

Contextual-cognitive convolutional attentive LSTM approach for sentiment analysis of Tamil review

M. Kavitha^{1*} and Dr. A. John Sanjeev Kumar²

^{1*}Assistant Professor, Department of Computer Applications, Mannar Thirumalai Naicker College, India, kavitha.rml@gmail.com

²Head and Assistant Professor, Department of Data Science, The American College, India, johnsanjeevkumar@gmail.com

Corresponding Author: kavitha.rml@gmail.com

ABSTRACT

Movie goers use Twitter and other websites to share their thoughts and feelings regarding Tamil films in this age of social networking and online interaction. It is essential to examine these feelings for realizing the responses of the public and assess a film's level of success. In this research, a novel contextual-cognitive convolutional attentive long/short-term memory (C3-ALSTM) strategy for assessing Tamil movie reviews via Tamil tweets is presented. We first gather the Tamil movie review dataset through an open database to assess the suggested C3-ALSTM approach. We normalize the gathered data by eliminating extraneous information, noise and redundant data. Tokenization and the removal of stop words and special characters are employed in this phase. Following data cleansing, term frequency-inverse document frequency (TF-IDF) is applied for extracting the attributes. The suggested approach evaluates the Tamil reviews based on these retrieved features. The suggested approach is put into practice using the Python platform and its accuracy, precision, recall and f-measure metrics are examined. According to the research's findings, the suggested method outperforms other used techniques for sentiment assessment of Tamil reviews. In addition, this research attempts to offer insightful information about how the public reacts to particular Tamil films, advancing knowledge of public preferences and attitudes toward the Tamil cinema industry.

Keywords: Tamil movie, Sentiment Analysis (SA), Deep learning (DL), contextual-cognitive convolutional attentive long/short term memory (C3-ALSTM) technique

INTRODUCTION

Sentiment analysis (SA), which refers to the process of gathering information from a large number of individuals to get a better understanding of the attitudes and responses of those people stated on the internet in relation to a variety of topics pertaining to the globe [1]. With more people using the internet, there is a wealth of knowledge on various goods, movies, books, technology and other topics that are accessible on the internet [2]. Individual audiences share their thoughts, ideas and perspectives on various books, services and items on the internet. When a consumer purchases a smart phone, for instance, they are quick to provide feedback about their impressions of the device and the features that they found appealing or unappealing [3]. The industry is from these kinds of user or consumer evaluations and feedback. Social media is a lively platform where people can share their thoughts, feelings, opinions and opinions on a range of issues [4]. Politics, social awareness, conflict, film reviews, school systems, comments on new or current products and other topics are debated on Twitter [5]. Tweets that indicate people's opinions about a business or topic can have an effect on the choices that are made by the firm, the government and the private sector. The fact that Twitter has a large public data collection that allows academics to examine user viewpoints is another factor that contributes to its success [6]. Technology has made it possible for people to create online content to share knowledge and experience, voice views and leave comments on a variety of interesting subjects as the Internet becomes more important in our daily lives [7]. The study on SA has grown more relevant in the domains of online and text mining, analysis of social media as well as natural language processing (NLP). It is a type of research that is included in the subfield of psychology that is known as social psychology. This branch of psychology investigates people's thoughts, feelings, evaluations, perspectives and sentiments in relation to a wide variety of subjects and entities, including goods and services, organizations, people, problems, events and topics, as well as the characteristics of these things. [8].Tamil is one of the world's oldest living languages, dating back to at least the 2nd century. Throughout seventy eight million individuals throughout the globe are able to communicate with one another via the use of this language [9]. It has been recognized as an official language in Singapore and Sri Lanka, in addition to Tamilnadu (India). It is one of the old Dravidian languages that use the subject-object-verb (SOV) pattern. There is a possibility that the sequence will shift when the phrases are constructed in Tamil [10]. In Tamil, it is possible to form whole sentences using only the verb, subject and the object of phrase. This is referred to as SOV construction. Implementing sentiment

analysis on reviews of Tamil movies using the C3-ALSTM methodology is to achieve precise and comprehensive categorization of sentiments.

1.1. KEY CONTRIBUTIONS

- The study provides a Tamil movie review dataset. This dataset from an open database is useful for testing Tamil sentiment analysis methods.
- Research uses tokenization for data normalization to resolve noisy and redundant data. To improve dataset quality, remove noise, redundancy and superfluous data.
- TF-IDF extracts features from extracted data in the study. This method helps to evaluate Tamil movie reviews' sentiment by identifying significant phrases.

2. RELATED WORK

The study [11] examined Lexicon-based, Supervised Machine Learning-based, Hybrid, K-means with Bag of Word (BoW) and K-modes with BoW were the five different types of SA approaches. They tested these methods on five corpora with various feature representation methodologies to determine the optimum SA method for Tamil literature. They employed basic characteristics like word count and punctuation count together with standard features like Bag of Words (BoW) and TF-IDF to test their impact on prediction. The paper [12] provided an area of SA in general and an examination of Indian languages in particular. The present position of Indian languages in the field of SA was categorized according to the families of Indian languages. It contained sources of publications selected for their relevance to the periodic development of Indian languages in the region of SA. SA taxonomy for Indian languages has been provided for categorizing languages according to method, domain, sentiment level and class. The paper [13] described an Indian language sentiment analysis with an emphasis on coding mixed languages. Natural Language Processing (NLP) academics have a lot of text data from Indian consumers who express their sentiments in several languages. Code-mixed text SA benefits politics, marketing, business, health, sports and more.

The research [14] was to provide a critical analysis on the comparison of the problems associated with SA of Tweets published in English Language against Indian Regional Languages. Tamil, Malayalam, Telugu and Hindi, four of India's official languages, were considered in the study. In order to identify and conceptualize in the form of a framework has a certain issues that related to the analysis of Twitter sentiments in those languages, a systematic review was used. The article [15] aimed to provide a comparative examination of sentiment analysis in several Indian languages. The categorization approaches discussed in this paper were based on lexicons, dictionaries and machine learning. They provided a list of the lexical resources that was used to carry out SA on Indian languages and the difficulties that might arise when attempting to generate linguistic resources for Indian languages that has limited resources. The study [16] suggested a method for interpreting tweets written in one of the 3 Indian languages (Tamil, Bengali and Hindi). With the best possible parameter values, 39 sequential models have been developed utilizing three distinct neural network layers: convolutional neural networks (CNNs), long short-term memory (LSTM) and recurrent neural networks (RNNs) (to minimize over-fitting and error accumulation). The goal of the research [17] was to use linguistic norms and a NLP toolkit to develop a tool that would be useful in anticipating the movie genre as perceived by the audience. The goal of the research was to develop a system for classifying the tone of tweets written in Tamil using rules. An NLP toolkit and Python has been used in the tool's creation. The study's [18] goal was to develop a system that can harvest Twitter data from actual users. User thoughts, reviews and comments about a product are becoming more commonplace on social media as its reach expands. In this paper, their contribution on user tweets to determine the feelings conveyed by people regarding Tamil movies based on grammar rule. They had chosen to focus on the Tamil film industry. The purpose of the work [19] was to provide a comparative examination of SA carried out in several Indian languages. The paper discusses a categorization techniques based on dictionaries, lexicons and machine learning. It provided a list of the lexical resources that can be used to carry out SA on Indian languages and the difficulties when attempting to generate lexical resources for Indian languages that has limited resources.

3. METHODOLOGY

The stages involved in the suggested strategy are outlined in this section. The proposed techniques of conceptual representation that covers tokenization processes for preprocessing, term frequency-inverse document frequency (TF-IDF) is employed for feature extraction. Figure 1 shows the process of proposed method.

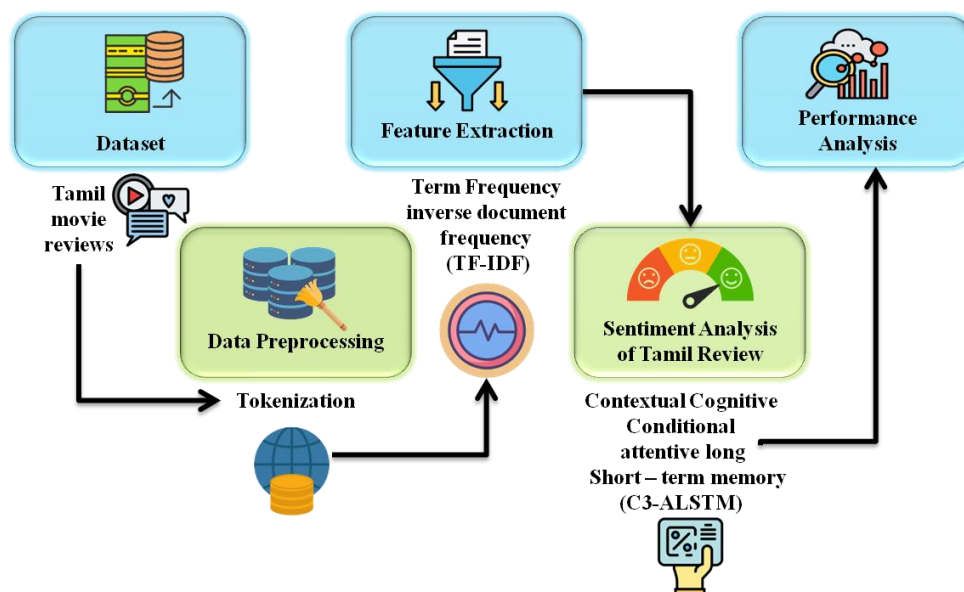


Figure 1: Overview of the study

3.1. DATASET

In this part, the study gathered the Tamil movie reviews dataset from Tamil websites for the popular Indian film Soorai pottru (2020) <https://www.indiaglitz.com/soorai-pottru-review-tamil-movie-22727>. The indiaglitz website collects and annotates positive and negative reviews.

3.2. PRE-PROCESSING USING TOKENIZATION

Pre-processing is the process of organizing a dataset by applying fundamental operations on it before submitting it to a model. Examples of these operations include removing gaps and meaningless words, rearranging word forms so they correspond to their original terms, eliminating duplicate words and so forth. It transforms the unprocessed material into a structured and helpful dataset for further usage. A method that separates a piece of writing into its component words, phrases, or other key components is known as tokens. Tokens are separated from one another by whitespace, punctuation marks and line breaks. When using tokenization, removing characters such as punctuation marks and other similar characters is common practice. Tokenization is regarded as a reasonably simple technique compared to other preprocessing methods. The process of tokenization involves chopping up each text or document into smaller bits of words that can be filed away into different categories. Convert the content to a reduced form and repair any spelling errors as part of the normalization process, which is guaranteed that the data is consistent.

3.3. TF-IDF FOR FEATURE EXTRACTION

TF-IDF is one of the most common approaches used in information retrieval and text mining. The TF-IDF is a weight measure that assesses the value of a word for that particular document. This approach helps to organize formal writings such as news articles, blogs and reviews of films. TF-IDF is insufficient to categorize tweets because of the informal nature of the format of tweets. Figure 2 depicts the flow of TF-IDF. As a benchmark for our work, we choose TF-IDF since it reveals the significance of the term based on the tweet data set. From the

Tamil movie dataset, we select a small number of important Tamil keywords. We selected the pertinent keywords in order to match the Tamil movie dataset. Tamil tweets are categorized using the top n TF-IDF keywords for each movie. Consider the movie o_i and its accompanying tweets $\{x_{(1)}, x_{(2)} \dots x_{(m)}\}$. Tweets include a collection of words $\{x_{(1)}, x_{(2)} \dots x_{(j)}\}$ we compute $T_f(x_{(j)}, o_i)$ and $idf(x_{(j)}, o_i)$ for Tamil movie tweets using keywords.

$$T_f(x_{(j)}, o_i) = (\text{frequency in overall tweets}) / (\text{total movie tweets}) \tag{1}$$

$$idf(x_{(j)}, o_i) = \log[\text{Movietweetcount} / (\text{Tweets that include the word count } x_j)] \tag{2}$$

We focused on the essential elements of the review's keyword list during our conversation with the film reviewers. They proposed the following key words: “வெற்றி (success), தோல்வி (loss), வசூல் (collection), அருமையான படம் (excellent movie), நகைச்சுவை (comedy)”. To perform the TF-IDF calculation, we considered each term individually.

$$[TF]_{\text{வெற்றி}}(\text{success}) = 21.0 / 495.1 = 0.04 \tag{3}$$

$$[IDF]_{\text{வெற்றி}}(\text{success}) = \log_{10}(495.1 / 19.1) = 1.41 \tag{4}$$

$$[TF-IDF]_{\text{வெற்றி}}(\text{success}) = 0.042 * 1.41 = 0.060 \tag{5}$$

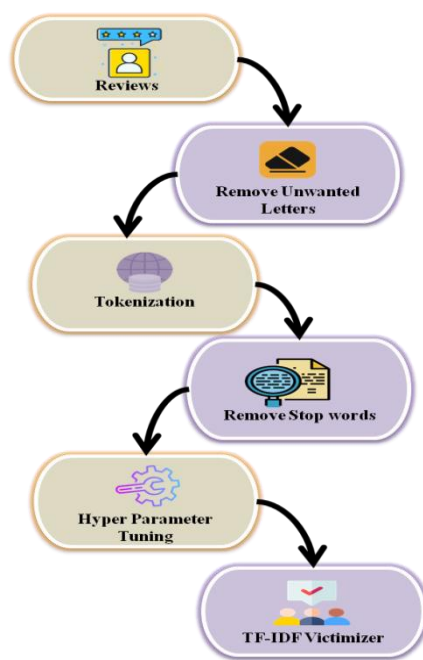


Figure 2: Steps involved in TF-IDF

3.4. CONTEXTUAL-COGNITIVE CONVOLUTIONAL ATTENTIVE LONG/SHORT TERM MEMORY (C3-ALSTM)

The temporal correlations that exist between the words in a phrase can be captured by recurrent neural networks (RNN). Textual information is a kind of time-series data in which the sequence of words and sentences is a crucial component in determining the meaning of the information. ALSTM is a form of RNN model that utilizes three cells an input cell, an output cell and a forget cell to analyze the supplied input sequence and create the desired

output. Figure 3 depicts an individual C3-ALSTM unit. ALSTM networks find use in various fields, including machine translation, spelling checking, text categorization, handwritten recognition and many more areas.

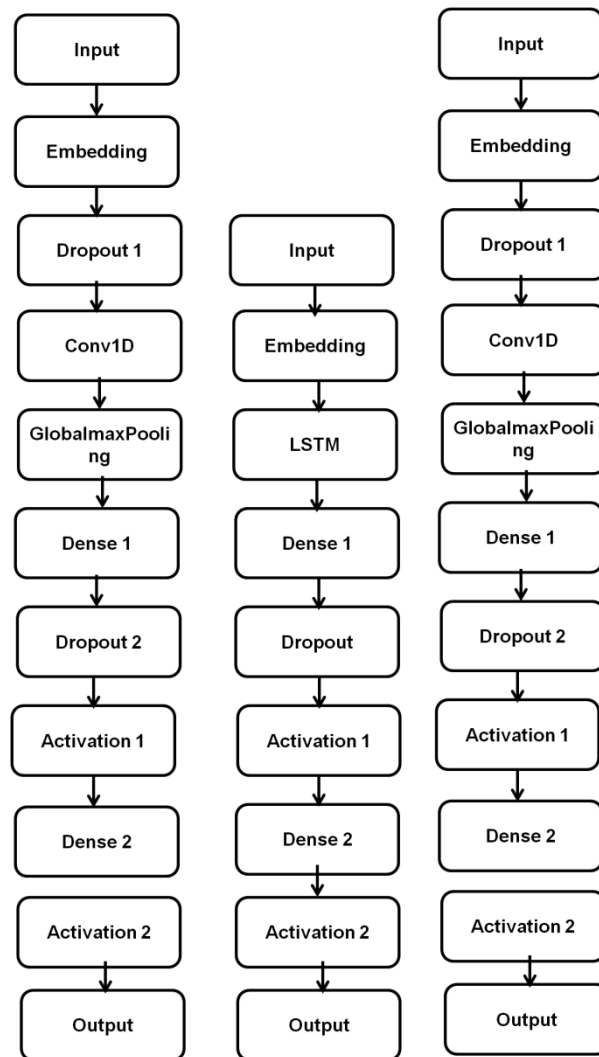


Figure 3: C3-ALSTM diagram

The Contextual ALSTM was the tool that we used for vulnerability detection. The contextual ALSTM model is an extension of the recurrent neural network in which contextual variables are included into the model. Automatically learning good representations allows a convolutional filter to function without representing the whole language. CNNs simulate the connection between scales using varying window widths, which results in the loss of most background information. C3-ALSTM classifies text using CNN as well as RNN and feeds ALSTM of the new features from CNN training. The standard ALSTM layer is shown in equation (6):

$$j_s = \sigma(X_{wj} w_s + X_{gj} g_{(s-1)} + X_{dj} d_{(s-1)} + a_s)$$

$$e_s = \sigma(X_{we} w_s + X_{ge} g_{(s-1)} + X_{de} d_{(s-1)} + a_e)$$

$$p_s = \sigma(X_{wp} w_s + X_{gp} g_{(s-1)} + X_{dp} d_{(s-1)} + a_p)$$

$$d_s = e_s \cdot d_{(s-1)} + j_s \cdot \tanh(X_{wf} w_s + X_{gd} g_{(s-1)} + a_d) \tag{6}$$

$$g_s = p_s \cdot \tanh(\tilde{f}_0)(d_s)$$

Equation (6) uses, j, e, p as input, forget as well as output gates, w as input, a as bias, d as cell memory and g as output. Adding S to the input gate creates a CLSTM layer modified as shown in equation (7).

$$j_s = \sigma(X_{wj} w_s + X_{gj} g_{(s-1)} + X_{dj} d_{(s-1)} + a_s + X_{Sj} S)$$

$$e_s = \sigma(X_{we} w_s + X_{ge} g_{(s-1)} + X_{de} d_{(s-1)} + a_e + X_{Sj} S)$$

$$p_s = \sigma(X_{wp} w_s + X_{gp} g_{(s-1)} + X_{dp} d_{(s-1)} + a_p + X_{Sj} S)$$

$$d_s = e_s \cdot d_{(s-1)} + j_s \cdot \tanh(X_{wf} w_s + X_{gd} g_{(s-1)} + a_d + X_{Sj} S)$$

$$g_s = p_s \cdot \tanh(\tilde{f}_0)(d_s) \tag{7}$$

Figure 4 depicts the cognitive load level of LSTM architectures used in this research along with two distinct weighting schemes: logistic regression weights and the weighting system that we propose liency-based weights.

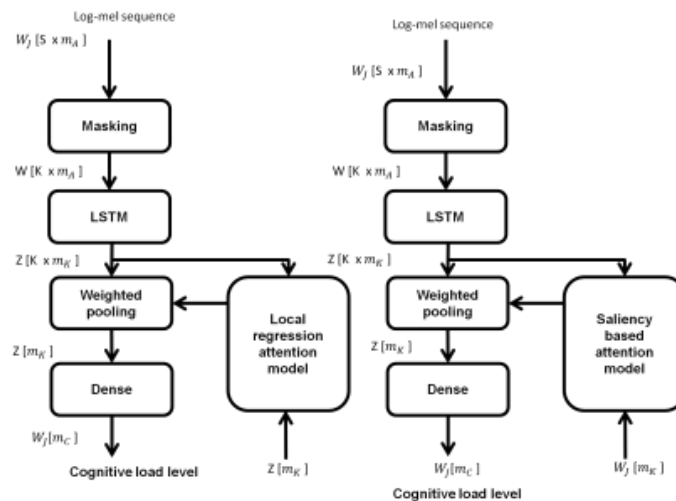


Figure 4: Cognitive load level

ALSTMs get sequences from CNNs. After the embedding layer, we can simply add a Conv1D and max pooling layer that feed the ALSTM layer consolidated features. The Conv1D and LSTM layers have the same hyper parameters as the previous models. After two deep layers, a sigmoid activation layer produces the final output. Convolutional ALSTM is a variant of ALSTM with a peephole connection that enables the gates to use the previous memory cell content even when the output gate is closed. When the output gate is closed, the earlier memory cells continue to have an effect because each gate accepts an input as an extra parameter, the prior memory content. Even when there is a lengthy chain of information, this relationship ensures that the earliest input will have an effect. The following description applies to each of the Conv-ALSTM's four gates.

$$e_s = \sigma_h(w_e^* w_s + v_e^* g_{s-1} + u_e \times d_{s-1} + a_e)$$

$$j_s = \sigma_h(w_e^* w_s + v_e^* g_{s-1} + u_e \times d_{s-1} + a_j)$$

$$\begin{aligned}
 d_s &= e_s \times d_{s-1} + j_s \times \sigma_g(w_d^*w_s + v_d^*g_{s-1} + a_d) \\
 o_s &= \sigma_h(w_p^*w_s + v_e^*g_{s-1} + u_e \times d_{s-1} + a_p) \\
 g_s &= p_s \times \sigma_g(d_s)
 \end{aligned}
 \tag{8}$$

With the ALSTM network connected to the same output layer, this network reads the input sequence forward and backward. One layer of an ALSTM processes the input in a single direction, while the second layer processes the sequence in reverse. The C3-ALSTM method predicts Tamil movie review sentiment. Its capacity to collect rich contextual information and long-range relationships, which helps to determine emotion in dynamic Tamil movie reviews.

4. PERFORMANCE ANALYSIS

In this section, Python is a programming language that is used for the evaluation of the results. Python was used to improve the C3-ALSTM algorithm that was previously used to build. Considering the earlier discussion, we built a new method C3-ALSTM to discover better performance of the accuracy, precision, recall and f-measure for comparing the K-nearest neighbor (KNN) [20], random forest (RF) [20] and support vector machine (SVM) [20], are the existing approaches.

4.1. ACCURACY

A classification model of accuracy can be evaluated based on the proportion of its overall predictions that turn out to be accurate. It is determined by taking the ratio of the number of occurrences that were properly predicted to the total number of instances included in the dataset. SA accuracy is the percentage of valid sentiment predictions a model makes. It's a frequent sentiment analysis model performance indicator. Figure 5 and Table 1 illustrate the accuracy comparison between the existing and recommended approaches. Compared to the KNN, RF and SVM approaches, which achieved scores of 86.2%, 83.8% and 79.12% respectively, the C3-ALSTM strategy demonstrated the highest score of 97.3%. It shows that our proposed method is superior to the existing methods. Equation (9) predicts the accuracy level.

$$Accuracy = (TP + TN) / (TP + TN + FP + FN)
 \tag{9}$$

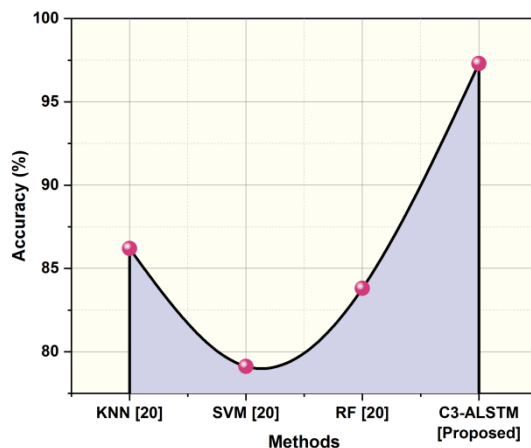


Figure 5: Comparisons of Accuracy

Table 1: Numerical Outcomes of Accuracy

Methods	Accuracy (%)
KNN [20]	86.2
SVM [20]	79.12
RF [20]	83.8
C3-ALSTM [Proposed]	97.3

4.2. PRECISION

Precision is a statistic used in evaluating the performance of a classification model, specifically in the context of binary classification issues. It assesses the actual accuracy of the optimistic projections made by the model. The precision can be determined with the use of the following equation (10). In the context of performing SA on Tamil reviews, precision is ascertaining what fraction of the predicted quantity of positive emotions really exists. Figure 6 and Table 2 illustrate the comparison of accuracy between the existing and recommended approach. In comparison to the KNN, SVM and RF approaches achieved scores of 81.3%, 78.8% and 88.2%, respectively, yet the C3-ALSTM strategy demonstrated a highest score of 97.6%. It shows that our proposed method is superior to the other existing methods.

$$Precision = TP / (TP + FP) \tag{10}$$

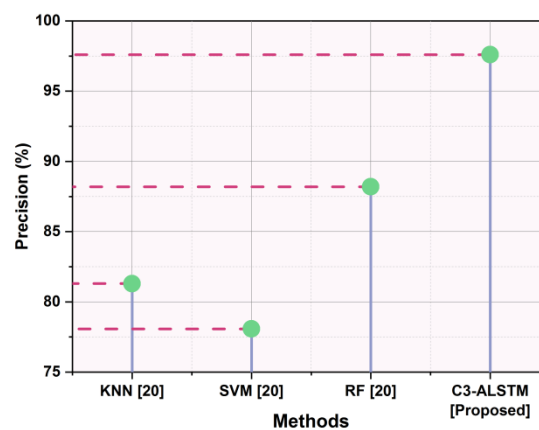


Figure 6: Comparisons of Precision

Table 2: Numerical Outcomes of Precision

Methods	Precision (%)
KNN [20]	81.3
SVM [20]	78.08
RF [20]	88.2
C3-ALSTM [Proposed]	97.6

4.3. RECALL

Recall refers to the proportion of relevant sentences that are recovered from memory compared to the total quantity of phrases in the database. The recall of equation (11) is used to determine the accuracy of the suggested text summarization approach. Figure 7 and Table 3 illustrate the comparison of accuracy between the existing and recommended method. In contrast to the KNN, SVM and RF approaches, which achieved 92.5%, 72.15% and 85%, respectively, the C3-ALSTM strategy demonstrated the highest score of 97.2%. It shows that our proposed method is superior to the other existing methods.

$$Recall = TP / (TP + FN) \tag{11}$$

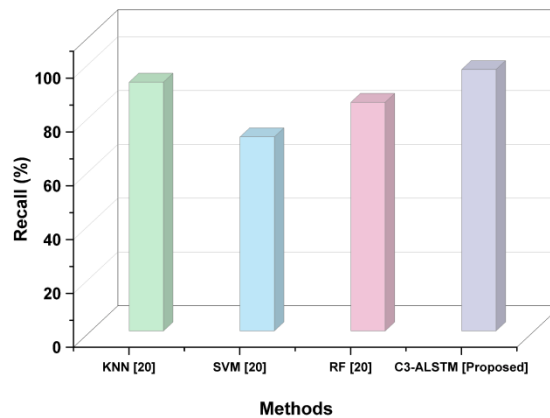


Figure 7: Comparisons of Recall

Table 3: Numerical Outcomes of Recall

Methods	Recall (%)
KNN [20]	92.5
SVM [20]	72.15
RF [20]	85
C3-ALSTM [Proposed]	97.2

4.4. F-MEASURE

To calculate the F-measure value for the whole dataset, the accuracy and recall values are considered. The F-measure can be described as follows in equation (12). Figure 8 and Table 4 illustrate the comparison of accuracy between the existing and recommended approach. Compared to the KNN, RF and SVM approaches, which achieved 86.5%, 74.9% and 91.3%, respectively, the C3-ALSTM strategy demonstrated the highest score of 98.0%. It shows that our proposed method is superior to the other existing methods.

$$F - Measure = 2 \times (Precision \times Recall) / (Precision + Recall) \quad (12)$$

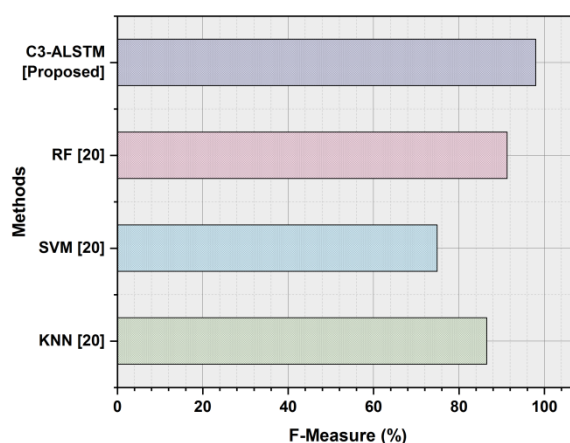


Figure 8: F-Measure Comparisons

Table 4: Numerical outcomes of F-Measure

Methods	F-Measure
KNN [20]	86.5
SVM [20]	74.9
RF [20]	91.3
C3-ALSTM [Proposed]	98

5. CONCLUSION

The research presents a unique method for assessing Tamil movie reviews taken from Tamil tweets: contextual-cognitive convolutional attentive long/short-term memory (C3-ALSTM). A Tamil movie review dataset is obtained from an open database as part of the approach. A thorough data pretreatment step is conducted to remove noise, duplicate data and unnecessary information. Standardizing to normalize the collected data, methods such as tokenization, stop word removal and special character handling are used. TF-IDF is used in the feature extraction process to find relevant features for the study. Utilizing these characteristics, the suggested C3-ALSTM method assesses Tamil reviews' sentiments. The technique is implemented on the Python platform and important performance measures, including accuracy (97.3%), precision (97.6%), recall (97.2%) and f-measure (98%), are used to assess the strategy's effectiveness. The results show that the proposed C3-ALSTM approach works better

than other widely used algorithms for sentiment analysis of Tamil reviews. The suggested approach's performance shows that it is superior to existing systems and can be useful for evaluating feelings in the context of Tamil movie reviews. SA of Tamil movie reviews is projected to improve further using cutting-edge models like C3-ALSTM in the future, increasing their precision in understanding the complex feelings communicated by this dynamic film industry.

REFERENCES

- [1] Zunic, A., Corcoran, P. and Spasic, I., 2020. Sentiment analysis in health and well-being: systematic review. *JMIR medical informatics*, 8(1), p.e16023.
- [2] Anandhan, A., Shuib, L., Ismail, M.A. and Mujtaba, G., 2018. Social media recommender systems: review and open research issues. *IEEE Access*, 6, pp.15608-15628.
- [3] Krishnan, I.A., Ching, H.S., Ramalingam, S., Maruthai, E., Kandasamy, P., De Mello, G., Munian, S. and Ling, W.W., 2020. Challenges of learning English in the 21st century: Online vs. traditional during Covid-19. *Malaysian Journal of Social Sciences and Humanities (MJSSH)*, 5(9), pp.1-15.
- [4] Sima, V., Gheorghe, I.G., Subić, J. and Nancu, D., 2020. A systematic review of the influences of the Industry 4.0 revolution on human capital development and consumer behavior. *Sustainability*, 12(10), p.4035.
- [5] Gutiérrez-Martín, A. and Torrego-González, A., 2018. The Twitter games: media education, popular culture and multiscreen viewing in virtual concourses. *Information, Communication & Society*, 21(3), pp.434-447.
- [6] Reyes-Menendez, A., Saura, J.R. and Alvarez-Alonso, C., 2018. Understanding#WorldEnvironmentDay user opinions in Twitter: A topic-based sentiment analysis approach. *International journal of environmental research and public health*, 15(11), p.2537.
- [7] Durriyah, T.L. and Zuhdi, M., 2018. Digital Literacy with EFL Student Teachers: Exploring Indonesian Student Teachers' Initial Perception about Integrating Digital Technologies into a Teaching Unit. *International Journal of Education and Literacy Studies*, 6(3), pp.53-60.
- [8] Dreisbach, C., Koleck, T.A., Bourne, P.E. and Bakken, S., 2019. A systematic review of natural language processing and text mining of symptoms from electronic patient-authored text data. *International journal of medical informatics*, 125, pp.37-46.
- [9] Nag, S. and Narayanan, B., 2019. Orthographic knowledge, reading, and spelling development in Tamil: The first three years. *Handbook of literacy in Akshara orthography*, pp.55-83.
- [10] Boruah, D.M., 2020. Language Loss and Revitalization of Gondi language: An Endangered Language of Central India. *Language in India*, 20(9).
- [11] Thavareesan, S. and Mahesan, S., 2019, December. Sentiment analysis in Tamil texts: A study on machine learning techniques and feature representation. In *2019 14th Conference on Industrial and Information Systems (ICIIS)* (pp. 320-325). IEEE.
- [12] Rani, S. and Kumar, P., 2019. A journey of Indian languages over sentiment analysis: a systematic review. *Artificial Intelligence Review*, 52, pp.1415-1462.
- [13] Ahmad, G.I., Singla, J. and Nikita, N., 2019, April. Review on sentiment analysis of Indian languages with a special focus on code mixed Indian languages. In *2019 International Conference on Automation, Computational and Technology Management (ICACTM)* (pp. 352-356). IEEE.
- [14] Soman, S.J., Swaminathan, P., Anandan, R. and Kalaivani, K., 2018. A comparative review of the challenges encountered in sentiment analysis of Indian regional language tweets vs English language tweets. *International Journal of Engineering & Technology*, 7(2), pp.319-322.
- [15] Shelke, M.B. and Deshmukh, S.N., 2020. Recent advances in sentiment analysis of Indian languages. *International Journal of Future Generation Communication and Networking*, 13(4), pp.1656-1675.

- [16] Bhargava, R., Arora, S. and Sharma, Y., 2019. Neural network-based architecture for sentiment analysis in Indian languages. *Journal of Intelligent Systems*, 28(3), pp.361-375.
- [17] Lakshmi Devi, B., Varaswathi Bai, V., Ramasubbareddy, S. and Govinda, K., 2020. Sentiment analysis on movie reviews. In *Emerging Research in Data Engineering Systems and Computer Communications: Proceedings of CCODE 2019* (pp. 321-328). Springer Singapore.
- [18] Ravishankar, N. and Shriram, R., 2018. Grammar rule-based sentiment categorization model for classification of Tamil tweets. *International Journal of Intelligent Systems Technologies and Applications*, 17(1-2), pp.89-97.
- [19] Shelke, M.B. and Deshmukh, S.N., 2020. Recent advances in sentiment analysis of Indian languages. *International Journal of Future Generation Communication and Networking*, 13(4), pp.1656-1675.
- [20] Sunitha, P.B., Joseph, S. and Akhil, P.V., 2019, October. A study on the performance of supervised algorithms for classification in sentiment analysis. In *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)* (pp. 1351-1356). IEEE.