Analysis of Human and AI-generated Text Classification

Md. Rony[1], Jubair Khan Shahos [1], Adnan Mahmud Fuad[2], Sadman Sadik Khan[1*], Asiful Islam [1], Md Minhajul Islam[1]

[1]Department of Computer Science and Engineering, Daffodil International University, Dhaka 1216, Bangladesh
[2]Department of Information and Communication Technology Cell, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh

{rony15-4967, islam15-5919, islam2305101257}@diu.edu.bd, {jubairkhanshahos}@gmail.com, {11adnan}@live.com, sadman15-13696@diu.edu.bd*

Abstract. Artificial Intelligence plays a vital role in generating Text that is used in our daily lives. However, the rapidly increasing usefulness of AI technology is a concern, according to need. This includes biases in the response of the generative AI. There are lots of generative AI that contribute to generating text, documents, articles, etc. However, the generative model is also concerned about the authentication of text. This paper surveys to collect human and AI responses to a question. The same question is set for AI tools to generate responses. This study evaluates and examines the linguistically and semantically important features of human and AI-generated text by using machine learning and deep learning classification. To achieve an optimal performance with the most suitable machine learning model, we also need a neural network model, which is based on deep learning. Then the response of this model classifies the text as human or AI-generated. The human-generated text will be much better than the AI-generated text. In the present digital world, AI captures 70% of information from humans, so it provides text like humans. In this paper, the Contribution of this novel work development of a methodology that can identify AI-generated text from human-authored text using different machine learning models. It has become significantly difficult to tell AI-generated text apart from human writing, and the existing detectors are restricted to minor datasets and superficial features. The study introduces a combined ML–DL approach that merges linguistic characteristics with VGG-16 semantic embeddings for the categorization of AI and human text. With the training on 487k samples, the model can reach 0.999 accuracy, thus proving to be very robust and able to generalize well. The findings validate the hybrid feature approach as being effective for the reliable detection of AI-generated texts.

Keywords: AI, Human, Artificial Intelligence, Text, Generative AI. Prompting, ChatGPT, AI in Education, Natural Language Processing. Neural Network, Generated, AI-generated content, NLP.

M. S. Arefin et al. (eds.), *Proceedings of the International Conference on Intelligent Data Analysis and Applications (IDAA 2025)*, Advances in Intelligent Systems Research 206,
https://doi.org/10.2991/978-94-6239-664-7_72

1 Introduction

In Computer Science and Engineering, AI (Artificial Intelligence) plays a vital role. AI is a major branch of computer science and engineering that deals with the functional and non-functional behavior of the machine. That can perform tasks similar to human intelligence. AI thinks like human intelligence and acts like humans. The central principles of AI revolve around reasoning, knowledge, planning, learning, communication, ability to move and manipulate objects. Machine learning is another concept of Computer Science and Engineering. It is capable of solving any problem by using different algorithms. AI is implemented in various ways as -

Machine Learning: It is one of the most common applications of AI where the machine learns and improves from its experience and prediction automatically, rather than being explicitly programmed to perform certain tasks. Another term is Deep learning. Deep learning is the advanced version of Machine learning, which includes the ANN (Artificial Neural Network). There are various types of machine learning algorithms implemented here, such as Supervised, Semi-Supervised, Unsupervised, and Reinforcement Machine learning.

In Unsupervised Learning, the machine is trained on an unlabeled dataset. On the other hand, the label dataset is used in Supervised machine learning. Natural Language Processing: It is the field of AI whose target is to develop algorithms that enable computers to understand, interpret, generate, and manipulate language. Neural Network: It is the part of the artificial system that is inspired by the human brain's neural network. It consists of a large connected network of computational neurons in layers. At present, the Application of AI is found in every sector, from healthcare and medicine to make better and faster diagnoses to engineering, like neural networks. In agriculture, for crop monitoring and predictive analysis. Since AI helps automate tasks, low-skilled and routine jobs are the most susceptible to being displaced by AI. Within the context of generative AI, focus on productive content across text, image, and audio. It applies to different concepts of classification, like Text, Image, Document, and Article. On the other hand, Humans also generate this type of Document, text, and image.

The study aims to gain insights into the difference between human language use and AI-generated text, and how this difference can improve the accuracy of classifying human and AI-generated text. Based on prior knowledge, first, key features first classified the important, and the machine learning model that is best for the selected feature. AI is now so powerful that it has become difficult for humans to determine whether the content is generated by AI or humans. An illustrative experiment in the form of a game of trivia was conducted to analyze how efficiently humans can detect whether the content generated is by the AI or a human. It has been shown that there is a 50% chance of correct detection of the content, which is an alarming situation, as AI can be misused to generate false and provocative content to spread hate in the community. The rise of deepfake images and videos, especially in the areas of adult movies and politics, has posed serious threats. The ethical considerations of DALL-E involve concerns about bias and discrimination, privacy, and unintended outcomes. For instance, the images it creates could uphold unrealistic beauty standards, strengthen gender stereotypes, or add to the objectification of women. To address the issues posed by the generative AI tools, this paper aims to analyze the responses generated by AI when compared with responses of humans using various machine learning algorithms on several parameters, like vocabulary richness, spelling errors, grammatical score, readability score, etc.

2 Literature Review

Researchers have previously implemented various methodologies and approaches to Human and AI-generated text. Recently, the authors Ali et al. Bataineh [2] proposed a benchmark dataset to distinguish AI and human-generated generated of 10,000 records, comparing both human and AI-generated text, and addressing the challenge between the two contexts of AI-generated text in academic domains. It highlights the ethical implementation and the need for detecting AI-generated text in various concepts, including the academic perspective. The study introduces various machine learning algorithms and models. If here include an advanced AI model, it improves the accuracy of detection. According to Mindner et. L [6], Ensure exploration of popular and new important features for finding text generated by AI from scratch and text rephrased by AI. Since they found that classification is more complex when the AI has been instructed to create the text in a way that a human would not recognize that it was generated by an AI, they also investigated this more advanced case of the content of AI-generated text. Their results show that the new features substantially help to improve the performance of many classifiers.
Researcher Ahmed M. Elkhatat [5] evaluated the efficacy of AI content detection tools in differentiating between human and AI-generated text. This study investigates the capabilities of various AI content detection tools in discerning human and AI-authored content. Fifteen paragraphs each from ChatGPT Models 3.5 and 4 on the topic of cooling towers in the engineering process, and five human-written control responses were generated for evaluation. AI content detection tools developed by OpenAI, Writer, Copyleaks, GPTZero, and CrossPlag were used to evaluate these paragraphs. Bellini, V., Semeraro, F., Montomoli, [9] introduces differentiating between human-written and AI-generated texts using linguistic features automatically extracted from an online computational tool. The limited research comparing linguistic features of human-written and AI-generated texts. Previous studies, such as those by Alexander [7] and Herbold [10], identified distinct language patterns between the two, with AI texts exhibiting structural uniformity and predictable patterns Bellini [11] further explored how language models like ChatGPT replicate human language but noted discrepancies in word choice and syntactic ambiguity resolution. The review emphasizes the need for comprehensive analyses to understand these differences better, which can enhance AI training methodologies and improve the quality of AI-generated content in various applications. Shashank Kumar [1] This includes biases in the response of generative AI, interventions affecting the privacy of individuals, etc. Also, with easy access to emerging technologies like ChatGPT, Bard, DALL-E, etc., there has been a significant rise in the generation of fake content like fake images and deep-fake videos. Bashardoust, A., Feuerriegel, S. [15] This paper involves comparing the willingness to share human-generated vs AI-generated fake news detection by using various machine learning models that can identify by machine learning and detect fake news. Mindner, L., Schlippe, T., & Schaaff, K. [4] This paper investigates the classification of human and AI-

generated text using various features. It analyzes the two scenarios AI AI-generated text and human-generated text, for classifications. This paper also evaluates the features of AI or human-generated content across multiple languages. In contrast, our experiment introduces a hybrid approach that combines machine learning and deep learning techniques. Including Random Forest, Gradient Boosting, and VGG-16 CNNs. And we use the cosine similarity to perform clustering, vocabulary finding, and grammar finding for the text classification process. For further validation of our models, we use the largest dataset, which includes 487,000+ samples which providing robustness and accurate results.

2.1 Justification of the Proposed Approach

Our proposed methodology is needed to build an accurate system to detect AI-generated text. Which is closer to human-generated text but different in semantic patterns. Our approach depends on some factors:

1. Explainability: It includes the feature-level explainability by vocabulary, grammatical errors, and spelling issues that are human-interpretable. It is the common difference between human and AI-generated text.
2. Deep feature Representation: It is represented by VGG-16, which extracts the deep feature representation from the text.
3. Cosine similarity: It represents the metric for comparing semantic vectors, like different answers to the same question.
4. Model Performance: Random forest and Gradient boosting can handle complex and high-dimensional features with low overfitting, and SVM, VGG-16 work for Deep learning.

Overall, the technique provides a balance between performance, scalability, and accuracy.

3 Proposed Methodology

The proposed approach is based on a clearly defined workflow comprising data preprocessing, dual feature extraction, model training, and evaluation. Initially, the text is cleaned and normalized, and subsequently, two parallel pipelines are deployed: one for extracting handcrafted linguistic features and the other for deep semantic embeddings. The text is transformed into embedding matrices, so the VGG-16 CNN model is able to identify semantic patterns; meanwhile, ML models like Random Forest and Gradient Boosting are trained on structured linguistic features. The combination of these features increases the system's resistance and classification correctness. The overall system flow is illustrated in Figure 1. Looking at past work, various activities have been conducted for the participants to differentiate between human and AI-generated content.

3.1 Data Collection

For data collection, various methods were used. First of all, the Data was initially collected from multiple sources, then we also pushed some random text to it, which is human-generated, and I also had to create text from AI differently. total sample 487,000 where human written text involved 243,500 and ai generated text 243,500 and text are repositories academic forum and GTP based generations which language is English. However, the data collected is from a specific demographic location in the English language (which is not the primary language for most of those who responded). Following the collection of 65% human responses, about 35% of responses are generated using various AI tools like ChatGPT, Bard, and Aichatting for the dataset.

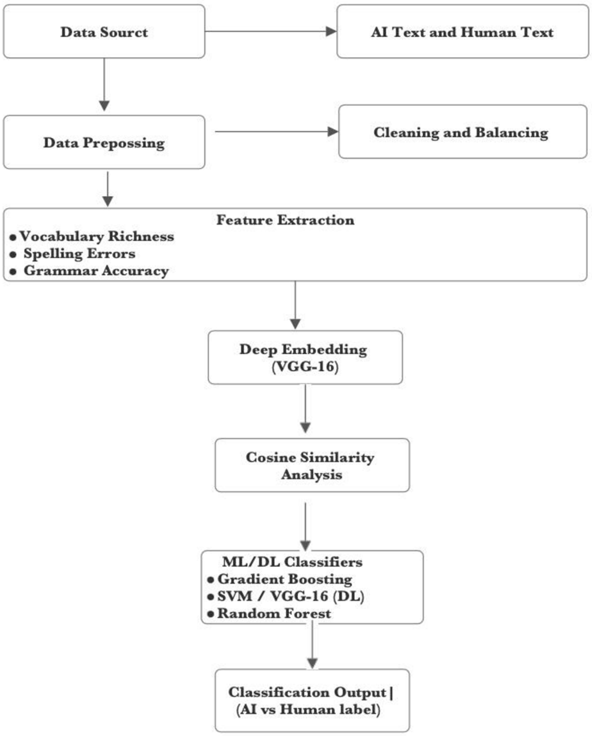

Fig. 1. Proposed System Architecture for Human and AI Text Classification

As shown in Fig. 1 Proposed System Architecture for Human and AI Text Classification where start from data source then it moves to the Classification output of AI vs Human label.

3.2 Content Selection and Curation

An equal number of human and AI-generated responses were selected to create a balanced dataset. To minimize the bias between human and AI-generated content, data is curated keeping the following factors in mind, which is written as a projection of (A).

As shown in Fig. 2 illustration of Cosine Similarity between vector embeddings where projB(A) and projA(B) make a angle.

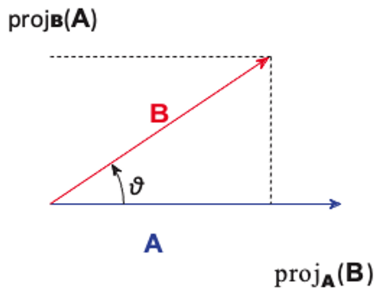

Fig. 2. Illustration of Cosine Similarity between vector embeddings.

3.3 Similarity score analysis

Once the responses were collected, Similarity Score Analysis was performed to analyze the distribution between AI-AI, Human-Human, and UL-box responses. It was performed by generating the embeddings for each response and evaluating the similarity scores.

1) VGG-16_Model: VGG-16 is a convolutional neural network model. It has performed 10/10 epochs. Epoch 1/10 provides an Accuracy of 0.84 and a loss of 0.36, and then on the continuous update, its result is where the Accuracy increases and provides 0.99 and the loss decreases to 0.006. They consist of 16 weight layers, 13 convolutional, and 3 fully connected layers. It is one of the top-performing learning models based on CNN.
Architecture Design of VGG-16: Input: The Input is provided in a different way from the builder. Convolutional Layer: It uses a 3*3 filter and 1 padding, which will be 0. Max Pooling: Max pooling can also be implemented here, which will be a 2*2 filter and a stride of 2. And a fully connected layer. This process will continue when it finds any image data that is created by AI as Text.

2) Random Forest Model: The score of the Random Forest model is 0.98 Random Forest model is another best model for analyzing the accuracy of the training data. In my research, Random Forest performs classification and regression, and it works by building a decision tree on my model. It also combines their output to get the best accuracy and prevents overfitting.
Working procedure: The first one is data bootstrapping, which creates multiple decision trees using a subset of the training data. In this process name is bagging. Feature Randomness: In this section, split into a tree. Which increases the feature of the model.

3) Gradient Boosting: Gradient boosting is another boosting algorithm. Which is to update the result and provide a satisfactory result. In this Survey, Gradient boosting provides a 0.99 accuracy result. We also implemented a boosting algorithm to get better results. It is not necessary to apply, but we are applying it to get better performance with accuracy. Finally, we get better accuracy by using this boosting algorithm. Which is a notable point.

4) Similarity Score - Cosine Similarity: This technique evaluates how alike two embeddings are by measuring their similarity. The score is determined by the cosine angle formed between the two vectors being compared. A greater cosine similarity value indicates that the vectors are closer, implying a higher degree of similarity between the two text corpora represented by the embeddings. The cosine similarity between two vectors **u** and **v** is given using the formula:

$$\text{Cosin Simimilarity (U.V)} = \frac{|\mathrm{U.V}|}{||\mathrm{U}||.||V||} \tag{1}$$

Where · denotes the dot product and **u** represents the Euclidean norm (magnitude) of vector **u**.

In this paper, the VGG-16 model has been used to generate the embeddings of all the responses. The cosine similarity score is calculated for all the possible human-human, human-AI, and AI-AI response pairs for respective questions and topics. Following this, corresponding mean scores are evaluated. The mean Human-Human similarity score indicates how similar the two different human responses to a particular question are. Similarly, the mean Human-AI similarity score indicates the similarity between human and AI responses, and the mean AI-AI similarity score indicates the similarity of two AI-generated responses to a particular question.

3.4 Clustering

An unlabeled dataset containing humans and AI is balanced by 50% AI-generated and 50% is Human human-generated text. Assume here that AI-generated clustering on class 1 and Human-generated clustering on class 2 responses for a particular dataset. Clustering is the process of grouping where the same group of data exists in the same cluster. In this survey paper, we have two clusters. Clustering finds the related text that is similar to both human and AI-generated. If 1 is represented on the cluster, it takes it as an AI-generated text, and if it finds 0 takes it as a human-generated dataset. If any value exists from 0 to 1, like 0.8, it takes it as AI-generated, it focuses and takes the 50%, and if 0.35, it takes it as 0, which is Human-generated. The cluster will be in 2 classes: the 0 and 1 cluster takes all the relevant AI-generated data, and the 0 cluster takes on all the Human-generated text.

3.5 Classification

The unlabelled dataset is analyzed using various classification algorithms to classify the responses into two classes, namely, human and AI, using various features like vocabulary richness, spelling errors, readability score, etc. Classification is a supervised machine-learning method in which a classifier is trained on a labeled dataset and is subsequently used to predict the class of a new data point.

Random Forest: Building on the large methodology, this method creates a diversified forest of decision trees by incorporating both large and feature randomness. Using different random selections of the data and features, the technique creates several decision trees. Each decision tree's predictions are then calculated, and the most frequent outcome is taken into account. The following are the hyperparameters for the Random Forest classifier

- Max_depth: This limits the depth needed for the random forest to grow by representing the largest path between a root node and a leaf node
- Max_features: The largest number of features that each tree in the random forest can have is indicated by the value max_features.
- Min_samples_split: This option establishes the bare minimum of observations needed in a decision tree node to start a split. Two is the default value.

3.6 Feature Extraction

From the responses collected through the survey, key features like best linguistic, format, and semantic attributes are extracted. Being used to this methodology forms the foundation of the research to differentiate between human and AI-generated text. Here is the overview of each feature of the method. Let's calculate them given below.

- **Vocabulary richness:** This feature determines the difference in the words within the text.Vocabulary richness = Number of unique words / Total number of words.
- **Spelling Errors:** Spelling errors are indicative of the accuracy of the text. To determine the percentage of spelling errors in the text. Spelling error (%) = Spelling error word/ total words *100
- **Grammatical error:** Grammatical Mistakes. Another crucial component of text quality is grammatical accuracy. Using the language-tool-python library and the Language Tool Package, we determine the proportion of grammatical errors.

3.7 Sample Dataset:

```
 Preview of the dataset:
                                                 text  generated
0  Cars. Cars have been around since they became ...        0.0
1  Transportation is a large necessity in most co...        0.0
2  "America's love affair with it's vehicles seem...        0.0
3  How often do you ride in a car? Do you drive a...        0.0
4  Cars are a wonderful thing. They are perhaps o...        0.0
```

Fig. 3. Sample dataset

In this sample dataset as shown in Fig. 3, there are some datasets randomly, which are previewed as sample datasets as ranging from 0 to 4, and Text is represented by 0.0, which means Human-generated text, and 1.1 is represented by AI-generated text. As shown in Fig. 4 Proposed System Architecture goes through the block diagram where we can see the working flow.

3.8 Pipeline

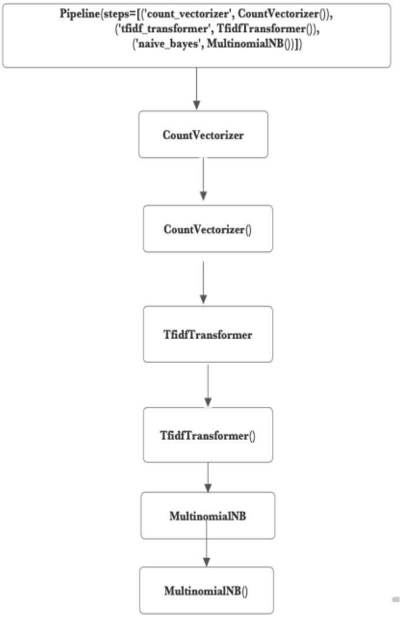

Fig. 4. Proposed System Architecture for Human and AI Text Classification

3.9 Data visualization

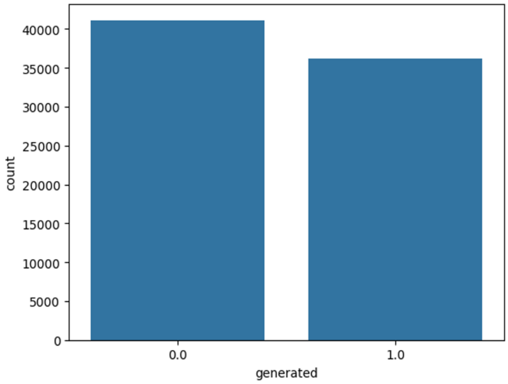

Fig. 5. Distribution of Human and AI-Generated Text

As shown in Fig. 5 data visualization according to the Generated and count of Distribution of Human and AI-Generated Text. Table 1. presents the preview of the dataset, Text, and Generated based on the generated text like Cars. Cars have been around since they became. Table 2 presents the similar scores for different responses based on model and similarity score.

Table 1. Preview of the dataset, Text, and Generated

Generated	Text
0.0	Cars. Cars have been around since they became..
0.0	Transportation is a large necessity in most co...
0.0	America's love affair with it's vehicles seem...
0.0	How often do you ride in a car? Do you drivea...
0.0	Cars are a wonderful thing. They are perhapso...

Table 2. Similarity Scores for Different Responses

Model	Similarity Score
Gradient Boosting	0.98
Random Forest	0.97
VGG-16	0.99

3.10 Novelty of the Proposed Approach

The novelty of our proposed method integrates with multiclass classification, which includes machine and deep learning methods for detecting AI-generated text. We propose a hybrid approach for feature-based classifications.
Novel contribution of our research work:

- Integration of Dual Feature: As surface-level feature vocabulary, spelling, and grammar. And a deep learning VGG-16 convolutional neural network, here we use for text comparison.
- Surface-level interpretable features for vocabulary richness and grammatical accuracy, and spelling mistakes.
- A large-scale and balanced dataset was used for the experiment.
- Multi-model performance evaluation involving a random forest, CNN CNN-based classifier.
- Cosine similarity: We use cosine similarity for semantic evaluation across Human-Human, AI-AI, and Human-AI, including the semanticity.
- Dataset size balancing (Robustness): we build datasets with 487,000+ samples, which is significantly larger. So we balance it.
- Evaluation of the model: To evaluate the performance of the model, we compare it with the Random Forest, Gradient Boosting, and VGG-16 with Cosine similarity for semantic comparison between AI and human-generated text.

4 Result Analysis

4.1 Similarity Score Analysis

Analysis of Similarity Scores Table II shows that responses produced by AI are likely to be those with a very high similarity score between two or more responses. It indicates that AI answers are meaning-consistent and typically employ the same vocabulary while responding to a given query. The accuracy, precision, recall, F1 score, confusion metrics, and AUROC value. A performance comparison table was added to the result.

Table 3. Similarity Scores with precision, recall, and f1 score

Classifier	Accuracy	Precision	Recall	F1 Score
Random Forest	0.97	0.97	0.98	0.98
Gradient Boost	0.98	0.99	0.97	1.00
VGG-16	0.999	-	-	-
SVM (SVC)	0.87	0.84	0.93	0.88

We can see that Table 3, Random forest model provides an accuracy of 97%, Gradient descent provides 98% and SVM does not work as well on accuracy as others, but VGG-16 proves the maximum accuracy on 10/10 epoch, which is 99%, and the low loss is 0.001. Precision is also notable here we can see precision is 97% on Random Forest and Gradient descent 97% and SVM only 93%. The maxim Precision is 99% which is provided by the Gradient boosting it boosts the Precision the Recall value indicate where Random Forest Provide 98% which is better than the boosting algorithm and F1 score of the boosting algorithm is maximum 1.00 as the noticeable point on this case study Boosting algorithm perform well and VGG-16 also is good result provider.

4.2 Classification Result:

1. Random Forest: The model obtained an accuracy of 0.97, a precision of 0.97, a recall of 0.98, and an F1 score of 0.98 with the hyperparameters adjusted. Significantly, the accuracy increased in comparison to Logistic Regression, improving the outcomes overall.

2. Gradient Boosting: The gradient boosting algorithm boosts the accuracy of the model, with an obtained accuracy of 0.98, a precision of 0.99, a recall of 0.97, and an F1 score of 1.00 with the hyperparameters adjusted. Significantly, the accuracy increased in comparison to Logistic Regression, improving the outcomes overall.

3. VGG-16: It is another model that is part of the neural network.

It obtained an accuracy of 0.996, a test loss of 0.006, and it used or moved here epoch 10/10, but the best result was provided epoch 7/10 with an accuracy of 0.999 and a loss of 0.001which one is the best?

4. SVM (SVC): The model obtained an accuracy of 0.87, a precision of 0.84, a recall of 0.93, and an F1 score of 0.88 with the hyperparameters adjusted.

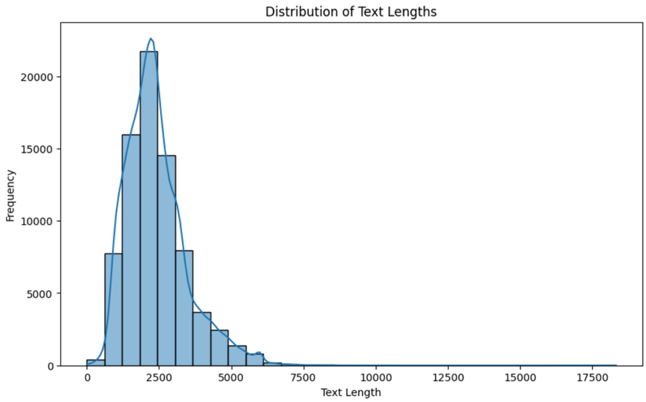

Fig. 6. Distribution of Text Length and Frequency

As shown in Fig. 6, which provides a view of the Distribution of text length on the Y axis and Frequency on the X axis. Their frequency is found to be more than 20000, and the lowest is almost 0. The text length is 75000 on the endpoint, and it goes up and down on different text lengths according to the frequency.

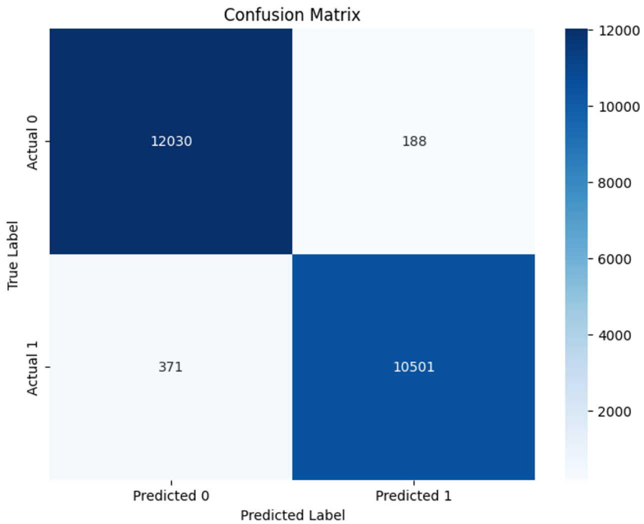

Fig. 7. Confusion Matrix VGG-16 classifier

As shown in Fig. 7, on this confusion matrix plot, the True level of the VGG-16 model Actual 0 is found here at 12030, and the 188 others, and the Actual 1 is found here at 371 on prediction is 0, but when the Actual 1 is 10501 then the prediction is 188 we can see on the predicted 1.

Comparison: In our evaluation of the classifiers, Random forest classifiers provide better results than the SVM model, which provides only 0.87 accuracy; on the other hand, Random forest provides 98. According to the model, VGG16 provides the best accuracy with a low loss of only 0.001.

5 Limitations and Feature Work

The limitation of this work is model different models perform differently results on the same work. Data is also another limitation when the data is run on the model, it takes different results in different ways like when it takes 100000 text data it or more than 100000 it provides a better result. On the other hand, when it takes less than 30000 text as input randomly, it does not work properly. The model is also another limitation; all of the models are from a well. Like a machine learning model. Future improvement may be achieved by applying a Deep Neural Network. Without using any boosting algorithm. If applying more better model, it proves a better result.

6 Conclusion

The research introduces a comparative study of different pre-trained models with the help of a Secondary dataset. Human and AI generative text can be classified by using different classifiers and machine learning models. A neural network model has also been implemented here, which provides good accuracy. In this survey, the paper focuses on Human and AI generative Datasets, which can be classified closely. If the model is found as 1, it is found as AI-generated data, and if it is found as 0, it means human-generated data. The proposed success of this paper is a hybrid model applying for the comparative classification accuracy, advising potential for reliable detection of AI-generated text or content. The study demonstrates that accuracy and outstanding performance are comparable approaches.

References

1. A. Lermann Henestrosa and J. Kimmerle, “The effects of assumed AI vs. Human authorship on the perception of a GPT-generated text,” Journalism and Media, vol. 5, no. 3, pp. 1085–1097, 2024.
2. A. Fiedler and J. Döpke, “Do humans identify AI-generated text better than machines? Evidence based on excerpts from German theses,” Int. Rev. Econ. Educ., vol. 49, no. 100321, p. 100321, 2025.
3. A. A. Bataineh, R. Sickler, K. Kurcz, and K. Pedersen, “AI-generated vs. Human text: Introducing a new dataset for benchmarking and analysis,” IEEE Trans. Artif. Intell., pp. 1–11, 2025
4. F. Retkowski and A. Waibel, “From text segmentation to smart chaptering: A novel benchmark for structuring video transcriptions,” arXiv [cs.CL], 2024.
5. L. Mindner, T. Schlippe, and K. Schaaff, “Classification of human- and AI-generated texts: Investigating features for ChatGPT,” arXiv [cs.CL], 2023.
6. L. Mindner, T. Schlippe, and K. Schaaff, “Classification of human- and AI-generated texts: Investigating features for ChatGPT,” in Artificial Intelligence in Education Technologies: New Development and Innovative Practices, Singapore: Springer Nature Singapore, 2023, pp. 152–170.

7. A. M. Elkhatat, K. Elsaid, and S. Almeer, “Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text,” Int. J. Educ. Integr., vol. 19, no. 1, 2023.
8. S. A. Hussain, R. Schmälzle, S. Lim, and N. Bouali, “Comparing AI and human-generated health messages in an Arabic cultural context,” Glob. Health Action, vol. 18, no. 1, p. 2464360, 2025.
9. A. Lermann Henestrosa and J. Kimmerle, “The effects of assumed AI vs. Human authorship on the perception of a GPT-generated text,” PsyArXiv, 2024.
10. H. Fang et al., “Could AI trace and explain the origins of AI-generated images and text?,” arXiv [cs.CL], 2025.
11. V. Bellini, F. Semeraro, J. Montomoli, M. Cascella, and E. Bignami, “Between human and AI: assessing the reliability of AI text detection tools,” Curr. Med. Res. Opin., vol. 40, no. 3, pp. 353–358, 2024.
12. A. Bhattacharjee and H. Liu, “Fighting fire with fire: Can ChatGPT detect AI-generated text?,” SIGKDD Explor., vol. 25, no. 2, pp. 14–21, 2024.
13. Y. Zhang, Y. Ma, J. Liu, X. Liu, X. Wang, and W. Lu, “Detection vs. Anti-detection: Is text generated by AI detectable?,” in Lecture Notes in Computer Science, Cham: Springer Nature Switzerland, 2024, pp. 209–222.
14. H. T. Hakam et al., “Human-written vs AI-generated texts in orthopedic academic literature: Comparative qualitative analysis,” JMIR Form. Res., vol. 8, p. e52164, 2024.
15. L. M. König, M. Podszun, W. Gaissmaier, and H. Giese, “Perceptions of AI-generated informational texts and effects of the source: An online experiment,” PsyArXiv, 2025.
16. G. D. Fisk, “AI or human? Finding and responding to artificial intelligence in student work,” Teach. Psychol., 2024.
17. A. Bashardoust, S. Feuerriegel, and Y. R. Shrestha, “Comparing the willingness to share for human-generated vs. AI-generated fake news,” arXiv [cs.SI], 2024.
18. J. Milička, A. Marklová, O. Drobil, and E. Pospíšilová, “Humans can learn to detect AI-generated texts, or at least learn when they can’t,” arXiv [cs.CL], 2025.
19. A. Hashemi, W. Shi, and J.-P. Corriveau, “AI-generated or AI touch-up? Identifying AI contribution in text data,” Int. J. Data Sci. Anal., 2024.
20. Z. Zeng et al., “Detecting AI-generated sentences in human-AI collaborative hybrid texts: Challenges, strategies, and insights,” in Proceedings of the Thirty-ThirdInternational Joint Conference on Artificial Intelligence, 2024.