International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Adaptive Multiscale Transformer Network with Bi-LSTM-based Neural Machine Translation Model using Attention Vector for Named Entity Recognition with Adolescent Suicidal Text

[1]K. Soumya*, [2]Dr. Vijay Kumar Garg

Submitted:11/03/2024 **Revised**: 26/04/2024 **Accepted**: 03/05/2024

Abstract- Suicide thoughts impact language usage stated on the internet. Many at-risk people utilize social discussion sites offering information about similar tasks. Our study aims to share ongoing research on automatically identifying suicidal comments. Deep learning classification techniques identify adolescents with suicidal thoughts in their early stages. The text data is gathered from standard data sources. The obtained textual data undergoes data pre-processing to remove redundant and inappropriate data. The pre-processed text data is given as input to the Adaptive Multiscale Transformer Network with Bidirectional Long Short Term Memory (AMTN-Bi-LSTM) for Named Entity Recognition (NER). The developed AMTN-Bi-LSTM consists of a Transformer Unit and a Bi-LSTM unit. At first, the pre-processed text data is given to the Transformer Network for Neural Machine Translation (NMT). During the translation time, the identified named entities are to be monitored via the specific process of translation model that helps to improve the quality. This Transformer Network with an inbuilt self-attention mechanism produces the pre-processed text's attention vectors as output. This attention vector of the text is now given to the encoder section of the Bi-LSTM. The encoded vector is then fed to the decoder section of the Bi-LSTM, from which the essential suicidal text words are recognized. For improved recognition, the parameters in the developed AMTN-Bi-LSTM model are tuned with the help of the Fitness Improved COOT (FICOOT) algorithm. The recognized text is given as input to the encoder unit of the Trans-Bi-LSTM, from which the given text is classified as a non-suicidal or suicidal class. The potential operation of the developed NER model for suicidal word recognition is verified by comparing the recommended method with the conventional models regarding various performance metrics.

Keywords- *Named Entity Recognition; Machine Translation; Adaptive Multiscale Transformer Network with Bidirectional Long Short Term Memory; Neural Machine Translation; Fitness Improved COOT; Trans-Bi-LSTM based Recognition;*

1. Introduction

Data extraction's NER task categorizes and extracts particular types of organizations, such as proper numbers, temporal expressions and titles [1]. In most research, names of places, organizations and people are typically regarded as correct titles. Money, percentage expressions, and numbers are often covered under numeric gestures, whereas time and date expressions are categorized as temporal phrases [2]. Due to the unlimited number of members of these kinds, employing simple search lists for recording the phrases in the flowing text is inadequate, making the NER task a non-trivial operation. It is not practical to create a huge list that contains all potential person names that might appear in a text [3]. The organization and geographical names also hold to this truth. A large NLP system typically includes a NER system as an initial processing step [4]. The NER network's quality directly impacts the effectiveness of the whole NLP system. Numerous research attempts were made to demonstrate the value of NER to additional NLP tasks including decision-making support system, Question Answering (QA) and knowledge graph[5]. Using NER as an initial processing stage enhances the search accuracy for the result clustering task and the previous NLP tasks [6].

Suicidal thoughts are defined as a propensity to take one's own life and can commit suicide to have a strong obsession with suicide. Suicide is the next most common cause of mortality among teenagers and is among the top causes of death across all ages [7]. Social media, which young people mostly use, has become a potent "window" into the emotional and physical well-being of those who use it in the past few years [8]. The most significant risk factors for thoughts or attempts at suicide in teenagers are emotion deregulation and social exclusion linked to depression and anxiety. Considering that the circle of friends becomes the main social touchstone during teens, individuals may be especially conscious of the need to belong. Treatments that work to stop teenage thoughts of suicide before they manifest into actual suicidal behavior are crucial. Social media platforms enable anonymous involvement in many online communities, creating a forum for public discourse on socially taboo subjects. Typically, 50% of people who successfully commit themselves and more than 20% of those who attempt suicide write suicide notes [9].

[1] *Research Scholar,School of Computer Science and Engineering, Lovely Professional University - Punjab, and Vignana Bharathi Institute of Technology – Hyderabad, India.* [ORCID:0000-0001-7364-0795]

[2] *Professor,School of Computer Science and Engineering, Lovely Professional University - Punjab, India.* [ORCID: 0000-0001-5926-71620]

* *Corresponding author's Email: kothapallisowmya@gmail.com*

Therefore, any written indication of suicidality is considered concerning, and the writer should be asked about their thoughts. It can reduce any incorrect text interpretations provided by a retrospective examination compared to an offline text [10]. The study of social media and associated venues for discussing mental health has grown in digital linguistics. It offers a useful research environment for creating cutting-edge technical innovations that could revolutionize suicide risk reduction and suicide detection. It can provide a useful entry point for intervention. The purpose of NER is to automatically gather and categorize NEs into predefined classes, such as time, place, person, organization and date [11]. In many NLP applications, including information retrieval and machine translation NER is a vital preprocessing stage that enhances the application's performance [12].

One genuine language is transformed automatically into another natural language using MT, a branch of NLP. A well-known method of MT that exhibits impressive translation performance due to its contextual analyzing potential is DNN-based MT, additionally referred to as NMT. Multimodal NMT (MNMT) also tries to gather data from many modalities. For example, while training an NMT approach, image and text data are combined. Systems for NER are being developed that combine mathematical models like neural networks with linguistic grammar-based approaches [13]. Handmade grammar-based systems often attain greater precision, albeit at the expense of recall and months of labor by skilled computational linguists [14]. The majority of training data for statistical NER systems must be manually marked. Semi-supervised methods have been proposed to reduce some of the annotation work. Deep learning techniques have already been achieved in pattern recognition and computer vision, independent of conventional text categorization techniques [15]. Conventional methods for machine learning mainly rely on frequently insufficient handmade features and are labor-intensive; however, neural networks based on dense representations of vectors can perform better on various NLP tasks [16]. Compared to more conventional machine learning methods DCN and word embedding perform better at estimating the probability of suicide [17].

The advanced model's major significance is clarified in the following manner.

- To develop a NER with adolescent suicidal text using AMTN-Bi-LSTM; it helps to identify and classify specific expressions related to suicidal thoughts, behaviors, or emotions expressed by adolescents.
- To gather the required data from standard datasets, this involves gathering relevant textual information from established and reliable data sources that are suitable for the intended analysis.
- To perform data pre-processing on the collected text data to remove redundant and inappropriate information. This step aims to clean the text and ensure its quality and consistency before further analysis.
- To utilize the AMTN-Bi-LSTM model for NER. The pre-processed text data is inputted into the model to identify and extract essential suicidal text words, finally obtained recognized text and optimize the parameters of the developed model using the FICOOT algorithm.
- To provide the recognized text from the previous step as input to the encoder unit of the Trans-Bi-LSTM method, where the machine translation is taken place. The method classifies the given text into the categories of suicidal or non-suicidal based on the learned patterns and characteristics.
- To verify the potential operation of the developed NER model for suicidal word recognition by comparing it with conventional models using various performance metrics. This comparison aims to assess the superiority and efficiency of the suggested model in terms of its performance on suicidal word recognition.

The parts that follow have the right names. The study's assessment of the most recent NER with adolescent suicidal text is covered in Section II. The transformer-based network presented in Section III: the adaptive NER idea with adolescent suicide text. The AMTN-Bi-LSTM is provided in Section IV to acquire the attention vectors. The FICOOT algorithm and encoder unit of the Trans-Bi-LSTM for identification is explained in Section V. Section VI presents the simulation's operation and results. Section VII presents the results of the planned task.

2. Existing Works

2.1 Related Works

In 2023, Ghosh et al. [18] have presented Bangla social media texts using an attention-based Bi-LSTM-CNN that performed superior to conventional approaches and was lighter and more resilient. To address the shortage, a dataset of these Bengali texts was developed in this effort. Three embedded data was employed in this job along with various processing phases. The suggested model had a 94.3% error rate, 92.63% sensibility, and 95.12% specificity. The suggested model performed wonderfully when tested on language other than its native tongue, including English. This paper also explored the suggested solution model's explainability and resilience. The suggested model outperformed traditional ensemble

methods, machine learning models, transformers, comparable designs, and current architectures.

In 2023, Rahman et al. [19] have investigated by resolving word-order divergence problems and data scarcity, and multimodal NMT was performed for the English-Assamese languages couple, a low-resource variety of languages pair. A transliteration-based phrase enhancement strategy was put out to address these problems. It took advantage of the sub-word frequency token sharing between source-target sequences during the training and added phrase pairs to offer additional word alignment information. Additionally, the pertinent image attributes related to the word pairings were increased by considering a filtering phase. Modern multimodal NMT results was achieved for each direction of English-Assamese pair translating using the suggested methodology.

In 2022, Zhao et al. [20] have proposed Region-Attentive NMT (RA-NMT) was a MNMT approach utilizing semantic picture areas. Studies on MNMT that was already conducted primarily concentrated on using equally sized grid local or global visual features generated by CNNs to enhance translating performance. However, they disregard the impact of semantic data encoded in visual characteristics. This study combined textual and visual data with three modality-dependent attention techniques, using semantic picture areas recovered by object identification for MNMT. Self-Attention Network (SAN) and "Recurrent Neural Network (RNN)," two neural architectures used in NMT, was used to build and verify the suggested technique. The suggested method beat most of the cutting-edge MNMT methods, according to findings from experiments on various language pairs of the multi30k dataset. Further investigation revealed that the suggested methods improved usage of visual features led to better-translating efficiency.

In 2019, Allen et al. [21] have explored how to solve the underlying issues in the immediate prediction of risk for suicide by utilizing new advancements in real-time surveillance techniques and computer analysis. With little participant burden, they was undertake extensive longitudinal assessments of potential risk factors using wearable computing, smart home and smartphone technology. They was also created algorithms that predicted for STBs using cutting-edge computational methods. Because of the emerging capabilities of novel technologies, recent experiment on immediate risk management for STBs revealed that this was a crucial study area for the future. Despite the huge potential for new information that these methodologies hold, there is little existing empirical research.

In 2019, Ali and Tan [22] have proposed a bilateral decoder-decoder system that uses bidirectional LSTMs as the decoder and encoder to handle the issue of Arabic NER using contemporary deep learning research. Character-level embedded data was used in word-level embedded data, coupled using an embedding-level attentiveness method. Through this attention system, the model dynamically decided what data from a character or word level element must be used. The framework with a bi-encoder-decoder network and embedded attentiveness layer produced a high F-score measurement of over 92%, according to the experiments conducted on the AQMAR and ANERCorp datasets.

In 2023, Priyamvada et al. [23] have worked relating to the automatic flagging and identification of suicidal posts. It described a method for analyzing Twitter, a social networking site, to find signals suggesting someone could consider suicide. The method's main goal was to automatically detect aberrant changes in a user's online conduct. Identified and spotted the numerous risk variables or indicators that preceded the occurrence posed obstacles for suicide prevention. Several NLP techniques was applied to this assignment to measure textual and linguistic changes and pass them via a unique framework that was widely used. With categorization models developed using machine learning and deep learning applied to tweets on the social media platform Twitter, the initial detection of self-harm thoughts was accomplished.

In 2021, Belouali et al. [24] have developed a classification system assisted on speech indicators for NLP and machine learning to check for suicidal thoughts among US soldiers. For suicide prevention and early identification, screening for thoughts of suicide in high-risk populations like American veterans was essential. At present, clinician interviews or self-report tests are used for testing. Both methods depend on the participants coming up with their suicide ideas. To create therapeutically useful tests, novel methods was required. Using language as an empirical marker to comprehend various emotions, including thoughts of suicide, was studied.

In 2022, Kodati and Tene [25] have identified unpleasant feelings in suicide posts that discussed the mental state of the poster. They suggested two models for identifying unfavorable emotions on social media: tension, rage, guilt, melancholy, despair, and anxiety. The first model, called C-BiGRU-MHA-CNN, was designed to preserve contextual data. The Self-Attention (SA) and Masked Language Modelling (MLM) mechanism procedures were suggested in this research to train systems and identify contextual characteristics over free-of-context systems. They also recommended the L-BiLSTM-MHA-CNN method, which was lexicon-based bilateral long short-term memory plus multiple heads of attention. When comparing the lexicon-based technique to more traditional methods, just one channel-based model outperformed all.

2.2 Research Gaps and Challenges

Since most of the people with suicidal ideation isolate themselves and prefer to be alone, there is a minimum chance in identifying their actual ideation. Since not much data are available, the sentiment and behavior analysis process of an individual is not much effective. Also, the existing suicide detection methods fall back of highlighting certain key factors. So, deep learning technique has been introduced in detecting suicidal ideation in individuals. The features and challenges of the conventional NER-based suicidal ideation detection model are shown in Table 1. Att- BiLSTM [18] performs better than the existing ensemble model. But, the various nature of the text cannot be classified provided with a huge dataset.BRNN and Transformer [19] can effectively translate the text in multimodal data both in the forward and backward direction. Yet, this method cannot process huge amounts of multi-modal data. RA-RNN [20] shows more effective and competitive performance than many existing state-of-the-art methods. This model is insensitive to semantic features in the image region. This model is highly stable. This method provides good image information. However, more fine visual information cannot be used in this model. STB [21] can perform intensive longitudinal evaluation to monitor the risk aspects with fewer participants. But, there is no necessary amount of empirical literature to execute this method to its full potential. Bidirectional encoder–decoder, and LSTM [22] models have higher F1-score. However, this model requires a large amount of annotated data for processing. "Stacked CNN-2 layer LSTM "[23] can detect the abnormal behavior changes in the online user automatically. But, the mixed of deep learning with the machine learning methods makes this process computationally complex. Ensemble model [24] allows the detection of veterans with suicidal ideation by using their everyday speech behaviors. Yet, only the RF technique is able to achieve a better result. C-BiGRU-MHA-CNN, BERT, and L-BiGRU-MHA-CNN [25] efficiently identify users with suicidal thoughts in online platforms. This method works effectively on identifying people with suicidal thoughts even by capturing their emotions on text sequence data. Yet, the computational time required by this model is high. Therefore, using cutting-edge methods of deep learning, a simple, efficient, and straightforward model for identifying teenagers with thoughts of suicide is going to be created.

Table 1. Features and challenges of existing NER-based suicidal ideation detection models

Author [citation]	Methodology	Features	Challenges
Ghosh *et al.* [18]	Attention-based BiLSTM-CNN	This model performs significantly better than the current ensemble approach.	The various nature of the text cannot be classified provided with a huge dataset.
Rahman *et al.* [19]	BRNN and Transformer	This model can effectively translate text in multimodal data forward and backward directions.	A huge size of multi-modal data cannot be processed by this method.
Zhao *et al.* [20]	RA-RNN	This method shows more effective and competitive performance than many existing state-of-the-art methods. This model is insensitive to semantic features in the image region. This model is highly stable. This method provides good image information.	More fine visual information cannot be used in this model.
Allen *et al.* [21]	STB	This method can perform intensive longitudinal evaluations to monitor the risk aspects with fewer participants.	There is no necessary empirical literature to execute this method to its full potential.
Ali and Tan [22]	Bidirectional encoder-decoder and LSTM	The F1-score obtained by this model is much higher.	This model requires a large amount of annotated data for processing.
Priyamvada *et*	Stacked CNN-2	The abnormal behavior changes in the online user can be detected	This procedure is highly complex computationally since deep

al. [23]	layer LSTM	automatically by this method.	learning and machine learning techniques are combined.
Belouali *et al.* [24]	Ensemble model	This method allows the detection of veterans with suicidal ideation using their everyday speech behaviours.	Only the RF technique can achieve a better result.
Kodati and Tene [25]	C-BiGRU-MHA-CNN, BERT, and L-BiGRU-MHA-CNN	The performance of this model in identifying users with suicidal thoughts online is higher. This method works effectively in identifying people with suicidal thoughts even by capturing their emotions on text sequence data.	The computational time required by this model is high.

3. Adaptive Concept of Named Entity Recognition with Adolescent Suicidal Text: Transformer-based Network

3.1 Architecture Representation of Named Entity Recognition

Adolescent suicide is a serious and distressing issue that requires prompt attention and intervention. Identifying individuals who may be at risk is essential to provide them with the necessary support and assistance. With the NLP techniques, entity recognition has emerged as a valuable tool in identifying potential suicidal ideation and related sentiments within text data. Detecting entities related to suicidal ideation and self-harm in the adolescent text can provide significant benefits. It can aid mental health professionals, counselors, and parents identify warning signs and potential risk factors, allowing for timely intervention and preventive measures.

On the other hand, entity recognition with adolescent suicidal text raises significant privacy and ethical concerns. Analyzing sensitive text data related to suicide risk requires strict adherence to privacy regulations and guidelines to protect the confidentiality and well-being of individuals involved. Care must be taken to ensure that data is handled securely and that informed consent is obtained when applicable. Reducing these kinds of problems by using the proposed model helps improve the entity recognition system. Fig 1 shows the framework of the developed FICOOT NER with the adolescent suicidal text model.

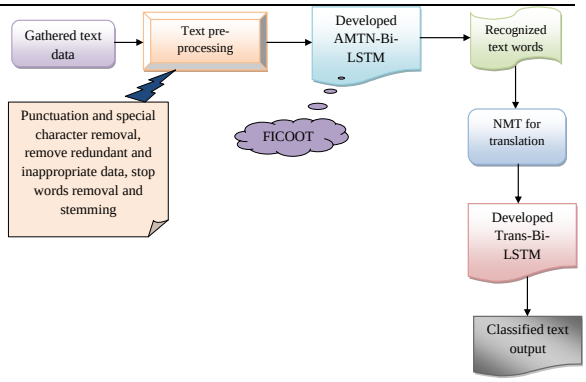


Fig 1. The framework of developed FICOOT NER with adolescent suicidal text model

The proposed system consists of several steps; initially, it gathers the required data from standard datasets. The collected data is fed as input to the pre-processing; this helps to clean the text and ensure its quality and consistency before further analysis. The pre-processed text data is inputted into the AMTN-Bi-LSTM model to identify and extract essential suicidal text words. This step involves NMT using the Transformer Network, which produces attention vectors as output. Feed the attention vectors obtained from the Transformer Network into the encoder section of the Bi-LSTM model. The encoder section will process and encode the vectors, then passed to the decoder section of the Bi-LSTM. The decoder section aims to recognize essential suicidal text words based on the encoded vector and optimize the parameters of the developed model using the FICOOT algorithm. Once the recognition is to be done, the MT is required for enriching the system efficiency. It also helps to resolve the data sparsity issue and preserve such context in the word document. The recognized text is fed as input to the encoder unit of the Trans-Bi-LSTM model; the model classifies the given text into the categories of suicidal or non-suicidal based on the learned patterns and characteristics. In summary, the objectives include data collection, data pre-processing, NER using AMTN-Bi-LSTM, incorporating a Transformer Network for NMT, parameter tuning with FICOOT algorithm, classification of

suicidal vs. non-suicidal text, and performance evaluation of the proposed model compared to traditional models.

3.2 Details of Raw Text

Dataset 1 ("Stress Analysis in Social Media Dataset"): The data from the link "https://www.kaggle.com/datasets/ruchi798/stress-analysis-in-social-media?select=dreaddit-train.csv: access date: 2023-07-04". This is a dataset of lengthy multi-domain social media data from five different categories of Reddit communities, specifically focused on identifying stress. This dataset can be valuable for developing stress detection models and gaining insights into stress-related discussions across different domains.

Dataset 2 ("SDCNL Dataset"): The data from the link "https://github.com/ayaanzhaque/SDCNL: access date: 2023-07-04". This dataset specifically addresses the classification problem of distinguishing between depression and more severe suicidal tendencies using web-scraped data, particularly from Reddit. The dataset aims to tackle this issue by leveraging Reddit data and implementing a novel label correction method to mitigate inherent noise in the data using unsupervised learning techniques.

From the mentioned data resources, the gathered data are signified as, x_m here $m = 1,2,...M$, and total acquired data is expressed as M .

3.3 Text Pre-processing

The textual data x_m is given as input to this phase; for preparing raw textual information for analytic or modeling tasks, a set of methods and procedures are referred to as text preprocessing. Unstructured text is converted into an organized form better suited for NLP activities. The purpose of text pre-processing is to eliminate noise, standardize the text, and reduce the complexity of the data. Raw text often contains punctuation and special character removal, remove redundant and inappropriate data, stop words removal and stemming, which can hinder analysis and affect the accuracy of models. Text pre-processing techniques help address these challenges and enhance the quality of the data. The definition of these methods is explained below.

Punctuation and Special Character Removal: The textual data x_m is given as input to this phase; an important stage in the pre-processing of text is the removal of commas and special characters, which removes those characters from the raw text information. Commas, periods, points of exclamation, markup, and question marks are punctuation symbols. Special characters encompass non-alphanumeric or non-textual symbols, such as dollar signs, hash tags, mathematical symbols or percent signs. Removing punctuation marks and special characters during text pre-processing makes the resulting clean and standardized text more amenable to various "NLP tasks, including sentiment analysis, text classification, information retrieval, and language modeling". Finally obtained, the special and punctuation removed data, which is represented as ps_m^{data} .

Remove Redundant and Inappropriate Data: The special and punctuation removed data ps_m^{data} is the input to this phase; process of removing redundant and inappropriate data involves careful comparison, assessment and examination of the dataset. Techniques like similarity analysis, deduplication, human moderation and text matching can be employed to identify and remove redundant or inappropriate instances. By eliminating redundant and inappropriate data during the preprocessing stage, the resulting dataset becomes more refined, reliable, and suitable for subsequent analysis tasks, promoting accurate insights and ethical considerations in text analysis and NLP applications. Finally it obtained the redundant and inappropriate removed data that is presented as Ri_m^{data} . This is the input to the next process.

Stop words Removal: The redundant and inappropriate removed data Ri_m^{data} is the input to this phase; stop words removal refers to eliminating commonly used words considered insignificant or non-informative in NLP tasks. These words, known as stop words, typically include common articles, prepositions, pronouns, and conjunctions that appear frequently in a language but carry little or no contextual meaning. Finally obtained the stop word removed data, which is represented as Sw_m^{data} .

Stemming: The stop word removed data Sw_m^{data} is the input and stemming is a useful technique for reducing word variations and capturing the underlying meaning of words, enabling more effective text analysis and information retrieval. However, it's essential to consider its limitations and potential impact on specific NLP tasks, as stemming may not always be suitable or sufficient for certain applications that require more precise linguistic analysis. Finally obtained the preprocessed data Pr_m^{data} .This step aims to clean the text and prepare it for further analysis.

4. Adaptive Multiscale Transformer Network with Bidirectional Long Short Term Memory for Obtaining the Attention Vectors

4.1 Trans-Bi- LSTM

Transformer [26]: The encoding component and decoder are the two primary parts of the transformer design. Each

part comprises numerous levels of self-attention neural networks and feed-forward. While the decoding device creates the output series based on the coded data, the encoding process analyzes the input series. The transformer's self-attention system makes it possible to represent long-range relationships in the input series than conventional serial models. It enables the model to focus on pertinent input irrespective of location or separation from the currently analyzed word.

Initially, the input of preprocessed data $\Pr_m^{data}$ is transformed into tokens by lowering it into several 2D patches $\{V_q^a \in G^{q2} \mid b = 1,.....v\}$ that are numerous in both the patch count $v = \frac{mk}{Y^2}\Pr_m^{data}$ and patch size $k \times k$.

The Multi-Layer Perceptron (MLP) and several patching blocks make up each layer of the Transformers encoder. Thus, the following can be written as Eq. (1) and (2).

The components that make up each layer Y of the Transformers decoder are the "Multi-Layer Perceptron (MLP) and Multi-head Self-Attention (MSA) a"blocks. As a result, what follows can be written as Eq. (1) and (2).

$$H_p^1 = MSA(SD(\Pr_m^{data} - 1)) + \Pr_m^{data} - 1 \quad (1)$$

$$H_p = MmP(SD(\Pr_m^{data})) + \Pr_m^{data} \quad (2)$$

The normalization operations layer in the equation above is called SD the picture encoder. One of the key advantages of transformers is their parallelizability, as the self-attention mechanism can be computed in parallel for all positions in the sequence. This significantly speeds up training and inference compared to sequential models like RNNs.

Bi-LSTM [27]: A variation of the LSTM design frequently employed in deep learning for serial information processing is called Bi-LSTM. Bi-LSTMs are very good at collecting future and past context, which makes them suitable for jobs demanding comprehension of directional relationships.

Two different layers of LSTM operate simultaneously as a Bi-LSTM. The input sequence is processed by one layer in the initial sequence, capturing connections from the past toward the future, and by another layer in the opposite sequence, catching connections from the distant future to the past. The last depiction of an input series is created by concatenating or combining the results will of the two layers.

The LSTM's loop core is depicted, and Eq. (3) outlines the procedures for selecting each among the several gate kinds.

$$Y(h) = \sigma(R_i T(h) + W_i T(h-1) + \Pr_m^{data} \quad (3)$$

The expressions used in Eq. (9), the term $T(h)$ which refers to the "input traits, the output of the preceding cell $T(h-1)$, the function's sigmoid value σ, the quantity of weighing from the gate" that accepts the parameter, the input data is $\Pr_m^{data}$, each symbol for different input characteristics. Eq. (4) depicts an ignored gate.

$$Y(h) = \sigma(R_g T(h) + W_g T(h-1) + \Pr_m^{data} \quad (4)$$

The terms R_g and W_g are each cell in the equation above must be negligible to ignore the gate's weight. The gate of energy is expressed in Eq. (5) and Eq.(6).

$$\overline{F}(h) = \tan m(R_E T(h) + U_E T(h-1) + \Pr_m^{data}) \quad (5)$$

$$j(h) = N(h) * d(h-1) + a(h) * j(h) \quad (6)$$

Eq. (5) and Eq. (6) characterize the cell's updating procedures in accordance; Eq. (5) describes a possible store component that provides alternative update data, while Eq.(6) describes a state update procedure. A knowledge base and fresh conditions from the gates of recalling are mixed to produce the most recent information. These values correspond to the proportional weight of the Hadamard product x_D and alternative new state R_D. The outcome of the gate computation is shown in Eq. (7).

$$\sigma(h) = \sigma(x_D T(k) + R_D N(h-1) + \Pr_m^{data} \quad (8)$$

A bi-LSTM may be better able to collect context data than a standard LSTM. Time-series data sets are utilized to gather information about a specific time from the past to the future. The neural network can more properly believe time sequences because to this. This serves as the following model's input. Fig 2 shows the architecture representation of the Trans-Bi- LSTM model for attention vectors.

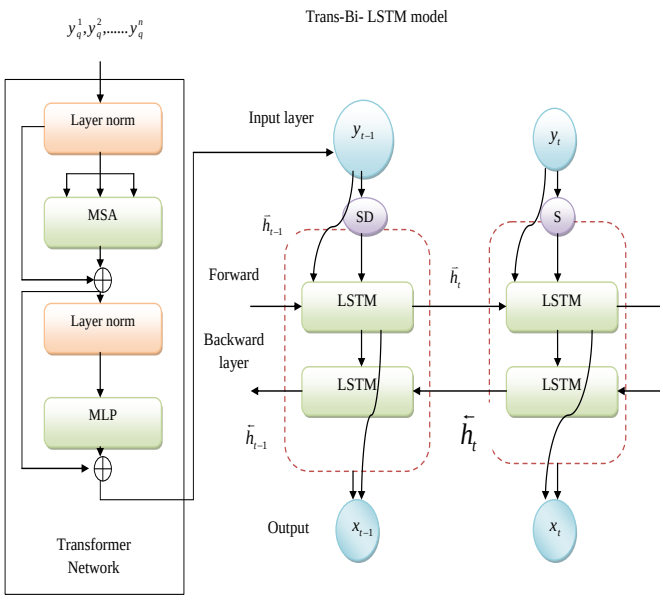


Fig 2. The architecture representation of the Trans-Bi-LSTM model for attention vectors

The model parameters of transformer and Bi-LSTM are given in Table 2.

Table 2. Model Parameter Settings

Network	Model Parameters	
Transformer	Layers	2
	Attention Heads	4
	Hidden neuron counts	64
	Epsilon	1e-6
	Activation Function	ReLU
Bi-LSTM	Gates	6 gates(2 input gates,2 forget gates,2 output gates)
	Hidden neuron counts	64
	Activation Function	Softmax
	Epoch	5
	Steps Per Epoch	5

4.2 Multiscale Trans-Bi-LSTM

Multiscale [28]: The preprocessed data $\Pr_m^{data}$ is given to the input to this phase in data processing; multiscale refers to the analysis or manipulation of data at multiple levels of resolution or granularity. It involves examining data at different scales to capture both global and local patterns, allowing for a more comprehensive understanding of the information. Another approach to multiscale data processing is using filters or transformations with different window sizes or scales. Applying filters of varying sizes to the data makes it possible to extract information at different levels of detail. For example, in image processing, a multiscale filtering approach may involve applying filters of different sizes to capture fine textures and larger structures.

The Trans-Bi-LSTM is mathematically defined in section A above. The output of multiscale is given to a transformer network with a self-attention mechanism to produce attention vectors of the text. This step involves NMT using the transformer network, which produces attention vectors as output. Feed the attention vectors obtained from the transformer network into the encoder section of the Bi-LSTM model. The encoder section process and encode the vectors, which is then passed to the decoder section of the Bi-LSTM. The decoder section aims to recognize essential suicidal text words based on the encoded vector. Finally, the model obtains the recognized text that is represented as Rt_m^{data} .

4.3 Adaptive MTN-BiLSTM

The AMTN-Bi-LSTM system is designed to address the challenges of the NER system by leveraging the strengths of both architectures. AMTN-Bi-LSTM combines the power of two popular deep learning architectures: the Bi-LSTM model with the Transformer network. Identifying and classifying identified entities in text, including names of people, companies, places, and further, is the goal of a NER system, a subtask of data extraction. The computationally demanding nature of a model rises with the number of hidden neurons, needing greater storage capacity and longer training cycles. If the steps per epoch are too high and the model repeatedly sees the same training examples, it can increase the risk of overfitting and also longer training times are required when using a higher number of epochs, especially for large datasets or complex models. To reduce these kinds of errors, optimized using some parameters like hidden neuron count, steps per epochs, no. of epochs in Bi-LSTM this helps to improve the word error rate. The objective function of the adaptive MTN-BiLSTM is mathematically modeled in Eq. (9).

$$O_{b1} = \underset{\left\{hd^{Bi-lstm}, nep^{Bi-lstm}, sep^{Bi-lstm}\right\}}{\arg\max} \left[WER\right] \tag{9}$$

From the above equation, the variable $hd^{Bi-lstm}, nep^{Bi-lstm}$ and $sep^{Bi-lstm}$ signifies the hidden neuron count, steps per epochs, no. of epochs in Bi-LSTM. The range of $hd^{Bi-lstm}, nep^{Bi-lstm}$ and $sep^{Bi-lstm}$ is [5-225], [5-50] and [5 - 50]. The term WER signifies the word error rate and the overall number of mistakes is

divided by the total number of phrases in the source text, and the result is expressed as a percentage. The number of substitutions, insertions, and deletions required to transform the system output into the reference text are counted to compute the WER. The mathematical model of WER is shown in Eq. (10).

$$WER = \frac{s_k + D_k + I_k}{N_k} \tag{10}$$

From the above equation, the variable s_k, D_k and I_k shows the substitution, deletion and insertion also the term N_k signifies the number of words. Fig 3 shows the implemented adaptive MTN-BiLSTM model framework with the parameter optimization of FICOOT.

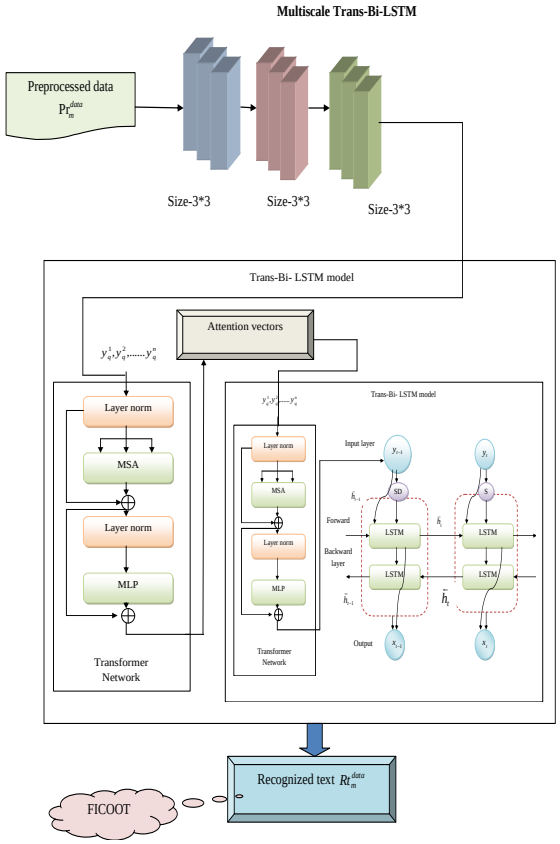


Fig 3. Implemented adaptive MTN-BiLSTM model framework with the parameter optimization of FICOOT

4.4 Machine Translation

The named entity recognized words may have the possibility of creating the data sparsity. Since it occurs as the same context, it affects the quality of the NER and translation as well. Therefore, the NMT is suggested to address the issues by replacing the named entities. Some of the examples are given to represent the process of translation.

When the source [29] is taken as "Such article explains the "Co-operation in agriculture" of the Programme of the European Initiative Interreg II Italy - Albania, being implemented in Apulia since 2000".

Here, the named entities are identified and replaced with the holders and its NE type (Date, Location, Organization). This is further used to translate into "reduced source sentences and original named entities". Reduced Src: Such article explains the "Co-operation in agriculture" of the Programme of the European Initiative Interreg II +NE_LocOrg_Country - + NE_LocOrg_Country, being implemented in +NE_LocOrg_City since +NE_Date.

NEs: Italy [LocOrg_Country], Albania [LocOrg_Country], Apulia [LocOrg_City], 2000 [Date].

Reduced translation model: It is then applied to the acquired source of reduced sentence that generates a reduced translation sentences. It is given as "cet article illustre "la coopération en agriculture " du programme de l' initiative interreg II + NE_LocOrg_Country - + NE_LocOrg_Country, mis en oeuvre à + NE_LocOrg_City depuis +NE_Date".

Furthermore, an external NE translation is employed for translating the replaced NEs; in turn, the multiple NE translators are being used to depend on the nature of NE. Thus, it can said to be as untranslated or transliterated (eg. PERSON). The translated words are shown below:

Translated words: "Albania=Albanie, Italy=Italie, Apulia=Pouilles, 2000=2000".

Hence, the completed translation is written as "cet article illustre "la coopération en agriculture" du programme de l' initiative interreg II Italie - Albanie , mis en oeuvre à Pouilles depuis 2000" .

Like as the above example, the named entities are obtained from the input text data. For implementing the model, the below given text is taken.

"Ex Wife Threatening SuicideRecently I left my wife for good because she has cheated on me twice and lied to me so much that I have decided to refuse to go back to her. As of a few days ago, she began threatening suicide. I have tirelessly spent these paat few days talking her out of it and she keeps hesitating because she wants to believe I'll come back. I know a lot of people will threaten this in order to get their way, but what happens if she really does? What do I do and how am I supposed to handle her death on my hands"

In the above context, the named entities like suicide, threatening or threaten, cheated and death on my hands are identified through the translation approach, which is further used in the proposed model.

This is the way of translation mechanism is included with the NER to get the precise outcome and also enhance the system effectiveness. Like the above example given, NERs

are denoted as Rt_m^{data} that is fed into the MT model to get the translated sentence. It is marked as Tr_m^{data} .

5. Elucidation of Fitness Improved COOT Algorithm and Encoder unit of Trans-Bi-LSTM for Recognition

5.1 Novel Heuristic Approach: FICOOT

The conventional algorithm COOT created the recommended FICOOT approach [30]. COOT algorithm is designed to efficiently solve large-scale optimal transport problems. It employs an iterative optimization approach that converges to the optimal solution while minimizing computational overhead. On the other hand, the COOT algorithm focuses on solving the optimization problem and finding the optimal solution but does not provide explicit interpretability. Understanding the underlying reasons behind the optimal transport mappings may require additional analysis and interpretation steps. To reduce these drawbacks, the new algorithm is proposed. The new formulation for the variable R_{and} is mathematically modeled in Eq. (11)

$$R_{and} = \frac{B_f / W_f}{W_f / B_f} \tag{11}$$

Here, the variable W_f and B_f denotes the worst and best fit. The term R_{and} defines the random number.

A Metaheuristic algorithm for optimization was used in the COOT method to model the various collective behaviors of Coots, a small aquatic bird in the rail family. Coots move regularly and erratically on the surface of the water, with the ultimate goal of getting to or away from food. The specific process of implementation is as follows, and the sample is generated using a random initialization method following Eq. (12):

$$Coot(i) = [R_{and}(1,d) \times (bu - bl) + bl \tag{12}$$

Here, the term $Coot(i)$ denotes the coot's location; d is the total amount of factors or problem measurements. To address the complexity and optimality issues associated with random values ranging from 0 to 1 in conventional algorithms, reduced using an updated formulation for generating random numbers R_{and} and the word bu and bl denote the lower and upper bounds of the search area and shown in Eq.(13) and Eq.(14).

$$bu = [bu_1, bu_2 ..... bu_d] \tag{13}$$

$$bl = [bl_1, bl_2 ..... bl_d] \tag{14}$$

The number of animals is initialized, and then the coot's location is updated based on one of four patterns of motion.

Random Movement: Applying Eq. (15), a location Q for this trend is initialized arbitrarily first.

$$W = R_{and}(1,d) \times (bu - bl) + bl \tag{15}$$

To avoid getting stuck in the vicinity of the optimal, the location is updated following Eq. (16):

$$Cootpos(i) = \frac{CSpos(i-1) + CSpos(i)}{2} \tag{16}$$

Adjusting position: A coot bird's location changes depending on the leader's location during each group, therefore one who follows goes in the leader's direction. The person in charge is chosen by Eq. (17).

$$L = (1 + (i \bmod KL) \tag{17}$$

Here, the term L is shown as the leader indexing number, i is the coot bird following number, and KL represents the total number of leads.

Leader Movement: Eq. (18) is used to update leadership locations based on the transition from local to worldwide optimal positions:

$$LPos(i) = \begin{cases} c \times c_3 \times \cos(r \prod R) \\ f_{bes} - Lpos(i) + f_{bes} \qquad R_4 < 0.5 \\ c \times c_3 \times \cos(r \prod R) \\ f_{bes} - Lpos(i) + f_{bes} \qquad R_4 \geq 0.5 \end{cases} \tag{18}$$

Here, R_3 as well as R_4 are the numbers chosen at random within [0, 1], R is a random number between [-1, 1], and f_{bes} refers to the most advantageous location that may be determined. Fig 4 shows the flowchart for recommended FICOOT.

Algorithm 1 : FICOOT

Consider the $\max_{iter}$

Determine the pop_{siz}

Updated random value R_{and} using Eq. (11)

Do while

$i > I_{\max}$

The generated random initialization method is specified in Eq. (12):

If $R_{and} > W$

R_3 and R_4 are random vectors along the dimension of the problem

Else

R_3 and R_4 are random values

End

If $R_{and} > 0.5$

The random movement of both upper and lower bounds are in Eq.(15)

Else

The adjusting position is shown in Eq.(16)

End

If $R_{and} < 0.5$

Eq. (18) is used to update leadership locations based on the transition from local to worldwide optimal positions:

End

End

End

Output the best fitness and best individual

Returns the optimal results

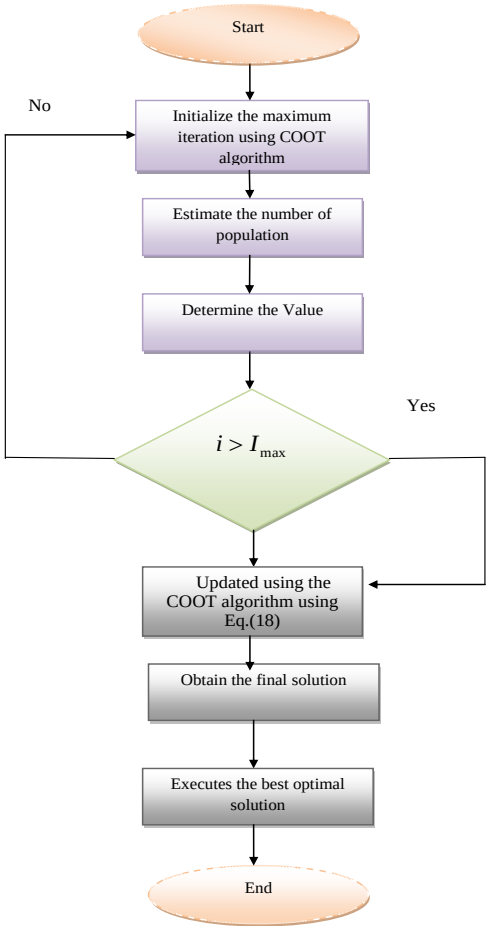


Fig4. Recommended FICOOT flowchart

5.2 Trans-Bi-LSTM based Recognition

The encoder unit of a Trans-Bi-LSTM model is a component that processes input sequences and generates a fixed-length representation or context vector capturing the information from the input sequence. In the context of the provided information, the translated sentence Tr_m^{data} obtained from the previous step of NER using the AMTN-Bi-LSTM model, is given as input to the encoder unit of the Trans-Bi-LSTM. In summary, the encoder unit of the Trans-Bi-LSTM takes the recognized text as input, applies a transformer-based encoder, and generates a context vector. This context vector is then used to classify the input text into suicidal or non-suicidal categories. Fig 5 shows the pictorial presentation of Trans-Bi-LSTM based recognition for NER with adolescent suicidal text.

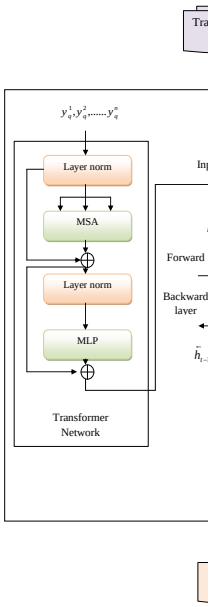


Fig 5. A pictorial presentation of Trans-Bi-LSTM based recognition for NER with adolescent suicidal text

6. Results and Discussions

6.1 Experimental setup

The suggested FICOOT-based NER with adolescent suicidal text system was conducted in Python, and the necessary research was done. The model that was created has 10 populations and the 50 best iterations. Additionally, the chromosomal length for the system in question was 3. Several techniques were used. The conventional models such as "Reptile Search Optimization (RSA)-AMTN-Bi-LSTM [31], Honey Badger Optimization (HBA)-AMTN-Bi-LSTM [32], Galactic swarm optimization (GSO)-AMTN-Bi-LSTM [33], COOT-AMTN-Bi-LSTM [30] was utilized. Then the traditional classifier models such as "Atten-BiLSTM [18], RA-RNN [20], Bi-EnDe [22] and MTN-Bi-LSTM were applied.

6.2 Performance measures

Accuracy: It is estimated in Eq. (18).

$$Ac = \frac{BC + DE}{BC + DE + NM + PQ} \tag{18}$$

Precision: It is formulated in Eq. (19).

$$P = \frac{BC}{BC + DE} \tag{19}$$

NPV: It is expressed in Eq. (20).

$$NPV = \frac{BC}{BC + NM} \tag{20}$$

FPR: It is shown in Eq. (21).

$$FPR = \frac{PQ}{PQ + BC} \tag{21}$$

F1-Score: It is estimated in Eq (22).

$$F1Score = 2 \times \frac{QP \times NM}{QP + NM} \tag{22}$$

FDR: It is estimated in Eq (23).

$$FDR = \frac{PQ}{BC + PQ} \tag{23}$$

Sensitivity: It is estimated in Eq (24).

$$Sen = \frac{NM}{NM + PQ} \tag{24}$$

Specificity: It is estimated in Eq (25).

$$spec = \frac{BC}{BC + DE} \tag{25}$$

Recall: It is estimated in Eq (26).

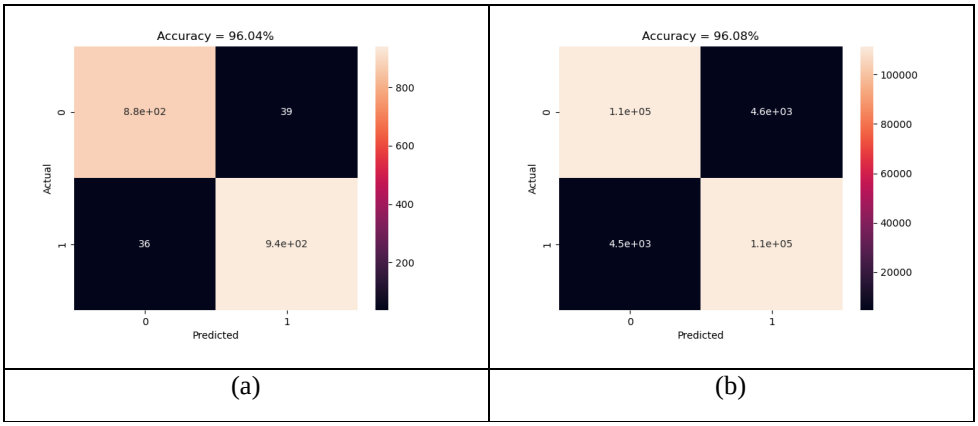
$$\mathrm{Re} = \frac{NM}{PQ + NM} \tag{26}$$

MCC: It is estimated in Eq (27).

$$MCC = \frac{PQ \times NM - MN \times NM}{\sqrt{(PQ + BC)(PQ + BE)(BC + DE)(Ik + PQ)}} \tag{27}$$

6.3 Evaluation of the confusion matrix for the developed FICOOT model over two datasets

In Fig. 6, a comparison of the proposed NER with adolescent suicide text technique with other methods utilizing various actual and expected rates can be seen. The NER matrix of confusion has a 96.04% correctness rate. When called the entity identification model, this suggests that it has greater accuracy and a lower percentage of false positives and false negatives.



(a) (b)

Fig 6. Confusion matrix analysis of the suggested FICOOT-based NER with adolescent suicidal text system concerning "(a) Dataset 1, and (b) Dataset 2".

6.4 Cost function analysis of two datasets for the developed FICOOT model over various algorithms

The expense measurement for both data sets has been used to examine the suggested approach, and the results are shown in Fig. 7. The expense measure is contrasted to many conventional methods in the suggested model. The cost function is changed from 1.00 to 1.10 while the iteration counts range from 0 to 25. When the iteration value is 25 for Fig. 7 (b), the cost function of the generated model is lowered by 54.9% of RSA-AMTN-Bi-LSTM, 45.5% of HBA-AMTN-Bi-LSTM, 78.9% of GSO-AMTN-Bi-LSTM, and 89% of COOT-AMTN-Bi-LSTM, correspondingly. This result demonstrated that the suggested approach has very limited cost effectiveness.

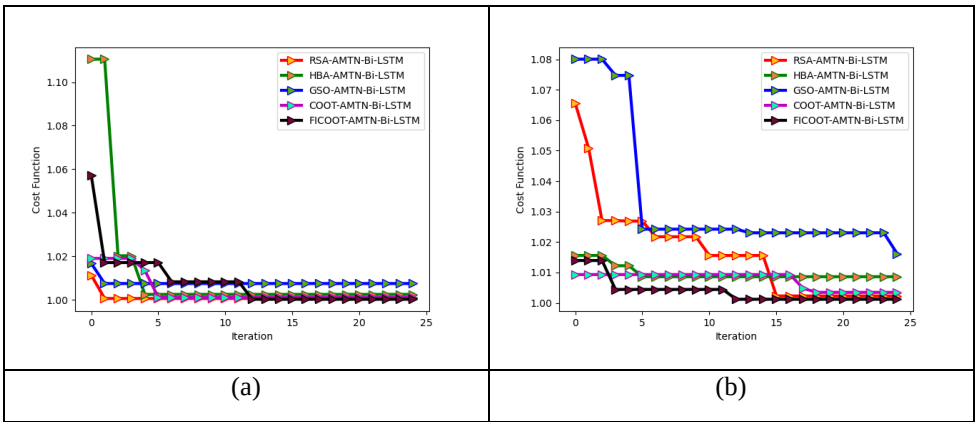


Fig 7. Cost function analysis of the suggested FICOOT-based NER with adolescent suicidal text system regarding " (a) Dataset 1, and (b) Dataset 2".

6.5 The ROC analysis of the proposed NER with the adolescent suicidal text system

The suggested method's ROC analysis is contrasted with those of other popular classifiers in Fig. 6. The true positive rate of the developed technique is shown in Fig 8. Using Atten-BiLSTM, RA-RNN, Bi-EnDe and MTN-Bi-LSTM, the true positive rate value was 56%, 89.2%, 24.8%, and 67.6% higher than the suggested method. Consequently, the system's true-positive rate ensures better recognition of the named entity.

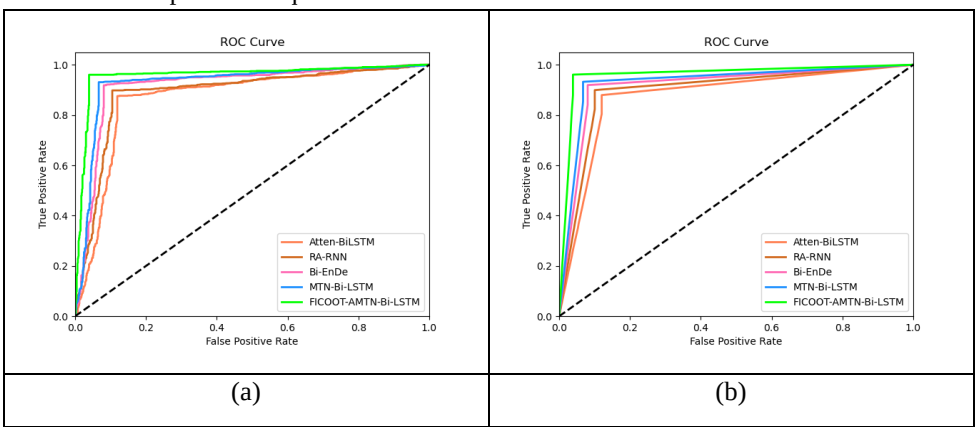
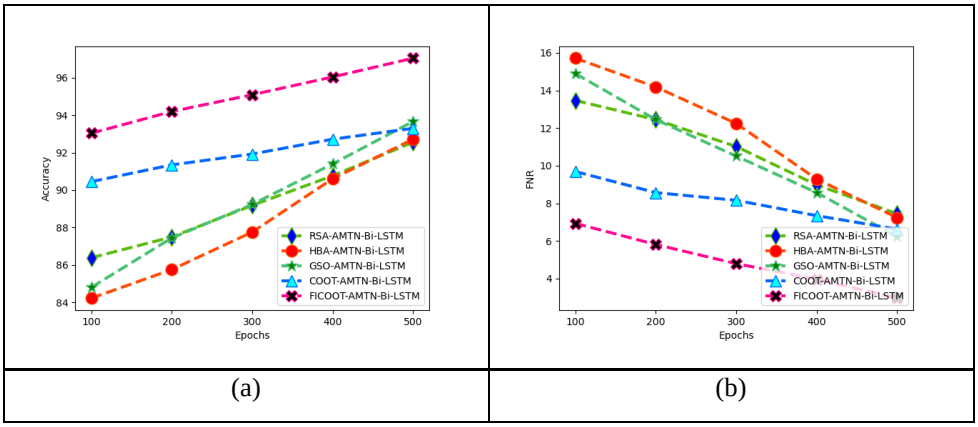


Fig 8. ROC analysis of the suggested FICOOT-based NER with adolescent suicidal text system regarding " (a) Dataset 1, and (b) Dataset 2".

6.6 Comparative analysis of the proposed NER system.

For dataset 1, Fig. 9 shows a comparison of the recommended method, and Fig. 10 shows a comparison of the recommended classifier. The accuracy analysis in comparison to the classification model is shown in Fig. 9(a). Following epochs of 400, the accuracy of the assessment obtains an accuracy level greater than 1.2% of RSA-AMTN-Bi-LSTM, 3% of HBA-AMTN-Bi-LSTM, 2% of GSO-AMTN-Bi-LSTM, and 3% of COOT-AMTN-Bi-LSTM. As a consequence, the likelihood of being able to identify the identified entity is higher in those with more precise outcomes.



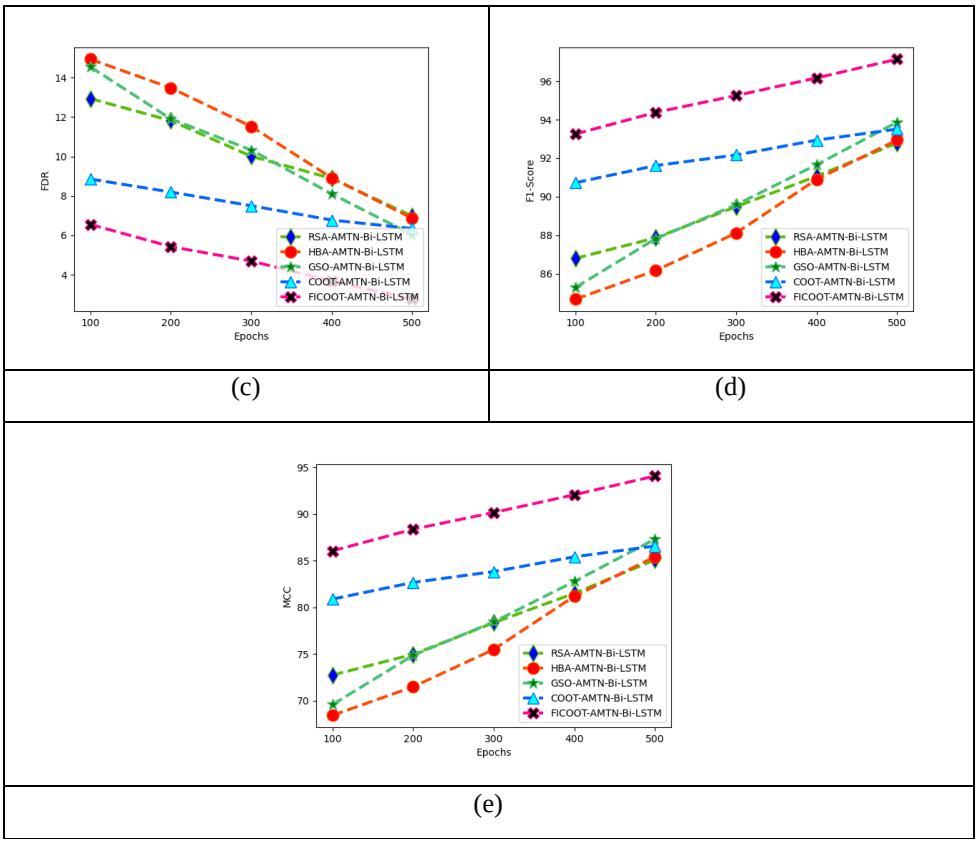
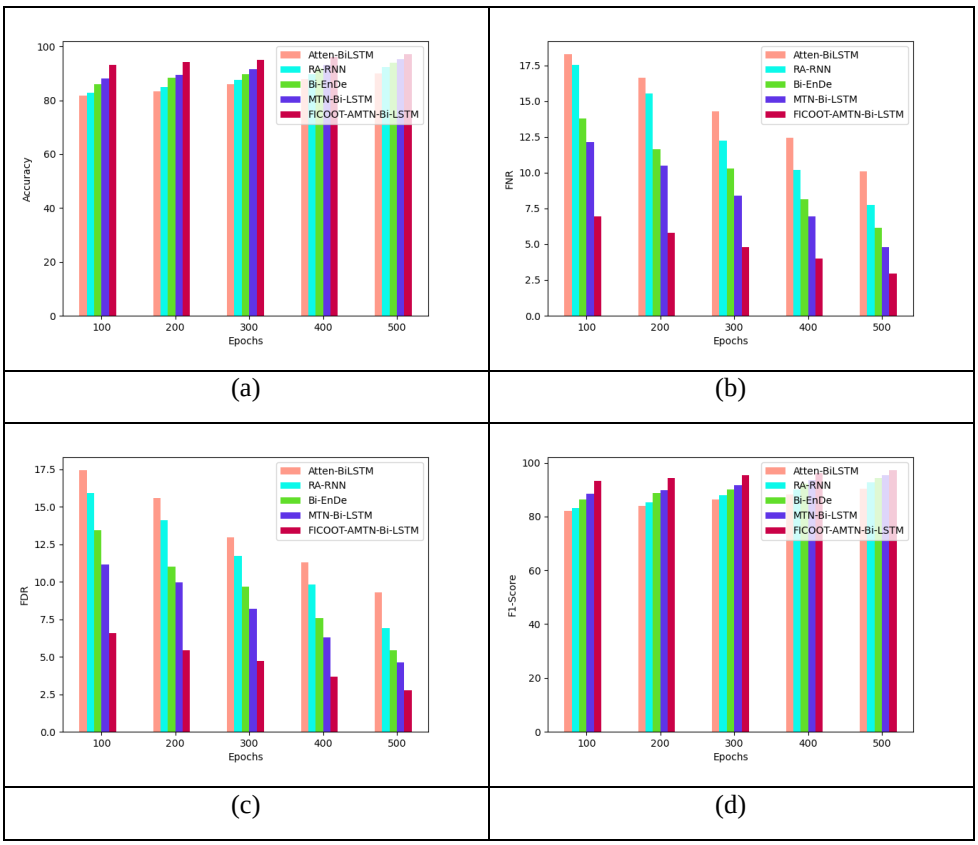
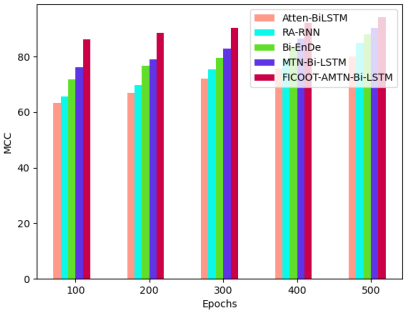


Fig 9. Comparative analysis of the suggested NER with adolescent suicidal text system compared with a classical algorithm in dataset 1 regarding " (a) accuracy (b) FNR, (c)FDR, (d)F1-score and (e)MCC".



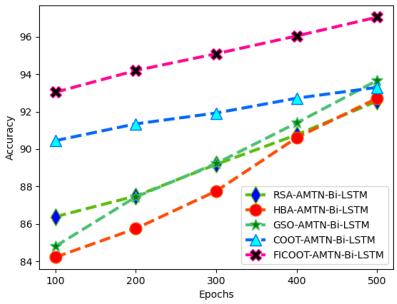


(e)

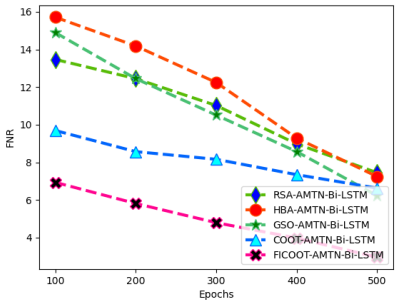
Fig 10. Comparative analysis of the suggested NER with adolescent suicidal text system compared with classical classifier in dataset 1 regarding " (a) accuracy (b) FNR, (c)FDR, (d)F1-score and (e)MCC".

6.7 Comparative analysis of the proposed NER system

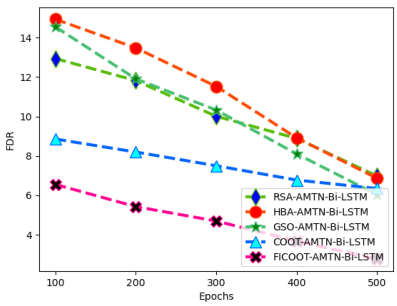
For dataset 2, Fig. 11 shows a comparison of the recommended method, and Fig. 12 shows a comparison of the recommended classifier. The accuracy analysis in comparison to the classification model is shown in Fig. 11(a). With an epoch's value of 300, the accuracy assessment obtains an accuracy level greater than 2.4% of RSA-AMTN-Bi-LSTM, 3.5% of HBA-AMTN-Bi-LSTM, 2.9% of GSO-AMTN-Bi-LSTM, and 3.0% of COOT-AMTN-Bi-LSTM. Consequently, the more precise outcomes have a higher chance of correctly identifying the named entity.



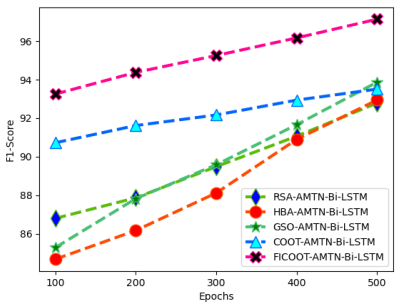
(a)



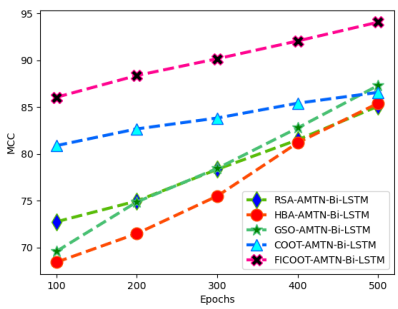
(b)



(c)



(d)



(e)

Fig 11. Comparative analysis of the suggested NER with adolescent suicidal text system compared with a classical algorithm in dataset 2 regarding " (a) accuracy (b) FNR, (c)FDR, (d)F1-score and (e)MCC".

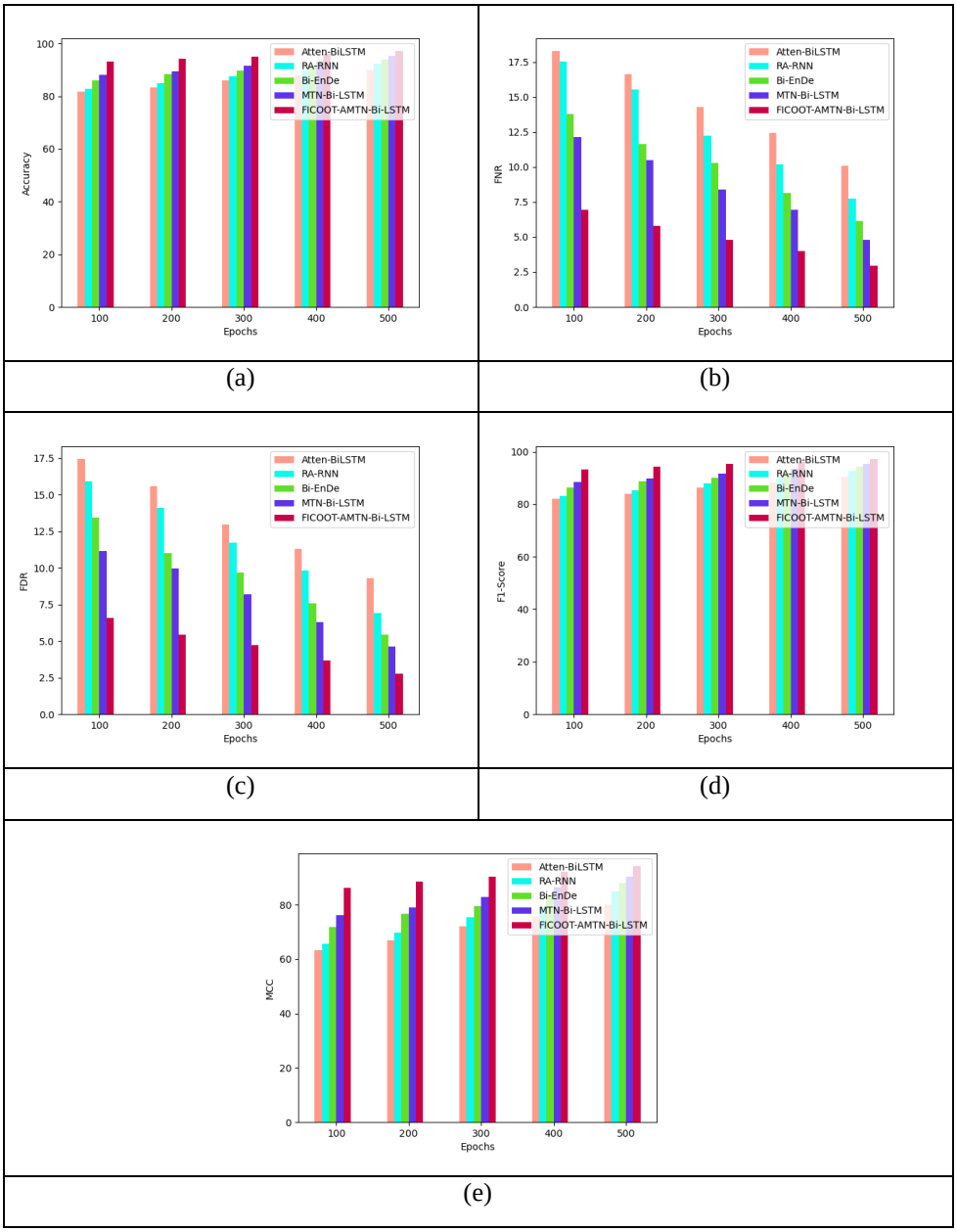


Fig 12. Comparative analysis of the suggested NER with adolescent suicidal text system compared with classical classifier in dataset 2 regarding " (a) accuracy (b) FNR, (c)FDR, (d)F1-score and (e)MCC".

6.8 Comparative classifier analysis of the proposed NER system

The comparative evaluation of the recommended classification for the first dataset is shown in Fig. 13, and the comparative evaluation of the recommended classifier for the second dataset is shown in Fig. 14. Fig. 13 (b) displays the precision and FDR analysis compared with the algorithm model. From the graph, the precision value achieves a higher value than 7.5% of RSA-AMTN-Bi-LSTM, 6.3 % of HBA-AMTN-Bi-LSTM, 4.9 % of GSO-AMTN-Bi-LSTM and 3.4 % of COOT-AMTN-Bi-LSTM; also the FDR value is lesser than proposed FICOOT-AMTN-Bi-LSTM. Hence, the precision and lesser FDR results contain a better potential to recognize the named entity.

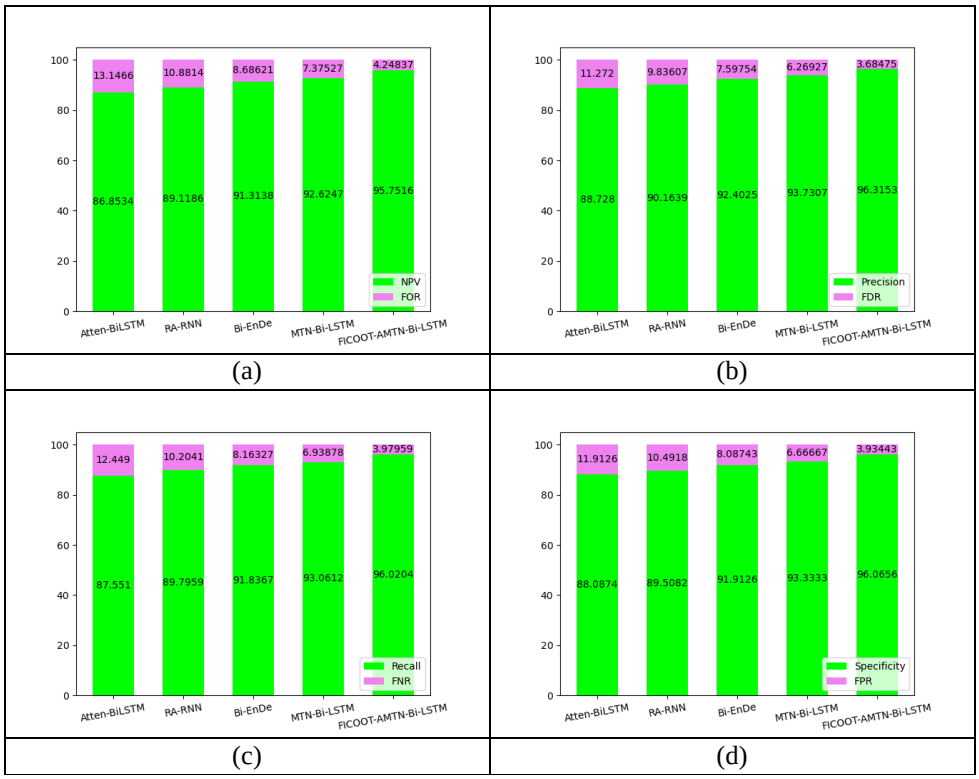


Fig 13. Comparative analysis of the suggested NER with adolescent suicidal text system compared with classical classifier in dataset 1 regarding " (a) NPV vs FOR (b) Presicion vs FDR, (c)Recall vs FNR, and (d)specificity vs FPR".

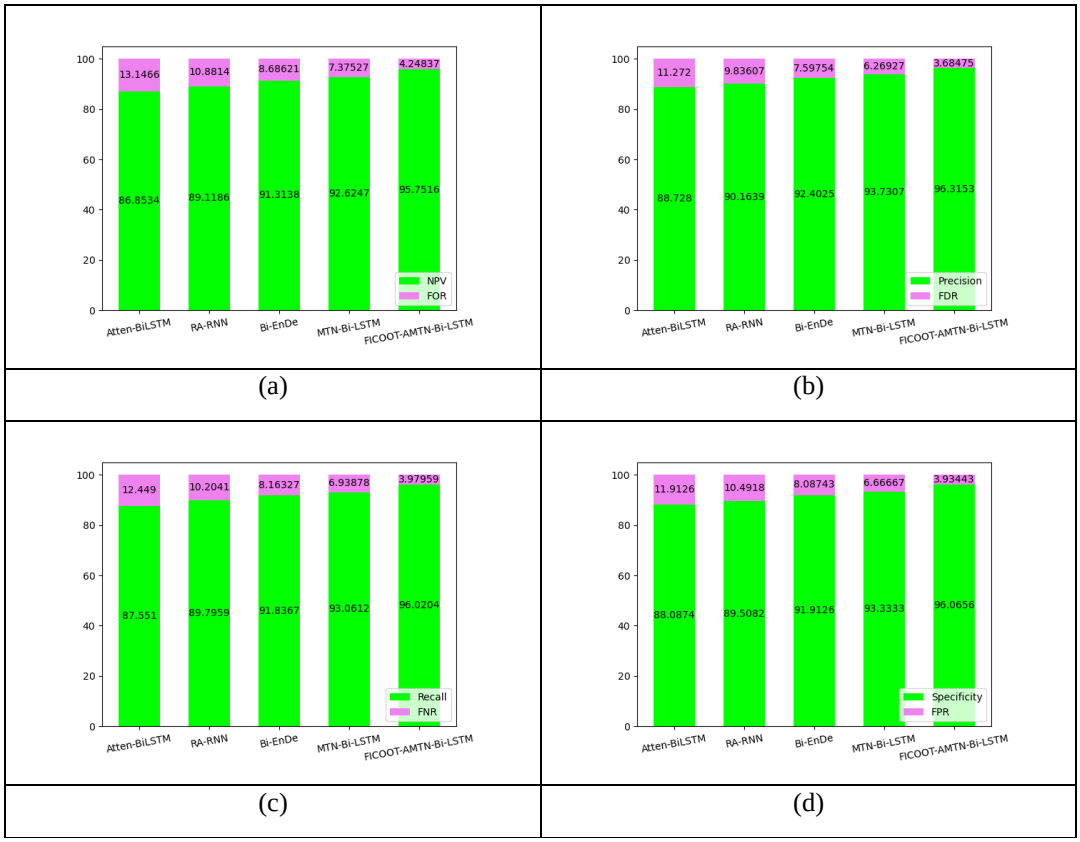


Figure 14. Comparative analysis of the suggested NER with adolescent suicidal text system compared with classical classifier in dataset 2 regarding " (a) NPV vs. FOR, (b) Precision vs. FDR, (c)Recall vs. FNR, and (d)specificity vs. FPR".

6.9 An overall analysis of proposed NER with the adolescent suicidal text system

Tables 3 and 4 give a comprehensive assessment of the structure compared to various methods and classifiers. Concerning data 2, the recommended model's accuracy is superior to that of the RSA-AMTN-Bi-LSTM, the HBA-AMTN-Bi-LSTM, the GSO-AMTN-Bi-LSTM, and the COOT-AMTN-Bi-LSTM by 6.5%, 4.5%, 2.9%, and 4.9%, respectively. These findings demonstrate that the system is

more capable of identifying named entities.

Table 3. The overall analysis of enhanced NER with adolescent suicidal text system against conventional algorithms for three datasets

TERMS	RSA-AMTN-Bi-LSTM [31]	HBA-AMTN-Bi-LSTM [32]	GSO -AMTN-Bi-LSTM [33]	COOT-AMTN-Bi-LSTM [30]	FICOOT-AMTN-Bi-LSTM
Dataset 1					
"Accuracy"	90.76517	90.60686	91.39842	92.71768	96.04222
"Recall"	91.02041	90.71429	91.42857	92.65306	96.02041
"Specificity"	90.4918	90.4918	91.36612	92.78689	96.06557
"Precision"	91.11338	91.08607	91.89744	93.22382	96.31525
"FPR"	9.508197	9.508197	8.63388	7.213115	3.934426
"FNR"	8.979592	9.285714	8.571429	7.346939	3.979592
"NPV"	9.606987	9.902067	9.130435	7.81759	4.248366
"FDR"	90.39301	90.09793	90.86957	92.18241	95.75163
"F1-Score"	8.886619	8.913934	8.102564	6.776181	3.684749
"MCC"	91.06687	90.8998	91.6624	92.93756	96.1676
Dataset 2					
"Accuracy"	90.84215	90.32464	91.55959	92.69026	96.08573
"Recall"	90.83568	90.31602	91.57252	92.69112	96.07625
"Specificity"	90.84861	90.33326	91.54666	92.6894	96.09521
"Precision"	90.84743	90.33159	91.54885	92.68953	96.09447
"FPR"	9.151391	9.666744	8.453338	7.310599	3.904789
"FNR"	9.164318	9.68398	8.427484	7.308876	3.923748
"NPV"	9.163134	9.682311	8.429664	7.309002	3.923005
"FDR"	90.83687	90.31769	91.57034	92.691	96.077
"F1-Score"	9.152575	9.668411	8.451153	7.310473	3.905529
"MCC"	90.84155	90.3238	91.56068	92.69033	96.08536

Table 4. An overall analysis of NER with adolescent suicidal text system against conventional classifiers for three datasets

Metrics	Atten-BiLSTM [18]	RA-RNN [20]	Bi-EnDe [22]	MTN-Bi-LSTM	FICOOT-AMTN-Bi-LSTM
Dataset 1					
"Accuracy"	87.81003	89.65699	91.87335	93.19261	96.04222
"Recall"	87.55102	89.79592	91.83673	93.06122	96.02041
"Specificity"	88.08743	89.5082	91.91257	93.33333	96.06557
"Precision"	88.72802	90.16393	92.40246	93.73073	96.31525
"FPR"	11.91257	10.4918	8.087432	6.666667	3.934426
"FNR"	12.44898	10.20408	8.163265	6.938776	3.979592

“NPV”	13.14655	10.88139	8.686211	7.375271	4.248366
“FDR”	86.85345	89.11861	91.31379	92.62473	95.75163
“F1-Score”	11.27198	9.836066	7.597536	6.26927	3.684749
“MCC”	88.13559	89.97955	92.11873	93.39478	96.1676
Dataset 2					
“Accuracy”	87.956	89.92433	91.93016	93.23492	96.08573
“Recall”	87.93574	89.91011	91.9181	93.2263	96.07625
“Specificity”	87.97625	89.93855	91.94223	93.24353	96.q09521
“Precision”	87.97138	89.93569	91.94028	93.24237	96.09447
“FPR”	12.02375	10.06145	8.057775	6.756466	3.904789
“FNR”	12.06426	10.08989	8.081905	6.773701	3.923748
“NPV”	12.05937	10.08702	8.079955	6.772534	3.923005
“FDR”	87.94063	89.91298	91.92004	93.22747	96.077
“F1-Score”	12.02862	10.06431	8.05972	6.75763	3.905529
“MCC”	87.95356	89.9229	91.92919	93.23433	96.08536

6.10 Statistical comparison of the developed model over distinct algorithms and classifiers

The suggested system's statistical assessment is provided in Table 5, along with several conventional techniques. When taking into account the mean measurement, the proposed model is created by 64.9% of RSA-AMTN-Bi-LSTM, 54.5% of HBA-AMTN-Bi-LSTM, 42.9% of GSO-AMTN-Bi-LSTM, and 65% of COOT-AMTN-Bi-LSTM. Thus, it attests to the improved algorithm's effectiveness.

Table 5. Statistical analysis of the developed NER with adolescent suicidal text system over distinct algorithms for two datasets

TERMS	**RSA-AMTN-Bi-LSTM [31]**	**HBA - AMTN-Bi-LSTM [32]**	**GSO - AMTN-Bi-LSTM [33]**	**COOT-AMTN-Bi-LSTM [30]**	**FICOOT-AMTN-Bi-LSTM**
Dataset 1					
“Best”	1.000676	1.002564	1.007549	1.001193	1.00032
“Worst”	1.011435	1.110537	1.016827	1.019094	1.057336
“Mean”	1.001107	1.012621	1.00792	1.004551	1.007892
“Median”	1.000676	1.002564	1.007549	1.001193	1.00032
“Standard Deviation”	0.002108	0.029269	0.001818	0.006789	0.012034
Dataset 2					
“Best”	1.002277	1.008616	1.016169	1.003473	1.001199
“Worst”	1.06559	1.015593	1.080053	1.009315	1.01393
“Mean”	1.016473	1.009741	1.034116	1.007497	1.003888
“Median”	1.015573	1.008616	1.024221	1.009315	1.001199
“Standard Deviation”	0.015699	0.002368	0.021988	0.002661	0.003996

7. Conclusion

This research paper has implemented NER with adolescent suicidal text using AMTN-Bi-LSTM. Initially, the required data has been gathered from

standard datasets for analysis. The pre-processing uses the obtained data as input, which helps clean the text and assure its quality and consistency before further analysis. The AMTN-Bi-LSTM model was used to detect and extract key suicidal text words from the pre-processed text data. This stage entailed NMT employing the Transformer Network, which generates attention vectors as an output. Input the attention vectors obtained from the Transformer Network into the encoder section of the Bi-LSTM model. The encoder section process and encode the vectors, which was then passed to the decoder section of the Bi-LSTM. The decoder section aims to recognize essential suicidal text words based on the encoded vector and uses the FICOOT algorithm to optimize the parameters of the generated AMTN-Bi-LSTM model. The Trans-Bi-LSTM model's encoder unit receives the recognized text as input, and based on previously acquired patterns and traits, the model divides the text into suicidal and non-suicidal categories. This model was comparing it with conventional models using various performance metrics. This comparison aims to assess the superiority and effectiveness of the proposed model in terms of its performance on suicidal word recognition. Regarding dataset 2, the suggested model's precision was higher than 4.9 % of RSA-AMTN-Bi-LSTM, 4.5 % of HBA -AMTN-Bi-LSTM, 2.9 % of GSO-AMTN-Bi-LSTM, and 6.5 % of COOT-AMTN-Bi-LSTM. These results validated the system's has better capacity to recognize named entity system

Acknowledgements

I would like to express my very great appreciation to the co-author of this manuscript for their valuable and constructive suggestions during the planning and development of this research work.

Author Contribution

All authors have made substantial contributions to conception and design, revising the manuscript, and the final approval of the version to be published. Also, all authors agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Conflict of Interest

The authors declare no conflict of interest

References

[1] Q. Qiu, Z. Xie, L. Wu, L.Tao and W. Li , "BiLSTM-CRF for geological named entity recognition from the geoscience literature," *Earth Science Informatics*, vol. 12, pp. 565–579, 2019.

[2] S. K.Gorla, S. S. Tangeda, L. B. M. Neti and A. Malapati, "Telugu named entity recognition using bert," *International Journal of Data Science and Analytics*, vol. 14, pp. 127–140, 2022.

[3] M. Affi, and C. Latiri, "BE-BLC: BERT-ELMO-Based Deep Neural Network Architecture for English Named Entity Recognition Task," *Procedia Computer Science*, vol. 192, pp. 168-181, 2021.

[4] C. Wu, G. Luo, C. Guo, Y. Ren, A. Zheng, and C. Yang, "An attention-based multi-task model for named entity recognition and intent analysis of Chinese online medical questions," *Journal of Biomedical Informatics*, vol. 108, pp. 103511, 2020.

[5] D. Peng, D. Zhang, C. Liu, and J. Lu, "BG-SAC: Entity relationship classification model based on Self-Attention supported Capsule Networks," *Applied Soft Computing*, vol. 91, pp. 106186, 2020.

[6] Z. Li, J. Yang, X. Gou, and X. Qi, "Recurrent neural networks with segment attention and entity description for relation extraction from clinical texts," *Artificial Intelligence in Medicine*, vol. 97, pp. 9-18, 2019.

[7] X. Meng and J. Zhang, "Anxiety Recognition of College Students Using a Takagi-Sugeno-Kang Fuzzy System Modeling Method and Deep Features," *IEEE Access*, vol. 8, pp. 159897-159905, 2020.

[8] A. Kumar J, T. E. Trueman, and A. K. Abinesh, "Suicidal risk identification in social media," *Procedia Computer Science*, vol. 189, pp. 368-373, 2021.

[9] T. Zhang, A. M. Schoene, and S. Ananiadou, "Automatic identification of suicide notes with a transformer-based deep learning model," *Internet Interventions*, vol. 25, pp. 100422, 2021.

[10] G. Berkelmans, L. Schweren, S. Bhulai, R. V. D. Mei, and R. Gilissen, "Identifying populations at ultra-high risk of suicide using a novel machine learning method," *Comprehensive Psychiatry*, vol. 123, pp. 152380, 2023.

[11] S.Ghosal, and A. Jain, "Depression and Suicide Risk Detection on Social Media using fastText Embedding and XGBoost Classifier," *Procedia Computer Science*, vol. 218, pp. 1631-1639, 2023.

[12] N. J.C. Stapelberg, M. Randall, J. Sveticic, P. Fugelli, H. Dave, and K. Turner, "Data mining of hospital suicidal and self-harm presentation records using a tailored evolutionary algorithm," *Machine Learning with Applications*, vol. 3, pp. 100012, 15 2021.

[13] D. Lekkas, R. J. Klein, and N. C. Jacobson, "Predicting acute suicidal ideation on Instagram using ensemble machine learning models," *Internet Interventions*, vol. 25, 100424, 2021.

[14] J. Du, Y. Zhang, J. Luo, Y. Jia, Q. Wei, C. Tao and H. Xu, "Extracting psychiatric stressors for suicide

from social media using deep learning," *BMC Medical Informatics and Decision Making*, vol. 18, no. 43, 2018.

[15] S. Ghosh, A. Ekbal and P. Bhattacharyya, "Deep cascaded multitask framework for detection of temporal orientation, sentiment and emotion from suicide notes," *Scientific Reports*, vol. 12, no. 4457, 2022.

[16] P. Burnap, G. Colombo, R. Amery, A. Hodorog, and J. Scourfield, "Multi-class machine classification of suicide-related communication on Twitter," *Online Social Networks and Media*, vol. 2, pp. 32-44, 2017.

[17] A. Chadha, and B. Kaushik, "A Hybrid Deep Learning Model Using Grid Search and Cross-Validation for Effective Classification and Prediction of Suicidal Ideation from Social Network Data," *New Generation Computing*, vol. 40, pp. 889–914, 2022.

[18] T. Ghosh, Md. H. A. Banna, Md. J. A. Nahian, M. N. Uddin, M. S. Kaiser, and M. Mahmud, "An attention-based hybrid architecture with explainability for depressive social media text detection in Bangla," *Expert Systems with Applications*, vol. 213, Part C, pp. 119007, 1 2023.

[19] S. R. Laskar, B. Paul, P. Pakray, and S. Bandyopadhyay, "English-Assamese Multimodal Neural Machine Translation using Transliteration-based Phrase Augmentation Approach," *Procedia Computer Science*, vol. 218, pp. 979-988, 2023.

[20] Y. Zhao, M. Komachi, T.Kajiwara, and C. Chu, "Region-attentive multimodal neural machine translation, *Neurocomputing*, vol. 476, pp. 1-13, 1 2022.

[21] N. B. Allen, B. W. Nelson, D. Brent, and R. P. Auerbach, "Short-term prediction of suicidal thoughts and behaviors in adolescents: Can recent developments in technology and computational science provide a breakthrough?," *Journal of Affective Disorders*, vol. 250, pp. 163-169, 2019.

[22] M. N. A. Ali and G. Tan, "Bidirectional Encoder–Decoder Model for Arabic Named Entity Recognition," *Arabian Journal for Science and Engineering*, vol. 44, pp. 9693–9701, 2019.

[23] B. Priyamvada, S. Singhal, A. Nayyar, R. Jain, P. Goel, M. Rani and M. Srivastava, "Stacked CNN - LSTM approach for prediction of suicidal ideation on social media," *Multimedia Tools and Applications*, 2023.

[24] A. Belouali, S. Gupta, V. Sourirajan, J. Yu, N. Allen, A. Alaoui, M. A. Dutton and Matthew J. Reinhard, "Acoustic and language analysis of speech for suicidal ideation among US veterans," *BioData Mining*, vol. 14, no. 11, 2021.

[25] D. Kodati and R. Tene, "Identifying suicidal emotions on social media through transformer-based deep learning," *Applied Intelligence*, 2022.

[26] A. S. Farag, M. Mohandes and A. A. Shaikh, "Diagnosing failed distribution transformers using neural networks," *in IEEE Transactions on Power Delivery*, vol. 16, no. 4, pp. 631-636, 2001.

[27] R. Zhong, R. Wang, Y. Zou, Z. Hong and M. Hu, "Graph Attention Networks Adjusted Bi-LSTM for Video Summarization," *IEEE Signal Processing Letters*, vol. 28, pp. 663-667, 2021.

[28] C. Li, J. Zheng, H. Pan, J. Tong and Y. Zhang, "Refined Composite Multivariate Multiscale Dispersion Entropy and Its Application to Fault Diagnosis of Rolling Bearing," *IEEE Access*, vol. 7, pp. 47663-47673, 2019.

[29] N.Vassilina & S.Agnes & D. Marc, "Hybrid Adaptation of Named Entity Recognition for Statistical Machine Translation", Conference: Proceedings of the Second Workshop on Applying Machine Learning Techniques to Optimise the Division of Labour in Hybrid MT, pp. 1-16. 2012.

[30] I. Naruei, and F. Keynia, "A new optimization method based on COOT bird natural life model," *Expert Systems with Applications*, vol. 183, pp. 115352, 30.

[31] Z. Elgamal, A. Q. M. Sabri, M. Tubishat, D. Tbaishat, S. N. Makhadmeh and O. A. Alomari, "Improved Reptile Search Optimization Algorithm Using Chaotic Map and Simulated Annealing for Feature Selection in Medical Field," *IEEE Access*, vol. 10, pp. 51428-51446, 2022.

[32] M. Obayya, J. M. Alsamri, M. A. Al-Hagery, A. Mohammed and M. A. Hamza, "Automated Cardiovascular Disease Diagnosis Using Honey Badger Optimization With Modified Deep Learning Model," *IEEE Access*, vol. 11, pp. 64272-64281, 2023, doi: 10.1109/ACCESS.2023.3286661.

[33] B. M. Nguyen, T. Tran, T. Nguyen and G. Nguyen, "Hybridization of Galactic Swarm and Evolution Whale Optimization for Global Search Problem," *IEEE Access*, vol. 8, pp. 74991-75010, 2020.