Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606

Volume 08 Issue 01
2024

Research Article
https://doi.org/10.56979/801/2024

Sentiment Analysis Classification of ChatGPT Tweets Using Machine Learning and Deep Learning Algorithms

Imtiaz Hussain[1], Syed Muhmmad Hassan Zaidi[1], Usman Khan[2], and Adnan Ahmed[3]

[1]Department of Artificial Intelligence and Mathematical Sciences, SMIU, Karachi-74000, Pakistan.
[2]Department of Computer Sciences, Karachi Institute of Economics and Technology, Karachi, Pakistan.
[3]Department of Computer Sciences, Bahria University Karachi, Pakistan.
[*]Corresponding Author: Syed Muhammad Hassan Zaidi. Email: m.hassan@smiu.edu.pk

Received: June 02, 2024 **Accepted:** September 30, 2024

Abstract: Sentiment analysis is essential for understanding public opinion and emotional responses to specific topics. In this study, we conduct sentiment analysis on a dataset comprising tweets related to ChatGPT. The dataset includes two primary columns: tweets and sentiment labels (positive, negative, and neutral_l). We developed and evaluated machine learning (ML) models to classify these tweets' sentiments. To preprocess the data, we applied standard text cleaning techniques such as removing special characters, tokenization, and stop word removal. The textual data was converted via Count Vectorizer to numerical features, and the labels were encoded using Label Encoder to transform categorical sentiment labels into numerical values. The Convolutional Neural Network (CNN) captured sequential patterns within the tweets and achieved a noteworthy accuracy of 88.25%. The Long Short-Term Memory (LSTM) network has captured temporal dependencies and yielded an accuracy of 89.24%. Logistic Regression (LR) achieved an accuracy of 85.74%, while Decision Tree (DT) and Multinomial Naive Bayes (MNB) models achieved 71.60% and 67% accuracy, respectively. The results demonstrate the efficacy of machine learning models, particularly CNN and LSTM, in accurately classifying the sentiment of ChatGPT-related tweets and effectively capturing sequential and temporal characteristics of social media text, offering insights into public sentiment towards ChatGPT. Our findings have practical implications for understanding user feedback on ChatGPT to enhance its performance and user experience on social platforms.

Keywords: Sentiment Analysis; ChatGPT Tweets; Machine Learning (ML); CNN; LSTM.

1. Introduction

The field of sentiment analysis computes individuals' views, emotions, assessments, and attitudes expressed on social media platforms and other online resources [1]. Generative artificial intelligence understands the inflexion points of factual dependability and the potential of OpenAI Inc. An AI-based conversational agent delivers contextually relevant responses to complex inquiries using advanced natural language processing (NLP) techniques [2]. As social media platforms have become integral to daily life, offering a venue for individuals to express their thoughts and emotions, the need to accurately understand public sentiment toward specific topics, products, or services has grown in importance for businesses, organizations, and researchers. Computation techniques were potentially used to derive the quantitative values of the people who shared their experiences with Twitter platforms regarding various online/offline products and services for further use by individuals, organizations, and others [1]. In the current era, the advent of large language models (LLMs) like ChatGPT has redefined human-AI interaction, making ChatGPT a key figure in conversational AI. It emphasizes the techniques of NLP to generate human-like responses, fundamentally altering perceptions of AI communication and analyzing public sentiments from tweets related to ChatGPT offers insights into user engagement and opinions about this AI system, which is critical for improving AI-human interaction. In tweet data, sentiments paired with positive, negative, or neutral for the classification of public opinion related to ChatGPT, further we removed the special characters and stopped words, and used Count Vectorized to convert textual data into numerical form [3]. For categorical sentiments, we applied a label encoder for numerical values. The CNN model was applied to classify the sentiment of ChatGPT-related tweets for image [4], object detection [5], and speech recognition [6]. The architecture behind CNNs makes it superior as compared to traditional fully connected networks [7]. The LSTM networks are used to capture sequential information from text data. Both CNN and LSTM models demonstrate remarkable accuracy, underscoring their effectiveness in sentiment classification [8]. An LR model, a DT Classifier and an MNB model were used to rule out the insights from the sentiments of ChatGPT-related tweets [9]. LSTM network was integrated to capture the temporal dependencies and contextual relationship [10] from the sequential data allowing to account for the chronological structure of the text and enabling a nuanced analysis of expressed tweets related to ChatGPT [11]. LSTM models have been demonstrated to outperform traditional Recurrent Neural Networks (RNNs) in context-free and context-sensitive languages [12]. Moreover, researchers successfully implemented LSTM and RNNs for sequential and labelled tasks for language modelling whereas, RNNs have shown substantial reductions in perplexity when compared to standard n-gram models [13-15] .

Various researchers used supervised and unsupervised learning techniques to extract the various meaningful features and sentiments from the documents, sentences, and phrases. A supervised learning approach was used by [16], to extract the sentiments on movie reviews through document-level study utilizing Support Vector Machines (SVM), Naïve Bayes (NB), and Maximum Entropy to the feature spaces. The development of opinion mining systems reflects many challenges due to the complexity of the relevancy of text contents of the natural language and its orientation [17]. Opinions are often defined as expressions of positive or negative sentiments, viewpoints, emotions, or appraisals of a subject.

In mathematical terms, an opinion is represented as a quintuple (ej, ajk, soijkl, hi, tl), where ej is the entity, ajk represents the k-th attribute of the entity, soijkl is the opinion sentiment score, hi represents the opinion holder, and tl denotes the time at which the opinion is expressed. The sentiment values can be positive, negative, neutral, or even more nuanced, such as a range of intensities [18].

An assessment method was conducted by [19] to analyze the Twitter messages of users who were interested in Hillary Clinton and Donald Trump. An LR machine learning model was implemented to predict tweet polarity. Results revealed that SVM and Naïve Bayes outperformed and achieved an accuracy of 85.23%. Similarly. [20-21] Applied supervised learning techniques for automatic sentiment classification in the context of highly noisy customer feedback data and found interesting results which were later used for growing the customer volume for businesses. Further motivated the necessitation of an automated system capable of processing large-scale noisy data. A key step in their methodology was the selection of appropriate features for sentiment analysis before classifier implementation.

In continuation of the previous findings proposed in their research to cope with the challenges in text categorization and to address the high dimensionality of feature spaces [22]. Furthermore, they suggested that the features which are used in the text categorization are typically based on a bag-of-words (BoW) model such as unigrams or n-grams, with the dimensionality of the feature space determined by the vocabulary size of the corpus. A simple but more effective unsupervised learning algorithm approach was used by [23] for classifying online reviews and used binary outcomes either recommended (thumbs up) or not recommended (thumbs down).

2. Materials and Methods

2.1. Performance Evaluation of Sentiment Analysis Models:

The performance of sentiment analysis was evaluated using CNN, LR, DT Classifier, Multinomial NB, and LSTM Networks to assess the efficacy of models (Table 1). The other key metrics were examined to provide more nuanced and comprehensive findings. The precision, recall, and the F1 score depict valuable insights into the models' ability to balance false positives (FP) and false negatives (FN) used to facilitate a deeper understanding of classification effectiveness.

2.1.1. Convolutional Neural Network (CNN)

The CNN model exhibits impressive performance in the sentiment classification of ChatGPT-related tweets, achieving an accuracy of 88.25% on the test dataset. This prominent level of accuracy underscores the model's capacity to effectively capture the sequential patterns and contextual relationships within the tweet text, enabling it to accurately discern and classify sentiment.

2.1.2. Logistic Regression (LR)

The LR model demonstrates robust performance in the task of sentiment classification, achieving an accuracy of 85.74% on the test dataset. This result highlights the model's effectiveness in accurately classifying sentiment, despite its simplicity. LR is widely regarded for its interpretability and computational efficiency, attributes that render it well-suited for sentiment analysis applications.

2.1.3. Decision Tree Classifier

The DT Classifier attained an accuracy of 71.60% on the test dataset, a comparatively lower performance than that of the CNN and Logistic Regression models. However, D T offer valuable insights into the hierarchical structure of features, enabling an understanding of their relative importance in the classification process. As an interpretable model, the DT Classifier is capable of capturing non-linear relationships between features and sentiments in ChatGPT tweets, making it a useful tool for exploring complex feature interactions in sentiment analysis.

2.1.4. Multinomial Naive Bayes

An accuracy of 67% on the test dataset was achieved by Naive Bayes classifiers under the assumption that feature independence characteristics enhanced their computational efficiency and rendered them well-suited for text classification tasks. Due to the simplicity of the Multinomial Naive Bayes model, it demonstrates an acceptable performance for the test data making it a viable option for handling large-scale text data in a computationally efficient manner.

2.1.5. Long Short-Term Memory (LSTM) Networks

The LSTM networks captured the temporal dependencies and contextual information inherent within the sequential tweet data. It demonstrated a superior performance with an accuracy of 89.24% surpassing the other models. In the chronological order of the textual content, the LSTM network provides a more nuanced and sophisticated analysis of sentiment in ChatGPT-related tweets, highlighting its effectiveness in handling complex sequential data for sentiment classification tasks.

Table 1. The performance of models and their accuracy

Algorithm	Accuracy
Convolutional Neural Network (CNN)	88.25%
Logistic Regression (LR)	85.74%
Decision Tree (DT) Classifier	71.60%
Multinomial Naive Bayes (NB)	68.00%
Long Short-Term Memory (LSTM) Networks	89.24%

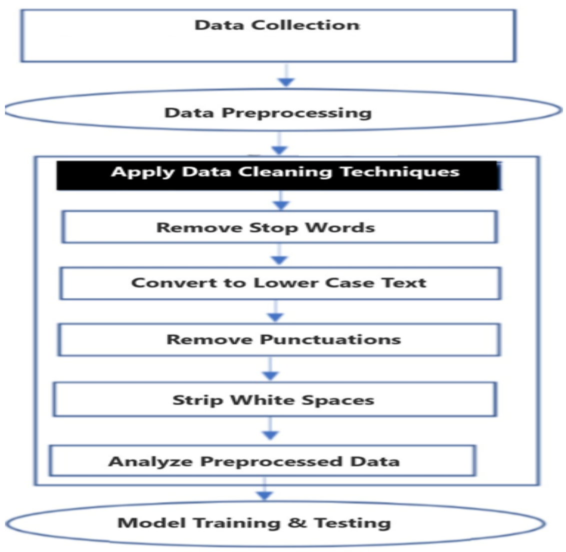

Figure 1. Flow diagram of Data Preprocessing

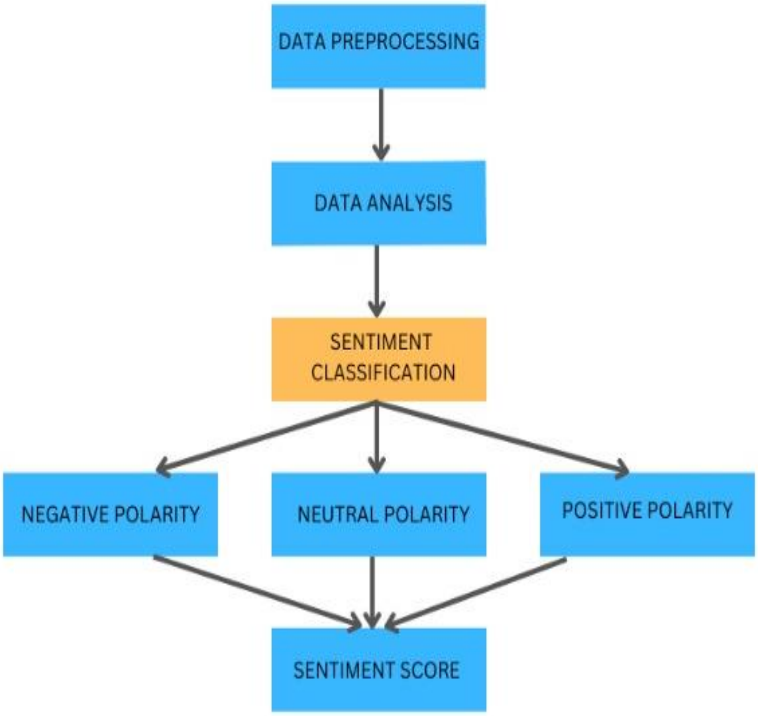

Figure 2. Flow diagram of Sentiment Analysis

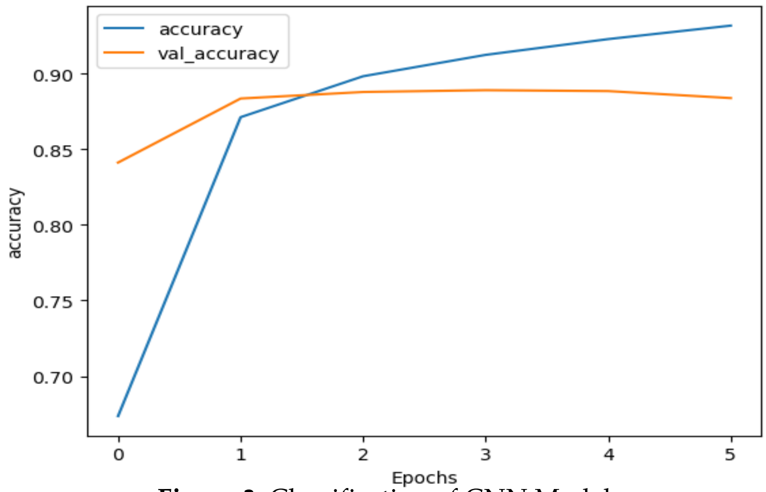

Figure 3. Classification of CNN Model

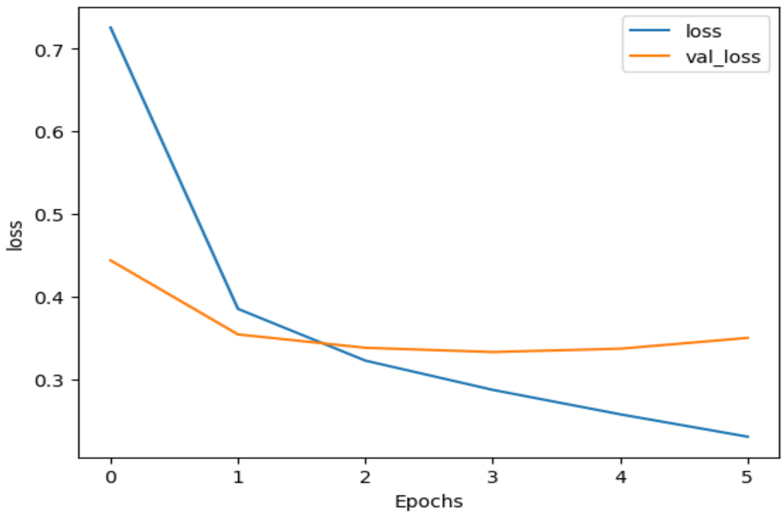

Figure 4. Classification LR Model

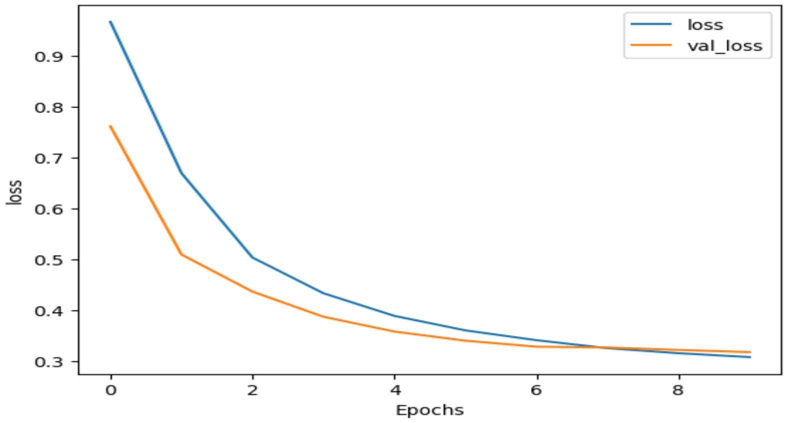

Figure 5. Classification LSTM

	precision	recall	f1-score	support
0	0.89	0.92	0.90	22962
1	0.88	0.86	0.87	18667
2	0.79	0.78	0.78	18078
accuracy			0.86	59707
macro avg	0.85	0.85	0.85	59707
weighted avg	0.86	0.86	0.86	59707

Figure 6. Classification report LR Classifier

	precision	recall	f1-score	support
0	0.80	0.82	0.81	22962
1	0.71	0.68	0.70	18667
2	0.61	0.62	0.62	18078
accuracy			0.72	59707
macro avg	0.71	0.71	0.71	59707
weighted avg	0.72	0.72	0.72	59707

Figure 7. Classification report DT Classifier

	precision	recall	f1-score	support
0	0.85	0.72	0.78	22962
1	0.56	0.88	0.69	18667
2	0.63	0.39	0.48	18078
accuracy			0.67	59707
macro avg	0.68	0.66	0.65	59707
weighted avg	0.69	0.67	0.66	59707

Figure 8. Classification report Multinomial Naive Bayes

3. Results

The sentiment classification was evaluated using multiple metrics, including accuracy, precision, recall, and the F1 score performance using machine learning models. Results revealed efficacy and accurately classified sentiments in ChatGPT-related tweets. CNN model demonstrated a notable accuracy of 88.25% on the test dataset, which highlights its capacity to capture sequential information embedded in tweet text. Whereas the LR model performed robustly and achieved an accuracy of 85.74%, further affirming its suitability for sentiment classification tasks. Further, the DT Classifier, though comparatively less effective in achieving accuracy, represented 71.60%, while the Multinomial Naive Bayes model reached an accuracy of 67% on the test dataset.

Notably, the LSTM network is adept at the temporal dependencies in sequential data, achieving the highest accuracy of 89.24%. This result underscores the value of utilizing the chronological order of tweet text for enhanced sentiment analysis, making the LSTM model the most effective among those evaluated.

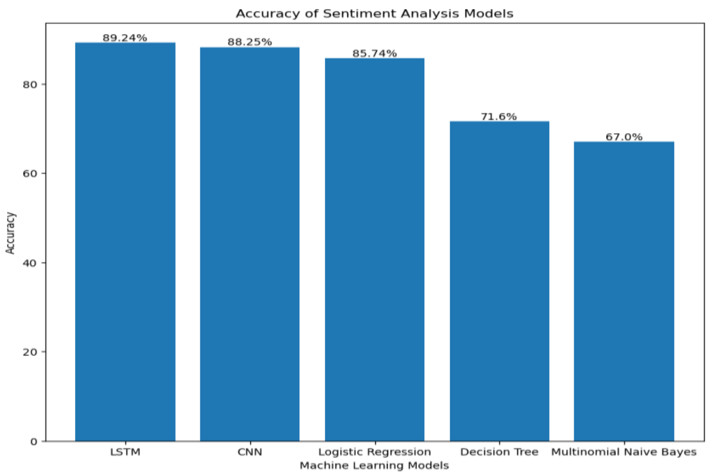

Figure 9. Models Accuracy Bar Chart

4. Discussion

Our results affirm the effectiveness of machine learning algorithms in the sentiment analysis of ChatGPT-related tweets. Among all models, CNN demonstrated superior accuracy, underscoring its ability to capture the sequential nature of tweet text and extract salient patterns for sentiment classification. The LR model also performed commendably, indicating its robustness and reliability in handling sentiment classification tasks. When we incorporated the LSTM network for capturing the temporal dependencies inherent in the tweet data. The strong performance highlighted by the LSTM considering the chronological sequence of text when conducting sentiment analysis, particularly in datasets characterized by dynamic and sequentially dependent information. Our findings contribute to a more nuanced understanding of public sentiment toward ChatGPT, offering valuable insights into user perceptions and interactions with this advanced language model. The ability to accurately classify sentiment from user-generated content, such as tweets, provides actionable insights for developers, researchers, and businesses aiming to enhance the performance and user experience of ChatGPT. Further, these findings illustrate the practical value of machine learning techniques in social media sentiment analysis, particularly in the context of evolving conversational AI systems.

5. Conclusions

Notably, the LSTM network is adept at the temporal dependencies in sequential data, achieving the highest accuracy of 89.24%. This result underscores the value of utilizing the chronological order of tweet text for enhanced sentiment analysis, making the LSTM model the most effective among those evaluated.

References

1. Liu, B. Sentiment analysis. (2015). “Mining opinions, sentiments, and emotions”. Cambridge University Press.
2. Aydın, Ö., & Karaarslan, E. (2022). “OpenAI ChatGPT generated literature review: Digital twin in healthcare”. Available at SSRN 4308687.
3. Hassan, Syed Muhammad, Imtiaz Hussain Haque Nawaz, Basit Hassan, and Mashooque Ali Mahar. "Impact of Climate Change on Fish Species Classification Using Machine Learning and Deep Learning Algorithms.” 2024, 14(2): 9-19
4. Wang, J., Lin, J., and Wang, Z. (2016). “Efficient convolution architectures for convolutional neural network”. In Wireless Communications and Signal Processing (WCSP), 8th International Conference, pages 1-5.
5. Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). “ImageNet classification with deep convolutional neural networks.” In Advances in neural information processing systems, pages 1097-1105.
6. Szegedy, C., Toshev, A., and Erhan, D. (2013). “Deep neural networks for object detection”. In Advances in neural information processing systems, pages 2553-2561.
7. Timoshenko, D., and Grishkin, V. (2013). “Composite face detection method for automatic moderation of user avatars”. Computer Science and Information Technologies (CSIT'13).
8. Sainath, T. N., Kingsbury, B., Mohamed, A., Dahl, G. E., Saon, G., Soltau, H., Beran, T., Aravkin, A. Y., & Ramabhadran, B. (2013). “Improvements to deep convolutional neural networks for LVCSR”. ArXiv. https://arxiv.org/abs/1309.1501
9. Luo, X., Shen, R., Hu, J., Deng, J., Hu, L., and Guan, Q. (2017). “A Deep Convolution Neural Network Model for Vehicle Recognition and Face Recognition”. Procedia Computer Science 107: pages 715-720.
10. Uçar, A. (2017). “Deep Convolutional Neural Networks for facial expression recognition”. In Innovations in Intelligent Systems and Applications (INISTA), IEEE International Conference, pages 371-375.
11. Nebauer, C. (1998). “Evaluation of convolutional neural networks for visual recognition”. IEEE Transactions on Neural Networks, 9(4): pages 685- 696.
12. Gers, F. A. and Schmidhuber, J. (2001). “LSTM recurrent networks learn simple context free and context-sensitive languages, IEEE Transactions on Neural Networks, 12(6), pages 1333–1340.
13. Mikolov, T., Karafiat, M. L., Burget, J. Cernock, Y, and Khudanpur, S. (2010). “Recurrent neural network-based language model”. In Proceedings of International Speech Communication Association, 9, pages 1045–1048.
14. Hassan, S. M., Ali, F., Wasi, S., Javeed, S. Hussain, I. and Ashraf, S.N. (2019). “Roman-Urdu news headline classification with IR models using Machine Learning Algorithms”. Indian Journal of Science and Technology, 12(35), pages 1-9. DOI: 10.17485/ijst/2019/v12i35/146571
15. Pang, B., & Lee, L. (2004). “A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts”. ArXiv. https://arxiv.org/abs/cs/0409058
16. Pang, B. and Lee, L. (2008). “Opinion mining and sentiment analysis”. Found. Trends Inf. Retr. 2(1-2), pages 1–135.
17. Liu, B. (2011). “Web data mining: exploring hyperlinks, contents, and usage data”. Springer Science & Business Media.
18. Razia, S. A., Jaithunbi, A. K., and Sai, R.L. (2018). “Sentiment analysis in Twitter data using data analytic techniques for predictive modelling”. Journal of Physics: Conference Series, 1000, page 12130.
19. Gamon, M. (2004). “Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis”. Proceedings of the 20th International Conference on Computational Linguistics, page 841, Geneva, Switzerland.
20. Yang, Y. and Pedersen, J. O. (1997). “A comparative study on feature selection in text categorization”. In International Conference on Machine Learning, pages 412–420.
21. Turney, P., and Pantel, P. (2010). “From Frequency to Meaning: Vector Space Models of Semantics”. Journal of Artificial Intelligence Research 37, pages 141-188.
22. Shaukat, S.S., Siddiqui, I. A., and Imtiaz Husain. (2006). “Essentials of Mathematical Ecology, Computer Programs in BASIC, FORTRAN, and C++”, Farquan Publisher., University of Karachi Publication, Karachi. 243pp, 2005 1, 350
23. Zoha, A., Ejaz, F., Tabbasum, T., & Sheikh, M. N. (2024). Prevalence of the Trigger Finger in Barbers of Multan. Journal of Computing & Biomedical Informatics.