



Category: STEM (Science, Technology, Engineering and Mathematics)

ORIGINAL

eHyPRETo: Enhanced Hybrid Pre-Trained and Transfer Learning-based Contextual Relation Classification Model

eHyPRETo: modelo híbrido mejorado de clasificación de relaciones contextuales basado en aprendizaje previo y transferencia

Jeyakodi G¹  , Sarojini B²  , Shanthi Bala P¹  

¹Department of Computer Science, Pondicherry University. Puducherry, India.

²Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education For Women. Coimbatore, India.

Cite as: JG, BS, Bala PS. eHyPRETo: Enhanced Hybrid Pre-Trained and Transfer Learning-based Contextual Relation Classification Model. Salud, Ciencia y Tecnología - Serie de Conferencias. 2024; 3:758. <https://doi.org/10.56294/sctconf2024758>

Submitted: 28-12-2023

Revised: 19-03-2024

Accepted: 11-05-2024

Published: 12-05-2024

Editor: Dr. William Castillo-González 

ABSTRACT

Introduction: relation classification (RC) plays a crucial role in enhancing the understanding of intricate relationships, as it helps with many NLP (Natural Language Processing) applications. To identify contextual subtleties in different domains, one might make use of pre-trained models.

Methods: to achieve successful relation classification, a recommended model called eHyPRETo, which is a hybrid pre-trained model, has to be used. The system comprises several components, including ELECTRA, RoBERTa, and Bi-LSTM. The integration of pre-trained models enabled the utilisation of Transfer Learning (TL) to acquire contextual information and complex patterns. Therefore, the amalgamation of pre-trained models has great importance. The major purpose of this related classification is to effectively handle irregular input and improve the overall efficiency of pre-trained models. The analysis of eHyPRETo involves the use of a carefully annotated biological dataset focused on Indian Mosquito Vector Biocontrol Agents.

Results: the eHyPRETo model has remarkable stability and effectiveness in categorising, as evidenced by its continuously high accuracy of 98,73 % achieved during training and evaluation throughout several epochs. The eHyPRETo model's domain applicability was assessed. The obtained p-value of 0,06 indicates that the model is successful and adaptable across many domains.

Conclusion: the suggested hybrid technique has great promise for practical applications such as medical diagnosis, financial fraud detection, climate change analysis, targeted marketing campaigns, and self-driving automobile navigation, among others. The eHyPRETo model has been developed in response to the challenges in RC, representing a significant advancement in the fields of linguistics and artificial intelligence.

Keywords: Pre-trained Model; RC (Relation Classification); Contextual; ELECTRA; RoBERTa; Domain Adaptable; eHyPRETo.

RESUMEN

Introducción: la clasificación de relaciones (RC) juega un papel crucial en la mejora de la comprensión de relaciones intrincadas, ya que ayuda con muchos PNL (Lenguaje Natural Procesamiento) solicitudes. Para identificar sutilezas contextuales en diferentes dominios, se podría hacer uso de modelos previamente entrenados.

Métodos: para lograr una clasificación de relaciones exitosa, se recomienda un modelo llamado Se debe utilizar eHyPRETo, que es un modelo híbrido previamente entrenado. El sistema comprende varios componentes, incluidos ELECTRA, RoBERTa y Bi-LSTM. La integración de los modelos pre-entrenados permitieron la utilización de Transfer Learning (TL) para adquirir información contextual. Información y patrones complejos.

Por lo tanto, la fusión de modelos previamente entrenados ha gran importancia. El objetivo principal de esta clasificación relacionada es manejar eficazmente entradas irregulares y mejorar la eficiencia general de los modelos previamente entrenados. El análisis de deeHyPRETo implica el uso de un conjunto de datos biológicos cuidadosamente anotados centrados en la India. Agentes de biocontrol de mosquitos vectores.

Resultados: el modelo eHyPRETo tiene una estabilidad y efectividad notables en la categorización, como evidenciado por su alta precisión continua del 98,73 % lograda durante el entrenamiento y evaluación a lo largo de varias épocas. La aplicabilidad de dominio del modelo eHyPRETo fue juzgado. El valor p obtenido de 0,06 indica que el modelo es exitoso y adaptable. en muchos dominios.

Conclusión: la técnica híbrida sugerida es muy prometedora para aplicaciones prácticas como diagnóstico médico, detección de fraude financiero, análisis del cambio climático, focalización campañas de marketing y navegación en vehículos autónomos, entre otros. El eHyPRETo El modelo se ha desarrollado en respuesta a los desafíos en RC, representando una importante avances en los campos de la lingüística y la inteligencia artificial.

Palabras clave: Modelo Preentrenado; RC (Clasificación de Relaciones); Contextual; ELECTRA; RoBERTa; Dominio Adaptable; eHyPRETo.

INTRODUCTION

RC facilitates the extraction of correct question answers, structured data, and sentiment interpretation from unstructured text. Information retrieval, knowledge graph construction, sentiment analysis, named entity identification, question-answering systems, and other applications depend on accurate and efficient relation categorization. The integration of transfer learning-based techniques, the advent of hybrid approaches, and the ongoing development of pre-trained models have all contributed to significant advancements in relation categorization. The relation classification enhancements are considered to be the overall objective of those developments.

Actively evolving BERT models, (RoBERTa) (Y. Liu et al., 2019), ELECTRA (Efficiently Learning an Encoder that Classifies Token Replacements Accurately) (Clark et al., 2020), and Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2019) were the pre-trained models for enhancing the NLP systems significantly, and also revolutionizing the classification efficiency. Additional data and optimisation approaches are utilized by RoBERTa throughout FT (Fine-Tuning) for the classification efficiency optimization. Then, ELECTRA provides alternative applicable synonyms in the particular tokens corpus in the input text, thereby enhancing the efficiency of the pre-trained procedure also maintaining the high accuracy. The other PTMs, considerably subsidized for the relation classification: DistilBERT (Sanh et al., 2019), A Lite BERT (ALBERT) (Lan et al., 2020), eXtreme Learning with Transformers (XLNet) (Yang et al., 2019), Enhanced Representation through knowledge Integration (ERNIE) (Zhang et al., 2020), and TAPAS (Tabular Pre-trained Language Model) (Herzig et al., 2020). For the relation classification, numerous pre-trained models are employed and it is presented in figure 1. Through the utilization of large-scale datasets, rich contextual embeddings are obtained for attaining textual semantic meanings. These pre-trained models are trained on several datasets can be obtained by FT such models. Researchers have modified them for the relation classification tasks thereby attaining superior outcomes.

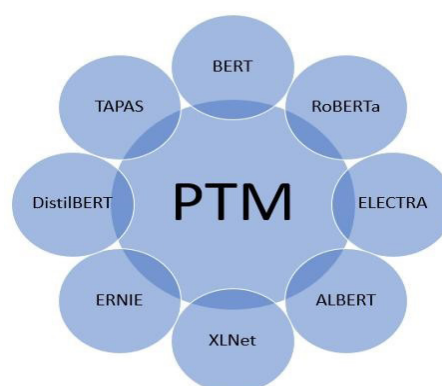


Figure 1. Pre-trained Models

Combination of several NN (Neural Networks) models by the hybrid models for enhancing the RC accuracy. Researches have suggested numerous hybrid models as it integrates Convolutional Neural Networks (CNN) (Hui et al., 2020), Long Short-Term Memory (LSTM), Bidirectional Long Short-Term Memory (Bi-LSTM) (Harnoune

et al., 2021), Bidirectional Gated Recurrent Unit (Bi-GRU) (Yi & Hu, 2019), etc. To create a Personality Trait Classification model from text, the advantages of CNN and LSTM have been combined (Ahmad et al., 2021) for capturing contextual data. The hybrid pre-trained models based on TL make use of pre-trained and FT data for the enhancement of the accuracy of classification. Employing generalised language understanding, models like SpanBERT (Joshi et al., 2020) and ERNIE (Zhang et al., 2020) provides the notable accuracy in RC.

Upgraded Hybrid PRE-trained and TL-RC Model, associates the advantages of ELECTRA, RoBERTa, and Bi-LSTM was suggested in this research. The pre-training phase of ELECTRA's token replacement and the TL phase of RoBERTa's FT are united, and Bi-LSTM is utilized for enhancing the contextual information capture, sequential dependency management, and RC model's accuracy.

Related work

PTM took advantage through the substantial attention that classes have received through the course of time. Figure 2 illustrates the techniques used for relation classification over time. Machine learning algorithms such as Support Vector Machines (SVMs), Random Forest, Logistic Regression, Gradient Boosting, and Hidden Markov Models (HMMs), etc. were used for developing the relation classification model (Formisano et al., 2008). For continuous improvement, deep learning neural networks such as CNN, Recurrent Neural Networks (RNN), Transformerbased models, etc. were used to learn the features automatically (Itamar Arel, Derek C. Rose, 2010) (He et al., 2010). These methods, however, often struggled with capturing complex linguistic patterns and generalizing them across various domains.

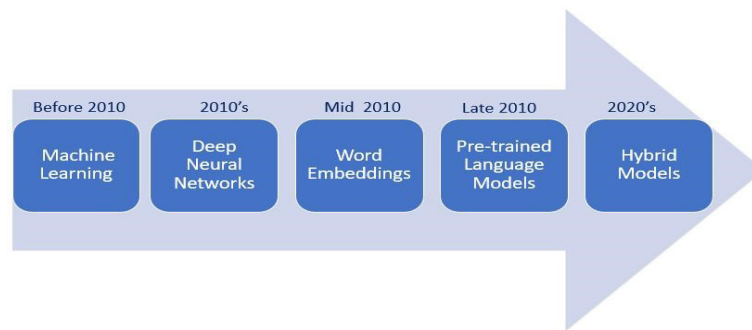


Figure 2. Relation Classification Techniques

To address this limitation, the word embeddings such as Word2Vec, GloVe, FastText, and ELMo, etc, and syntactic features such as Dependency Parse Trees, Part-of-Speech (POS) tags were explored to enhance classification accuracy through semantic understanding and contextual information (Xu et al., 2015)(Agarwal et al., 2015). Furthermore, the utilization of word embeddings and syntactic features helps to make significant improvements in accuracy and the ability to handle diverse datasets. However, recent advancements in the field have witnessed a paradigm shift towards the utilization of large-scale transformer-based PTM, such as GPT, BERT, RoBERTa, ELECTRA, MarianMT and so on (S. Yu et al., 2019) (Du et al., 2022). Their outstanding capacity to use contextual information made them the preferred choices for RC.

An evident focus on multitask learning and domain adaptation named innovative TL approaches are presented in this review, and it is considered as critical to enhance the model performance and extending RC abilities. Also, novel trainings indicate the significance. Including pretrained models in technical NN layers because they give greater outcomes and leads to the development of more consistent and accurate hybrid models for diverse applications (Aydođan & Karci, 2020). Many transfer learning hybrid models, including BERT-BiLSTM, BERT-BiLSTM-CRF, BERT-BiLSTM-LSTM, BERT-BiGRU, BERT-BiGRU-CRF, RoBERTa-BiLSTM-CRF, etc., have been developed for a number of applications and have improved relation classification performance (Q. Yu et al., 2021)(Qin et al., 2021). These models combine pre-trained models and deep neural networks. The hybrid pre-trained models for text categorization that researchers built using transfer learning are listed in table 1. In order to handle noisy words and increase the network's resilience, Tianyi et al. employed transfer learning to initialise the network with prior information gained via entity categorization (T. Liu et al., 2018).

The eHyPRETo, the integration of ELECTRA, RoBERTa, and Bi-LSTM is proposed to address the limitation of existing hybrid models in capturing fine-grained contextual information and complex sequential dependencies, especially in the context and specialized relation classification tasks. Section 3 explores the proposed model eHyPRETo, detailing its working architecture. Section 4 describes the results and discussion, and Section 5 provides the conclusion.

Table 1. Transfer Learning based Hybrid Pre-trained Models			
Sl.No	Reference	Dataset	Hybrid Models
1	Gou, 2023	MELD	BERT-BiLSTM-CNN BERT-BiGRU-CNN BERT-BiLSTM-attention BERT-(BiLSTM + CNN) BERT-(BiLSTM + attention + CNN) BiBERT-BiLSTM
2	Q. Yu et al., 2021	THUCNews	BERT-BiGRU ELMo-BiGRU BERT-CNN BERT-RNN
3	Qin et al., 2021	Electronic Medical Record	BERT-BiGRU-CRF BERT-BiLSTM-CRF
4	Tan et al., 2022	IMDb, Sentiment140 and Twitter US Airline Sentiment Dataset	RoBERTa-LSTM
5	Q. Xu et al., 2023	Chinese Online DoctorPatient Q&A Platform	RoBERTa-BiLSTM-CRF RoBERTa-CRF RoBERTa-Softmax
6	Zhou et al., 2021, Sun et al., 2022	DocRED	BERT-ATLOP RoBERTA-ATLOP BERT-DHGCN TDGAT-BERT TDGAT-ROBERTa
7	Geng et al., 2020, Zhao et al., 2019	SemEval2010	BLSTM+BiLSTM+Att Bi-GRU+MCNN-Att + Sum BERT- Graph-based Neural Network

METHODOLOGY

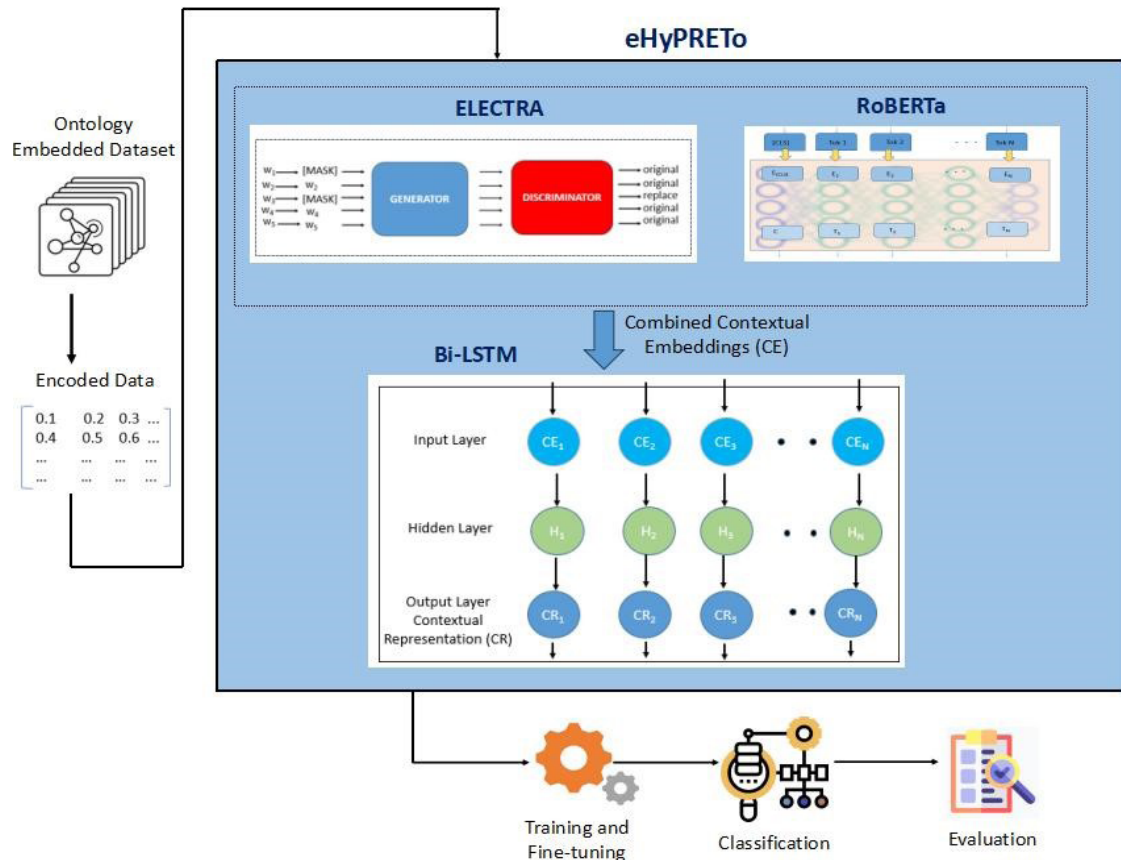


Figure 3. eHyPRETo Architecture

The proposed eHyPRETo suggests a novel approach that synergizes their unique capabilities, resulting in understanding and classifying relationships in text across various contexts and improving accuracy. The eHyPRETo consists of various components such as ontology embedded dataset for training and testing, data encoding for embedded vector transformation, eHyPRETo architecture for relation classification model building, Training and Fine-tuning to optimize parameters, Classification for categorizing relations, and Evaluation for the model's performance assessment. Figure 3 illustrates the eHyPRETo model components.

Ontology Embedded Dataset

The input dataset comprises 632 manually annotated sentences related to mosquito vectors and biocontrol agents which are collected from research articles in PubMed database. During annotation, six distinct relation labels were assigned to capture different relations between entities. The domain-specific concepts such as mosquito name, and biocontrol agents are embedded using MVAOnt, Mosquito Vector Association Ontology¹. The entity pairs $\langle e1 \rangle \langle /e1 \rangle$ and $\langle e2 \rangle \langle /e2 \rangle$ are used for incorporating concepts. The dataset is partitioned into subsets such as for training, validation, and testing to ensure a balanced distribution of relation labels for eHyPRETo evaluation.

Data Encoding

An essential step in getting input data ready for the eHyPRETo model is data encoding. It entails converting input phrases that are ontology embedded into numerical embedded vectors useful for model processing. During this procedure, the model's embeddings are used to translate tokenized sentences where each token represents a word or subword to numerical representations. From the sentence "The mosquito is a vector," which is encoded into ["The", "mosquito", "is", "a", and "vector"], a vector with numerical values, such as [12, 345, 45, 7, 89], is captured. Each number is an index of tokens. The attention masks make it easier to focus on significant concepts and ignore unnecessary ones. The encoded vectors and attention mask together make up the model's input, permitting for managing information for RC effectively.

eHyPRETo Architecture

eHyPRETo model integrates the advantages of RoBERTa, BiLSTM, and ELECTRA for incrementing RC performance and contextual knowledge. While RoBERTa supports in comprehending the general semantics of phrases, ELECTRA assists in understanding the way of words that fit together. This special combination makes use of RoBERTa embeddings to capture the wider context, ELECTRA embeddings for FT, and Bi-LSTM for capturing the bidirectional context. By integrating ELECTRA, RoBERTa, and Bi-LSTM is a crucial enhancement that greatly progresses the model's contextual comprehension of ideas and RC.

Table 2. eHyPRETo Specifications

SI.No	Specification	Details
1	Model Architecture	eHyPRETo
2	Pre-trained Models	ELECTRA, RoBERTa
3	Additional Layers	Bi-LSTM
4	ELECTRA Hidden Layers	12
5	RoBERTa Hidden Layers	12
6	Attention Heads	12
7	LSTM Hidden Dimension	128
8	Batch Size	32
9	Learning Rate	1e-5
10	Dropout Rate	0,25
11	Number of Epochs	10
12	Input Representation	Tokenized Encoded Sentences
13	Training Data	Manually Annotated Dataset
14	Relation Labels	Six Distinct Labels for Six Relations

1 <https://bioportal.bioontology.org/ontologies/MVAONT>

In order to capture bidirectional context, the eHyPRETo Bi-LSTM layer includes an input layer that receives tokenized and contextual embeddings from ELECTRA and RoBERTa. Setting input and output dimensions, configuring hidden units, and taking advantage of the LSTM's bidirectional nature to process input sequences in a bidirectional manner are all part of the layered design. Long-range dependencies are captured, and contextual awareness is improved. More RC accuracy is possible with the eHyPRETo model thanks to dynamic contextual awareness. The combined contextual embedding vectors of ELECTRA, RoBERTa, and Bi-LSTM are the outcomes of the eHyPRETo. This technique improves knowledge of the way of ideas are related to one another, which in turn progresses the model's RC ability. The comprehensive eHyPRETo requirements for RC tasks are presented in table 2.

Training and Fine-tuning

In the Training and Fine-tuning, eHyPRETo iteratively refines parameters related to the weights and biases of its neural network layers for optimization. This optimization involves minimizing the cross-entropy loss function by Adam optimizer enhancing the model's capacity to learn and generalize from training data. Consequently this process enhances complex relationship capturing and ensuring better performance in relation classification.

Classification

The classification layer processes the contextual embeddings obtained from the eHyPRETo model. It takes the encoded tokenized sentences, which have been enriched with ELECTRA, RoBERTa, and Bi-LSTM, and uses rich contextual information to classify the type of relationship between the concepts in the input sentence more accurately. The classification layer outputs a set of probabilities, indicating model's confidence in each possible relation label, thereby facilitating accurate relation classification within the given text. The high probability is the appropriate classified relation. For example, if the probability values generated for “<e1>Pinus kesiya</e1> leaf essential oil is against <e2>Aedes aegypti</e2>” for six relations is [0,730, 0,135, 0,025, 0,020, 0,045, 0,45], then the relation label corresponding to the highest probability is the classified output.

Evaluation

In the evaluation phase, the eHyPRETo model's predictions are compared with the test dataset. It assesses the performance by statistical measurers such as accuracy, precision, recall, F1-score, and Cohen's Kappa. Evaluation shows how well the model is performing in terms of relation classification. After evaluation, the numerical predictions or probabilities generated by the eHyPRETo model for the raw data are transformed into human-readable and semantically meaningful relation labels. This step connects model's output with real-world understanding by mapping numerical representations to linguistic expressions, such as "is_a," "biocontrol", "cause_effect", etc. The resulting structured output provides a clear representation of the relationships between concepts within the input text to facilitate tasks like knowledge extraction, information retrieval, and decision support based on the context.

The next section delivers the comprehensive results and discussion of eHyPRETo relation classification performance in mosquito vectors and biocontrol agents context.

RESULTS AND DISCUSSION

The eHyPRETo performs remarkably in relation classification by enhancing contextual understanding and capturing complex relationships. The eHyPRETo model is developed using Python. The eHyPRETo model is assessed using a manually annotated mosquito vectors and biocontrol agents dataset, categorized into six distinct relation labels. This model exhibits enhanced performance in effectively classifying and distinguishing between these relation labels, particularly in the context of the unbalanced dataset. Table 3 shows the eHyPRETo performance over 10 epochs. Even after 10 epochs, there has been no noticeable improvement in accuracy. Figure 4 illustrates the graph depicting epoch and accuracy.

Epoch	Accuracy (%)	Precision	Recall	F-Score	Cohens Kappa
Epoch 1	92,86	0,9201	0,9286	0,9218	0,8947
Epoch 2	89,68	0,9033	0,8968	0,8932	0,8507
Epoch 3	96,03	0,9459	0,9603	0,9528	0,9398
Epoch 4	97,62	0,9773	0,9762	0,975	0,9642
Epoch 5	97,62	0,9773	0,9762	0,975	0,964
Epoch 6	98,73	0,9861	0,9890	0,9852	0,9994
Epoch 7	98,73	0,9861	0,9890	0,9852	0,9994
Epoch 8	98,73	0,9861	0,9890	0,9852	0,9994
Epoch 9	98,73	0,9861	0,9890	0,9852	0,9994
Epoch 10	98,73	0,9861	0,9890	0,9852	0,9994

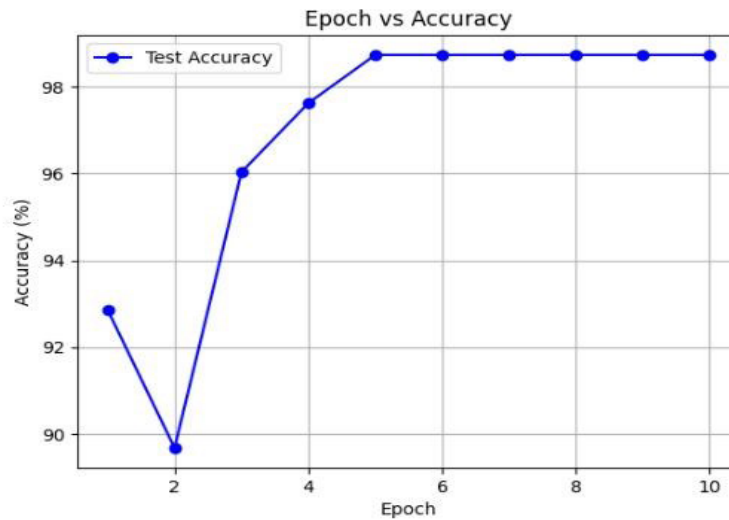


Figure 4. Epoch Vs Accuracy

The eHyPREto model’s performance across ten epochs, consistently demonstrates high accuracy and precision, indicating the model's effectiveness in making accurate predictions. The accuracy increases from 92,86 % in the first epoch to 98,73 % in subsequent epochs, suggesting the model's learning and adaptation over time. Precision, Recall, and F1score also maintain high values, with minimal variations across epochs, emphasizing the model's strength and stability. Cohen's Kappa score demonstrates strong agreement between predicted and actual values, with values ranging from 0,8507 to 0,9761 across the ten epochs. The constant performance across each epoch is a positive reflection of the eHyPREto model’s stability.

Further, the eHyPREto is assessed with traditional machine learning models and PTM. Table 4 shows the eHyPREto comparison with BERT and its variants. eHyPREto outperforms all the models, achieve high values for each metrics. Figure 5 illustrates the corresponding comparison chart. The outstanding performance of eHyPREto suggests the model effectiveness in accurately classifying complex relationships in the unbalanced dataset. After assessing and refining the model's performance, the final classified relation in terms of probabilities is mapped into human-readable relation labels for understanding as shown in figure 6.

Sl. No	Model	Precision	Recall	Accuracy(%)	F-score	Cohen’s Kappa
1	BERT	0,962	0,965	96,50	0,963	0,951
2	BioBERT	0,958	0,956	95,64	0,954	0,949
3	ALBERT	0,942	0,94	93,97	0,938	0,914
4	RoBERTa	0,795	0,824	90,48	0,803	0,706
5	ELECTRA	0,961	0,964	96,43	0,962	0,927
6	BioELECTRA	0,955	0,956	95,874	0,955	0,941
7	eHyPREto	0,986	0,989	98,73	0,985	0,999

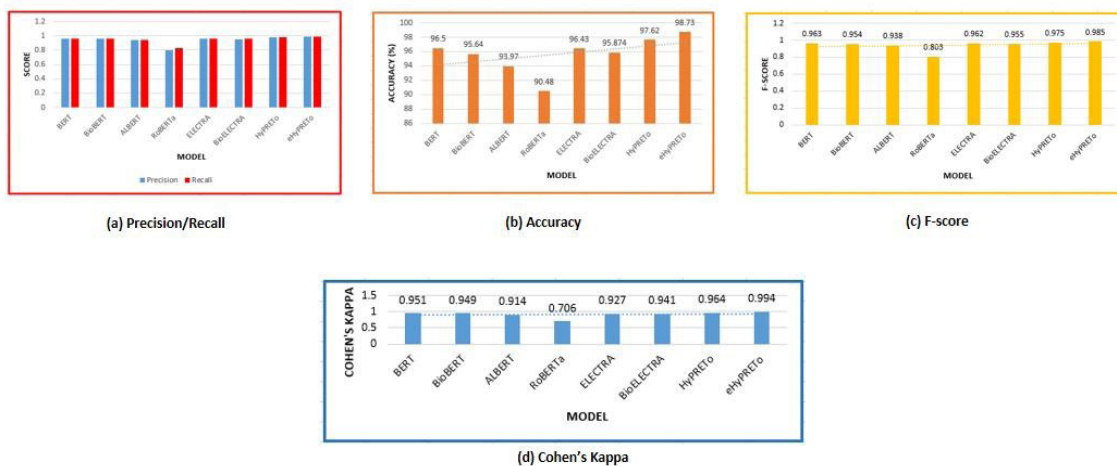


Figure 5. eHyPREtoComparison

```

get_reply("cassia occidentalis eliminate mosquito larvae")

'lifespans'

Code for get-reply Function

def get_reply(msg):
    model.eval()
    input_msg, _ = prepare_features(msg)
    if torch.cuda.is_available():
        input_msg = input_msg.to(device)
    output = model(input_msg)[0].to(device)
    _, pred_label = torch.max(output.data, 1)
    prediction=list(label_to_ix.keys())[pred_label]
    return prediction

Relation Labels

label_to_ix.keys()

dict_keys(['is', 'biocontrol', 'lifespans', 'predator', 'cause_effect', 'repellent'])
    
```

Figure 6. Illustration of Model’s Classified Relation

Finally, the eHyPRETo's performance is examined across different domains by Mosquito Vector Biocontrol Agents as the source domain and SemEval 2010 Task 8 dataset as the target domain. The combined data is used for training, and the target data is used for testing and evaluating model's performance. The p-value (0,06) that is obtained by t-test indicates the model's ability to adapt in diverse domains, exposing its effectiveness beyond specific contexts. Figure 7 visualizes the domain adaptability performance of eHyPRETo model in terms of accuracy.

```

Source Metrics: [0.944, 0.921, 0.912, 0.918, 0.946]
Target Metrics: [0.942, 0.919, 0.911, 0.917, 0.94]
P-value: 0.060812856154522504
No significant difference in performance 0.060812856154522504
Model is domain adaptable and generalised
    
```

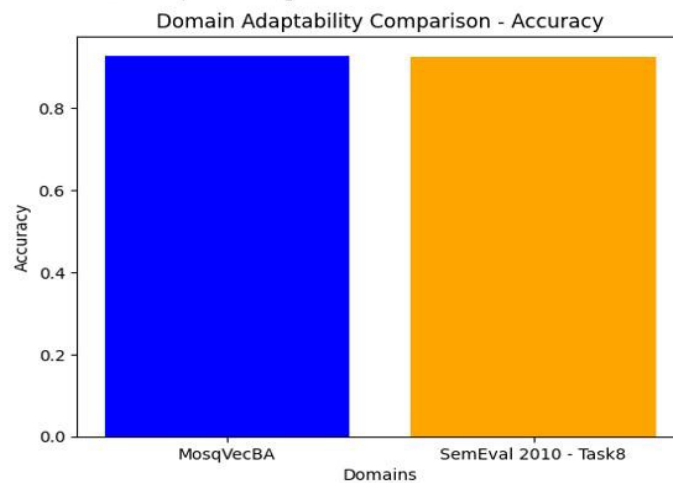


Figure 7. eHyPRETo’s Domain Adaptability Visualization

The eHyPRETo model's performance highlights its substantial potential in decisionmaking, knowledge extraction, question-answering systems, semantic search engines, and more applications related to mosquito vectors and biocontrol agents. The model has demonstrated effectiveness in the specific domain of mosquito vectors and biocontrol agents.

CONCLUSION

In conclusion, relation classification in the field of mosquito vector biocontrol agents was successfully carried out using the eHyPRETo, a unique hybrid pre-trained model combining ELECTRA, RoBERTa, and Bi-LSTM. Through precise classification of interactions among different entities in this setting, the model shows promise for greatly improving our knowledge of vector-borne illnesses and biocontrol agents. Using a variety of metrics, the eHyPRETo model demonstrates the significance of relation categorization. The significance test also demonstrates the domain applicability of the model. The created model can be useful in a number of fields, including agriculture, healthcare, the military, the air force, and more. Overall, the combination of

PTM and neural networks in the eHyPRETo model signifies a potential development in the field of relation categorization. Subsequent investigations will focus on integrating contextual elements like syntactic data and knowledge graph embeddings to improve the model's contextual comprehension.

Dataset and Code: the dataset and code will be provided based on the requirement.

REFERENCES

1. Agarwal B, Poria S, Mittal N, Gelbukh A, and Hussain A. Concept-level sentiment analysis with dependency-based semantic parsing: a novel approach. *Cognitive Computation*, 7, pp. 487-499. <https://doi.org/10.1007/s12559-014-9316-6v>.
2. Ahmad H, Asghar MU, Asghar MZ, Khan A, and Mosavi, AH. A hybrid deep learning technique for personality trait classification from text. *IEEEAccess*, 9, pp. 146214-146232. <https://doi.org/10.1109/ACCESS.2021.3121791>.
3. Aydođan M, and Karci A. Improving the accuracy using pre-trained word embeddings on deep neural networks for Turkish text classification. *Physica A: Statistical Mechanics and its Applications*, 541, p.123288. <https://doi.org/10.1016/j.physa.2019.123288>.
4. Clark K, Luong MT, Le QV, and Manning, CD. Electra: Pre-training text encoders as discriminators rather than generators. *arXiv preprint arXiv:2003.10555*, pp. 1-18. <https://doi.org/10.48550/arXiv.2003.10555>.
5. Devlin J, Chang MW, Lee K, and Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, pp. 1-16. <https://doi.org/10.48550/arXiv.1810.04805>.
6. Du M, He F, Zou N, Tao D, and Hu X. Shortcut learning of large language models in natural language understanding. *Communications of the ACM*, 67(1), pp. 110-120. <https://doi.org/10.1145/3596490>.
7. Formisano E, De Martino F, and Valente G. Multivariate analysis of fMRI time series: classification and regression of brain responses using machine learning. *Magnetic resonance imaging*, 26(7), pp. 921-934. <https://doi.org/10.1016/j.mri.2008.01.052>.
8. Geng Z, Chen G, Han Y, Lu G, and Li F. Semantic relation extraction using sequential and tree-structured LSTM with attention. *Information Sciences*, 509, pp. 183-192. <https://doi.org/10.1016/j.ins.2019.09.006>.
9. Gou Z, and Li Y. Integrating BERT embeddings and BiLSTM for emotion analysis of dialogue. *Computational Intelligence and Neuroscience*, pp. 1-8. <https://doi.org/10.1155/2023/6618452>.
10. Harnoune A, Rhanoui M, Mikram M, Yousfi S, Elkaimbillah Z, and El Asri B. BERT based clinical knowledge extraction for biomedical knowledge graph construction and analysis. *Computer Methods and Programs in Biomedicine Update*, 1, pp. 100042. <https://doi.org/10.1016/j.cmpbup.2021.100042>.
11. He X, Chen W, Nie B, and Zhang M. Classification technique for danger classes of coal and gas outburst in deep coal mines. *Safety science*, 48(2), pp. 173-178. <https://doi.org/10.1016/j.ssci.2009.07.007>.
12. Herzig J, Nowak PK, Müller T, Piccinno F, and Eisenschlos, JM. TaPas: Weakly supervised table parsing via pre-training. *arXiv preprint arXiv:2004.02349*, pp. 4320 - 4333. <https://doi.org/10.18653/v1/2020.acl-main.398>.
13. Hui B, Liu L, Chen J, Zhou X, and Nian Y. Few-shot relation classification by context attention-based prototypical networks with BERT. *EURASIP Journal on Wireless Communications and Networking*, pp. 1-17. <https://doi.org/10.1186/s13638-020-01720-6>.
14. Arel I, Rose DC, and Karnowski TP. Deep machine learning-a new frontier in artificial intelligence research [research frontier]. *IEEE computational intelligence magazine*, 5(4), pp. 13-18. <https://doi.org/10.1109/MCI.2010.938364>.
15. Joshi M, Chen D, Liu Y, Weld DS, Zettlemoyer L, and Levy O. Spanbert: Improving pre-training by representing and predicting spans. *Transactions of the association for computational linguistics*, 8, pp. 64-77. https://doi.org/10.1162/tacl_a_00300.

16. Lan Z, Chen M, Goodman S, Gimpel K, Sharma P, and Soricut R. Albert: A lite bert for self-supervised learning of language representations. arXiv preprint arXiv:1909.11942, pp. 1-17. <https://doi.org/10.48550/arXiv.1909.11942>.
17. Liu T, Zhang X, Zhou W, and Jia W. Neural relation extraction via inner-sentence noise reduction and transfer learning. arXiv preprint arXiv:1808.06738, pp. 2195- 2204. <https://doi.org/10.18653/v1/d18-1243>.
18. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, Levy O, Lewis M, Zettlemoyer L, and Stoyanov V. Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692, pp. 1-13. <https://doi.org/10.48550/arXiv.1907.11692>.
19. Mehta SB, Chaudhury S, Bhattacharyya A, and Jena A. Handcrafted fuzzy rules for tissue classification. *Magnetic Resonance Imaging*, 26(6), pp. 815-823. <https://doi.org/10.1016/j.mri.2008.01.021>.
20. Qin Q, Zhao S, and Liu C. A BERT-BiGRU-CRF model for entity recognition of Chinese electronic medical records. *Complexity*, pp. 1-11. <https://doi.org/10.1155/2021/6631837>.
21. Sanh V, Debut L, Chaumond J, and Wolf T. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv preprint arXiv:1910.01108, pp. 1-5. <https://doi.org/10.48550/arXiv.1910.01108>.
22. Sun Q, Xu T, Zhang K, Huang K, Lv L, Li X, Zhang T, and Dore-Natth D. Dual-channel and hierarchical graph convolutional networks for document-level relation extraction. *Expert Systems with Applications*, 205, pp. 117678. <https://doi.org/10.1016/j.eswa.2022.117678>.
23. Sun Q, Zhang K, Huang K, Xu T, Li X, and Liu Y. Document-level relation extraction with two-stage dynamic graph attention networks. *Knowledge-Based Systems*, 267, pp. 110428. <https://doi.org/10.1016/j.knosys.2023.110428>.
24. Tan KL, Lee CP, Anbananthen KSM, and Lim KM. RoBERTa-LSTM: a hybrid model for sentiment analysis with transformer and recurrent neural network. *IEEE Access*, 10, pp. 21517-21525. <https://doi.org/10.1109/ACCESS.2022.3152828>.
25. Xu R, Chen T, Xia Y, Lu Q, Liu B, and Wang X. Word embedding composition for data imbalances in sentiment and emotion classification. *Cognitive Computation*, 7, pp. 226-240. <https://doi.org/10.1007/s12559-015-9319-y>.
26. Yang Z, Dai Z, Yang Y, Carbonell J, Salakhutdinov RR, and Le QV. Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems*, 32, pp. 1-11.
27. Yi R, and Hu W. Pre-trained BERT-GRU model for relation extraction. In *Proceedings of the 8th international conference on computing and pattern recognition*, pp. 453-457. <https://doi.org/10.1145/3373509.3373533>.
28. Yu Q, Wang Z, and Jiang K. Research on text classification based on bert-bigru model. In *Journal of Physics: Conference Series*, 1746(1), pp. 012019). IOP Publishing. <https://doi.org/10.1088/17426596/1746/1/012019>.
29. Yu, S., Su, J. and Luo, D., 2019. Improving bert-based text classification with auxiliary sentence and domain knowledge. *IEEE Access*, 7, pp.176600-176612. <https://doi.org/10.1109/ACCESS.2019.2953990>.
30. Zhang Z, Han X, Liu Z, Jiang X, Sun M, and Liu Q. ERNIE: Enhanced language representation with informative entities. arXiv preprint arXiv:1905.07129, pp. 1-11, <https://doi.org/10.18653/v1/p19-1139>.
31. Zhao Y, Wan H, Gao J, and Lin Y. Improving relation classification by entity pair graph. In *Asian Conference on Machine Learning*, pp. 1156-1171.
32. Zhou W, Huang K, Ma T, and Huang J. Document-level relation extraction with adaptive thresholding and localized context pooling. In *Proceedings of the AAAI conference on artificial intelligence*, 35(16), pp. 14612-14620. <https://doi.org/10.1609/aaai.v35i16.17717>.

FINANCING

"The authors did not receive financing for the development of this research".

CONFLICT OF INTEREST

"The authors declare that there is no conflict of interest".

AUTHORSHIP CONTRIBUTION

Conceptualization: Jeyakodi G.

Data curation: Sarojini B.

Formal analysis: Sarojini B.

Research: Shanthi Bala P.

Methodology: Jeyakodi G.

Drafting - original draft: Shanthi Bala P.

Writing - proofreading and editing: Jeyakodi G.