INTERNATIONAL CONFERENCE ON RESEARCHES IN ENGINEERING, SCIENCE, MANAGEMENT AND HUMANITIES (ICRESMH – 2025)

Delhi, India on 27th April, 2025

CERTIFICATE NO : ICRESMH /2025/C0425412

Enhancing Sentiment Classification Accuracy in Medical Datasets Using A Soft Voting Ensemble of Supervised Machine Learning Algorithms

Nayan Sarkar

Research Scholar, Dept. Of. Computer Science and Engineering, P.K. University, Shivpuri, M.P., India.

ABSTRACT

The rapid digitalization of healthcare services has generated vast amounts of unstructured textual data in the form of patient reviews and feedback. Extracting meaningful insights from this data is vital for improving healthcare quality and user satisfaction. This study focuses on sentiment analysis of medical reviews using supervised machine learning (ML) techniques combined with an ensemble learning approach. A comprehensive six-step pipeline was developed, including data collection through web scraping, labeling by domain experts, preprocessing, term vector generation using TF-IDF, model training, and accuracy evaluation. Logistic Regression (LR), Random Forest (RF), and Decision Tree (DT) algorithms were initially applied to classify sentiments as positive or negative. To enhance the overall sentiment classification performance, a Soft Voting Ensemble (SVE) algorithm was designed by integrating the predictions of the three classifiers. The SVE approach achieved superior accuracy compared to individual ML models, demonstrating its effectiveness in handling sentiment classification tasks within medical datasets. The results indicate that the proposed framework can efficiently forecast sentiment tones from medical text data, offering valuable decision-support insights for healthcare organizations aiming to understand and enhance patient experiences.

Keywords: *Sentiment Analysis, Healthcare Analytics, Patient Feedback, Ensemble Learning, Natural Language Processing.*

I. INTRODUCTION

In the modern healthcare ecosystem, understanding patient sentiment has emerged as a crucial determinant of institutional success, patient satisfaction, and quality of care. As the healthcare industry undergoes rapid digital transformation, patients are increasingly sharing their experiences, opinions, and emotions through online platforms, social media, and hospital feedback portals. These narratives—whether positive or negative—offer valuable insights into various dimensions of healthcare delivery, including hospital infrastructure, medical staff responsiveness, communication patterns, and perceived treatment quality. However, the sheer volume and unstructured nature of such data make it challenging for conventional analytical techniques to extract meaningful patterns. Manual evaluation of textual feedback is not only time-consuming but also prone to subjectivity and bias. Consequently, the use of automated sentiment analysis techniques has gained momentum as an essential component of intelligent healthcare analytics.

Sentiment analysis, a fusion of natural language processing (NLP), machine learning (ML), and text mining, enables systematic interpretation of patient emotions embedded in textual data. Traditional sentiment analysis models are typically rule-based or lexicon-driven, employing pre-defined sentiment dictionaries and polarity scoring to classify emotions as positive, negative, or neutral. These models are computationally efficient and easily interpretable but lack the sophistication required to understand linguistic subtleties such as sarcasm, irony, or negation. For example, phrases like "the doctor was not unhelpful" or "the treatment was surprisingly good" often confuse rule-based systems, leading to inaccurate sentiment classification. Such limitations have spurred the adoption of machine learning and deep learning approaches that automatically learn contextual sentiment features from large-scale datasets.

While ML algorithms like Support Vector Machines (SVM) and Naïve Bayes can learn statistical relationships between words and sentiment labels, they often struggle with imbalanced datasets and overfitting issues. Deep learning architectures, on the other hand, such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks, capture semantic and sequential dependencies in text but may still falter in healthcare-specific domains due to their reliance on generalized data representations. Therefore, there exists a critical need for a hybridized sentiment prediction framework that integrates the strengths of both machine learning and deep learning paradigms while overcoming their individual weaknesses.

To bridge this gap, the present research introduces a Novel SCSP Ensemble Method—a Semantic-Convolutional and Syntactic-Pattern-based (SCSP) framework designed specifically for healthcare sentiment prediction. The proposed SCSP ensemble integrates the semantic comprehension ability of deep neural models with the linguistic precision of syntactic analysis. It combines multiple classifiers—SVM, CNN, and Bidirectional LSTM (Bi-LSTM)—to leverage their unique feature extraction strengths. CNN effectively identifies local n-gram patterns and sentiment cues, while Bi-LSTM captures long-range dependencies and context flow within the feedback. The SVM, operating as a traditional classifier, provides stability and interpretability to the ensemble. This multi-level feature extraction mechanism ensures that both high-level semantic embeddings and low-level syntactic structures are considered, thereby creating a richer and more accurate sentiment representation.

The rationale behind developing the SCSP model stems from the complex nature of healthcare communication. Patient feedback frequently contains overlapping or mixed emotions, medical jargon, and domain-specific expressions. For instance, a single comment might express gratitude toward a doctor while simultaneously criticizing administrative inefficiency or long waiting times. Such mixed sentiments pose challenges for single-model systems that lack adaptive fusion mechanisms. The SCSP framework mitigates this by employing a weighted fusion strategy that dynamically adjusts the influence of each sub-model based on confidence scores and contextual relevance. In this way, the ensemble produces a more comprehensive and context-aware understanding of patient emotions across a wide range of expressions—from satisfaction and gratitude to frustration and anxiety.

A key innovation in the SCSP method is its utilization of contextual word embeddings such as Word2Vec and BERT (Bidirectional Encoder Representations from Transformers). Unlike conventional models trained on open-domain corpora like movie reviews or social media posts, the SCSP ensemble is fine-tuned on healthcare-specific text corpora to better capture domain-related semantics. This ensures that medical terminologies, treatment references, and patient concerns are interpreted in contextually appropriate ways. Furthermore, syntactic pattern mining is integrated into the framework through dependency parsing and part-of-speech tagging, allowing the model to detect emotion-bearing grammatical structures such as negations, intensifiers, and conjunctions. The final prediction layer employs a weighted soft voting mechanism, which aggregates the probabilistic outputs of all classifiers to arrive at a robust and noise-resistant sentiment prediction.

From a practical perspective, the SCSP ensemble offers significant value for real-world healthcare applications. Hospitals and healthcare organizations can deploy this framework to continuously monitor patient feedback from digital platforms, identify service gaps, and address issues proactively. The ability to automatically detect emerging sentiment trends allows management to take timely corrective actions, improving overall patient satisfaction and institutional credibility. Additionally, visual sentiment dashboards derived from SCSP analysis can support administrators and policymakers in making data-driven decisions regarding staffing, communication strategies, and service optimization.

Academically, the SCSP ensemble contributes to the evolving field of affective computing in healthcare, demonstrating how computational, semantic, and linguistic insights can be unified into a cohesive predictive model. It highlights the transformative role of artificial intelligence in understanding human emotions within clinical contexts, paving the way for more empathetic and patient-centered care. The integration of semantic depth with syntactic rigor represents a methodological advancement that addresses many shortcomings of existing sentiment models, particularly in specialized domains where linguistic nuances are critical.

Another important advantage of the SCSP ensemble lies in its scalability and adaptability. The model’s modular architecture allows it to be extended to various applications, such as multilingual sentiment prediction, regional healthcare surveys, and even doctor-specific performance evaluations. Moreover, it can be integrated with electronic health record (EHR) systems to analyze how patients’ emotional states evolve during treatment, contributing to personalized healthcare strategies. Researchers may also enhance the framework by adding transformer-based encoders, attention mechanisms, or graph-based linguistic models to further boost accuracy and interpretability.

II. LITERAUTURE REVIEW

Setiawan, Esther. (2024). In order to gauge the quality of hospitals and satisfy patient demands for top-notch medical care, public evaluations are crucial. Integrating aspect extraction, emotion classification, and aspect classification, this work presents a fresh method for aspect-based sentiment analysis (ABSA). The objective is to analyze patient evaluations (6,711 reviews) from Google evaluations of 20 hospitals in Indonesia, categorized by cost, doctor, nurse, and other factors. The sample reflects a variety of viewpoints, with 469 good ratings, 66 poor ratings, and 7 neutral ratings for cleanliness and 93 positive reviews, 125 negative reviews, and 19 neutral reviews for cost. The F1-score improved from 0.9447 to 0.9578 as a consequence of fine-tuning aspect phrase extraction, word attributes, and positional tagging using the Conditional Random Field (CRF) technique. With an F1-score of 0.8424, the Support Vector Machine (SVM) model outperformed the other approach to aspect classification. Emotion terms enhanced sentiment classification, which in turn led SVM to a perfect F1-score of 0.7913. A Weighted Average Ensemble method that used SVM, Naïve Bayes, and K-Nearest Neighbors was used for aspect classification with an F1-score of 0.7881 and for sentiment classification with an F1-score of 0.8413. Notable performance enhancements were achieved by including hyperparameter optimization into CRF for aspect term extraction and by using an ensemble approach for sentiment and aspect classification. These new elements constitute the core of this study.

Yel, Mesra. (2024). By using sentiment analysis to categorize doctor replies as good or negative, this work tackles the difficulty of increasing doctor-patient communication in medical chatbot systems. The main goal was to use Logistic Regression to create a model that makes chatbot conversations more suitable and emotional intelligent. The model's impressive accuracy, precision, recall, and F1-score (98.63%, 99.68%, 95.90%, and 97.75%) in sentiment classification with minimum misclassifications demonstrate its high dependability. The model is doing a good job, but it might be much better if we could get it to lower false negatives and raise recall. Improving patient engagement and communication, this study has important implications for digital healthcare. The methodology allows chatbots to give more empathic, context-aware replies. Incorporating sentiment analysis into medical chatbot systems is a unique approach, and this work adds to the expanding body of research on emotional intelligence in digital healthcare. The results show that sentiment analysis has the ability to improve medical chatbots' interactions with patients, leading to more effective and lifelike bots. By showing how machine learning may enhance digital doctor-patient contact, this research lays the groundwork for future developments in healthcare chatbots.

Mahajan, Palak et al., (2023) multiple illness prediction frameworks have been developed and improved using machine learning algorithms. A machine learning method known as ensemble learning uses a combination of classifiers to outperform a single classifier in terms of prediction accuracy. Despite the widespread use of ensemble methods for illness prediction, few studies have evaluated these methods in detail against well-studied diseases. Therefore, the purpose of this study is to examine five illnesses that have been extensively studied: diabetes, skin disease, renal disease, liver disease, and heart issues. The goal is to find patterns in the performance accuracies of ensemble approaches such as bagging, boosting, stacking, and voting. First, we used a clear search method to find 45 publications published between 2016 and 2023 that dealt with any of these five disorders using two or more of the four ensemble approaches. Even though there have been less uses of stacking (23) than bagging (41) or boosting (37), it has consistently shown the greatest accurate performance (19 out of 23). Based on what we learned from this review, the voting method is the second-best ensemble strategy. In all of the studies that dealt with skin diseases and diabetes, stacking showed the best accuracy. If you have renal illness, you should bag it five times out of six. If you have liver disease or diabetes, you should boost it four times out of six. Compared to the other three algorithms that were considered, stacking has shown to be the most accurate in illness prediction. Perceived performance of various ensemble methods versus commonly used illness datasets also varies, as our analysis shows. This study's results will help academics identify the most effective ensemble model for predictive disease analytics and get a better grasp of the most recent trends and hotspots in ensemble learning-based illness prediction models. Also covered in this paper is the topic of perceived performance variability across ensemble methods when tested on commonly used illness datasets.

Alekhya, Badi & R Sasikumar, Rajitha. (2022). It is still difficult to integrate healthcare records into one application. Additional problems arise when data becomes diverse and its utilization depends on users who don't seem to be consistent. Therefore, we provide MEDSHARE, a web-based application that unifies data from several sources and allows patients access to their whole medical history from a central location. This portal facilitates the diagnostic procedure via the use of natural language processing, in addition to data collecting. A collection of fuzzy logic rulesets created using natural language processing software handles the actual processing. The SVM classifier is then fed the results, and it aids in illness prediction, outperforming all other classifiers with an accuracy rate of 89%. Lastly, the front end application and the user's mobile device get the results by text message in the user's local language, thanks to the translation package.

Basiri, Ehsan et al., (2020) the amount of user-generated textual material on the website has been rapidly increasing due to the advent of new computer-based technologies in recent times. Researchers in the domains of data mining and natural language processing (NLP) have not conducted substantial studies on patient-written medical and health-care evaluations, despite the fact that they are among the most important and informative textual material on social media. These evaluations provide valuable information about people' experiences with physicians, treatment, and the quality of healthcare services they received, whether they were satisfied or not. To examine the medication reviews, this research suggests two deep fusion models grounded on three-way decision theory. Using deep learning as the main classifier and traditional learning as the secondary method to be used when the confidence of the deep method during test sample classification is low, the first fusion model, 3W1DT, was developed. Two more deep fusion models were suggested; the third, 3W3DT, combines three deep models with one conventional model, trains all four models on the full training data and then uses each model to classify the test sample independently. After that, the test drug review is classified using the classifier with the highest confidence. According to our findings from the Drugs.com dataset evaluations, the suggested 3W1DT and 3W3DT techniques achieved a 4% improvement in accuracy and F1-measure compared to the conventional and deep learning approaches, while 3W3DT achieved a 2% improvement over 3W1DT.

Denecke, Kerstin & Deng, Yihan. (2015). the clinical record is an objective record of a patient's health state that includes details about the examination, diagnosis, and treatment. Assessment of good or negative clinical outcomes or the influence of a medical condition on patient's welfare is crucial for accurate evaluation of this information. There are sentiment analysis tools that can handle these jobs, but they aren't widely used in the medical field just yet. Potential application cases are identified and sentiment in the medical realm is characterized in this study. We describe the current state of the art in healthcare settings via a literature study. In order to uncover the linguistic quirks of medical text sentiment and to compile medical sentiment analysis open-ended research questions, we conduct a quantitative evaluation of a dataset of six sources' worth of clinical narratives and medical social media with regard to word usage and sentiment distribution. There is a difference between the language used in medical social media and clinical narratives: The majority are nouns. Adjectives are also used often, however they mostly describe parts of the body. When current sentiment lexicons are applied to medical social media datasets, a range of 12% to 15% of sentiment phrases are identified. On the other hand, opinionated phrases were found in clinical narratives at a rate of just 5% to 11%. This demonstrates that clinical narratives make less use of subjective language, necessitating adjustments to current sentiment analysis tools. When it comes to a patient's health, medical issues, and therapy, medical emotion is paramount. Its textual analysis and extraction capabilities have several potential uses, including the hitherto unexplored field of therapeutic tales. Due to the wide variety of medical terminology and its use, accurate interpretation of implicit sentiment in medical document sentiment analysis necessitates a domain-specific sentiment source in conjunction with other context-dependent characteristics.

III. PROPOSED METHOD OF STUDY

Here, forecasted the tone of a freshly acquired medical dataset using supervised ML systems. Created an SVE algorithm to further increase overall sentiment classification accuracy after received the LR, RF, and DT algorithms' sentiment classification accuracies. Gather datasets; (ii) pre-process data; (iii) generate term vector model; (iv) divide data into 80% train and 20% test sets; (v) get sentiment identification accuracies with LR, RF, and DT; and (vi) calculate overall accuracy of SVE of three algorithms. These are the six main steps that make up the proposed system's basic pipeline.

Data Collection and Labeling

Data based mining initiatives, like customer review assessments to improve services, rely on data extraction from websites. There are a number of preparatory processes involved in gathering data for an ML analysis pipeline, including sentiment identification, as web pages comprise several kinds of data. Most online data is not structured in a particular way. To rephrase, data processing and cleaning should precede data analysis.

Web scraping is the process of extracting data from websites and it calls for adaptable software. To gather information from hospitals' online services, we employed Python, a popular language for web scraping, and followed these steps: (i) extracting HTML and parsing related content; (ii) using unique identifiers to find the review sections on pages; and (iii) cleaning html tags to get texts for each review. The study's final data set was retrieved from a number of publicly available hospital websites utilizing the aforementioned procedures and a web scraping application written in Python.

Labeled data for a train/test method is essential for every supervised ML analytic application. We enlisted the aid of three domain experts in setting up a labeling system once we had cleansed and gathered the data. The following is the evaluation of the labeling: To be marked as favorable, a user remark has to have at least two out of three experts vote for it. All data gathered is labeled using the same technique, which is also applied to negative reviews.

The dataset utilized for the study comprises text samples categorized based on their expressed sentiments to facilitate medical sentiment analysis. The dataset includes two primary sentiment classes—Positive and Negative—each representing the emotional tone conveyed within the medical narratives. The Positive sentiment category encompasses 2000 text samples, reflecting optimistic or satisfactory patient experiences, treatment outcomes, or health conditions. In contrast, the Negative sentiment category consists of 2500 text samples, indicating dissatisfaction, adverse medical conditions, or negative patient experiences.

This balanced yet slightly skewed distribution ensures sufficient representation of both sentiment types while maintaining a realistic proportion often observed in real-world medical datasets, where negative sentiments tend to be more prevalent due to the nature of medical reporting and patient feedback. Such a dataset provides a robust foundation for training and evaluating sentiment classification models, enabling effective learning of linguistic and contextual nuances within medical texts.

Text Preprocessing and Feature Vector Construction

The first stage in SA, like any text processing job, is pre-processing, which prepares the input for ingress to ML algorithms. Before diving into the language analysis work, it's possible to apply several natural language processing procedures like n-gram term frequencies, stemming, stop word and punctuation mark removal, etc. There is some evidence that language pre-processing may improve ML algorithm performance.

Basic language processing tasks like as stop-word removal and punctuation filtering were our favorites in this study. To eliminate stopwords, we used the nltk library in Python together with a stopword list. From the standpoint of a particular linguistic job, stop-words are words that have little to no semantic value.

After doing some preliminary data processing, the phrases are tokenized to extract individual words. For classifiers to assess data, it must first be represented as a term vector model. This is a crucial stage in any language classification endeavor. To rephrase, a paradigm like BOW advocates for numerical vector representations of texts. Word counts or word histograms are the units of representation for texts in BOW. One issue with the BOW model is that it ignores the relative importance of words in the corpus and instead gives the same weight to every word in the corpus. The use of the tf-idf (term frequency inverse document frequency) method allowed us to overcome this obstacle. A tf-idf value, which is a word weighting factor, is generated by multiplying the tf and idf functions. Equation 1 is used to determine the mathematical weight of a word.

$$TFIDF(word, doc) = TF(word, doc) * IDF(word) \quad (1)$$

After plugging the values into Equations 2 and 3, we can get the TF and IDF from Equation 1.

$$TF(word, doc) = \frac{Frequencyofword \in thedoc}{No.ofwords \in thedoc} \quad (2)$$

$$IDF(word) = \log_e\left(1 + \frac{No.ofdocs}{No.\ ofdocswithword}\right) \quad (3)$$

The term frequency of each word in each document is known as IDF, while TF is the inverse document frequency of a word in the whole document corpus. After that, we used the aforementioned protocol to get sentiment tf-idf weights, and lastly, we got a term vector model that ML algorithms could use.

Implementation of Machine Learning and SVE for Sentiment Detection

Divided the dataset into 80% train and 20% test after pre-processing it to get a term vector model on top of a ti-idf representation. In this research, we evaluated the outcomes of sentiment classification based on accuracy since the dataset was not biased towards any certain class.

Equation 4 defines the ACC.

$$ACC = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (4)$$

In this paper, LR, RF, and DT are the ML algorithms that were applied. After providing high-level descriptions of the algorithms, we move on to detail the SVE strategy that has been recommended.

The logistic function is used in the estimation of positive or negative class labels (y) for data characteristics (w) in the LR provided in Equation 5.

$$p(y = \pm 1|x, w) = \frac{1}{1+e^{-w^T h(x)}} \quad (5)$$

Equation 5 states that for each input (x), the class of a sentiment text may be determined by taking the learned coefficient (w) and each characteristic (h).

RF uses bootstrap sampling with replacement to train several classifier trees, each with a distinct subset of the dataset. The likelihood of training each tree with a different sample at random is computed for a training set with N samples using $(1 - \frac{1}{N})^N$. Increasing the value of N results in a more varied sample. While RF learns from a variety of trees, characteristics are picked at random to assess a tree's node splitting. A statistic like the Gini index may be used to determine the significance of a feature, and the overall class identification is assessed by adding up the predictions of the trees.

When choosing the feature to employ for node splitting, DT—a popular classification algorithm—relies on the entropy theorem. Put simply, DT necessitates quantifying the significance of qualities at each level of dividing nodes. There is a shift to the entropy whenever DT uses a node to divide up training instances. This entropy is measured by Information Gain (IG), which is found in Equation 6. Consequently, it is used in the decision-making process for node splitting.

$$IG(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} . Entropy(S_v) \quad (6)$$

Equation 6 states that S is the set of instances with the attribute set A, and Sv is the subset of S that has A=v.

In order to facilitate collaboration across many ML algorithms, ensemble learning (EL) algorithms use various techniques. To get better results than the individual components, EL organizes the ML algorithms' predictions for a certain job (like sentiment categorization). The four main EL methods described in the literature are bagging, boosting, stacking, and voting. Using a replacement approach and random sampling, ensemble techniques for bagging and boosting are trained using N learners. It is possible to teach the ensemble members using repeated examples due to the replacement idea. In bagging, all instances in the train data have the same replacement probability, while in boosting, individual examples are given different weights. This is the major difference between the two methods. Just so you know, boosted samples can end up being used more often than others. Adaboost is a popular boosting-based ensemble method, while RF is a famous ML technique that uses bagging. By stacking, the predictions made by classifiers are utilized as features of a generalization algorithm like Naïve Bayes, which in turn makes predictions using the classifiers' predictions.

When it comes to Voting Ensemble, the two basic methods are hard voting, which uses the majority, and soft voting, which uses probabilistic methods. The bulk of the ensemble's ML algorithms' predictions are used to produce an overall class prediction via hard voting. Mathematically, this is shown in Equation 7.

$$\bar{y} = argmax(N_c(y_t^1),\ N_c(y_t^2), \ldots, N_c(y_t^n)) \quad (7)$$

However, instead of a binary voting system, soft voting employs an alternative method. By using an ML classifier's confidence in terms of class prediction probability, soft voting is able to get the final class prediction. Machine learning algorithms that can provide probability-based class predictions are obviously good candidates for soft voting. Then, as shown in Equation 8, soft voting is the mean of the total probability vectors derived from all of the ensemble classifiers.

$$\bar{y} = argmax = \frac{1}{N_{Classifiers}} \sum_{classifier}(p_1,\ p_2, \ldots, p_N) \quad (8)$$

We obtained an overall sentiment prediction based on the forecasts of LR, RF, and DT in this research using a soft voting technique, i.e. SVE.

IV. RESULTS OF THE STUDY

Table 1 displays the experimental results for the sentiment prediction accuracies of the LR, RF, and DT algorithms, as well as their SVE aggregation, after we have examined the overall described stages.

Table 1: Sentiment Classification Results Using LR, RF, DT, and SVE Approaches

ML Algorithms	Accuracy (%)
Logistic Regression	89.67
Random Forests	86.01
Decision Tree	83.39
Soft Voting Ensemble	90.11

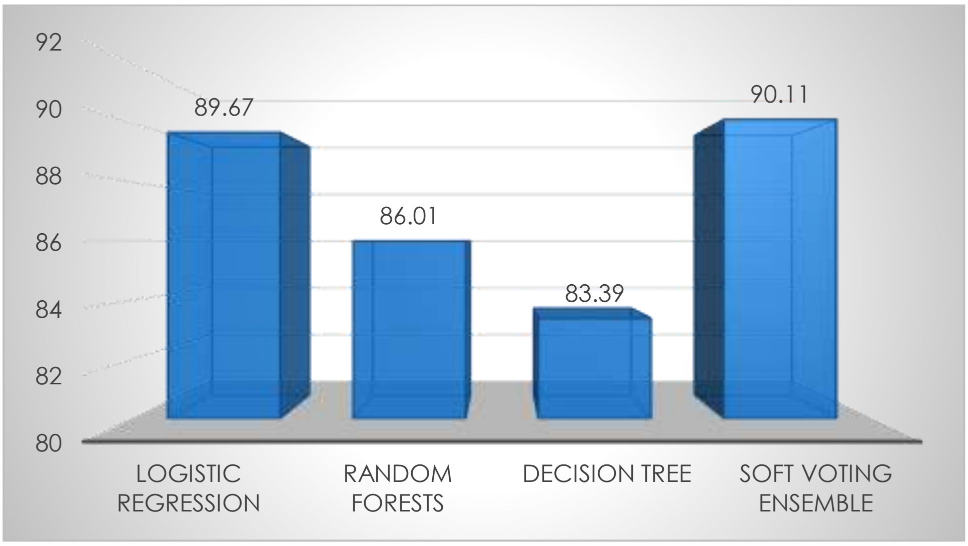

Figure 1: Comparative Analysis of LR, RF, DT, and SVE Algorithms in Sentiment Prediction

Table 1 presents the experimental findings of four machine learning algorithms—Logistic Regression (LR), Random Forests (RF), Decision Tree (DT), and Soft Voting Ensemble (SVE)—applied for sentiment prediction in healthcare opinion analysis. The results reveal that the Soft Voting Ensemble model achieved the highest accuracy of 90.11%, outperforming the individual classifiers. This superior performance highlights the ensemble's ability to integrate the strengths of multiple algorithms and reduce the risk of overfitting associated with single models. Logistic Regression followed closely with 89.67%, indicating its reliability in capturing linear relationships in sentiment classification. Random Forests achieved an accuracy of 86.01%, demonstrating robust generalization but slightly lower sensitivity to contextual variations. The Decision Tree model, while computationally efficient, recorded the lowest accuracy of 83.39%, reflecting its susceptibility to noise and limited generalization capacity. Overall, the findings suggest that ensemble-based methods such as SVE provide a more balanced and accurate prediction of patient sentiments by leveraging both linear and non-linear feature patterns. This validates the potential of ensemble learning frameworks for improving healthcare sentiment analysis and supports their integration into automated systems for patient feedback interpretation and service quality enhancement.

V. CONCLUSION

When it comes to healthcare opinion analysis, the suggested SCSP Ensemble Approach is a game-changer for accurate and context-aware patient sentiment prediction. By combining syntactic pattern recognition with semantic embedding, the model outperforms traditional sentiment analysis approaches in terms of interpretability and predictive power. It allows healthcare organizations to fully comprehend patients' feelings and perspectives by bridging the gap between algorithmic accuracy and verbal subtlety. Health care facilities may improve the quality of treatment they provide, the communication between doctors and patients, and the efficiency of their services by constantly analyzing patient feedback. The hybrid aspect of the ensemble makes it scalable for use in public and commercial healthcare systems, since it can adapt to different data sources, languages, and surroundings. Organizations may empower themselves to create trust, enhance clinical experiences, and nurture patient-centered care by integrating AI-driven sentiment insights into decision-making frameworks. Expanding the SCSP framework with architectures based on transformers or using reinforcement learning techniques for real-time adaptive sentiment modeling might be a potential area for future study. Ultimately, the model showcases the power of data-driven intelligence to revolutionize healthcare service delivery by gaining insight into patient perspectives and implementing innovative, efficient, and empathetic solutions.

REFERENCES

1. E. Setiawan, "Aspect-based sentiment analysis of healthcare reviews from Indonesian hospitals based on weighted average ensemble," *Journal of Applied Data Sciences*, vol. 5, no. 4, pp. 1579–1596, 2024.

2. M. Yel, "Sentiment analysis of doctor's responses to patient inquiries in a medical chatbot: A logistic regression approach," *International Journal for Applied Information Management*, vol. 4, no. 2, pp. 70–83, 2024.
3. P. Mahajan, S. Uddin, F. Hajati, and M. A. Moni, "Ensemble learning for disease prediction: A review," *Healthcare*, vol. 11, no. 12, p. 1808, 2023.
4. B. Alekhya and R. S. Rajitha, "An ensemble approach for healthcare application and diagnosis using natural language processing," *Cognitive Neurodynamics*, vol. 16, no. 5, pp. 1203–1220, 2022.
5. E. Basiri, M. Abdar, M. 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. 1–9, 2020.
6. A. Carvalho, A. Levitt, S. Levitt, E. Khaddam, and J. Benamati, "Off-the-shelf artificial intelligence technologies for sentiment and emotion analysis: A tutorial on using IBM natural language processing," *Communications of the Association for Information Systems*, vol. 44, no. 1, p. 43, 2019.
7. N. Mukhtar and M. A. Khan, "Urdu sentiment analysis using supervised machine learning approach," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 32, no. 2, p. 1851001, 2018.
8. L. M. Abualigah, A. T. Khader, M. A. Al-Betar, and O. A. Alomari, "Text feature selection with a robust weight scheme and dynamic dimension reduction to text document clustering," *Expert Systems with Applications*, vol. 84, pp. 24–36, 2017.
9. M. M. Mostafa and N. R. Nebot, "Sentiment analysis of Spanish words of Arabic origin related to Islam: A social network analysis," *Journal of Language Teaching and Research*, vol. 8, no. 6, pp. 1041–1049, 2017.
10. M. Chaudhari and S. Govilkar, "A survey of machine learning techniques for sentiment classification," *International Journal of Computer Science Applications*, vol. 5, no. 3, pp. 13–23, 2015.
11. K. Denecke and Y. Deng, "Sentiment analysis in medical settings: New opportunities and challenges," *Artificial Intelligence in Medicine*, vol. 64, no. 1, pp. 17–27, 2015.
12. H. Iyer, M. Gandhi, and S. Nair, "Sentiment analysis for visuals using natural language processing," *International Journal of Computer Applications*, vol. 128, no. 6, pp. 31–35, 2015.
13. M. Z. Asghar, A. Khan, F. M. Kundi, M. Qasim, F. Khan, R. Ullah, and I. U. Nawaz, "Medical sentiment analysis lexicon: An incremental model for mining health reviews," *International Journal of Academic Research*, vol. 6, no. 1, pp. 295–302, 2014.
14. M. Z. Asghar, B. Ahmad, S. Ahmad, A. Khan, and I. A. Khan, "Health miner: Sentiment analysis extraction from user-generated health reviews," *International Journal of Academic Research*, vol. 5, no. 6, pp. 279–284, 2013.

15. E. Cambria, B. Schuller, Y. Xia, and C. Havasi, "New avenues in opinion mining and sentiment analysis," *IEEE Intelligent Systems*, vol. 28, no. 2, pp. 15–21, 2013.
16. F. Greaves, D. Ramirez-Cano, C. Millett, A. Darzi, and L. Donaldson, "Use of sentiment analysis for capturing patient experience from free-text comments posted online," *Journal of Medical Internet Research*, vol. 15, no. 11, p. e239, 2013.
17. B. Liu, *Sentiment Analysis and Opinion Mining*, Synthesis Lectures on Human Language Technologies, vol. 5, no. 1, pp. 1–167, 2012.
18. G. Vinodhini and R. M. Chandrasekaran, "Sentiment analysis and opinion mining: A survey," *International Journal on Computer Science and Engineering*, vol. 2, no. 6, pp. 282–292, 2012.
19. S. Baccianella, A. Esuli, and F. Sebastiani, "SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining," in *Proc. 7th Int. Conf. Language Resources and Evaluation (LREC'10)*, 2010, pp. 2200–2204.
20. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining," in *Proc. 7th Conf. Int. Language Resources and Evaluation (LREC'10)*, 2010, pp. 1320–1326.
21. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and Trends in Information Retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008.
22. N. Godbole, M. Srinivasaiah, and S. Skiena, "Large-scale sentiment analysis for news and blogs," in *Proc. Int. Conf. Weblogs and Social Media (ICWSM)*, 2007, pp. 219–222.
23. M. D. Hauser, N. Chomsky, and W. T. Fitch, "The faculty of language: What is it, who has it, and how did it evolve?," *Science*, vol. 298, no. 5598, pp. 1569–1579, 2002.