DOI: 10.21917/ijsc.2026.0580

A MODIFIED APPROACH OF ANALYSING FEATURE EXTRACTION FROM PATIENT SENTIMENTS USING WORD EMBEDDING AND VECTOR REPRESENTATION

S. Punithavathy and J.M. Dhayashankar
Department of Computer Science, Sri Ramakrishna Mission Vidyalaya College of Arts and Science, India

Abstract

Sentiment analysis has been increasingly popular in the present digital era which attempts to analyse the consumer reviews acquired from websites, blogs and social media platforms. In hospitals and other healthcare organizations, understanding patient feedback helps to exceed in providing top-notch care. Sentiment analysis to enhance patient care is the way to know how patients feel about different service aspects, including processes, infrastructure, treatment, and healthcare professionals. Enhancing healthcare with sentiment analysis means removing human bias through consistent analysis, gaining real-time insights about patient satisfaction, and improving standards of care by incorporating patient feedback. The proposed work is structured into four phases to ensure a systematic approach to sentiment analysis using deep learning and NLP techniques. Each phase is designed to enhance data processing, feature extraction, feature selection, and sentiment classification, ensuring accurate and interpretable sentiment analysis in the healthcare domain. The proposed framework improves accuracy, interpretability, and feature selection while addressing key challenges in healthcare sentiment classification. This work will contribute significantly to the fields of Natural Language Processing, Healthcare AI, and Sentiment Analysis.

Keywords:

Sentiment Analysis, Healthcare, Feedback, Feature Extraction, Feature Selection

1. INTRODUCTION

Sentiment analysis, also known as opinion mining or emotion analysis, uses NLP and other methods to identify and extract subjective information from text, including attitudes, opinions, and emotions. In healthcare, it helps in understanding patients' feelings and opinions about treatments, medications, and overall care [1]. Integrating sentiment analysis into DTx can provide real-time insights into patients' emotional and psychological states, allowing for more personalized and responsive care. This contributes to the importance of sentiment analysis in healthcare.

One of the critical aspects of sentiment analysis in general and patient care sentiment analysis, in particular, is sentiment scoring and polarity classification.

Healthcare sentiment analysis focuses on diagnosing healthcare-related issues that people have discovered. To create rules and changes that could directly address patients' issues by considering their input. Machine learning techniques analyze millions of review documents and conclude them towards an efficient and accurate decision. Technology is moving at a faster and more innovative pace these days and because of this advancement and increase in the availability of internet tools and a growth in social media platforms for personal blogs, opinion sites, and websites that evaluate products online. People now use several platforms to convey their thoughts, feelings, and ideas and take advice on them. Which have drawn interest from interested parties like clients, businesses, and governments to examine and examine these viewpoints. Individuals use blogs and forums to talk about their ailments, symptoms, drugs, and other health-related subjects. There is also discussion on the accessibility, service, atmosphere, comfort, contentment, and other characteristics of the local healthcare facilities visited. Hearing about other people's experiences choosing a healthcare institution, taking medications, or making decisions regarding their health is often beneficial for new patients. Medical facilities require this information to identify and treat their patients' problems [2]. This information can help hospitals and healthcare providers better understand and handle their patients' interests and concerns. The study's strongest point is the patients' sharing of experiences covered by their sentiment analysis and passions. Sentiment analysis, also known as opinion mining, analyzes users' opinions, attitudes, and emotions expressed in natural language text. It categorizes this data as positive, negative, or neutral. Utilizing machine learning and other types of artificial intelligence technologies along with computational linguistics, sentiment analysis elicits valuable insights from various online communication channels, such as websites, reviews, forum discussions, blogs, and social media. This helps businesses understand customer sentiments [3].

Sentiment analysis in healthcare has gained significant attention due to its potential to improve patient satisfaction, clinical outcomes, and mental health monitoring. Existing methods primarily focus on rule-based and machine learning approaches, which often fail to capture complex emotional nuances and contextual sentiments in patient feedback [4]. Additionally, traditional sentiment analysis models lack explainability, making it difficult for healthcare providers to interpret the reasons behind sentiment classifications. This research aims to develop an advanced sentiment analysis framework integrating deep learning models with explainable AI (XAI) techniques to enhance the accuracy, interpretability, and reliability of patient sentiment classification.

2. RELATED WORK

In current scenario, healthcare industries are more focused on improving the experiences of patients. Patients have a lot of options to choose from varieties of services provided to them. So, Healthcare industries need to be more active in attracting and retaining patients. Sentiment analysis or opinion mining is a field that classifies comments or opinions into positive or negative that can be used to improve the quality. Sentiment analysis applied to patient's experiences is the systematic analysis of patient comments. Patient comments are broken into various parts and analysis is done on these components to measure how positive or negative is a patient about a particular component [5]. Patient's Word of Mouth is the key success of the any healthcare organizations. Some empirical literature review has been done that covers various sentiment analysis, emotion recognition, handwritten/typed text analysis, and Natural language processing, specifically.

Serrano-Guerrero [6] address the problem of detecting multiple emotions from patient reviews. The authors analysed, First, a large set of patient opinions was collected from a website that allowed patients to publish their experiences when visiting hospitals. Second, each opinion was labeled with the corresponding conveyed emotions. Third, a deep learning architecture based on a bidirectional gated recurrent unit with a multichannel convolutional neural network layer was proposed to detect multiple emotions from these reviews. Finally, the hyper parameters of this architecture were fine-tuned and different pertained word embedding models were configured to test its performance. To compare the results, different deep learning and machine learning algorithms have been implemented and configured to detect multiple emotions.

Khanbhai et al. [7] undertake a systematic review of the literature on the use of natural language processing (NLP) and machine learning (ML) to process and analyse free-text patient experience data [7]. Due to the heterogeneous nature of the studies, a narrative synthesis was deemed most appropriate. Data related to the study purpose, corpus, methodology, performance metrics and indicators of quality were recorded. Comments extracted from social media were analysed using an unsupervised approach, and free-text comments held within structured surveys were analysed using a supervised approach. Reported performance metrics included the precision, recall and F-measure, with support vector machine and Naïve Bayes being the best performing ML classifiers. To ensure that every patients' voice is heard, healthcare organisations must react and mould their language analysis strategy in line with the various patient feedback platforms.

Chen et al. [8] aimed to identify common discussion topics related to health care experience from the public's perspective and to determine areas of concern from patients' perspectives. Social media platforms allow individuals to openly gather, communicate, and share information about their interactions with health care services, becoming an essential supplemental means of understanding patient experience [8]. This study conducted a spatiotemporal analysis of the volume, sentiment, and topic of patient experience–related posts on the Weibo platform developed by Sina Corporation. The authors applied a supervised machine learning approach including human annotation and machine learning–based models for topic modeling and sentiment analysis of the public discourse. A multiclassifier voting method based on logistic regression, multinomial naïve Bayes, and random forest was used. The results of this study highlighted the interpersonal and functional aspects of care, especially the interpersonal aspects, which are often the "moment of truth" during a service encounter in which patients make a critical evaluation of hospital services.

Azam et al. [9] deals the bad emotional health can prompt social or psychological well-being issues. Recognizing or detecting feelings in online health care data gives important and helpful information regarding the emotional state of patients [9]. This work proposed a method for the automatic detection of patient's emotions in healthcare data using supervised machine learning approaches. It also performed a detailed comparison of the chosen six machine learning algorithms using different feature vectors on our dataset. We achieved the highest 87% accuracy using MultiLayer Perceptron as compared to other state of the art models. The proposed methodology can be implemented for the development of real-time web-based system. This work can be extended by using some embeddings and deep learning models.

Rastogi et al. [10] aimed to study and validate a method for automatic emotion recognition in PwPD during daily living through autonomic signals acquired by acceptable and low-cost consumer technology [10]. In each group of participants, it was possible to find a combination of feature set and algorithm to reach a classification accuracy greater than 90%. The Random Forest reached the best performance in both groups and for both valence and arousal. For each classification task (valence or arousal, PwPD or controls), the best model was selected and the minimal feature set was found by performing a recursive feature elimination based on the Shapley value. A lower accuracy of appraisal emerged for arousal compared to valence. This work investigated for the first time automatic emotion recognition in PwPD through physiological autonomic signals demonstrating that it is possible to recognize emotions in these people by low-cost, widely acceptable and usable methods with high accuracy.

Chakrapani et al. [11] finds that, sentiment analysis plays a role in seeing their perspective; the significant broad medical application does not yet meet the analysis of patient mindset [11]. This work introduced, a practical framework to analyse patients' perspectives using a socio-medical dataset that contains various reviews and feedback of critical diseases-affected people. Initially, applied a pre-processing technique, including Lowercase Conversion, removing special characters, removing stop words, Number to word conversion, Stemming, and lemmatization over dataset. Next, N-gram tokenization methodology is used to extract the valuable features followed by assigning polarity score to each sentiment we extract and calculate the overall polarity of the context. Finally, a probabilistic LDA model was employed to combine the review.

Thus, from the above related work, the importance of sentiment analysis in the area of healthcare domain and the various aspects and approaches related to sentiment analysis has been touched by various authors in the different domains of healthcare. Overall, these papers highlight the importance of sentiment analysis in the healthcare domain and hence motivate us to apply robust machine learning techniques which can provide accurate results and provide insights into the latest techniques and approaches used in this area.

3. RESEARCH METHODOLOGY

Medical and clinical reports, patients' feedback, and sentiments about medical systems and services are among the most valuable and useful textual content. The goal of healthcare sentiment analysis is to identify the health facilities and identify what people like or dislike about them. Patients complaints and suggestions can motivate the organizations to improve their facilities for the future. The various challenges in this field that can be derived from the work are as follows:

1. To find the authentic set of patient opinions for sentiment analysis.
2. Constructions of domain specific lexicon for health care.
3. Applicability of hybrid methods on aspect level to search the health related solutions closely matching to the need of patients.
4. Detection of irony and sarcastic sentences in patient comments.
5. Analysis of opinions in language other than English, i.e. multilingual sentiment analysis.

These are the challenges faced by researchers while working on sentiment analysis in health care.

3.1 PROBLEM STATEMENT

Sentiment analysis has been increasingly popular in the present digital era which attempts to analyse the consumer reviews acquired from websites, blogs and social media platforms. The rich textual information contained data sources, thus understood as consumers' reviews are very important to any particular domain as businesses are able to improve themselves in several aspects. Current sentiment analysis techniques in healthcare face the following challenges:

- Difficulty in processing complex medical and patient-related language, leading to misclassification.
- Limited feature extraction techniques, making it hard to capture nuanced sentiment variations.
- Inefficient feature selection, which reduces classification accuracy by including irrelevant features.
- Lack of an explainable AI component, making it difficult for healthcare providers to understand model decisions.

It is ML and DL models were trained and tested on the drugs.com dataset, which is a rich dataset for medication reviews.

3.2 RESEARCH OBJECTIVES

Sentiment analysis in healthcare is categorized based on the source of text (such as medical websites, biomedical publications, clinical notes), the method used (including polarity-based, classification, rules-based, machine learning), and the type of analysis conducted (like outcome classification). Moreover, sentiment analysis can also be categorized based on the level of analysis, such as word-level or sentence-level analysis. The primary objectives of this research are:

- Develop an effective sentiment analysis pipeline that systematically collects, preprocesses, and classifies healthcare-related sentiment data.
- Extract relevant features from text using NLP-based methods like TF-IDF, word embeddings (Word2Vec, GloVe, BERT), and topic modeling.
- Apply feature selection techniques enhanced feature selection to retain only the most meaningful features.
- Develop a hybrid deep learning-based classifier to improve sentiment classification accuracy.
- Enhance interpretability using Explainable AI (XAI) techniques like SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations).

Healthcare services have a number of drawbacks that need to be recognized. Patients have complaints about a variety of aspects (features) of hospitals, such as the cleanliness of the facility, the friendliness of the staff, the cost of the medications, and other factors. As a result, it's important for all of these aspects to remain excellent. The hospital needs to be the primary focus of improvement in healthcare services. The healthcare service is also changing as time passes, i.e., sometimes it is good and sometimes it is very bad. As a result, timestamp-based healthcare service analysis is critical.

4. PROPOSED SCHEME

Proposed research in sentiment analysis for healthcare can focus on enhancing patient outcomes by predicting mood disorders, personalizing treatment, improving patient-provider communication, and boosting treatment adherence. Other key areas include using sentiment analysis of patient feedback to improve services, developing more accurate machine learning models for analyzing clinical narratives, and using sentiment analysis for public health surveillance.

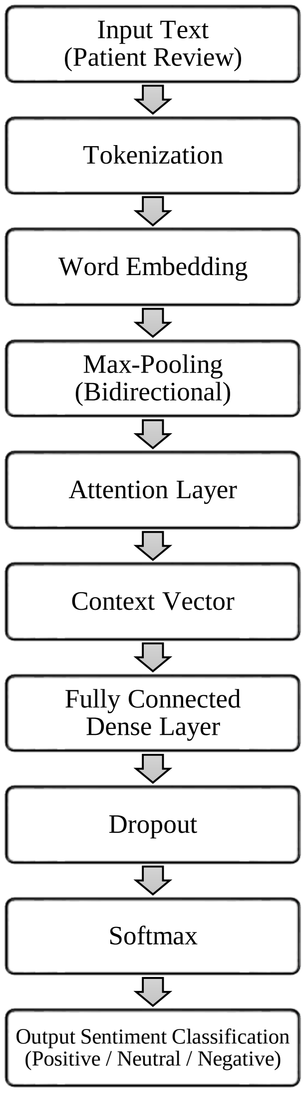

Fig.1. Architectural Design

4.1 PREPROCESSING THE DATA

The proposed research focuses on gathering patient feedback data from various sources such as social media, hospital surveys, and online healthcare forums. A novel pre-processing algorithm will be developed to clean and normalize the text data [12]. Preprocessing steps include tokenization, stopword removal, stemming, lemmatization, and handling negations to enhance sentiment polarity detection. This phase ensures that the input text is structured and noise-free for further analysis.

Preprocessing for healthcare sentiment analysis involves several key steps, including cleaning text data, tokenizing it into smaller units, and normalizing it to handle domain-specific language like medical jargon, abbreviations, and complex sentences. Other crucial steps include identifying and handling negation, dealing with sentiment-bearing entities and their aspects (like treatment side effects), and using domain-specific lexicons or models to improve accuracy. The preprocessing steps are:

- **Data collection**: Gather text data from various sources like patient reviews, social media, clinical notes, and online forums.
- **Text cleaning**: (1) **Remove noise**: Remove irrelevant information like HTML tags, URLs, and special characters. (2) **Correct spelling**: Fix common spelling mistakes and typos. Domain-specific misspellings can occur frequently.
- **Tokenization**: Break the clean text into individual words or tokens.
- **Lemmatization/Stemming**: Reduce words to their root form (e.g., “running,” “ran” become “run”) to reduce the vocabulary size and group similar words.
- **Stop word removal**: Remove common words that don’t carry significant sentiment, such as “the,” “is,” and “and.” However, be cautious with medical contexts, as some stop words might be part of a larger medical phrase.
- **Negation handling**: Identify and handle negation words (e.g., “not,” “no,” “never”) that can flip the sentiment of a phrase (e.g., “not good” is negative).

Denote $P\mathscr{R}$ as a collection of patient reviews containing documents dr_1,dr_2,…,dr_n, and $\mathcal{TK}$ as a set of word lists in each review containing post-processed tokens tk_1, tk_2,…,tk_m. In this study, for the purpose of review cleaning, text processing functions involving spell check (SpellChecker*()),* stemming (stem*()),* tokenization *(*tokenize()), Part-of-Speech (POS) tagging(*tag*()), and lemmatization (*lemmatize*()) are implemented.

4.2 FEATURE EXTRACTION USING WORD EMBEDDING AND VECTORIZATION

It refers to the process of using domain knowledge to select and transform the most relevant features from raw data during the creation of a predictive model using ML. The goal of feature engineering is to improve the performance of ML algorithms. The pre-processed data is represented using various combinations of traditional vector representation models such as BoW, Count Vectorizer/TF-IDF, N-grams and word embeddings like Word2Vec and BERT.

The process of creating text vectors for the input review dataset is given below

BoW:

- Step1: Split each review into individual tokens or words.
- Step2: Calculate the frequency of each word in the input review.
- Step3: Represent the review in the vector format where each element in the review corresponds to the count of a particular word.

TF-IDF:

- Step1: Count the number of times each word has occurred in the review.
- Step2: Calculate the inverse document frequency.
- Step3: Compute the TF-IDF score.
- Step4: Represent every review as a vector of TF-IDF scores where every element in the review corresponds to one word in the vocabulary.

Word2Vec:

- Step 1: Train the Word2Vec model on the Drug Review dataset.
- Step 2: Choose between the architectures (Skip-gram/ CBOW).
- Step 3: Represent each word as a dense vector based on the context within the reviews.

4.2.1 BERT (Bidirectional Encoder Representations from Transformers):

The BERT model is trained to identify positive and negative sentiments in the dataset. Hyper paramters of the model such as the number of iterations (epochs) and the dropout rates were fine-tuned to optimize the performance of SA tasks. The BERT parameters used is given below:

- ***Activation function***: For binary classification problems, the sigmoid activation function is a popular choice since it can transform input values into a probability distribution with a range of 0 to 1. This range is essential for the output to be understood as a likelihood of belonging to one of the two classes. Since the sigmoid function’s output may also be thought of as a probability, it is easy to set a decision threshold for class assignment. Due of these reasons, the sigmoid activation function is frequently used to solve binary classification problems.
- ***Optimizer*****:** The Adam optimizer is widely used in deep learning as it combines the advantages of momentum-based optimization techniques and adaptive learning rate approaches. It efficiently modifies the learning rate for each parameter based on the gradient size and records both the first and second moments of the gradients. This adjustable learning rate reduces the likelihood of encountering local minima while accelerating convergence. The momentum term in the Adam optimizer accelerates convergence and enhances generalization by collecting prior gradients. Because it accommodates sparse gradients well and is less susceptible to hyper paramters modification than other optimization methods, Adam is a preferred choice in this work.

- ***Binary cross-entropy loss***: A typical loss function in binary classification issues is logistic loss, often known as binary log loss. It calculates the difference between the actual binary labels and the expected probability. The formula for binary cross-entropy loss can be expressed using Eq.(1):

$$L=-[y*\log(p)+(1-y)*\log(1-p)] \quad (1)$$

- ***Dropout*:** A useful tool for neural network training, providing a straightforward but efficient regularization strategy to enhance generalization performance and fend off overfitting. Underfitting in neural networks can result in raising the dropout rate. Increased dropout rates can result in information loss, a decreased network's ability to learn complicated representations, and an overly regularized learning process. This may lead to a model that is improperly fitted and misses significant characteristics and trends in the data. Furthermore, overly high dropout rates can cause gradients to become unstable, which would impede convergence and the learning process. To avoid both overfitting and underfitting, the proper balance between regularization and model capacity must be struck. To get optimal performance, the dropout rate and other hyperparamters must be tuned properly. In this work, a 10% dropout rate has been considered thereby statistically giving the most accurate classification.
- ***Number of iterations*:** Up to a certain point, increasing the number of iterations or epochs can improve the accuracy of a model's training. This is the reason behind improved convergence. The model has more possibilities to fine-tune its weights and biases as it trains for more iteration, gradually minimizing the loss function. As a result, the model can converge towards a better answer, potentially increasing accuracy. By optimizing the model's parameters, training for more iteration makes the model more generalizable. It enhances the model's performance on unobserved examples by enabling it to capture more complex patterns and dependencies in the data. To fully identify the underlying patterns in some datasets or activities, several iterations may be necessary. Longer training times may be necessary for complex datasets with significant variation to attain higher accuracy. It is crucial to remember that continuing to iterate endlessly does not ensure that progress will be made. If the model begins to memorize the training data too thoroughly, overfitting might happen, which results in poor generalization of fresh data. To avoid overfitting and determine the right number of iterations to maximize accuracy, regularization approaches like early halting or learning rate decay may be required.

4.2.2 Word Embedding:

Due to sensitivity and privacy issues, publicly available medical text datasets are limited and merely used for non-commercial purposes. With adequate file processing steps, objective records and reports are removed from initial database and a corpus composed of around 4143 clinical documents is formed. The corpus is fed into a word2Vec object with skip–gram model in gensim, a well-developed open source Python topic and vector space modeling library to train medical domain specific word embeddings.

Post-trained medical word2Vec model is treated as similarity measurement to ameliorate current SWN lexicon by adding medical specific sentiment phases. Meanwhile, medical word embeddings are stored as a list of vector for further processing. Given a specific sentiment phrase, we compute the cosine similarity using word2Vec. If any returned association term is not in the existing SWN, it computes the sentiment value by multiplying a similarity value with the corresponding sentiment value of specific sentiment phrases in SWN.

4.2.3 Position Encoding in Vector Representation:

As discussed previously, drug reviews are rich in relational information. Hence, it is inadequate to use frequency or term-weighted feature representation methods. Considering the order of sequence of sentiment phases in feature representation is one direction of solving this variable length issue.

In this study, we revise the position encoding technique proposed to incorporate sentiment sequence into document-based sentiment mining, and we conduct experiments on the comparison among non-sequence and simple chronological sequence representation methods. Position encoding is derived from the attention model developed by which takes the order of the words into consideration and returns the scaled position of the words based on sentence indexing and dimension of the embeddings. As suggested, high dimensionality and feature redundancy are two major problems concerning current feature extraction methods [15]. Therefore, to reduce the computational complexity and minimize feature redundancy, it adjusts the embedding dimensionality of the above mentioned position encoding method to two.

5. PERFORMANCE EVALUATION

Out of the eight features in the drug review data, the algorithm was used to select the top N features. It should be mentioned that adding recent options from the previous set results in the new best feature set since the value of N is increased. This demonstrates that adding a feature has an impact on that particular subset's classification capacity, either positively or negatively. As a result, there is an optimal feature set where accuracy is maximized. The drug review dataset is subjected to a variety of machine learning techniques in order to evaluate the suggested implementation.

5.1 DATASET DETAILS

The dataset is named as Drug Reviews. The dataset provides patient reviews on specific drugs along with related conditions. Reviews and ratings are grouped into reports on the three aspects benefits, side effects and overall comment.

The dataset provides patient reviews on specific drugs along with related conditions. Furthermore, reviews are grouped into reports on the three aspects benefits, side effects and overall comment. Additionally, ratings are available concerning overall satisfaction as well as a 5 step side effect rating and a 5 step effectiveness rating. The data was obtained by crawling online pharmaceutical review sites. The intention was to study,

- Sentiment analysis of drug experience over multiple facets, i.e. sentiments learned on specific aspects such as effectiveness and side effects,

- The transferability of models among domains, i.e. conditions, and
- The transferability of models among different data sources (see 'Drug Review Dataset (Drugs.com)').

The data is split into a train (75%) a test (25%) partition (see publication).

5.2 PERFORMANCE ANALYSIS

5.2.1 Accuracy Comparison

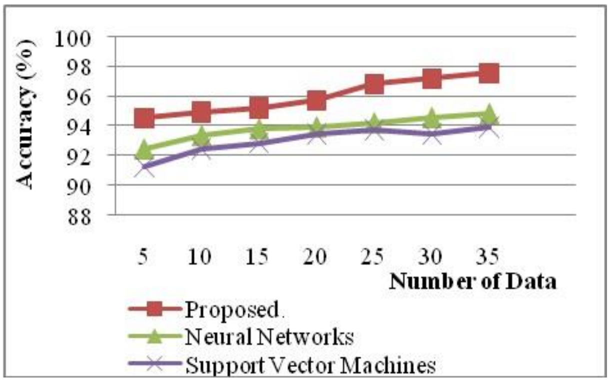

Fig.2. Prediction and Detection accuracy comparison

Since the suggested methods pick the key features in the dataset, the prediction and detection accuracy of the suggested proposed Word Embedding and Vector Representation based method achieves higher classification accuracy than the current classification methods like SVM and Neural Network. Accuracy result is shown in Fig.2.

5.2.2 Error Rate Comparison:

The Fig.3 shows the data classification error rates. According to the graph, the suggested Word Embedding and Vector Representation method produces generally low error rates. The reason is that the proposed algorithm, which is used in the proposed work for feature subset selection, allows a group of concurrent distributed agents to jointly identify the best features for a given dataset.

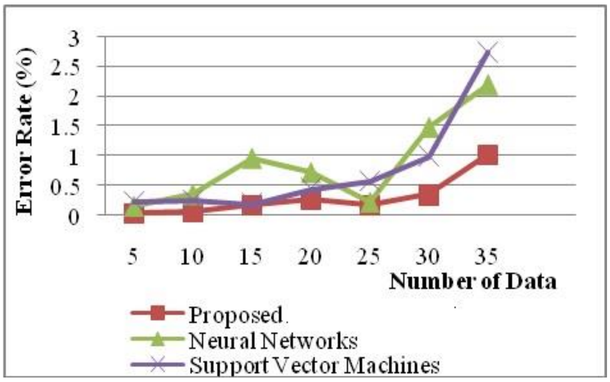

Fig.3. Error Rate Comparison Results

Additionally, compared to current techniques like SVM and neural networks, the classification methodology is far more effective.

5.2.3 Execution Time Comparison:

The results of comparing the execution times of the suggested Word Embedding and Vector Representation approach and the current techniques, including SVM and Neural Network, are displayed in Fig.4.

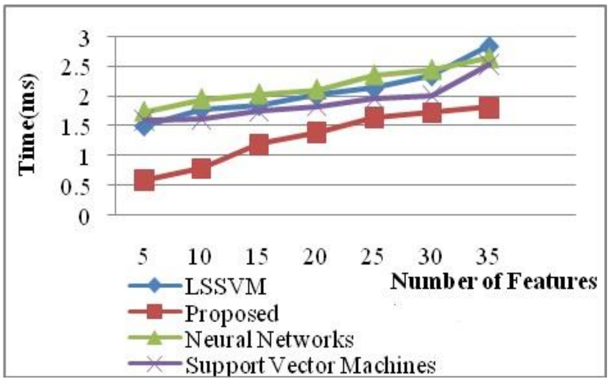

Fig.4. Execution Time comparison

Since the suggested method can accurately predict heart illness at an early stage, the suggested Word Embedding and Vector Representation methodology has a shorter execution time for diagnosing heart disease.

6. CONCLUSION

Lastly, Word Embedding and Vector Representation is used to define a decision support system for classifying cardiac disease. The best feature subset to maximize Word Embedding and Vector Representation's classification accuracy with fewer characteristics is found using a genetic algorithm. Support vectors are found quickly and iteratively using the Word Embedding and Vector Representation technique. The Word Embedding and Vector Representation performance was greatly improved by using genetic algorithms for feature selection, and a high accuracy of 97.55% was attained. In the future, the system can be improved to diagnose heart disease more accurately by employing a better optimization algorithm to build a model that produces an effective outcome and by adding more categorization techniques and algorithms.

REFERENCES

[1] S. Jadhav, B. Shiragapur, R. Purohit and N. Jha, "A Systematic Review on the Role of Sentiment Analysis in Healthcare", *Authorea*, Vol. 1, pp. 1-17, 2024.

[2] P.K. Nag, A. Bhagat, R.V. Priya and D.K. Khare, "Emotional Intelligence through Artificial Intelligence: NLP and Deep Learning in the Analysis of Healthcare Texts", *Proceedings of International Conference on Artificial Intelligence for Innovations in Healthcare Industries*, Vol. 1, pp. 1-7, 2023.

[3] A.M. Shah, X. Yan, S.A.A. Shah and G. Mamirkulova, "Mining Patient Opinion to Evaluate the Service Quality in Healthcare: A Deep-Learning Approach", *Journal of Ambient Intelligence and Humanized Computing*, Vol. 11, No. 7, pp. 2925-2942, 2020.

[4] S.T. Lai, R. Mafas, M. Wang and Y. Yu, “Sentiment Analysis in Healthcare: Motives, Challenges and Opportunities Pertaining to Machine Learning”, *Proceedings of International Conference on Distributed Computing and Electrical Circuits and Electronics*, Vol. 11, No. 11, pp. 1166-1174, 2020.

[5] F.J. Ramirez-Tinoco, G. Alor-Hernández, J.L. Sánchez-Cervantes, M.D.P. Salas-Zarate and R. Valencia-García, “Use of Sentiment Analysis Techniques in Healthcare Domain”, *Proceedings of International Conference on Current Trends in Semantic Web Technologies: Theory and Practice*, pp. 189-212, 2019.

[6] J. Serrano-Guerrero, M. Bani-Doumi, F.P. Romero and J.A. Olivas, “Understanding What Patients Think about Hospitals: A Deep Learning Approach for Detecting Emotions in Patient Opinions”, *Artificial Intelligence in Medicine*, Vol. 128, pp. 1-11, 2022.

[7] M. Khanbhai, P. Anyadi, J. Symons, K. Flott, A. Darzi and E. Mayer, “Applying Natural Language Processing and Machine Learning Techniques to Patient Experience Feedback: A Systematic Review”, *BMJ Health and Care Informatics*, Vol. 28, No. 1, pp. 1-8, 2021.

[8] X. Chen, Z. Shen, T. Guan, Y. Tao, Y. Kang and Y. Zhang, “Analyzing Patient Experience on Weibo: Machine Learning Approach to Topic Modeling and Sentiment Analysis”, *JMIR Medical Informatics*, Vol. 12, pp. 1-13, 2024.

[9] N. Azam, T. Ahmad and N.U. Haq, “Automatic Emotion Recognition in Healthcare Data using Supervised Machine Learning”, *PeerJ Computer Science*, Vol. 7, pp. 1-6, 2021.

[10] A. Rastogi, R. Singh, M.Z. Khan, I. Aljubayri and F.M. Ghabban, “Patient Emotion and Sentiment Analysis using Deep Learning”, *Journal of Innovative Image Processing*, Vol. 7, No. 3, pp. 582-601, 2025.

[11] K. Chakrapani, M. Kempanna, M.I. Safa, T. Kavitha, M. Ramachandran, V. Bhaskar and A. Kumar, “An Enhanced Exploration of Sentimental Analysis in HealthCare”, *Wireless Personal Communications*, Vol. 128, No. 2, pp. 901-922, 2023.

[12] M.E. Basiri, M. Abdar, M.A. Cifci, S. Nemati and U.R. Acharya, “A Novel Method for Sentiment Classification of Drug Reviews using Fusion of Deep and Machine Learning Techniques”, *Knowledge-based Systems*, Vol. 198, pp. 1-8, 2020.

[13] A.O. Almagrabi and S. Ahmad, “Sentiment Analysis System for Classification of Patient-Generated Health Reviews using Rough Set Theory and Machine Learning Technique”, *Journal of Medical Imaging and Health Informatics*, Vol. 10, No. 10, pp. 2361-2368, 2020.

[14] J. Polisena, M. Andellini, P. Salerno, S. Borsci, L. Pecchia and E. Iadanza, “Case Studies on the Use of Sentiment Analysis to Assess the Effectiveness and Safety of Health Technologies: A Scoping Review”, *IEEE Access*, Vol. 9, pp. 66043-66051, 2021.

[15] A.I.A. Rahim, M.I. Ibrahim, S.L. Chua and K.I. Musa, “Hospital Facebook Reviews Analysis using a Machine Learning Sentiment Analyzer and Quality Classifier”, *Healthcare*, Vol. 9, No. 12, pp. 1-11, 2021.