Indonesian Journal of Electrical Engineering and Computer Science
Vol. 41, No. 1, January 2026, pp. 191~199
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v41.i1.pp191-199

An enhanced NLP approach for BI-RADS extraction in breast ultrasound reports using deep learning

Ahmed Sahl[1], Shafaatunnur Hasan[2], Maie M. Aboghazalah[3]
[1]Department of Computer Science, College of Computer Science, King Khalid University, Abha, Saudi Arabia
[2]Faculty of Computing, Universiti Teknologi Malaysia, Johor Bharu, Malaysia
[3]Computer Science College, Nahda University, Beni Suef, Egypt

Article Info

Article history:

Received Nov 11, 2024
Revised Dec 4, 2025
Accepted Dec 13, 2025

Keywords:

BI-RADS
Deep-learning
LSTM
Named entity recognition
NLP
RNN

ABSTRACT

Breast cancer stands as one of the top causes of death around the globe, making the accurate interpretation of breast ultrasound reports vital for early diagnosis and treatment. Unfortunately, key findings in these reports are often buried in unstructured text, complicating automated extraction. This study presents a deep learning-based natural language processing (NLP) approach to extract breast imaging reporting and data system (BI-RADS) categories from breast ultrasound data. We trained a recurrent neural network (RNN) model, specifically using a BiLSTM architecture, on a dataset of reports that were manually annotated from a hospital in Saudi Arabia. Our approach also incorporates uncertainty estimation techniques to tackle ambiguous cases and uses data augmentation to boost model performance. The experimental results indicate that our deep learning method surpasses traditional rule-based and machine-learning techniques, achieving impressive accuracy in classification tasks. This research plays a significant role in automating radiology reporting, aiding clinical decision-making, and pushing forward the field of breast cancer research.

Corresponding Author:

Maie M. Aboghazalah
Computer Science Collage, Nahda University
Beni Suef, Egypt
Email: eng_maie@yahoo.com

1. INTRODUCTION

Breast cancer continues to be one of the most widespread cancers globally, posing a substantial burden on both individual patients and public health systems [1]. Early detection is crucial in reducing mortality rates and alleviating financial strain. Consequently, medical guidelines advocate for routine mammography screening to assess breast cancer risk [2]. To minimize discrepancies and standardize radiologists' reporting of mammographic results, the American College of Radiology (ACR) introduced the breast imaging reporting and data system (BI-RADS) [3]. This system provides a standardized lexicon for reporting mammographic findings and a six-category classification framework to assess malignancy risk as shown in Table 1 [4]. Although BI-RADS was initially designed for mammography, it has since been adapted for other imaging modalities, such as magnetic resonance imaging (MRI) and breast ultrasound. Nevertheless, these essential radiology discoveries are frequently recorded in unstructured narrative formats, rendering them unavailable to computational systems dependent on organized data. Numerous investigations have utilized natural language processing (NLP) strategies to extract BI-RADS findings and final assessment categories from diverse English-language breast radiography reports, encompassing mammograms and breast ultrasound data. Early approaches were predominantly rule-based [5], such as MedLEE, one of the pioneering clinical NLP systems developed for extracting abnormal findings from mammography reports [6].

Table 1. A description of BI-RADS six-category mammography findings

BI-RADS category	Findings	Likelihood of breast cancer	Management
0	Need additional imaging or prior examination	N/A	Recall for additional imaging and/or await prior examinations
1	Negative	Negligible	Routine screening
2	Benign	Negligible	Routine screening
3	Probably benign	<2%	Short interval-follow-up
4	Suspicious	23-34%	Tissue diagnosis
5	Highly suggestive of malignancy	≥95%	Tissue diagnosis
6	Malignancy confirmed biopsy	100%	Surgical excision when clinically appropriate

Other rule-based methods were implemented to classify BI-RADS breast tissue composition from mammography records [7] to examine ambiguity in BI-RADS assessment categories utilizing the GATE NLP framework [8]. Later advancements introduced machine learning-based NLP systems, such as those employing support vector machines (SVMs) and Naïve Bayes (NB) to extract BI-RADS categories and laterality classifications, achieving an F1-score of 0.95 and surpassing rule-based approaches in performance [9]. Additionally, alternative NLP pipelines have been utilized to extract BI-RADS evaluation categories [10], while statistical testing has been employed to support clinical decision-making using extracted findings [11], [12].

There are now enormous digital archives of clinical documents in Saudi Arabia due to the widespread use of electronic health records, necessitating NLP-driven methods to extract meaningful insights. Several studies have explored NLP techniques for symptom extraction and disease progression analysis within Saudi medical records [13]-[15]. More recently, research on Saudi clinical text has expanded to address broader NLP challenges, including negation detection and tumor-related information extraction from surgical notes. Despite these contributions, no prior study has specifically addressed the structured extraction of BI-RADS findings from Saudi breast ultrasound reports, highlighting a critical gap in the field.

Structured BI-RADS extraction is crucial for clinical decision-making. To address this gap, we present a deep learning approach to Saudi Arabian breast ultrasound data that can extract all BI-RADS finding classifications. In light of the growing interest in deep learning models, we explicitly investigate the application of a bidirectional long-short term memory (BiLSTM) network for this task. We specifically explore the application of a BiLSTM-based recurrent neural network (RNN) for extracting BI-RADS findings from breast ultrasound reports [16]-[19]. We use an annotated dataset of 465 reports to illustrate that deep learning methodologies surpass conventional conditional random fields (CRF) based machine learning techniques. This underscores their potential to enhance breast cancer research and clinical decision support.

Recent NLP advancements show that RNN-based models surpass CRF models in named entity recognition (NER) tasks, especially when integrating human-generated features and domain-specific dictionaries [19]. In the clinical domain, RNNs have been effectively applied to medical event detection [20], medical concept extraction [21], extraction of temporal information in clinical contexts [22], and disease name recognition [23]. However, there is still a lack of study on using deep learning and machine learning to extract BI-RADS in Saudi healthcare. Here is the outline for the rest of the paper: first, we examine the methodology and experiment; second, we present the results and discuss what we found; and finally, we wrap up the study in section 4.

2. METHOD AND EXPERIMENT

2.1. Dataset and annotation

Data source:

We utilized breast ultrasound reports from Khamis Mushayt Maternity Hospital in Aseer Province, Saudi Arabia, covering the period from 2015 to 2020. All reports were anonymized to protect patient confidentiality by replacing sensitive information (e.g., patient names, addresses, telephone numbers, and medical staff names) with surrogates or pseudonyms. This ensured the text remained coherent without unusual gaps. Notably, our dataset comprised textual reports without accompanying ultrasound images.

Annotation process:

An iterative approach was employed to develop the annotation guidelines:

i) Initial drafting: collaborated with domain experts to create the initial annotation guidelines.
ii) Pilot annotation: two annotators independently annotated a subset of 65 reports using the initial guidelines. The inter-annotator agreement yielded an F-measure of 0.821, highlighting the task’s complexity.

iii) Guideline refinement: based on discrepancies observed, the guidelines were refined for clarity and comprehensiveness.
iv) Re-annotation: the same 65 reports were re-annotated using the updated guidelines, resulting in an improved F-measure of 0.942.
v) Full annotation: the finalized guidelines were applied to annotate the remaining 400 reports, culminating in 465 annotated reports. Figure 1.

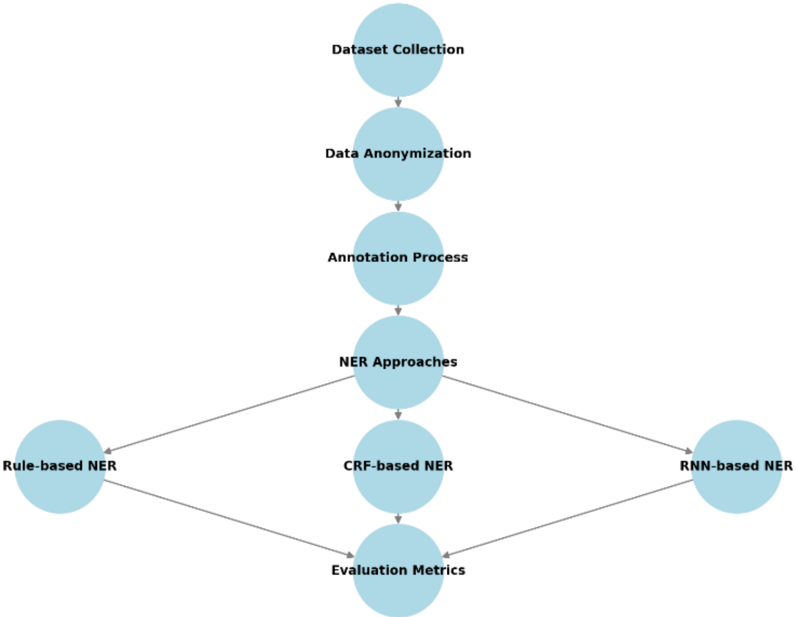

Figure 1. The method flowchart is the entire methodology pipeline

A web-based annotation application called BRAT, which is open-source, made the annotation process easier [24]. Figure 2 in addition to entity annotation, we marked the negation state of every entity. Afterward, the annotated corpus was divided into two halves: a training set comprising 310 reports (about two-thirds) and a testing set comprising 155 reports (about one-third).

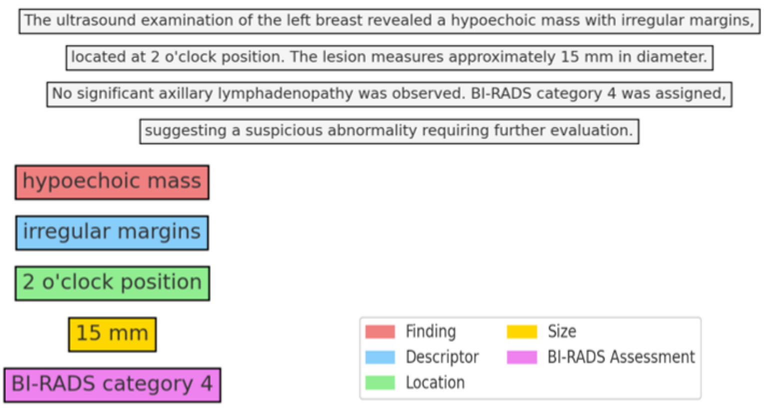

Figure 2. BART annotation for a brest cancer radiology report

2.2. NER methods

NER is a crucial problem in NLP that entails identifying and classifying entities in text into established categories [25], such as names of people, organizations, locations, and specific terminologies pertinent to a domain Figure 3. We explored three NER approaches:

2.2.1. Rule-based method:

Established as a baseline, this method integrates manually generated rules with an entity dictionary and is implemented through the UIMA Ruta framework. The entity dictionary was constructed from the annotated development set, and regular expressions were designed to identify specific patterns [26]. Ruta rules handled complicated permutations, including combining integers and units.

2.2.2. CRF-based method:

CRFs are probabilistic models adept in sequence labeling tasks. We utilized the CRF++ package to incorporate essential NER components, including bag-of-words and n-grams, to construct the NER model. CRFs include contextual information, making them suitable for tasks like ours.

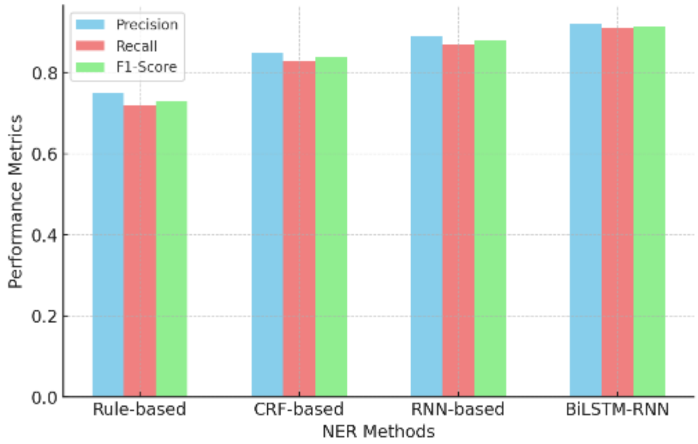

Figure 3. Performance comparison on NER methods

2.2.3. RNN-based deep learning method:

Capturing long-term dependencies in sequential data is a strong suit of RNNs, especially LSTM networks. We implemented an RNN architecture with LSTM units drawn from Lample's creations *et al.* [18] enhancements explored included:

– Character embeddings: capturing morphological features of words.
– BiLSTM: radiology reports are structured as a sequential medical text, making BiLSTM effective for capturing long-range dependencies in textual patterns. Processing sequences in forward and backward directions to utilize past and future context Figure 4.

Finally, what was needed for the RNN model were:

– Character embedding dimension: 50
– LSTM layer size: 100 units at the word level
– Learning rate: 0.005
– Dropout probability: 0.5

Several epochs passed before the training and validation losses began to change. Figure 5 illustrates the BiLSTM model's loss convergence curve, demonstrating the stability and effectiveness of the training process. While transformer-based models such as BERT have shown strong performance in NLP tasks, we opted for a BiLSTM-based approach due to its effectiveness in handling sequential medical text, lower computational requirements, and better generalizability on a moderate-sized dataset.

Evaluation measures: We used standard metrics measures to assess the NER systems' efficacy:

– Precision (PRE): the proportion of accurately anticipated positive observations to the projected positives.
– Recall (REC): the proportion of accurately predicted positive instances to the total instances in the actual category.
– F1-score (F1): the sum of precision and recall, calculated using weights.
– Accuracy (ACC): the proportion of accurately anticipated observations to the total observations.

Mathematically, these are defined as:

$$\text{PRE} = \frac{TP}{TP+FP} \quad (1)$$

$$\text{REC} = \frac{TP}{TP+FN} \tag{2}$$

$$\text{F1} = \frac{(2*PRE*REC)}{PRE+REC} \tag{3}$$

$$\text{ACC} = \frac{TP+TN}{TP+FP+FN} \tag{4}$$

TP = true positives, FP = false positives, FN = false negatives, TN = true negatives.
The training set was utilized to develop and train the NER models, while the test set served to evaluate their performance based on the metrics mentioned above.

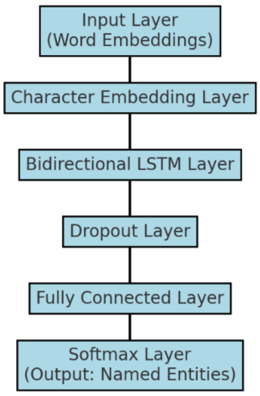

Figure 4. BiLSTM-based model architecture for BI-RADS extraction

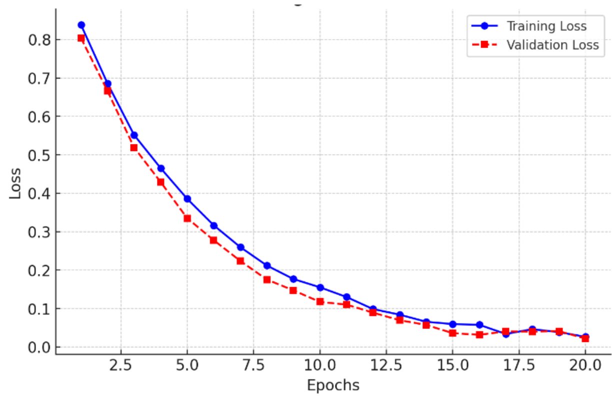

Figure 5. Loss convergence of BiLSTM model

3. RESULTS AND DISCUSSION

3.1. Result

In alignment with BI-RADS standards, experts identified 20 entity types in the breast ultrasound reports, with an additional category labeled “Other” to group infrequent occurrences. Out of 465 reports, 9,132 entities were annotated. Four entity types—location, echo, size, and vascularity—were the most frequent, each appearing over 1,000 times. Conversely, ten entity types, such as architectural distortion, calcifications, and tissue composition, had fewer than 100 occurrences as shown in Table 2.

Table 2. Distribution of BI-RADS entity categories in the annotated corpus

Entity type	Number of entities
Alder	15
Architectural-distortion	22
Calcifications	30
Ductchanges	52
Echo	1,308
Elasticity-assessment	60
Hardness-ratio	19
Location	2,301
LymphNode	399
Margin	801
Masses	158
Negation	655
Orientation	26
Posterior-features	31
Resistance-index	231
Shape	461
Size	1,440
Skin	131
Tissue-composition	14
Vascularity	958
Other	20
Total	**9,132**

We aimed to determine which of three methods—the rule-based approach, the CRFs-based model, and the RNN-based model (BiLSTM)—was most effective in NER evaluation. The deep learning model using BiLSTM accomplished the greatest F1-score of 0.908, followed by the model based on CRFs with an F1-score of 0.885, and the approach based on rules with an F1-score of 0.864, as illustrated in Table 3. Compared to more conventional methods, these results demonstrate that deep learning-based systems are more effective in extracting BI-RADS conclusions from breast ultrasound records. The evaluation metrics used to assess the performance of the NER models include PRE, REC, F1, and ACC, which are calculated using the (5)-(8):

$$\mathrm{PRE} = \frac{TP}{TP+FP} \tag{5}$$

$$\mathrm{REC} = \frac{TP}{TP+FN} \tag{6}$$

$$\mathrm{F1} = \frac{(2*PRE*REC)}{PRE+REC} \tag{7}$$

$$\mathrm{ACC} = \frac{TP+TN}{TP+FP+FN} \tag{8}$$

- Precision (PRE) is defined as in (5).
- Recall (REC) is defined as in (6).
- F1-measure (F1) is defined as in (7).
- Accuracy (ACC) is defined as in (8).

Table 3. Performance comparison of NER approaches (Rule-based, CRF, and BiLSTM)

Method	Precision	Recall	F1-score
Rule-based	0.871	0.820	0.864
CRFs-based	0.902	0.862	0.885
BiLSTM	0.913	0.895	0.908

The superior performance of the RNN-based model suggests that deep learning techniques can effectively capture complex linguistic patterns in radiology reports. However, the model's performance varied across different entity types, particularly struggling with those with fewer occurrences in the dataset.

3.2. DISCUSSION

3.2.1. Addressing gaps in previous research

While previous studies have explored BI-RADS entity extraction using rule-based and statistical methods, the application of deep learning for structured information extraction from breast ultrasound reports remains underexplored. Additionally, most prior research has focused on mammography rather than ultrasound imaging. To bridge these gaps, our study utilizes a deep-learning NER system based on BiLSTM to find breast ultrasound data with BI-RADS information.

3.2.2. Key findings and novel contributions

Our study demonstrates that RNN-based deep learning models (BiLSTM) outperform traditional CRFs and rule-based approaches in extracting BI-RADS entities. The best possible F1-score of 0.908, achieved by the RNN-based model, underscores the effectiveness of deep learning in this domain. Additionally, our annotation process, involving 18 BI-RADS entity types, sets this study apart by providing a more comprehensive labeled dataset than previous efforts.

3.2.3. Comparison with existing literature

Previous research has shown that deep learning is useful in medical text processing, and our results are in line with that [27]. For instance, An *et al.* [27] demonstrated that BiLSTM models improve entity recognition in clinical text. However, our approach extends this by applying deep learning to breast ultrasound reports precisely rather than broader clinical narratives. Unlike earlier methods that relied solely on hand-crafted rules or statistical models, our deep learning model effectively captures contextual dependencies, enhancing accuracy.

3.2.4. Limitations of the study

While the results show promise, our study has certain limitations. First, while the dataset includes a diverse set of reports, the relatively low frequency of some entity types may have impacted the model’s ability to generalize. Additionally, the absence of image-text alignment means that entity extraction was performed solely on textual data, limiting multimodal insights. Future studies should consider integrating imaging features alongside text-based analysis.

3.2.5. Implications for future research

Future research should explore hybrid models incorporating deep learning and rule-based techniques to improve performance on rare entity types. Expanding the dataset with reports from multiple institutions could enhance the model’s robustness and generalizability. Exploring transformer-based architectures like BERT or BioBERT may improve entity extraction accuracy.

3.2.6. Conclusion

Our findings confirm that deep learning approaches, particularly RNN-based models, extract BI-RADS results from breast ultrasound reports more effectively. The study contributes to the field by presenting a comprehensive annotated dataset and demonstrating the feasibility of deep learning for structured information extraction in radiology. Future advancements in multimodal learning and dataset expansion could further enhance automated BI-RADS classification and clinical decision support.

4. CONCLUSION

This study successfully addressed the challenge of extracting structured BI-RADS findings from unstructured reports generated by deep learning for breast ultrasounds. As anticipated in the Introduction, our BiLSTM model outperformed rule-based and CRF methods in BI-RADS extraction, demonstrating its effectiveness in capturing linguistic patterns in radiology reports. It achieved an F1-score of 0.908, confirming its potential for automated clinical decision support.

Our findings align with previous research on deep learning in medical text processing but extend its application to breast ultrasound, an underexplored area. While our model demonstrates high accuracy, limitations such as dataset diversity and the lack of image-text alignment highlight areas for further study. Future work should explore hybrid deep learning-rule-based models, expand datasets across institutions, and integrate transformer models like BERT for improved accuracy. Incorporating multimodal learning—aligning text with ultrasound images—could further enhance artificial intelligence (AI)-driven radiology decision support.

ACKNOWLEDGMENTS

A heartfelt “thank you” from all of us goes out to Dr. Abd Almotea Mahomud and Dr. Belgasim Mohammed for their invaluable assistance during this research. Their expert guidance and insightful feedback have significantly enhanced the quality of this work. We are deeply appreciative of their time and support.

FUNDING INFORMATION

No funding is involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Ahmed Sahl	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Shafaatunnur Hasan		✓		✓		✓		✓	✓	✓	✓	✓	✓	
Maie M. Aboghazalah	✓		✓	✓			✓			✓	✓		✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

As the writers of this paper, we affirm that no ties or financial interests have any bearing on the content or conclusions drawn from it.

DATA AVAILABILITY

- You can request a copy of the data used to support the study's conclusions from the corresponding author, [Maie M. Aboghazalah]. Restriction measures prevent the data, which may reveal research participants' personal information, from being publicly available.
- If you would like derived data to back up the study's conclusions, you can get in touch with the corresponding author, [Maie M. Aboghazalah].

REFERENCES

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, Nov. 2018, doi: 10.3322/caac.21492.

[2] A. O. Berg *et al.*, "Screening for breast cancer: recommendations and rationale," *Annals of Internal Medicine*, vol. 137, no. 5 I, pp. 344–346, Sep. 2002, doi: 10.7326/0003-4819-137-5_part_1-200209030-00011.

[3] S. J. Magny, R. Shikhman, and A. L. Keppke, "Breast imaging reporting and data system," *StatPearls*, 2023. https://www.ncbi.nlm.nih.gov/books/NBK459169/ (accessed Aug. 28, 2023).

[4] M. M. Eberl, C. H. Fox, S. B. Edge, C. A. Carter, and M. C. Mahoney, "BI-RADS classification for management of abnormal mammograms," *Journal of the American Board of Family Medicine*, vol. 19, no. 2, pp. 161–164, Mar. 2006, doi: 10.3122/jabfm.19.2.161.

[5] N. L. Jain and C. Friedman, "Identification of findings suspicious for breast cancer based on natural language processing of mammogram reports," *Journal of the American Medical Informatics Association*, vol. 4, no. SUPPL., pp. 829–833, 1997.

[6] B. Percha, H. Nassif, J. Lipson, E. Burnside, and D. Rubin, "Automatic classification of mammography reports by BI-RADS breast tissue composition class," *Journal of the American Medical Informatics Association*, vol. 19, no. 5, pp. 913–916, Sep. 2012, doi: 10.1136/amiajnl-2011-000607.

[7] B. Selen and R. Daniel, "Automated detection of ambiguity in BI-RADS assessment categories in mammography reports," in *Studies in Health Technology and Informatics*, vol. 197, 2014, pp. 35–39.

[8] H. Gao, E. J. A. Bowles, D. Carrell, and D. S. M. Buist, "Using natural language processing to extract mammographic findings," *Journal of Biomedical Informatics*, vol. 54, pp. 77–84, Apr. 2015, doi: 10.1016/j.jbi.2015.01.010.

[9] S. M. Castro *et al.*, "Automated annotation and classification of BI-RADS assessment from radiology reports," *Journal of Biomedical Informatics*, vol. 69, pp. 177–187, May 2017, doi: 10.1016/j.jbi.2017.04.011.

[10] S. Bozkurt, F. Gimenez, E. S. Burnside, K. H. Gulkesen, and D. L. Rubin, "Using automatically extracted information from mammography reports for decision-support," *Journal of Biomedical Informatics*, vol. 62, pp. 224–231, Aug. 2016, doi: 10.1016/j.jbi.2016.07.001.

[11] D. A. Sippo *et al.*, "Automated extraction of BI-RADS final assessment categories from radiology reports with natural language processing," *Journal of Digital Imaging*, vol. 26, no. 5, pp. 989–994, Oct. 2013, doi: 10.1007/s10278-013-9616-5.

[12] T. A. Patel *et al.*, "Correlating mammographic and pathologic findings in clinical decision support using natural language processing and data mining methods," *Cancer*, vol. 123, no. 1, pp. 114–121, Jan. 2017, doi: 10.1002/cncr.30245.

[13] D. Al-Saleh, M. Bin Al-Amir, and S. Larabi-Marie-Sainte, "SNAD Arabic dataset for deep learning," in *Advances in Intelligent Systems and Computing*, vol. 1250 AISC, 2021, pp. 630–640.

[14] S. Alotaibi, R. Mehmood, I. Katib, O. Rana, and A. Albeshri, "Sehaa: a big data analytics tool for healthcare symptoms and diseases detection using twitter, apache spark, and machine learning," *Applied Sciences (Switzerland)*, vol. 10, no. 4, p. 1398, Feb. 2020, doi: 10.3390/app10041398.

[15] A. Sahl and S. Hasan, "Radiology reports automated annotation performance: rule-based machine learning Vs deep learning," *IET Conference Proceedings*, vol. 2020, no. 6, pp. 433–436, Jun. 2020, doi: 10.1049/icp.2021.0893.

[16] Colert *et al.*, "Natural language processing (almost) from scratch," *Journal of Machine Learning Research*, vol. 12, no. Aug, pp. 2493–2537, 2011, [Online]. Available: http://www.jmlr.org/papers/v12/collobert11a.html.

[17] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences," in *52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014 - Proceedings of the Conference*, 2014, vol. 1, pp. 655–665, doi: 10.3115/v1/p14-1062.

[18] G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, and C. Dyer, "Neural architectures for named entity recognition," *2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT 2016 - Proceedings of the Conference*, pp. 260–270, 2016, doi: 10.18653/v1/n16-1030.

[19] A. N. Jagannatha and H. Yu, "Bidirectional RNN for medical event detection in electronic health records," in *2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT 2016 - Proceedings of the Conference*, 2016, pp. 473–482, doi: 10.18653/v1/n16-1056.

[20] Z. Liu *et al.*, "Entity recognition from clinical texts via recurrent neural network," *BMC Medical Informatics and Decision Making*, vol. 17, no. S2, p. 67, Jul. 2017, doi: 10.1186/s12911-017-0468-7.

[21] J. A. Fries, "Brundlefly at SemEval-2016 task 12: recurrent neural networks vs. joint inference for clinical temporal information extraction," *SemEval 2016 - 10th International Workshop on Semantic Evaluation, Proceedings*, pp. 1274–1279, 2016, doi: 10.18653/v1/s16-1198.

[22] S. K. Sahu and A. Anand, "Recurrent neural network models for disease name recognition using domain invariant features," *54th Annual Meeting of the Association for Computational Linguistics, ACL 2016 - Long Papers*, vol. 4, pp. 2216–2225, 2016, doi: 10.18653/v1/p16-1209.

[23] J. Li, A. Sun, J. Han, and C. Li, "A survey on deep learning for named entity recognition," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 1, pp. 50–70, 2022, doi: 10.1109/TKDE.2020.2981314.

[24] "Brat rapid annotation tool," 2012. http://brat.nlplab.org/ (accessed Nov. 08, 2012).
[25] A. Mansouri, L. S. Affendey, and A. Mamat, "Named entity recognition approaches," *Journal of Computer Science*, vol. 8, no. 2, pp. 339–344, 2008.
[26] P. Kluegl, M. Toepfer, P.-D. Beck, G. Fette, and F. Puppe, "UIMA ruta: rapid development of rule-based information extraction applications," *Natural Language Engineering*, vol. 22, no. 1, pp. 1–40, Jan. 2016, doi: 10.1017/S1351324914000114.
[27] Y. An, X. Xia, X. Chen, F. X. Wu, and J. Wang, "Chinese clinical named entity recognition via multi-head self-attention based BiLSTM-CRF," *Artificial Intelligence in Medicine*, vol. 127, p. 102282, May 2022, doi: 10.1016/j.artmed.2022.102282.

BIOGRAPHIES OF AUTHORS

Ahmed Sahl he earned his B.Sc. in computer science from Sudan University for Science and Technology, Sudan, in 2006, followed by an M.Sc. in computer science from University of Khartoum in 2012. Now he is a Ph.D. candidate in computer science at Faculty of Computing, Universiti Teknologi Malaysia. He currently serves as lecturer at King Khalid University. His professional focus is firmly anchored in machine learning, data analysis, and NLP. This dedication is evident in his continuous research, where he explores emerging areas of technology. He remains at the forefront of innovation, with a particular interest in advanced fields like deep learning and NLP. His commitment to staying up-to-date with the latest developments highlights his drive to push technological boundaries and make meaningful contributions to the rapidly evolving tech landscape. He can be contacted at email: akaldin@kku.edu.sa.

Dr. Shafaatunnur Hasan received a bachelor's degree in computer science (artificial intelligence) from Universiti Malaya, a master's degree in computer science, and a Ph.D. in GPU-based machine learning from Universiti Teknologi Malaysia. She worked at Shimano Components Malaysia for five years as a planning and analytics officer dealing with fishing tackle components. She is researching GPU-based Hadoop technology for big data and Pelangi deep learning analytics for big data research and development. She has developed GPUMLiB: machine learning library for big data processing, an open-source library using Cuda language to deal with Nth dimensional space under GPU platform (http://gpumlib.sourceforge.net/), with professor Noel Lopes from Portugal. Her experiences in solving industry problems ranging from fundamental issues to practical applications, such as polymer degradation analytics, retail sport analytics, desk help service analytics, malware analytics, and others. She has published in many journals and conference papers. Her research interests include big data computing, machine learning, soft computing, Cuda programming, Hadoop engineering, and big data platforms. She is also the managing editor of Scopus-Index Journal, International Journal of Advances in Soft Computing and Its Applications (http://www.home.ijasca.com), and reviewer of many international reputable journals related to her field. She can be contacted at email: Shafaatunnur@utm.my.

Dr. Maie M. Aboghazalah she earned her B.Sc. in Computer Science from El, Menoufia University, Egypt, in 2006, followed by an M.Sc. from Ain Shams University in 2017. After that she received her Ph.D. in computer science from the Faculty of Science, Menufia University, Egypt in 2022. She currently serves as an assistant professor and data analyst at Elnahda University. Her professional pursuits are deeply rooted in machine learning, data analysis, and computer vision. His dedication to these fields is reflected in his ongoing research endeavors, where he continually explores new dimensions of technology. Moreover, he remains at the forefront of innovation and is deeply interested in cutting-edge domains such as deep learning and computer vision (CV). This commitment to staying abreast of the latest advancements underscores his dedication to pushing the boundaries of technology and contributing significantly to its ever-evolving landscape. She can be contacted at email: eng_maie@yahoo.com.