

Available online at www.qu.edu.iq/journalcm

JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS

ISSN:2521-3504(online) ISSN:2074-0204(print)



Adaptive Fusion of DeBERTa-Base and MC-CNN for Enhanced Text Classification by Using AG News Dataset

Ali Jaber Tayh Albderi^{a*}, Rasha Falah Kadhem^b

^aComputing Department, College of Computer Science and Information Technology, University of AL-Qadisiyah, Iraq. ali.jaber@qu.edu.iq

^bComputing Department, College of Computer Science and Information Technology, University of AL-Qadisiyah, Iraq. rasha.kadhem@qu.edu.iq

ARTICLE INFO

Article history:

Received: 00 /mm/2026

Revised form: 00 /00/2026

Accepted : 27 /09/2026

Available online: 30 /09/2026

Keywords:

Natural Language Processing (NLP), MC-CNN and DeBERTa-base.

ABSTRACT

Text classification are an significant task in natural language processing(NLP). Different deep learning models can capture different types of information from text. This study proposes a new adaptive fusion method that combines MC-CNN and DeBERTa-base for text classification. MC-CNN are used to obtain local text features while DeBERTa-base captures contextual and semantic information. An adaptive fusion mechanism is proposed to dynamically combine the outputs for both models and improve classification performance. The presented model achieved the best results, with 95.62% accuracy, 95.01% precision, 94.38% recall, and 94.69% F1-score. The aim of this study is develop a simple and effective deep learning framework that combines the complementary strengths of CNN and Transformer-based models for improved text classification.

MSC..

<https://doi.org/10.29304/jqcm.2026.18.33219>

1. Introduction

Text classification are an significant task in natural language processing(NLP)[1] [2]. It is designed to automatically classify text documents to specified categories[3]. It has many applications such as news classification statement analysis topic detection and information retrieval[4][2]. With the growth of online text, accurate and efficient text classification methods have become increasingly important[5][6].

1.1. Motivation

Deep learning has improved text classification through enable models to automatically identify useful features from text[7][8]. Convolutional neural networks (CNN) can efficiently identify local patterns and important word combinations[9][1]. In contrast, Transformer-based models can understand the context and relationships between words in a sentence[10][11]. Therefore, these models can provide different and complementary types of information[12][13].

1.2. Proposed solution

In this study, we use Multi-Channel Convolutional Neural Network (MC-CNN) and DeBERTa-base for text classification. MC-CNN is used to extract local textual features, while DeBERTa-base is used to capture contextual and semantic information. To combine the strengths of both models, we present a new Adaptive Fusion Mechanism which dynamically determines this contribution of each model during classification.

1.3. Evaluation Strategy

The suggested method is tested on the AG News data set. The performance of MC-CNN, DeBERTa-base and the proposed adaptive fusion model is compared using common performance metrics including accuracy, precision, recall, and F1-score. The main goal of this research is to enhance text classification by effectively combining local and contextual features from two different deep learning models.

1.4. Paper Organization

The structure of the article organized as the following: section 2 review the related work. Section 3 describe the suggest model. Section 4 present the experiment and result. Finally, Section 5 present Conclusion.

2. Related Work

Victor Kwaku Agbesi et al[14]Has discussed the existing text classification models may fail to capture information at different levels. Single CNN models canmiss longer dependencies, while transformer models can require highcomputational resources. The problem is also harder for multilingual and low-resource languages such as Ewe, Arabic, and Urdu The authors proposed MuTCELM which combines five transformer models with a proposed Multi-TextCNN. Different CNN kernels and the five classifiers are combined using a weighted voting method. The presented model was tested based on Accuracy, Precision, Recall, and F1. Moreover, statistical analysis using confidence intervals and significance testing was also performed. Use Five datasets: AJGT (Arabic sentiment), ALJ-News (Arabic news), ArSarcasm-V2 (Arabic sarcasm), Ewe (Ewe news), and Urdu (Corpus Urdu fake-news detection). The best total result was obtained on the Ewe dataset, where MuTCELM achieved 97.7% accuracy, 96.9% precision, 97.3% recall, and 97.1% F1-macro. Across the five datasets, accuracy was 93.0% (AJGT), 94.9% (ALJ-News), 95.7% (ArSarcasm-V2), 97.7% (Ewe), and 91.3% (Urdu).(The proposed model has used five models, which may increase the computational cost.

Chrifi Alaoui Mehdi et al[15] Existing hybrid models often use simple fusion methods such as concatenation or averaging. These methods cannot adapt the significance of local and global features to input text. CNNs may miss long-range dependencies, while attention models may miss important local patterns. The authors proposed GEODENSECONVNET, which combines a multiscale CNN and multihead for global features. A new Adaptive Geometric Fusion mechanism based on Bregman divergence and an adaptive element wise geometric mean is used to combine the two representations. The performance of the models was assessed using Accuracy, Precision, Recall, and F1-score. The paper remember IMDB, AG News, and EmoBank as benchmark datasets. The model achieved the best reported performance with 80.12% accuracy and 0.8011 F1. the popularization of the proposed fusion method to other datasets and NLP tasks is not fully demonstrated.

Pengcheng He et al[16] BERT combines word content and position in one representation, which may limit its ability to model the relationship between content and relative position. Also, standard BERT does not use absolute position information effectively during masked-token decoding. The authors propose DeBERTa, using two main ideas: Disentangled Attention and an Enhanced Mask Decoder (EMD). They also introduce Scale-invariant Fine-Tuning (SiFT) to improve generalization. DeBERTa was evaluated on GLUE, SuperGLUE, SQuAD v1.1/v2.0, RACE, ReCoRD, SWAG, CoNLL-2003 NER, and WikiText-103. Use datasets: OpenWebText, Stories and WikiText-103. The DeBERTa ensemble achieved 90.3% macro-average on SuperGLUE, exceeding the reported human baseline of 89.8%. The single DeBERTa 1.5B model achieved 89.9%, also exceeding the human baseline. On GLUE, DeBERTa-Large achieved an average score of 90.00, outperforming BERT-Large and RoBERTa-Large.

Rasha Kadhem et al[17]Have discussed the Arabic sentiment analysis has received less attention than English, and Arabic text has challenges such as morphology, slang, and dialect variation. This article has a study that compares CNN, RNN, and LSTM using three feature extraction methods: BoW, TF-IDF, and word embedding. It also uses three classification schemes: rating, polarity, and balance polarity. The presented model was tested based on using Precision, Recall, F1-score, and Accuracy. The dataset was LABR (Large-Scale Arabic Book Reviews). The best reported value is 86.09%. The paper's main conclusion states that CNN was the most effective overall among CNN, LSTM, and RNN for the task.

Xiang Zhang et al[18] Traditional text classification mainly depends on words and hand-crafted features. The authors aimed to classify text directly from raw characters .without relying on word-level features or linguistic knowledge. The authors propose deep Character-Level Convolutional Neural Network (Char-CNN) that converts characters into

one-hot vectors and uses convolution, Max pooling and fully connected layers to classify text. Compared with bag of words, TF-IDF, n-grams, Bag-of-Means, LSTM, and word-based CNN models using large-scale text classification datasets. Use datasets AG's News, DBpedia, Sogou News, Yelp Review Polarity, Full, Amazon Review, Yahoo! Answers and Amazon Review. The best Character-CNN result varied by dataset. For example, the large Char-CNN with thesaurus augmentation achieved 4.88% error on Yelp Polarity (95.12% accuracy), while on Amazon Polarity it achieved 4.93% error (95.07% accuracy) and AG News Polarity it achieved 9.51% error (90.49% accuracy). The study showed that Character-CNN alone does not consistently outperform all word-based models.

3. Proposed Models

Although previous studies have combined DeBERTa-base with Multi-TextCNN for text classification, the existing approaches mainly focus on ensemble learning and conventional model combination. This section presents the proposed models for text classification. It describes the MC-CNN model for extracting local textual features and the DeBERTa-base model for capturing contextual and semantic information. Moreover, an Adaptive Fusion Mechanism is proposed to combine the outputs of both models dynamically and improve the final classification performance. This study addresses this gap by proposing an adaptive fusion framework for text classification on AG News dataset.

3.1. Multi-Channel Convolutional Neural Network (MC-CNN)

The first part of the proposed approach is the MC-CNN utilized to extract important local feature for the input text. The model applies different convolution filters to detect useful word patterns and local relationships between words. These features help the model identify important information related to the text class [14].

3.2. DeBERTa-base

The second part of the proposed approach is DeBERTa-base. It is a Transformer-based model that is used to capture contextual and semantic information from the input text. DeBERTa-base can understand the meaning of words according to their context and their position in the text. Therefore, it provides a different type of information compared with MC-CNN [19].

3.3. Confidence-based Adaptive Fusion Mechanism

The third part is the proposed Confidence based Adaptive Fusion Mechanism. This mechanism combines the outputs of MC-CNN and DeBERTa-base to obtain a final classification result. Instead of using fixed weights for the two models, the proposed mechanism dynamically determines the contribution of each model according to the input text. This allows the system to give more importance to the model that provides more useful information for a particular input.

3.4. Proposed Operating Framework

The complete proposed framework consists of three main stages. First, the input text is processed by MC-CNN to extract local textual features. At the same time, the same input text is processed by DeBERTa-base to obtain contextual and semantic features. Second, the outputs of both models are passed to the Adaptive Fusion Mechanism. Finally, the mechanism combines the two outputs and produces the final text classification result. The overall operation of the proposed system is illustrated in the proposed framework in Figure (1). The Adaptive Fusion Mechanism is defined by the following equations:

$$W_C = \frac{\max(P_C)}{\max(P_C) + \max(P_D)} \quad (1)$$

$$W_D = \frac{\max(P_D)}{\max(P_C) + \max(P_D)} \quad (2)$$

Where:

W_C : the adaptive weight assigned to the MC-CNN model.

W_D : the adaptive weight assigned to the DeBERTa-base model.

P_C : the probability vector produced by MC-CNN.

P_D : the probability vector produced by DeBERTa-base.

$$P_{AF} = W_C P_C + W_D P_D \quad (3)$$

Where PAF is the final probability vector produced by the Adaptive Fusion Mechanism.

Then

$$Y = \arg \max(P_{AF}) \quad (4)$$

Where: Y Final Classification for Adaptive Fusion.

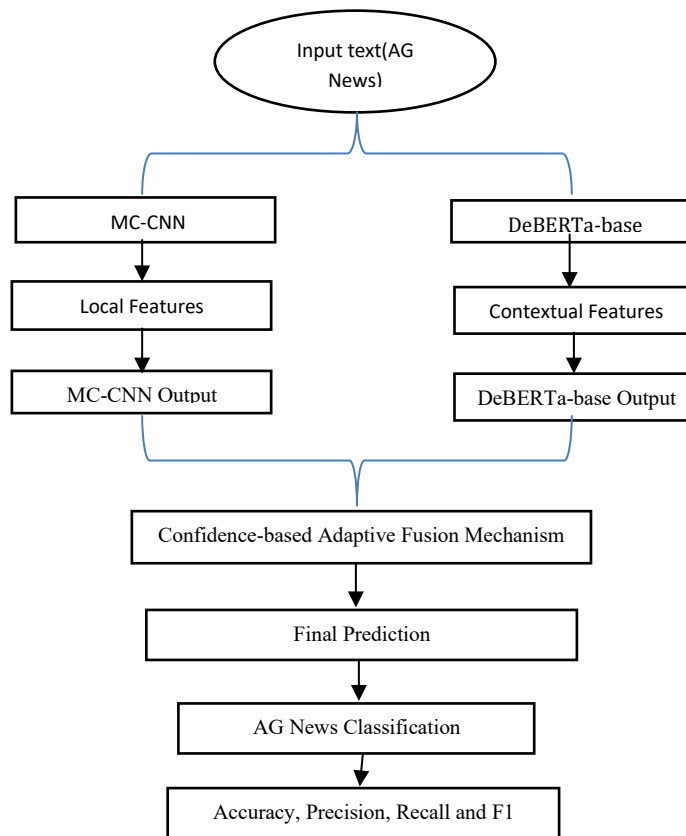


Fig 1: Proposed operating framework based on MC-CNN, DeBERTa-base, and the Adaptive Fusion Mechanism.

4. Experiments And Result

This section presents the experimental setup and the evaluation results of the proposed model. The experiments were conducted using the AG News dataset to evaluate the performance of MC-CNN, DeBERTa-base, and the proposed Adaptive Fusion Mechanism.

4.1. Dataset

The AG News corpus dataset was used as the benchmark dataset for evaluating the proposed text classification model. The dataset contains 120,000 samples of training and 7,600 samples of testing divided into four categories: World, Business, Sports and Sci/Tech [18]. AG News is suitable and sufficiently large for the task study. The main dataset is used widely in text classification study to evaluate traditional deep learning.

4.2. Evaluation metrics

To evaluate the performance of the proposed model four commonly used evaluation matrices are considered: Accuracy, Precision, Recall and F1 score. This metrics provide a comprehensive evaluation of the classification performance on the AG News dataset.

- Accuracy measures the proportion of classified samples among all samples[20]. It is calculated as:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

- Precision measure is proportion of correctly predicted positive samples among all samples predicted as positive[21]:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (6)$$

- Recall measures the proportion of correctly predicted positive samples among all actual positive samples[22]:

$$\text{Recall} = \frac{TP}{TP+FN} \quad (7)$$

- F1- Score represents the harmonic mean of precision and recall[23]:

$$\text{F1-Score} = \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8)$$

4.3. Analysis and Comparisons

Table 1, further explained in Fig 2 shows the performance of MC-CNN, DeBERTa-base, and the proposed Adaptive Fusion model on the AG News dataset. The MC-CNN model achieved 92.40% accuracy, while DeBERTa-base achieved 94.73%. The proposed Adaptive Fusion model obtained the best performance, by 95.62% Accuracy, 95.01% Precision, 94.38% Recall, and 94.69% F1-score. Compared with MC-CNN, the proposed model improves accuracy by 3.22 percentage points, while it improves accuracy over DeBERTa-base by 0.89 percentage points. This improvement indicates that combining local features from MC-CNN with contextual and semantic information from DeBERTa-base can provide more effective text representations than using either model independently. For comparison with previous work, Table 2 and further explained in Fig 3 show that Zhang et al. (2015) reported a character-level CNN error rate of 9.51% on AG News, corresponding to approximately 90.49% accuracy. Their results demonstrate the effectiveness of character-level CNNs, but the proposed model suggest a higher accuracy of 95.62% on the same database.

Table1: The result of metrics of the designed model with standalone MC-CNN model and the standalone DeBERTa-base model of AG news database.

Model	Accuracy	Precision	Recall	F1-Score
MC-CNN Model	92.40%	92.1%	91.9%	92.0%
DeBERTa-base Model	94.73%	93.9%	93.1%	93.9%
PROPOSED MODEL	95.62%	95.01%	94.38%	94.69%

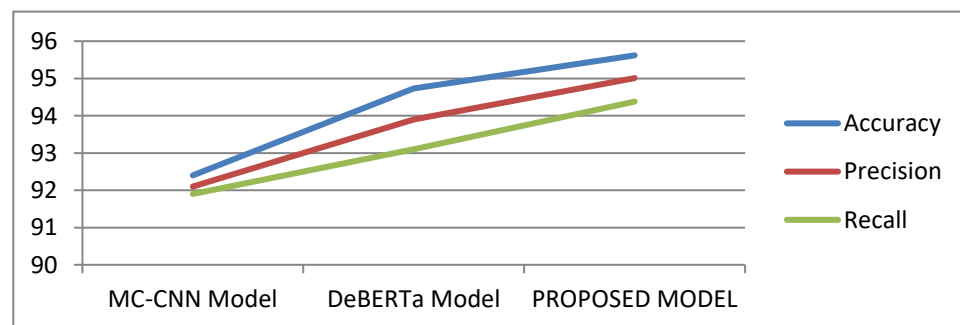
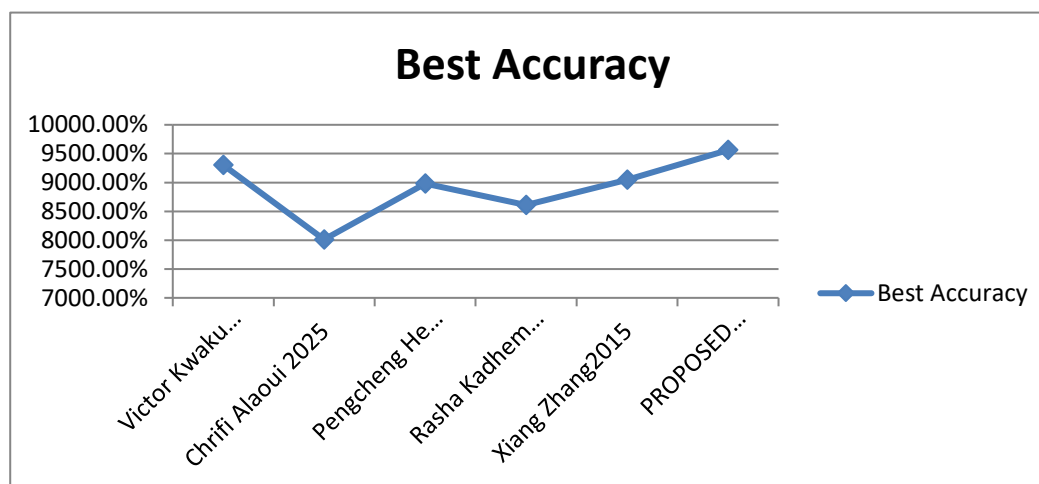


Fig2: The outcome metrics of the designed model with standalone MC-CNN model and the standalone DeBERTa-base model of AG news dataset.

Table2: Comparison between the proposed model and previous studies for AG news dataset.

Reference	Models	Best Accuracy
Victor Kwaku Agbesi 2024	proposed MuTCELM model which combines five transformer models[15]	93.0%
Chrifi Alaoui Mehdi 2025	authors proposed GEODENSECONVNET model which combines a CNN with multi-head for global[16]	80.12%
Pengcheng He 2021	authors propose DeBERTa with using two main ideas[17]	89.8%
Rasha Kadhém 2024	study that compares CNN, RNN, and LSTM using three feature extraction methods: BoW, TF-IDF, and word embedding[18]	86.09%
Xiang Zhang2015	A deep Character-Level Convolutional Neural Network (Char-CNN) that converts characters into one-hot vectors[19]	90.49%
PROPOSED MODEL	comparing MC-CNN, DeBERTa-base, and the proposed Adaptive Fusion model.	95.62

**Fig3: Comparison between the proposed model and previous studies for AG news dataset**

5. Conclusion

This study proposed an Adaptive Fusion framework that combines MC-CNN and DeBERTa-base for text classification. MC-CNN extracts local textual patterns, while DeBERTa-base captures contextual and semantic information. The Adaptive Fusion mechanism dynamically combines the outputs of both models. Evaluation results for the AG News database show the proposed system achieved 95.62% accuracy, 95.01% precision, 94.38% recall, and 94.69% F1-score, outperforming both standalone MC-CNN and DeBERTa-base models. These results indicate that adaptive fusion can effectively exploit the complementary strengths of CNN-based and Transformer-based models. Overall, the proposed framework provides a simple and effective approach for improving text classification performance by dynamically integrating local and contextual information.

References

- [1] Z. M. Koli, J. Alam, Z. Sultana, and A. A. Khan, "BanglaVerb : A sentence-level dataset for transitivity classification in Bangla NLP," *Data Br.*, vol. 65, p. 112605, 2026, doi: 10.1016/j.dib.2026.112605.
- [2] E. Boitel, A. Mohasseb, and E. Haig, "MIST : Multimodal emotion recognition using DeBERTa for text , Semi-CNN for speech , ResNet-50 for facial , and 3D-CNN for motion analysis," vol. 270, no. April 2024, 2025.
- [3] A. Assiri, A. Gumaei, F. Mahmood, T. Abbas, and S. Ullah, "DeBERTa-GRU : Sentiment Analysis for Large Language Model," no. M1, 2024, doi: 10.32604/cmc.2024.050781.
- [4] B. S. Al-smadi, "DeBERTa-BiLSTM : A multi-label classification model of Arabic medical questions using pre-trained models and deep learning," *Comput. Biol. Med.*, vol. 170, no. January, p. 107921, 2024, doi: 10.1016/j.combiomed.2024.107921.

- [5] A. Seerat, S. Nasir, M. Wasim, and N. M. Garcia, "ScienceDirect Biomedical Biomedical Natural Natural Language Language Inference Inference on on Clinical Clinical trials using using the the Models BERT-based Models," *Procedia Comput. Sci.*, vol. 241, no. 2019, pp. 576–581, 2024, doi: 10.1016/j.procs.2024.08.083.
- [6] Q. Zhang, X. Xiang, Y. Li, and Y. Hou, "ScienceDirect Machine Learning and NLP Fusion Algorithms for Short Text Classification," vol. 285, pp. 10–18, 2026, doi: 10.1016/j.procs.2026.07.165.
- [7] H. Black, R. Kurt, and O. Turan, "A domain-adapted NLP pipeline for human factors classification of maritime accidents," *Ocean Eng.*, vol. 357, no. P3, p. 125625, 2026, doi: 10.1016/j.oceaneng.2026.125625.
- [8] P. Guleria, "NLP based text classification using TF-IDF enabled fine-tuned long short-term memory : An empirical analysis," *Array*, vol. 27, no. July, p. 100467, 2025, doi: 10.1016/j.array.2025.100467.
- [9] T. Meyer, T. Schlosser, S. Moniri, M. Friedrich, D. Kowerko, and A. Sampath, "Engineering Applications of Artificial Intelligence Medical text generation and classification using large language models : A feasibility study in ophthalmology," *Eng. Appl. Artif. Intell.*, vol. 178, no. P2, p. 114962, 2026, doi: 10.1016/j.engappai.2026.114962.
- [10] M. Arslan and C. Cruz, "Leveraging NLP approaches to define and implement text relevance Leveraging NLP approaches to define and implement text relevance hierarchy framework for business news classification hierarchy framework for business news classification," *Procedia Comput. Sci.*, vol. 225, pp. 317–326, 2023, doi: 10.1016/j.procs.2023.10.016.
- [11] X. Ren, M. Jiang, and Y. Wang, "ScienceDirect Text Style Recognition and Classification Model with Multi- Feature Fusion," *Procedia Comput. Sci.*, vol. 285, pp. 366–374, 2026, doi: 10.1016/j.procs.2026.07.205.
- [12] M. Bourdin, A. Neumann, T. Paviot, R. Pellerin, and S. Lamouri, "A Framework for Selecting the Optimal NLP Solution for Classification Tasks in Industry 4.0 Based on Data and Business Constraints," *IFAC Pap.*, vol. 59, no. 10, pp. 1850–1855, 2025, doi: 10.1016/j.ifacol.2025.09.311.
- [13] I. Didier and R. Luiz, "Technovation AI-powered innovation tournament evaluations : A hybrid NLP approach," vol. 156, no. February, 2026, doi: 10.1016/j.technovation.2026.103622.
- [14] V. Kwaku, W. Chen, S. Banaamwini, C. C. Ukwuoma, Y. Hyeon, and M. A. Al-antari, "Heliyon MuTCELM : An optimal multi-TextCNN-based ensemble learning for text classification," *Heliyon*, vol. 10, no. 19, p. e38515, 2024, doi: 10.1016/j.heliyon.2024.e38515.
- [15] C. A. Mehdi, M. Oubrahim, N. Joudar, and M. Ettaouil, "GEODENSECONVNET : A BREGMAN DIVERGENCE-INSPIRED FRAMEWORK FOR ADAPTIVE FEATURE FUSION IN TEXT CLASSIFICATION," vol. 38, no. 11, pp. 1908–1919, 2025.
- [16] P. He et al., "D E BERT A : D ECODING - ENHANCED BERT ENTANGLED A TTENTION WITH D IS -," pp. 1–21, 2021.
- [17] S. Mallat and M. Zrigui, "ScienceDirect The optimization of deep learning classification through the The optimization of deep learning classification through the comparison between di fferent algorithm by using LABR comparison between di fferent algorithm by using," *Procedia Comput. Sci.*, vol. 246, pp. 3575–3584, 2024, doi: 10.1016/j.procs.2024.09.200.
- [18] X. Zhang, J. Zhao, and Y. Lecun, "Character-level Convolutional Networks for Text," pp. 1–9, 2015.
- [19] H. Sebbag and N. El Faddouli, "MTBERT-Attention : An Explainable BERT Model based on Multi-Task Learning for Cognitive Text Classification," vol. 21, 2023, doi: 10.1016/j.sciaf.2023.e01799.
- [20] J. F. C. M. A. Rizzardi, S. Sicari, and A. C. Porisini, "Computers & Security NERO : NEural algorithmic reasoning for zeRO-day attack detection in the IoT : A hybrid approach," *Comput. Secur.*, vol. 142, no. April, p. 103898, 2024, doi: 10.1016/j.cose.2024.103898.
- [21] E. E. Abdallah, W. Eleisah, A. F. Ootom, and I. I. Ahmed, "ScienceDirect ScienceDirect Intrusion Detection Systems using Supervised Machine Learning Techniques : survey Machine Learning Intrusion Detection Systems using A Supervised Techniques : A survey," *Procedia Comput. Sci.*, vol. 201, pp. 205–212, 2022, doi: 10.1016/j.procs.2022.03.029.
- [22] A. Zohourian, S. Dadkhah, H. Molyneaux, E. Carlos, P. Neto, and A. A. Ghorbani, "Computers & Security IoT-PRIDS : Leveraging packet representations for intrusion detection in IoT networks," *Comput. Secur.*, vol. 146, no. August, p. 104034, 2024, doi: 10.1016/j.cose.2024.104034.
- [23] Z. Wang, Y. Liu, D. He, and S. Chan, "Intrusion detection methods based on integrated deep learning model," vol. 103, 2021, doi: 10.1016/j.cose.2021.102177.