CNO-LSTM: A Chaotic Neural Oscillatory Long Short-Term Memory Model for Text Classification

NUOBEI SHI, ZHUOHUI CHEN, LING CHEN, AND RAYMOND S. T. LEE, (Member, IEEE)

Faculty of Science and Technology, Beijing Normal University–Hong Kong Baptist University United International College, Zhuhai 519000, China
Faculty of Innovation Engineering, Macau University of Science and Technology, Taipa, Macau, China
School of Applied Science and Civil Engineering, Beijing Institute of Technology Zhuhai, Zhuhai 519000, China

Corresponding author: Raymond S. T. Lee (raymondshtlee@uic.edu.cn)

This work was supported in part by the Beijing Normal University–Hong Kong Baptist University United International College (UIC) under Grant R202008, in part by the Key Laboratory for Artificial Intelligence and Multi-Model Data Processing of Department of Education of Guangdong Province, and in part by the Science and Technology Commissioner Project of Research on Offline Handwritten Signature Verification from the Department of Science and Technology of Guangdong Province under Grant GDKTP2021031000.

ABSTRACT Long Short-Term Memory (LSTM) networks are unique to exercise data in its memory cell with long-term memory as Natural Language Processing (NLP) tasks have inklings of intensive time and computational power due to their complex structures like magnitude language model Transformer required to pre-train and learn billions of data performing different NLP tasks. In this paper, a dynamic chaotic model is proposed for the objective of transforming neurons states in network with neural dynamic characteristics by restructuring LSTM as Chaotic Neural Oscillatory-Long-Short Term Memory (CNO-LSTM), where neurons in LSTM memory cells are weighed in substitutes by oscillatory neurons to speed up computational training of language model and improve text classification accuracy for real-world applications. From the implementation perspective, five popular datasets of general text classification including binary, multi-classification and multi-label classification are used to compare with mainstream baseline models on NLP tasks. Results showed that the performance of CNO-LSTM, a simplified model structure and oscillatory neurons state in exercising different types of text classification tasks are above baseline models in terms of evaluation index such as Accuracy, Precision, Recall and F1. The main contributions are time reduction and improved accuracy. It achieved approximately 46.76% of the highest reduction training time and 2.55% accuracy compared with vanilla LSTM model. Further, it achieved approximately 35.86% in time reduction compared with attention model without oscillatory indicating that the model restructure has reduced GPU dependency to improve training accuracy.

INDEX TERMS Long short-term memory, chaotic neural network, Lee-oscillator, text classification, natural language processing.

I. INTRODUCTION

Text Classification, also known as text categorization, is an integral branch of Natural Language Processing. The purpose of Text Classification is tagging raw data text with pre-set labels from websites, online chat messages, emails, question-answer conversations, articles in different characters, words, and sentences against humans thinking [1]. Generally, it includes Sentiment Analysis (SA), News Classification (CA), Topic Classification (TC) and Question Answering (QA) to reasoning and Natural Language Inference (NLI) with binary classification, multi-classification and multi-label classification [2] For example, 1) sentiment analysis has simple happy and sad emotion categories [3], [4], [5], [6], [7], 2) email spam filter is considered as binary classification problem to distinguish email contents based on texts understanding for further classification, 3) publication database with multi-label such as WOS-11967 illustrated by index terms is classification of different topics in research areas.
There are increasingly amounts of raw data analysis from online platforms with users’ expressions or online queries since internet and mobile apps development. To verify the proposed CNO-LSTM model, five popular and large text classification datasets include 1) Question-Answer binary classification of BoolQ, 2) sentiment analysis of IMDB, 3) Multi-label datasets of WOS-11967, 4) AG News for new categorization and 5) Topic Classification of DBpedia are selected to represent different branches of text classification. These datasets characteristics consist of pure texts, human natural language in general format which require data preprocessing for model training.

Text Classification traditional automatic methods consist of two parts 1) rule-based hand-craft labeling for pattern matching and 2) machine learning (data-driven)-based models with pre-define labels like Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Naïve Bayes (NB) methods. These methods inhibited high-accuracy performance compared with state-of-art deep learning methods because 1) dataset formats are limited to structured data means algorithms like KNN, SVM cannot deal with pure text and relies highly on data transformation which cannot follow the understanding of human thinking to handle industrial data instantly, and 2) dataset magnitude is larger and required to pre-train by state-of-art technologies like Transformer for accuracy improvement in data mining and machine learning problems, hence, have inhibited efficiency and performance to accomplish current NLP tasks. On the other hand, deep learning method fed with different datasets for text classification or generation become mainstream for NLP tasks nowadays [8], [9] because it can integrate various tasks in the same model, for example pre-trained language model, Transformer of BERT [10], [11] and GPT, as compared with traditional model which required to identify functions or tasks in forecasting or classification.

A language model allows to pre-train sufficient knowledge and enable neural network can match with NLP tasks. The first language model with neural network is proposed by Bengio [12] in 2001 based on Feed-Forward Neural Network to train 14 million words with embedding technology for word-level meaning representation [13] to map text into low-dimension vector in continual spatial space. The advent of word embedding technology development [14] and word2vec [15] proposed by Google in 2013 with 6 billion words are widely adopted for data preprocessing. By using state-of-art NLP preprocessing technology for machine understanding, data preprocessing method for the proposed CNO-LSTM is to transfer text from word to vector by NLTK, with Stanford NLP tool and tokenizer for sequential training. The pre-trained sequential language model with NLP data fed has high accuracy on text classification and primary stage text generation as long-text sequence-to-sequence model. Meaning representation, in deep learning with different formats of text meaning extraction from natural language for model training consists of three levels: character, word, and phrase or sentence to match the understanding of different language models for model training.

Additionally, baseline model like Recurrent Neural Network (RNN), word embedding, attention mechanism combine with neural network can handle lengthy contexts of magnitude data. RNN inherent recurrent feature for sequential data in memory help to build up the relationship with internal inputs data, output and its varieties are widely used in NLP against fully connected Artificial Neural Network (ANN) or Convolutional Neural Network (CNN). However, RNN, as a sequence-to-sequence model connects words within sequence is unable to handle text sequential processing, which means it only maintains relative message but foregoes previous information. It also encounters exploding and gradient vanishing problems during language model training. Therefore, LSTM, a revised RNN is proposed to solve this problem focusing on long-term memory control and prediction [15] The dynamic neural state of memory cell in LSTM of the proposed CNO-LSTM are altered by oscillatory neurons. From gradient vanishing aspect, oscillatory connection state depends on pulse signal in the network with neurons relationships. From gradient exploding aspect, a one-layer LSTM weight refreshment is coherent to avoid the exploding than other deeper networks. Performance results in accuracy and efficiency will be presented in experiment section.

Embedding models using Transformer or so-called Transfer Learning technology with parallelization computation are received by many industrial NLP tasks aside network matching in recent years. The main difference between these two models is the pre-trained segment. Transformer improves the efficiency in application because of this segment, which uses self-attention mechanism to train the model with magnitude data as knowledge base before applying into model predication or generation. However, the training procedure relies highly on GPU or TPU regardless pre-training or fine-tuning, leading the model lacks a suitable way to improve from neural network structure than relies on external factors like GPU for inhibiting acceleration. To illustrate the contribution of model training acceleration, the proposed CNO-LSTM is compared with attention mechanism to improve computation efficiency and model accuracy with industrial data under the same condition with and without revising structure from oscillator.

The increasing rate depends on deeper layers and more neurons for accuracy improvement. However, complex computation involves time and computational redundancy. Thus, hybrid models [16], [17] or variant model such as TextCNN is the combination of CNN [18] and RNN [19], or revise attention mechanism are considered by researchers to improve performance. The proposed CNO-LSTM is a cross-discipline model to research neuron state with artificial intelligence (AI) to revise LSTM model as dynamic neural networks.

LSTM structure has input, output layers, recurrent segment, forget gate and memory cell for propagation of the latest state. It uses cells states [20] to store and forgo information to retain useful information in memory storage against traditional RNN. In other words, LSTM structure is an imitation...
of human brain to remember, update, or select significant information mechanism to regulate memory.

From biological science and neuroscience perspective, Chaotic Neural Oscillator (CNO) are neurons in the proposed CNO-LSTM structure to substitute memory cell in network for periodic information transferring within recurrent network, it is the combination of biology and physics to build transient nonlinear neural dynamics [21] and oscillations with nonlinear and highly-dimension properties. The significance of oscillator is the transient for time efficiency and progressive memory of network joint together for task processing, so called progressive memory recalling [22].

Neural network training for large language model relies on GPU acceleration for industrial usage nowadays. Unlike research use many layers to pursue higher accuracy, some reduce layers or delete structures’ unnecessary segments to speed up training, for example, Gated Recurrent Units (GRU) is a simplified version of RNN.

Further, the performance of a preferable language model is to balance deeper neural network with higher accuracy and computation efficiency. Hence, the proposed CNO-LSTM considered these factors using Lee-Oscillator structure, one of the author previous work, with vanilla LSTM structure to replace neuron state when absorbing new knowledge in training as a dynamic network to update memory and process information with pattern association rapidly.

In this paper, the main contributions include: 1) rework vanilla LSTM as dynamic oscillatory network, CNO-LSTM, to improve accuracy and reduce training time, 2) generalize model for text classification and test 5 popular datasets readily for real-world applications, 3) solve gradient vanishing and exploding problems with LSTM using neurons state and signals exchange in periodic, discrete features 4) the proposed CNO-LSTM is a cross-discipline research in computer science, neuron-science, biology and physics.

This paper is organized as follows and match with Figure 1 several layers: Section II presents detail review of different models in text classification, vanilla LSTM structure and formula for data training. Section III presents an architecture of the proposed CNO-LSTM with structure revision and parameter settings from LSTM and Lee-Oscillator separately applied in the model as transient dynamic state transferring. Section IV presents CNO-LSTM implementation with substantial representative dataset for model comparisons and testing in different text classification. Section V is discussion on model extension and conclusion.

II. LITERATURE REVIEW

A. TEXT CLASSIFICATION

NLP is unlimited in handcraft or labelling in linguistic, computational linguistic or statistic methods after decades. It is a cross-discipline or branches of techniques to text classification. For instance, rule-based matching from pattern recognition, machine learning also known as statistic learning method with supervised learning and unsupervised learning schemes, deep learning [23] modeling in AI and its hybrid alternatives [24], [25], [26]. The former two categories depend on hand-craft labeling as the training target attributed to traditional methods, despite automatic systems are built, they are unable to exercise industrial projects with billions of data by self-detection or self-judgement.

Prior language model was widely implemented, simple text classification or categorization depend on machine learning methods such as NB, KNN, SVM, Logistic Regression(LR), and Random Forest(RF). NB is amongst conditional probability improvement for classification with structured data, the classification feature not specialized in NLP like language model but covering most fields require classification if data can be transformed to match system requirement. LR is a general classification algorithm, where its features are correlated but its traditional algorithms are prolonged for training with overfitting problems. KNN has high dependency for a chosen K property, which measures the target distance with neighbors and use k for decision. It is not limited in data size but is subjected by the k circle. SVM has high dimension hyper-planes used for spatial space classification with high speed and accuracy, however, it was designed for binary classification initially as such, multi classification and multi-label classification are required to modify. RF is like a decision tree structure but consist of many decision tree branches for higher accuracy. However, data format restriction lead to algorithm remain on statistic learning method and rely on humans to label text understanding. Deep learning method is an imitation of human brain structure with reasoning algorithms.

Traditional text classification methods contain input data format requires adjustment to match with data algorithm requirement. In NLP, Natural Language Understanding (NLU) reformat data with different types effect the understanding of text classification accuracy, especially in general classification to process pure text resemble to humans understanding, word embedding and tokenization, adds contexts interaction for meaning representation. Table 1 shows the input data requirement for each algorithm as abovementioned.

There are many LSTM variants for industrial application [27], [28], [29]. Tai [30] revised chain-LSTM model structure

![FIGURE 1. Structure of paper for the comparison and application of the proposed CNO-LSTM model.](image-url)
with Tree-LSTM model because the properties of natural language syntactic are combined with word and phrase levels, which is the same in BERT theory for meaning representations to improve model efficiency. The tree structure proves that the model is validated on sentiment classification and correlations prediction between two utterances. The revision from Zhu [31] is identical to show a tree-based structure can store more information in child cells branch connections for root in a recursive process, which aims to long-distance interaction more than hierarchies. Multi-Timescale LSTM (MT-LSTM) neural networks [32] divided hidden states of vanilla LSTM into groups to capture long texts information and update at different timescales. It achieved satisfactory performance than standard LSTM and RNN. Some LSTM application research focus on data preprocessing to modify input dataset format. Johnson [33] used text region embedding without LSTM word-embedding with LSTM, Wan [34] used bi-directional LSTM to semantic matching with multiple positional sentence representations. Text classification applications is extended to specific topics such as medical history [35], [36], [37], news categorization [27] and sentiment analysis [18] in recent years. These comparison are shown in Table 2.

It is noted that variants models skilled for one or two task types attributed to text classification but gates construction remain unchanged, which means the calculation requires modification. Lee-Oscillator is applied as a neural structure network substitution especially for memory association and recalling long-term memory within a short time in high-frequency oscillation [38].

### B. LONG SHORT-TERM MEMORY

Based on RNN, Hochreiter, Sepp and Schmidhuber [39], Jürgen proposed Long Short-Term Memory in 1997 to handle long-term data with sequence-to-sequence model. The recurrent of LSTM [40] added memory cell to process retain and forgo information during long-time intervals in order to solve RNN long-term dependency problem. Also, gradient exploding or vanishing problems on information capture and update from memory cells are solved simultaneously.

The structure of vanilla LSTM consists of six formulations, which represent structure gates, activation functions and cells states. A standard LSTM structure has input gate, output gate, forget gate as shown in Figure 2 to exercise the flow of information processing, input modulation gate is denoted as $g$ to refresh memory cells. The first step for LSTM is to decide what should be dissipated during updating and controlled by forget gate, the outcome number is between 0 to 1 to be stored or released. Next, the new information is received by input gate to select useful information for memory cells on prediction. The input gate has 1) sigmoid function to select the expected value and 2) tanh activation function to update cell state. It will be activated when previous state multiples forget gate to release meaningless information, and add new information as candidate value, which follows memory cell states for modification. The last step is to decide output value based on memory cell state. The sigmoid function will be used initially to confirm output and the tanh activation function will obtain value between -1 and 1 to multiple with sigmoid outcome.

$$i_t = \sigma (x_t W_{ix} + b_x + h_{t-1} W_{ih} + b_h + c_{t-1} W_{ic} + b_c) \quad (1)$$

$$f_t = \sigma (x_t W_{fx} + b_x + h_{t-1} W_{fh} + b_h + c_{t-1} W_{fc} + b_c) \quad (2)$$

$$o_t = \sigma (x_t W_{ox} + b_o + h_{t-1} W_{oh} + b_o + c_{t-1} W_{oc} + b_c) \quad (3)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \quad (4)$$

$$h_t = o_t \odot \tanh(c_t) \quad (5)$$

In equations of (1)-(6), where, $x_t$ is $x$ at time $t$. $W_{ix}, W_{ih}, W_{ic}, W_{fx}, W_{fh}, W_{fc}, W_{ox}$, $W_{gh}, W_{gc}, W_{gh}$, $W_{oh}$ and $W_{oc}$ are weights matrix, the shape of $W_{ix}, W_{fx}, W_{gx}, W_{ox}$ are $(H_{in}, H_{out})$, the shape of are, the shape of $W_{ih}, W_{fh}, W_{gh}, W_{oh}$ are $(H_{out}, H_{out})$, the shape of $W_{ih}, W_{fh}, W_{gh}, W_{oh}$ are $(H_{out}, H_{out})$, the shape of $b_{ix}, b_{ih}, b_{ic}, b_{fx}, b_{fh}, b_{fc}, b_{gx}, b_{gh}, b_{gc}, b_{gh}, b_{gc}, b_{ox}, b_{oh}$ and $b_{oc}$ are biases vector, their length are $H_{out}$.

### TABLE 1. Comparison of traditional model and CNO-LSTM.

| Methods | Data Type | Application Focusing |
|---------|-----------|----------------------|
| NB      | Text      | Text Classification  |
| LR      | Structured Data | Structured Classification & Predication |
| KNN     | Structured Data | Structured Classification & Predication |
| SVM     | Structured Data | Structured Classification & Predication |
| RF      | Structured Data | Structured Classification & Predication |
| CNO-LSTM | Text & Structured Data | Text Classification, Structured Classification & Predication |

### FIGURE 2. Structure of vanilla LSTM.
⊗ is the Hadamard product. Additionally, all the weights and biases are initialized from $U(-\sqrt{a}, \sqrt{a})$, where $a = 1/H_{\text{out}}$. The output of LSTM model is the result of merging $h_t(t = 1, 2, \ldots, T)$, and $\sigma$ is the sigmoid function.

In these formulas, $\delta$ is a sigmoid activation functions to activate $f_t, i_t, o_t$ at time $t$ for forget gate, input gate and output gate respectively. The tanh is also an activation function target for memory cell refreshment at $g_t$ and $h_t$ respectively. The core component for information storage is the memory cell denoted as $c_t$.

LSTM model is naturally designed for long-range context feature dependency extraction as RNN improvement, it has distinct performance with other models and applied widely in industrial projects and research.

## C. CHAOTIC NEURAL MODEL

Chaos is a widespread feature existing in neurons and nonlinear systems as a type of neural dynamic. The advantage of chaotic neural networks covers computational time and memory for numerical analyses, where the complex dynamic patterns are represented by equations. There are different types of oscillatory in neurons with higher accuracy as compared with its baseline model due to variant dynamic adaptiveness for feature learning. The different between static and dynamic is that static is invariable, whereas dynamic absorbs information from nearby to make inference or decision.

Traditional chaotic neural model inspired by Hopfield Network for memory recalling in auto-associate in computer vision and pattern recognition in spatial domain, where only one-layer network structure consists of neurons to store information to recall memory efficiency. However, current study on dynamic neural networks is not limited in vision imitation, which illustrates different methods can achieve neural dynamic in spatial and temporal models.

From neuroscience perspective, brain neural circuits generate activity in complex patterns of extreme spatial and temporal structure but always sensitive to sensory input. Neural network dynamic reviews three types of stimulus-driven neural dynamics with sustained response to transient stimuli for working memory, oscillatory network and chaotic activity. Considering neuronal firing rates, the continuous and oscillatory and chaotic activities with spiking model neurons are discussed. The model structure has sensitivity, selectivity properties and other features, similarly, current dynamic neural networks also have these characteristics such as skip reading to speed up training and testing in text classification. Like attention mechanism with preference and context, the skip pattern follows experiences to ignore punctuations and meaningless single words which have no influence on the understanding of whole sentence.

Neural circuitry procedure generates complex activity patterns are complex as model types are sensitive to sensory input. A comprehensive internal state with variables is displayed, at the same time, when stimulus occurs, external variables allow oscillator to variant. The proposed CNO-LSTM has tested 8 different types of variants for modification as listed in Table 3, its dynamic version to adapt temporal dimension for sequential data on text classification.

### Table 2. Comparison of CNO-LSTM and other variant LSTM in literature review.

| Methods          | Feature Extraction  | Dataset                  | Application Focusing                  | Limitations                  |
|------------------|---------------------|--------------------------|---------------------------------------|------------------------------|
| Tai et al.[30]   | 300-dimensional Glove vectors | 1 SICK                   | Sentiment Analysis                    | Ignore the deeper layers will influence text |
| Zhu et al.[31]   | Stanford parser     | 1 Stanford Sentiment Treebank | Sentiment Analysis                    | Only focusing on Sentiment Analysis |
| Liu et al.[32]   | Embedding layer     | SST-1, SST-2, QC, IMDB   | Sentiment Analysis, Question Classification | Multi-scale LSTM more complex than before |
| Johnson and Zhang[33] | One-Hot Encoding    | IMDB, Elec, RCV1, 20NG  | Sentiment Analysis, News Categorization, Topic Classification | One hot encoding lack of explainable |
| Wan et al. [34]  | Wiki corpus         | 1 Hu et al.              | Question Answering, Sentence completion | Not available in text classification |
| CNO-LSTM         | Tokenizer Embedding layer | IMBD, BoolQ, WOS-11967, AG News, DBpedia | Natural Language Inference, Sentiment Analysis, News Categorization, Topic Classification | Only one-layer can be extended |

In neural network dynamic review, Tim P. Vogels used four categories to classify long-term behavior of dynamical systems - fixed, periodic, quasi-periodic and chaotic. Fixed-point dynamics means the system variables do not change with time but remain in a state. Periodic means variables in a specific interval will alter against time and repeat again following the period. Quasi-periodic has non repetitive property at an interval but consist of more than one periodic pattern with incommensurate frequencies. During processing, the main state of stable fixed, oscillatory and chaotic recognized as attractor will close by the nearby state over time. From neuroscience perspective, it can easily categories these four systems or behaviors into 1) fixed-point and 2) oscillatory. When dynamic behavior links to memory or information processing, the fixed-point resembles short-term memory with a sustained activity feature. Regardless of system complexity and repetitiveness, the frequency of periodic behavior is totally different from one to another.
which is the common feature in oscillatory. Also, periodic
dynamic means the limited cycle attractor, based on that,
 quasi-periodic is complex and non-repetitive for different
frequency properties. Chaotic system consists of oscillatory
elements. In other words, the short-term memory is inoper-
ative for memory recalling, where the final state of system
should retain indications of the initial state set sensitiveness
to stimulus. A series points is the essential form as fixed-
points line attractor to trace back memory. To summarise, the
dynamic behavior or dynamic model of chaotic is developed
by numbers of neurons and its dynamic behavior compet-
itiveness and non-repetitiveness with different frequency for
patterns activities of each neuron. In addition, the oscillatory
state initially changed by internal situation, however, external
stimulus for inhibition and excitation are quintessential to
synchronous or asynchronous the network.

The static network in deep learning with encouraging per-
formances engage high computational cost, memory band-
width and long inference latency in recent years [44] limiting
many industrial and mobile projects with billions of data
required training and efficiency. Neural network research
direction involves deeper layers, more neurons or filters on
computer vision models such as VGG, GoogLeNet, ResNet,
AlexNet. All achieved outstanding performances in task
accuracy as compared with single layer and less neurons.
On the other hand, time-consuming with lot of GPUs devices,
GPUs, memory and efficiency for project executions are con-
cerned. The structures revision of neural model is better than
highly dependence for GPU or TPU economically. However,
from brain science and neuroscience perspective, ANN is imi-
tation of human brain structure for reasoning and inference.
Thus, attention mechanism proposed for NLP tasks is based
on humans’ habitual to relate reading contexts focusing on
useful information. The neural dynamic is like sources from
cognitive science, psychology and even physics of a human’s
mind.

A dynamic neural network survey [43] reviews that the
dynamic network has improved performance to general static
model with its adaptiveness of structural, parameters of dif-
ferent inputs to accuracy and computational efficiency. Tra-
ditional neural dynamic focuses on neurons activity against
abovementioned neural dynamics as current research divides
dynamic networks into 1) instance-wise on data-driven sam-
ple, 2) spatial-wise on image data and 3) temporal-wise on
videos and texts sequential data. These properties comprise of
efficiency, representation and adaptiveness. Efficiency refers
to computation resources allocation for layers or channels to
maximize computational power usage. Representation relates
to weight assigned to model structures and parameters to
expand model capacity that allows the expected application
can composite with attention mechanism in dynamic net-
work. Adaptiveness meets with the balance between accuracy
and efficiency as compared with fixed computational cost.

Dynamic networks applications in NLP nowadays include
associated with Transformer, Question-Answer Chatbot Sys-
tems, Text Classification and Sentiment Analysis. Most of
dynamic models focus on the skipping and jumping to com-
prehend texts like attention mechanism to select key points
based on the context. Skip is to forego meaningless from
the whole document or text to improve text reading effi-
ciency. Furthermore, an abstract method called early exit-
ing [45] only interprets abstract or text introduction to decide
continuity in text classification. For efficiency, vector is
sentence-level or called phrase-level, different with word-
level in the proposed CNO-LSTM model. The dynamic
activation applied to model adjustments. Cao’s team [46]
introduced pos-TAN and LSTM to design an English Transla-
tion Model against chaotic neural network using the dynamic
loss function to adjust class samples contribution during train-
ing procedure.

Oscillatory pattern and theory resemble RNN cycles [21]
from external input, the state of system will change with
time. A chaotic activity is achieved by large number of
neurons and their individual frequency with complex and
non-repetitive situations. The proposed CNO-LSTM model
is composited by original LSTM basic structure but used
Lee-oscillatory dynamic model to replace LSTM cell state
computation by oscillatory non-static state with excitation
and inhibition in series data and oscillatory variants to test
model performance. Both structure and variables are changed
garding to dynamic network.

III. METHODOLOGY
A. CHAOTIC NEURAL OSCILLATORS

Chaotic neural oscillator research began from neuroscience,
neurophysiology, non-linear neural dynamic and oscillations
development in past decades. Chaotic neural network simu-
lates biological neurons ‘real’ neuron behaviors and human
brain with high efficiency for tasks are dynamic, which is a
dynamic neural unit for transient information processing as
compared with ANN simple imitation. Thus, it naturally fits
to high-efficiency or high-frequent tasks in temporal reaction.

Many chaotic neural models based on computational neu-
roscience models are proposed. For example, Hoshino [47]
proposed cortical network for long-term memory recalling,
transient chaotic neural network by Chen [48] for combina-
tional optimization problems, and Wang-oscillator [49], [50]
on spatial-temporal information processing. Since chaotic
neural behaviors simulation with complex ANN application
are understated in brain, Lee [51] proposed Lee-oscillator
based on composite neural-oscillatory model. The common
features of these model are memory related and applied to
pattern association memory recalling, natural language infor-
mation processing and ontology.

\[ E, I, \Omega \text{ and } L \text{ neurons in equations (7)-(10) at time } t, \]
which is the state variables denoted as excitatory, inhibitory,
input and output neurons are shown in Figure 3. The basic
activation function sigmoid function is the used, revealing
start point of equations with \( f() \) and assisted by trained weight
parameters, \( e_1, e_2, l_1 \text{ and } l_2, \) to obtain oscillator best per-
formance in bifurcation diagram. \( \xi_{eu} \text{ and } \xi_e \) are the thresholds for
excitatory and inhibitory neurons. The external input stimulus work with \( I_2 \), and \( k \) is the decay constant at decay operator.

Neural dynamic and temporal information processing properties refined Lee-oscillator structure functions like human brain on LSTM modification. There are four aspects for potential oscillator application and research: chaotic neural elements, chaotic auto-associator, chaotic neural oscillatory model and chaotic bifurcation transfer unit (BTU).

1) CHAOTIC NEURAL ELEMENTS FOR INFORMATION PROCESSING

Information processing requires to process data such as text, images for recognition, classification, and prediction. The feature of oscillator is a two-state for input space identical to LSTM. Hence, oscillator facilitates information processing and neural dynamic into chaotic neural elements to simulate data processing variants of human brain science behavior.

2) CHAOTIC AUTO-ASSOCIATOR

The structure of oscillator stores information into memory recalling schemes. Hopfield Network are combined with 2D layer Lee-Oscillators in pattern association for information processing in progressive memory association and memory recalling scheme for oscillator implementation [51].

3) CHAOTIC NEURAL OSCILLATORY MODEL

Lee-Oscillator is equipped with model’s quintessential elements. It can integrate with other models to explore a complicated chaotic neural oscillatory model to compensate shortcomings with its properties. The proposed CNO-LSTM model is a Lee-Oscillator with LSTM advantages but can solve LSTM time redundancy problem to improve accuracy against vanilla model.

4) CHAOTIC BIFURCATION TRANSFER UNIT AS ACTIVATION FUNCTION

Since neuron state is effected when the activation function is implemented, the original Lee-Oscillator activation function is sigmoid function. However, the role of Lee-Oscillator with bifurcation transfer unit can also be used as activation function with current functions to improve contemporary neural networks efficiency and chaotic growth in neural dynamics.

Equations (7)-(10) formulated the parameters of bifurcation transfer unit, internal and external stimulus will change oscillator shapes to adapt the data feature in neural dynamic model, provide competitive performance for comparison to identify the best to qualify neural network efficiency.

The proposed CNO-LSTM substitutes LSTM for temporal or transient information processing by Lee-Oscillator neural dynamic state variables to accelerate models training against the relationships within neurons for structure revision from external stimulus. The chaotic feature of dynamic neural model with Lee-Oscillator dynamic equations are given by:

\[
\begin{align*}
E_t &= \sigma (e_1 E_{t-1} - e_2 I_{t-1} + x_t - \xi_E) \quad (7) \\
I_t &= \sigma (i_1 E_{t-1} - i_2 I_{t-1} - \xi_I) \quad (8) \\
\Omega_t &= \sigma (x_t) \quad (9) \\
Lee_t &= (E_t - I_t) e^{-k \xi^2} + \Omega_t \quad (10)
\end{align*}
\]

In equations of (7)-(10), where \( E, I, \Omega \) and \( Lee \) is the excitatory, inhibitory, input and output neurons respectively, \( e_1, e_2, i_1 \) and \( i_2 \) is weights, \( \xi_I \) and \( \xi_E \) is threshold, \( x_t \) is input value. And the Lee-Oscillator output is the result of merging \( Lee_t (t = 1, 2, \ldots, T) \).

LSTM structure and Lee-Oscillator are combined to modify neural dynamics instead of using the loss or activation function, where Lee-Oscillator parameters are used for the best fitting with CNO-LSTM. Table 3 shows the method of author’s previous work on CT2TFDNN model [53] using different bifurcation parameters to qualify the best parameters for neural oscillator. These figures in different types are tested by the loop from a classic oscillator.

**B. CNO-LSTM MODEL**

RNN is a sequence-to-sequence model for time-series or NLP task designed for the correlation within sequence, such as next words predication in text generation. It is used as language model to model features in NLP. However, the gradient is unable to transmit with more information because of gradient exploding and vanishing problems reverting RNN...
to acquire long-distance information to process the magnitude of sequential data despite with bi-directional or Deep RNNs variants. Hence, sequence representation learning is a meaning presentation method to prove LSTM has remarkable performance for NLP tasks [39].

In order to fulfill various Text Classification scenario requirements, LSTM is a sequential model to handle contextual data with long-range sequence and solve traditional RNN with memory lost and gradient exploding or vanishing problems. However, a standard LSTM model is too complex to compute especially with billions of data for industrial project as compared with RNN, GRU, or CNN. Figure 2 shows a vanilla LSTM. It is an artificial computation architecture for memory store and update with quintessential selection as filter [20].

From application perspective, RNN integrates the sequential part with all encoder-decoder pipeline calculation for information processing as a language model. In LSTM, memory cell is to refresh information storage but processing updates computation is time-consuming.

Therefore, a Chaotic Neural Oscillatory LSTM is proposed to simplify gates computation to process new information, extend memory cells storage to accelerate memory recalling and state updates. CNO-LSTM formulation is to first reserve input gate and forget gate, second the output gate is from input with $\sigma()$, where input gate affects output directly to avoid influences from hidden layers and memory cells. At the same time, the input modulation gate $g_t$ is deleted to expedite updates of whole system and reduce the complexity of vanilla LSTM, which means memory cell updates are decided by input and forget gates to ensure that output $h_t$ is influenced by $x_t$, $h_{t-1}$ and $c_{t-1}$ simultaneously. A CNO-LSTM workflow is shown below:

$$i_t = \sigma (i_1 h_{t-1} + i_2 c_{t-1} + x_t + \xi_t)$$  \hspace{1cm} (11)

$$f_t = \sigma (f_1 h_{t-1} + f_2 c_{t-1} + x_t + \xi_f)$$  \hspace{1cm} (12)

$$o_t = \sigma (x_t)$$  \hspace{1cm} (13)

$$c_t = f_t \otimes c_{t-1} + i_t$$  \hspace{1cm} (14)

$$h_t = o_t \otimes \tanh(c_t)$$  \hspace{1cm} (15)

Equations (11)–(15) have similar equations expression of LSTM and Lee-Oscillator, where $\sigma$ is denoted as Sigmoid activation function and is the original activation function of Lee-Oscillator and standard LSTM. The comparison table for the parameters in Lee-Oscillator and proposed CNO-LSTM shown at Table 4.

| Parameters | Type1 | Type2 | Type3 | Type4 | Type5 | Type6 | Type7 | Type8 |
|------------|-------|-------|-------|-------|-------|-------|-------|-------|
| $i_1$      | 5     | 5     | 5     | 5     | 1     | 1     | 5     | 5     |
| $i_2$      | 5     | 5     | 5     | 5     | 1     | 1     | 6     | 6     |
| $f_1$      | 5     | 5     | 5     | 5     | 1     | 1     | 1     | 1     |
| $f_2$      | 5     | 5     | -5    | -5    | -1    | -1    | 1     | 1     |
| $\xi_t$    | 0     | 1     | 0     | 1     | 0     | 1     | 0     | 1     |
| $\xi_f$    | 0     | 1     | 0     | 1     | 0     | 1     | 0     | 1     |

**Algorithm 1 CNO-LSTM**

1. Begin
2. Initialize weights and bias of network, and parameters of Lee-oscillator
3. Load the training data
4. While (training times < total training times)
5. For each sample of the training data
6. Compute the hidden neurons utilizing from Eq. (11) to Eq. (15)
7. End for
8. Obtain network output and calculate the loss function
9. Update weights and bias of network
10. End While
11. End

It showed that original six formulas are reduced to five due to input modulation gate $g_t$ is deleted. Also, Lee-Oscillator is neural dynamic and good at memory recalling to replace LSTM structure for high frequency updates and accuracy for memory recalling.

In Lee-Oscillator model, the Excitatory and Inhibitory neurons exercise the flow of new information from input neuron, like LSTM, new information are selected by the model to
than QA task, whose dataset is separated to 9,427 labeled ‘no’. In other words, BoolQ is a binary classification rather than a multiple classification and multi-label classification for model evaluation. CNO-LSTM calculation consists of input model structure, output and forget gate are substituted by Lee-Oscillator dynamic structure. Hence, the calculation of CNO-LSTM is the combination of Lee-Oscillator and vanilla LSTM to reserve the functions of quintessential structure.

IV. IMPLEMENTATION AND COMPARISON

This section shows CNO-LSTM performance for Text Classification with five different datasets, where Lee-Oscillator parameters contribute to the proposed model with eight forms adaption from original paper [53].

A. EXPERIMENTAL SETUP

1) DATASET

There are five datasets used for Text Classification including 1) BoolQ for Question-Answer Classification, 2) IMDb dataset for Sentiment Analysis, 3) WOS-11967 for Multi-label Classification, 4) AG News for News Categorization, and 5) DBpedia for Topic Classification covering binary, multiple classification and multi-label classification for model evaluation.

BoolQ (Boolean Questions) is a question-answer dataset containing 15,942 sample pairs, its performance is different from Question-Answer interaction system. It has a triple form, question-passage-answer denoted as queries selected by search engine with Wikipedia background information as passage information assistant to judge QA answer is ‘yes’ or ‘no’. In other words, BoolQ is a binary classification rather than QA task, whose dataset is separated to 9,427 labeled training, 3,270 labeled development, and 3,245 unlabeled test examples. QA for machine, to some extent, is the natural language inference to understand the contents of given texts.

IMDB is a benchmark dataset consist of non-repetitive movie reviews for binary sentiment analysis. According to classification rules, where sentiment label is 0, rank less than 5 will be classified as negative, in contrast, rank more than 7 will be classified as positive, where sentiment label is 1. It has 25,000 English comments dataset size with website ID, sentiment labels and original reviews in natural language for training and corresponding 25,000 is test dataset without any labels. That means it provides 25,000 movie reviews for training, and 25,000 for testing.

Web of Science-11967 (WOS-11967) is a subset of WOS-46,985. WOS is a multi-label dataset, which means one instance has more than one labels against other datasets used. Its source data of WOS is from the web of science, whose samples are abstracts of published papers cited by WOS. The samples in WOS-11967 has two labels, where dataset includes 11,967 documents with 35 categories from 7 parent categories.

AG News is a subset of Agriculture News (AG)’s Group, according to titles and descriptions of news articles to allocate classes, which include World, Sports, Business, Sci/Tech, for News Categorizations. It has 120,000 English News for training in total, where for each categorization of 30,000 samples and test 7,600 datasets.

DBpedia is the largest multi-language knowledge base for Topic Classification created by Wikipedia with 14 labels for each sample. Its source data is the structured data with 9 different classification schemata for things. It provides 308,375 texts for training, and 34,367 texts for testing.

2) DATA PREPROCESSING

The dataset in CNO-LSTM model contains both structured data and pure text, the data preprocessing with basic NLP technologies or called NLP pipeline should be completed before further processing against traditional methods in handling structured data. The steps are follows:-

Firstly, Natural Language Toolkit (NLTK) is used for word segmentation and remove all punctuations simultaneously. Also, upper case is transformed to lower case in alphabetical order.

Secondly, tokenization is to convert text into sequence-as-sequential data for LSTM target format as a sequential model. Here, tokenizer splits text in each dataset into a set of words and reorder word frequency in descending order to form a corresponding dictionary with unique index for each word.

Due to each dataset has different sequence length, each dataset is assigned with its own dictionary standardized with fixed length which shows IMDB, BoolQ, WOS-11967, AG News and DBpedia are set at 100, 15, 150, 50 and 50 as shown in Table 5 respectively.

Before training, data preprocessing split Training and Validation ratio for each dataset. The ratio for these 5 dataset is 9:1, which means 90% is for training and 10% is for...
TABLE 5. Description of FIVE data sets used.

| Datasets    | Sample size | Number of words | Class | Sequence length |
|-------------|-------------|-----------------|-------|-----------------|
| IMDB        | 50,000      | 105,637         | 2     | 100             |
| BoolQ       | 15,942      | 8,601           | 2     | 15              |
| WOS-11967   | 11,967      | 181,762         | 7     | 150             |
| AG News     | 127,600     | 86,488          | 4     | 50              |
| DBpedia     | 342,742     | 722,817         | 9     | 50              |

FIGURE 5. Distribution of datasets.

validation within one dataset. The testing dataset follows the original data splitting. The distribution is shown in Figure 5.

3) TRAINING ENVIRONMENT
The hardware environment is Intel Core i5-10200H CPU with 2.4GHz and 16GB memory for the acceleration of NVIDIA GeForce RTX 2060 GPU. The framework of Python is PyTorch 1.80 based on Window 10.

B. MODEL PERFORMANCE
According to chaotic neural networks characteristics, the prominent results focus on computation efficiency against baseline models, especially for vanilla LSTM model. Due to LSTM structure, neuron features are modified to aim for oscillatory and chaotic generation to reduce LSTM computational time, power and improve model accuracy.

The selected 4 baseline model is RNN, CNN, LSTM and Bi-direction LSTM (Bi-LSTM) is showed in Table 8. Since LSTM is a RNN variant modification to solve gradient vanishing and exploding problems, when LSTM model is revised, it is naturally to compare with RNN to retain the advantage of RNN with property of recurrent with original one. TextCNN [54] means CNN model is applied to NLP tasks, where the kernel with different sizes are used to extract key information and identify the local relations within text. TextCNN is simpler than traditional one because it has only one-layer convolution and max-pooling model to calculate efficiency with satisfying accuracy in multi-classification.

CNO-LSTM has a similar one-layer model setting as baseline model under the condition listed in Table 6, where LSTM is a one-layer setting with 128 neurons to extract models features without any Dropout or L2 index that may influence model accuracy.

The advantages of baseline model mainly focus on text relations within models. As LSTM used long-term memory to store meaningful information to decide or improve prediction accuracy, the same objective is to revise RNN for time-series or CNN for image feature extraction to the model, which is more suitable to LSTM on Text Classification in NLP tasks. CNO-LSTM with correlations of input data computed in neurons to achieve high frequency for information processing in a chaotic state.

According to authors’ previous work, 8 types of oscillator parameters for model are used. Experimental results are shown in Table 7. The results showed that bold figures achieved the best accuracy among 8 variants, where different datasets match with various oscillator parameters, which reflect the contents in memory recalling correspond to unlike neural states. The contents are considered as a set of word patterns representation which express the feature of text from dataset by the parameters in oscillator. Thus, in this step, the best type among 8 oscillator variants can compare with baseline model and expected further improvement under the condition of attention mechanism.

The 4 baseline models (RNN, TextCNN and LSTM and Bi-direction LSTM) are compared with CNO-LSTM variants with most suitable parameters as shown in Table 8. It showed that the simplified CNO-LSTM model requires less time to compute against complex vanilla LSTM model. It also showed that LSTM requires approximate double the time to compare with RNN and TextCNN due to its complex structure to exercise memory from input to output for judgement and selection. LSTM performed better in accuracy than other three baseline models within 4 baseline models, apart from

TABLE 6. Model parameters.

| Parameter                       | Value |
|---------------------------------|-------|
| Epoch                           | 20    |
| Maximum Sequence Length         | 50    |
| Word Vector Dimension           | 128   |
| Validation Set Ratio            | 0.1   |
| Test Set Ratio                  | 0.1   |
| Batch Size                      | 64    |
TABLE 7. Model training results for eight different types parameters with Lee oscillator.

| Parameters                      | LSTM with Lee-oscillator |
|---------------------------------|---------------------------|
|                                 | Type1 | Type2 | Type3 | Type4 | Type5 | Type6 | Type7 | Type8 |
| Accuracy (%)                    | 61.56 | 62.57 | 61.13 | 62.42 | 57.83 | 61.87 | 61.93 | 61.38 |
| Training Time (seconds)         | 158   | 155   | 153   | 163   | 157   | 160   | 159   | 152   |
| Accuracy (%)                    | 73.28 | 72.73 | 63.67 | 72.47 | 73.42 | 72.67 | 72.28 | 73.60 |
| Training Time (seconds)         | 1244  | 1214  | 1222  | 1222  | 1213  | 1228  | 1237  | 1282  |
| Accuracy (%)                    | 70.50 | 70.65 | 69.76 | 70.05 | 70.82 | 67.34 | 64.13 | 70.91 |
| Training Time (seconds)         | 1978  | 1959  | 1927  | 1885  | 1910  | 1899  | 1968  | 1923  |
| Accuracy (%)                    | 89.38 | 89.22 | 88.58 | 88.72 | 89.04 | 88.70 | 89.38 | 89.29 |
| Training Time (seconds)         | 2748  | 2734  | 2745  | 2799  | 2780  | 2785  | 2786  | 2782  |
| Accuracy (%)                    | 98.89 | 98.87 | 98.88 | 98.86 | 98.87 | 98.65 | 98.81 | 98.85 |
| Training Time (seconds)         | 13600 | 13341 | 13208 | 13538 | 13099 | 13115 | 13230 | 13209 |

TABLE 8. Comparison of baseline models and CNO-LSTM.

| Datasets    | Parameters      | RNN  | TextCNN | Bi-LSTM | LSTM | CNO-LSTM |
|-------------|-----------------|------|---------|---------|------|----------|
| BoolQ       | Accuracy (%)    | 61.74| 62.17   | 63.49   | 61.01| 62.57    |
|             | Training Time (seconds) | 100  | 115     | 705     | 241  | 155      |
| IMDB        | Accuracy (%)    | 50.00| 72.11   | 71.06   | 72.74| 73.60    |
|             | Training Time (seconds) | 914  | 934     | 2530    | 1805 | 1282     |
| WOS-11967   | Accuracy (%)    | 17.59| 68.63   | 67.24   | 69.15| 70.91    |
|             | Training Time (seconds) | 1345 | 1391    | 3811    | 2565 | 1923     |
| AG News     | Accuracy (%)    | 25.00| 88.54   | 87.01   | 88.61| 89.38    |
|             | Training Time (seconds) | 2706 | 2799    | 9058    | 5162 | 2748     |
| DBpedia     | Accuracy (%)    | 51.76| 98.46   | 98.29   | 98.11| 98.89    |
|             | Training Time (seconds) | 10287| 10638   | 27361   | 19618| 13600    |

BoolQ for QA pairs and DBpedia for multi-labels because of long text processing with time-series property.

It is noted that when CNO-LSTM is implemented with 8 types of variants, the performance of best matching variant with specific dataset performed better in computation efficiency and time than almost all baseline models with overall datasets. Accuracy improvement is 1.56 for BoolQ, 0.86 for IMDB, 1.76 for WOS-11967, 0.77 for AG news and 0.78 for DBpedia respectively compared with LSTM model. It showed that different datasets and neural models with neurons correlations in neural dynamic states of chaotic neural models are adaptive with improved accuracy.

Further, CNO-LSTM 8 variants training time has achieved remarkable progress compared with baseline models because high-frequency non-period oscillatory to perform information processing within neurons and redundant cell in vanilla LSTM. The highest time improvement is 78.01% with 550s, which is calculated by the subtraction of training time with CNO-LSTM 155s and Bi-LSTM 705s for BoolQ, where accuracy improvement depends highly on complex computation. It also showed that CNO-LSTM can not only solve vanilla LSTM training time problem in broader application in industrial project that relies on GPU for acceleration, but also maintains almost the same or relative higher accuracy.
TABLE 9. Comparison of different composition of models.

|          | BoolQ  | IMDB   | WOS-11967 |
|----------|--------|--------|-----------|
| LSTM     | 61.74  | 60.18  | 62.17     |
| TextCNN  | 50.00  | 57.39  | 72.11     |
| Bi       | 914    | 1089   | 934       |
| Lee-oscillator | 17.59  | 26.28  | 68.63     |
| Attention| 1345   | 1547   | 1391      |

C. COMPARISON WITH BASELINE MODELS

LSTM chaotic state, attention mechanism [56]. Chaotic RNN are used for text classification to compare with first 3 datasets: BoolQ, IMDB and WOS-19967, to demonstrate CNO-LSTM model performance. In this part, CNO-LSTM with and without attention are compared with vanilla LSTM with attention mechanism in the last three comparison at Table 9. Comparison of different models’ composition is shown in Table 9.

For the first layer, it showed that although RNN and TextCNN performed better than CNO-LSTM in time consumption, but accuracy performance are below CNO-LSTM. Also, LSTM and CNO-LSTM performed better in time consumption and accuracy.

For the second layer, it showed that CNO-LSTM and Chaotic RNN models are restructured. As mentioned in methodology section, Lee-oscillator was added to CNO-LSTM which means the model state in a dynamic way and different from other baseline models. In this section, Chaotic RNN is used to compare with other composition.

The original RNN formula is calculated by:

\[ h_t = \tanh (x_t W_{hx} + b_{hx} + h_{t-1} W_{hh} + b_{hh}) \] (16)

In equation (16) where, \( x_t \) is x at time t, \( W_{hx} \) and \( W_{hh} \) are weights matrix, the shape of \( W_{hx} \) is \((H_{in}, H_{out})\), the shape of \( W_{hh} \) is \((H_{out}, H_{out})\), \( b_{hx} \) and \( b_{hh} \) are biases vector, their length are \( H_{out} \). In addition, all the weights and biases are initialized by \( U(-\sqrt{a}, \sqrt{a}) \), where \( a = 1/H_{out} \). And RNN model output is the result of merging \( T h_t (t = 1, 2, \ldots, T) \).
When RNN is combined with equations (7)-(10), Chaotic RNN is shown below:

\[ h_t = \tanh \left( hh_{t-1} + x_t + \xi_h \right) \tag{17} \]

In equations of (17), input weights are removed to reduce computation complexity.

It is also the reasons LSTM is selected to avoid gradient vanishing and exploding problems, and with CNO-LSTM property in memory recalling and computation to improve accuracy. The results also showed it can compensate inadequate parts with each other like oscillatory in neural dynamic for memory recalling and LSTM in neural network for long-term memory.

For the third layer, similar phenomenon occurred when attention mechanism is added, the results showed that CNO-LSTM has satisfactory performance in accuracy and time consumption.

When analyzing the performance of best variants for CNO-LSTM and LSTM under condition of attention compared with baseline models, accuracy with attention has improved significantly. The training time for CNO-LSTM remained advantageous, and added attention is the growth point for accuracy. For example, the time reduction for BoolQ dataset binary classification is maximized to 35.86%, which is calculated by 237 of LSTM subtract 152 of CNO-LSTM under attention. The other two dataset accuracy has increased 3.47 account to 4.73% of WOS-11967 from 73.30 to 76.77, corresponding time reduction is 835 seconds, 30.69% from 2721s to 1886s. For IMDB is increased 1.81 account to 2.42% from 74.67 to 76.48, corresponding time reduction is 455 seconds, 26.75% from 1,701s to 1,246s. Disregarding other components such as the Dropout or L2 for model accuracy improvements, which not be added in modeling, the acceleration and precision in experiment totally attributed to revised structure of CNO-LSTM and its neuron computation variants.

The visualization of different types of performance is shown on Figure 6. First one is BoolQ dataset where attention LSTM is without any additional settings in neural network. The revision of CNO-LSTM attention with Type 1 is the parameters for CNO-LSTM model to transfer neural network dynamics for high-frequency oscillation. Second is the IMDB dataset and last is WOS-11967, where the visualization in Figure 6 is the result listed in Table 9 with attention mechanism.

The series visualization of bifurcation oscillator model comparison of 8 variants, apart from Type 0 means non-oscillator added in model, LSTM with attention is applied. The best parameter with lowest time consumption and highest accuracy is Type 1 for BoolQ, Type3 for IMDB and Type6 for WOS-11967. The disparity of LSTM with attention and CNO-LSTM with attention are contained in all 3 sub-figures

It also showed that regularity in line chart has high accuracy with minimized time consumption in computation efficiency and model accuracy.

**D. EVALUATION**

Further to general index, accuracy and computation time span, evaluation index from confusing matrix [55] is used for Precision, Recall and F1 Score to compensate accuracy inadequacy for model evaluation.

In confusing matrix, there are four types True or False denoted as True Positive, False Positive (Type I error), False Negative (Type II error) and True Negative.

Precision represents True Positive percentage among True Positive and False Positive, where the correct prediction
with inaccurate predication for False to be True for positive predictive value.

Recall is the contrast of Precision to calculate the real True predicate, which means the percentage of True Positive among all True values including True Positive and False Negative.

F1 Score is the combination of Precision and Recall.

The output figures from WOS-11967, of multi-label classification with 7 class for evaluation are shown in Table 10. CNO-LSTM achieved satisfactory performance than all four indices as compared with baseline models. It also has satisfactory performance when attention is added than LSTM without oscillator, which proves the contribution of the restructure of our proposed model.

\[
\text{Accuracy} = \frac{1}{K} \sum_{i=1}^{K} (y_i = \hat{y}_i) \\
\text{Precision} = \frac{TP}{TP + FP} \\
\text{Recall} = \frac{TP}{TP + FN} \\
F1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}
\]

V. DISCUSSION AND CONCLUSION

In this paper, CNO-LSTM model is proposed for Text Classification to apply five datasets for testing and comparison with state-of-art models for Question-Answer Classification Sentiment Analysis, News Categorization and Topic Classification with binary, multi classification and multi-label classification.

CNO-LSTM is a compositied neural network by Lee-Oscillator architecture with chaotic oscillatory property for time efficiency and LSTM for memory storage with Memory Recalling to improve accuracy on relations within chaotic neurons in text classification.

Although CNO-LSTM model structure has improvement in computational efficiency for time consumption, but further modification is required. Firstly, regardless of LSTM and other baseline models, CNO-LSTM model with 8 variants have only one-layer model without additional setting for higher accuracy. Thus, the matching for CNO-LSTM model with other parameters or methods required further testing. Secondly, model generalization ability required to further improve despite model implementation has improved accuracy but test datasets remained to have accuracy disparities. Hence, dataset features have influence in model performance, which means model selection should align with dataset features for application and model specialization.

In conclusion, the proposed CNO-LSTM model has the following contributions: 1) Speed improvement – a simplified model has reduced training time between static and dynamic models, 2) Higher accuracy - generated higher or approximate equal accuracy with series variants from Lee-Oscillator, 3) Pure contribution from model itself – is tested by using only one-layer network without other methods to verify model contribution, 4) Extensive applications - do not restricted in structured data but extend to all pure texts with binary, multi and multi-label classification for Natural Language Understanding.

In this paper, text classification for different branches of NLP tasks and dataset in various disciplines with QA dataset, sentiment analysis, news categorization and topic classification, the anticipated. For the next step, CNO-LSTM can have the ability for data generation and implement it into text generation for NLP Chatbot in connection with ontology knowledge.

REFERENCES

[1] Z. Liu, X. Di, W. Song, and W. Ren, “A sentence-level joint relation classification model based on reinforcement learning,” Comput. Intell. Neurosci., vol. 2021, pp. 1–10, May 2021, doi: 10.1155/2021/5557184.
[2] I. Ameen, G. Sidorov, H. Gomez-Adorno, and R. M. A. Nawab, “Multi-label emotion classification on code-mixed text: Data and methods,” IEEE Access, vol. 10, pp. 8779–8789, 2022, doi: 10.1109/ACCESS.2022.3143819.
[3] F. Tao and G. Liu, “Advanced LSTM: A study about better time dependency modeling in emotion recognition,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), Apr. 2018, pp. 2906–2910, doi: 10.1109/ICASSP.2018.8461750.
[4] F. Es-Sabery, I. Es-Sabery, A. Hair, B. Sainz-De-Abajo, and B. Garcia-Zapirain, “Emotion processing by applying a fuzzy-based vader lexicon and a parallel deep belief network over massive data,” IEEE Access, vol. 10, pp. 87870–87899, 2022, doi: 10.1109/ACCESS.2022.3200389.
[5] Y. Li and N. Li, “Sentiment analysis of Weibo comments based on graph neural network,” IEEE Access, vol. 10, pp. 23497–23510, 2022, doi: 10.1109/ACCESS.2022.3154107.
[6] F. Arias, M. Z. Nunez, A. Guerra-Adames, N. Tejedor-Flores, and M. Vargas-Lombardo, “Sentiment analysis of public social media as a tool for health-related topics,” IEEE Access, vol. 10, pp. 74850–74872, 2022, doi: 10.1109/ACCESS.2022.3187406.
[7] X. Cao, J. Yu, and Y. Zhuang, “Injecting user identity into pretrained language models for document-level sentiment classification,” IEEE Access, vol. 10, pp. 30157–30167, 2022, doi: 10.1109/ACCESS.2022.3158975.
[8] X. Chen, P. Cong, and S. Lv, “A long-text classification method of Chinese news based on BERT and CNN,” IEEE Access, vol. 10, pp. 34046–34057, 2022, doi: 10.1109/ACCESS.2022.3162614.
[9] I. Fursov, A. Zaytsev, P. Burnychev, E. Dmitrieva, N. Klyuchnikov, A. Kravchenko, E. Artemova, E. Komleva, and E. Burnaev, “A differentiable language model adversarial attack on text classifiers,” IEEE Access, vol. 10, pp. 17966–17976, 2022, doi: 10.1109/ACCESS.2022.3148413.
[10] G. Althari and M. Alsulmi, “Exploring transformer-based learning for negation detection in biomedical texts,” IEEE Access, vol. 10, pp. 83813–83825, 2022, doi: 10.1109/ACCESS.2022.3197772.
[11] Q. Qi, L. Lin, R. Zhang, and C. Xue, “MEDT: Using multimodal encoding-decoding network as in transformer for multimodal sentiment analysis,” IEEE Access, vol. 10, pp. 28750–28759, 2022, doi: 10.1109/ACCESS.2022.3157712.
[12] Y. Bengio, R. Ducharme, P. Vincent, and C. Janvin, “A neural probabilistic language model,” J. Mach. Learn. Res., vol. 3, pp. 1137–1155, Feb. 2003.
[13] E. Únal, Y. Ji, and A. Papafragou, “From event representation to linguistic meaning,” Topics Cognit. Sci., vol. 13, no. 1, pp. 224–242, Jan. 2021, doi: 10.1111/TPS.12475.
[14] A. Alsharief, K. Aggarwal, D. Koundal, H. Alyami, and D. Ameyed, “An automated toxicity classification on social media using LSTM and word embedding,” Comput. Intell. Neurosci., vol. 2022, pp. 1–8, Feb. 2022, doi: 10.1155/2022/8467349.
[15] T. Mikolov, L. Sutskever, K. Chen, G. Corrado, and J. Dean, “Distributed representations of words and phrases and their compositionality,” in Proc. Adv. Neural Inf. Process. Syst., vol. 2, Dec. 2013, pp. 3111–3119, doi: 10.5555/2999792.2999959.
[16] J. Wang and J. Wang, “A new hybrid forecasting model based on SW-LSTM and wavelet packet decomposition: A case study of oil futures prices,” Comput. Intell. Neurosci., vol. 2021, pp. 1–22, Jul. 2021, doi: 10.1155/2021/7653091.

[17] Z. Cui, Y. Zhou, S. Guo, J. Wang, H. Ba, and S. He, “A novel hybrid XAI-LSTM model for multi-step-ahead flood forecasting,” Hydrol. Res., vol. 52, no. 6, pp. 1436–1454, Dec. 2021, doi: 10.2166/NH.2021.016.

[18] F. Huang, X. Li, C. Yuan, S. Zhang, J. Zhang, and S. Qiao, “Attention-emotion-enhanced convolutional LSTM for sentiment analysis,” IEEE Trans. Neural Netw. Learn. Syst., vol. 33, no. 9, pp. 4332–4345, Sep. 2022, doi: 10.1109/TNNLS.2021.3056664.

[19] J. Du, C. Vong, and L. C. P. Chen, “Novel efficient RNN and LSTM-like architectures: Recurrent and gated broad learning systems and their applications for text classification,” IEEE Trans. Cybern., vol. 51, no. 3, pp. 1586–1597, Mar. 2021, doi: 10.1109/TCYB.2020.2969705.

[20] Y. Yu, X. Si, S. Hu, and Z. Jianxun, “A review of recurrent neural networks: LSTM cells and network architectures,” Neural Comput., vol. 31, no. 7, pp. 1235–1270, Jul. 2019, doi: 10.1162/neco_a_01199.

[21] T. P. Vogels, K. Rajan, and L. F. Abbott, “Neural network dynamics,” Annu. Rev. Neurosci., vol. 28, pp. 357–376, Jul. 2005, doi: 10.1146/annurev.neuro.28.061604.135637.

[22] R. S. T. Lee, “A transient-chaotic autoassociative network (TCAN) based on Lee oscillators,” IEEE Trans. Neural Netw., vol. 15, no. 5, pp. 1228–1243, Sep. 2004, doi: 10.1109/TNN.2004.832729.

[23] A. S. Aljaloud, D. M. Uliyan, A. Alkhalil, M. A. Elrhman, A. F. M. Alogoli, Y. M. Altameemi, M. Altamimi, and P. Kwan, “A deep learning model to predict Student learning outcomes in LMS using CNN and LSTM,” IEEE Access, vol. 10, pp. 85255–85265, 2022, doi: 10.1109/ACCESS.2022.3196784.

[24] H. K. Maragheh, F. S. Gharachopogh, K. Majidzadeh, and A. B. Sangar, “A new hybrid based on long short-term memory network with spotted hyena optimization algorithm for multi-label text classification,” Mathematika, vol. 10, no. 3, p. 488, Feb. 2022, doi: 10.1112/math.1003488.

[25] B. A. H. Murshed, J. Abawajy, S. Mallappa, M. A. N. Saif, and H. K. Maragheh, F. S. Gharehchopogh, K. Majidzadeh, and A. B. Sangar, “Dynamic neural networks: A survey,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 11, pp. 7436–7456, Nov. 2022, doi: 10.1109/TPAMI.2021.3117837.

[26] W. Xia, H. Yin, X. Dai, and N. K. Jha, “Fully dynamic inference with deep neural networks,” IEEE Trans. Emerg. Topics Comput., vol. 10, no. 2, pp. 962–972, Jun. 2022, doi: 10.1109/TEC.2021.3056031.

[27] X. Liu, L. Mou, H. Cui, Z. Lu, and S. Song, “Finding decision jumps in text classification,” Neurocomputing, vol. 371, pp. 177–187, Jan. 2020, doi: 10.1016/j.neucom.2019.08.082.

[28] Q. Cao and H. Hao, “A chaotic neural network model for english machine translation based on big data analysis,” Comput. Intell. Neurosci., vol. 2021, pp. 1–12, Jul. 2021, doi: 10.1155/2021/5274326.

[29] H. Luo, Y. Huang, X. Du, Y. Zhang, A. L. Green, T. Z. Aziz, and S. Wang, “Dynamic neural network with a spatiotemporal self-learning model with delayed feedback,” Neural Comput. Appl., vol. 37, no. 5, pp. 1735–1750, Sep. 2020, doi: 10.1007/s00521-019-04836-1.

[30] R. S. T. Lee, “A transient-chaotic autoassociative network (TCAN) based on Lee oscillators,” IEEE Trans. Neural Netw., vol. 15, no. 5, pp. 1228–1243, Sep. 2004, doi: 10.1109/TNN.2004.832729.

[31] Y. Han, G. Huang, S. Song, L. Yang, H. Wang, and Y. Wang, “Dynamic neural networks,” IEEE Access, vol. 10, pp. 1228–1243, Sep. 2004, doi: 10.1109/TNN.2004.832729.

[32] Q. Cao and H. Hao, “A chaotic neural network model for english machine translation based on big data analysis,” Comput. Intell. Neurosci., vol. 2021, pp. 1–12, Jul. 2021, doi: 10.1155/2021/5274326.

[33] H. Luo, Y. Huang, X. Du, Y. Zhang, A. L. Green, T. Z. Aziz, and S. Wang, “Dynamic neural network with a spatiotemporal self-learning model with delayed feedback,” Neural Comput. Appl., vol. 37, no. 5, pp. 1735–1750, Sep. 2020, doi: 10.1007/s00521-019-04836-1.

[34] R. S. T. Lee, “A transient-chaotic autoassociative network (TCAN) based on Lee oscillators,” IEEE Trans. Neural Netw., vol. 15, no. 5, pp. 1228–1243, Sep. 2004, doi: 10.1109/TNN.2004.832729.

[35] Y. Han, G. Huang, S. Song, L. Yang, H. Wang, and Y. Wang, “Dynamic neural networks,” IEEE Access, vol. 10, pp. 1228–1243, Sep. 2004, doi: 10.1109/TNN.2004.832729.
NUOBEI SHI received the M.Sc. degree (Hons.) in data science from the Beijing Normal University–Hong Kong Baptist University United International College (UIC), Hong Kong Baptist University, in 2020, where she is currently pursuing the M.Phil. degree in computer science and technology. Her research interests include natural language processing, dynamic neural networks, ontology graph, text classification, chatbot in natural language processing, and explainable artificial intelligence (XAI).

ZHUOHUI CHEN is currently pursuing the master’s degree in intelligent technology with the Macau University of Science and Technology. He is a Research Assistant with the Beijing Normal University–Hong Kong Baptist University United International College (UIC). His research interests include neural networks, object detection, and pattern recognition.

LING CHEN (Member, IEEE) received the B.Sc. degree in mathematics from Jinan University, China, in 1982, and the M.Sc. degree in mathematics from the University of Windsor, Canada, in 1995. He has over 35 years of experience in telecommunications, IT, accounting firms, government, and universities in multiple countries. He is currently a Professor with the Beijing Institute of Technology, Zhuhai, China. His research interests include computer vision, natural language processing, speech recognition, optical character recognition, and signatures verification. He is a member of CCF.

RAYMOND S. T. LEE (Member, IEEE) received the B.Sc. degree in physics from The University of Hong Kong, Hong Kong, in 1989, and the M.Sc. degree in information technology and the Ph.D. degree in computer science from The Hong Kong Polytechnic University (HKPolyU), Hong Kong, in 1997 and 2000, respectively. He is the Founder of Quantum Finance Forecast Center (QFFC) with more than 25 years IT consultancy and Research and Development experiences in AI, chaotic neural networks, intelligent fintech systems, and quantum finance. He is currently an Associate Professor with the Beijing Normal University–Hong Kong Baptist University United International College. His research interests include quantum finance, quantum anharmonic oscillators, chaotic neural oscillators, fuzzy-neuro financial systems, chaotic neural networks, and severe weather modeling and prediction.