Journal for Technology and Science
Homepage : https://journal.antispublisher.com/index.php/IPTEKS/index
Email : admin@antispublisher.com

e-ISSN : 3047-4337
JOTECH, Vol. 3, No. 3, July 2026
Page 1-14

Revolutionizing Cardiovascular Care: An AI-Driven Approach to Early Intervention

Hussein Ali Al-jashamy[1], Hashim Adnan[2], Hussein Majid[3]
[1]General Directorate of Education in Karbala, Iraq
[2]Open Educational College, Iraq
[3]General Directorate of Education in Al-Muthanna, Iraq

DOI : https://doi.org/10.61796/ipteks.v3i3.505

Sections Info

Article history:
Submitted: March 15, 2026
Final Revised: April 10, 2026
Accepted: May 05, 2026
Published: June 10, 2026

Keywords:
NLP
CVD
Catheterization
Deep learning

ABSTRACT

Objective: *Cardiovascular diseases (CVDs) continue to be a primary cause of early death globally, with both their prevalence and the costs associated with healthcare consistently increasing. Epidemiological Researches has pinpointed a range of risk factors, including high cholesterol levels, elevated blood pressure, diabetes, obesity, smoking, and lack of physical activity, which together account for more than 90% of the risk linked to CVDs. The integration of artificial intelligence (AI) into healthcare has revolutionized medical diagnosis and treatment, particularly in the field of cardiology. Natural Language Processing (NLP) algorithms further enhance this by converting unstructured clinical notes into structured data, thus supporting clinical decision-making processes. This study explores the implementation of both traditional machine learning methods – such as Decision Trees (DT), Multilayer Perceptron (MLP)and advanced deep learning techniques in conjunction with NLP to diagnose heart conditions requiring catheter intervention.* ***Method:*** *This study explores the implementation of both traditional machine learning methods – such as Decision Trees (DT), Multilayer Perceptron (MLP)and advanced deep learning techniques in conjunction with NLP to diagnose heart conditions requiring catheter intervention.* ***Results:*** *Our findings suggest that the hybrid model employing deep learning methods outperforms traditional models, demonstrating the potential of AI in advancing cardiovascular healthcare.* ***Novelty:*** *Our findings suggest that the hybrid model employing deep learning methods outperforms traditional models, demonstrating the potential of AI in advancing cardiovascular healthcare.*

INTRODUCTION

Cardiovascular diseases (CVDs) affect over 6 million individuals in the US and rank among the leading causes of premature death worldwide, despite limited evidence supporting cardiac catheterization [1], [2]. Governments face considerable social and economic burdens, including a substantial increase in healthcare expenditures. One of the main underlying causes of many CVDs, such as peripheral vascular disease, venous thromboembolism, coronary artery disease, and cerebrovascular disease, is atherosclerosis. These conditions might eventually result in myocardial infarction, arrhythmia, or stroke. Numerous epidemiological studies have shown that a combination of variables, including physical inactivity, obesity, diabetes, hypertension, smoking, and hyper lipidaemia, accounts for about more than 90% of the risk associated with cardiovascular diseases, as demonstrated in numerous epidemiological studies [3].

Artificial intelligence technologies have become an integral part of various aspects of life these days, stimulating development and improvement in many fields. It is worth noting that artificial intelligence has made significant contributions to the medical sector by increasing the speed and accuracy of diagnosis and developing assistive technologies for treating various diseases. Artificial intelligence mimics the human brain's ability to

process data, and plays a crucial role in medicine by identifying, processing, integrating and analyzing vast amounts of medical information. This includes medical records, ultrasound images, drug data, drug interactions, clinical studies and experimental results [4]. Decision-making algorithms may soon help operators and planners. experts predict implementation within a few years in interventional cardiology [5]. Studies show that robotic-assisted PCI reduces radiation exposure for lab staff. It improves safety and precision in cardiac procedures [6].

Predictive modeling in cardiology has advanced significantly, with researchers developing several models to predict CVDs prognosis based on diverse patient groups. Notable examples include the Seattle Heart Failure Model (SHFM) [7], the Enhanced Efficient Cardiac Therapy (EFFECT) model [8], and the National Acute Decompensated Heart Failure Registry (ADHERE) [9], each offering unique insights into cardiovascular understanding. These models leverage distinct patient cohorts and emphasize different aspects of cardiac health, improving prognostic precision and guiding therapeutic strategies in clinical cardiology, thus representing substantial progress in evidence-based CVDs medicine.

One of the most significant applications is in the use of electronic health records (EHRs) that systematically gather and retain a variety of medical information from specific individuals throughout time [10]. Complicated, unstructured clinical notes may be parsed and transformed into organized, useable data using natural language processing (NLP) techniques. NLP reads, interprets, and systematically arranges vital health information embedded in free-text by utilizing advanced language tools [11]. NLP algorithms perform a variety of linguistic tasks such as information extraction, syntactic processing, and semantic analysis [12]. Despite limitations and challenges in the healthcare sector due to complex medical terminology, The wide use of NLP's in healthcare applications highlights its potential to improve data management and clinical decisions [13], [14]. Our study explored traditional and advanced machine learning with NLP for heart problem diagnosis, focusing on patients needing catheter intervention and determining intervention types: diagnostic, therapeutic, or renal. We examined traditional methods like Decision Trees (DT) and Multilayer Perceptron (MLP), alongside a designed hybrid deep learning model, with results suggesting the hybrid approach outperforms traditional models.

Related Work

The negative impact of heart attack and stroke has spurred extensive research to enhance its management and diagnosis through data-driven approaches. There are many studies on the prediction of strokes. In [15], a study conducts a systematic analysis of various features within EHRs for stroke prediction, employing principal component analysis (PCA) to reduce the high-dimensional feature space into a lower-dimensional subspace. Three widely-used classification models—neural network (NN), DT, and random forest (RF)—were implemented, with the NN demonstrating superior performance, achieving an accuracy of 78% and a miss rate of 19%. These findings offer valuable insights for enhancing precision in patient management within clinical practice.

The researchers in [16] focused on applying machine learning (ML) and deep learning (DL) techniques for early prediction of stroke. Several models, such as LR and DT, were compared with a hybrid Neural Network-Random Forest (NN-RF) approach. To handle class imbalance problem in stroke prediction datasets, oversampling techniques including SMOTE and ADASYN were employed. The hybrid NN-RF model with ADASYN oversampling showed superior performance with an F1 score of 75% and an accuracy of 84%, outperforming other models. These results highlight the promise of ML technique to improve stroke prediction and healthcare outcomes.

Several studies have investigated the link between stroke and heart attacks. Ref [17]. introduces the UCO (Undersampling-Clustering-Oversampling) algorithm, which enhances heart attack prediction in stroke patients by balancing the data through under sampling, clustering, and oversampling techniques. On the MIMIC-III dataset, the random forest classifier achieved the highest accuracy (70.29%) and precision (70.05%), significantly improving machine learning model performance.

Other studies have explored machine learning models for cardiovascular disease (CVD) risk prediction. Researchers in [18] developed a Genetic Algorithm (GA)-based model for CVD prediction using the Isfahan Cohort Study (ICS) data, showing superior performance over traditional models like Framingham, PROCAM, and Deep Learning, with an AUROC of 0.76.

Furthermore [19], developed machine learning models, such as XGBoost and random forests, to predict late recurrence of atrial fibrillation (LRAF) following catheter ablation. in sample comprised of 201patients. These models use clinical, biomarker, and procedural data to enhance patient selection, achieving high predictive accuracy and improving clinical decision-making. A recent study of [20] aimed to optimize risk stratification for major adverse cardiac events (MACE) using ML models derived from myocardial perfusion imaging (MPI) data, integrating both clinical and imaging variables. While imaging variables are automatically generated, clinical data requires manual input, which is time-consuming. The ML models were incrementally trained, identifying the minimum number of manually collected and imaging variables required to maintain prognostic accuracy. Models with fewer input variables, achieved comparable performance to full models (AUC 0.798 vs. 0.799) and did better than traditional methods, making them more practical for use in the clinic implementation. The study [21] has demonstrated the effectiveness of DL models such as RNN-LSTM for the prediction of CVD. These models utilizing time-series health data and outperformed traditional Cox regression models providing significant clinical risk factors. Our literature review revealed a lack in the research regarding the specific intervention needed for patients undergoing catheter based procedures. This gap complicates the ability to make proactive decisions and prepare the necessary equipment beforehand. Our study aims to fill this gap by investigating this issue.

RESEARCH METHOD

Figure 1 illustrates the method of this research. As shown, the achievement of the objectives of the study includes three stages with several steps. The first stage is called the pre-processing step which includes filtering and cleaning unstructured data and then tokenizing and padding sequences. The second stage is about feature extraction and selection, to find the most relevant features. Finally, the third step is classification and evaluation.

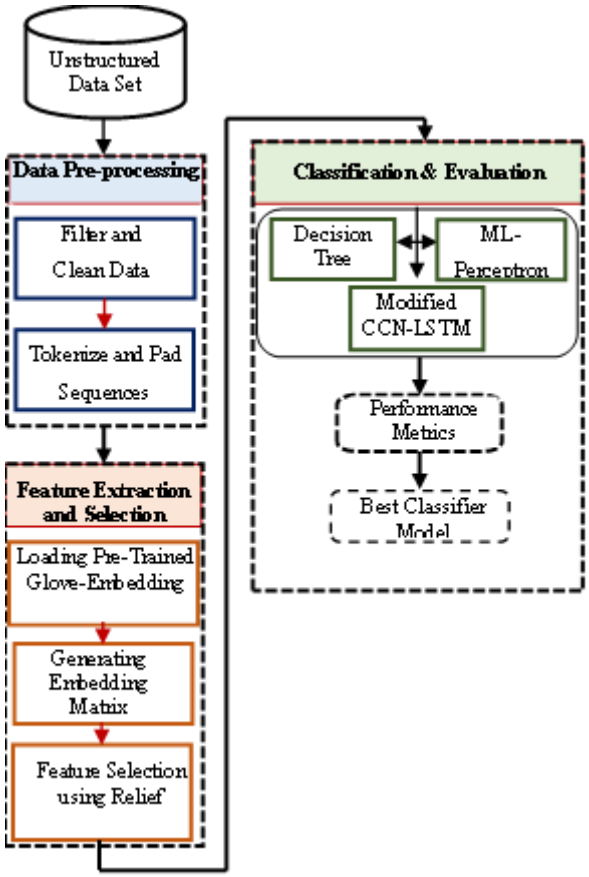

Figure 1. Illustration of the architecture of the proposed model.

2.1 Patients and Data Source

EHRs data was collected from hospitalized patients at Alkafeel Super Specialty Hospital in Karbala, Iraq, over a two-year period from June 2022 to June 2024. This study included patients who underwent diagnostic, therapeutic, or renal catheterization, all treated according to hospital guidelines. The database comprises information on 1,000 hospitalized patients who received medical care. The dataset includes various patient-level variables, such as patient history, medications, investigations, diagnoses, problem lists, and vital signs. In addition to structured data, EHRs encompass a significant amount of unstructured data, often referred to as 'free text', such as progress notes. This unstructured data may hold crucial patient information that could be missing, complementary, or even contradictory to the structured data fields. The inherent complexities of unstructured data, combined with the limitations of existing text mining

tools and NLP applications in accurately extracting information due to low data quality (e.g., missing data and input errors), pose a major challenge in achieving this task.

2.2 Data Preprocessing

Preprocessing is primarily intended to prepare data appropriately and make it usable for data mining techniques. This stage is divided into two separate phases. The first was the pre-processing of unstructured data, while the second was tokenization and pad sequences, which made it suitable for text mining algorithms. These pre-processing steps included:

2.2.1 Filter and Clean Data

The process consists of four steps as shown in Figure 2:

1. **Conversion to Lower Case**: All characters are converted to lowercase to simplify the NLP task.

Removing Symbols, Special Characters, and Punctuation: It is essential for cleaning and preparing text data from un unnecessary elements such as '!', '"', '#', '$', '%', '&', "'", '(', ')', '*', '+', ',', '-', '.', '/', ':', ';', '?', '[', '\', ']', '_', '`', '{', '|', '}', and '~' that do not contribute to the semantic meaning of the text. These irrelevant symbols, if retained, can introduce noise and skew model predictions. By eliminating them, the complexity of the text is reduced, facilitating more accurate feature extraction and improving tokenization accuracy. This preprocessing step ensures a more standardized and reliable input for downstream text analysis.

2. **Lemmatization**: Words are returned to their common base forms, excluding additions due to linguistic rules.

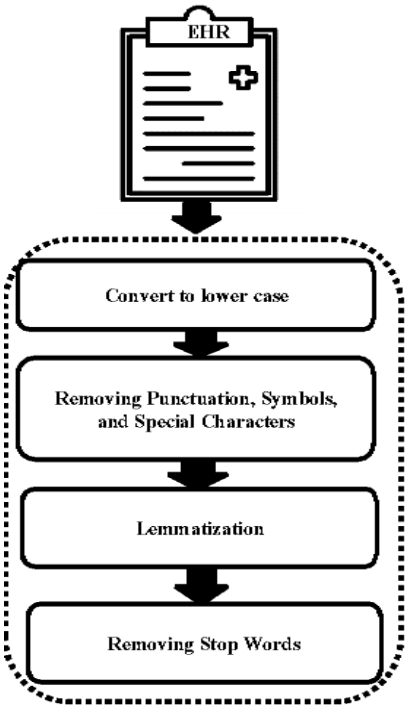

Figure 2. Illustration of the architecture of the preprocessing stage.

3. **Removing Stop Words**: Both English and nondomain stop words, which contribute little to the overall meaning of the text (e.g., "the", "to",…, "this"), are filtered out.

For this study, the NLTK library was used along with a predefined set of medical terms (e.g., "tb", "mg", "patient", …, "lt") to clean the data of stop words.

2.2.2 Tokenize and Pad Sequences:

Tokenizer transforms each text in a list into a sequence of integers, each unique word being assigned a distinct integer that corresponds to its index in the tokenizer's dictionary. The dictionary(attribute) has words as keys and their corresponding integer indices as values. The pad_sequences function ensures that all sequences have a uniform length by padding them with zeros or truncating them if they exceed the specified maximum length. In this work, sequences are padded to a maximum length of 1000. This step is crucial for making sure that all input data has the same shape, which is necessary for ML models.

2.3 Feature Extraction and Selection

- **Feature Extraction:** In this step, we exploit the previously known pre-trained pre-word embeddings from the GloVe (Global Vectors for Word Representation) model, specifically utilizing the 'glove.6B.100d.txt' resource. (to build a matrix of the vocabulary found in the EHR dataset from the previous step. Each row generated in this matrix represents a vector of coefficients that incorporate semantic information from GloVe into the ML model, thereby enhancing its performance on NLP tasks.
- **Feature Selection (*Relief*):** The use of feature selection techniques is crucial for improving the predictive performance of advanced models like convolutional and recurrent neural networks, particularly in text classification. ReliefF is an effective feature selection method that determines which features are most relevant by assessing their significance concerning the target variable. Applying the ReliefF feature selection method to the CNN-LSTM model significantly streamlines the input data by removing unnecessary or redundant features that might create noise or lead to overfitting.

As a result, the CNN-LSTM model operates with a more curated set of features, allowing it to learn more effectively and concentrate on the most important elements of the data, which enhances its classification capabilities. Conversely, without feature selection the model processes a higher volume of input features, including those that may not contribute meaningfully to the task. This can dilute the model's learning capacity, increase computational complexity, and potentially degrade its accuracy. Therefore, the integration of ReliefF feature selection with CNN-LSTM models not only optimizes the feature set but also enhances the model's accuracy by ensuring that only the most informative features are utilized during the training process.

2.4 Classification & Evaluation

2.4.1 Decision Tree

DT is one of the widely used powerful techniques in various fields, such as ML, NLP, image processing and pattern recognition [22]. Each tree composed of nodes and branches, nodes represent features of the text data to be classified and each subset defines a value that can be taken by the node [23]. DT commonly used versions are:

Iterative Dichotomies 3 (ID3), Successor of ID3 (C4.5), Classification And Regression Tree(CART)[24]. It optimally partitions a space of possible observations by subsequent recursive splits.

In the context of NLP, DT algorithms function through the iterative segmentation of text based on attributes value, including but not limited to: the presence or absence of specific lexical units, term frequency-inverse document frequency (TF-IDF) metrics, n-gram sequences, and various other syntactic or semantic properties[25]. This recursive partitioning process facilitates the hierarchical organization of textual data, enabling the construction of a tree-like structure that can be utilized for classification tasks in NLP applications.

2.4.2 Modified CCN-LSTM

This architecture implements a hybrid CNN-LSTM model as shown in Figure 3. It is designed for multiclass classification tasks. The model's initial layer is an embedding layer, which transforms input sequences into dense vectors of fixed dimensions. Following this, a convolutional layer with 128 filters and a kernel size of 5 is applied to capture local patterns. The subsequent max pooling layer, with a pool size of 4, performs feature down-sampling by selecting maximum values. These CNN and max pooling layers collectively extract localized features from the input. To capture long-range dependencies within the sequences, a bidirectional LSTM layer is incorporated. The model employs dropout layers at strategic points to combat overfitting. The final layer utilizes softmax activation, enabling multiclass classification by outputting probability distributions across the target classes.

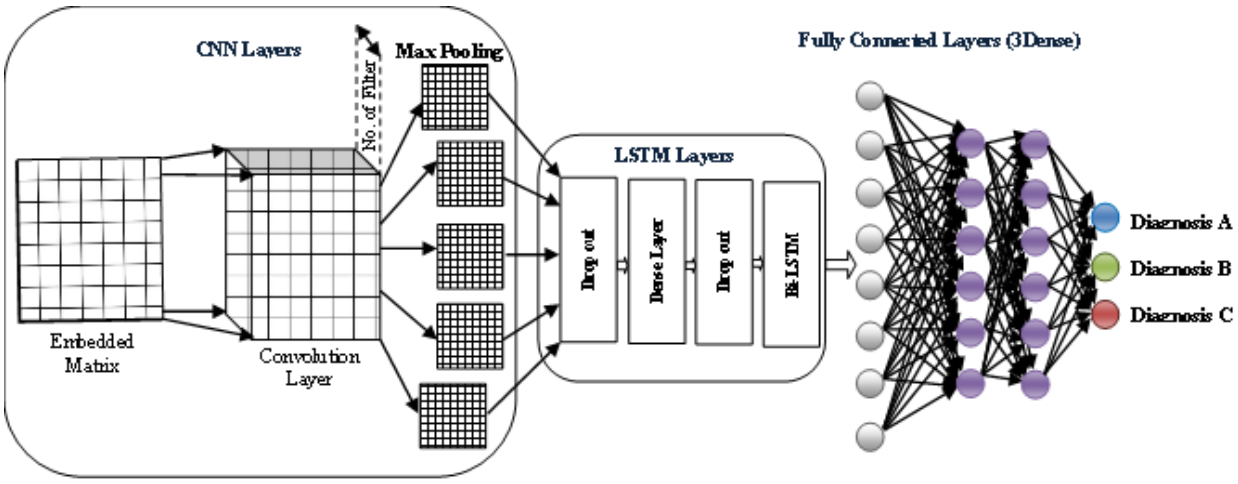

Figure 3. Illustration of the architecture of the proposed model.

2.4.3 Multilayer Perceptron (MLP)

MLP represents a category of artificial neural networks characterized by the presence of one or more hidden layers. In this architecture, each layer comprises nodes or neurons that employ nonlinear activation functions, with the sigmoid function being a prevalent choice. An MLP consists of an input layer, one or more intermediate hidden layers, and an output layer. The nodes are connected in a feed-forward manner [16].

Generally, MLPs are trained using the backpropagation (BP) algorithm that has two phases. The forward phase is the propagation of input signals through the network from the input layer to the output layer, passing through the hidden layers, where each node performs a calculation according to its activation function. The backward phase uses the difference between the desired and actual output values to propagate error signals from the output layer to the input layer, so that synaptic weights can be modified. This bidirectional process allows the network to iteratively optimize its parameters and enhance its predictive accuracy.

RESULTS AND DISCUSSION

Results

The dataset is made up of 998patients; 67.4% of them were males and 32.6% were females. 24.23% of participants were aged between 1- 50 years old, 55.4% were aged 51-70 years old, and 19.7% were greater than 70 years old. 98.5% of our sample were Iraqi who underwent catheterization procedures highlighted in Figure 4. These operations were collected from Alkafeel super specialty hospital, Iraq, it is already sorted qualitatively with Diagnosis, Therapeutic, Dialysis and Spinal canal catheterization. Spinal canal catheterization was excluded due to the criterion set to disregard any category with fewer than 100 samples.

The study included 1194 patients;

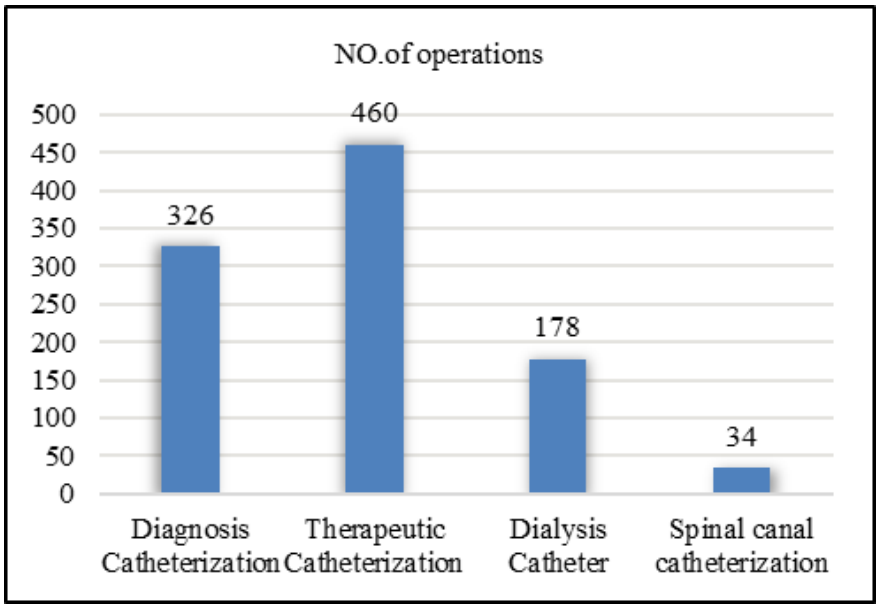

Figure 4. Class Distribution of Catheterization Dataset.

The classification accuracy metric of the three-implemented algorithm is depicted in Table 1 and Figure 5. The results show that CNN-LSTM model achieved the highest accuracy (93.93%) when the full extracted features from unstructured dataset is utilized. In addition, FS enhanced the accuracy of our model to (94.48%) and achieved the best classification performance in comparison with other methods.

Table 1. The Performance Metric for Textual data without FS(Relief).

TEXT MINING ALGORITHM	*METRICS*			
	Accuracy	*Precision*	*Recall*	*F1-Score*
CNN-LSTM	93.93%	94.12%	93.93%	93.90%
DT	59.47%	61.84%	59.47%	59.49%
MLP	50.90%	50.90%	50.75%	49.81%

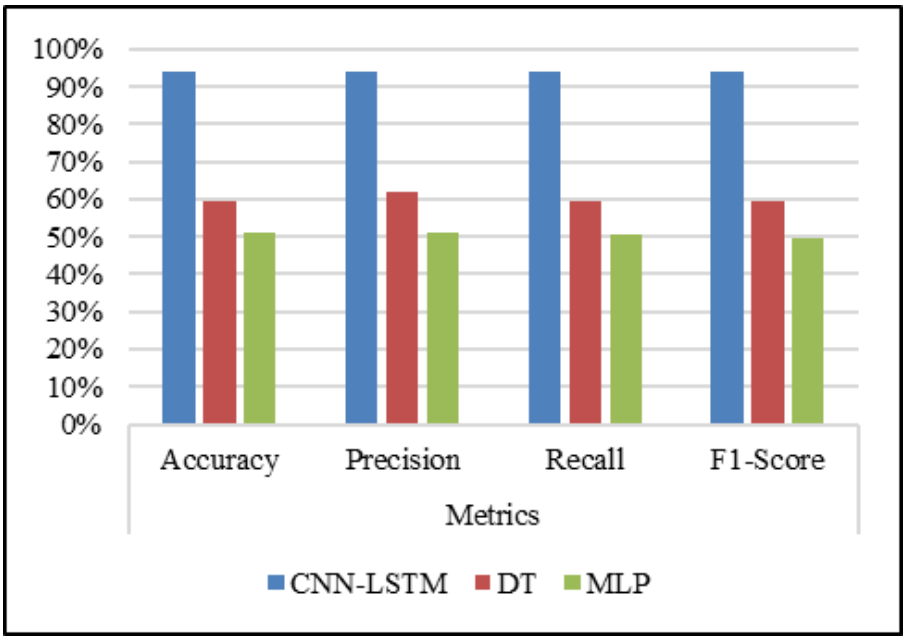

Figure 5. The Accuracy of the Four Classifiers without the FS technique.

The proposed CNN-LSTM model's superior performance is due to its ability to handle sequential data, utilize pre-trained embeddings, and leverage the strengths of both convolutional and recurrent layers to capture both local and long-range dependencies in the text. Its complex architecture allows it to learn and generalize better from the data, resulting in higher accuracy and balanced precision, recall, and F1 scores.

Table 2. The Performance Metric for Textual data with FS (Relief).

TEXT MINING ALGORITHM	*METRICS*			
	Accuracy	***Precision***	***Recall***	***F1-Score***
CNN-LSTM	94.48%	94.64%	94.48%	94.45%
DT	61.26%	62.79%	61.26%	61.01%
MLP	47.00%	47.82%	47.00%	46.85%

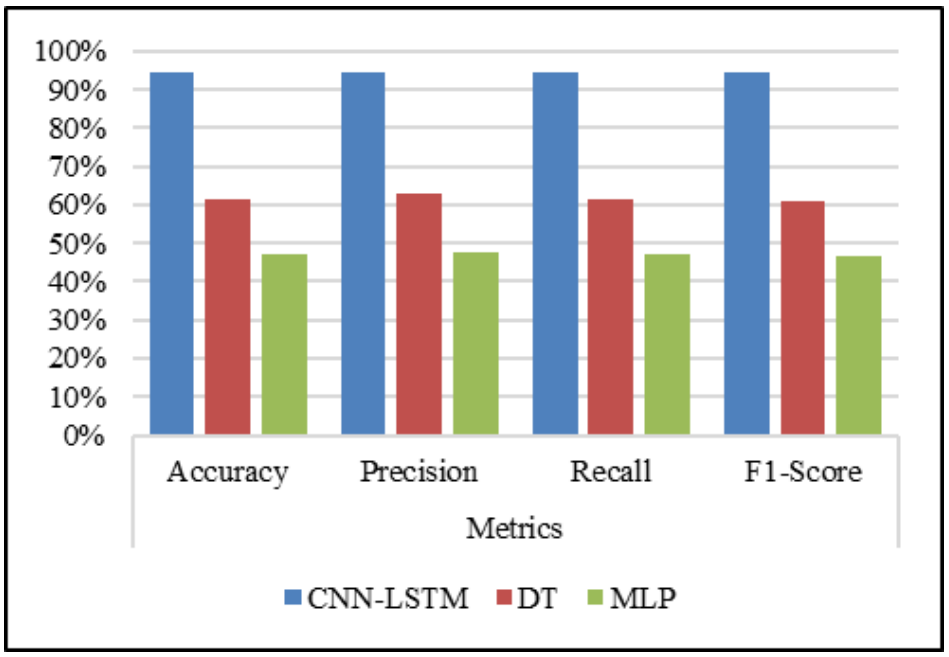

Figure 6. The Accuracy of the Four Classifiers with the Relief FS technique.

The classification accuracy metric of the three-implemented algorithm with FS is depicted in Table 2 and Figure 6.

Discussion

The findings of this study demonstrate that the proposed hybrid CNN-LSTM model provides a stronger predictive performance than the traditional Decision Tree and Multilayer Perceptron models in classifying catheterization intervention types from unstructured electronic health record data. The model achieved an accuracy of 93.93% without feature selection and improved to 94.48% after applying the ReliefF feature selection technique. This improvement indicates that selecting the most relevant textual features can enhance model performance by reducing redundant or noisy information in clinical narratives. In contrast, the Decision Tree and MLP models produced considerably lower performance, suggesting that conventional machine learning models may be less effective in handling complex and sequential clinical text data.

The superior performance of the CNN-LSTM model can be explained by its ability to combine local feature extraction and sequential dependency learning. The convolutional layer captures important local patterns and phrase-level information from the medical text, while the LSTM component preserves contextual relationships across longer sequences. This capability is particularly important in clinical documentation, where the meaning of a diagnosis or intervention may depend on the relationship between symptoms, medical history, investigations, and procedural notes. Therefore, the hybrid architecture is more suitable for processing unstructured EHR text compared with simpler models that rely mainly on surface-level feature separation.

The results also confirm the importance of NLP-based preprocessing in improving the quality of clinical text analysis. The original dataset contained noisy, unstructured, multilingual, and inconsistent textual data, including Arabic, English, Latin symbols, missing values, and input errors. Such characteristics are common in real-world hospital information management systems and often create difficulties for automated

classification. By applying lowercasing, punctuation removal, lemmatization, stop-word removal, tokenization, padding, and word embedding, the study transformed raw textual data into a more structured and machine-readable form. This process allowed the model to extract meaningful clinical information from otherwise fragmented records.

The application of ReliefF feature selection contributed positively to the CNN-LSTM and Decision Tree models, although its effect was not equally beneficial for all algorithms. For CNN-LSTM, ReliefF improved accuracy, precision, recall, and F1-score, showing that the model benefited from a more focused feature representation. Similarly, the Decision Tree model showed a slight improvement after feature selection, increasing from 59.47% to 61.26% accuracy. However, the MLP model decreased from 50.90% to 47.00%, suggesting that the selected feature subset may not have been sufficient for this architecture or that MLP required a different feature representation strategy. This finding highlights that feature selection methods should be carefully matched with the characteristics of each learning algorithm.

From a clinical perspective, the proposed model has practical potential to support early decision-making in catheterization procedures. Accurate identification of whether a patient requires diagnostic, therapeutic, or renal catheterization can help medical teams prepare appropriate resources, equipment, and procedural plans in advance. This is especially relevant in busy hospital environments, where delayed classification of intervention type may affect workflow efficiency and patient management. The model may also reduce the burden on clinicians by assisting in the rapid interpretation of large volumes of unstructured EHR data.

Despite these promising results, several limitations should be acknowledged. First, the study was conducted using data from a single hospital, which may limit the generalizability of the findings to other healthcare institutions with different documentation styles, patient populations, or clinical workflows. Second, the dataset included multilingual and noisy text, but the study did not deeply explain how language-specific medical terms, abbreviations, and spelling variations were standardized. Third, the study mainly evaluated classification performance using accuracy, precision, recall, and F1-score, but did not include external validation, explainability analysis, or clinical usability assessment. These aspects are important before the model can be implemented in real clinical decision-support systems.

Another important issue is model interpretability. Although CNN-LSTM achieved the best performance, deep learning models are often considered black-box systems. In medical applications, clinicians need not only accurate predictions but also clear explanations of why a model produces a specific output. Therefore, future work should incorporate explainable AI techniques, such as SHAP, LIME, or attention visualization, to identify which words, phrases, or clinical features most strongly influence the classification decision. This would increase clinician trust and support safer integration into hospital decision-making processes.

Future studies should also expand the dataset by including multiple hospitals and larger patient populations. A multicenter dataset would allow the model to learn from

more diverse clinical documentation patterns and improve its robustness across different healthcare settings. In addition, future research could compare CNN-LSTM with more recent transformer-based language models, such as BERT, ClinicalBERT, or multilingual medical language models. These models may provide better contextual understanding, especially for complex and multilingual clinical records. Finally, prospective validation in a real hospital environment is recommended to evaluate whether the proposed model can improve workflow efficiency, intervention planning, and patient outcomes.

Overall, the findings suggest that integrating NLP, deep learning, and feature selection is a promising approach for improving cardiovascular care, particularly in classifying catheterization intervention needs from unstructured EHR data. The study contributes to the growing evidence that AI-driven clinical text mining can support early intervention, enhance diagnostic workflows, and improve decision-making in cardiology. However, broader validation, stronger interpretability, and clinical implementation studies are required to ensure that the model is reliable, transparent, and applicable in real-world healthcare practice.

CONCLUSION

Fundamental Finding : In this study, ML algorithms such as (CNN, LSTM,DT, and MLP), were employed to diagnose patients who need catheterization interventions and recognize the type of intervention: whether diagnostic, therapeutic, or renal. The dataset was obtained from HIMS of Al-Kafeel Specialized Hospital located in Karbala, Iraq. The proposed model consists of multiple stages. Initially, data preprocessing techniques were utilized to filter and clean the dataset. NLP methods were then applied to extract relevant features from the textual data. Subsequently, the most impactful features were selected using a ReliefF feature selection technique. These features were fed into traditional AI algorithms and a hybrid deep learning model. The findings of the study reveal that the hybrid deep learning model achieved a high level of performance (94.48%) as it produced the best results when compared to other approaches on metrics such as (accuracy, F1-measure, recall, and precision). **Implication :** This research indicates a great opportunity to improve heart disease patients' diagnostic process and provide accurate recommendations for app usage DL techniques to enhance it. **Limitation :** It is noted that this dataset lacks of structure, noisy, input errors and multilingual (Arabic, English and Latin symbols) words. **Future Research :** The results of this study may serve as a theoretical and practical foundation for future scientific research and for the development of effective youth support strategies in Uzbekistan.

REFERENCES

[1] F. L. J. Visseren *et al.*, "ESC National Cardiac Societies; ESC Scientific Document Group. 2021 ESC Guidelines on cardiovascular disease prevention in clinical practice," *Eur Hear. J*, vol. 42, no. 34, pp. 3227–3337, 2021.

[2] P. Patel *et al.*, "Cardiac Catheterization and Outcomes for Elderly Patients Hospitalized With Heart Failure," *Heal. Serv. Insights*, vol. 17, p. 11786329231224616, 2024.

[3] C. McGorrian *et al.*, "Estimating modifiable coronary heart disease risk in multiple regions of the world: the INTERHEART Modifiable Risk Score," *Eur. Heart J.*, vol. 32, no. 5, pp. 581–589, 2011.

[4] L. Montull, A. Slapšinskaitė-Dackevičienė, J. Kiely, R. Hristovski, and N. Balagué, "Integrative proposals of sports monitoring: subjective outperforms objective monitoring," *Sport. Med.*, vol. 8, no. 1, p. 41, 2022.

[5] P. Sardar, J. D. Abbott, A. Kundu, H. D. Aronow, J. F. Granada, and J. Giri, "Impact of artificial intelligence on interventional cardiology: from decision-making aid to advanced interventional procedure assistance," *Cardiovasc. Interv.*, vol. 12, no. 14, pp. 1293–1303, 2019.

[6] T.-H. Lin *et al.*, "Clinical outcomes of multivessel coronary artery disease patients revascularized by robot-assisted vs conventional standard coronary artery bypass graft surgeries in real-world practice," *Medicine (Baltimore).*, vol. 100, no. 3, p. e23830, 2021.

[7] S. Li, P. Marcus, J. Núñez, E. Núñez, J. Sanchis, and W. C. Levy, "Validity of the Seattle Heart Failure Model after heart failure hospitalization," *ESC Hear. Fail.*, vol. 6, no. 3, pp. 509–515, 2019.

[8] D. S. Lee, P. C. Austin, J. L. Rouleau, P. P. Liu, D. Naimark, and J. V Tu, "Predicting mortality among patients hospitalized for heart failure: derivation and validation of a clinical model," *Jama*, vol. 290, no. 19, pp. 2581–2587, 2003.

[9] P. C. Austin, J. V Tu, and D. S. Lee, "Logistic regression had superior performance compared with regression trees for predicting in-hospital mortality in patients hospitalized with heart failure," *J. Clin. Epidemiol.*, vol. 63, no. 10, pp. 1145–1155, 2010.

[10] I. Li *et al.*, "Neural natural language processing for unstructured data in electronic health records: a review," *Comput. Sci. Rev.*, vol. 46, p. 100511, 2022.

[11] N. Mehta and A. Pandit, "Concurrence of big data analytics and healthcare: A systematic review," *Int. J. Med. Inform.*, vol. 114, pp. 57–65, 2018.

[12] T. A. Koleck, C. Dreisbach, P. E. Bourne, and S. Bakken, "Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review," *J. Am. Med. Informatics Assoc.*, vol. 26, no. 4, pp. 364–379, 2019.

[13] Y. Wang *et al.*, "Clinical information extraction applications: a literature review," *J. Biomed. Inform.*, vol. 77, pp. 34–49, 2018.

[14] W. W. M. Fleuren and W. Alkema, "Application of text mining in the biomedical domain," *Methods*, vol. 74, pp. 97–106, 2015.

[15] M. S. Pathan, A. Nag, M. M. Pathan, and S. Dev, "Analyzing the impact of feature selection on the accuracy of heart disease prediction," *Healthc. Anal.*, vol. 2, p. 100060, 2022.

[16] V. S. Elangovan, R. Devarajan, O. I. Khalaf, M. S. Sharif, and W. Elmedany, "Analysing an imbalanced stroke prediction dataset using machine learning techniques," *Karbala Int. J. Mod. Sci.*, vol. 10, no. 2, p. 8, 2024.

[17] M. Wang, X. Yao, and Y. Chen, "An imbalanced-data processing algorithm for the prediction of heart attack in stroke patients," *IEEE Access*, vol. 9, pp. 25394–25404, 2021.

[18] M. Ordikhani, M. Saniee Abadeh, C. Prugger, R. Hassannejad, N. Mohammadifard, and N. Sarrafzadegan, "An evolutionary machine learning algorithm for cardiovascular disease risk prediction," *PLoS One*, vol. 17, no. 7, p. e0271723, 2022.

[19] J. Budzianowski *et al.*, "Prediction of late atrial fibrillation recurrence after catheter ablation using machine learning," *Eur. Heart J.*, vol. 44, no. Supplement_2, pp. ehad655-509, 2023.

[20] R. Rios *et al.*, "Determining a minimum set of variables for machine learning cardiovascular event prediction: results from REFINE SPECT registry," *Cardiovasc. Res.*, vol. 118, no. 9, pp. 2152–2164, 2022.

[21] J. M. Sung *et al.*, "Development and verification of prediction models for preventing cardiovascular diseases," *PLoS One*, vol. 14, no. 9, p. e0222809, 2019.

[22] G. Stein, B. Chen, A. S. Wu, and K. A. Hua, "Decision tree classifier for network intrusion detection with GA-based feature selection," in *Proceedings of the 43rd annual Southeast regional conference-Volume 2*, 2005, pp. 136–141.

[23] B. Mahesh, "Machine learning algorithms-a review," *Int. J. Sci. Res. (IJSR).[Internet]*, vol. 9, no. 1, pp. 381–386, 2020.

[24] C. E. Brodley and P. E. Utgoff, "Multivariate decision trees," *Mach. Learn.*, vol. 19, pp. 45–77, 1995.

[25] T. Pranckevičius and V. Marcinkevičius, "Comparison of naive bayes, random forest, decision tree, support vector machines, and logistic regression classifiers for text reviews classification," *Balt. J. Mod. Comput.*, vol. 5, no. 2, p. 221, 2017.

*__Hussein Ali Al-jashamy (Corresponding Author)__
General Directorate of Education in Karbala, Karbala, Iraq
Email: hussein_ali_tuama@karbala.edu.iq

__Hashim Adnan__
Open Educational College, Department of Mathematics, Al-Muthanna Study Centre, Iraq

__Hussein Majid__
General Directorate of Education in Al-Muthanna, Iraq