Classifying Student Learning Styles Using Machine Learning: A DistilBERT-Based Approach

Fong Yee Lee 1, 2, a), S Prabha Kumaresan 1, 3, b), Chee Kong Wong 4, c) and Mohammad ShadabKhan 1, 5, d)

1Faculty of Computing and Informatics, Multimedia University, Persiaran Multimedia, 63100 Cyberjaya, Selangor, Malaysia

2Center of Digital Innovation, CoE for Immersive Experience, Multimedia University, Persiaran Multimedia, 63100 Cyberjaya, Selangor, Malaysia

3Center of Natural Language Processing (NLP), CoE for Artificial Intelligence, Multimedia University, Persiaran Multimedia, 63100 Cyberjaya, Selangor, Malaysia

4School of Diploma and Professional Studies, Taylor’s College, 1, Jalan Taylors, 47100 Subang Jaya, Selangor, Malaysia

5Center for Advanced Analytics, CoE for Artificial Intelligence, Multimedia University, Persiaran Multimedia, 63100 Cyberjaya, Selangor, Malaysia

*a) Corresponding author: lee.fongyee@mmu.edu.my  
b)prabha.kumaresan@mmu.edu.my*

*c)cheekong.wong@taylors.edu.my   
d)shadab.khan@mmu.edu.my*

**Abstract.** This study investigates classifying student learning styles (Auditory, Visual, Kinesthetic) based on textual input using the DistilBERT transformer model. This study is needed because traditional methods like observation of individual learners and manual assessment of learning styles are not only time-consuming but also subjective. In order to tackle the issues addressed, the study employed Natural Language Processing (NLP) for automated classification, using DistilBERT and comparing its performance to a past study that used an LSTM model with GloVe embeddings. The results of the study demonstrate DistilBERT's performance is better than LSTM’s: (i) achieving higher accuracy (98.06% vs. 97.84%); and (ii) achieving lower loss (0.0593 vs. 0.0609). These findings shows that DistilBERT’s outperformed LSTM in capturing contextual nuances in text for the classification task. The methodology included: (i) preparing a dataset of text samples labelled with learning styles; (ii) exploratory data analysis; (iii) feature engineering to optimize the text for machine learning. The DistilBERT model is then trained and fine-tuned for sequence classification. The evaluation metrics show the model can accurately classify new and unseen sentences. This approach informs educators on individual learning styles and would help in create more personalised learning experience that adapt to each learner’s needs. In-addition, the study also details the data preparation, data cleaning, data labelling processes as well as the experimental setup. Besides these, the fine-tuning procedures and evaluation metrics used to assess the models were also explained in detail. The comparison between DistilBERT and LSTM models highlights how the models handle the complexities of language for learning style classification.

# INTRODUCTION

Machine Learning is transforming education by offering personalised learning to cater different needs and learning styles. Educator must understand the learner’s preferred learning styles able to curate the personalised education content. Educators usually will do the manual assessment via survey and in-class observation to find out individual student learning preferences and all these take times and the result can be subjective [1], [2]. Automation on the learning style assessment is made possible now by inferring preferences from written sentiment due to the advancements in Natural Language Processing (NLP) [3],[4]. This study investigates the potential of deep learning using transformer architectures to classify learning styles based on natural language inputs in hoping this will eventually contribute to advance adaptive learning. The tremendous development of Artificial Intelligence (AI) has pushed instructional technology to new heights especially with the surge of Generative AI uses in education. According to the authors of study [5], in academic environments, nearly 92% of the students use AI tools in education. A significant percentage of academic scholars and educators have also been adopting AI to support research, course content creation, assessment, curriculum development. An AI expert, Andrew Ng co-founded an edtech, named Kira Learning that provides AI-powered teaching solutions [6]. The intelligent platform effortlessly generating teaching materials, automating class organization, simplifying assessments and enabling individualised learning based on student performance analytics. These help to reduce educator’s workloads, allowing them to have more time for one-one-one coaching which enabling them to focus on enhancing student understanding and foster an adaptive and personalised learning system. Several studies have shown that AI-driven models can categorize different learning styles utilising the machine learning techniques. These algorithms utilising decision trees and Convolution Neural Networks (CNN) to determine the learner’s preferred learning methods and curate the content accordingly. These studies successfully classified learners into their respective learning styles with high accuracy using Natural Language Processing (NLP) techniques [3],[7]. For instance, named entity recognition and Long Short-Term Memory (LSTM), a classical type of Recurrent Neural Network (RNN) designed to remember on information for long periods while avoiding the vanishing gradient problem. Beyond these studies, AI-driven learning style prediction models have many potentials in online education which foster better and more effective learning experiences. In fact, AI-based prediction models are capable of detecting discern learning styles from the analysis of distinctive characteristics like attention patterns, meditation levels, cognitive workload, facial expressions and even emotional states [8], [9]. With these additional attributes included in the analysis of individual learner preferences, AI-based prediction models can be further refined to deliver truly personalised educational content that is tailored to each learner’s need. The objective of this study is to investigate learning styles using a state-of-art transformer-based deep learning model particularly DistilBERT in the classification of student learning styles from textual data. This approach seeks to enhance previous AI-based prediction models that utilised LSTM networks model with GloVe embeddings [10],[11],[12]. DistilBERT is optimised and suited for the learning styles classification in resource-limited educational environments because of its compact size and speed. By leveraging cutting-edge transformer’s NLP capabilities, this study endeavours to create an efficient model that can predict learning styles with greater accuracy and improved learning styles identification.

# Literature Review

The study of learning styles has undergone substantial development over the years. It was initially tracked using manual surveys and continual behavioural studies where the learner’s preferences were deduced from the rule-based and statistical models. However, such manual critical examination has inherent limitations and the validity of its results is questionable. Kirschner and C. Hendrick (2024) had studied the validity of these research methods on learning styles and pointed out the limitations of these traditional approaches [13], [14]. Although the traditional methods are deemed fundamental, these early approaches were found inflexible and failed to account for the dynamic nature of individual learning preferences. Machine learning techniques play an important role in learning style classification as technology developed over the years. According to García et al. (2019), their research showed machine learning algorithms of Random Forest (RF), Support Vector Machines (SVMs) and shallow neural networks may be effective at predicting learning styles given variety of features [15],[16]. These AI-based prediction models were superior to early approached because they manage to handle complicated, non-linear relationships in data for better generalisation of the findings. Even though these models have outperformed the preceding methods, they are limited to capture nuances in textual data and often required substantial manual feature engineering.

The introduction of Bidirectional Encoder Representations from Transformers (BERT) has brought about a transformative era of NLP tasks. With the BERT’s deep bidirectional training, text classification became more accurate through a more profound comprehension of the context relationship. The transformer architecture empowers AI models to take into account of a word’s entire context by simultaneously examining words that come before and after it; thereby capturing the linguistic nuances [17],[18]. This sophisticated deep learning approach to understanding words and their contextual meanings potentially enhance the field of learning style classification. Although transformer models are widely used in many NLP tasks, their applications in specific domain of learning style classification remained largely underexplored. The work of Silva Barbon and Akabanes’ (2022) used BERT and its variants for automatic text classification across diverse languages emphasising the models' inherent adaptability and versatility [19],[20]. Despite of many promising success, comprehensive research on learning style classification with transformer architectures is still limited. To address this gap, this study proposes the use of DistilBERT, a highly efficient distilled version of BERT to categorise learning styles based on textual data. DistilBERT is a lightweight yet powerful advanced version of BERT. DistilBERT is 40% smaller in size and 60% faster in speed compares to BERT while retaining 97% of BERT’s robust language understanding capabilities [21],[22]. This makes DistilBERT is well-suited for real-time applications and environments with limited computational resources. With DistilBERT’s proven effectiveness and deep contextual understanding, this study aims to achieve improvement both the accuracy and scalability of learning style classification thereby contributing to adaptive learning technologies.

# Research Method

The methodological framework comprised six phases which systematically optimized model performance. These start from data preparation and exploratory analysis for linguistic understanding to feature engineering and fine-tuning on DistilBERT architecture, model training and evaluation that finally led to learning style predictions.

## Dataset Preparation

Dataset preparation for model training and evaluation was the initial step of this research. The dataset was originated from Kaggle [11], previous researcher utilized an LSTM model with GloVe word vectors for learning style classification. The downloaded CSV file comprising 15,450 sentences labelled with three distinct learning styles- Auditory, Visual and Kinesthetic (AVK). Data preparation also include removing duplication items to ensure data quality. All texts were converted to lowercase to make sure they are consistent. Data cleaning procedures implemented including (i) lemmatization to reduce words to their base forms; (ii) stop word removal to eliminate noise; and (iii) use of regular expression to eliminate non-alphabetic characters. Additionally, sentences were split into individual word after tokenisation was applied and this forms a cleaned word list for every sentence. This tokenised data was used for feature engineering and exploratory data analysis later. These essential preprocessing steps were crucial for standardising the input data to ensure that the model would not be skewed by noise or inconsistency in the textual input.

## Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) was employed to learn about the distribution and structure of the dataset. The frequencies of AVK classes were counted. In-addition, a sentence length was examined to see the maximum sequence length during the tokenisation for model training. Besides, word frequencies of each learning styles was counted and both raw word counts and Term Frequency-Inverse Document Frequency (TF-IDF) scores were generated. These dual approaches highlight the most frequent occurring and statistically significant words associated with each learning style. The linguistic features inherent to each class were learned through the textual EDA which gave the model the semantic richness it needed for accurate and nuanced classification.

## Feature Engineering

In this study, the main goal of feature engineering was to improve the representational quality of textual data. It starts with raw word frequencies were calculated to provide a simple statistical overview of learning style occurrences. TF-IDF scores, a crucial indicator was computed to reflect the significance of individual words relative to the entire corpus. Label encoding was applied and three categorical learning style labels - (Auditory, Visual, Kinesthetic - AVK) were then converted into numerical format suitable for model training. Subsequently, in order to ensure compatibility with the CategoricalCrossentropy loss functions used in deep learning models, these numerical labels underwent one-hot encoding by converting each label into a binary vector (e.g., [1,0,0] for Auditory). These steps were carried out to ensure that both input features and target variables were in optimal format for processing by transformer-based neural network.

## Model Architecture

The study fine-tunes a pretrained transformer, DistilBERT in the task of sequence classification by looking at a sentence and classifying in into one of the labelled learning style categories. The Huggingface Transformers library was used to get both the tokenizer and the DistilBERT model. The tokenizer breaks down sentences into individual words or parts so it can understand it better. In this study, DistilBERT was selected because it’s a streamlined, faster version of bigger and more complex BERT. The model was adapted to predict which of the three learning styles a sentence belongs to. Moreover, in order to speed up the learning process, the coding included strategies to leverage hardware acceleration, such as running on a TPU if available. Besides, TensorFlow distribution strategy was used to setup how the model would learn; Categorical Crossentropy loss function was used to measure how wrong its prediction is for the different learning styles and learn from its mistakes. In-addition, an Adam optimizer was initialised with a warm-up learning schedule which smartly adjusting how the learning pace of the model to start slow and speed up later. Leveraging the capabilities of DistilBERT, the model to able to achieve high classification accuracy while also being faster and requiring fewer resources.

## Training and Evaluation

The model was trained over five “rounds” (epochs), showing it 32 examples at a time (batch size of 32). It learned from a prepared dataset which had been tokenised and labelled with learning styles. In order to optimise model parameters, it begins with a gentle learning rate of 2e-5 along with a learning rate scheduler to gradually adjust the learning rate over time. During the training, the model’s performance metrics including categorical accuracy can be monitored concurrently at every epoch. After training, the model’s performance was evaluated using the test dataset; therefore, providing an unbiased estimate of generalization. The evaluation metrics encompassed loss, accuracy, precision, recall, and F1-score were reported upon completion of the training. The results confirmed the model’s ability to correctly classify unseen sentences into the correct learning style categories. Hence, this validates its potential to support adaptive learning technologies based on natural language data.

## Prediction

Upon training and evaluation of the model, the model is ready for deployed to interpret new textual data from student written responses and the model was able to predict the learning styles. Hence, several sentences were fed into the trained model and the model outputted raw scores (logits) for each learning style. A softmax function was used to turn them into probabilities and the model then picked the learning style with the highest probability as its guess. After the model made its numerical guesses (like 0, 1, or 2 for the different styles), the numerical predictions were inverse transformed back into their previously label encoder ("Auditory," "Visual," "Kinesthetic"). This workflow demonstrated the practical usability of the trained model, whereby educators could take what a learner writes and effortlessly get an idea of how that learner learns best.

# Results

The key metric to measure how well the model performed was accuracy as it shows the proportion of correct predictions among all predictions made. The LSTM model that used GloVe word vectors reached an accuracy of **97.84%. This result indicates that the LSTM model** did a fairly good job at identifying learning styles in sentences. In contrast, the DistilBERT-based model outperformed the LSTM model and reached a higher accuracy of **98.06%**. This result shows even lightweight versions like DistilBERT captured deeper contextual relationships in text. In other words, DistilBERT model understand the small differences in how people described different learning styles better than LSTM’s. Besides accuracy metric, when looking at lass, the LSTM with GloVe word vectors had a higher loss value of **0.0609**. This means LSTM model had more trouble reducing in its prediction. On the other hand, the DistilBERT-based model had a much lower loss of **0.0593**. This indicates that DistilBERT model learned faster nad made more accurate predictions. The lower loss also shows how well DistilBERT model learn fewer parameters to adjust. Although it is a smaller model, it still understands the rich semantic of text very well. This is crucial when the task involved choosing from many categories like learning style prediction. In addition, a paired T-test was done to compare the LSTM and DistilBERT on (i) mean accuracy and (ii) mean loss. The T-test result for mean accuracy was t (3) = 3.4, p < 0.042 indicates that it was statistically significant in accuracy improvement from **0.965 ± 0.018 to 0.977 ± 0.004 (*p* < 0.05)**. An important indicator must be highlighted was the lower standard deviation (SD = 0.004 compared to LSTM's SD = 0.02) indicates a significant improvement in the consistency and stability of its performance, making DistilBERT a more reliable classifier. On the other hand, the paired T-test result for loss comparison was t (3) = 2.1, p < .127 indicates that it was statistically non-significant in loss reduction from **0.096 ± 0.048 to 0.071 ± 0.016 (*p* >0.05).** Although the mean loss for DistilBERT is noticeably smaller than LSTM’s; however, this loss reduction cannot be attributed to the model’s architecture benefits may be due to small sample size.

Table 1 shows the performance results of DistilBERT model. It reached a high precision of **98%**, indicating it was good correctly identifying all three learning styles. This also means the model made trustworthy prediction as it made fewer false alarms. DistilBERT model also had a recall score of **98%**. This shows it was strong at finding most relevant cases across learning styles of Auditory and Visual. The recall for Kinesthetic was slightly lower at 97% but is still considered good. DistilBERT’s self-attention mechanisms focus on important parts of the text and keep strong performance even it is a smaller version of BERT using fewer resources.

**TABLE 1.** Classification report

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Learning Styles** | **Precision** | **Recall** | **F1-Score** | **Support** |
| Auditory | 0.98 | 0.98 | 0.98 | 0.98 |
| Kinesthetic | 0.98 | 0.97 | 0.98 | 0.98 |
| Visual | 0.98 | 0.98 | 0.98 | 0.98 |

The confusion matrix for the DistillBERT model showed clearer class separations between the classes. It had high true positives rates for all three learning styles. This means it made accurate prediction most of the time. Most of the correct guesses appeared along the diagonal of the matrix. This pattern shows that DistillBERT was able to difference between the learning styles with very few mistakes. Figure 1 shows the confusion matrix of the model. It highlights strong classification performance across three learning style categories (Auditory, Kinesthetic and Visual). The model correctly predicted Auditory **910** times, Kinesthetic **863** times and Visual **1109** times (diagonal elements). The numbers outside the are small, and this shows small number of misclassifications. Among the three, the model performed best on Visual, which had the highest correct predictions (**1109** times). In a nutshell, the confusion matrix shows that the model did a good job all three learning styles.

A blue squares with white text

AI-generated content may be incorrect.

**FIGURE 1.** Confusion matrix of the model

# DISCUSSION AND CONCLUSION

The result of the study revealed that the DistilBERT-based model is more robust and performed better in terms of generalization capabilities. In simpler terms, the DistilBERT-based model perform better in classifying learning style of learners based on their text inputs. Indeed, this are proven by the results that consistently beat LSTM in terms of the data loss and accuracy during the cross validation and evaluation on unseen data. The reason for its success can be attributed to DistilBERT’s capability to retain a large portion of BERT’s contextual understanding while offering lightweight and efficient architecture. The LSTM model with GloVe embeddings provided reasonable results, its capacity to adapt to different sentence structures and linguistic settings was constrained by its sequential processing and reliance on static word vectors. However, DistilBERT, on the other hand, could better generalize to diverse inputs by utilizing pre-trained knowledge and self-attention mechanisms. As both models of LSTM-GloVe and DistilBERT are compared, a distinct advantage for transformer-based models are notable. DistilBERT, being a lightweight version of its predecessor BERT, not only retains its core architecture especially offers strategic multi-head self-attention but also generation of deep contextual embeddings. The model is able to capture complex dependencies between words in a sentence regardless of their placements. “Multi-head self-attention” means the model examines each word in a sentence from multiple perspectives (or "heads") at once in order to understand how each words relates to every other word. Having said that, this empowers the model to identify important patterns, such as whether a word refers to a visual or auditory preference, even if it is far away in the sentence. This is particularly valuable in learning style classification as subtle linguistic cues (e.g., preference verbs like “listen”, “see”, “move”) can appear in varied contexts. The main difference of both models underlies in its processes; LSTM processes sequences token by token and relies on static word embeddings whereas DistilBERT processes entire sequences simultaneously that enabling it to model relationships across the whole sentence more effectively and efficiently.

Moreover, DistilBERT’s architecture prioritizes efficiency: (i) retains 97% of BERT’s performance (ii) use 40% fewer parameters (iii) running 60% faster [21], [22]. LSTM processes input sequentially, word by word, in a slower manner; in contrast, DistilBERT leverages a transformer-based parallel processing mechanism having all words in a sentence can be analysed at once, a faster computation on modern hardware (like GPUs or TPUs). Furthermore, DistilBERT uses a distillation process to compress knowledge from a larger BERT model into a smaller network without affecting its accuracy. As a result, DistilBERT consumes less memory for both training and inference with high level of contextual understanding [23]. LSTM model needs to maintain internal memory states and suffer from vanishing gradient issues over long sequences. As a result of these, the efficiency and performance on tasks involving long or complex text are thus limited. In short, for learning style classification, DistilBERT appears to be a more scalable and reliable approach due to its rapid and context-aware text interpretation. These findings strongly support the adoption of compact transformer models like DistilBERT in educational NLP tasks due to its exceptional performance in terms of speed and efficiency. The benefits of DistilBERT allow adaptive learning system to intelligently curate educational curriculum and delivery like choosing visual, audio or interactive content based on user preferences; therefore, it boosts learning efficiency.

Although DistilBERT model achieved impressive result in learning style classification, it has few limitations. Although DistilBERT is less computationally costly than full BERT; however, significant resources for fine-tuning and inference are needed for fine-tuning and inference. Therefore, this requirement may hinder deployment in low-resource environments. Moreover, the dataset used in this study was relatively small in both size and diversity, the generalizability of the results across broader educational contexts might be affected. Hence, future research should utilise larger, more heterogeneous datasets to include multiple languages, educational levels, and demographic groupings as well as evaluating model performance in real-world scenarios would further validate its effectiveness. Future research can also explore the application of other efficient transformer cousins like ALBERT, MobileBERT or Longformer which are the optimised models designed for different constraints and input lengths. In-addition, multimodal data like student interaction patterns, video or audio cues can be included which would enhance the accuracy of learning style classification. Another promising approach is to leverage on knowledge distillation by using large transformer models to train smaller models with negligible performance trade-off. With all these optimised solutions, personalised learning technologies more accessible, even in educational settings with limited resources.

# Acknowledgments

# We would like to thank anonymous reviewers for their insightful criticism for improvement of this manuscript.

# References

1. G. Brown, “The past, present and future of educational assessment: a transdisciplinary perspective,” *Frontiers in Education* **7**, 106063 (2022).
2. J. Arribas, V. Pastor, and A. Picos, “External constraints on the development of quality assessment of students’ learning in higher education,” *Education Sciences* **15**(1), 20 (2024).
3. T. Hussain, L. Yu, M. Asim, A. Ahmed, and M. A. Wani, “Enhancing E-Learning Adaptability with Automated Learning Style Identification and Sentiment Analysis: A Hybrid Deep Learning Approach for Smart Education,” *Information* **15**(5), 277 (2024).
4. A. Sayed, M. Khafagy, M. Ali, and M. Mohamed, “Exploring the VAK model to predict student learning styles based on learning activity,” *Intelligent Systems with Applications* **25**, 200483 (2025).
5. J. Freeman, *Student Generative AI Survey 2025* (Higher Education Policy Institute, London, UK, 2025).
6. Kira Learning, “AI-powered teaching solutions,” available at: <https://www.kira-learning.com/> (accessed May 4, 2025).
7. S. G. Essa, T. Celik, and N. E. Human-Hendricks, “Personalized Adaptive Learning Technologies Based on Machine Learning Techniques to Identify Learning Styles: A Systematic Literature Review,” *IEEE Access* **11**, 48392–48409 (2023).
8. M. S. Hasibuan, R. R. Isnanto, D. A. Dewi, T. B. Kurniawan, M.-L. Yeh, and A. Wijaya, “A proposed model for detecting learning styles based on the Felder-Silverman model using KNN and LR with electroencephalography (EEG),” *Journal of Applied Data Sciences* **6**(2), 1129–1139 (2025).
9. R. Lawpanom, W. Songpan, and J. Kaewyotha, “Advancing facial expression recognition in online learning education using a homogeneous ensemble convolutional neural network approach,” *Applied Sciences* **14**(3), 1156 (2024).
10. J. Pennington, R. Socher, and C. D. Manning, “GloVe: Global Vectors for Word Representation,” arXiv preprint (2014), available at: <https://nlp.stanford.edu/pubs/glove.pdf>.
11. A. Adell, *Learning Style (VAK)* (Kaggle Dataset, 2021), available at: <https://www.kaggle.com/code/ahmedadell30/classification-of-learning-style-using-dl/notebook> (accessed May 4, 2025).
12. A. Adell, *Classification of Learning Style Using DL* (Kaggle Notebook, 2023), available at: <https://www.kaggle.com/code/ahmedadell30/classification-of-learning-style-using-dl/notebook> (accessed May 4, 2025).
13. P. A. Kirschner and C. Hendrick, *How Learning Happens: Seminal Works in Educational Psychology and What They Mean in Practice*, 2nd ed. (Routledge, London, 2024).
14. H. Pashler, M. McDaniel, D. Rohrer, and R. Bjork, “Learning styles: Concepts and evidence,” *Psychological Science in the Public Interest* **9**(3), 105–119 (2008).
15. S. F. Noorani, M. Karimi, and Z. Gholijafari, “Using ensemble machine learning and feature engineering to increase the accuracy of predicting learners' performance in an online educational environment,” *Interdisciplinary Journal of Virtual Learning in Medical Sciences* **15**(4), 369–387 (2024).
16. A. Hananto, “Identifying student learning styles using support vector machine in Felder-Silverman model,” *Journal of Applied Data Sciences* **5**(3), 1495–1507 (2024).
17. N. Gardazi, A. Daud, M. Malik, A. Bukhari, T. Alsahfi, and B. Alshemaimri, “BERT applications in natural language processing: a review,” *Artificial Intelligence Review* **58**(6) (2025).
18. J. Wang, J. Huang, X. Tu, J. Wang, A. Huang, and M. Laskar et al., “Utilizing BERT for information retrieval: survey, applications, resources, and challenges,” *ACM Computing Surveys* **56**(7), 1–33 (2024).
19. R. Barbon and A. Akabane, “Towards transfer learning techniques—BERT, DistilBERT, BERTimbau, and DistilBERTimbau for automatic text classification from different languages: a case study,” *Sensors* **22**(21), 8184 (2022).
20. S. Ruder, A. Conneau, and N. Y. Du, “Are We Really Making Much Progress in Text Classification? A Comparative Review,” arXiv preprint arXiv:2204.03954 (2024).
21. V. Sanh, L. Debut, J. Chaumond, and T. Wolf, “DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter,” arXiv preprint arXiv:1910.01108 (2019).
22. C.-Y. Shin, J.-T. Park, U.-J. Baek, and M.-S. Kim, “A Feasible and Explainable Network Traffic Classifier Utilizing DistilBERT,” *IEEE Access* **11**, 70216–70237 (2023).
23. J. Jayapradha, Y. Kulkarni, L. Vadhanie G, P. Naveen, and E. Abdulwahab Anaam, “Treatment Recommendation using BERT Personalization,” *Journal of Informatics and Web Engineering* **3**(3), 41–62 (2024).