Context-Aware Deep Learning Feature Extraction for Domain-Independent Aspect-Based Sentiment Analysis Using BERT

***Sadineni Venugopal, **Dr. Kamakshaiah Kolli**

*Research Scholar, **Supervisor,

Department of Computer Science

Faculty of Computing & Information Technology

Himalayan University, Itanagar, Arunachal Pradesh, India

[1]*Received: 08 February 2025; Accepted: 15 April 2025; Published: 08 May 2025*

ABSTRACT

Aspect-Based Sentiment Analysis (ABSA) finds aspect terms in text and calculates the sentiment polarity for each aspect to offer fine-grained opinion mining. Real-world applications are still difficult because of contextual ambiguity, overlapping aspect expressions, domain-specific language, and noise in user-generated content, even though deep learning techniques have significantly improved ABSA performance. Strong contextual representation learning has been shown by transformer-based language models, especially Bidirectional Encoder Representations from Transformers (BERT). However, traditional fine-tuning techniques frequently rely only on the final hidden layer, which might not produce the best representations for aspect-level tasks. In order to improve domain-independent ABSA, this research suggests a context-aware deep learning feature extraction system based on BERT. By using a weighted layer pooling strategy to take advantage of multi-layer BERT representations, the suggested method enables the model to include linguistic data from various abstraction levels. To improve contextual aspect representations, an attention-based feature refinement module is also utilised. The suggested approach increases aspect extraction accuracy, lowers susceptibility to noisy inputs, and effectively generalises across domains, according to extensive studies carried out on several benchmark datasets. The findings verify that BERT-based context-aware feature extraction is appropriate for scalable and domain-independent ABSA systems.

1. Introduction

Aspect-Based Sentiment Analysis (ABSA), which enables in-depth examination of viewpoints expressed in online reviews, consumer feedback, and social media content, has emerged as a key task in natural language processing [1][2]. ABSA concentrates on identifying particular elements and figuring out the sentiment expressed towards each aspect, in contrast to conventional sentiment analysis techniques that provide a single polarity label to an entire document or sentence [3][4]. For applications like market analysis, recommendation systems, and customer experience management, this fine-grained viewpoint offers more useful insights.

Even with deep learning methods success ABSA is still difficult in practical situations. Semantic overlap, in which various lexical terms (such as "battery life" and "power backup") may refer to the same thing, is a significant problem. Because the sentiment attached to an element frequently depends on the surround ding context, contextual ambiguity makes the task much more difficult. Furthermore, loud and informal language in user-generated text negatively impacts model robustness, and domain-specific variance restricts the transferability of models trained on a particular domain [5].

.A potent pre-trained language model that can capture bidirectional contextual dependencies is called Bidirectional Encoder Representations from Transformers (BERT). A task-specific output layer is added to the final hidden representation in the majority of downstream applications to refine BERT. Nevertheless, BERT is made up of several

[1] *How to cite the article:* Venugopal S., Kolli K (2025); Context-Aware Deep Learning Feature Extraction for Domain-Independent Aspect-Based Sentiment Analysis Using BERT; *International Journal of Technology, Science and Engineering*; Vol 8 Issue 2; 18-26

transformer layers, each of which encodes a different level of linguistic information [6][7]. Higher layers concentrate on task-agnostic semantic abstractions, intermediate layers encode contextual linkages, and lower layers record lexical and grammatical patterns. Therefore, for aspect-based sentiment tasks, depending solely on the last layer may limit the expressiveness of the learnt representations. According to recent research, intermediary BERT layers frequently offer more discriminative and transferable properties for downstream applications. Inspired by this finding, this work presents a weighted layer pooling technique that builds richer contextual representations by combining concealed states from several BERT layers [8]. The model dynamically highlights the most informative representations for ABSA by giving each layer learnable weights. Additionally, aspect-specific contextual interactions are improved through the use of an attention-based feature refining module

2. Literature Review

- BERT for Sentiment Based on Aspects In order to facilitate domain-independent ABSA, we present a BERT-based context-aware feature extraction system in this research. This paper makes three main contributions: (1) the incorporation of weighted layer pooling to take advantage of multi-level BERT representations; (2) the application of attention mechanisms to improve aspect-aware features; and (3) thorough testing on several datasets to show improved cross-domain generalisation, reduced noise sensitivity, and robustness. For practical ABSA applications, the suggested method provides a scalable and efficient solution.

- The application of Bidirectional Encoder Representations from Transformers (BERT) for Aspect-Based Sentiment Analysis was investigated by Hoang, Tuan, Pham, and Nguyen (2019). Their research showed that BERT improves out-of-domain sentiment classification performance by efficiently capturing contextual word representations. The authors obtained competitive results on SemEval ABSA benchmark datasets by fine-tuning the pre-trained BERT model with supplementary text creation objectives. In comparison to conventional neural models, this work offered preliminary empirical evidence that BERT's deep bidirectional encoding generates richer aspect and sentiment representations [9].

- The contribution of hidden representations in pre-trained BERT models on ABSA performance was thoroughly examined by Xu, Liu, Shu, and Yu (2020). Their results showed that while certain self-attention heads largely concentrate on domain-level semantics, others exclusively encode aspect-related and opinion-oriented information. Instead than depending solely on the final transformer layer during fine-tuning, the authors stressed the significance of utilizing multi-layer representations [10]

- To better utilize intermediate BERT layer representations for ABSA, Karimi, Rossi, Prati, and Freitas (2021) suggested augmentation modules, such as Parallel Aggregation and Hierarchical Aggregation. Their method showed that, when properly aggregated, intermediate transformer layers collect important linguistic and contextual information that greatly enhances both aspect extraction and sentiment classification [11]

- Abdel wad, Moussa, and Badr (2021) investigated the application of BERT to Aspect-Based Sentiment Analysis in Arabic. Their experiments showed that contextualized BERT embeddings consistently outperform static word representations such as Word2Vec across multiple domain-specific datasets. This study highlights the effectiveness of pre-trained transformer models in supporting fine-grained sentiment analysis for low-resource and morphologically rich languages [12]

- Zhang, Li, and Zhou (2024) analysed over 700 ABSA-related studies in a thorough systematic analysis that was published in Artificial Intelligence analysis. According to their poll, there has been a noticeable trend toward transformer-based architectures and deep learning, with BERT becoming the most popular representation model. The authors also pointed out enduring issues with current ABSA techniques, including domain mismatch, dataset bias, and generalization constraints [13]

- The efficacy of optimized transformer variations, such as Distil BERT and Roberta, for ABSA tasks was investigated by Li, Chen, and Wang (2023). According to their findings, these lighter-weight models dramatically reduce computational overhead while achieving accuracy and F1-score performance that is on par with or better than normal BERT models. The adaptability and scalability of transformer-based representations for sentiment analysis are highlighted in this work [14].

- Drasković, Vasilyevich, Panic, and Marković (2025) used ABSA to extract fine-grained sentiment at the aspect level from extensive microblog data. In order to balance local and global contextual information and enhance aspect-specific sentiment prediction, they combined BERT embeddings with Context Dynamic Masked (CDM) and Context Dynamic Weighted (CDW) layers. Their research shows how well context-aware transformer models handle informal social media text. The findings demonstrate the usefulness of hybrid ABSA pipelines for real-world datasets by highlighting the predominately negative sentiment across COVID-19-related elements [15]

- A hybrid ABSA framework for Hindi was presented by Singh, Verma, and Gupta (2023). It combines linguistic elements like part-of-speech tags with multilingual BERT embeddings. Their findings showed enhanced aspect extraction accuracy in low-resource language environments, demonstrating transformer models' versatility and the need of integrating linguistic knowledge into deep learning frameworks [16]

- ABSA, Liu, Zhao, and Sun (2024) presented a multi-layer augmented graph convolutional network that combines biaffine attention mechanisms with BERT embeddings. Their model improves performance on certain ABSA subtasks by efficiently capturing structured aspect-context interactions. This field of study demonstrates how graph-based and attention-driven modelling might enhance transformer representations for more profound semantic comprehension [17]

3. Problem Statement

Even while transformers like BERT have been successful in ABSA, a number of issues still exist:

- Suboptimal representation selection: Because intermediate layers frequently encode important syntactic or semantic information, using only the final hidden layer may restrict model performance. CSDN Blog • Domain bias: Sentiment models that are optimized for one domain frequently don't perform well when applied to datasets from other domains.

- Text noise and ambiguity: User-generated content degrades performance by introducing misspellings, acronyms, and contextual ambiguity. Therefore, to develop a domain-agnostic ABSA model, an effective, context-rich representation training approach that fully utilizes multi-layer transformer characteristics is required.

4. Proposal Work

Aspect-Based Sentiment Analysis (ABSA) links opinion statements to particular components inside a sentence in order to capture fine-grained sentiment information. Transformer-based models, like BERT, are very useful for this purpose since they describe bidirectional token dependencies and produce rich contextualized embeddings. A single representation, like the [CLS] token or the final hidden layer, is frequently used in conventional fine-tuning techniques, which may not be able to fully utilize the variety of contextual information encoded across several layers. Token-level representations are aggregated by context-aware feature extraction approaches, such as pooling and attention processes, to provide robust aspect embeddings that are less susceptible to noise, domain-specific terminology, and contextual ambiguity.

The multi-layer architecture of BERT naturally encodes a hierarchy of linguistic information: lower layers capture surface-level features, middle layers encode syntactic patterns, and higher layers represent deep semantic content. However, fine-tuning usually adds a task-specific output layer to the final transformer layer. The operation of hidden layers in the context of BERT for Aspect-Based Sentiment Analysis (ABSA) is depicted in the following diagram.

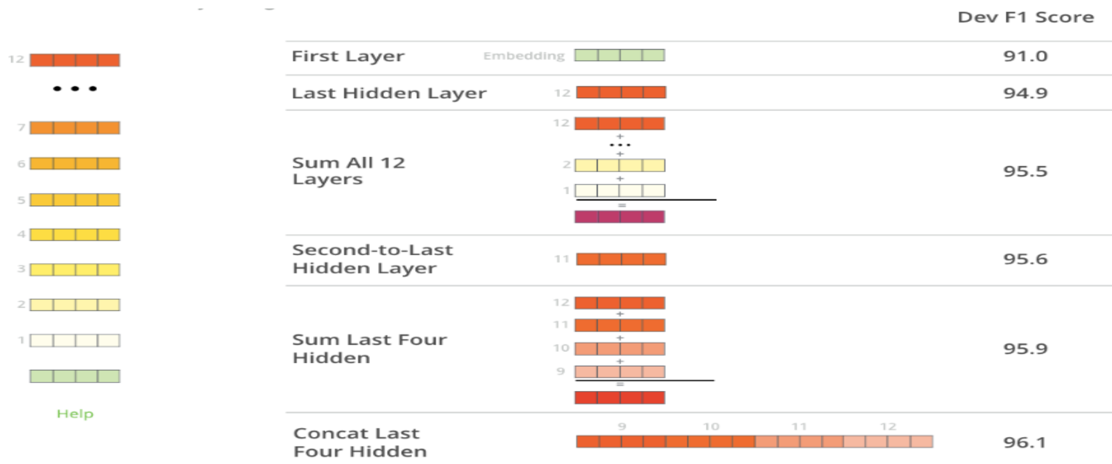

Fig 1: Different layers of BERT.

The diagram above shows that by capturing a variety of linguistic factors, concatenating BERT's final four layers improves performance. Relying just on the last layer may overlook important context since different layers encode different information. Richer, more complex token representations are made possible by multi-layer aggregation. This method enhances accurate, domain-independent, and semantically aware ABSA.

5. Working of Hidden Layers in BERT for ABSA for Features Extraction

- **Step 1:** Input Preparation

Tokenizing input sentences into smaller words is necessary for BERT, a transformer-based approach. The beginning of a sentence and the space between sentences are indicated by the special symbols [CLS] and [SEP]. BERT is better able to concentrate on the pertinent portion of the phrase when the aspect term is explicitly included. Segment IDs, make attention masks, and convert tokens to IDs.

Example: Sentence: "The battery life is amazing, but the camera is disappointing "Aspect: "battery life" Input tokens: [CLS] The battery life is amazing, but the camera is disappointing [SEP] battery life [SEP]

- **Step 2:** Embedding Layer

By adding token embeddings, positional embeddings, and segment embeddings, BERT transforms tokens into contextual embeddings. This enables the model to comprehend tokens' positions in the sequence as well as their meanings.

$$E = E^{\text{token}} + E^{\text{position}} + E^{\text{segment}}, E \in \mathbb{R}^{n\times d}$$

Where n = sequence length, and d = embedding dimension

- **Step 3:** Pass Through Transformer Layers

Each of the "L" stacked transformer layers in BERT has feed-forward networks and self-attention. In this case, lexical features are captured by lower layers, syntactic relations are captured by middle layers, and semantic and contextual information—which is essential for ABSA to comprehend aspect-specific sentiment—is captured by higher layers

$$H^{(l)} = \text{TransformerLayer}(H^{(l-1)}), l = 1,2,\dots,L$$

Where $H^{(l)}$ token embedding's at layer l.

Step 4: Extract Token-Level Representations

In BERT, the hidden state of each token is a contextualized representation that is acquired via self-attention throughout the sentence. The aspect representation for Aspect-Based Sentiment Analysis is calculated by averaging the aspect tokens' final-layer embeddings:

$$h_{\text{aspect}} = \frac{1}{k} \sum_{i \in \text{aspect tokens}} h_i^{(L)}$$

The special [CLS] token aggregates global sentence information through all transformer layers. Its final hidden state represents the sentence-level semantics:

$$h_{\text{CLS}} = h_1^{(L)}$$

Combining h_{aspect} and h_{CLS} enables effective aspect-specific sentiment prediction

- **Step 5:** Context-Aware Feature Aggregation

Token-level contextual embeddings are combined using pooling techniques to create a fixed-length vector that captures the impact of words surrounding the target aspect and provides the entire sentence context. Pooler output is the final layer hidden-state of the sequence's initial token (classification token), which is then processed by a Tanh activation function and a linear layer. During pretraining, the next sentence prediction (classification) goal is used to train the linear layer weights. The pooler receives the final hidden state and returns the pooled output. By setting the add pooling layer to False in the model configuration and submitting that to the model, we may deactivate pooler outputs.

Contextualized embeddings for every token generated by the final transformer layer are included in the final hidden state output of BERT. Each token representation uses self-attention to encode information from the entire sentence.

$$H^{(L)} = \{h_1^{(L)}, h_2^{(L)}, \dots, h_n^{(L)}\} \in \mathbb{R}^{n \times d}$$

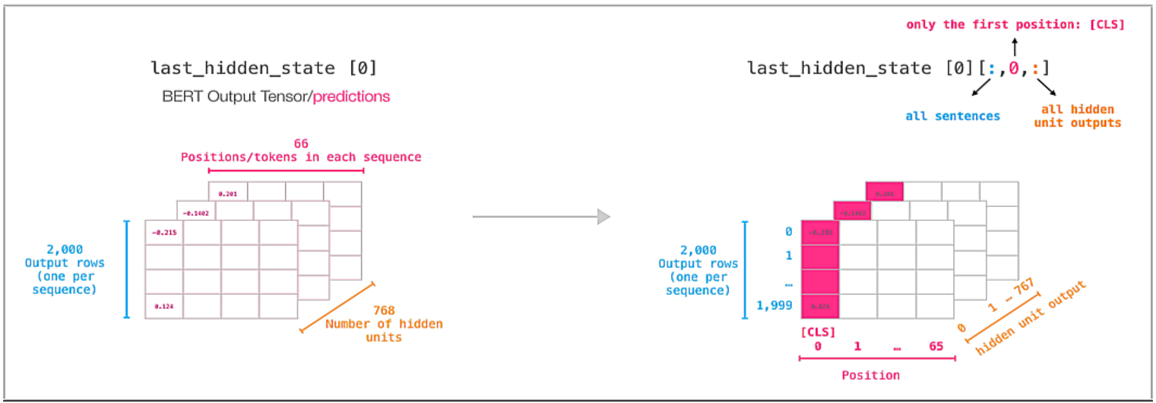

Fig-2: Final hidden state output of BERT.

The concept behind the [CLS]: Transformers contextual model is that a [CLS] token would have fully captured the context and would be adequate for straightforward downstream tasks like classification. Therefore, you can utilize [batch, hidden state] for applications like classification utilizing sentence representations. For classification tasks, the [CLS] token embedding, which is the first vector of the final hidden state, aggregates global sentence-level semantics.

$$h_{\text{CLS}} = h_1^{(L)} \in \mathbb{R}^d$$

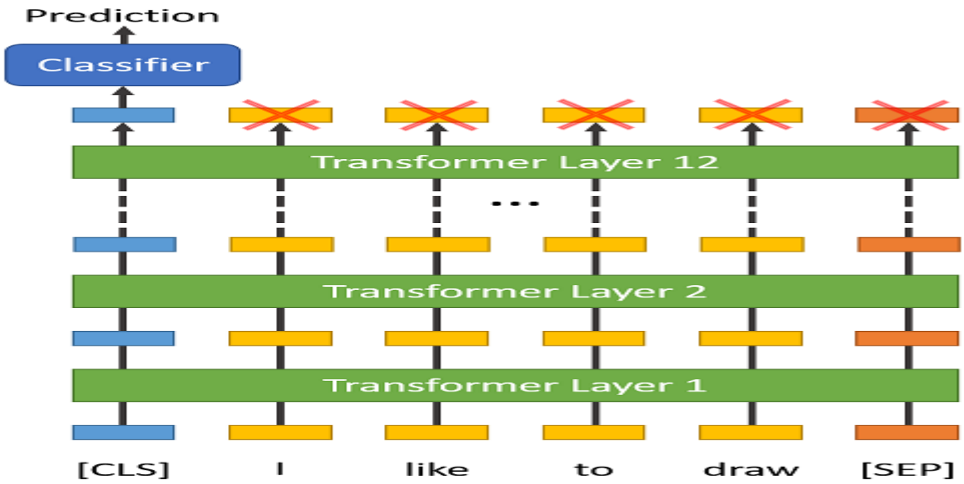

Fig-3: For classification tasks of [CLS] token embedding

Pooling layers are applied over token embeddings to provide fixed-length sentence representations. Mean pooling ignores padding tokens and averages contextual embeddings.

$$h_{\text{mean}} = \frac{\sum_i M_i\, h_i^{(L)}}{\sum_i M_i}$$

Max pooling selects the most salient feature in each embedding dimension:

$$h_{\text{max}}[j] = \max_i \left(h_i^{(L)}[j] \cdot M_i\right), j = 1, \ldots, d$$

Mean–Max pooling combines global and salient information by concatenating mean and max pooled vectors:

$$h_{\text{mm}} = [h_{\text{mean}}; h_{\text{max}}] \in \mathbb{R}^{2d}$$

Additionally, Conv1D pooling captures local n-gram features by applying convolution over consecutive token embeddings followed by max pooling:

$$c_i = f(W_c H_{i:i+k-1}^{(L)} + b_c), h_{\text{conv}} = \max_i (c_i)$$

Finally, the pooler output produces a sentence-level representation by applying a linear layer and *tanh* activation to the [CLS] embedding:

$$h_{\text{pool}} = \tanh(W_p h_{\text{CLS}} + b_p)$$

- **Following table is represented that notation, which are used in above para.**

Table 1: Notation of formula.

Notation	Meaning	Shape	Purpose
$h_i^{(L)}$	Token embedding of token i (last BERT layer)	$\mathbb{R}^d$	Contextual representation
h_{mean}	Mean pooling of all tokens	$\mathbb{R}^d$	Sentence-level feature
h_{max}	Max pooling of all tokens	$\mathbb{R}^d$	Salient features

Notation	Meaning	Shape	Purpose
h_{mm}	Mean-Max pooling	$\mathbb{R}^{2d}$	Rich combined context
h_{final}	Final feature vector for classification	$\mathbb{R}^{4d}$	Combines sentence, aspect, and pooled info
$H^{(L)}_{i:i+k-1}$	Token embedding's from i to i+k-1	$\mathbb{R}^{k\times d}$	Local n-gram features for Conv1D

- **Step 6:** Combine Features

Combining sentence-level, aspect-level, and pooled context embeddings provides a comprehensive feature vector for sentiment classification.

$$h_{\text{final}} = [h_{\text{CLS}}; h_{\text{aspect}}; h_{\text{mm}}]$$

- **Step 7:** Classification of BERT

The final feature vector is fed to a fully connected layer (MLP) followed by softmax to predict sentiment toward the aspect.

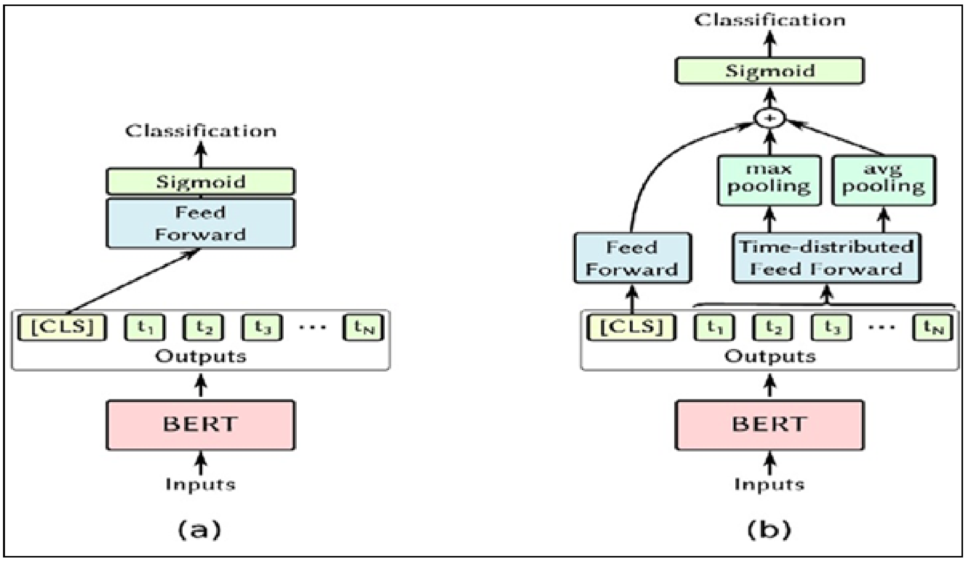

Fig-4: Classification of BERT

References

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

da Costa, K. A., Papa, J. P., Lisboa, C. O., Munoz, R., & de Albuquerque, V. H. C. (2019). Internet of Things: A survey on machine learning-based intrusion detection approaches. *Computer Networks, 151*, 147–157. https://doi.org/10.1016/j.comnet.2019.01.023

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). 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: Human Language Technologies, Volume 1 (Long and Short Papers)* (pp. 4171–4186). Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423

Drasković, V., Vasiljević, P., Penić, B., & Marković, I. (2025). Aspect-based sentiment analysis of user-generated content from a microblogging platform. *Journal of Big Data, 12*, Article 186. https://doi.org/10.1186/s40537-025-01244-0

Foster, D. (2019). *Generative deep learning: Teaching machines to paint, write, compose, and play*. O'Reilly Media.

Frustaci, M., Pace, P., Aloi, G., & Fortino, G. (2018). Evaluating critical security issues of the IoT world: Present and future challenges. *IEEE Internet of Things Journal, 5*(4), 2483–2495. https://doi.org/10.1109/JIOT.2017.2767291

García, S., Grill, M., Stiborek, J., & Zunino, A. (2015). An empirical comparison between generative and discriminative classifiers on a malware explanation task. *Virus Bulletin*, 1–8.

García-Magariño, I., Muttukrishnan, R., & Lloret, J. (2019). Human-centric AI for trustworthy IoT systems with explainable multilayer perceptrons. *IEEE Access, 7*, 125562–125574. https://doi.org/10.1109/ACCESS.2019.2937521

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Haria, S. (2019). The growth of the Hide and Seek botnet. *Network Security, 2019*(3), 14–17. https://doi.org/10.1016/S1353-4858(19)30030-5[1]

Hinton, G. E., Krizhevsky, A., & Sutskever, I. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*. https://doi.org/10.48550/arXiv.1207.0580

Hodo, E., Edinburgh, X., Branch, J. W., & Getachew, A. T. (2016). Threat analysis of IoT networks using artificial neural network intrusion detection system. In *2016 International Symposium on Networks, Computers and Communications (ISNCC)* (pp. 1–6). IEEE. https://doi.org/10.1109/ISNCC.2016.7750961

Hoang, M., Bihorac, O. A., & Rouces, J. (2019). Aspect-based sentiment analysis using BERT. In *Proceedings of the 22nd Nordic Conference on Computational Linguistics* (pp. 187–196). Linköping University Electronic Press. https://aclanthology.org/W19-6120

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning* (pp. 448–456). PMLR.

International Data Corporation (IDC). (2019). *The growth in connected IoT devices is expected to generate 79.4ZB of data in 2025, according to a new IDC forecast*.

Javaid, A., Niyaz, Q., Sun, W., & Alam, M. (2016). A deep learning approach for network intrusion detection system. In *Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS)* (pp. 21–26). ICST.

Kang, K., Pang, Z., Wang, C., & Zhou, L. (2013). Security and privacy mechanism for health internet of things. *The Journal of China Universities of Posts and Telecommunications, 20*(Suppl 2), 64–68. https://doi.org/10.1016/S1005-8885(13)60219-8[1]

Karimi, A., Rossi, L., & Prati, A. (2021). Improving BERT performance for aspect-based sentiment analysis. In *Proceedings of the 4th International Conference on Natural Language and Speech Processing (ICNLSP 2021)* (pp. 39–46). Association for Computational Linguistics. https://aclanthology.org/2021.icnlsp-1.5

Kolosnjaji, B., Eraisha, G., Webster, G., Zarras, A., & Eckert, C. (2016). Deep learning for classification of malware system call sequences. In *Australasian Joint Conference on Artificial Intelligence* (pp. 137–149). Springer. https://doi.org/10.1007/978-3-319-50127-7_12

Kumar, J. S., & Patel, D. (2014). A survey on internet of things: Security and privacy issues. *International Journal of Computer Applications, 90*(14), 20–26. https://doi.org/10.5120/15764-4454

Li, S., Lee, S. Y., Chen, Y., Huang, C.-R., & Zhou, G. (2010). Sentiment classification and polarity shifting. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010)* (pp. 635–643).

Li, Z., Qin, D., Carlson, M., & Lee, K.-H. (2018). VulDeePecker: A deep learning-based system for vulnerability detection. *arXiv preprint arXiv:1801.01681*. https://doi.org/10.48550/arXiv.1801.01681

Mane, S., & Rao, D. (2021). *Explaining network intrusion detection system using explainable AI framework*. arXiv. https://doi.org/10.48550/arXiv.2103.07110

Marino, D. L., Wickramasinghe, C. S., & Manic, M. (2018). An adversarial approach for explainable AI in intrusion detection systems. In *IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society* (pp. 4183–4188). IEEE. https://doi.org/10.1109/IECON.2018.8592143

Moustafa, N., & Slay, J. (2015). UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). In *2015 Military Communications and Information Systems Conference (MilCIS)* (pp. 1–6). IEEE. https://doi.org/10.1109/MilCIS.2015.7348942

Moustafa, N., Slay, J., & Turnbull, B. (2019). An ensemble intrusion detection technique based on proposed statistical flow features for protecting network traffic of Internet of Things. *IEEE Internet of Things Journal, 6*(3), 4815–4830. https://doi.org/10.1109/JIOT.2018.2875467

Pascanu, R., Stokes, J. W., Santhanam, S., & Turner, A. (2015). Malware classification with recurrent networks. In *2015 IEEE