Advancing Preauthorization Task in Healthcare: An Application of Deep Active Incremental Learning for Medical Text Classification

Aishat Salau
Nile University of Nigeria, Nigeria
salauaishat@gmail.com (corresponding author)

Nnanna Agwu Nwojo
Nile University of Nigeria, Nigeria
nagwu@nileuniversity.edu.ng

Musa M. Boukar
Nile University of Nigeria, Nigeria
musa.muhammed@nileuniversity.edu.ng

Osasumwen Usen
Independent Researcher, Nigeria
osas.usen@gmail.com

Received: 27 August 2023 | Revised: 14 September 2023 | Accepted: 29 September 2023

 DOI: https://doi.org/10.48084/etasr.6332

ABSTRACT

This study presents a novel approach to medical text classification using a deep active incremental learning model, aiming to improve the automation of the preauthorization process in medical health insurance. By automating decision-making for request approval or denial through text classification techniques, the primary focus is on real-time prediction, utilization of limited labeled data, and continuous model improvement. The proposed approach combines a Bidirectional Long Short-Term Memory (Bi-LSTM) neural network with active learning, using uncertainty sampling to facilitate expert-based sample selection and online learning for continuous updates. The proposed model demonstrates improved predictive accuracy over a baseline Long Short-Term Memory (LSTM) model. Through active learning iterations, the proposed model achieved a 4% improvement in balanced accuracy over 100 iterations, underscoring its efficiency in continuous refinement using limited labeled data.

Keywords-medical text classification; deep learning; active learning; incremental learning; preauthorization

I. INTRODUCTION

The healthcare industry generates extensive data, including electronic health records, clinical reports, medical claims, and administrative documents. Extracting insights from these data is crucial for patient care, medical decisions, and cost management. However, the volume and complexity of healthcare data pose challenges to effective analysis and interpretation. To address this issue, the use of machine learning and Natural Language Processing (NLP) techniques has attracted a growing interest in medical text classification. This involves automatically categorizing medical texts or documents into predefined classes to facilitate the identification of relevant information and enhance relevant medical tasks [1].

An illustrative application of medical text classification is in medical preauthorization tasks. These tasks involve obtaining insurance approval before providing certain services or medications to patients. This process is usually time-consuming, but medical text classification offers an avenue for process optimization and automation by categorizing preauthorization requests. Despite potential benefits, medical text classification faces significant challenges, such as the complexity and dynamic nature of medical data, the high cost and complexity of data annotation, and the unbalanced and noisy nature of medical text data. These make the development of robust classification models challenging [2-3]. Although conventional machine learning algorithms have been applied to such classification tasks, limitations are encountered, particularly when handling large volumes of data [4]. As such, new approaches are needed to overcome the limitations.

Recent studies have focused on the use of deep learning approaches for medical text classification. These approaches take advantage of the power of neural networks to learn complex and dynamic representations of medical data. For example, in [5-6] Convolutional Neural Networks (CNNs) were used for medical text classification tasks, and the results showed that the models performed better than state-of-the-art methods. In [7], a CNN model based on weak supervision and deep representation was proposed for clinical text classification, achieving an F1-score of 0.97. In [8], a unified deep neural network based on a convolutional layer, a Bidirectional Gated Recurrent Unit (Bi-GRU), and an attention mechanism was used for medical text classification. The proposed method was evaluated on various datasets, achieving accuracies of 89.09%, 93.75%, and 97.73%. In [2], another approach was presented, based on CNN, Bi-LSTM, and Multi-head attention, achieving an accuracy and F1-score of 91.99% and 92.03%, respectively.

In [1], a hybrid text classification model was proposed that combined a gated Attention-Based Bidirectional Long Short-Term Memory (ABLSTM) and a regular expression-based classifier for medical text classification. This method demonstrated an accuracy of 89% and an F1-score of 0.92. These studies demonstrate the potential of deep learning approaches for medical text classification but also highlight some challenges, such as the requirement for large amounts of annotated data. Active Learning (AL) is a technique capable of minimizing the volume of data required to train a machine learning model and works by iteratively selecting the most informative examples from a large pool of unlabeled data for annotation by a human annotator [9]. This approach can save time and money in data annotation while still achieving high classification accuracy. The effectiveness of AL is evident in its application to medical text classification. For instance, in [10], AL based on a Support Vector Machine (SVM) was used to classify medical text, achieving accuracy levels of more than 90% across four datasets. This result underscores the use of AL as a powerful tool to improve the performance of medical text classifiers. Similarly, in [11], an active learning framework was proposed based on SVM, leading to a 64% reduction in annotation efforts. In [12], active learning was used to classify radiology reports, achieving a classifier performance of 98.25% sensitivity and 96.14% specificity, while also saving up to 92% of the training data required for supervised machine learning. These studies demonstrate the potential of active learning approaches for medical text classification but also highlight the importance of careful selection of informative samples in the AL process.

Deep Learning (DL) is known for its need for large data and its ability to process highly dimensional and unstructured data while inherently performing feature extraction [13]. On the other hand, AL is known for its ability to increase annotation efficiency and reduce annotation costs [14]. Therefore, the combination of DL and AL, known as Deep Active Learning (DAL), combines the advantages of both methods, leveraging the strengths of each one while addressing their limitations. Therefore, better results are expected, and the application potential of both approaches can be expanded. In [15], a comparative analysis of 11 AL strategies was carried out on the classification of cancer pathology reports. Using CNN as the base classification model, the results showed that using less than half of the data, AL can obtain a similar performance compared to a supervised learning model trained on all available data. In the context of medical text classification, the application of DAL remains limited despite initial success.

This study aimed to bridge this significant research gap by exploring the potential of DAL in medical preauthorization tasks. This study proposes a novel approach to medical preauthorization, which combines the power of deep neural networks with Active Incremental Learning (AIL) strategies to actively select and query uncertain or informative samples for annotation and, consequently, update the model with newly annotated data samples, ensuring continuous model refinement and improved performance. Additionally, this study used a real-world dataset of medical preauthorization requests. The results of this study have far-reaching implications, potentially improving the preauthorization process by enabling faster and more accurate decisions, optimizing cost, and ultimately improving patient care. In addition, these findings offer insights and set the stage for future research in the field.

II. METHODOLOGY

A. Data Collection and Pre-processing

The first step was to collect and pre-process the data used for medical text classification. A dataset of real preauthorization requests was obtained from a Nigerian Health Management Organization, containing 117,342 preauthorization requests. Ethical aspects were a paramount concern, particularly in the collection and use of patient data, and data privacy and confidentiality were ensured by de-identifying the patient information and complying with HIPAA regulations. The preauthorization request data includes descriptions of diagnoses along with related procedures, treatments, or medications prescribed by healthcare providers. The data were pre-processed using a combination of natural language processing techniques and manual review. At first, duplicates, null spaces, and punctuation were removed, and the textual data were tokenized. Proper encoding is essential to enable the effective processing of categorical features in a neural network. One-hot coding is a conventional method for converting categorical variables into vectors, but it comes with a drawback: notably, a substantial increase in dataset dimensionality and the inability to capture intricate semantic relationships between words [16]. To address these challenges, word embedding was used, which considered the context of words for vectorizing text data and one-hot encoding of categorical features [17].

Class balancing techniques were used as a part of the pre-processing steps as the resulting data exhibited the class rarity challenge, a situation in which the distribution of classes in the training data is not balanced, meaning that some classes have significantly fewer examples compared to others [18, 19]. The class distribution of the dataset was extremely imbalanced with

117,068 authorized requests and 279 unauthorized requests. Most machine learning algorithms have difficulty creating a model that accurately classifies examples in a minority class, as a result of bias toward the majority class samples due to its increased occurrence in the dataset [20]. Therefore, data-level class balancing methods, including Random Undersampling (RUS) and Synthetic Minority Oversampling Technique (SMOTE) [21], were explored to address this challenge. Random undersampling was used to reduce the number of majority samples with a higher frequency of occurrence or highly similar samples, as this can be detrimental to the performance of the model. When a model is repeatedly exposed to similar samples, it can be biased towards those samples, leading to overfitting, misleading evaluation, and reducing generalization to new data points [20]. SMOTE was also applied for oversampling to achieve a balanced dataset. The pre-processed dataset was of the form $D=\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$. D is the dataset containing a total number of n samples, where each sample is a pair consisting of the feature vector x_i and corresponds to a ground truth label y_i, (x_i, y_i). $y_i \in \{0, 1\}$, since it is a binary classification problem.

B. Model Design

The proposed model consists of a Bi-LSTM neural network that can effectively handle the dynamic and complex nature of medical data. To implement AIL, a stream-based AL scenario and a combination of uncertainty sampling were used as a selective strategy and online learning for model update. In a stream-based AL scenario, unlabeled samples arrive in a data stream, uncertainty sampling involves selecting the most informative samples from the unlabeled data, and querying a human expert for the ground truth is subsequently used to update the model. Online learning involves updating the model with new data in real time, without the need to retrain the entire model from scratch. Figure 1 presents an overview of the proposed framework.

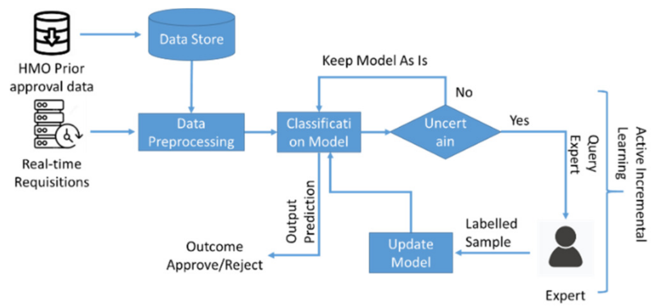

Fig. 1. High-level architecture of the proposed framework.

1) Bi-LSTM

Bi-LSTM, a variant of Recurrent Neural Networks (RNNs), is particularly adept at handling sequential data, such as text. Bi-LSTM enhances the standard LSTM by integrating both forward and backward hidden layers. This configuration allows it to capture information from both preceding and subsequent contexts, giving it a significant edge in sequential modeling tasks compared to the unidirectional LSTM approach that only exploits the historical context [22-23].

2) Active Learning

A stream-based AL approach was used, in which preauthorization requests come in a stream of data. For the selection strategy, the entropy-based uncertainty sampling strategy was used, which is an adaptive sampling strategy that evaluates the model's current state while adjusting the criteria for sample selection based on the model performance. This involves using the entropy of the predicted probabilities as a measure of uncertainty, which is calculated as the sum of the probability of each class multiplied by the log of its probability [24]. In entropy-based uncertainty sampling, following the training of a model on labeled data and its subsequent use to predict labels for unlabeled data, the data points exhibiting the greatest entropy emerge as the most informative ones and are consequently selected for labeling. By selecting samples with the highest uncertainty, the goal is to improve the performance of the model by reducing the uncertainty in its predictions.

3) Incremental Learning

For incremental learning, online learning was used, which involves updating the model with new data in real-time without the need to completely retrain the model. Combining online learning and AL can overcome the main drawbacks of traditional supervised online learning, which are labeled data dependency and concept drift, by actively selecting informative samples for annotation and continuously updating the model as new data become available [25]. Online AL functions through a series of iterative steps [26]. In each iteration, an unlabeled sample is presented to the model, which determines whether to request its label. If the label is requested, this labeled sample is incorporated and used to refine the model; otherwise, the model maintains its current state [27]. This is contrary to the typical online learning process, where the learner requests class labels for all incoming instances. Additionally, during the model update, class weights are applied to samples to effectively address the challenge of class imbalance within the data stream.

C. Model Training

The deep learning model was trained on the preprocessed preauthorization request data. To improve the performance of these models, as well as address overfitting issues common with imbalanced datasets, cross-validation and regularization techniques were used, including stratified k-fold cross-validation, dropout, and L2 regularization. The model was trained using the Adam optimizer, the binary cross-entropy loss function, and the sigmoid activation function. Following the initial model training and evaluation, the active incremental learning phase was activated. For the incoming stream of data samples, the initial model makes a prediction on each sample, and based on the entropy of the model prediction, the most informative samples were iteratively selected for annotation by a human expert. The threshold for the entropy measure was set to 0.8, indicating that only samples with a prediction confidence of at least 80% are considered.

Consequently, the model was consistently updated with the newly labeled samples. After 100 iterations, the performance of the resultant model was assessed in the test set. To simulate the preauthorization request stream, 10% of the dataset was set aside, and the active learner reads and processes the data samples.

Let D_{init} be the initial dataset containing labeled data samples to be used for model pretraining, represented in (1), where x_i is the i-th data sample and y_i is its corresponding label.

$$D_{init} = \{(x_i y_i)\}_{i=1}^{N_{init}} \quad (1)$$

Let f_{init} represent the initial deep learning model that takes an input data sample x and gives a prediction output $\hat{y}$:

$$f_{init}(x) \rightarrow \hat{y} \quad (2)$$

Given D_{init}, the model is trained to minimize the chosen loss function L that quantifies the difference between $\hat{y}$ and y:

$$\theta^* = argmin_{\theta} \sum_{(x_i y_i) \in D_{init}} L(f(x_i; \theta), y_i) \quad (3)$$

where θ represents the parameters of the model and θ^*, the optimal parameters of the model after training. Let X_{str} be the subsequent unlabelled data stream:

$$X_{str} = \{x_j\}_{j=1}^{N_{str}} \quad (4)$$

For all $x_j \in X_{str}$, the initial model f_{init} predicts the class probability $P(y=c|x_j)$ for each class c. The entropy of x_j is calculated by :

$$H(x_j) = -\sum_c P(\mathrm{y} = \mathrm{c}|x) log P(y = c|x) \quad (5)$$

where $P(y=c|x)$ is the predictive probability of class c for input x_j. If $H(x_j) > \sigma$, where σ is a predefined threshold, then select the sample x_j for annotation by the oracle O, $O \xleftarrow{H(x_j > \sigma)} x_j$; query O for the true label y_j^* of the selected sample. The oracle responds with the true label,

$$y_j^*: \; x_j \overset{y_j^*}{\leftarrow} \mathrm{O}.$$

Update model f_{init} with the newly labeled data, (x_j, y_j^*):

$$f_{update} = argmin_f \sum_{(x_j y_j)} \sum_{c=1}^{C} W_{cy_j} \,.\; L(f(x_j), y_j) \quad (6)$$

where C is the number of classes, and W_{cy_j} is the class weight for class c in sample x_j. The term $W_{ic} \,.\; L(f(x_j), y_j)$ adjusts the loss contribution for each sample and class, based on the assigned class weight. This encourages the model to focus on and improve performance for the less represented class during training. The integration of AL and incremental learning strategies offers the advantage of an iterative process, systematically selecting the most instructive examples to improve the model's training, consequently leading to progressive improvement in its performance over time.

D. Evaluation Metrics

Although accuracy is a commonly used metric to assess classifier performance in machine learning, it becomes inadequate when dealing with highly imbalanced datasets. This is because a classifier giving preference to the majority class can yield an optimistic estimate, leading to misleading results [28]. Here, accuracy fails to differentiate between the correct labels of different classes [29]. To address this limitation, balanced accuracy emerges as a more suitable evaluation metric. Balanced accuracy computes the arithmetic mean of sensitivity and specificity, treating both classes equally. By doing so, it delivers a more reliable assessment of model performance on imbalanced data [30]. Thus, this study used the balanced accuracy metric to evaluate the performance of the proposed model.

III. RESULTS

The proposed deep AIL model was evaluated on a real preauthorization task dataset obtained from a health insurance company. First, the performance of the initial Bi-LSTM model was compared to a baseline LSTM model, and subsequently, the augmented Bi-LSTM with AIL was compared with them for a more nuanced evaluation. Table I presents the preliminary evaluation results, comparing the balanced accuracy of both models in the validation and test datasets. The results underscore the inherent superiority of the Bi-LSTM model in dealing with the intricacies of medical text data with a balanced accuracy of 0.79. Figure 2, shows the balanced accuracy of the models in eight folds. Furthermore, the adaptive potential of the augmented Bi-LSTM model was evaluated in stages in more than 100 sample queries. This sets the stage for the iterative learning process, where the model progressively refines its predictions over subsequent AL iterations. Table II presents the results of this experiment, indicating that the proposed model achieved increasing performance with a significantly smaller number of labeled samples, underscoring its ability to leverage sparsely labeled data and offering a practical solution to the often resource-constrained medical domain. Figure 3 also shows the incremental improvement of the proposed model with an increasing number of queries and model updates, demonstrating its potential to improve over time after several iterations of training over new data. The proposed model showed a 4% increase in performance throughout the query and training iterations.

TABLE I. RESULTS OF MODELS BEFORE AIL

Models	Balanced Accuracy	
	Validation Set	Test Set
Baseline	0.704	0.737
Bi-LSTM	0.753	0.793

TABLE II. PERFORMANCE REPORTING BALANCED ACCURACY OF MODELS WITH AIL AFTER MULTIPLE QUERY AND TRAINING ITERATIONS

	Number of Query Iterations			
Models	20 queries	40 queries	60 queries	100 queries
Baseline + AIL	0.739	0.702	0.731	0.764
Bi-LSTM + AIL	0.752	0.756	0.769	0.795

The inherent capability of the proposed model to learn consistently from newly annotated data results in gradual performance improvement over time. This intrinsic ability allows the model to evolve and enhance its predictive accuracy with each subsequent set of newly labeled samples. This dynamic learning mechanism enables the model to quickly adapt to the evolving landscape of requests, providing the preauthorization process with increased adaptability and responsiveness. Furthermore, the proposed approach enables real-time prediction and effectively tackles the issue of inaccurate predictions by incorporating human involvement in the model training process. This empowers insurers to make faster and more accurate decisions about preauthorization requests, ultimately leading to better patient care.

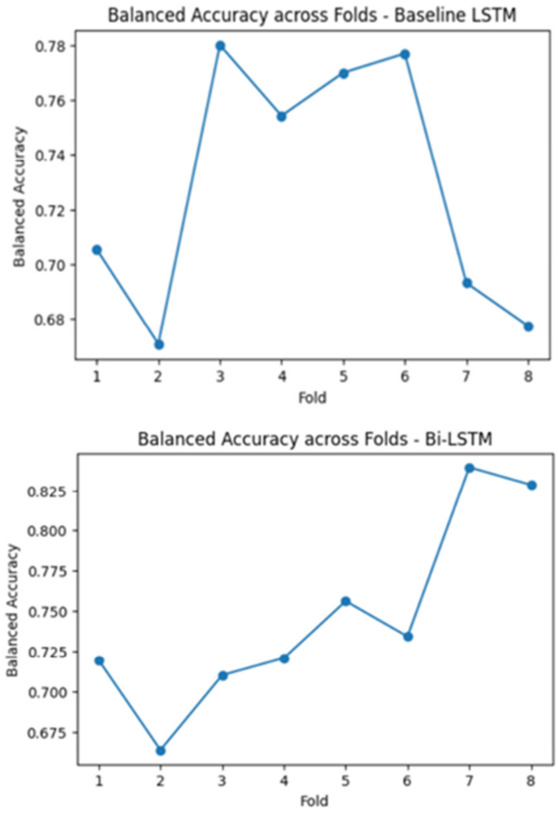

Fig. 2. Balanced accuracy of initial models across folds.

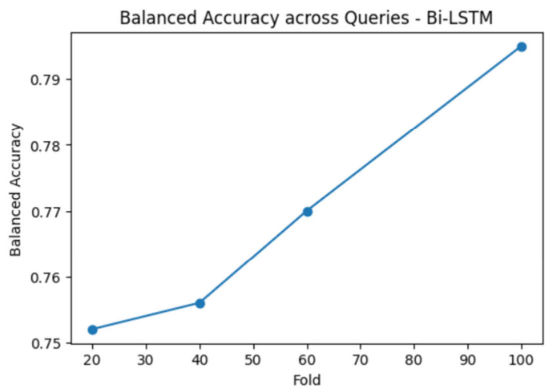

Fig. 3. Balanced accuracy increased across 100 queries and training iterations.

Previous studies on medical preauthorization in health insurance [31-32] relied on traditional machine-learning methods. Although the proposed approach shows potential, it is important to note that previous studies exhibited higher performance than the initial performance of the proposed model, and this could be attributed to the limited amount of labeled data used to train it. Furthermore, the models in previous studies are rather static, lacking the ability of continuous model updates, and require retraining on an entire dataset to implement any updates. In essence, they are not capable of adapting to the dynamic nature of preauthorization requests, a key feature of the proposed model.

IV. CONCLUSION

This study proposed a novel deep AIL framework for medical preauthorization tasks, addressing critical issues in the field, and reducing human annotation effort by selecting only informative samples for annotation instead of requesting class labels for all incoming instances. The proposed model can learn continuously from new data, which improves its performance over time, thus alleviating the strong dependence on labeled data for model training. The experimental results indicate the model's ability to learn from limited data and exhibit ongoing improvements when updated with newly labeled data. Additionally, the issue of class imbalance in the preauthorization request stream was addressed. The model's adaptability demonstrated in its initial performance and subsequent iterative improvements, substantiated its potential as a valuable tool for accurate and evolving medical data analysis, enabling insurers to make faster preauthorization decisions and, in turn, improve patient care.

Although this study yielded promising results, there is room for further research. Future work could investigate alternative DL models, including the use of specialized medical pre-trained word embeddings within a transfer-learning framework, potentially resulting in further improvement in model performance. Furthermore, to address the challenges associated with accessing preauthorization data, it would be valuable to develop a publicly accessible deidentified preauthorization dataset, as such an initiative would not only facilitate collaborative research but also allow effective benchmarking of research results.

REFERENCES

[1] X. Li, M. Cui, J. Li, R. Bai, Z. Lu, and U. Aickelin, "A hybrid medical text classification framework: Integrating attentive rule construction and neural network," *Neurocomputing*, vol. 443, pp. 345–355, Jul. 2021, https://doi.org/10.1016/j.neucom.2021.02.069.

[2] Z. Shen and S. Zhang, "A Novel Deep-Learning-Based Model for Medical Text Classification," in *Proceedings of the 2020 9th International Conference on Computing and Pattern Recognition*, Xiamen, China, Jan. 2021, pp. 267–273, https://doi.org/10.1145/3436369.3436469.

[3] C. Y. Kesiku, A. Chaves-Villota, and B. Garcia-Zapirain, "Natural Language Processing Techniques for Text Classification of Biomedical Documents: A Systematic Review," *Information*, vol. 13, no. 10, Oct. 2022, Art. no. 499, https://doi.org/10.3390/info13100499.

[4] M. Naeem *et al.*, "Trends and Future Perspective Challenges in Big Data," in *Advances in Intelligent Data Analysis and Applications*, Singapore, 2022, pp. 309–325, https://doi.org/10.1007/978-981-16-5036-9_30.

[5] M. Hughes, I. Li, S. Kotoulas, and T. Suzumura, "Medical Text Classification Using Convolutional Neural Networks," *Studies in health technology and informatics*, vol. 235, pp. 246–250, Jan. 2017, https://doi.org/10.3233/978-1-61499-753-5-246.

[6] B. He, Y. Guan, and R. Dai, "Classifying medical relations in clinical text via convolutional neural networks," *Artificial Intelligence in*

Medicine, vol. 93, pp. 43–49, Jan. 2019, https://doi.org/10.1016/j.artmed.2018.05.001.

[7] Y. Wang *et al.*, "A clinical text classification paradigm using weak supervision and deep representation," *BMC Medical Informatics and Decision Making*, vol. 19, no. 1, Jan. 2019, Art. no. 1, https://doi.org/10.1186/s12911-018-0723-6.

[8] L. Qing, W. Linhong, and D. Xuehai, "A Novel Neural Network-Based Method for Medical Text Classification," *Future Internet*, vol. 11, no. 12, Dec. 2019, Art. no. 255, https://doi.org/10.3390/fi11120255.

[9] B. Settles, "Active Learning Literature Survey," University of Wisconsin-Madison Department of Computer Sciences, Technical Report, 2009. [Online]. Available: https://minds.wisconsin.edu/handle/1793/60660.

[10] R. L. Figueroa, Q. Zeng-Treitler, L. H. Ngo, S. Goryachev, and E. P. Wiechmann, "Active learning for clinical text classification: is it better than random sampling?," *Journal of the American Medical Informatics Association*, vol. 19, no. 5, pp. 809–816, Sep. 2012, https://doi.org/10.1136/amiajnl-2011-000648.

[11] N. Nissim *et al.*, "An Active Learning Framework for Efficient Condition Severity Classification," in *Artificial Intelligence in Medicine*, Pavia, Italy, 2015, pp. 13–24, https://doi.org/10.1007/978-3-319-19551-3_3.

[12] D. H. M. Nguyen and J. D. Patrick, "Supervised machine learning and active learning in classification of radiology reports," *Journal of the American Medical Informatics Association*, vol. 21, no. 5, pp. 893–901, Sep. 2014, https://doi.org/10.1136/amiajnl-2013-002516.

[13] T. Kumar and R. Ponnusamy, "Robust Medical X-Ray Image Classification by Deep Learning with Multi-Versus Optimizer," *Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 111406–11411, Aug. 2023, https://doi.org/10.48084/etasr.6127.

[14] C. Schröder and A. Niekler, "A Survey of Active Learning for Text Classification using Deep Neural Networks." arXiv, Aug. 17, 2020, https://doi.org/10.48550/arXiv.2008.07267.

[15] K. De Angeli *et al.*, "Deep active learning for classifying cancer pathology reports," *BMC Bioinformatics*, vol. 22, no. 1, Mar. 2021, Art. no. 113, https://doi.org/10.1186/s12859-021-04047-1.

[16] F. K. Khattak, S. Jeblee, C. Pou-Prom, M. Abdalla, C. Meaney, and F. Rudzicz, "A survey of word embeddings for clinical text," *Journal of Biomedical Informatics*, vol. 100, Jan. 2019, Art. no. 100057, https://doi.org/10.1016/j.yjbinx.2019.100057.

[17] B. Ahmed, G. Ali, A. Hussain, A. Baseer, and J. Ahmed, "Analysis of Text Feature Extractors using Deep Learning on Fake News," *Engineering, Technology & Applied Science Research*, vol. 11, no. 2, pp. 7001–7005, Apr. 2021, https://doi.org/10.48084/etasr.4069.

[18] R. M. Mathew and R. Gunasundari, "A Cluster-based Undersampling Technique for Multiclass Skewed Datasets," *Engineering, Technology & Applied Science Research*, vol. 13, no. 3, pp. 10785–10790, Jun. 2023, https://doi.org/10.48084/etasr.5844.

[19] S. Rodda and U. S. R. Erothi, "Class imbalance problem in the Network Intrusion Detection Systems," in *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, Chennai, India, Mar. 2016, pp. 2685–2688, https://doi.org/10.1109/ICEEOT.2016.7755181.

[20] M. Reza, S. Miri, and R. Javidan, "A Hybrid Data Mining Approach for Intrusion Detection on Imbalanced NSL-KDD Dataset," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 6, pp. 20–25, 2016, https://doi.org/10.14569/IJACSA.2016.070603.

[21] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, Jun. 2002, https://doi.org/10.1613/jair.953.

[22] S. Siami-Namini, N. Tavakoli, and A. S. Namin, "The Performance of LSTM and BiLSTM in Forecasting Time Series," in *2019 IEEE International Conference on Big Data (Big Data)*, Sep. 2019, pp. 3285–3292, https://doi.org/10.1109/BigData47090.2019.9005997.

[23] G. Liu and J. Guo, "Bidirectional LSTM with attention mechanism and convolutional layer for text classification," *Neurocomputing*, vol. 337, pp. 325–338, Apr. 2019, https://doi.org/10.1016/j.neucom.2019.01.078.

[24] R. R. Yager, "Entropy Measures Under Similarity Relations," *International Journal of General Systems*, vol. 20, no. 4, pp. 341–358, May 1992, https://doi.org/10.1080/03081079208945039.

[25] J. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy, and A. Bouchachia, "A survey on concept drift adaptation," *ACM Computing Surveys*, vol. 46, no. 4, pp. 44:1-44:37, Nov. 2014, https://doi.org/10.1145/2523813.

[26] N. Cesa-Bianchi, C. Gentile, and L. Zaniboni, "Worst-Case Analysis of Selective Sampling for Linear Classification," *Journal of Machine Learning Research*, vol. 7, no. 44, pp. 1205–1230, 2006.

[27] J. Lu, P. Zhao, and S. C. H. Hoi, "Online Passive-Aggressive Active learning," *Machine Learning*, vol. 103, no. 2, pp. 141–183, May 2016, https://doi.org/10.1007/s10994-016-5555-y.

[28] M. F. Jojoa Acosta, L. Y. Caballero Tovar, M. B. Garcia-Zapirain, and W. S. Percybrooks, "Melanoma diagnosis using deep learning techniques on dermatoscopic images," *BMC Medical Imaging*, vol. 21, no. 1, Jan. 2021, Art. no. 6, https://doi.org/10.1186/s12880-020-00534-8.

[29] M. Sokolova, N. Japkowicz, and S. Szpakowicz, "Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation," in *AI 2006: Advances in Artificial Intelligence*, Hobart, Australia, 2006, pp. 1015–1021, https://doi.org/10.1007/11941439_114.

[30] P. Thölke *et al.*, "Class imbalance should not throw you off balance: Choosing the right classifiers and performance metrics for brain decoding with imbalanced data," *NeuroImage*, vol. 277, Art. no. 120253, Aug. 2023, https://doi.org/10.1016/j.neuroimage.2023.120253.

[31] K. Farias, P. Santos Neto, A. Santana, and R. Bezerra Neto, "Using Historical Information of Patients for Prior Authorization Learning," in *2019 8th Brazilian Conference on Intelligent Systems (BRACIS)*, Salvador, Brazil, Jul. 2019, pp. 598–603, https://doi.org/10.1109/BRACIS.2019.00110.

[32] G. Veras Magalhães Junior, J. P. Albuquerque Vieira, R. Lira de Sales Santos, J. L. Nascimento Barbosa, P. de Alcântara dos Santos Neto, and R. Santos Moura, "A Study of the Influence of Textual Features in Learning Medical Prior Authorization," in *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*, Cordoba, Spain, Jun. 2019, pp. 56–61, https://doi.org/10.1109/CBMS.2019.00021.