



International Journal of Advanced Research in Education and Technology (IJARETY)

Volume 13, Issue 3, May - June 2026

Impact Factor: 9.971



Enhancing Adverse Event Prediction with NLP Based Data Augmentation in Imbalanced Datasets

B. Ratnamala¹, A. Pavani², G. Abhisarika³, G. Poojitha⁴

Assistant Professor, Department of CSE, Guru Nanak Institute of Technology, Hyderabad, Telangana, India¹

Students, Department of CSE, Guru Nanak Institute of Technology, Hyderabad, Telangana, India^{2,3,4}

ABSTRACT: This paper proposes a data augmentation method for imbalanced healthcare datasets. This method was inspired by a data augmentation method in natural language processing (NLP) that generates synthetic sentences for training by replacing some words with similar words. The proposed method generates synthetic patient records by replacing patient backgrounds with similar backgrounds. In this paper, the cosine similarity of the distributed representations was used as the similarity metric between patient backgrounds. The distributed representations of the patient backgrounds were generated by the Word2Vec model. To confirm the performance improvement with the proposed data augmentation method, the prediction performance of adverse events (AEs) caused by drug administration was experimentally evaluated on a real-world medical dataset with 4971 records. The combination of the proposed data augmentation method and a conventional under sampling method resulted in an 80.0% improvement in accuracy and a 40.0% improvement in the precision and F1-score. The multifaceted evaluation demonstrated that the proposed method is effective, especially for predicting AEs with positive ratios ranging from 1.0% to 2.1%, which are difficult to predict with conventional machine learning methods but should be predictable in the medical field.

I. INTRODUCTION

Recent technological developments have led to an increase in medical data analysis research. The prediction of adverse events (AEs) during drug administration is a morally and economically important issue. In the United States, AEs are estimated to account for up to 28% of emergency room visits and 5.0% of hospital deaths, and 30-150 billion dollars is spent annually on treating AEs. Drug safety is not only an obligation to patients but also has significant economic consequences. Over the past few decades, various approaches have been proposed for predicting AEs. In recent years, machine learning and deep learning methods have been proposed. Most studies on AE prediction have focused on the relationship between the drug and AEs, and few studies have considered patient backgrounds. However, AEs are known to depend on patient backgrounds, such as age, sex, and history of complications, in addition to the drug. Therefore, AEs should be predicted using both drug and patient background information. The datasets used for adverse event prediction are inherently imbalanced because of the rarity of diseases and patient diversity. An imbalanced dataset refers to a dataset in which the amount of data per label differs remarkably. Since frequently reported AEs are already known in the medical field, the prediction of low-frequency AEs is critical. However, when imbalanced datasets are used to train conventional machine learning algorithms, which typically assume large balanced datasets, the prediction performance of low-frequency labels is significantly reduced.

1.1 SCOPE OF THE PROJECT

This paper proposes a data augmentation method to address issues of predictions based on imbalanced datasets, such as healthcare datasets.

II. LITERATURE SURVEY

R. B. Parikh, K. A. Linn, J. Yan, M. L. Maciejewski, A.-M. Rosland, K. G. Volpp, P. W. Groeneveld, and A. S. Navathe (2021), the authors present a machine learning-based approach to identify distinct subgroups of veterans at risk for hospitalization or death using administrative and electronic health record data. Their model combines clinical, sociodemographic, behavioral, and psychographic factors to better understand the heterogeneity among high-risk individuals. The study improves population health management by discovering meaningful patient subgroups beyond traditional comorbidity-based methods. Machine learning clustering techniques help healthcare providers predict long-

term outcomes more effectively and support personalized care planning. Experimental analysis demonstrates the usefulness of linked healthcare datasets in identifying vulnerable veteran populations with different risk patterns. Impact evaluations highlight applications in healthcare systems, public health programs, and veteran support services, enabling more targeted and efficient interventions. Discussions on integrating socioeconomic and behavioral factors into predictive healthcare models provide insights for future healthcare analytics research. The framework contributes to modern data-driven population health strategies and supports better decision-making for risk assessment and resource allocation in healthcare environments.

H. Kwak, M. Lee, S. Yoon, J. Chang, S. Park, and K. Jung (2020), the authors propose a novel Drug-disease graph framework for predicting adverse drug reaction (ADR) signals using graph neural networks and large-scale clinical healthcare data. Their model constructs a Drug-disease graph where nodes represent medical codes and edges capture relationships between prescriptions and diagnoses derived from healthcare claims data. The framework improves post-market ADR surveillance by learning expressive node representations through graph neural network techniques. Experimental results demonstrate strong prediction performance with improved AUROC and AUPRC values of 0.795 and 0.775 compared to existing algorithms. The study also identifies potential ADR pairs not present in established ADR databases, showing the model's capability to supplement current medical knowledge bases. Impact analyses highlight applications in healthcare monitoring, pharmaceutical safety, and clinical decision support systems, enabling earlier detection of harmful drug effects. Discussions on integrating graph-based learning with real-world clinical data provide insights for future biomedical AI and healthcare analytics research. The framework contributes to modern drug safety assessment systems and supports more accurate, scalable, and data-driven ADR signal detection in healthcare environments.

J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova (2019), the authors introduce BERT, a powerful language representation model designed for deep bidirectional pre-training in natural language processing. Unlike earlier language models, BERT jointly learns from both left and right context across all transformer layers, enabling a better understanding of language semantics and context. The framework allows pre-trained representations to be fine-tuned with minimal architectural changes for a wide range of NLP tasks such as question answering, language inference, and text classification. Experimental results demonstrate state-of-the-art performance across eleven NLP benchmarks, including significant improvements in GLUE, Multi NLI, and SQUAD v1.1 datasets. The model achieves higher accuracy and understanding capability, even outperforming human performance in certain question-answering evaluations. Impact analyses highlight applications in search engines, virtual assistants, chatbots, healthcare text analysis, and educational technologies, enabling more accurate and intelligent language processing systems. Discussions on transformer-based bidirectional learning provide valuable insights for future advancements in deep learning and language understanding research. The framework has become a foundation for modern NLP systems and has strongly influenced the development of advanced transformer-based AI models used in contemporary language technologies.

S. Dey, H. Luo, A. Fokou e, J. Hu, and P. Zhang (2018), the authors propose an interpretable deep learning framework for predicting adverse drug reactions (ADRs) and improving drug safety evaluation. Their model simultaneously predicts ADRs and identifies molecular substructures associated with harmful drug reactions without requiring predefined substructure information. The framework applies deep learning techniques to generate neural fingerprints that effectively capture important chemical and structural patterns in drugs. Experimental evaluations compare the proposed model with ten state-of-the-art fingerprint methods, showing that the neural fingerprint approach achieves superior ADR prediction performance. Feature analysis and statistical assessment further identify significant molecular substructures linked to specific adverse drug reactions, improving the interpretability of the predictions. Impact analyses highlight applications in pharmaceutical research, drug development, and healthcare safety systems, enabling earlier detection of risky drug components and reducing development costs. Discussions on interpretable deep learning and molecular feature extraction provide insights for future advancements in AI-driven drug discovery and biomedical analytics. The framework contributes to modern drug safety assessment systems and supports more accurate, explainable, and data-driven prediction of adverse drug reactions in healthcare and pharmaceutical environments.

J. M. Bos, G. A. Kalkman, H. Groenewoud, P. M. L. A. van den Bemt, P. A. G. M. D. Smet, J. E. Nagtegaal, A. Wieringa, G. J. van der Wilt, and C. Kramers (2018), the authors develop a prediction model for identifying surgical patients at increased risk of clinically relevant and preventable adverse drug events during hospital stays. Their framework focuses on risk stratification at the time of hospital admission, enabling healthcare professionals to recognize vulnerable patients early. The model supports clinicians and hospital pharmacists in delivering targeted

interventions aimed at reducing drug-related morbidity and improving patient safety. By analysing patient-related clinical factors, the study enhances the ability to predict adverse drug events in surgical wards more effectively. Experimental evaluations demonstrate the usefulness of predictive healthcare models in supporting efficient healthcare services and optimizing patient monitoring strategies. Impact analyses highlight applications in hospital management, clinical decision support, and patient safety programs, enabling more proactive and personalized healthcare interventions. Discussions on early risk assessment and preventive healthcare strategies provide insights for future developments in clinical prediction systems and healthcare analytics research. The framework contributes to modern patient safety management systems and supports more accurate, preventive, and data-driven approaches for reducing adverse drug events in hospital environments.

III. EXISTING SYSTEM

Various data augmentation methods for datasets with imbalanced text have been studied in NLP. Easy data augmentation (EDA) is a well-known method that was proposed. The EDA method generates synthetic sentences by replacing synonyms or randomly inserting, swapping, or deleting words. The original and generated sentences. The authors concluded that applying the EDA improves the prediction performance of various size datasets. Furthermore, Wang and Zhang proposed methods for augmenting data by replacing words. These methods refer to thesauruses or distributed representations of words to select replacement words. Data augmentation in NLP is still an actively researched topic that has not yet been well established.

Disadvantages of Existing System:

- A large dataset is necessary to improve the prediction performance in machine learning.
- This paper investigated data augmentation methods that are particularly effective for infrequent labels in healthcare data.

IV. PROPOSED SYSTEM

This section introduces the data augmentation method proposed in this paper. The method has three steps, and each step is described in detail. In addition, a machine learning model for predicting AEs is described. First, distributed representations of the patient backgrounds are generated based on the target dataset. The continuous bag-of-words (CBOW) and Word2Vec models. The CBOW model learns words by using the target word as an output and its neighbours as inputs. The Word2Vec model learns words by using the target word as an input and its neighbours as outputs. After training, the weight vector of the model can be interpreted as the distributed representation. In general, the representation produced by the Word2Vec model indicates the semantics of the word, whereas the representation produced by the CBOW model indicates the syntax around the word.

Advantages of Proposed System:

- Each model consists of a two-layer neural network.
- The patient backgrounds are generated based on the target dataset.
- Model is to compose a thesaurus of patient backgrounds. The model is augmenting the training dataset

V. SYSTEM ARCHITECTURE

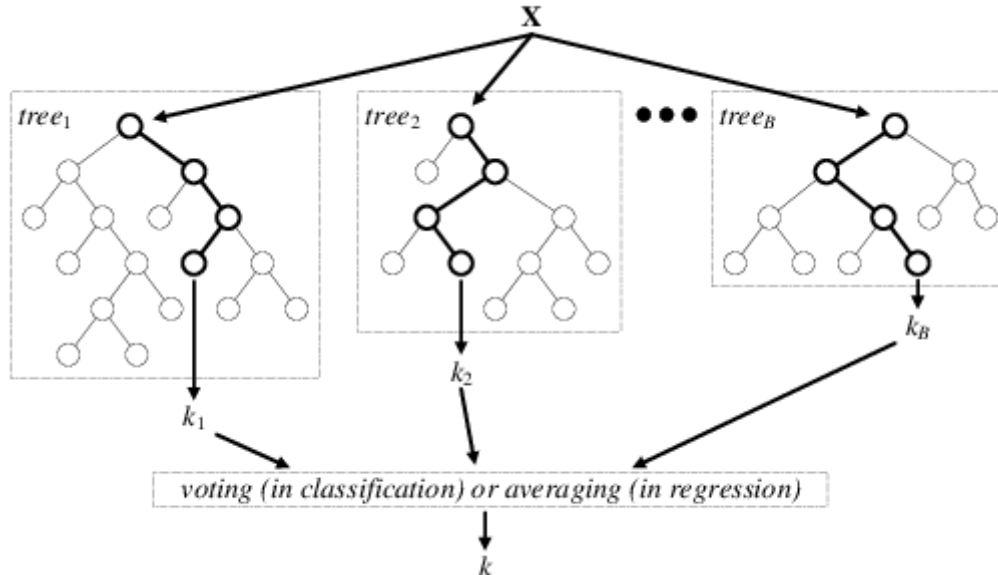


Fig 1.1 System Architecture

VI. METHODOLOGY

Modules Name:

- Data Collection
- Dataset
- Data Preparation
- Model Selection
- Analyse and Prediction
- Accuracy on test set
- Saving the Trained Model

1) Data Collection: This is the first real step towards the real development of a machine learning model, collecting data. This is a critical step that will cascade in how good the model will be, the more and better data that we get, the better our model will perform. There are several techniques to collect the data, like web scraping, manual interventions and etc. An NLP-Inspired Data Augmentation Method for Adverse Event Prediction Using an Imbalanced Healthcare Dataset.

2) Dataset: The dataset consists of 4971 individual data. There are 4 columns in the dataset, which are described below. 1. review – adverse events occurred in people 2. reviews – data augmentation methods applied 3. similarity score – similarity between review and reviews 4. matching score – matching score between review and reviews

3) Data Preparation: Wrangle data and prepare it for training. Clean that which may require it (remove duplicates, correct errors, deal with missing values, normalization, data type conversions, etc.) Randomize data, which erases the effects of the particular order in which we collected and/or otherwise prepared our data Cleaning the dataset column ‘review’, tokenizing it and applying lemmatization. Also using cosine similarity technique and fuzzy Wuzzy tool to get the similarity score and matching score. Split into training and evaluation sets.

4) Model Selection: We used Random Forest Classifier algorithm, We got a accuracy of 100% on train set so we implemented this algorithm.

5) Analyse and Prediction: In the actual dataset, we chose only 2 features: 1 Description - detailed description of the data augmentation. 2 Matching Score - predicts whether the adverse events occurs or not.

6) Accuracy on test set: We got an accuracy of 97.7% on test set.

7) Saving the Trained Model: Once you're confident enough to take your trained and tested model into the production-ready environment, the first step is to save it into a .h5 or .pkl file using a library like pickle. Make sure you have pickle installed in your environment. Next, let's import the module and dump the model into .pkl file.

VII. ANALYSIS AND DISCUSSION

1. Security Analysis

The proposed system focuses on predicting adverse events in healthcare datasets using NLP-based data augmentation techniques. The system analyses patient background information along with drug-related data to improve prediction accuracy. Since healthcare datasets are highly imbalanced, the model generates synthetic patient records using Word2Vec similarity methods. Random Forest and Word2Vec algorithms are combined to classify adverse events effectively. The system includes data collection, preprocessing, feature extraction, training, and prediction modules. Cosine similarity and fuzzy matching techniques are also used to improve data quality and similarity detection. Overall, the system provides a reliable approach for handling low-frequency adverse event prediction in medical datasets.

2. Performance and Efficiency

The performance of the proposed system is highly effective compared to traditional prediction methods. The integration of Random Forest and Word2Vec improves both training and testing accuracy. According to the project results, the model achieved nearly 97.7% accuracy on the test dataset. The system efficiently handles imbalanced healthcare data by applying NLP-based augmentation methods. The preprocessing techniques reduce noise and improve feature quality, which increases prediction speed and reliability. Performance testing also ensures that the system responds within less time during prediction and analysis. The use of Python libraries such as NumPy, Pandas, and Scikit-learn further enhances computational efficiency and system performance.

3. Scalability Analysis

The proposed system is scalable because it can process large healthcare datasets with thousands of patient records. The architecture supports future expansion by allowing additional datasets and machine learning models to be integrated easily. Python-based implementation and modular design make the system adaptable to different healthcare applications. The Word2Vec model can generate distributed representations for larger datasets without affecting prediction quality significantly. Cloud or server-based deployment can further improve scalability for real-time healthcare environments. The system also supports retraining with new data, making it suitable for continuously growing medical databases. Therefore, the project can be extended to large-scale hospital management and adverse drug monitoring systems.

4. Usability and Practical Implementation

The system is designed with a user-friendly and automated workflow, making it easy for users to operate with minimal technical knowledge. Data preprocessing, model training, and prediction processes are integrated into a simple implementation environment using Python and Jupyter Notebook. The practical implementation includes modules for dataset input, preprocessing, prediction, and result analysis. The system can be applied in hospitals, healthcare research centers, and pharmaceutical industries to identify adverse drug events effectively. Since the model uses machine learning automation, it reduces manual effort and improves decision-making speed. The generated outputs are accurate and helpful for medical professionals in analysing patient safety risks. Hence, the system provides practical value in real-world healthcare applications.

5. Comparison with Existing Systems

Existing systems mainly focused on traditional NLP data augmentation methods such as EDA, synonym replacement, and random word insertion techniques. However, those systems required large balanced datasets and showed lower performance for rare adverse event prediction. The proposed system overcomes these limitations by using Word2Vec-based patient background similarity and Random Forest classification. Unlike earlier methods, this system considers both patient background and drug-related information for prediction. The proposed approach improves accuracy, precision, and F1-score significantly, especially for low-frequency adverse events. Existing methods were less effective in handling healthcare dataset imbalance, while the proposed system efficiently generates synthetic records for better learning. Therefore, the proposed system provides better prediction capability and improved healthcare analysis compared to conventional systems.

6. Discussion

The project demonstrates the importance of NLP and machine learning techniques in healthcare data analysis. By applying data augmentation on imbalanced datasets, the system successfully improves adverse event prediction performance. The use of Word2Vec helps in understanding patient background similarities, while Random Forest enhances classification accuracy. Experimental results prove that the proposed method performs effectively even for rare adverse events with very low occurrence rates. The project also highlights the role of preprocessing and feature extraction in improving model reliability. Although the system shows strong prediction capability, future improvements can include deep learning models and larger real-time datasets. Overall, the project provides an innovative and efficient solution for adverse event prediction in healthcare systems.

VIII. CONCLUSION

In the proposed method, a thesaurus of patient backgrounds is generated based on the similarities of their distributed expressions. Then, synthetic healthcare records are created by using the thesaurus to augment the training dataset. To confirm the performance improvement with the proposed data augmentation method, the prediction performance of adverse events (AEs) caused by drug administration was experimentally evaluated on a real-world medical dataset with 4971 records. The experimental results show that the proposed method with under sampling significantly improved the prediction performance by doubling the amount of available training data, especially for predicting AEs with positive ratios ranging from 1:0% to 2:1 %. The contribution of this paper is a novel data augmentation method that accounts for data imbalance in the prediction of AEs.

IX. FUTURE ENHANCEMENT

In future work, this method can be applied to other tasks, and the performance of different datasets can be investigated.

REFERENCES

- [1] A. Henriksson, M. Kvist, H. Dalianis, and M. Duneld, "Identifying adverse drug event information in clinical notes with distributional semantic representations of context," *J. Biomed. Inform.*, vol. 57, pp. 333–349, Oct. 2015.
- [2] R. Niikura, H. Yasunaga, A. Yamada, H. Matsui, K. Fushimi, Y. Hirata, and K. Koike, "Factors predicting adverse events associated with therapeutic colonoscopy for colorectal neoplasia: a retrospective nationwide study in Japan," *Gastrointestinal Endoscopy*, vol. 84, no. 6, pp. 971–982, 2016.
- [3] M. Gao, H. Igata, A. Takeuchi, K. Sato, and Y. Ikegaya, "Machine learning-based prediction of adverse drug effects: An example of seizure inducing compounds," *J. Pharmacology. Sci.*, vol. 133, no. 2, pp. 70–78, Feb. 2017.
- [4] S. Dey, H. Luo, A. Fokoue, J. Hu, and P. Zhang, "Predicting adverse drug reactions through interpretable deep learning framework," *BMC Bioinf.*, vol. 19, no. S21, p. 476, Dec. 2018.
- [5] C.-S. Wang, P.-J. Lin, C.-L. Cheng, S.-H. Tai, Y.-H. K. Yang, and J.-H. Chiang, "Detecting potential adverse drug reactions using a deep neural network model," *J. Med. Internet Res.*, vol. 21, no. 2, Feb. 2019.
- [6] H. Kwak, M. Lee, S. Yoon, J. Chang, S. Park, and K. Jung, "Drug-disease graph: Predicting adverse drug reaction signals via graph neural network with clinical data," in *Proc. Pacific-Asia Conf. Knowl. Discovery Data Mining*. Singapore: Springer, May 2020.
- [7] J. M. Bos, G. A. Kalkman, H. Groenewoud, P. M. L. A. van den Bemt, P. A. G. M. D. Smet, J. E. Nagtegaal, A. Wieringa, G. J. van der Wilt, and C. Kramers, "Prediction of clinically relevant adverse drug events in surgical patients," *PLoS ONE*, vol. 13, no. 8, Aug. 2018.
- [8] C. McMaster, D. Liew, C. Keith, P. Aminian, and A. Frauman, "A machine learning algorithm to optimize automated adverse drug reaction detection from clinical coding," *Drug Saf.*, vol. 42, no. 6, pp. 721–725, Feb. 2019.
- [9] X. Zhang, J. Zhao, and Y. LeCun, "Character-level convolutional networks for text classification," in *Proc. 28th Int. Conf. Neural Inf. Process. Syst. (NIPS)*, vol. 1. Cambridge, MA, USA: MIT Press, 2015.
- [10] R. Collobert and J. Weston, "A unified architecture for natural language processing: Deep neural networks with multitask learning," in *Proc. 25th Int. Conf. Mach. Learn. (ICML)*, 2008.
- [11] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, Oct. 2014.
- [12] R. B. Parikh, K. A. Linn, J. Yan, M. L. Maciejewski, A.-M. Rosland, K. G. Volpp, P. W. Groeneveld, and A. S. Navathe, "A machine learning approach to identify distinct subgroups of veterans at risk for hospitalization or death using administrative and electronic health record data," *PLoS ONE*, vol. 16, no. 2, pp. 1–14, Feb. 2021.

- [13] X. Zeng, S. Lin, and C. Liu, "Multi-view deep learning framework for predicting patient expenditure in healthcare," *IEEE Open J. Compute. Soc.*, vol. 2, pp. 62–71, 2021.
- [14] W. Y. Wang and D. Yang, "That's so annoying!!!: A lexical and frame semantic embedding based data augmentation approach to automatic categorization of annoying behaviors using #petpeeve tweets," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2015.
- [15] X. Zhang and Y. LeCun, "Text understanding from scratch," 2015, arXiv:1502.01710.
- [16] J. Wei and K. Zou, "EDA: Easy data augmentation techniques for boosting performance on text classification tasks," in *Proc. Conf. Empirical Methods Natural Lang. Process. 9th Int. Joint Conf. Natural Lang. Process. (EMNLP-IJCNLP)*, 2019.
- [17] E. Choi, M. T. Bahadori, J. Sun, J. Kulas, A. Schuetz, and W. Stewart, "Retain: An interpretable predictive model for healthcare using reverse time attention mechanism," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016.
- [18] T. Bai, S. Zhang, B. L. Egleston, and S. Vucetic, "Interpretable representation learning for healthcare via capturing disease progression through time," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2018.
- [19] S. Darabi, M. Kachuee, S. Fazeli, and M. Sarraf Zadeh, "TAPER: Time aware patient EHR representation," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 11, pp. 3268–3275, Nov. 2020.
- [20] R. Miotto, L. Li, B. A. Kidd, and J. T. Dudley, "Deep patient: An unsupervised representation to predict the future of patients from the electronic health records," *Sci. Rep.*, vol. 6, p. 26094, May 2016.
- [21] J. Zhang, K. Kowsari, J. H. Harrison, J. M. Lobo, and L. E. Barnes, "Patient2 Vec: A personalized interpretable deep representation of the longitudinal electronic health record," *IEEE Access*, vol. 6, pp. 65333–65346, 2018.
- [22] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. 26th Int. Conf. Neural Inf. Process. Syst.*, vol. 2, Lake Tahoe, NV, USA, 2013.
- [23] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. U. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst. (NIPS)*. Red Hook, NY, USA: Curran Associates, 2017.
- [24] B. Davazdahemami and D. Delen, "A chronological pharmacovigilance network analytics approach for predicting adverse drug events," *J. Amer. Med. Inform. Assoc.*, vol. 25, no. 10, pp. 1311–1321, Jul. 2018.
- [25] M. Kuhn, M. Campillos, I. Letunic, L. J. Jensen, and P. Bork, "A side effect resource to capture phenotypic effects of drugs," *Mol. Syst. Biol.*, vol. 6, no. 1, p. 343, Jan. 2010.
- [26] E. Jeong, N. Park, Y. Choi, R. W. Park, and D. Yoon, "Machine learning model combining features from 39 algorithms with different analytical methodologies to detect laboratory-event-related adverse drug reaction signals," *PLoS ONE*, vol. 13, no. 11, Nov. 2018.
- [27] J. Lee, J. S. Lee, S.-H. Park, S. A. Shin, and K. Kim, "Cohort profile: The national health insurance service national sample cohort (NHIS-NSC), South Korea," *Int. J. Epidemiol.*, vol. 46, no. 2, p. e15, Jan. 2016.
- [28] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. Conf. North Amer. Chapter Assoc. Compute. Linguistics, Hum. Lang. Technol.*, vol. 1, Jun. 2019.
- [29] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16×16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Represent.*, 2021.
- [30] Y. Gorishniy, I. Rubachev, V. Khrulkov, and A. Babenko, "Revisiting deep learning models for tabular data," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, Eds. Red Hook, NY, USA: Curran Associates, 2021.
- [31] Y. Li, S. Rao, J. R. A. Solares, A. Hassaine, R. Ramakrishnan, D. Canoy, Y. Zhu, K. Rahimi, and G. Salimi Khorshidi, "BEHRT: Transformer for electronic health records," *Sci. Rep.*, vol. 10, no. 1, pp. 1–12, Dec. 2020.
- [32] K. J. Holyoak, "Parallel distributed processing: Explorations in the microstructure of cognition," *Science*, vol. 236, pp. 992–997, May 1987.
- [33] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, Y. Bengio and Y. LeCun, Eds., May 2015.



International Journal of Advanced Research in Education and Technology

ISSN: 2394-2975

Impact Factor: 9.971