilitator to empower private sector to meet global standards through hand holding by introducing schemes for up liftmen of MSME [1]. The Indian auto-components industry can be broadly categorised into organized and unorganized sectors. The organized sector mostly caters to original equipment manufacturers (OEMs) categorized as vendor or supplier. Now-a-days every OEM is having a vendor development cell(VDC) that comprises of specialized persons from design, purchase, quality, assembly etc. This roll of the team is to understand the manufacturing capability of the vendor based on the available plant and machinery, testing and inspection facilities etc. Moreover, this is important to know the capability/capacity of a vendor because the components are made according to the company's drawings and specification.The unorganized sector due to the limited resources focuses on products that caters mostly to the aftermarket category.

The automobile component industry's turnover was ₹3.40 lakh crore (US$ 45.9 billion) in FY21, a 3% decrease from the previous year and is expected to reach US$ 200 billion by FY26. Exports of auto components declined by 8.28% to ₹0.98 lakh crore (US$ 13.3 billion) in FY21, ₹1.02 lakh crore (US$ 14.5 billion) recorded in FY20. As per Automobile Component Manufacturers Association (ACMA), automobile components export from India is predicted to attain US$ 80 billion by 2026. Strong global demand and rejuvenation of the local original equipment and aftermarket segments are predicted to help the Indian auto component industry grow by 20-23% in FY22.
(https://www.ibef.org/industry/autocomponents-india.aspx).

2.ENTERY OF GLOBAL AUTOMOBILE COMPANIES

Globalization and easing trade restrictions had a positive effect on both Indian industry as well as consumers due to direct investments of many global players in the way of establishing manufacturing facilities pan India. Pre liberalization, the consumer had limited choices

due to limited brands available in automobile sector. The list of some global players that established their manufacturing facilities in India from early 1980's to late 1990's are as follows:

Table I
Details of Global Players

Year	Name of the Global Player	Country	Product
1982	Suzuki Motor Corp. Joint Venture with Govt. of India	Japan	Car
1982	Mitsubishi Motors Corp. Joint Venture Eicher Goodearth Ltd	Japan	Light Commercial Trucks
1983	Mazda Motor Corp. Joint Venture with Punjab Tractors Ltd	Japan	Light Commercial Trucks, Buses
1983	Nissan Motor Comp. Joint Venture with Hyderabad Allwyn Ltd	Japan	Light Commercial Trucks
1995	Honda Motor Company Limited	Japan	Cars
1995	Ford Motor Company	USA	Cars
1996	Hyundai Motor India Limited	South Korea	Cars
1995	Daewoo Motors	South Korea	Cars
1997	Toyota Motor Corp. Joint Venture with Kirloskar India Ltd	Japan	Cars
1997	Fabbrica Italiana Automobili Torino (F.I.A.T.)	Italy	Cars

Source: https://www.google.com

The success of these companies laid on the foundation of practicing tool and techniques such Lean manufacturing (replica of Toyota Production System developed by Taiichi Ohno and Eiji Toyoda in 1948 and 1975 respectively), Six Sigma (Introduced by American engineer Bill Smith working at Motorola in 1986), Total Quality Management (TQM was developed by William

Deming, a management consultant).The requirement of meeting closer machining tolerances led to the adoption of these tested methodologies also in Indian components manufacturing industries.

3. METHODOLOGY OF THE STUDY

This case study is based on an Indian MSME that is a tier-1 supplier to a tractor manufacturing company based in the state of Punjab, India. The component under study is a DOL component against which a customer complaint was received through mail citing fitment issues at assembly line. The road map for the study is as follows:
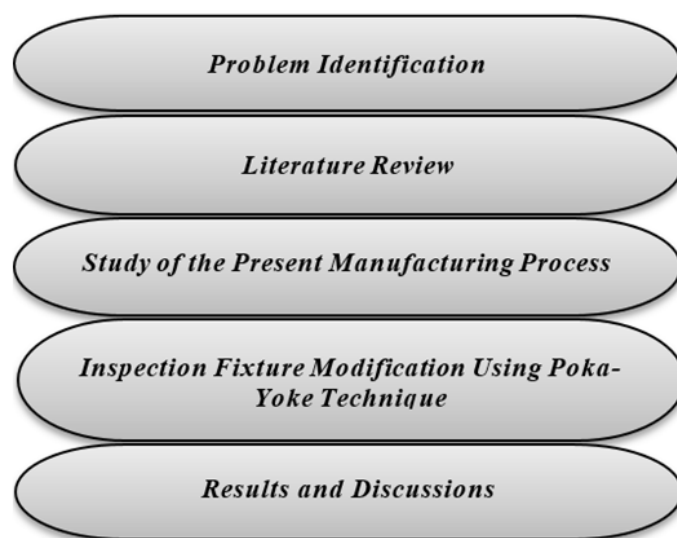
4. COMPANY PROFILE

M/s BI Pvt. Ltd (a company), established in 1980, has earned a good reputation in manufacturing intricate and precise components. While catering to a growing OEM segment, company has achieved excellence through state-of-the-art technology, superior manufacturing capabilities and dedicated personnel. This makes them one of preferred supplier in the domestic markets. However, there are certain issues which act as bottle necks and catch the attention of top management.

The company is ISO: 9001:2015 certified with acceptance as a quality supplier among the auto component manufacturers. The company supplies components to many reputed OME's such as Manindra & Mahindra (Swaraj Tractors Div.), Mahindra & Mahindra (Farm Equipment Div.), Swaraj Engines Ltd., SML Isuzu Ltd, Indo Farm Equipment's Ltd etc. These OEM's are manufacturing Tractors, Harvester Combines, Tractor Engines, Cranes etc. The company started with basic manufacturing technology that comprised of conventional

machines such as lathe machines, universal milling machines, drilling machines, thread rolling machines, thread milling machines etc. The company upgraded its manufacturing setup by adding Computer Numerically Controlled (CNC) Lathe machines, Vertical milling machines (VMC), automatic band saw cutting etc. Post 2005 technological upgradation helped the company to meet customer demand with shorter lead time and enhanced component quality.

5. PROBLEM STATEMENT

The company manufactures around 150 odd components for different OEMs. The component selected for this study is an important child part of hydraulic control valve assembly that is used in tractor lift application. This assembly controls the hydraulic lift functions of the tractor. Presently, the company is suppling 8000 to 10000 pieces per month depending on the customer demand that are direct on line (DOL) assured. During the assembly one piece found out of spec. that lead to line stoppage due to fitment problem. An official mail was received from the ReceiptQuality Assurance (RQA), describing the problem and raising of Corrective action and Preventive action (CAPA) on supplier portal for investigation and closure. (Ref. Mailbox of deepakbubber).

6. LITERATURE REVIEW

6.1 Origins of Lean Manufacturing

Henry Ford, of the Ford Motor Company, and Alfred Sloan, with General Motors, shifted from craft manufacturing to the mass manufacturing in the early part of the twentieth century as a cost reduction approach. Ford's mass-production system focused heavily on production rate rather than the voice of customer. In the mid-twentieth century, the Japanese manufacturer Toyota developed more efficient production methodologies that were based on the deep analysis of western production systems. The Japanese industries lacked resources and their productivity was far less than their western counterparts. Taiichi Ohno, an industrial engineer is credited with developing a new production system that focused on cost reduction, waste elimination, first time right (FTR) by producing quantity as per customer demand [2]. The outcome of the new system made Toyota a pioneer in auto products through demonstration of world class quality at highly competitive prices. [3]. The term Lean became used for the first time in 1988, when an International Motor Vehicle Program was launched to understand the productiveness among Japanese and Western industries. The term was then popularized by Womack, Jones and Roos in their book "The Machine That Changed the World"[4]. The source of Lean Manufacturing came from the Toyota Production System.[5]. Lean is thus a weapon against waste elimination that focuses on both manufacturing inefficiencies and underutilization of people enterprise.

6.2 Lean Manufacturing Competitive Scheme

The GOI had setup "National Manufacturing Competitive Council (NMCC)" in the year 2004 under the Ministry of Micro, Small and Medium Enterprises (MSMEs) with the vision to motivate, facilitate and support the growth of manufacturing industries in India. Under this program following schemes were launched:

1. Technology and Quality Upgradation Support for MSMEs.
2. Credit Linked Capital Subsidy Scheme.
3. IPR Building Awareness On Intellectual Property Rights for MSMEs.
4. Lean Manufacturing Competitiveness Scheme for MSMEs.
5. Design Clinic Scheme for Design Expertise to MSMEs.
6. Incubation Centre Support for Entrepreneurial and Managerial Development of SMEs Through Incubators.
7. Zero Effect Zero Defect.

Mailbox of deepakbubber
Subject: Fw: Hyd. Shop Issue-M/S- BUBBER INDUSTRIES PRIVATE LIMITED
From: bubber industries pvt ltd <info@bubberindustries.com>on Fri, 14 Jan 2022 14:12:31
To: "Deepak Bubber" <deepakbubber@rediffmail.com>
From: SINGH HARJOT – SWARAJ
Sent:Wednesday, October 6, 2021 4:50 PM
To: bubber industries pvt ltd ; Gulshan Babber
Subject: FW: Hyd. Shop Issue-M/S- BUBBER INDUSTRIES PRIVATE LIMITED
Dear Sir,
Kindly find below quality issue raised in Plant-1 by RQA team. Kindly send a representative for resolution of given issue.
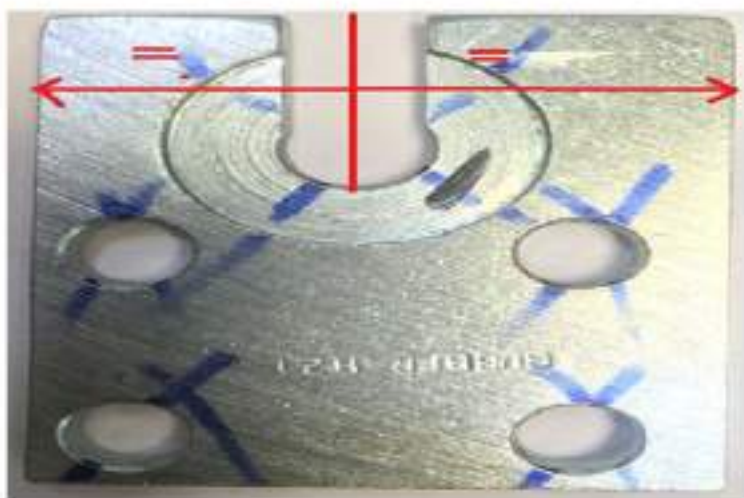Regards,
Harjot Singh

From: KUMAR JATINDER -KUMAR.JATINDER3@mahindraswaraj.com>

Sent: Wednesday, October 6, 2021 2:23 PM
To: KUMAR MAYANK - SWARAJ <KUMAR.MAYANK@mahindraswaraj.com>; BABBAR SUMIT - MVML 123152983@MAHINDRA.COM>; SINGH HARJOT - SWARAJ- <SINGH.HARJOT2@mahindraswaraj.com>
Cc: KAUL RAKESH <KAUL.RAKESH@mahindraswaraj.com>; KUMRA KRISHAN - SWARAJ <KUMRA.KRISHAN@mahindraswaraj.com>; BISHT ANKIT - SWARAJ <BISHT.ANKIT@mahindraswaraj.com>
Subject: Hyd. Shop Issue-M/S- BUBBER INDUSTRIES PRIVATE LIMITED
Dear Sir,
Hyd. Shop Issue of with details below:

Part no	Part Name	Supplier	Required	Observed	Remarks
P108208C	Front Plate for Control Valve	Bubber Industries Pvt. Ltd.	Equal Dim Required 25.00 mm	Equal Dim Observed 27.38 – 22.85 (Slot Shift)	Stock Qty. 2000 pcs, Material Hold

complaint description

You are requested to take up the issue with supplier for necessary actions
Regards
Chandra Prakash Singh
RQA PLANT-1

The information contained in the mail as shown above have the type of quality problem and quantity laying at the customer end. As per mail around 2000 pieces were available for segregation i.e. the first step is to know the exact quantity of the faulty pieces laying in the assembly/store.

Lean Manufacturing Competitive Scheme is one of the salutaryfor the Indian MSME[6]. The objectives of this scheme is to reduce waste, increase productivity, promoting innovative practices for improving overall competitiveness, inculcating good management systems and imbibing a culture of continuous improvement. The scheme is operational since July, 2009. A three-tier structure is proposed for the implementation of the scheme with a group of ten MSME units, called a Mini Cluster. The scheme is being administer through nodal agency National Productivity Council (NPC), New Delhi as a coordinating body or National Monitoring and Implementing Unit (NMIU) and financially supported by GOI towards the cost of conducting awareness programs and implementation of LM techniques [7].

6.3 Lean Manufacturing: The Indian Scenario
Lately, India has become a preferred manufacturing hub for many reputed global brands that cater to domestic as well as global demands. The effects of this situation has helped Indian manufacturer to shelf orthodox production methodsand to adopt to the latest world class technology used for manufacturing [8].Lean manufacturing has become the most popular methodology among manufacturing and service industry that facilitate efficient working of the organization to produce world class quality at highly competitive prices[9][10].

Despite the overall benefits of implementing lean manufacturing
the success among Indian manufacturing industries still remains in the infancy stage, moreover Indian firms especially micro and smallenterprises are notable to reap the benefits derived from its applications due to lack of

resources and management commitments [10][11]. Many practitioners wrongly confine "lean" as an activity limited to the shop floor whereas lean is the way of thinking and attitude. The lean approach can be applied to every stage or situation in the company. The idea is to create a culture that promotes people to continuously innovate and improve their processes that are able to produce defect free products at globally commutative price[12][13].

7. LEAN TOOL: POKA-YOKE

Poka Yoke is another term that was conceived in Toyota, Japan. It was first coined in the 1961 by Shigeo Shingo, who was employed as an industrial engineer responsible for introducing popular systems such as Toyota Production System (TPS) and Just in Time (JIT) that created and streamlined processes to produce Zero defective components. The purpose of poka-yoke is to eliminate defects due to human negligence in products by way of prevention of mistakes in the process execution as early as possible. The original wording was Baka-Yoke is the Japanese term for fool-proofing which was retaliated by Japanese operators, as they seemed to be insulted by the word 'fool'. "In 1963, a lady operator at Arakawa body company refused to use Baka-Yokemethodology in her work station, because the term seemsdisgusting and offensive inkling" inviting Shingo, to rethink and rename the term Baka-Yoke that has the mass acceptance duly invented by him in 1961, keeping in mind the philosophy behind this concept[14].

The philosophythat drives the concept of poka-yoke methodology honors the human values, specifically human intelligence. In the process of repetitive operations which require beingvigilant and focused. Applying poka-yoke helps to reduce valuable human efforts due to occupancy that further is used to the utilize in creativity, innovation and overall value[15]. A defect exists in two states, firstly the defect either has already occurred, that is termed as defect detection, secondly that has the potential to occur in future that is called predictive defect detection[16].

Shingo conceptualized three types of poka-yoke methodology for detection and prevention of errors in a mass production system. The "control method"detects the defects in the product via shape, size, color, or other physical attributes. The "fixed-value method"is used where repetitive activities are made in production that is detected by sensors, counters etc. for missing movements. The "motion-step method"determines the adherence of prescribed steps as per defined process[17][18].

7.1 Examples of Poka Yoke

Poka-Yoke technique is used to eliminate mistakes due to human negligence. Figure 1(a) shows various designs of connectors and cables used in a computerapplication that are not interchangeable with each other i.e. specific connectors for specific applications and use. Figure 1(b) shows how wrong components are rejected through the application of online gauging installed on the conveyor. This process acts as a tollgate that only allows the right components reaches the next station
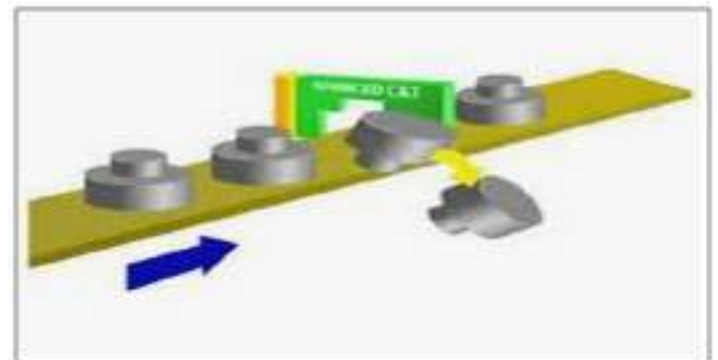

Figure. 1(a) Control Method.

Figure. 1(b) Motion-Step Method

7.2 Mistakes vs Defects

Shingo differentiated mistakes/errors that are mostly inevitable and defect generators. These defects are further passed on through the system undetected that is eventually passed on to the end customers. Poka-Yoke thus acts as mainstay formistake preventionin the process.

The Poka-Yoke techniques are implemented to ensure idealistic conditions in processes by ensuring defect free components before the subsequent operations.The main focus of lean methodology is to avoid any type of wastes; human negligence is one of them i.e. taken care by implementing poka-yoke. In case a mistake or defect cannot be prevented from occurring, Poka Yoke can also be used to come across any defects and rectify them as soon as possible.

7.3Types of Errors in Manufacturing

Poka-Yoke techniques are used in situations or places

that are prone to possible mistakes/errors. Poka-yoke thus can be applied both in manufacturing or services effectively. Some of the possible errorsare as follows:
- Errors in processing.
- Errors with the set up.
- Errors due to missing parts in assembly or welding in manufacturing.
- Errors due to wrong parts used in manufacturing.
- Errors with measurements.

7.4 When to use Poka-Yoke?
There are various types of human errors, some of the error types are listed in Table II along with some potential safeguards.

Table II
Types of Human Errors

| S. N	Type of Errors	Description	Safeguard	
1.	Forgetfulness	Sometimes we forget things when we are not concentrating e.g. A person forgets to set his/her alarm clock at night.	1.	Establish a routine which includes checking before going to bed.
2.	Errors due to a misunderstanding	Sometimes we make mistakes when we jump to the wrong conclusion before we are familiar with the situation e.g. A person not used to automatic transmission steps on the brake, thinking it is the clutch.	2.	Training, checking in advance, work standardization.
3.	Errors in identification	Sometimes we make mistakes when we view it too quickly e.g. A ₹10 note is mistaken for a ₹20 note.	4.	Training, attentiveness, vigilance.
4.	Errors made by amateurs	Sometimes we make mistakes through lack of experience e.g. New operator does not know the operation or is barely familiar with it.	6.	Skill building, work standardization.
5.	Willful errors	Sometimes we make mistakes when we decide we can ignore rules under certain circumstances e.g. Crossing a street against a red light because there are no cars in sight.	8.	Basic education and experience.
6.	Inadvertent errors	Sometimes we make mistakes without knowing how they happened e.g. Someone lost in thought tries to cross the street without noticing the light is in red.	Attentiveness, discipline, work standardization.	
7.	Errors due to slowness	Sometimes we make mistakes when our actions are slowed down by delays in judgment e.g. A person learning to drive is slow to step on the brake.	Skill building, standardization.	
8.	Errors due to lack of standards	Sometimes we make mistakes when there are no suitable standards or work instructions e.g. A measurement may be left to an individual worker's discretion.	Work standardization, work instruction.	
9.	Surprise errors	Sometimes we make mistakes when equipment runs differently than expected e.g. A machine may malfunction without warning.	Total Productive Maintenance, work standardization.	
10.	Intentional errors	Sometimes we make mistakes deliberately e.g. Crimes and industrial sabotage.	Fundamental education, discipline.	

8. CURRENT MANUFACTURING PROCESS

8.1 COMPONENT DRAWING

Company is manufacturing component based on the drawing supplied by the OEM. This drawing contains all the information of the component like component name, material, dimensions, tolerances etc. as shown in figure 2. The development process of the component starts with the submission of an estimated cost against a request for quote (RFQ).After the acceptance of the cost by the OEM,a development order (DO) containing terms and conditions such as annual requirement, warranty claims are listed i.e. accepted by both vendor and vendee. A formal purchase order (PO) is released by the OEM.

A joint activity between vendor, sourcing and supplier quality assurance is done i.e. properly called advance quality planning (AQP) that thoroughly reviews tolerances, critical parameters and checking methodology. Subsequently after the AQP closure the vendor has to submit Production Part Approval Process (PPAP). This is a set of documents as listed below:

1. Part Submission Warrant
2. Process sheet
3. Control Plan
4. Process Failure Mode Effects Analysis (PFMEA)
5. List of instruments and gauges
6. Calibration Plan
7. Process capability study
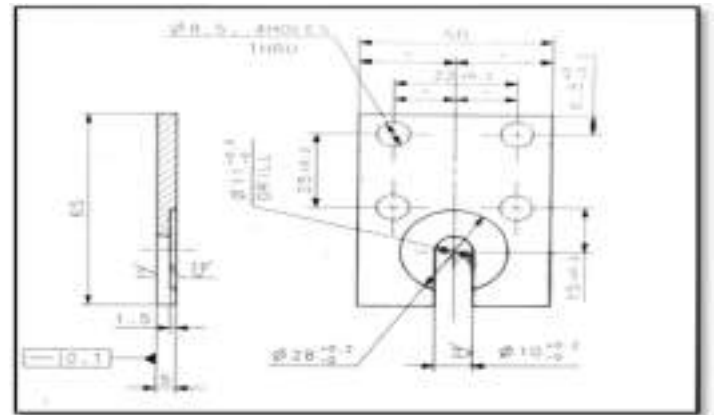
8. Material test report

Figure. 2 Component Drawing

8.2 PROCESS SHEET

Process sheet is an important document which have all manufacturing related information's as shown in figure 3(a) and 3(b). This document is jointly signed by the company and the OEM.

The sequence of process consists of cutting on band saw machine, total length machining on lathe machine, punching holes on power press with the help of punching die (outside job work), thickness grinding on surface grinding machine, counter bore and key way machining on vertical machining center (VMC), deburring and vendor/lot/month code punching on pneumatic press, final inspection with gauge as well as visual for any defect and finally packing and dispatch.

Figure. 3(a) Component Process Sheet

Figure. 3(b) Component Process Sheet

8.3 METHOD OF INSPECTION BEFORE DISPATCH

Final pre dispatch inspection (PDI) is done to ensure the detection of defects such as material fault (crack), missing operations, 100% fitment inspection with approved receiver gauge to ensure defect by SQA and designed by OEM's industrial engineering department is used as the part of quality assurance due to the component having a DOL status.

8.4 THE EFFECTIVENESS OF 100% INSPECTION

The least effective form of quality control is that of inspection by humans, monotonous and long working hours lead to low job engagement resulting in mistakes/defect. However, that does not mean that operator inspection should be dismissed as being of no value, having each and every operator inspect the work that they receive before use and the work that they produce is still an effective way to reduce the possibility of defects reaching the customer, although it is not the best.

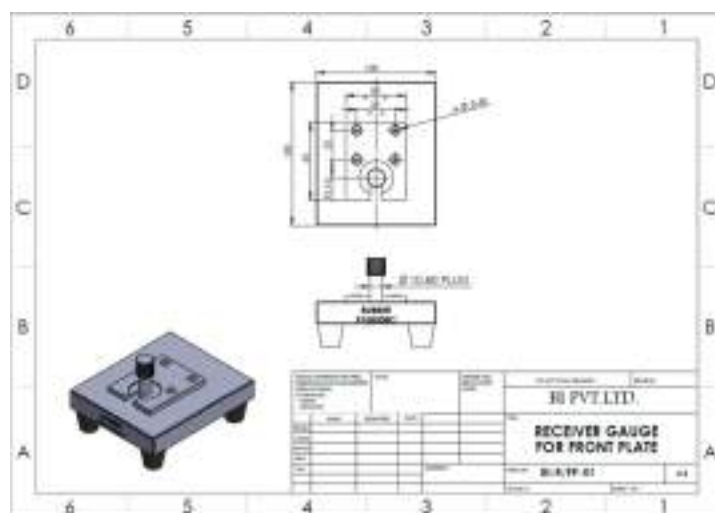
Figure. 4 Inspection Receiver Gauge

8.5 ROOT CAUSE OF REJECTION (COMPLAINT)

After the detailed analysis of the root cause based on the customer complaint a "why-why" analysis was done to analyze the occurrence, detection and the system fail-

ure. The outcome of the why-why analysis and the possible reasons are listed below:

1. OPN 50 (as shown in figure 3(b)) is the final machining of counter bore Ø28+0.2, depth 1.5, Drill Ø8.5 at 4 places and key way 10+0.2. It is observed that during the setting on VMC, first piece was shifted (as per first piece inspection report).
2. This setting piece was mixed with the entire lot due to operator mistake whereas it must be scrapped immediately or should have a red paint mark.
3. Design of final inspection receiver gauge checks only the PCD of holes and the slot.
4. The shift of the hole and slot w.r.t dim 50 i.e. width of the front plate cannot be checked with the receiver gauge (design fault).

5. Final inspection team have a mindset that no such problem observed earlier so they assume that lot is ok.

9. RESULTS AND DISCUSSIONS
9.1 INSPECTIONFIXTURE DESIGN USING POKA-YOKE

The permanent solution of the problem is to make the process so mistake-proof which eliminates the human error. The new and improved design of the inspection receiver gauge that is able to detect shifting of 4 holes and the key way slot w.r.t to component width. Operator/Inspector has to place the component on the receiver gauge as shown in the figure5.Defective pieces will be rejected/detected by the receiver gauge.
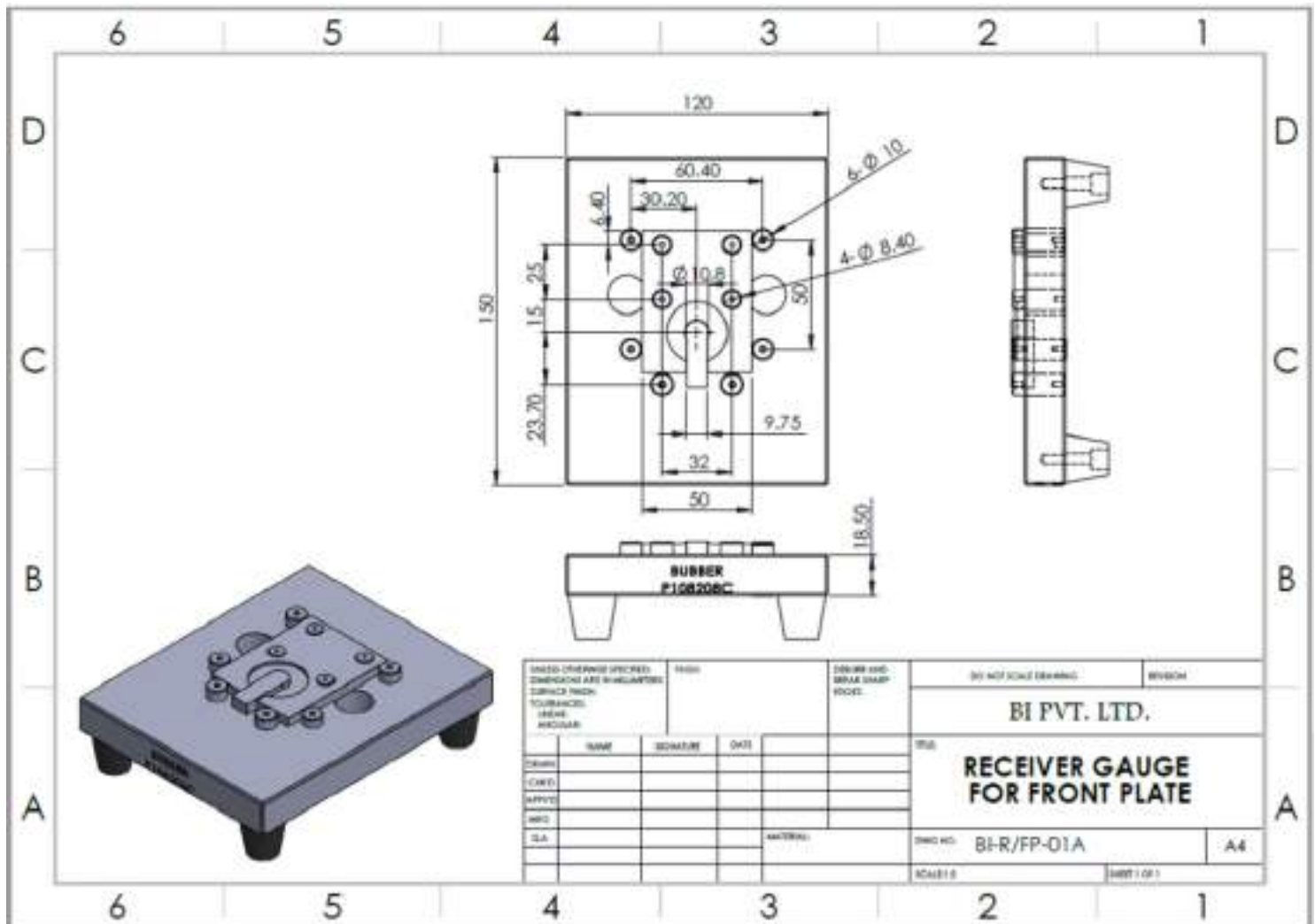

Figure. 5 Inspection Receiver Gauge

CONCLUSION

For the growth and sustainability of any organization, proper emphases have to be given to the voice of customer. The success of any manufacturing company is based on pillars of quality and competitive pricing that caters to the customer needs. The lean manufacturing tools and techniques are very effective use judiciously. In this tool kit, Poka-yoke is one of the most widely used and practiced tool that focuses on eliminating human errors. If a rejected component is returned by the

customer, it directly hits the companies' reputation and customer loyalty.

1-10-100 Rule

By this rule the cost for solve the defect is multiplying 10 times step by step;

1. If a defect found in a job at machine, then material and machining cost is waste.
2. Suppose at assembly station we found defect in the job then material, machining and assembly cost is waste.Which is approx. 10 times the defective job found at machining.
3. If customer found defect in the job then material, machining, assembly, logistic and profit is waste. Which is approx. 100 times the defective job found at machining and assembly.

Due to this we have to minimize the defect or error as early as possible. For that we have to implement proper poka-yoke atparticular stations where chances of errors are there due to mass production[19].

LIMITATION AND SCOPE OF THE FUTURE WORK

This study focuses on the effective use of Poka-yoke technique based on the customer complaint. The problem is solved bymodifying the design of the inspection receiver gauge using poka-yoke technique.

There is future scope to study the machining methods that will help to eliminate/minimize setup rejection caused by machine setters so that component quality is assured first time right and every time right.

REFERENCES

[1]   M. S. Ahluwalia, "Economic reforms in India since 1991: Has gradualism worked?," J. Econ. Perspect., vol. 16, no. 3, pp. 67–88, 2002, doi: 10.1257/089533002760278721.

[2]   A. G M, "A Brief Overview on Toyota Production System (TPS)," Int. J. Res. Appl. Sci. Eng. Technol., vol. 8, no. 5, pp. 2505–2509, 2020, doi: 10.22214/ijraset.2020.5415.

[3]   J. Keyes, "The Need for Lean Training Background of Lean Manufacturing," J. Manag. Policy Pract., vol. 14, no. 3, pp. 78–83, 2013.

[4]   D. Womack, J.P., Jones, D.T., Roos, The machine that changed the world. New York: Macmillan Publishing Company, 1990.

[5]   M. Holweg, "The genealogy of lean production," J. Oper. Manag., vol. 25, no. 2, pp. 420–437, Mar. 2007, doi: 10.1016/j.jom.2006.04.001.

[6]   "MSME Annual Report 2020-21," www.msmse.gov.in, 2021.

[7]   S. J. Thanki and J. Thakkar, "Status of lean manufacturing practices in Indian industries and government initiatives: A pilot study," J. Manuf. Technol. Manag., vol. 25, no. 5, pp. 655–675, 2014, doi: 10.1108/JMTM-05-2012-0057.

[8]   R. Narain, R. C. Yadav, and J. Antony, "Productivity gains from flexible manufacturing: Experiences from India," Int. J. Product. Perform. Manag., vol. 53, no. 2, pp. 109–128, 2004, doi: 10.1108/17410400410515025.

[9]   P. Saraswat, M. K. Sain, and D. Kumar, "A Review on Waste Reduction through Value Stream Mapping Analysis," Int. J. Res., vol. 1, no. 6, pp. 200–207, 2014.

[10]   R. Kumar and V. Kumar, "Lean manufacturing in Indian context: A survey," Manag. Sci. Lett., vol. 5, no. 4, pp. 321–330, 2015, doi: 10.5267/j.msl.2015.2.009.

[11]   A. Agarwal, "Lean manufacturing in India," https://www.projectguru.in.

[12]   D. Basak, T. Haider, and A. K. Shrivastava, "Lean Manufacturing – An Advanced Manufacturing Technique in Modern Operations Management," IJCSMS Int. J. Comput. Sci. Manag. Stud. ISSN 2231 –5268 www.ijcsms.com, vol. 13, no. 05, pp. 7–8, 2013.

[13]   D. D. Hari Parshad , Rajesh Kumar, "Benefits of Lean Manufacturing System in Indian Industry-A Study," Int. J. Eng. Sci. Res. Technol., no. January, pp. 76–82, 2017. ISSN: 2277-9655

[14]   P. Thareja, "Poka Yoke: Poking into Mistakes for Total Quality!," OmniScience A Multi-disciplinary J., vol. 6, no. 2, pp. 1–8, 2016.

[15]   S. Patel, B. G. Dale, and P. Shaw, "Set-up time reduction and mistake proofing methods: An examination in precision component manufacturing," TQM Mag., vol. 13, no.3,pp.175–179,2001,doi: 10.1108/09544780110385528.

[16]   M. Dudek-Burlikowska and D. Szewieczek, "The Poka-Yoke method as an improving quality tool of operations in the process Industrial management and organisation," J. Achiev. Mater. Manuf. Eng., vol. 36, no. 1, pp. 95–102, 2009.

[17]   P. Parikshit, V. Nair, and P. Nikunj, "A Review on use of Mistake Proofing (Poka Yoke) Tool in Blow Molding Process," Int. J. Sci. Res., vol. 2, no. 2, pp. 2319–7064, 2013, [Online]. Available: www.ijsr.net.

[18]   H. ewita, F. Ali Tosa, Y. Santoso, L. Herliani Kusumah, and H. yetti, "Application Poka-Yoke to Capture Defect (A Case Study in Industry Component Otomotive)," Int. J. Ind. Eng., vol. 6, no. 1, pp. 14–17, 2019, doi:

10.14445/23499362/ijie-v6i1p103.

[19]   A. K. J. Sushant S Shrigadi, Shivalee P. Pol, "International Journal of Innovation in Engineering, Research and Technology," in Poka:Yoke:Mistake Proofing to achive Zero Defect, pp. 1–6, ISSN No - 2394-3696

Faculty of Engineering & Applied Sciences
Desh Bhagat University,
Mandi Gobindgarh, Punjab