Enhancing Contextual Masking in Reversible Linguistic Steganography with Ensemble Methods

M Prasha Meena, N J S Deepalakshmi, R Dharsni, R Subashree

Abstract**:** ***Various cybercrimes can be prevented by text authentication that is responsible for preserving digital identities and contents. Digital signatures come in handy as a way of authenticating texts, which is an extensively used method. One approach to this problem is linguistic steganography, which allows hiding the signature in other words within the text and thereby facilitating efficient data management. However, it should be noted that there is a danger that these kinds of changes may result in inappropriate decisions being taken by automated computing systems not to mention change their final outputs (unseen). As such, many people are becoming more concerned with the possibility of reversing steganography so that it becomes possible to eliminate any distortions made during the process. This paper uses Contextual masking instead of masking randomly with BERT model. The goal behind this research was developing a natural language text specific Reversible Steganographic System. Our model uses pre-trained BERT as a transformer based masked language model and reversibly embeds messages through predictive word substitution. To quantify predictive uncertainty, we introduce an adaptive steganographic technique using Bayesian deep learning. This experiment shows us how our proposed system balances imperceptibility with capacity while maintaining near semantics at all times. Also, we integrate ensemble methods instead of Monte Carlo to balance the imperceptibility.***

Keywords**:** ***Contextual, Ensemble Methods, Reversibility, Steganography.***

I. INTRODUCTION

Every organization should be concerned about cyber security, especially when it comes to authentication. This process involves confirming the identities of users and the integrity of the digital content. In an era where cyberspace is growing at a very rapid pace, authentication has become increasingly crucial in maintaining trust and guarding against several kinds of fraud such as identity theft, spamming, fake news dissemination, malicious hyperlinking, and tampering with digital media. One widely used to mean of ensuring authenticity is through digital signatures which employ modern cryptographic techniques such as encryption and hashing. Security can be further reinforced by incorporating timestamps or employing other tamper-evident methods. However, this additional information may get lost or mishandled accidentally during storage, transmission or format translation. The secret data turns into ciphertext through a cryptography method called encryption. Nevertheless, the unreadability of ciphertext can provoke uninvited attention from anyone who tries to decrypt it improperly. Reversibility refers to the ability to remove steganographic distortion without any loss of information. A reversible steganographic method extracts hidden message perfectly while retaining entire original data elsewhere put across by this definition text below. Reversible computing is mentioned in it, which means that computational processes can at least be reversed in time. This concept has relevance to steganography because one may infer from it that information hiding can be done away without any loss. Even though text data based reversible steganographic methods are not well developed unlike digital imagery ones, this happens due to difficulties caused by the natural language where small variations even can be noticed. Even though text data based reversible steganographic methods are not well developed unlike digital imagery ones, this happens due to difficulties caused by the natural language where small variations even can be noticed. Digital photographs and images have been subjected to such algorithms (methods) for years but there’s virtually no progress related with reversible steganography of textual data until recently. This necessitates efficient development of approaches towards hiding information within texts. Due to the growing importance of textual data and advancements in NLP technologies, the field of reversible steganography for textual data has emerged as a promising area of research. Researchers are exploring ways to hide information in text while ensuring reversibility and minimal distortion. There is a main focus in the article: reversible linguistic steganography, which refers to hiding messages within natural language text so that the original text can be recovered without losing meaning. To make an array of predictive words, the article uses a masked language model. It is possible that this model utilizes techniques like BERT (Bidirectional Encoder Representations from Transformers), which predicts masked words based on the surrounding context.

Manuscript received on 25 April 2024 | Revised Manuscript received on 05 May 2024 | Manuscript Accepted on 15 May 2024 | Manuscript published on 30 May 2024.

*****Correspondence Author(s)

Mrs. M. Prasha Meena*, Department of Information Technology, Mepco Schlenk Engineering College, Sivakasi (Tamil Nadu), India. Email: prashameena@mepcoeng.ac.in, ORCID ID: 0009-0007-4503-0704

N J S Deepalakshmi, Department of Information Technology, Mepco Schlenk Engineering College, Sivakasi (Tamil Nadu), India. Email: deepasarav6122003@gmail.com

R Dharsni, Department of Information Technology, Mepco Schlenk Engineering College, Sivakasi (Tamil Nadu), India. Email: dharsni05@gmail.com

R Subashree, Department of Information Technology, Mepco Schlenk Engineering College, Sivakasi (Tamil Nadu), India. Email: subashree444@gmail.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license http://creativecommons.org/licenses/by-nc-nd/4.0/

The list of predictive words produced is later utilized to embed messages in texts using predictive word substitution. In other words, some particular words in the text are replaced by predicted ones that would retain their contexts and logical flow. Based on redundancy in language; most words in this type of steganography can be accurately guessed within a limited set of predictive words hence making them reversible as well. A more accurate model will result in better coverage and detection while concealment of hidden massage will increase. This paper suggests another way to do this by using uncertainty about future predictions rather than making use of carrier selection process for setting up a steganographic system which hides information inside plain texts by replacing each “carrier” word with one corresponding to its prediction uncertainty; where uncertainty here means how unsure the model was concerning what it thought would come next. The paper looks at quantifying uncertainty using Bayesian deep learning techniques, presumably to assess the reliability of predictions and make informed decisions about message integration.

The rest of this article follows this structure:

Part II: Related Works. Part III: Overview of masked language modeling versus context masking Part IV: Practical methods for hidden language techniques with reversible. Part V: Discusses steganographic routing and efficiency. Part VI: A Bayesian Framework for Uncertainty Quantification. Part VII: Ensemble methods an uncertainty analyzer. Part VIII: Experimental results demonstrate the proposed systems. Part IX: Concluding remarks summarize the conclusions and implications.

II. RELATED WORKS

Unlike recent language representation models (Peters et al., 2018a; Radford et al., 2018), BERT is designed to pretrain deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context in all layers. Several studies have explored the application of Bayesian methods for uncertainty estimation in neural networks. For instance, Sicheng [1][28][29][30][31][32] demonstrated the use of Bayesian neural networks to improve model calibration and robustness by capturing epistemic and aleatoric uncertainty. Ensemble methods such as Random Forest, Gradient Boosting, and Bagging have been extensively studied for uncertainty quantification in machine learning models. Notably, the work by Yang [2] introduced Random Forest as a powerful ensemble technique for capturing model uncertainty and improving predictive performance. Research in reversible linguistic steganography and its applications in secure communication and data hiding has been explored by numerous scholars. Majid Khan [3] proposed a novel steganographic method based on linguistic patterns to embed and extract hidden information from text corpora. Recent advancements in deep learning-based steganography methods have shown promising results in encoding and decoding hidden information in text or images. The work by Lee. [4] introduced a deep learning approach for steganographic image generation, achieving high levels of imperceptibility and robustness against detection algorithms. Gaussian Processes, Variational Autoencoders, and Bayesian Neural Networks have been widely used for uncertainty prediction tasks in machine learning. For example, Chuen [5] presented a comprehensive study on Gaussian Processes for uncertainty estimation, highlighting their flexibility and applicability in various domains. Ensemble learning techniques have been leveraged to improve model robustness, accuracy, and generalization across different domains. The work by Mienve [6] provided insights into the benefits of ensemble methods in enhancing model performance and handling complex data distributions. Bayesian inference techniques play a crucial role in capturing uncertainty and making reliable predictions in machine learning models. Dietterich [7]. discussed the principles of Bayesian inference and its applications in model interpretation, parameter estimation, and uncertainty quantification. Monte Carlo Dropout has emerged as a valuable technique for uncertainty estimation in deep learning models. Saraf Ali [8] introduced Monte Carlo Dropout as a scalable method for uncertainty estimation in neural networks, enabling improved model reliability and decision-making. Deep Generative Models for Uncertainty Estimation: Generative adversarial networks (GANs) and variational autoencoders (VAEs) have been explored for uncertainty estimation in generative models. Ming Wei. [9]. introduced GANs as a framework for generating realistic synthetic data and discussed their potential for uncertainty quantification. Transfer learning techniques have been applied to calibrate model uncertainty and improve generalization across domains. Dong [10] presented a transfer learning approach for uncertainty calibration in deep learning models, showcasing its effectiveness in diverse application scenarios. Advancements in quantum computing have opened avenues for exploring uncertainty analysis and probabilistic modeling at quantum scales. Wanqar and Osama [11] introduced quantum algorithms for sampling and probabilistic reasoning, laying the foundation for quantum-based uncertainty estimation techniques. Information-theoretic metrics such as entropy, mutual information, and Kullback-Leibler divergence have been utilized for quantifying uncertainty and information content in machine learning models. Cover and Lorsung [12] provided a comprehensive overview of information theory and its relevance to uncertainty quantification in data science. Robust optimization methods and uncertainty-aware planning techniques have been applied to decision-making under uncertainty. Xiao Xiang. [13] discussed robust optimization frameworks for handling uncertain parameters and optimizing decision strategies in dynamic environments. Natural language processing (NLP) techniques have been employed for sentiment analysis, opinion mining, and uncertainty detection in textual data. Khantawala and Patel [14] reviewed methods for sentiment analysis and uncertainty modeling in NLP applications, highlighting the challenges and opportunities in this domain. Research on adversarial attacks and defenses has contributed to understanding model robustness and uncertainty in deep learning models. Gupta. [15] introduced adversarial examples and discussed strategies for improving model robustness and resilience to adversarial perturbations.

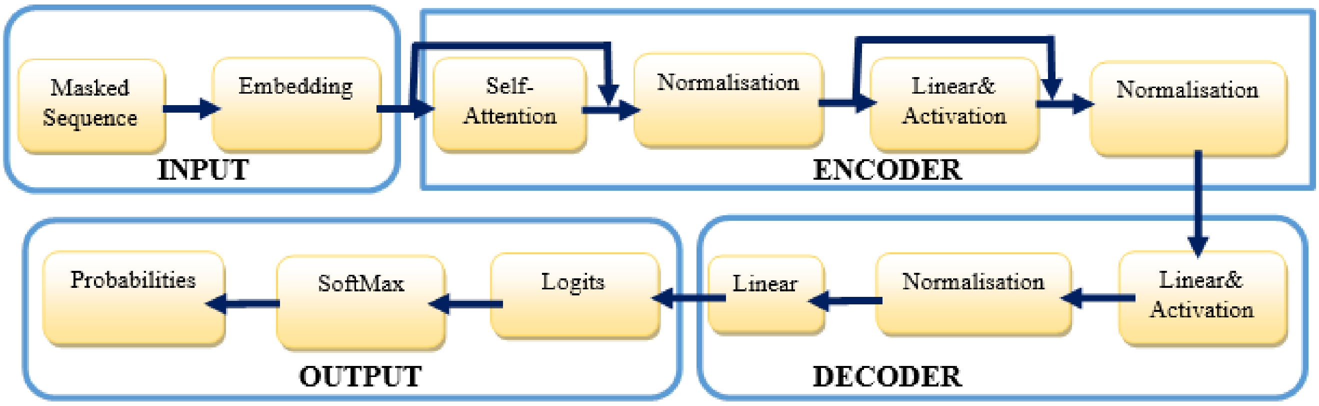

Figure 1: Architecture of BERT

III. MASKED LANGUAGE MODELLING

This is a task in which the model predicts a masked word based on surrounding contextual words. The model estimates the probability distribution of each mask token across the entire vocabulary.

Tokenization defines about the partitioning the string of text into tokens, which can be words, punctuation, numbers, etc. Tokens must be represented numerically for computational models. A popular representation is the one-hot vector, but it has limitations in capturing the semantics of words. Distributed representation, based on co-occurrence patterns, is another approach that considers words with similar contexts to have similar meanings. Early models such as word2vec and GloVe had limitations regarding polysemy (multiple meanings) and homonyms (same spelling, different meanings). Representing words in context, as in BERT, addresses these limitations by dynamically adapting word vectors to their context. In figure 1, Architecture of BERT is shown [27]. BERT is an advanced neural network architecture that uses a transformer architecture with a self-attention mechanism. It processes input sequences in two dimensions and is trained using a masked language model to learn how words are represented in context. Fine tuning of BERT’s adaptability is an inherent part of its design; users can train task-specific layers on top of the pre-trained model to be used in other tasks such as sentiment analysis, machine translation, and many others.

BERT was trained by predicting randomly masked words within a sentence. The steganographic system discussed in this context is based on BERT's implicit language modelling functionality. This involves leveraging BERT's ability to predict hidden words based on context, which can be used to encode and decode messages.

IV. CONTEXTUAL MASKING

Context masking is a technique that enhances mask language modelling capabilities. In contextual cloaking, a cover word in a text string is cloaked not only based on a fixed probability but also takes into account the surrounding context. This means that the decision to mask a word depends on its contextual relevance, making the masking process more dynamic and context-sensitive.

Unlike traditional cloaking methods that apply uniform probabilities to masked words, contextual masking introduces variation in the cloaking probability depending on the context of each word. Words that are more central in context or semantically important may have a higher probability of being hidden, while words that have less contextual meaning may have a lower probability of being hidden. By incorporating contextual information into the masking process, contextual masking improves the model's ability to capture nuanced semantic relationships and dependencies in text. This allows the model to focus more on important words for prediction while also providing the level of randomness needed to ensure reliability and unpredictability, which is important for applications of steganography. Figure 2 explains about the flow of contextual masking and ensemble methods. In the context of information hiding and steganography, contextual hiding adds an additional layer of complexity and security.

This ensures that the hidden message remains embedded in the text in a way that is not easily discernible without knowledge of the probability of obfuscation and the context in which the obfuscation occurred. This makes the steganographic system more resistant to detection and improves its effectiveness in covert communication situations.

V. COMPARISON WITH BERT AND CONTEXTUAL MASKING

A. Bert Masking

Masking in BERT offers simplicity and effectiveness, making it a popular choice in natural language processing (NLP) tasks. Its straightforward implementation and ability to handle general tasks where broader contextual

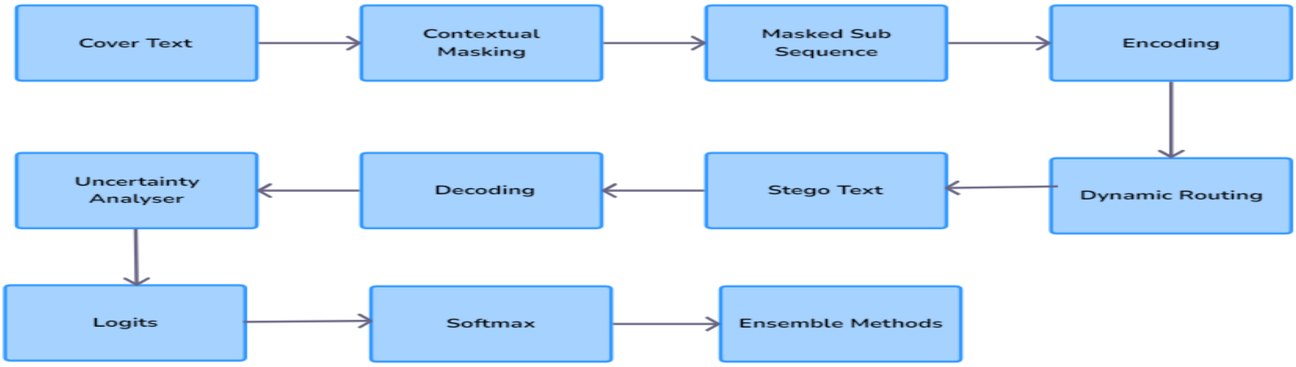

Figure 2: Reversible Linguistic Steganography based on Contextual Masking with Ensemble Methods

predictions suffice are notable advantages. Moreover, its widespread adoption and understanding within the NLP community contribute to its appeal. However, despite these strengths, BERT's masking technique has limitations. It may struggle to capture nuanced contextual information as effectively as methods that delve deeper into context, potentially leading to inaccuracies, particularly in tasks requiring fine-grained comprehension.

Additionally, its uniform masking probabilities might not appropriately prioritize important words within specific contexts, potentially impacting the model's ability to make accurate predictions. Furthermore, BERT's heavy reliance on pretraining data could limit its performance on tasks with domain-specific nuances or limited training data, necessitating additional fine-tuning or customization. Thus, while masking in BERT offers advantages, it's crucial to be aware of its limitations and consider alternative approaches for tasks requiring deeper context understanding or specialized knowledge.

B. Contextual Masking

Contextual masking is better than regular masking in some ways. It lets the model decide which words are most important in each sentence, which helps it understand the text better. This is useful for tasks where getting the small details right is really important. But it's harder to set up and needs more tweaking to work well. You have to be careful with how you adjust the settings to make sure the model learns properly. So, while it's a good method, it's a bit trickier to use than the simpler masking technique.

Contextual masking works well for tasks where getting every detail right is super important.

VI. REVERSIBLE LINGUISTIC STEGANOGRAPHY

Reversible linguistic steganography is a process that embeds a hidden message in the sequence while maintaining the minimum possible distortion of the sequence. The purpose is to enable the sender to communicate with the message recoveree by introducing recoverable distortions, which are

known as the steganographic distortion. The cover sequence termed as the “cover” is translated into a distorted version, which is referred to as the “stego.” The message that is hidden is a random binary sequence. The main objective of this entire operation is to provide maximum accurate extraction of the message and fruition of the text with the minimal steganographic distortion. To attain this, the approach uses a methodology rooted in contextual masking. More specifically, it does so by admitting a cover word in the text sequence; thereafter, the method masks the preferred word, generating a masked sub-sequence. The masked sub-sequence consists of a predetermined number of context words surrounding the masked word [27]. Afterward, the masked sub-sequence is channelled into language models that are pre-trained; Here, it is often engendered for masked language modelling. In the MLM framework, the method utilizes the masked sub-sequence to determine the probability distribution over the masked cover word. Subsequently, the probabilities generated are organized, which engenders a predictive permutation of words.

The intention of this arrangement is to discern an appropriate prediction; it should replace the masked cover word to represent a designated-message digit or part.

In linguistic steganography, with contextual masking uses techniques like masked language modelling and probability-based prediction to hide a message within a text while keeping the texts integrity and readability intact. Each main word is represented by an 'y' index in a list. These indices have a limit to ensure 'y' stays within bounds. If 'y' goes beyond this limit the word is skipped.

Here is the algorithm for both the encoding *Algorithm 1* and decoding *Algorithm 2*:

Algorithm 1 Theta Shift Cipher

```
Input: y, u, $msg_{rev}$, $count_d$
Output: y_prime, u, $count_d$
► encoding
u_prime = u. copy ()
if y< bound then
    if y< q then                                ◊carrier zone
        y_prime ← 2*y+msg[$count_d$]
        $count_d$ ← $count_d$+1
    else                                        ◊ non-carrier zone
        y_prime ← y+ q
        if (bound- q) ≤ y_prime < bound then
                                                ◊ambiguity zone
            u_prime. append (0)
        else if y_prime ≥ bound then            ◊ out of bound
            y_prime ← y
            u_prime. append (1)
else                                            ◊out of bound
    y_prime ← y
u. append(u)                                    ◊update flag list
```

Algorithm 2 Bound Decode Cipher

```
Input: y_prime, u, $msg_{rev}$, $count_d$
Output: y, $msg_{rev}$, $count_d$
► decoding
msg = [  ]
if y_prime< bound then
    if y< 2*θ then                         ◊carrier zone
        y ← int(y_prime/2)
        msg ←y_prime%2
    else                                   ◊ non-carrier zone
        y_prime ←y_prime-q
    if (bound- q) ≤y_prime < bound then
    ◊ambiguity zone
        if u[$count_d$]=1 then
            y_prime ←y
            $count_d$ ←$count_d$ - 1
else if y_prime ≥ bound then               ◊ out of bound
    y_prime ←y
$msg_{rev}$. append(msg)                   ◊update flag list
```

A threshold value 'θ' separates the index from main indices. The main indices form a set of size θ, which's crucial for ensuring reversibility in encoding. Encoding needs to be bijective to ensure a one-to-one match. Unique pairings between indices and message digits are created, with adjustments made to prevent confusion during message extraction. Flag bits are used to tell shifted and unshifted main indices within the confusion range. Shifting indices helps handle confusion and extra steps in the encoding process.

Following assumptions, about how index values occur less alteration is made for main indices when embedding information. If 'y' falls within θ both 'y'. The binary message digit 'm' is encoded into an index; if not, 'y' gets shifted by θ.

The encoding algorithm contains various zones like carrier, non-carrier, ambiguity and out of bound. Encoding condition check: if y is less than bound then proceed with encoding else exit the process. Carrier zone: It enters the carrier zone if y is less than 0., Refer to (1); non-carrier zone: If y is greater than or equal to 0, it enters non-carrier zone. Ambiguity zone: This depends on range y that is bound-0 to bound. Using (2) Out of bound: If y is greater than bound, it enters the out of bound. If none of the conditions met it enters out of bound condition.

$$y' = \{2y+m \quad if \quad y<\theta \qquad y+\theta \quad otherwise \tag{1}$$

Suppose the binary message digit to be embedded is is, denoted by msg. If y is within θ, we encode y and m into a stego index; otherwise, we shift x by θ .

$$v = \{0 \; if \; (\text{bound} - \theta) \leq y\acute{} < bound \qquad 1 \; if \; y\acute{} \geq bound) \tag{2}$$

If the stego index is out of bound, we reset it to its original value and record the cases by flag bit.

VII. STEGANOGRAPHIC ROUTING

Steganographic routing involves the concealment of hidden data, like messages or information, within the routing structures of a network or communication system. This process employs diverse methods to mask the presence of concealed data within the routing details, making it challenging for unauthorized entities to detect or interpret.

Dynamic routing, conversely, entails the automated modification of routing routes and decisions in response to real-time circumstances or fluctuations in network parameters. This approach enhances network efficiency, flexibility, and resilience by dynamically selecting the optimal routes for data transmission.

Steganographic routing could encompass the insertion of messages or information within routing protocols or data packets themselves, utilizing methods such as predictive word substitution or Bayesian uncertainty estimation. Dynamic routing might be employed to adapt routing paths based on imperceptibility metrics, capacity evaluations, or other considerations tied to steganographic communication.

VIII. BAYESIAN UNCERTAINTY QUANTIFICATION

Bayesian uncertainty quantification provides a way to incorporate and quantify uncertainty [21] in predictions by considering the distribution of model parameters and integrating over possible parameter values to compute the predictive distribution.

Bayesian statistics offers a probabilistic lens through which we can grasp and quantify the inherent uncertainty tied to model predictions. This predictive uncertainty, pivotal in Bayesian inference, is encapsulated within the prediction distribution. Specifically, for a masked sequence s and training set D, the predicted distribution of masked words can be mathematically represented as an integral, with θ symbolizing the model parameters within the Bayesian framework.

The concept of "average prediction over all plausible parameter settings based on the posterior parameter" delves into Bayesian inference's core philosophy. It involves weighing model parameter [22]-[23] values by their posterior probabilities, culminating in an average prediction that encapsulates the range of plausible outcomes.

Deriving posterior parameters analytically for deep learning models presents formidable challenges, often deemed "analytically difficult." To navigate this complexity, variational inference steps in. It approximates the intricate posterior distribution using a simpler variational distribution θ belonging to a family of distributions with computational advantages.

In tackling integrals, particularly those involving Bayesian models, Monte Carlo integration emerges as a potent technique. By sampling from the variational distribution, we obtain estimates of the integral, offering an approximation of the true predictive distribution. This approach empowers us to navigate the complexities of Bayesian statistics, capturing and navigating the uncertainties inherent in predictive modeling. Using (3) we calculate as follows

$$p\,(y=w_i \mid s, D) \approx \sum Tt=1 \; 1/T \; p\,(y= w_i \mid s, \theta\hat{}t) \tag{3}$$

where θ^t represents the sampled model parameters from variation distributed in each random transition.

In the masked language model, the possibilities p (y= w_i ∣ s, θ ^t) are calculated using the SoftMax function.

In the masked language model, the likelihood p (y= w_i ∣ s, θ ^t) is calculated using the SoftMax function.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)
© Copyright: All rights reserved.

where: f_i represents the ith logit (raw prediction) of the model for word w_i.

The SoftMax function normalizes the log exponential to obtain probabilities, ensuring that the sum of probabilities is 1 for all words in the dictionary.

Shannon's entropy is a measure of the uncertainty or information content in a probability distribution. Using (4) Shannon Entropy is calculated as follows:

$$H(X)=-\sum i \; P(x_i)\cdot \log 2 \; (P(x_i)) \qquad (4)$$

Specifically, when applied to a predictive distribution Shannon entropy quantifies the uncertainty associated with predicting the next word in a sequence given certain input information. The concept is rooted in Claude Shannon's work on information theory, where entropy represents the average amount of information produced by a stochastic process or the average surprise in an event's outcome. In the context of a predictive distribution, higher entropy indicates greater uncertainty or unpredictability in the predictions. Conversely, lower entropy suggests more certainty or predictability in the predictions [24]. For example, if a predictive distribution assigns similar probabilities to multiple possible outcomes (words in this case), the entropy would be higher because there is more uncertainty about which outcome will occur. On the other hand, if one outcome is highly probable and others are improbable, the entropy would be lower, indicating less uncertainty. So, when the text refers to measuring the uncertainty underlying the predictive distribution by Shannon entropy, it means quantifying how much uncertainty or information is contained in the distribution of predictions.

A. Drawbacks of Monte Carlo

Monte Carlo is not cost efficient. It is not applicable for real time data compared to ensemble methods. This technique involves applying dropout during inference and averaging the predictions over multiple dropout samples (stochastic forward passes) to estimate uncertainty. While Ensemble methods create diversity by training multiple models on different subsets of data or using [25] different algorithms, and then combining their predictions. This diversity helps capture uncertainty.

IX. ENSEMBLE METHODS

Techniques like bagging, boosting, and stacking can be used to create an ensemble of models. Each model in the ensemble may provide different predictions due to variations in training data, model architecture, or hyperparameters.

Improved Accuracy: Ensembles often yield better accuracy compared to individual models, especially when the individual models are diverse and complementary.

Uncertainty Estimation: Ensemble methods inherently capture uncertainty by considering [17] multiple predictions and aggregating them, providing a more robust estimate of uncertainty compared to a single model.

Monte Carlo Dropout: This technique involves applying dropout during inference and averaging the predictions over multiple dropout samples (stochastic forward passes) to estimate uncertainty.

Ensemble Methods: Ensemble methods create diversity by training multiple models on different subsets of data or using different algorithms, [15] and then combining their predictions. This diversity helps capture uncertainty.

Complexity: Monte Carlo dropout can be simpler to implement compared to creating and managing an ensemble of models.

Effectiveness: The effectiveness of Monte Carlo dropout vs. ensemble methods depends on the specific task, dataset, and model architecture. In some cases, ensemble methods may outperform Monte Carlo dropout, while in others, Monte Carlo dropout may be sufficient or more suitable. Monte Carlo dropout is a specific technique for uncertainty estimation, ensemble methods can also be used effectively to capture uncertainty and improve prediction accuracy.

A. Advantages of Ensemble Methods

Improved Accuracy: Ensembles often yield better accuracy compared to individual models, especially when the individual models are diverse and complementary [20].

Uncertainty Estimation: Ensemble methods inherently capture uncertainty by considering multiple predictions and aggregating them, providing a more robust estimate of uncertainty compared to a single model.

X. RANDOM FOREST CLASSIFIER

Random Forest is indeed a bagging technique, specifically an ensemble of decision trees created using bootstrap aggregation (bagging) [18]-[19]. Each tree in the random forest is trained on a random subset of data and features, and their predictions are combined by averaging (for regression) or voting (for classification) to make the final prediction.

Here is pseudo code for RANDOM FOREST CLASSIFIER:

- Initialize Parameters:
- Import necessary libraries (NumPy for array operations, RandomForestClassifier from sklearn. ensemble).
- Prepare Data:
- Convert SoftMax probabilities (softmax_probabilities_list) into a NumPy array X_train.
- Prepare labels (y_train) based on whether the sequence contains '[MASK]' (1 for yes, 0 for no).
- Train Random Forest Classifier:
- Initialize a Random Forest classifier (rf_classifier) with specified parameters
- Fit the classifier using the training data.
- Predict Uncertainty:
- Use the trained classifier to predict uncertainty (uncertainty predictions) on the training data (X_train).
- Print the uncertainty predictions.

Here is the comparison graph figure 3 showing the results of the Random Forest Classifier and Monte Carlo

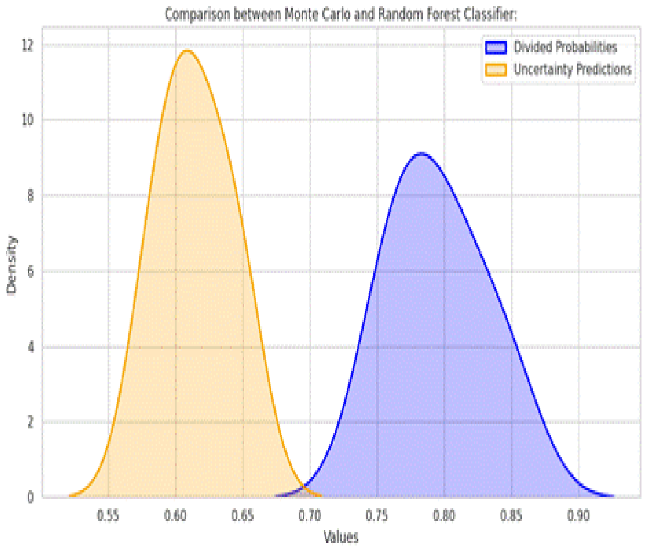

Figure 3: Random Forest and Monte Carlo Comparison

XI. EXPERIMENTS

We proposed a stego system with limit and threshold value parameters. We also use dynamic routing, which strikes a balance between imperceptibility, reversibility, and capacity. In addition, various improvements have also been provided.

A. Experimental Setup

Cover text refers to a piece of text used for analysis or processing. In this case, it is eight passages selected from "Alice's Adventures in Wonderland", a classic work of English literature. The cover text contains 711 words as well as punctuation. Text is converted to lowercase for consistency. The BERT model shows that the BERT (Bidirectional Encoder Representation from Transformer) model is used which has a vocabulary size of 30,522 tokens. This vocabulary represents the unique words and entities that the model understands. Specifies that the size of the context window around each target word is set to 32 words per side. This means the model considers the 32 words before and after the target word when making predictions, creating hidden sequences of 65 words each.

Monte Carlo dropout samples: Specifies that the number of Monte Carlo dropout samples, which are stochastic transitions used to estimate model uncertainty, is set to 1000. This facilitates Convenient for evaluating model prediction variability.

B. System Evaluation

Our steganographic system is comprehensively evaluated across different parameter configurations to evaluate the trade-off in capacity, invisibility, and reversibility. Capacity measurements, including payload bits and payload bits per word, are essential metrics. Indiscernibility is evaluated using a cosine similarity measure, which highlights the quality of concealment.

Reversibility, which is important for data integrity, is quantified by the flag bits used to resolve word index conflicts.

Capacity Analysis: System capacity, which is important for data integration, is evaluated using payload bit metrics.

Evaluation of imperceptibility: The effective data hiding ability of the system is evaluated using cosine similarity calculations.

Reversibility assessment: The system's reversibility, ensuring correct data recovery, is analyzed through the use of flag bits.

Exploring tradeoffs: Balancing capacity with imperceptibility and reversibility reveals nuanced tradeoffs important for system optimization.

Impact of routing strategy: Dynamic routing strategies outperform static routing strategies, especially in adaptive handling of predictive uncertainties.

Random forest classifier integration: Random Forest classifier integration helps predict uncertainties, thereby improving route decisions to improve system performance.

Impact of contextual hiding: [16] The impact of contextual hiding on imperceptibility and reversibility highlights its importance in improving the quality of data hiding.

C. Semantic Analysis

Semantic analysis evaluates the [27] similarity and quality of the stego text to the cover text and suggests refinement strategies to improve the naturalness and grammatical correctness of the stego text. Word clouds are used to visualize the frequency distribution of words in cover text. Stego text is generated using random parallel routing with specific parameters: $\theta = 1$ limit = 270 $\tau = 1$. These parameters affect affects how information is integrated into stego text. Capacity is set to 0.

Capacity-Reversibility and Capacity-Imperceptibility graph is shown in figure 4 and figure 5 respectively.

3 bits per word, allowing for a payload of more than 200 bits. This capability is considered sufficient for many authentication applications, demonstrating the potential utility of the steganographic method.

The text mentions that perfect reversibility is guaranteed without any overhead information.

This means that the original cover text can be completely recovered from the stego text without the need for additional information. Similarity and unnatural word usage: The analysis notes that the cover text and stego text are similar in terms of semantics and word frequency distribution. However, upon closer inspection, we noticed unnatural word usage and grammatical errors in the stego text.

To improve the quality of stego text, the analysis proposes a screening method: Filtering of grammatical words and named entities: This step aims to remove words without content can contribute to a lack of naturalness. Focus on content words (words that convey meaning) for manipulation that can improve the naturalness of stego text.

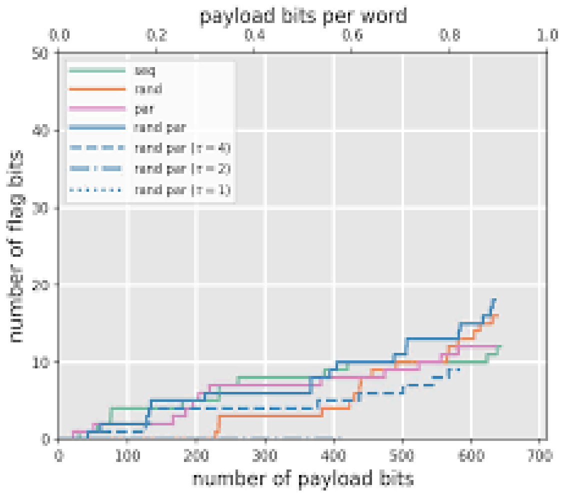

Figure 4: Capacity-Reversibility bound=27 & θ=bound/3

XII. CONCLUSION

In this work, we introduce a novel approach to linguistic steganography, leveraging advanced ML techniques for reversible coding and uncertainty quantification.

By embedding hidden data within the routing mechanisms of a network, we achieve enhanced security and privacy in data transmission. The use of dynamic routing further optimizes network efficiency and adaptability, ensuring robust performance even in changing network conditions.

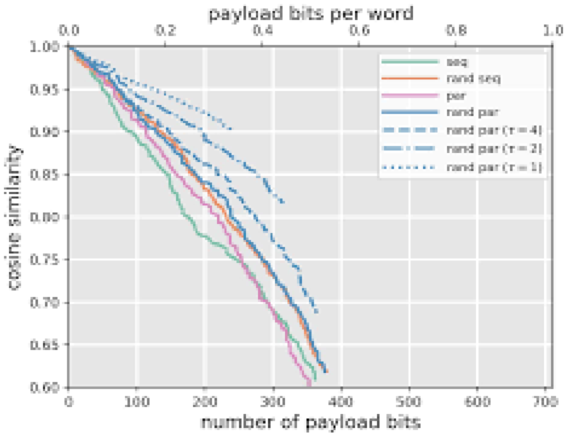

Figure 5: Capacity-Imperceptibility bound=27 & θ=1

Through the implementation of predictive word substitution and Bayesian uncertainty quantification, our steganographic routing system achieves imperceptibility while maintaining high capacity for data embedding. The dynamic routing component adapts routing paths based on real-time metrics, ensuring reliable and secure communication channels.

Our language stego system, based on predicted word substitution using a pre-trained masked language model, where a contextual mask is used to mask the text Remarkable inversion without auxiliary information, maintaining semantic and emotional consistency between cover text and stego. Using a Bayesian uncertainty quantification approach, we establish an adaptive integration pathway that ensures robust detection resistance while optimizing computational efficiency. Firstly, refining imperceptibility to mitigate even subtle steganographic distortions remains a priority, necessitating deeper analysis of linguistic patterns and syntactic structures. Additionally, advancing uncertainty analysis in real-time scenarios can optimize computational resources, making our system more practical for diverse applications. Exploring syntactic and generative [26] methods alongside lexical substitution promises a more comprehensive steganographic framework adaptable to varied linguistic contexts.

These enhancements collectively aim to fortify our stego system's efficacy, paving the way for robust and covert communication in digital environments. Overall, our proposed system presents a promising avenue for secure and efficient data communication, with potential applications in areas such as secure messaging, data privacy, and confidential information exchange.

Future work may focus on refining the routing algorithms, exploring additional steganographic techniques, and evaluating the system's performance in diverse network environments. The below figure 6 and figure 7 represents about the word cloud for the cover text and stego text respectively.

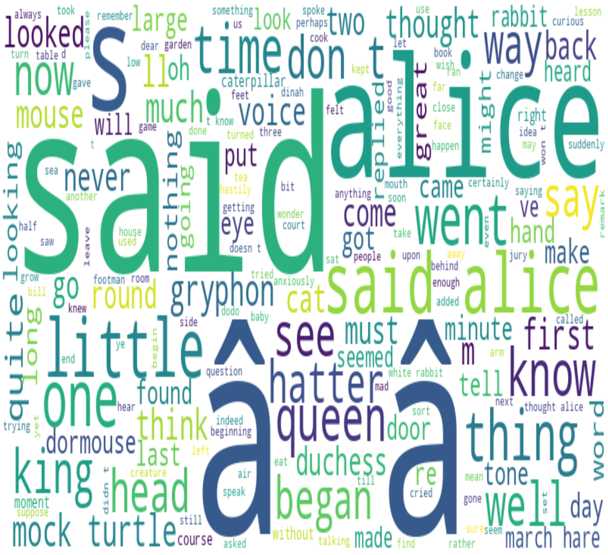

Figure 6:Word Cloud for Cover Text

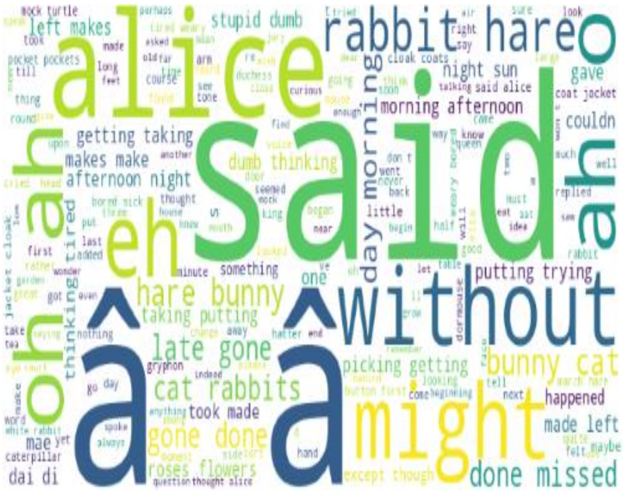

Figure 7:Word Cloud for Stego Text

DECLARATION STATEMENT

All authors have contributed equally to this article. To the best of our knowledge, there are no conflicts of interest among the authors regarding this paper. The article does not require ethical approval, and consent to participate is not applicable with evidence. The article did not receive any funds, grants, or financial support for this work.

Funding	No, I didn't receive.
Conflicts of Interest	No conflicts of interest to the best of our knowledge.
Ethical Approval and Consent to Participate	No, the article does not require ethical approval and consent to participate with evidence.
Availability of Data and Material/ Data Access Statement	Credit Card Transactions Fraud Detection Dataset (kaggle.com)
Authors Contributions	All authors have equal participation in this article.

REFERENCES

1. "Deep Learning for predictive Analysis in Reversible steganography" Ching-Chun Chang; Xu Wang; Sisheng Chen; Isao Echizen; Victor Sanchez; Chang-Tsun Li.,IEEE: 2020
2. "VAE Stega: Linguistic Steganography based on Variational Auto Encoder", Zhong-Liang Yang; Si-Yu Zhang; Yu-Ting Hu; Zhi-Wen Hu; Yong-Feng Huang.,IEEE: 2020
3. "Introduction to Linguistic Steganography", Majid Khan, Ali Shahab*, and Zeeshan Asghan
4. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," Jacob Devlin Ming-Wei Chang Kenton Lee Kristina Toutanova
5. "Reversible steganography techniques: A survey," Tzu-Chuen Lu, Thanh Nhan Vo Chaoyang University of Technology, Department of Information Management, Taichung, Taiwan, R.O.C
6. "A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects" Ibomoiye Domor Mienye; Yanxia Sun.,IEEE: 2020
7. "Ensemble Methods in Machine Learning", Thomas G. Dietterich.IEEE: 2020
8. "Ensemble learning methods for decision making: Status and future prospects", Shahid Ali; Sreenivas Sremath Tirumala; Abdolhossein Sarrafzadeh
9. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," Jacob Devlin Ming-Wei Chang Kenton Lee Kristina Toutanova
10. " Uncertainty in Bayesian Reinforcement Learning for Robot Manipulation Tasks with Sparse Rewards", Li Zheng; Yanghong Li; Yahao Wang; Guangrui Bai; Haiyang He; Erbao Dong
11. " Bayesian Learning for Uncertainty Quantification, Optimization, and Inverse Design", Madhavan Swaminathan; Osama Waqar Bhatti; Yiliang Guo; Eric Huang; Oluwaseyi Akinwande
12. " Understanding Uncertainty in Bayesian Deep Learning", Cooper Lorsung
13. " A Survey of Uncertainty in Deep Neural Networks", Jakob Gawlikowski, Cedrique Rovile Njieutcheu Tassi, Mohsin Ali, Jongseok Lee, Matthias Humt, Jianxiang Feng, Anna Kruspe, Rudolph Triebel, Peter Jung, Ribana Roscher, Muhammad Shahzad, Wen Yang, Richard Bamler, Xiao Xiang Zhu
14. " Theoretical Evaluation of Ensemble Machine Learning Techniques", Milind Shah; Kinjal Gandhi; Kinjal A Patel; Harsh Kantawala; Rohini Patel; Ankita Kothari
15. "Ensemble Learning Techniques and its Efficiency in Machine Learning: A Survey", Thomas N. Rincy; Roopam Gupta
16. " A Random Forest Classification Algorithm Based on Dichotomy Rule Fusion", Yueyue Xiao; Wei Huang; Jinsong Wang
17. " Review of random forest classification techniques to resolve data imbalance", A. S. More; Dipti P. Rana
18. " A novel improved random forest for text classification using feature ranking and optimal number of trees", Nasir Jalal, Arif Mehmood, Gyu Sang Choi, Imran Ashraf
19. " A Review on Random Forest: An Ensemble Classifier", Aakash Parmar, Rakesh Katariya & Vatsal Patel
20. " Application Research of Text Classification Based on Random Forest Algorithm", Yanxiong Sun; Yeli Li; Qingtao Zeng; Yuning Bian
21. "Deep Generative Models for Uncertainty Estimation", Ian J. Goodfellow et al.
22. "Transfer Learning for Uncertainty Calibration" Sinno Jialin Pan and Qiang Yang
23. "Information Theory and Uncertainty Metrics", Thomas M. Cover and Joy A. Thomas
24. "Robust Optimization and Uncertainty-Aware Planning", Dimitris Bertsimas, David B. Brown, and Constantine Caramanis
25. "Natural Language Processing for Uncertainty Analysis", Bo Pang and Lillian Lee
26. "Adversarial Attacks and Model Robustness", Christian Szegedy et al.
27. "Reversible Linguistic Steganography with Bayesian Masked Language Modeling", Ching-Chun Chang et al.
28. Younis, Z., Kafri, N., & Hasouneh, W. (2022). A Framework for Sentiment Analysis Classification based on Comparative Study. In International Journal of Soft Computing and Engineering (Vol. 12, Issue 2, pp. 7–15). https://doi.org/10.35940/ijsce.a3524.0512222
29. Kumthekar, A., & G, R. R. (2019). Ensemble Learning Technique for Cloud Classification. In International Journal of Engineering and Advanced Technology (Vol. 9, Issue 2, pp. 2582–2587). https://doi.org/10.35940/ijeat.b3957.129219
30. Jindam, S., Challa, S. T., Chada, S. J., B, N. S. B., & Malgireddy, S. (2023). Prediction of Software Defects using Ensemble Machine Learning Techniques. In International Journal of Recent Technology and Engineering (IJRTE) (Vol. 11, Issue 5, pp. 58–65). https://doi.org/10.35940/ijrte.e7421.0111523
31. Kumthekar, A., & G, R. R. (2019). Ensemble Learning Technique for Cloud Classification. In International Journal of Engineering and Advanced Technology (Vol. 9, Issue 2, pp. 2582–2587). https://doi.org/10.35940/ijeat.b3957.129219
32. S, D., & P L, L. (2020). Binary Class Classification of Software Faults in Software Modules using Popular Machine Learning Techniques. In International Journal of Innovative Science and Modern Engineering (Vol. 6, Issue 6, pp. 14–18). https://doi.org/10.35940/ijisme.f1221.046620

AUTHORS PROFILE

Mrs. M. Prasha Meena, M.E., working as Assistant Professor, Department of Information Technology, Mepco Schlenk Engineering College (Autonomous), Sivakasi. I have done my UG and PG research works on Image processing, video processing, Neural Networks, Deep learning and Pattern recognition. I have presented my works in 3 International conferences and published my works in 4 International Journals.

Deepalakshmi N.J.S, currently pursuing B.Tech in Information Technology at Mepco Schlenk Engineering College, embodies a passion for technology and academic excellence. With adept problem-solving skills and an analytical mindset, she contributes actively to her academic and research endeavors. Having conducted research on Big Data in 2023, her proactive learning approach and enthusiasm for innovation underscore her potential contributions to the IT domain. Eagerly anticipating her graduation in 2025, Deepalakshmi is poised to embark on a journey of impactful contributions, driven by a strong desire to excel in her academic and professional pursuits.

Dharsni R, currently pursuing a B-Tech in Information Technology at Mepco Schlenk Engineering College, Sivakasi. Throughout my coursework, I have delved into various subjects, including programming languages, database management systems, network security, and software engineering methodologies. I am so glad to be co-authored with this paper. I am currently in my penultimate year of study, with a projected graduation year of 2025. I have already co-authored a paper on medical data security. During my academic tenure, I have actively participated in several noteworthy projects, showcasing my practical skills and theoretical knowledge.

R Subashree currently pursuing B.Tech in IT at Mepco Schlenk Engineering College, embodies a fervent passion for the latest advancements in Information Technology (IT) and Artificial Intelligence (AI). With a voracious appetite for learning, I delve into the depths of these emerging technologies, eager to unravel their intricacies. Beyond IT and AI, my interests extend to Full Stack Development and Machine Learning (ML), where I aspire to master the art of creating seamless, innovative solutions. Fueled by curiosity and driven by ambition, I am poised to make meaningful contributions to the ever-evolving landscape of technology, shaping the future with each step I take.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of the Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)/ journal and/or the editor(s). The Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP) and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.