A Named Object Recognition Model of Health Records Using Improved Cuckoo Search Optimization Algorithm and Kernel Extreme Learning Machine

[a]R. Ramachandran, [b]Dr. K. Arutchelvan

[a]Assistant Professor, Department of Computer and Information Science, Annamalai University, Chidambaram.

[b]Assistant Professor, Department of Computer and Information Science, Annamalai University, Chidambaram.

E-mail:

rr15597@annamalaiuniversity.ac.in,ka05890@annamalaiuniversity.ac.inOrcid:https://orcid.org/0000-0003-4307-3800https://orcid.org/0000-0001-7233-3566

Abstract

7906

Named Object Recognition (NER) has become an important area in several natural language processing technologies such as information extraction and information retrieval. The benefits of NER have attracted more attention among researchers in diverse fields. But the class labels and extension of named entities considerably differ with respect to distinct application domains. Medical literature includes valuable details namely clinical signs, diagnosis, drug, and medication for particular diseases. As the knowledge gaining from health literature is a tedious task, this paper aims to develop a new NER model in medical literature using Improved Cuckoo search algorithm (ICSA) with a Kernel Extreme Learning Machine (KELM), called ICSA-KELM. The presented ICSA-KELM model involves three major stages namely preprocessing, classification, and parameter tuning. The presented model performs preprocessing to convert the raw medical data into a useful format. In addition, the KELM model is executed to perform a classification process. Finally, the two main parameters of KELM such as penalty parameter 'C' and kernel bandwidth 'γ' of the KELM are computed using the ICSA algorithm. An general set of recreations were carried out to determine the effectual classification performance of the ICSA-KELM model. The resultant experimental values ensured the betterment of the ICSA-KELM model over the existing methods.

Keywords: Machine learning, Named entity recognition, Parameter tuning, Medical literature, KELM model

DOI Number:10.14704/nq.2022.20.8.NQ44816 **NeuroQuantology 2022; 20(8): 7906-7914**

1. Introduction

Electronic Medical Record (EMR) [1] is normally used to keep the scientific statistics concerning the affected person in textual format. It brought about creating a top notch healthcare machine in which collective info can be stored in a unmarried database. In order to construct a popular EMR machine, distinctive fashions are crucial and a number of them are: global degree modules like Part of Speech (POS), Named Entity Recognition (NER), even as sentence degree levels are Dependency Parsing in addition to Semantic Role Labeling, and record degree modules are Classification and Summarization. In general, every module calls for awesome techniques. In the case of EMR summarization, EMR is compiled into 2 levels namely, Extractive in addition to Abstractive summaries. Some of the applicable procedures are Clinic Viewer and IHC Patient Worksheet. In the case of record class, records retrieved from EMR is implemented for diagnosing the coronary heart sickness and suicide threat stratum with the assist of Deep Learning (DL) techniques like Deep Patient, Doctor synthetic intelligence (AI), and EMR pushed Nonnegative Restricted Boltzmann Machines (NRBM). Especially, random records in EMR defines the kingdom of patients` fitness and info like remedy duration, call of the sickness, severity rate, and so on. Which is useful for the physicians to recognize the fitness circumstance and take a look at the virtual statistics regularly. Therefore, extracting records from EMR performs an critical function in medical applications. Therefore, obtaining medical topics is a hard and time-ingesting mission. In addition, the utility of the modern-day EMR era to the evaluation of scientific items is taken into consideration a complicated method as EMRs are amassed swiftly and aren't appropriate for pre-processing. Also, wrong syntax, most abbreviations, and devices after mathematical values have a tendency to create complicated medical tasks. Natural language processing (NLP) gadgets aren't green whilst utilized in EMR due to the fact NLP entity regulations aren't applicable. As a result, there may be a right away requirement to carry out Entity Recognition (ER) from EMR info. NER is described as a mission of evaluation and class of phrases or terms with particular semantics or meanings in a sentence. Biomedical applications, it consists of medical norms together with proteins, genomes, styles of diseases, etc. In the case of legislative applications, NAE is composed It is from unsigned phrase evaluation and it is able to be a important part of records extraction and retrieval, Q & A, in addition to change herbal language processing (NLP) fashions [2]. The predominant goal of NER in NLP has promoted severa builders of library records and laptop technological know-how and led them to recommend huge technologies. Therefore, the elegance and extensions of Named Entities (NAE) range from one another. Specifically, the NAE consists of the affected person's call, area, and time. In the case of prison phrases and provisions. In practice, the significance of NERs varies through area of look at and it's far very hard to migrate. In addition, the dearth of area-particular NERs and area entities together with law, historic Chinese poetry, etc. suggests that records for direct studying is constrained. Meantime, due to better records specialization, area expertise need to be enriched even as annotating the texts manually. Therefore, properly skilled specialists and huge burdens require huge personnel and resources. Hence, the use of a constrained quantity of annotated records schooling efficient NER is the important thing goal of NER studies. NER is taken into consideration to be an critical and complex studying problem for the reason that schooling datasets are uncommon in maximum applications, and eliminating “brute force” framework through exhaustive lexicons. Finally, maximum of the lately evolved fashions are primarily based totally accessible engineered phrases and semantic primarily based totally statistics to conquer those complexities. Irrespective, the above –said frameworks are time ingesting and highly-priced which relies upon upon language and the utility which isn't always powerful below the lifestyles of casual sentences, abbreviations and, no matter accomplishing most precision, it nevertheless stories minimal recall (lacking entities). Inversely, Machine Learning (ML) fashions are relevant to remedy the demerits because of the intrinsic and supremacy for variations. The latest modern-day ML technique follows 2step levels namely, Feature engineering and automated class. The preliminary step defines that, textual content through making use of mathematics vectors and area particular expertise. Secondly, it denotes that, class of a unmarried phrase into diverse NAE with acquainted alternatives for a classifier to be linear chain Conditional Random Fields (CRF), Structural Support Vector Machines (SSVM) in addition to most entropy. But, the hindrance of those fashions is that function engineering is probably time ingesting and laborious. Recently, the advent of DL has proven a more involvement in resolving this problem.

The most important contributions of the paper are given here. This paper designs an green NER approach in scientific literature making use of the Improved Cuckoo Search Optimization Algorithm (ICSA) with kernel intense studying machine (KELM), named ICSA-KELM. The supplied ICSA-KELM version incorporates 3 predominant sub procedures together with pre-processing, class, and parameter tuning. The supplied version consists of out the records preprocessing to extrade the uncooked scientific records right into a beneficial format. Moreover, KELM version is achieved to perform class method. Lastly, the 2 predominant parameters of KELM together with penalty parameter `C` and kernel bandwidth `γ` of the KELM are calculated making use of ICSA algorithm. A enormous set of experimentations takes area to compute the gifted classifier outcomes of the ICSA-KELM version.

The relaxation of the sections with inside the paper are prepared as given here. Section 2 critiques the lately evolved NER fashions. Section three introduces the operating of the supplied ICSA-KELM version and segment four validates the experimental evaluation of the supplied version. At last, segment five concludes the paper.

2. Literature Review

Presently, the fast improvement of Deep Neural Networks (DNN) for different NLP tasks like NER. Pretrained word inserting was conveyed in old style ML models and in NN, in which Deconourt et al. [3] has accomplished moderate execution when contrasted and traditional framework for de-distinguishing proof of patient subtleties. Cocos et al. [4] utilized Bidirectional Long Short-Term Memory (Bi-LSTM) procedure to name Adverse Drug Reactions (ADR) in pharmacovigilance. Wei et al. [5] joined the recreation result of Bi-LSTM and Conditional Random Fields (CRF) and took care of it as contribution to SVM to perform infection name investigation. A portion of the sensible restrictions are finished expectation is unstructured and not many significant connections may be lost from yield. Then, at that point, Jaganatha and Yu [6] used a Bi-LSTM-CRF for naming NAE from Electronic Health Records (EHR) of disease patients. This approach is changed from CRF yield while pairwise possibilities are created under the use of Convolutional Neural Network (CNN) rather than utilizing typical progress framework.

The significant obligation of NAE Recognition is to observe the bits of text which implies the targets like medications, signs, meds, and cancers. Rule-based as well as word reference based modalities are exceptionally fundamental in this application. For example, Fukuda et al. [7] introduced a standard based structure for removing material names like proteins from natural records. Sadly, these models experience the ill effects of nonappearance of speculation because of the prerequisite of hand-designing guidelines. Also, designers have figured out how to apply ML models to track down substances from irregular information. Wang et al. [8] made an examination among 6 biomedical NER materials as per the Hidden Markov Model (HMM) and CRF. In any case, ML approaches need to choose aggregate elements genuinely, which requires immense labor supply and time. As of late, profound learning (DL) approaches are appropriate in expanding the functioning capacity of NER with no component designing, have accomplished better consideration from the engineers. For example, Zhu et al. [9] inferred a start to finish DL model to biomedical NER activities that clout the neighborhood settings by utilizing CNN. If there should be an occurrence of Recurrent Neural Network (RNN). Chen et al. [10] applied word reference elements to track down the obscure and strange clinical NAE. In this way, DL approaches experience lacking preparation information.

Devlin et al. [11] assumed that NER is a downstream process of BERT to extract NAE from news (MSRA-NER). Pires et al. [12] examined zero-shot NER using multilingual BERT. Followed by, pre-training approaches have applied on domain-based NER, like biomedicine. Hakala and Pyysalo [13] used a CRF-based baseline framework as well as multilingual BERT to Spanish biomedical NER process. Therefore, the abilities of pre-training methods were not exploited completely. In addition, adopting pre-training approaches by excluding BERT in particular domains like NER in Chinese clinical studies have shown a greater impact recently.

3. The Proposed ICSA-KELM model

The working principle involved in the presented ICSA-KELM model is demonstrated in Fig. 1. Primarily, the input raw medical data is preprocessed in three stages such as segmentation, tokenization, and morphosyntactic analysis to make it compatible for further processes. Next, the preprocessed data are classified by the use of KELM model. Finally, the ICSA technique is executed to optimally tune the parameters of the KELM method. The detailed working of these processes is provided in the succeeding sections.

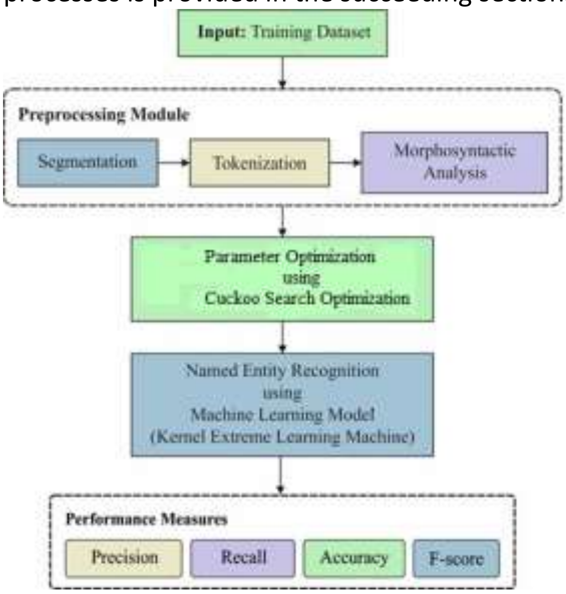

Fig. 1. Overall working process of ICSA-KELM model

3.1. Preprocessing

Basically, Preprocessing is one of the significant elements. It is in charge to pre-process the input data and perform collective preprocessing operations so that the text is facilitated as input for upcoming module. Moreover, preprocessing plays a vital role in NLP and Information Retrieval research works [14]. Here, preprocessing module is classified as 3 portions which are invariably and implemented subsequently.

Segmentation: In this model, a sentence is computed without a dependency on preceding content. It is simulated via way of means of classifying the entire textual content as sentences via way of means of finishing with punctuation marks like Period (.), query mark (?), exclamation (!), and suspension points (...).

Tokenization: It has a tendency to be a 2nd module withinside the chain. It orders the textual content as n-grams, words, and combination words. A remember of n-grams is defined, and the tokenization is produced from textual content as vector of combination words. In mild of this model, selections are made via accentuation, that's produced from disconnecting the non-alphanumeric characters from words. The accentuation marks via way of means of barring dash (-), at (@), and reduce (/) are segregated via way of means of void area from alphanumeric characters. Normally, definition implemented in processing the chain is reduce as unigrams.

Morphosyntactic Analysis: Once the tokenization is completed, morphosyntactic analysis is carried out to divide the unigrams. Followed by, text is scrutinized and categorized using POS Tagging with the help of various methods and models. The performance of POS Tagging is comprised of examining and tagging the words from text at syntactic level.

3.2. KELM based Classification Process

Generally, KELM has defined an extended variety of ELM, presented by Huang et al. [15]. It is the combination of the kernel function and ELM, which make sure that the system is comprised of optimal generalization function and enhances the forward network learning speed and eliminates the issues in Gradient Descent (GD) training models projected using Back Propagation Neural Networks (BPN), such that the simple drop in local optimal, indefinite iterations, and so on. KELM is a multi-dominance of ELM and unifies the kernel function for the purpose of mapping the linear attached pattern to the high-dimension feature space for accomplishing linear separability and enhance accuracy. Fig. 2 shows the structure of KELM. With the help of L hidden nodes from output layer, the resultant function of single-hidden-layer feed-forward neural networks (SLFNs) is depicted as:

$$f(u) = h(u)\beta \tag{1}$$

Where $\beta = [\beta_1, \beta_2, \ldots, \beta_L]^{\mathrm{T}}$ defines the final weight vector among hidden layer with L neurons and consequent neurons, $h(xu) = [h_1(u), h_2(u), \cdots, h_L(u)]$ implies the resultant vector related to hidden layer of input u, where mapping is carried out from input to ELM feature space [16].

In the recently deployed KELM, the establishment of positive coefficient to learning system tends to make stable ELM. When it can be non-singular, coefficient C is included in the diagonal of $\mathrm{HH^T}$ while determining the final weight β. Therefore, corresponding result of regular ELM is demonstrated in the following:

$$f(u) = h(u)\beta = h(u)\mathrm{H^T}\left(\frac{\mathrm{I}}{\mathrm{C}} + \mathrm{HH^T}\right)^{-1} T \quad (2)$$

It is clear that ELM embedded with a kernel matrix is expressed as follows.

$$\Omega_{KELM} = \mathrm{HH^T} : \Omega_{KELMi,\mathrm{j}} = h(u_i)\cdot h(u_j) = K(u_i, u_j) \quad (3)$$

The resultant function is illustrated as: $f(u) = h(u)\mathrm{H^T}\left(\frac{\mathrm{I}}{\mathrm{C}} + \mathrm{HH^T}\right)^{-1}\mathrm{T} = \left(\frac{\mathrm{I}}{\mathrm{C}} + \begin{bmatrix} K(u,u_1) \\ \cdots \\ K(u,u_N)\end{bmatrix}^T \Omega_{KELM}\right)^{-1} \mathrm{T}$ (4)

Here, there is no point to hidden layer feature map $h(u)$ due to the replacement of corresponding kernel function $K(u, v)$.

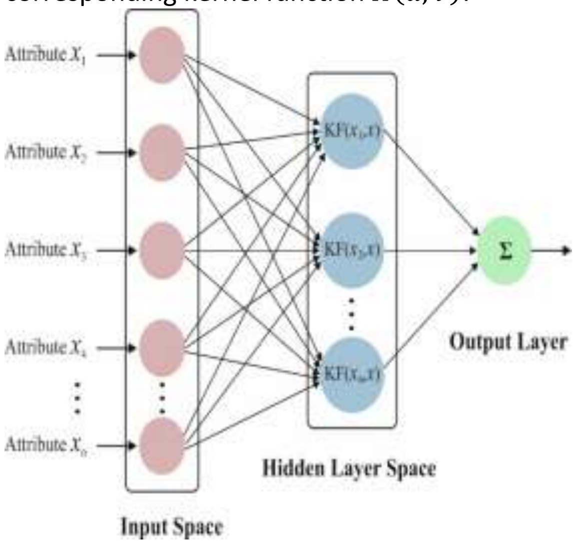

Fig. 2. Structure of KELM

3.3. Processes involved in ICSA

The important CS approach relies on the brood parasitism of cuckoo very own family that lays eggs inside aspect the homes of opportunity host birds. If there ought to upward thrust up an incidence of effortlessness, the CS method is worked in 3 best systems [14]: (1) A cuckoo lays single egg at an at once, in an arbitrarily decided on set; (2) the proper homes with first-rate fantastic eggs are carried out for humans with within the future; (3) don't forget of to be had host homes are distributed, and egg laid via cuckoo has been diagnosed via the host fowl with a opportunity p_a∈[0,1]. Now, the host fowl can also moreover depart the egg away or defend the residence and make a completely unique and muddled home. As in line with the predefined rules, the CS method is characterized as exhibited in Fig. 2 [14].

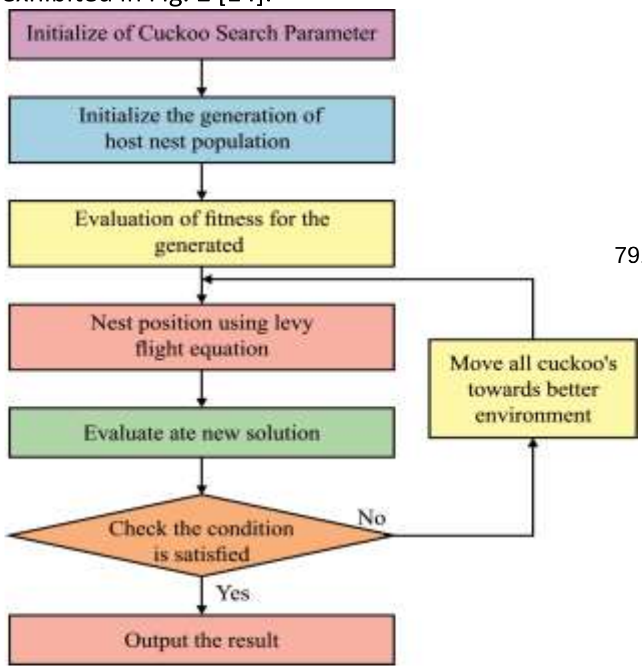

Fig. 2. Flowchart of ICSA model

In addition, the plan applied a superior blend of neighborhood arbitrary stroll as well as a worldwide explorative irregular walk, oversaw by utilizing exchanging boundary p_a. Subsequently, neighborhood arbitrary walk is communicated as,

$$\mathrm{x\_i\hat{}(t+1)} = \mathrm{x\_i\hat{}t} + \alpha s \otimes H(p\_a - \varepsilon) \otimes (\mathrm{x\_}j\mathrm{\hat{}t} - \mathrm{x\_k\hat{}t}) \quad (5)$$

Where x_j^t and x_k^t are 2 diverse answers selected randomly the use of random permutation, H defines a Heaviside performance, ε implies a random value, and s way the step size. Besides, international random stroll is carried out with the assist of Lévy fights [14–17]:

$$x_i^{\mathrm{t+1}} = \mathrm{x_i^t} + \alpha \oplus \mathrm{Lévy}(s, \lambda) \quad (6)$$

Where, α > zero connotes the development length scaling factor; Lévy (s, λ) represents the development lengths dispersed primarily based totally on given probability conveyance as portrayed in (6) that is concerned boundless distinction along infinite mean:

$$Levy(s,\lambda) = \frac{\lambda\Gamma(\lambda)\sin(\pi\lambda/2)}{\pi}\frac{1}{s^{1+\lambda}} \quad (7)$$

To improve the looking through capacity of the model, the symmetrical technique and reenacted toughening (SA) process has been joined with CS system. The basic standard of symmetrical plan is to apply the highlights of partial examination for deciding ideal blend of levels [17]. The symmetrical exhibit of K angles and Q levels as well as M mixes are projected as L_M (Q^K), where Q means the indivisible number, M = Q^J, and J characterizes a positive number by the detailing of K=(Q^J-1)/(Q-1)

4. Experimental Validation

The experimental consequences of the KELM-ICSA set of rules were confirmed towards datasets [19, 20]. The dataset posted with the aid of using I2B2 from 2007 to 2012 is utilized. For discarding the reproduction data, each corpus are polled to a group of 4610 files. The I2B2 2009 mission launched a complete of 1255 specific documents, with 262 of them annotated.

Table 1 Result Analysis of Proposed ICSA-KELM Model on Dataset-1

Entities	Precision	Recall	F-score
Medication	90.36	90.13	90.07
Dosage	91.89	90.79	90.12
Mode	90.32	91.72	90.73
Frequency	89.06	90.87	89.83
Duration	89.65	89.83	89.75
Reason	87.38	89.36	88.79
Average	**90.78**	**91.45**	**90.88**

Table 1 and Fig. 4 shows the class consequences assessment of the ICSA-KELM model on the carried out dataset 1. The furnished ICSA-KELM model has proficiently identified the entities. For instance, on the recognition of drugs entity, the ICSA-KELM model has obtained a precision of 90.36%, don`t overlook of 90.13%, and F-score of 90.07% respectively. Besides, on the recognition of dosage entity, the ICSA-KELM model has attained a precision of 91.89%, recollect of 90.79%, and F-score of 90.12% respectively. Additionally, on the recognition of Mode entity, the ICSA-KELM model has reached a precision of 90.32%, recollect of 91.72%, and F-score of 90.73% correspondingly. Moreover, on the recognition of frequency entity, the ICSA-KELM model has finished a precision of 89.06%, recollect of 90.87%, and F-score of 89.83% correspondingly. Furthermore, on the recognition of length entity, the ICSA-KELM model has obtained a precision of 89.65%, recollect of 89.83%, and F-score of 89.75% respectively. Simultaneously, on the recognition of purpose entity, the ICSA-KELM model has attained a precision of 87.38%, recollect of 89.36%, and F-score of 88.79% correspondingly

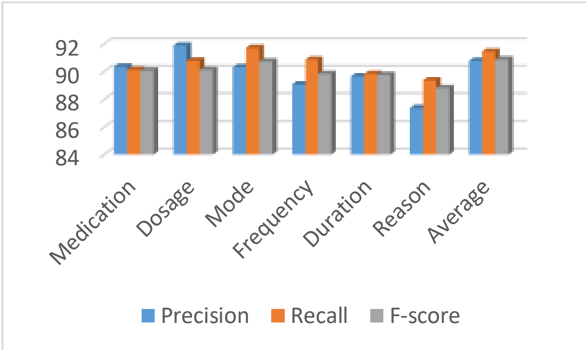

Fig. 4. Result analysis of KELM-ICSA model on Dataset-1

Fig. 5 investigates the average classifier analysis of the ICSA-KELM model on the applied dataset 1. The figure portrays that the presented ICSA-KELM model has reached a higher average precision of 90.78%, recall of 91.45%, and F-score of 90.88%.

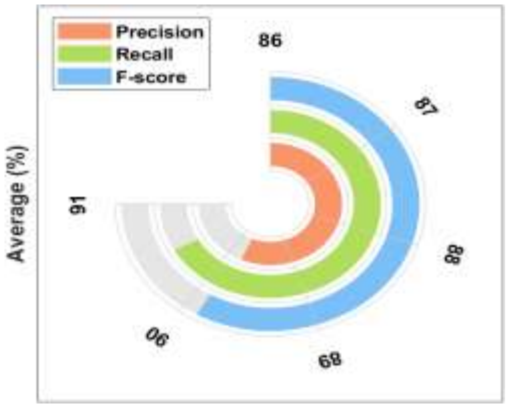

Fig. 5. Average analysis of KELM-ICSA model on Dataset-1

Table 2 Result Analysis of Existing with Proposed ICSA-KELM Model on Dataset-1

Methods	Precision	Recall	F-score
Proposed ICSA-KELM	89.78	90.45	89.88

CRF	82.33	70.68	75.51
Seq2Seq-RNN	84.60	81.49	82.40
RNN-SWE	84.74	85.01	84.87
BiGRU-CRF	72.14	78.20	74.99
Tree-CRF	80.20	75.50	78.90

Table 2 and Fig. 6 offers the comparative outcomes of the ICSA-KELM version with current strategies at the implemented dataset 1. The experimental outcomes said that the BiGRU-CRF version has acquired useless results with the least precision of 72.14%, do not forget of 78.20%, and F-rating of 74.99%. At the equal time, the Tree-CRF version has acquired a barely better precision of 80.20%, do not forget of 75.50%, and F-rating of 78.90%. Followed by, the CRF version has achieved a mild precision, do not forget, and F-rating of 82.33%, 70.68%, and 75.51%. Simultaneously, the Seq2Seq-RNN version has caused even higher overall performance with the precision, do not forget, and F-rating of 84.60%, 81.49%, and 82.40%. Along with that, the RNN-SWE version has attained aggressive precision, do not forget, and F-rating of 84.74%, 85.01 %, and 84.87%. At last, the supplied ICSA-KELM version has led to better precision, do not forget, and F-rating of 89.78%, 90.45%, and 89.88%.

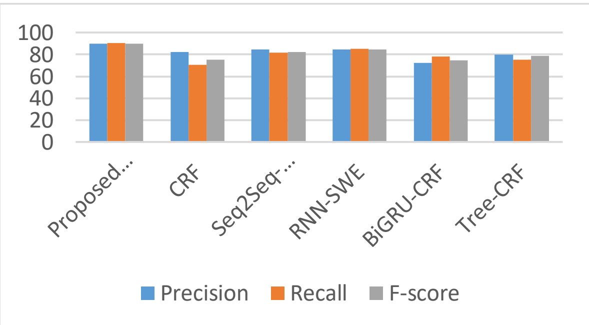

Fig. 6. Comparative result analysis of Dataset-1

Table three and Fig. 7 illustrates the type consequences evaluation of the ICSA-KELM version at the implemented dataset 2. The proposed ICSA-KELM version has proficiently diagnosed the entities. For instance, on the popularity of disorder entity, the ICSA-KELM version has attained a precision of 93.05%, consider of 95.78%, and F-rating of 94.86% correspondingly. Also, on the popularity of symptom entity, the ICSA-KELM technique has reached a precision of 96.26%, consider of 92.85%, and F-rating of 92.9% respectively. In addition, on the popularity of disorder organization entity, the ICSA-KELM version has done a precision of 93.82%, consider of 95.89%, and F-rating of 94.63% correspondingly. Furthermore, on the popularity of remedy entity, the ICSA-KELM version has received a precision of 95.89%, consider of 94.77%, and F-rating of 94.61% respectively.

Table 3 Result Analysis of Proposed ICSA-KELM Model on Dataset-2

Entities	Precision	Recall	F-score
Disease	93.05	95.78	94.86
Symptom	96.26	92.85	92.9
Disease group	93.82	95.89	94.63
Treatment	95.89	94.77	94.61
Average	**95.76**	**95.82**	**95.25**

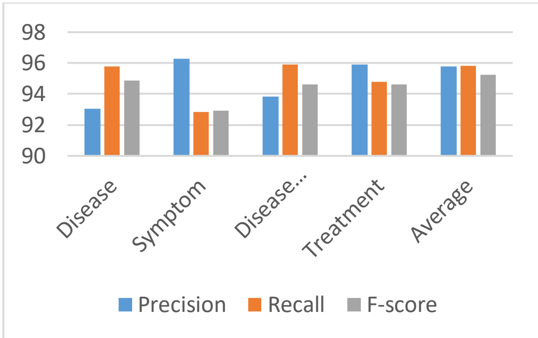

Fig. 7. Result analysis of KELM-ICSA model on Dataset-2

Fig. 8 examines the average classifier analysis of the ICSA-KELM method on the applied dataset 2. The figure showcases that the proposed ICSA-KELM model has attained a higher average precision of 95.76%, recall of 95.82%, and F-score of 95.25%.

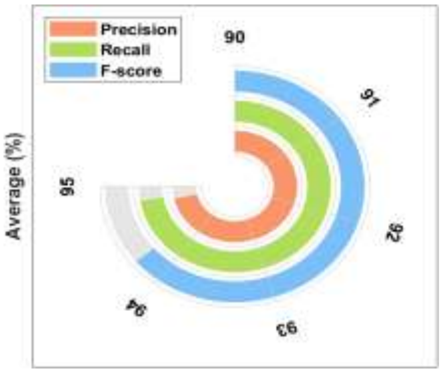

Fig. 8. Average analysis of KELM-ICSA model on Dataset-2

Table 4 and Fig. 9 offers the comparative outcomes of the ICSA-KELM model with existing techniques on the applied dataset 2. The experimental results stated that the NB model has attained ineffective results with the least precision of 78.07%, recall of 77.91%, and F-score of 77.99%. Likewise, the maximum entropy model has achieved a somewhat superior precision of 88.81%, recall of 88.81%, and F-score of 88.81%. Afterward, the SVM model has accomplished a moderate precision, recall, and F-score of 90.52%, 90.52%, and 90.52%. In line with, the CRF model has reached competitive precision, recall, and F-score of 93.15%, 93.15 %, and 93.15%. Finally, the presented ICSA-KELM model has resulted in a maximum precision, recall, and F-score of 95.76%, 95.82%, and 95.25%.

Table 4 Result Analysis of Existing with Proposed ICSA-KELM Model on Dataset-2

Methods	**Precision**	**Recall**	**F-score**
Proposed ICSA-KELM	95.76	95.82	95.25
NB	78.07	77.91	77.99
Maximum Entropy	88.81	88.81	88.81
SVM	90.52	90.52	90.52
CRf	93.15	93.15	93.15

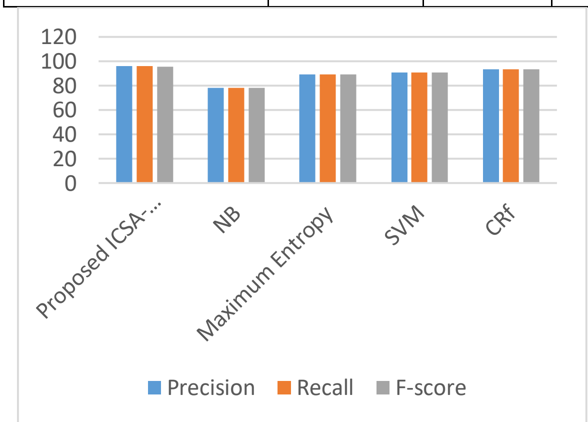

Fig. 9. Comparative result analysis of Dataset-2

5. Conclusion

This paper has evolved a brand new ICSA-KELM approach for NER in scientific literature. The provided ICSA-KELM version incorporates 3 principal sub tactics together with pre-processing, class, and parameter tuning. Primarily, the enter uncooked scientific statistics is preprocessed in 3 tiers together with segmentation, tokenization, and morphosyntactic evaluation to make it well suited for in addition tactics. Next, the preprocessed statistics are categorised via way of means of using KELM version. Finally, the ICSA set of rules is carried out to optimally track the 2 principal parameters of KELM together with penalty parameter `C` and kernel bandwidth `γ` of the KELM version. A complete effects evaluation become completed and the effects are tested interms of wonderful dimensions. The received simulation effects proven the advanced overall performance of the ICSA-KELM version over the opposite methods. As part of destiny scope, the class overall performance may be more suitable the use of superior deep learning (DL) architectures.

References

[1] An Efficient Named Entity Recognition Model in Medical Health Records using Water Wave Optimization with Kernel Extreme Learning Machine Dr. K. Arutchelvan, R. Ramachandran. Design Engineering, vol(2021),no 7, pp.8564- 8580

[2] Gunter TD, Terry NP. The emergence of national electronic health record architectures in the United States and Australia: models, costs, and questions. Journal of medical Internet research. 2005; 7(1).

[3] Z. Sun and H. Wang, “Overview on the advance of the research on named entity recognition,” New Technology of Library and Information Service, vol. 26, no. 6, pp. 42–47, 2010.

[4] F. Dernoncourt, J.Y. Lee, O. Uzuner, P. Szolovits, De-identification of patient notes with recurrent neural networks, J. Am. Med. Inform. Assoc. 156 (2016).

[5] A. Cocos, A.G. Fiks, A.J. Masino, Deep learning for pharmacovigilance: recurrent neural network architectures for labeling adverse drug reactions in Twitter posts, J. Am. Med. Inform. Assoc. 180 (2017).

[6] Q. Wei, T. Chen, R. Xu, Y. He, L. Gui, Disease named entity recognition by combining conditional random fields and bidirectional recurrent neural networks, Database (2016).

[7] A.N. Jagannatha, H. Yu, Structured prediction models for RNN based sequence labeling in clinical text, Conference on

Empirical Methods in Natural Language Processing (EMNLP) 2016 (2016) 856.
[8] K. Fukuda, T. Tsunoda, A. Tamura, T. Takagi et al., “Toward information extraction: identifying protein names from biological papers,” in Proceedings of the Pacific Symposium on Biocomputing, pp. 707–718, Maui, HI, USA, 1998.
[9] X. Wang, C. Yang, and R. Guan, “A comparative study for biomedical named entity recognition,” International Journal of Machine Learning and Cybernetics, vol. 9, no. 3, pp. 373– 382, 2018.
[10] Q. Zhu, X. Li, A. Conesa, and C. Pereira, “Gram-CNN: a deep learning approach with local context for named entity recognition in biomedical text,” Bioinformatics, vol. 34, no. 9, pp. 1547–1554, 2018.
[11] X. Chen, C. Ouyang, Y. Liu, and Y. Bu, “Improving the named entity recognition of Chinese electronic medical records by combining domain dictionary and rules,” International Journal of Environmental Research and Public Health, vol. 17, no. 8, 2020.
[12] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: pre-training of deep bidirectional transformers for language understanding,” in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), pp. 4171–4186, Minneapolis, MN, USA, 2019.
[13] T. Pires, E. Schlinger, and D. Garrette, “How multilingual is multilingual BERT?” in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL), pp. 4996–5001, Florence, Italy, 2019.
[14] K. Hakala and S. Pyysalo, “Biomedical named entity recognition with multilingual BERT,” in Proceedings of the 5th Work-shop on BioNLP Open Shared Tasks, pp. 56–61, Hong Kong, China, 2019.
[15] Dias, M., Boné, J., Ferreira, J.C., Ribeiro, R. and Maia, R., 2020. Named Entity Recognition for Sensitive Data Discovery in Portuguese. *Applied Sciences*, *10*(7), p.2303.
[16] Huang, G.-B.; Zhou, H.; Ding, X.; Zhang, R. Extreme learning machine for regression and multiclass classification. IEEE Trans. Syst. Man Cybern. Part B 2012, 42, 513–529.
[17] Lv, X., Chen, H., Zhang, Q., Li, X., Huang, H. and Wang, G., 2018. An improved bacterial-foraging optimization-based machine learning framework for predicting the severity of somatization disorder. *Algorithms*, *11*(2), p.17.
[18] Zheng, Y.-J. (2015). Water wave optimization: a new nature-inspired metaheuristic. Computers & Operations Research, 55, 1-11.
[19] Zhao, F., Liu, H., Zhang, Y., Ma, W. and Zhang, C., 2018. A discrete water wave optimization algorithm for no-wait flow shop scheduling problem. *Expert Systems with Applications*, *91*, pp.347-363.
[20] Gligic, L., Kormilitzin, A., Goldberg, P. and Nevado-Holgado, A., 2020. Named entity recognition in electronic health records using transfer learning bootstrapped neural networks. *Neural Networks*, *121*, pp.132-139.
[21] Dong, X., Chowdhury, S., Qian, L., Li, X., Guan, Y., Yang, J. and Yu, Q., 2019. Deep learning for named entity recognition on Chinese electronic medical records: Combining deep transfer learning with multitask bi-directional LSTM RNN. *PloS one*, *14*(5), p.e0216046.
[22] R. Ramachandran and K. Arutchelvan, "Optimized Version of Tree based Support Vector Machine for Named Entity Recognition in Medical Literature," 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), 2020, pp. 357-361, doi:10.1109/ICISS49785.2020.931605