

The social media sentiment analysis framework: deep learning for sentiment analysis on social media

Prasanna Kumar Rangarjan¹, Bharathi Mohan Gurusamy¹, Gayathri Muthurasu², Rithani Mohan¹, Gundala Pallavi¹, Sulochana Vijayakumar³, Ali Altalbe^{4,5}

¹Department of Computer Science and Engineering, Amrita School of Computing, Amrita Vishwa Vidyapeetham, Chennai, India

²Department of Electronics and Communication Engineering, College of Engineering and Technology, SRM Institute of Science and Technology, Vadapalani Campus, Tamil Nadu, India

³Anna Administrative Staff College, Government of Tamil Nadu, Chennai, India

⁴Department of Computer Engineering, Prince Sattam bin Abdulaziz University, Alkharj, Saudi Arabia

⁵Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

Article Info

Article history:

Received Feb 6, 2024

Revised Mar 4, 2024

Accepted Mar 5, 2024

Keywords:

Convolutional neural networks

Deep neural networks

Emotion

Recurrent neural networks

Sentiment analysis

Social media

ABSTRACT

Researching public opinion can help us learn important facts. People may quickly and easily express their thoughts and feelings on any subject using social media, which creates a deluge of unorganized data. Sentiment analysis on social media platforms like Twitter and Facebook has developed into a potent tool for gathering insights into users' perspectives. However, difficulties in interpreting natural language limit the effectiveness and precision of sentiment analysis. This research focuses on developing a social media sentiment analysis (SMSA) framework, incorporating a custom-built emotion thesaurus to enhance the precision of sentiment analysis. It delves into the efficacy of various deep learning algorithms, under different parameter calibrations, for sentiment extraction from social media. The study distinguishes itself by its unique approach towards sentiment dictionary creation and its application to deep learning models. It contributes new insights into sentiment analysis, particularly in social media contexts, showcasing notable advancements over previous methodologies. The results demonstrate improved accuracy and deeper understanding of social media sentiment, opening avenues for future research and applications in diverse fields.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Bharathi Mohan Gurusamy

Department of Computer Science and Engineering, Amrita School of Computing, Amrita Vishwa

Vidyapeetham

Chennai, India

Email: g_bharathimohan@ch.amrita.edu

1. INTRODUCTION

The term “opinion mining,” also known as “sentiment analysis,” refers to the computational analysis of people's opinions and assessments of a wide range of things, including but not limited to goods and services, companies, individuals, topics, and characteristics [1]. The birth and rapid growth of the field coincide with the birth and rapid growth of social media on the internet, such as reviews, forum debates, blogs, microblogs, twitter, and social networks, providing us with unprecedented access to a vast trove of opinionated material never before seen in human history [2]. They might share their thoughts on politics, current events, or a recently purchased product [3]. There are several uses for user data, including optimization techniques, research analysis within organizations, and election campaign methods [4]. In addition, governments value public opinion analysis because it elucidates why specific actions and behaviors

are popular and how those actions and behaviors are influenced by the views of others [5]. The proliferation of social media complements the internet as the most accessible and economical information hub [6]. Opinions are mined from social media sources such as blogs, reviews, tweets, posts, and discussions [7]. A person's sentiments: their attitude, ideas, feelings, and opinions; are crucial to understanding their actions [8].

The practice of sentiment analysis (sometimes called opinion mining) examines such feelings as they pertain to any given entity [9]. Objects consist of parts and a collection of characteristics [10]. An illustration [11] shows that the laptop has a cracked screen and poor battery life. It is a portable computer complete with a display and power. A screen's display quality is comparable to the battery's life expectancy in terms of its characteristics [12]. More nuanced classifications (highest, lowest, highest, lowest degree of positivity) might be applied to the thoughts or opinions represented in the text [13]. The case above portrays the laptop in a bad light. Moreover, various media, such as words, graphics, sounds, and moving pictures, can communicate feelings [14].

Using user-generated content tools and social networking capabilities, users of Web 2.0 can actively participate in the production of Web content and generate massive amounts of text content on a wide variety of themes [15]. Social media is an integral aspect of social networking and information sharing, which is why social networking is an integral part of the Web 2.0 platform and application service [16]. Since the explosion of social media messaging and data, issues such as online views and evaluations have received attention from the government, businesses, and academics [17]. Users, connections, and content are the three main areas that social media mining may shed light on [18].

Advances in data analytics and the skyrocketing volume of user-generated material on for-profit websites have created new difficulties and possibilities [19]. For many sentiment analysis applications, the ability to detect unusual and potentially beneficial patterns in a vast body of user-generated content is essential [20]. Businesses need to understand consumer sentiments regarding their brand, product, or service, and sentiment analysis techniques aim to identify and extract such information from the source data [21]. However, there are several obstacles to identifying emotions in user generated content (UGC) since it is often fragmentary, noisy, poorly structured, and full of idiosyncratic idioms, illogical constructions, and non-lexical terminology [22]. Moreover, the wide variety of linguistic concerns complicates the sentiment analysis process, making it difficult to investigate the association between opinion sentences [23]. There has been a rise in businesses that urge customers to participate in online focus groups to gain valuable insights and suggestions for increasing efficiency [24].

Researchers and practitioners generally address sentiment analysis across various granularities, from the broadest to the most specific levels. Determining the tone of an entire phrase falls under coarse-grained analysis, while attribute-level and phrase-level sentiment analysis falls under fine-grained analysis. Researchers have previously turned to tried-and-true natural language processing (NLP) and machine learning techniques when faced with the problem of sentiment categorization, harnessing sentiment analysis as a potent technique to accomplish research objectives if the appropriate approach is applied to it [25]. The main significant contributions to the paper are as follows: i) The research's primary aim was to develop a methodological framework for social media sentiment analysis (SMSA), introducing a custom-built sentiment dictionary to enhance emotion and sentiment classification; ii) It presents a comprehensive sentiment classification system and explores the implications of employing popular deep learning architectures in this context; and iii) The study rigorously examines the performance of various deep learning models across a spectrum of parameter calibration combinations, assessing their effectiveness in the realm of sentiment analysis.

The upcoming section will benchmark this research against similar studies. The second section conducts a literature survey in the field. The third part delves into the methodologies of the SMSA framework. The fourth section details the experimental setup, and the fifth part presents the results and discussion

2. LITERATURE REVIEW

Navigating numerous websites presents a challenge in identifying the most relevant ones tailored to individual interests. Lengthy blog posts and forum threads often contain extensive, subjective commentary, making it difficult to decipher. Consequently, average readers need help pinpointing valuable resources and extracting and consolidating opinions. Researchers have developed supervised and unsupervised strategies for various sentiment analysis tasks to address this issue. Additionally, previous research in a supervised setting has utilized many other supervised machine learning methods, including support vector machines (SVM), maximum entropy, and feature combinations.

Zhang *et al.* [26] developed deep learning (DL). The machine learning method known as deep learning has recently become a reliable tool for achieving cutting-edge prediction accuracy by acquiring data representations and features at several levels. Sentiment analysis is another area that has benefited from deep

learning in recent years, following in the footsteps of its widespread success in other application domains. This paper introduces readers to the fundamentals of deep learning and then surveys the field's existing applications, focusing on sentiment analysis.

Dang *et al.* [27] incorporated term frequency-inverse document frequency (TF-IDF) applications of word embedding and deep learning to various datasets, enhancing sentiment analysis. This paper provides a comprehensive overview of the most recent research that has used deep learning to address issues in sentiment analysis, such as the determination of sentiment polarity. However, limitations beyond those encountered in natural language processing hinder the efficacy and precision of sentiment analysis. In recent years, deep learning models have emerged as a practical solution to address the challenges of NLP.

Swathi *et al.* [28] delivered a taxonomy of sentiment analysis (TSA). It provides a comprehensive study of the most popular deep learning models used for sentiment analysis now. The article discusses the ramifications of well-known deep-learning architectures. Here, the focus is primarily on deep learning methodologies to highlight the significant contributions of various academics. An overview of the most critical applications of sentiment analysis is provided, and the range of languages in which this technique is used is established. Many studies have employed machine learning and NLP-based approaches to analyze these expressions. Many studies have used machine learning and NLP-based approaches to analyze these expressions. However, due to their superior performance as of late, methods based on deep learning are enjoying a meteoric rise in popularity.

Chen *et al.* [29] exhibited to build a system for analyzing sentiment (SAS). It uses social media data analysis to propose an in-house military sentiment lexicon to improve sentiment classification and assess the efficacy of multiple deep-learning models with different parameter calibration combinations. As a result of the proliferation of social media data on the web, government, industry, and academia have all taken an interest in opinion mining and sentiment analysis of text. Since its inception as a subject of knowledge fusion in the significant data era, sentiment analysis has been a hot topic in artificial intelligence and machine learning.

Wadawadagi *et al.* [30] aim to provide a comprehensive empirical analysis of deep neural networks (DNN) deployed in sentiment categorization tasks. In the first stage, they analyze many state-of-the-art DNN models and the theories behind them. They provide estimates of the efficacies of various DNN models described in the literature through sentiment dataset experiments. Ultimately, deep learning methods are crucial to the success of any model in the modern computational environment [31]. However, deep learning techniques offer a high degree of automation by deriving generalized rules for text and sentiment categorization tasks [32], [33]. The sentiment analysis need large articles to be condensed for effective analysis using topic modeling [34], clustering and text summarization [35] to reduce the content and complexity [36].

This review aims to surpass previous surveys by delving into the significant and widespread methods for sentiment analysis that have recently emerged using deep learning. However, older techniques are still integrated to understand sentiment analysis comprehensively. This review addresses the gap left by prior surveys by discussing and comparing deep learning-based approaches. It provides a taxonomy of current methodologies that analyzes how deep learning techniques can enhance the efficacy of sentiment analysis.

3. METHOD

3.1. Social media sentiment analysis framework

Computer models of the brain are used in deep learning, which consists of applying these networks to learning tasks with numerous layers of networks. Using the vastly increased learning (representation) capacity of neural networks, what was once thought feasible with one or two layers and a small amount of data is now a real possibility. Both feedforward and recurrent/recursive neural networks can be merged to form novel neural architectures.

In brief, deep learning extracts and transforms features using a series of layers of nonlinear processing units. The lowest layers, which are closest to the data input, learn the most straightforward features, while the higher layers, which are farthest from the input, learn the more complex features inferred from the simpler ones. Architecture's hierarchical feature representation is robust. DNNs are a subset of artificial neural networks (ANNs) in which the input and output layers are separated by many hidden layers, as shown in Figure 1. Hence, they differ from shallow single-layered networks by the number of levels at which information travels. DNNs can learn analogously to a real brain by inserting numerous hidden layers into the network and adjusting the weights of the connections between the neurons. Layers closer to the input layer, known as lower-order layers, learn less important features during operation, whereas layers further away from the input layer, features learned at lower-order levels, are used as the basis for further feature learning at higher-order layers.

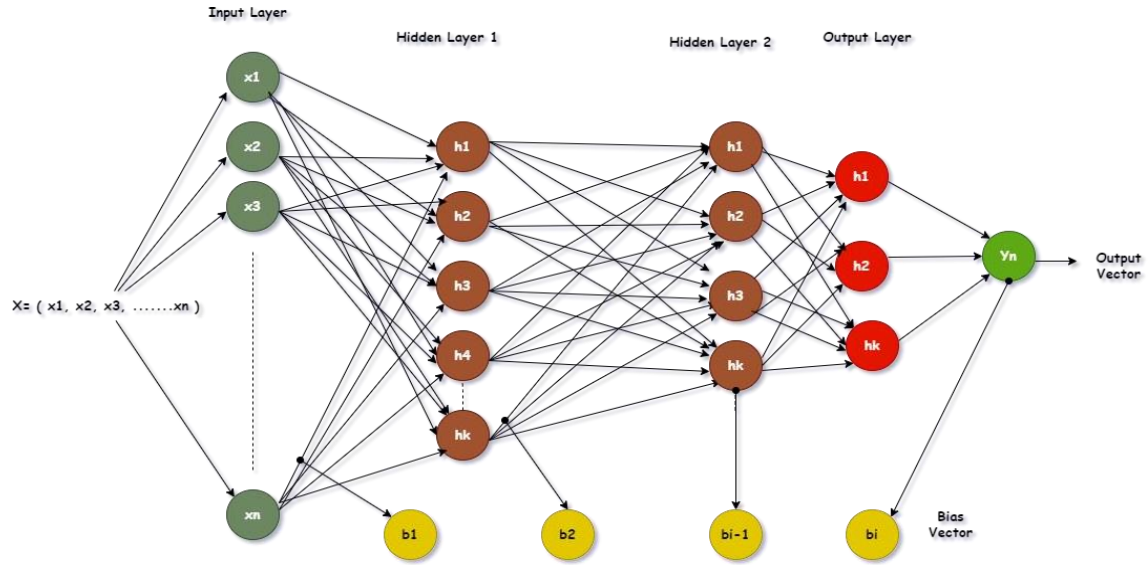


Figure 1. Framework of deep neural network architecture

The model has n layers, with l_1 being the input layer for the sequence $x = x_1$ through fx , connection weights w_1 through w_i , and an intercept bias vector b_1 through b_i . Outputs y_1, y_2, \dots, y_m are connected to the l^{th} layer, while the layers between l_2 and l^{n1} are the hidden layers whose results are not readily apparent. Values in the input layer represent data sent into the network, while neurons, the smallest building blocks of computation, are represented by components in the hidden and output layers. These neurons, which transfer inputs to outputs in a way that is not linear, are also referred to as activation functions. Directional edges in a network of connections determine how information moves between neurons. Moreover, the signal between every given pair of neurons is modulated by a weight associated with each connection in the network.

The computational process within the hidden layers is described by (1), where A represents the activation function at the i^{th} hidden layer, applied to the weighted sum of inputs from the previous layer, and the bias term b_i ,

$$h_i(x) = A(w_i h_i - 1 + b_i) \tag{1}$$

Equation (2) is used for the output layer, where OA denotes the output activation function. This function is applied to the output from the last hidden layer to compute the network's final output.

$$y_n(x) = OA(h_i(x)) \tag{2}$$

Training neural networks typically involves applying optimization techniques, necessitating a loss function to measure the model's inaccuracy. The negative log-likelihood or residual sum of squares is a standard metric for measuring loss, while the precise measure depends on the learning task's nature.

3.2. Sentimental analysis framework

Figure 2 presents a comprehensive visual representation of the SMSA framework. This architectural diagram elucidates the sequential process and integrated components that make up the SMSA system. The architecture is designed to handle, process, and analyze social media text data to accurately identify and categorize sentiments.

3.2.1. Data acquisition

At this stage, there are two critical activities: collecting community posts and creating a military emotion dictionary. Seventy-three negatively connotative words (such as bruise and heavenly army) and 53 positively connotative terms (such as lean and united) are included in the dictionary's content. These sentiment dictionaries were put to use in the subsequent model analysis to aid in the processes of learning and prediction.

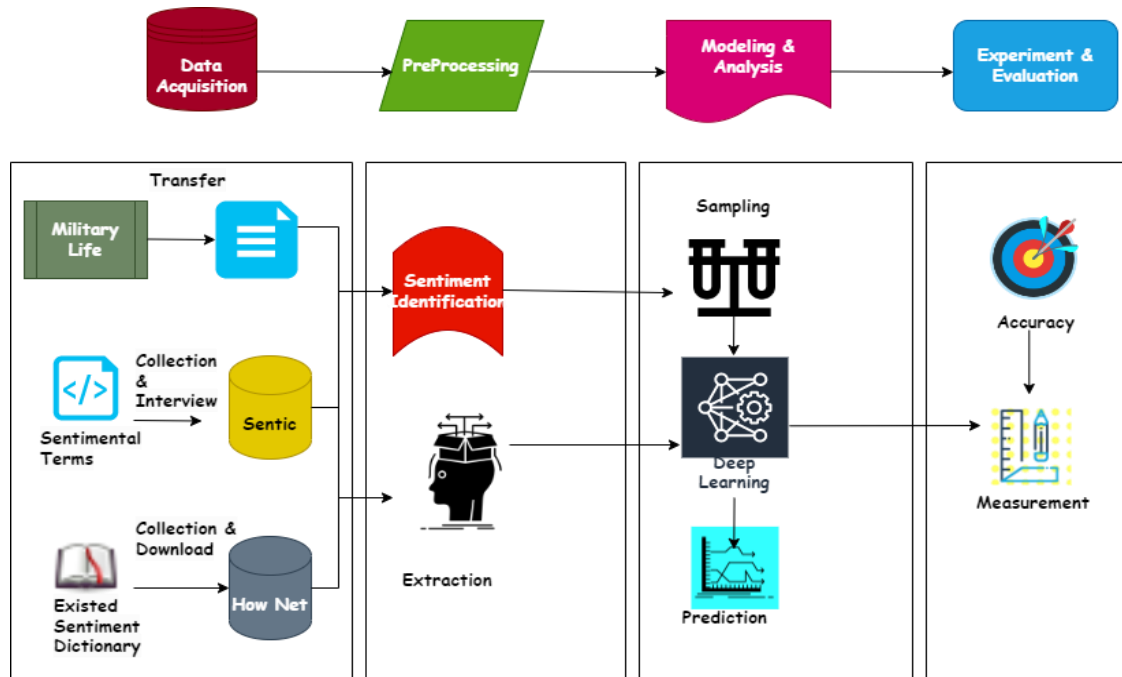


Figure 2. Overall architecture for SMSA framework

3.2.2. Preprocessing

Extraction and emotion recognition are two distinct processes during the preprocessing phase. The former focuses more on identifying the polarity of an article's tone, while the latter uses a sentiment dictionary to aid in segmenting words. Three modes are available when using a word segmentation system: complete, precise, and search engine. Analysis of several social media platforms revealed that both upbeat and downbeat stories were more popular than neutral ones among users.

The below workflow diagram depicts the sequence of NLP operations within the SMSA framework employed to categorize sentimental states from social media text data. The initial stage involves preprocessing the raw text through a suite of NLP tasks, graphically represented within the pink mobile phone icon. These tasks include tokenization, lemmatization, stemming, part-of-speech (POS) tagging, named entity recognition (NER), and removing stop words. Each circle within the mobile phone icon corresponds to a specific preprocessing function that prepares the text for intricate analysis.

3.2.3. Application of deep learning

Upon completion of text preprocessing, the data is conveyed to a deep learning algorithm, as indicated by the neural network icon. This model is responsible for discerning the sentiment from the processed text. The resultant sentiments are then categorized, as suggested by the elliptical 'Sentiment' symbol. The diagram also intimates additional analytical layers post-sentiment classification, represented by the series of rectangles, though these stages still need to be fully detailed in Figure 3.

Figure 3 illustrates the SMSA framework's operational pipeline. The process initiates with the raw social media text, which is subjected to comprehensive preprocessing to extract meaningful linguistic features essential for accurate sentiment analysis. This preprocessing is a multi-step process that involves tokenization to parse words and phrases; lemmatization and stemming to reduce words to their base or root form; POS tagging to classify words into their grammatical categories; NER to identify and categorize critical entities present in the text; and stop word removal, to eliminate common words that add noise to sentiment analysis.

The enriched text data enters the deep learning model following the preprocessing phase. The model, trained on extensive social media datasets, utilizes complex algorithms to discern patterns indicative of sentiment. The output of this model is a granular sentiment classification that reflects the nuanced emotional states expressed in the text. The SMSA framework addresses the challenges posed by the brevity and informal language commonly found in social media content. By leveraging a tailored deep learning approach, the framework achieves high accuracy in sentiment categorization, vital for applications such as market analysis, public opinion surveys, and social research.

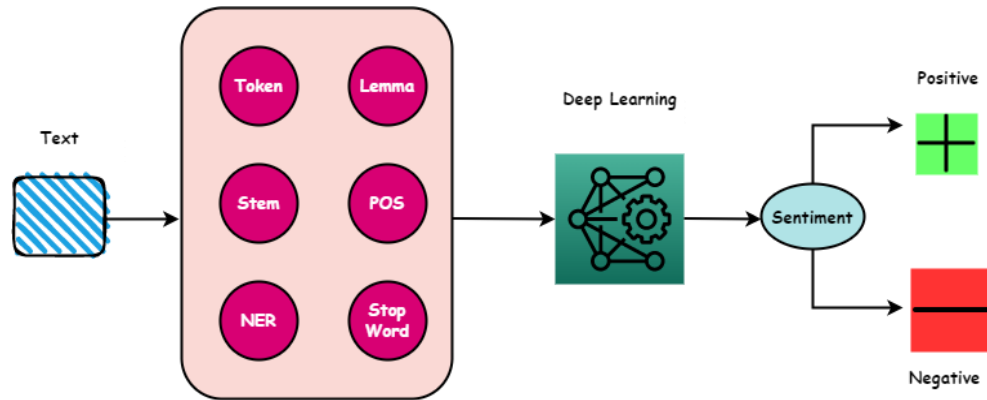


Figure 3. Categorizing sentimental states using SMSA framework

Conventional machine learning methods define and extract characteristics by hand and employ feature selection techniques. On the other hand, deep learning models use features in automated learning and extraction, leading to improved precision and efficiency. As a rule, hyper-automatically evaluating the classifier's parameters is also possible. Machine learning is the best answer to many issues in the fields currently provided by networks and deep learning used in NLP, image processing, and speech recognition. Features are not hardcoded in deep learning models; they are learned and improved precision and speed by being removed automatically. Typically, hyperparameters and other classifier models are also measured automatically in terms of accuracy.

3.2.4. Activation functions in neural networks

In neural network architectures, the activation function, denoted as f , is fundamental in determining a neuron's output from a given input or set of inputs. This function is predominantly nonlinear, expressed as $f(W^T x)$, where $W^T x$ represents the weighted sum of inputs x , weights W , and a bias term b . The introduction of non-linearity through these functions allows the model to discern and learn complex patterns within the data. Among the prevalent activation functions are the sigmoid, hyperbolic tangent (\tanh), and rectified linear unit (ReLU). Once widely used, the sigmoid function maps its input into a range between 0 and 1. However, its application has diminished over time due to the vanishing gradient problem it presents during backpropagation, where gradients shrink to negligible sizes, severely impeding the model's training process.

For the hyperbolic tangent function, given by (3), the \tanh function is more commonly used than the tan function since its output range is $[-1, 1]$ rather than $[0, 1]$. ReLU function has also gained in popularity recently. When the input is less than 0. It has a simple 0 threshold activation. Performance is greatly achieved from (3),

$$f(W^t x) = \tanh((W^t x) = \frac{e^{W^t(x)} - e^{-W^t(x)}}{e^{W^t(x)} + e^{-W^t(x)}} \tag{3}$$

ReLU is simple to calculate, quickly converges during training, and produces results that are on par with or better than those obtained using neural networks' sigmoid and tanh functions. It is activated only when the input exceeds 0, making it computationally efficient and often yielding results comparable or superior to those obtained using sigmoid and tanh functions in various neural network applications.

$$f(Wtx) = ReLU(W^t x) = \max (0, W^t x) \tag{4}$$

An additional noteworthy function is the softmax, often employed in the output neurons, particularly in classification tasks. It adapts the logistic function, transforming a K-dimensional vector of arbitrary absolute values into a K-dimensional vector of actual values within the range of $[0, 1]$ that collectively sum to 1. This is mathematically represented in (5). To encapsulate, the activation function in neural networks is a critical component that significantly influences the network's learning capability and proficiency in modeling intricate patterns. Each function's unique characteristics contribute to the network's overall performance and efficacy.

$$\Phi (x) = \frac{e^x}{\sum_k e^x} \text{ for } j = 1 \tag{5}$$

The softmax function, an extension of the logistic function, can be used as an output neuron in L3. This operation takes as input a K-dimensional vector X containing arbitrary absolute values and produces an identical K-dimensional vector (X) containing 0-1 fundamental values that add up to 1. Softmax is commonly used for classification in the hidden layer of a feedforward neural network. These gradients are used to modify the layer weights. A set of halting conditions must be completed before the process can be considered complete. Advances in hardware have made more processing power available, large volumes of training data are easily accessible, and powerful and versatile intermediate representations may be learned. Training a neural network to minimize the cross-entropy loss, a loss function for softmax output, is a prevalent task for which stochastic gradient descent via backpropagation is employed. Loss function gradients are calculated using output layer weights and hidden layer weights; we recursively apply the chain rule in a backward fashion to calculate expressions' gradients concerning inter-layer weights in a network.

4. RESULTS AND DISCUSSION

The simulation analysis of the proposed SMSA framework model rigorously evaluated several vital factors: accuracy, predictability, probability, robustness, and efficiency. These factors play a crucial role in determining the effectiveness and reliability of the model in analyzing sentiments from social media content. The data obtained from the simulations meticulously represent each of these aspects in the respective graphs provided below.

4.1. Accuracy analysis

Table 1 illustrates the comparative performance of six different analytical methods: deep learning (DL), term frequency-inverse document frequency (TF-IDF), time series analysis (TSA), statistical analysis system (SAS), deep neural networks (DNN), and social media sentiment analysis (SMSA); across a range of sample sizes (0 to 500). As the number of samples increases, there is a general trend of improvement in the performance of all methods. The SMSA method consistently outperforms other methods across all sample sizes. The DNN and SAS methods significantly improve as the sample size increases, particularly noticeably beyond 300 samples.

Figure 4 presents the accuracy study of the suggested model. Accuracy requires measurements to approach a reference value closely. It is a quantitative indicator of trustworthiness, which refers to whether a calculated or observed value is consistent with an accurate value. Both accuracy and precision are crucial in many contexts, but they are different.

Table 1. Accuracy analysis

Number of samples	DL	TF-IDF	TSA	SAS	DNN	SMSA
0	50	52	57	67	80	92
100	68	63	73	73	92	94
200	71	58	77	78	94	98
300	72	71	79	81	96	96
400	80	75	83	86	96.4	94
500	83	79	88	88	96.2	97.5

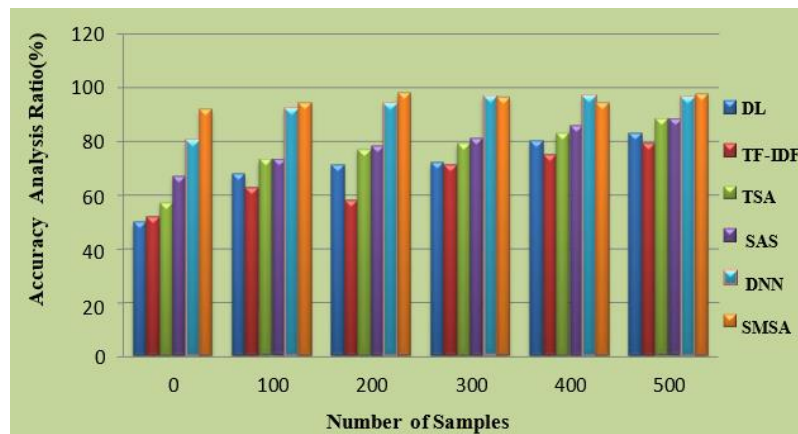


Figure 4. Accuracy analysis

Figure 5 presents a compare the accuracy of six analytical methods: DL, TF-IDF, TSA, SAS, DNN, SMSA; across a range of sample sizes from 0 to 500. The accuracy of all methods improves as the number of samples increases, suggesting that more data contributes to better model performance. The SMSA method consistently demonstrates high accuracy across all sample sizes, indicating its robustness and reliability. The DNN method shows a marked increase in accuracy as the sample size grows, especially after the 300-sample mark, suggesting its capacity to leverage larger datasets effectively. DL and TF-IDF methods show improvement with increased sample sizes but tend to plateau, hinting at potential limitations in these methods or an optimal data threshold.

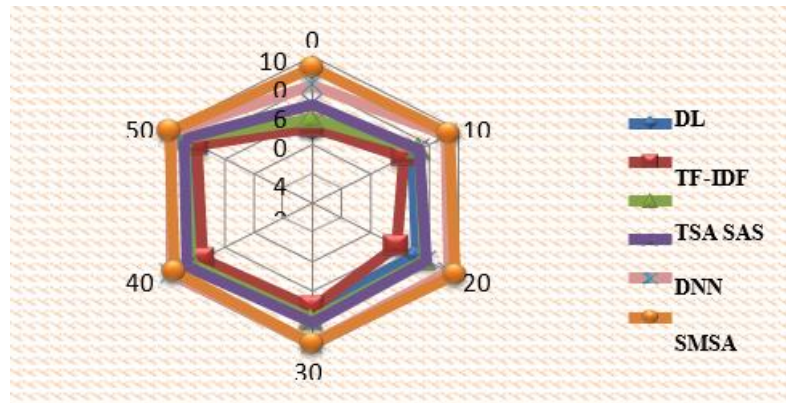


Figure 5. Probability analysis

4.2. Prediction analysis

Foresight analysis uses accumulated data and computational intelligence methods for statistical forecasting. Predictive analytics incorporates many statistical methods, such as data analysis, predictive modeling, and machine learning. These methods look at the here-and-now and the past to predict the unpredictable. The point is not to learn from history but to predict the future as accurately as possible. Figure 6 displays a line chart detailing the trend of accuracy for four analytical methods: DL, TF-IDF, TSA, and SAS; as the number of samples increases from 0 to 50. The accuracy of all methods displays a slight upward trend as the sample size grows, suggesting a positive correlation between sample size and accuracy, albeit with a modest slope.

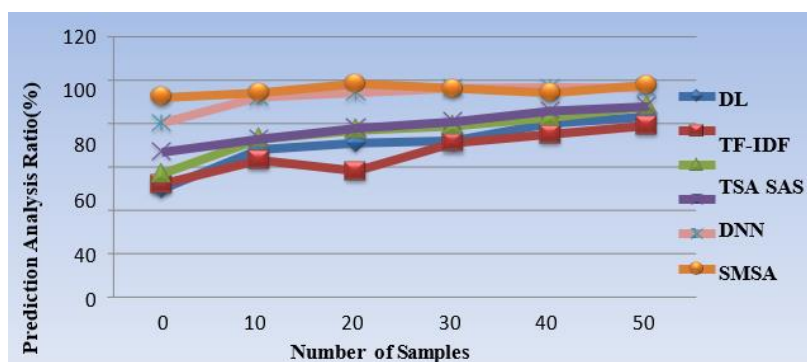


Figure 6. Prediction analysis

4.3. Probability analysis

Figure 5 provides a probability analysis of this word's spread across existing and proposed methods. This concept refers to the likelihood of someone continuing to exist or maintaining their vitality. Depending on the context, the phrase can have a more nuanced meaning. Persuasive speech encompasses all the reasons why words and vocabulary are vital to human existence.

4.4. Robustness analysis

Figure 7 presents a 3D bar chart detailing the accuracy of six analytical methods: DL, TF-IDF, TSA, SAS, DNN, and SMSA; for increasing sample sizes (0, 100, 200, 300, 400, and 500). The z-axis indicates cross-validation accuracy percentages, with each bar height depicting the average accuracy achieved by a method during the cross-validation process for a given sample size. The chart illustrates a general trend showing that cross-validation accuracy improves as the sample size increases for all methods. This improvement indicates that the models have an enhanced ability to generalize when trained on more data. The SMSA method consistently achieves the highest cross-validation accuracy, indicating its suitability for the data and its potential to generalize well to unseen data. In contrast, DL and TF-IDF exhibit the lowest cross-validation accuracy, which may imply that these methods are either more prone to overfitting or less capable of capturing the underlying pattern with limited data.

The modest improvements in cross-validation accuracy for DL and TF-IDF may point to scalability issues or a ceiling in performance, possibly due to model simplicity or complexity not aligning optimally with the data's structure. Meanwhile, the marked rise in the cross-validation accuracy of the DNN method with larger sample sizes suggests that the model benefits from more data points to avoid overfitting, thus achieving better generalization. The SMSA method's consistent performance across various sample sizes during cross-validation underscores its robustness. It has a solid algorithmic foundation capable of handling complex datasets while maintaining high accuracy.

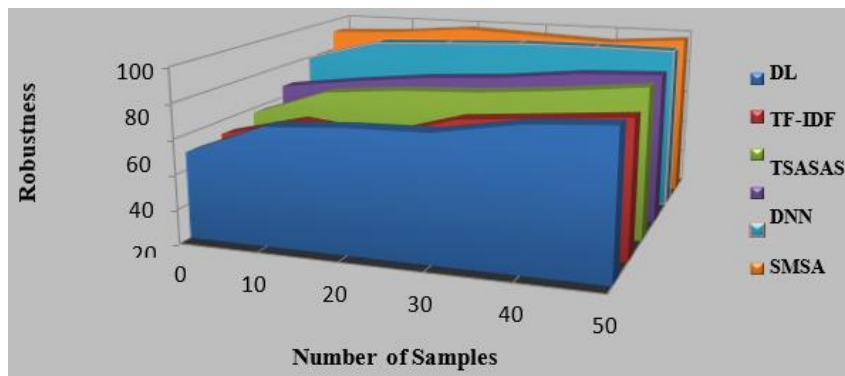


Figure 7. Accuracy analysis

4.5. Efficiency analysis

Efficiency entails maximizing all available resources to achieve optimal results. When considering efficiency, opting for the most straightforward and time-efficient approach is essential to accomplish goals with minimal time and effort. Figure 8 illustrates the outcomes of effectiveness tests. Compared to widely used models like DL, TF-IDF, TSA, SAS, and DNN, the proposed SMSA surpasses them in all areas. The study's findings led to new technologies to address the dual challenges.

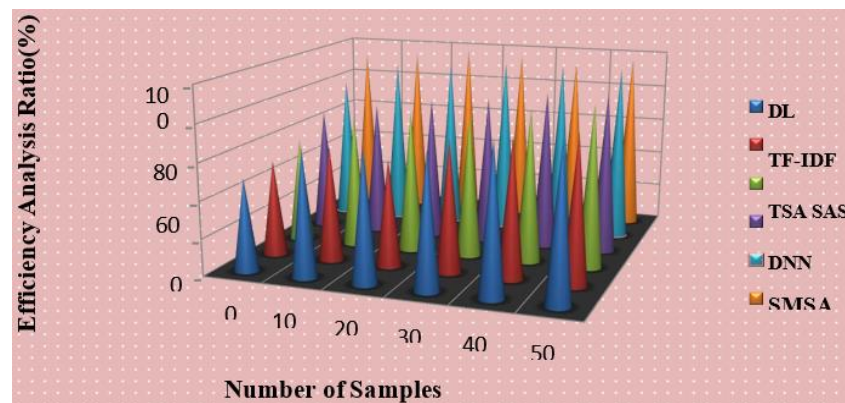


Figure 8. Efficiency analysis comparison

5. CONCLUSION

Rapid user engagement on social media platforms, coupled with user-friendly interfaces, has led to a substantial increase in community data. This growth has emphasized the need for effective sentiment analysis in internet and social media contexts, a problem that existing deep learning approaches have begun to address but with limitations in sentiment polarity determination and other challenges beyond conventional natural language processing. Our research addresses these challenges by focusing on deep learning-based sentiment analysis, augmented with a custom-developed sentiment dictionary, a novel approach not extensively explored in previous studies. The proposed solution involves fine-tuning deep learning models with our unique sentiment dictionary to enhance the accuracy and depth of sentiment analysis. This approach has demonstrated significant improvements in the precision of sentiment classification. The study not only reviews existing literature on sentiment analysis and its challenges, as highlighted in recent research, but also introduces new methodologies and data sets to overcome these challenges. Our results indicate a marked improvement in sentiment analysis, especially in areas where previous models faced limitations. This research contributes to the field by offering a comprehensive framework for SMSA and proposing a new sentiment dictionary. These advancements not only enhance the field of sentiment analysis but also extend the capabilities of artificial intelligent in understanding and interpreting complex human emotions on social media.

ACKNOWLEDGEMENT

The authors extend their appreciation to Prince Sattam bin Abdulaziz University for funding this research work through the project number (PSAU/2024/01/78918)




REFERENCES

- [1] R. Arulmurugan, K. R. Sabarmathi, and H. Anandakumar, "Classification of sentence level sentiment analysis using cloud machine learning techniques," *Cluster Computing*, vol. 22, no. S1, pp. 1199–1209, Sep. 2019, doi: 10.1007/s10586-017-1200-1.
- [2] R. Patel and K. Passi, "Sentiment analysis on Twitter data of World Cup Soccer Tournament using machine learning," *Internet of Things*, vol. 1, no. 2, pp. 218–239, Oct. 2020, doi: 10.3390/iot1020014.
- [3] M. Soleymani, D. Garcia, B. Jou, B. Schuller, S. F. Chang, and M. Pantic, "A survey of multimodal sentiment analysis," *Image and Vision Computing*, vol. 65, pp. 3–14, Sep. 2017, doi: 10.1016/j.imavis.2017.08.003.
- [4] K. H. Manguri, R. N. Ramadhan, and P. R. Mohammed Amin, "Twitter sentiment analysis on worldwide COVID-19 outbreaks," *Kurdistan Journal of Applied Research*, pp. 54–65, May 2020, doi: 10.24017/covid.8.
- [5] S. A. Abdu, A. H. Yousef, and A. Salem, "Multimodal video sentiment analysis using deep learning approaches, a survey," *Information Fusion*, vol. 76, pp. 204–226, Dec. 2021, doi: 10.1016/j.inffus.2021.06.003.
- [6] D. Li, R. Rzepka, M. Ptaszynski, and K. Araki, "HEMOS: a novel deep learning-based fine-grained humor detecting method for sentiment analysis of social media," *Information Processing and Management*, vol. 57, no. 6, Nov. 2020, doi: 10.1016/j.ipm.2020.102290.
- [7] U. Ö. Osmanoglu, O. N. Atak, K. Çağlar, H. Kayhan, and T. Can, "Sentiment analysis for distance education course materials: a machine learning approach," *Journal of Educational Technology and Online Learning*, vol. 3, no. 1, pp. 31–48, Jan. 2020, doi: 10.31681/jetol.663733.
- [8] N. Sultan, "Sentiment analysis of Amazon product reviews using supervised machine learning techniques," *Knowledge Engineering and Data Science*, vol. 5, no. 1, pp. 101–108, Jun. 2022, doi: 10.17977/um018v5i12022p101-108.
- [9] R. Chandra and A. Krishna, "COVID-19 sentiment analysis via deep learning during the rise of novel cases," *PLoS ONE*, vol. 16, Aug. 2021, doi: 10.1371/journal.pone.0255615.
- [10] E. Boiy and M. F. Moens, "A machine learning approach to sentiment analysis in multilingual web texts," *Information Retrieval*, vol. 12, no. 5, pp. 526–558, Sep. 2009, doi: 10.1007/s10791-008-9070-z.
- [11] J. Guo, "Deep learning approach to text analysis for human emotion detection from big data," *Journal of Intelligent Systems*, vol. 31, no. 1, pp. 113–126, Jan. 2022, doi: 10.1515/jisys-2022-0001.
- [12] M. T. J. Ansari and N. A. Khan, "Worldwide COVID-19 vaccines sentiment analysis through twitter content," *Electronic Journal of General Medicine*, vol. 18, no. 6, p. em329, Nov. 2021, doi: 10.29333/ejgm/11316.
- [13] A. Yadav and D. K. Vishwakarma, "Sentiment analysis using deep learning architectures: a review," *Artificial Intelligence Review*, vol. 53, no. 6, pp. 4335–4385, Dec. 2020, doi: 10.1007/s10462-019-09794-5.
- [14] L. Williams, C. Bannister, M. Arribas-Ayllon, A. Preece, and I. Spasić, "The role of idioms in sentiment analysis," *Expert Systems with Applications*, vol. 42, no. 21, pp. 7375–7385, Nov. 2015, doi: 10.1016/j.eswa.2015.05.039.
- [15] S. Kardakis, I. Perikos, F. Grivokostopoulou, and I. Hatzilygeroudis, "Examining attention mechanisms in deep learning models for sentiment analysis," *Applied Sciences*, vol. 11, no. 9, Apr. 2021, doi: 10.3390/app11093883.
- [16] W. Liu, B. A. Muthu, and C. B. Sivaparthipan, "Sentimental analysis in student-teacher communication for effective learning," *Aggression and Violent Behavior*, p. 101629, Jun. 2021, doi: 10.1016/j.avb.2021.101629.
- [17] H. H. Do, P. W. C. Prasad, A. Maag, and A. Alsadoon, "Deep learning for aspect-based sentiment analysis: a comparative review," *Expert Systems with Applications*, vol. 118, pp. 272–299, Mar. 2019, doi: 10.1016/j.eswa.2018.10.003.
- [18] K. Shakhovska, N. Shakhovska, and P. Vesely, "The sentiment analysis model of services providers' feedback," *Electronics (Switzerland)*, vol. 9, no. 11, pp. 1–15, Nov. 2020, doi: 10.3390/electronics9111922.
- [19] M. S. F. Alharbi and E.-S. M. El-kenawy, "Optimize machine learning programming algorithms for sentiment analysis in social media," *International Journal of Computer Applications*, vol. 174, no. 25, pp. 38–43, Mar. 2021, doi: 10.5120/ijca2021921169.
- [20] P. Chakriswaran, D. R. Vincent, K. Srinivasan, V. Sharma, C.-Y. Chang, and D. G. Reina, "Emotion AI-driven sentiment analysis: a survey, future research directions, and open issues," *Applied Sciences*, vol. 9, no. 24, Dec. 2019, doi: 10.3390/app9245462.
- [21] A. Gandhi, K. Adhvaryu, S. Poria, E. Cambria, and A. Hussain, "Multimodal sentiment analysis: a systematic review of history,




- datasets, multimodal fusion methods, applications, challenges and future directions,” *Information Fusion*, vol. 91, pp. 424–444, Mar. 2023, doi: 10.1016/j.inffus.2022.09.025.
- [22] M. S. Hossain, M. F. Rahman, M. K. Uddin, and M. K. Hossain, “Customer sentiment analysis and prediction of halal restaurants using machine learning approaches,” *Journal of Islamic Marketing*, vol. 14, no. 7, pp. 1859–1889, Jun. 2023, doi: 10.1108/JIMA-04-2021-0125.
- [23] J. Godara, R. Aron, and M. Shabaz, “Sentiment analysis and sarcasm detection from social network to train health-care professionals,” *World Journal of Engineering*, vol. 19, no. 1, pp. 124–133, Jun. 2022, doi: 10.1108/WJE-02-2021-0108.
- [24] M. Nissim and V. Patti, “Semantic aspects in sentiment analysis,” in *Sentiment Analysis in Social Networks*, Elsevier, 2017, pp. 31–48.
- [25] W. Medhat, A. Hassan, and H. Korashy, “Sentiment analysis algorithms and applications: a survey,” *Ain Shams Engineering Journal*, vol. 5, no. 4, pp. 1093–1113, Dec. 2014, doi: 10.1016/j.asej.2014.04.011.
- [26] L. Zhang, S. Wang, and B. Liu, “Deep learning for sentiment analysis: a survey,” *WIREs Data Mining and Knowledge Discovery*, vol. 8, no. 4, Jul. 2018, doi: 10.1002/widm.1253.
- [27] N. C. Dang, M. N. Moreno-García, and F. De la Prieta, “Sentiment analysis based on deep learning: a comparative study,” *Electronics*, vol. 9, no. 3, p. 483, Mar. 2020, doi: 10.3390/electronics9030483.
- [28] G. Swathi, R. P. Kumar, B. M. G., and E. R., “Ensemble integration of deep learning models for gender-based speech emotion recognition,” in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Jul. 2023, pp. 1–7, doi: 10.1109/ICCCNT56998.2023.10306801.
- [29] L. C. Chen, C. M. Lee, and M. Y. Chen, “Exploration of social media for sentiment analysis using deep learning,” *Soft Computing*, vol. 24, no. 11, pp. 8187–8197, Oct. 2020, doi: 10.1007/s00500-019-04402-8.
- [30] R. Wadawadagi and V. Pagi, “Sentiment analysis with deep neural networks: comparative study and performance assessment,” *Artificial Intelligence Review*, vol. 53, no. 8, pp. 6155–6195, May 2020, doi: 10.1007/s10462-020-09845-2.
- [31] R. P. Kumar *et al.*, “Sarcasm detection in Telugu and Tamil: an exploration of machine learning and deep neural networks,” in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Jul. 2023, pp. 1–7, doi: 10.1109/ICCCNT56998.2023.10306775.
- [32] C. Spandana, I. V. Srisurya, S. Aasha Nandhini, R. P. Kumar, G. Bharathi Mohan, and P. Srinivasan, “An efficient genetic algorithm based auto ML approach for classification and regression,” in *2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*, Jan. 2023, pp. 371–376, doi: 10.1109/IDCIoT56793.2023.10053442.
- [33] S. Padma, O. V. R. Murthy, and S. Veni, “Sentiment analysis by deep learning approaches,” *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 18, no. 2, pp. 752–760, Apr. 2020, doi: 10.12928/telkomnika.v18i2.13912.
- [34] G. Bharathi Mohan and R. Prasanna Kumar, “Survey of text document summarization based on ensemble topic vector clustering model,” in *Lecture Notes in Networks and Systems*, vol. 528, Springer Nature Singapore, 2023, pp. 831–847.
- [35] B. M. Gurusamy, P. K. Rengarajan, and P. Srinivasan, “A hybrid approach for text summarization using semantic latent Dirichlet allocation and sentence concept mapping with transformer,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 6, pp. 6663–6672, Dec. 2023, doi: 10.11591/ijece.v13i6.pp6663-6672.
- [36] M. Adraoui, S. Souabi, A. Retbi, M. K. Idrissi, and S. Bennani, “Towards a hybrid recommendation approach using a community detection and evaluation algorithm,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 6, pp. 6718–6728, Dec. 2023, doi: 10.11591/ijece.v13i6.pp6718-6728.

BIOGRAPHIES OF AUTHORS






Prasanna Kumar Rangarjan    received his Ph.D. degree from Anna University, Chennai, Tamil Nādu, and India. He is currently working as chairperson and associate professor in Amrita School of Computing, Amrita University, India. He has 20 years of experience in teaching. Since 2010 he has published scientific papers in data mining, data analysis, time series analysis and computer vision. His areas of interest include data analytics, machine learning, theory of computation, compiler design, and python programming. He can be contacted at email: r_prasannakumar@ch.amrita.edu.






Bharathi Mohan Gurusamy    pursuing Ph.D. degree in Amrita School of Computing, Amrita Vishwa Vidyapeetham Chennai, Tamil Nadu, India. He has 15 years of teaching experience and 3 experience of research experience. He is currently working as an assistant professor with the Amrita School of Computing, Amrita Vishwa Vidyapeetham University, and Chennai, India. Since 2021 he has been publishing scientific papers in text summarization. His research interests include natural language processing, computational intelligence, machine learning, and deep learning. He can be contacted at email: g_bharathimohan@ch.amrita.edu.






Gayathri Muthurasu    received a B.E. degree in electronics and instrumentation from the Department of Electronics and Instrumentation, Adhiyamaan College of Engineering, Hosur in 2011, and a master's degree in applied electronics from the Department of Electronics and Communication Engineering, Chennai, in 2014. She is currently pursuing her Ph.D. degree in electronics and communication engineering at SRM Institute of Science and Technology, Vadapalani. Her research interests include vehicular ad hoc networks, wireless communication, wireless networks, network security. She can be contacted at email: gayathri.muthurasu@gmail.com.






Rithani Mohan    pursuing Ph.D. degree in Amrita school of computing, Amrita Vishwa Vidyapeetham Chennai, Tamil Nadu, India. She has 2 years of teaching experience and 1.5 years of research experience. She is currently working as a Faculty Associate with the Amrita School of Computing, Amrita Vishwa Vidyapeetham University, Chennai, India. Since 2023 she has been publishing scientific papers in big data and deep learning. Her research area includes deep learning, bigdata, machine learning. She can be contacted at email: m_rithani@ch.amrita.edu.






Gundala Pallavi    pursuing Ph.D. degree in Amrita School of Engineering Amrita Vishwa Vidyapeetham Chennai, Tamil Nadu, India. She has 5 years of teaching experience and 1 year of research experience. Her research interests include natural language processing, machine learning, deep learning, quantum artificial intelligence, bioinformatics. She can be contacted at email: Pallavimalliswari@gmail.com.



Sulochana Vijayakumar    currently serving as a Course Director at the Anna Administrative Staff College, Government of Tamil Nadu, She hold a Ph.D. in computer science and engineering, along with a master's degree in the same fields. This educational background has equipped me with a profound understanding of computer science and engineering intricacies. In her role, she orchestrates the creation of meticulously tailored courses for professionals and students, using needs analysis, project management, course development, and evaluation methodologies. Specializing in information and communication technology, she designed programs on emerging technology and cyber security, like artificial intelligence for good governance and cybersecurity for organizational safety. She can be contacted at email: sulo62002@yahoo.com.



Ali Altalbe    received his Ph.D. in information technology from The University of Queensland in Australia, completed his M.Sc. in information technology from Flinders University, Australia. He is currently working as associate professor in the Department of IT, King Abdulaziz University, Jeddah, Saudi Arabia. She can be contacted at email: aaltalbi@kau.edu.sa.