International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7216
***Article Received**: 29 April 2023 **Revised:** 16 June 2023 **Accepted:** 29 June 2023*

Exploiting Emotions via Composite Pretrained Embedding and Ensemble Language Model

Shila S. Jawale[1], Dr. S. D. Sawarkar[2]
[1]Department of Computer Engineeringline
Datta Meghe college of Engineering
Airoli, Navi Mumbai,Mahrashtra,India
shilaph@gmail.com
[2]Department of Computer Engineeringline
Datta Meghe college of Engineering
Airoli, Navi Mumbai,Mahrashtra,India
sudhir_sawarkar @yahoo.com

Abstract—Decisions in the modern era are based on more than just the available data; they also incorporate feedback from online sources. Processing reviews known as Sentiment analysis (SA) or Emotion analysis. Understanding the user's perspective and routines is crucial now-a-days for multiple reasons. It is used by both businesses and governments to make strategic decisions. Various architectural and vector embedding strategies have been developed for SA processing. Accurate representation of text is crucial for automatic SA. Due to the large number of languages spoken and written, polysemy and syntactic or semantic issues were common. To get around these problems, we developed effective composite embedding (ECE), a method that combines the advantages of vector embedding techniques that are either context-independent (like glove & fasttext) or context-aware (like XLNet) to effectively represent the features needed for processing. To improve the performace towards emotion or sentiment we proposed stacked ensemble model of deep lanugae models.ECE with Ensembled model is evaluated on balanced dataset to prove that it is a reliable embedding technique and a generalised model for SA.In order to evaluate ECE, cutting-edge ML and Deep net language models are deployed and comapared. The model is evaluated using benchmark datset such as MR, Kindle along with realtime tweet dataset of user complaints . LIME is used to verify the model's predictions and to provide statistical results for sentence.The model with ECE embedding provides state-of-art results with real time dataset as well.

Keywords- Sentiment Analysis, Glove, Fasttext, XLNet, GRU, LSTM, LIME, MR, CNN, Kindle,real time complaint tweet dataset

I. INTRODUCTION

Internet technology has had profound effects on people's daily lives in the modern information age [1]. Every service and product we use regularly in our daily lives has review comments available on the Internet. As time goes on, the opinions expressed in customer reviews grow more useful to examine and expand a business [2]. As much as it benefits retailers to have insight about products' quality and flaws from customers' perspectives, there are downsides to this method of gathering information. On the other hand, it can be challenging to manually complete the review reports due to the sheer volume and diversity of information included in each report. Sentiment analysis (also called opinion mining or Intent Identification) is a technique in natural language processing (NLP) that uses text analysis and computational linguistics to identify, extract, and categories subjective information from unstructured text [3]. The ultimate purpose of determining user sentiment (positive, negative) from written feedback [4]. An important tool in many fields, sentiment analysis [16].

There are a number of language issues that affect automatic sentiment analysis's effectiveness, including word sparsity, polysemy, feature divergence, and polarity divergence [20],[21]. From early sentiment dictionary-based algorithms [5] to contemporary machine learning techniques like Support Vector Machine (SVM), Naive Bayes (NB), decision tree (DT), random forest (RF), logistic regression, etc. [6], [7], many methods have been developed and refined over time to address the aforementioned problems. Although statistical machine learning techniques perform well in simple applications of sentiment analysis [8], [9], these algorithms cannot be scaled to address more complex text categorization issues. The effectiveness of these techniques relies heavily on feature representation, and it is difficult to achieve satisfactory classification results [10],[11]. As interest in deep learning has grown, a growing number of researchers have applied various deep learning techniques to the problem of. In computer vision [12], opinion analysis [13], speech recognition [14], sentiment classification [15] and the models have shown to yield particularly impressive results. In contrast to more conventional machine learning methods, deep learning can function without the input of any labelled data or manually extracted features. However, the depth and generalizability of a model in deep learning are significantly affected by the availability of training data [17].

II. BACKGROUND AND RELATED WORK

Deep learning employs artificial neural networks to analyze and predict information from large and intricate datasets. It trains deep neural networks with multiple layers to automatically learn hierarchical representations and extract meaningful features from raw data. To represent these features, vector embedding techniques are utilized. Vector embedding, or word embedding, involves converting individual words in a text into numerical values and vectors. This process generates word vector representations that are semantically equivalent and internally consistent. [20],[21] There are two main categories of word embedding methods: probabilistic prediction and count-based approaches. Probabilistic prediction methods, such as Word2Vec [22], Skip-gram [23], and continuous bag of words (CBOW) [24], use a corpus for model training. CBOW predicts unknown words by employing skip-gram, which predicts neighboring words based on the given word. Count-based approaches learn vectors through a co-occurrence frequency matrix, with the GloVe [25] model being a well-known technique that uses the ratio of a word's co-occurrence probability to store its meaning. GloVe and Word2Vec models efficiently generate word vectors for tasks like word similarity. The character-based models such as FastText [26] and ELMO [27] also very popular. Transformer-based models [28], including BERT [29], ALBERT [30], XLNet [31], and GPT2 [32], have been developed to enhance contextualized word representations. These models capture the context and effectively represent word relationships. However, word2vec loses the linearity of word relationships, and it struggles to process words that are not in the dictionary [33][34]. While models like BERT and GPT2 can learn from cross-reference data, BERT lacks the ability to predict the next word based on previous words, which is crucial for considering contextual information [28],[33].

Majority of the time performance of sentiment analysis tasks is degraded due to a variety of reasons such as pre-processing techniques used, language construct, embedding techniques used, model architecture, type of dataset, number of samples available in dataset along with the computational resources. Out of this the embedding techniques plays a vital role in terms of representing the vector form of available text. Along with the language construct also matters such as syntax and semantic meaning of the available text [34], presence of polysemy words and handling out of vocabulary (OOV) words [35]. Several prior studies have proposed improving sentiment analysis performance by encoding sentiment knowledge into word embeddings [36].

Extensive research is the backbone of any reliable, objective, and compelling conclusion. Why not use a highly reliable and well-explained machine learning model that was developed by humans? It can be used either after the seeker has received a choice, as part of the model, or before any data is fed into the model, depending on the system and the seeker's needs. Decisions regarding categorization, forecasting, and aggregation might be justified. Justification of the model could occur either after or during the process. A task's explanation can be given either globally or locally. Criteria such as feature importance, the justification of a minor or major class, etc. To communicate with humans, it is normal practice to employ visual or graphical aids. [16],[37],[38],[39].

The limits of many NLP tasks have been pushed by pre-trained language models. This frequently comes at the expense of extraordinarily complex structures with hundreds of millions or billions of parameters that cost a lot of money to develop and implement. Because they lack access to a vast pool of resources, institutions, small and medium-sized businesses, and individual researchers face this obstacle, which effectively prevents the democratization of cutting-edge deep learning research [14,18,19,20,40,41].

In order to find the effective solution related with sentiment analysis with deep learning by taking care of low computational resource setting the following research questions has been decided.

RQ1. Finding the optimum approach to create the embedding that take care of syntactical, OOV words along with represent the context too.

RQ2.How it will perform with low computational resources in terms of model’s architecture requirements?

RQ3.Can we able to justify the interpret and decision given by the model and how?

To answer the research question stated above the paper is arranges in multiple section such as RQ1 is solved in section 2. RQ2, RQ3 is solved in section 3 and section 4 respectively. The sections described as section 2. Background and related work, Section 3. Methodology, Section 4. Result and discussion, Section 5. future work followed by Conclusion.

The reviews' emotions were analyzed using an approach called Effective Composite Embedding (ECE), which is based three different embedding such as glove, fasttext and XLNet. ECE represents the word embedding in turn feature representational layer. The further the sentiment categorization done with the help of novel language model built from stacked

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7216
***Article Received**: 29 April 2023 **Revised:** 16 June 2023 **Accepted:** 29 June 2023*

ensemble approach [42]. Ensemble model has better performance as compared to other models [43] which is helpful in getting low-cost solution with comparable performance that of transformer-based architectures [44].

As a result of the advancement and development of deep learning methods, a variety of structured deep learning networks have been presented for diverse applications and designs with respect to sentiment analysis tasks, as shown in Table 1. The contributions of can be summarized in table below

TABLE 1.SUMMRAIZED WORK

Author	**Methodology**	**Application coved**
Abdalraouf Hassan and Ausif Mamhmood [45]	Sentence modelling uses bag-of- words (BoW) along with CNN+LSTM(RNN)model foe sentence classification	sentence classification
Joachims, T [46]	sentence modelling is N-gram with the help of SVM to resolve data sparsity issues.	text categorization
J. Panthati et al.[47]	word2vec and CNN on pre processing of the input datasets, they are fed to word2vec to generate word embeddings for processing massive amount of review.	Amazon product review
Z. Z. Wint et al.[48]	Global Vectors for Word Representation (GloVe) word vector and deep convolution neural network to uses co-occurrence probability rather than the probability to learn word vector	The Stanford Twitter Sentiment Test (STSTd) data set, SemEval 2014 dataset, the Stanford Twitter Sentiment Gold dataset, The Sentiment Evaluation Twitter dataset and The Sentiment Strength Twitter dataset
Huy Nguyen et al.[49]	Deep CNN +Bi(Bidirectional)-LSTM" character-level embeddings to increase the information for word embeddings , tweet processor to remove the non-essential words from the tweet however retaining the emoticons and necessary information."	Twitter,Sanders and HCR datasets
Jianqiang, Z et.al [50]	introduce a word embeddings method obtained by unsupervised learning based on large twitter corpora, this method using latent contextual semantic relationships and co-occurrence statistical characteristics between words in tweets	Twitter data
Wei Xue, Tao Li [51]	Gated Tanh-ReLU units GTRU saves time as computation, Simple model	product and service review SemEval workshop on laptop and restaurant
Z. Jianqiang et al.[52]	hybrid deep learning architecture called Hybrid two Convolutional Neural Networks and Bidirectional LSTM (H2CBi) 1) its lack of semantic understanding 2) inability to understand misspelled words. Rectified	Amazon, movie review and Yelp and four social network site data from Twitter I and II, Facebook and FormSpring.me
A. Onaciu and A. Nicoleta Marginean [53]	ensemble of classifiers using deep learning strategies aspect detection and polarity identification done aspect is denoted as a tuple of <ENTITY,ATTRIBUTE>. Concept Net is used	SemEval-2016 a restaurant based review dataset
Ganaie, M. A et al.[54]	Provided in depth review on ensemble models such as bagging, boosting , implicit/explicit ensembles, heterogeneous ensembles.	-
N. Aslam et al.[55]	ensemble model LSTM-GRU designed that combines two recurrent neural networks applications including long short term memory (LSTM) and gated recurrent unit (GRU). LSTM and GRU are stacked where the GRU is trained on the features	Cryptocurrency Related Tweets

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7216
***Article Received**: 29 April 2023 **Revised:** 16 June 2023 **Accepted:** 29 June 2023*

	extracted by LSTM. They used tf-idf, word2vec, and bag of words (BoW) as embedding	

The table shows the different application a covered in sentiment analysis along with the different types of embedding used to represent text corpus in vector form. The study not limited to the as shown in Table 1. But it goes beyond that such as study of pretrained embedding, language models, computational requirement etc. with respect to that in [36] author gives us a highlight on semantic word embeddings that they unable to capture sentiment. For example, despite being opposite in attitude, 'right' and 'wrong' are employed in comparable settings, they are vector-space related. In similar settings, right and wrong are assigned similar linguistic representations [34].

With respect to modern architecture that can do better sentiment analysis by keeping the context to such as BERT and GPT etc. and their different approaches and architecture proposed by researchers too [19].

The automatic sentiment analysis processing implemented by different researchers by considering the pretrained word embedding gives us the idea and applicability on varied domain of natural language processing. The different embedding and their pros and cons described by [56]. Also, author described the advantages and limitation of each type of embedding. Due to variety of vector embedding strategy available.[57],[58] The embedding can be applied to the model in two different ways as 1. Train the model with respective embedding 2. use the existing pretrained embedding representation. The 1st approach takes a larger time in terms of creating the embedding representation [28] but in 2nd approach the training times saved as well as model could be more generalized and can be used for multiple types of problem to be sentiment related problem to be solved.

To make optimum use of available vector embedding techniques becomes an need and motivates us to combine them and have a good representation to handle the sentiment analysis in better way. The combined approach improves the performances on downstream tasks [19]. The various study and approaches described by different researchers.

In order to solve the problem of sentiment analysis in low resource setting it is better to go with 2nd approach of using the existing pretrained strategy. But to the different implementation strategy their representation ins different and each has different way to tackle language constraints such as syntax and semantics, handling Out-of-vocabulary (OOV) words and ambiguity and polysemy words.

The in-depth study of various embedding and how to combine them to improve the model performance discussed in [59]. The author experimentally demonstrated the ways to combine the different types of embedding. The method such as concatenation, PCE, AutoE were discussed with different kind of NLP tasks.

In their work, [60] introduced an ensemble model that integrates three hybrid deep learning models: RoBERTa, LSTM, BiLSTM, and GRU, to tackle sentiment analysis. RoBERTa is utilized to represent features, while LSTM, BiLSTM, and GRU capture dependencies within the embeddings based on the sentiment class. By combining the predictions from these models using averaging ensemble and majority voting techniques, the overall performance of sentiment analysis is improved. To address imbalanced datasets, the authors also employ data augmentation with pre-trained word embeddings from GloVe.

The work such as [57] summarized the different word embedding techniques together and applied on citation data. They got very good result with BERT. The experimental results proved that BERT based encoding provides the good result as compared to glove and other embedding techniques. Author also suggested the limitation of BERT as per as computational cost is concern.

The fasttext model perform better in case of Malware Detection and Classification proved by [61]. The author applied this on API call sequence information of the malware. They tested on both fasttext and BERT as pretrained embedding. Given us the thought towards absence of domain information in BERT too.

To solve long Chinese text sentiment problem [62] proposed an ensemble model based on BERT. The approach involves using BERT-based models as the foundation for classifiers to extract partial sentiments. Additionally, a BiGRU network is employed to capture the intricate and varying structures found in Chinese articles. The results from multiple BERT-based models are combined, and an attention layer is utilized to amplify the central sentiments within lengthy Chinese texts.

In [63] author used fasttext as embedding with first model single-layered Bidirectional Gated Recurrent Unit (BiGRU) along with second model designed from single-layered Convolutional Neural Network (CNN) model for sentiment analysis and got very comparative result.

In [64] in their proposed approach provides an ensemble model based on SVM as base learner and Adaboost as the Ensemble Boosting algorithm.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7216
***Article Received**: 29 April 2023 **Revised:** 16 June 2023 **Accepted:** 29 June 2023*

A proposal for a contextual sentiment embedding (CoSE) can be found in [65]. This paper presents a two-layer GRU model as a language model, trained concurrently to combine semantic and sentiment information from labelled corpora and lexicons, in response to the high need of computational like transformer architecture. To deal with OOV or informal-writing sentiment words, the WordPiece tokenizer was used to break down the text into token. They created embedding as a combination of external knowledge sentiment data, pretrained-embedding and random initialized vector.

With Static Character and Word Embeddings for Arabic Tweets, [66] builds Contextualized Language Models for Emotional Severity and Sentiment Intensity Detection. They combined character embedding to solve OOV problems and static embedding to improve performance of Arabic tweet.

To contribute to this study further improvement is needed in terms of creating the more simpler and effective embedding approach that can be served to language model for sentiment analysis tasks, which can be perform well in low resource setting and can justify the model decision.

III. METHODOLOGY

We present an approach that combines character and word-based embeddings (fasttext, glove) with dynamic context-based embeddings (XLNet). Following a similar strategy as proposed in [63], we merge these embedding. Syntax and semantics are handled with static word-level embeddings (glove), the OOV problem is conquered with fasttext, and the context is preserved with XLNet. Fig. 1 provides an overview of our proposed method. In the following section explains the concept and how our strategy combines them. This section addresses the research question RQ2 with proposed ECE embedding and Ensemble model designing with performance evaluation.

A Preliminary treatment for text:

The raw corpus is polluted with unwanted and useful token. Such as special character, repeated words, digits, stop words, hashtags, email-id, repeated numbers in short noise. These creates the overhead of processing. To tackle this text pre-processing is an important and crucial step. So, we applied all necessary processing such to remove not needed character such as number, punctuation mark, emoji, stop words. To represent in meaningful way the contradiction also applied with their long form such as didn't as did not etc. This improves the context handling capabilities of the model

B Static word embedding:

The static embedding used for the experimentation purpose is glove and fast due to their word representation property. They can handle syntax and semantical problems [67] very perfectly in the vector form of words and characters

GloVe: GloVe, or Global Vectors [25] for Word Representation, was created by a Stanford group led by Pennington et al. It is a learning algorithm for extracting word vector representations without human supervision. Matrix factorization methods, including Latent Semantic Analysis, were used to create traditional vector space model representations of words (LSA). Seven distinct word embedding models have already been trained, each with its own unique set of parameters for token size, vectors, and vocabularies. We tried them all, however a model trained on 840 billion tokens in the validation set produced the best results. Glove has 7 different representations [68].

Fasttext: FastText [26] expands on the Skip-gram model by employing n-grams of characters to improve word representations. The vector representation of a word is built by adding up the vectors associated with each n-gram character that appears in the word. This characteristic enables it to learn words that are not just uncommon but also unfamiliar that also called as Out-of-vocabulary (OOV words).

C Context aware Embedding :

These embedding have a capability to represent the textual representation by keeping the context preserved. These are sentence vector and can able to effectively represent the context information effectively. Also called as transformer-based model trained on huge cross reference data from different sources.

XLNet [31]: is a recent transformer-based language representation model that was developed using extensive cross-domain data. XLNet and BERT uses a masked language model to predict words that are randomly masked in a sequence and then uses the task of predicting the next sentence to learn the associations between sentences. XLNet and BERT [33] both are under transformer-based architecture. Due to model trading done on large corpus, the model can identify the distinction of meaning with varying context. XLNet is an " Auto-Regressive " model (AR). Here the model must predict randomly masked language tokens.

When it comes to modelling the mutual dependence of the XLNet succeeds where BERT falls short. The example shows that BERT can learn some dependency pairs, like (Pali, Hill) and (Hill, known), but XLNet always learns more pairings and has "denser" effective training signals for the same targets.

D Combining XLNet and Static embedding :

The ECE (Effective composite embedding) is created from static and contextual embedding (generated from XLNet). Here

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7216
***Article Received**: 29 April 2023 **Revised:** 16 June 2023 **Accepted:** 29 June 2023*

we combined them by creating their corresponding vector representation generated as Mean vector for static embedding and Normalized vector from XLNet. Combined them with concat strategy given in [65]. To reduce the dimensionality further we applied SVD (Singular value Decomposition) [69] technique.

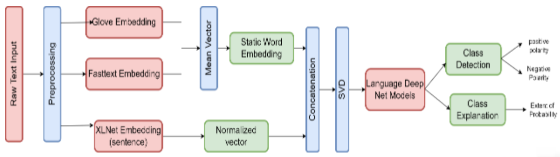

Figure 1. ECE and Language deep Net architecture for SA

As shown above, each embedding approach has unique characteristics, thus it is crucial to understand what feature vectors will be supplied by utilizing them separately or in combination.

The embedding layer, represented by e in the sentiment classification workflow, applies one or more pre-trained embeddings to produce word representations for the classifier.

As we used three different embedding each represented a single sentence at three different ways.

The sentences S can be represented with glove as Sg={st1,st2,st3……..stn} , Similar way for S can be represented by fasttext as Sf={sf1,sf2,sf3……sfn}.For the XLNet each sentence is represented with default 768 dimension. But by creating the separate embedding via tokenizer and encoding we can make it custom dimensional. Thus created sentence embedding as Sxl={sx1…...sxn} represented with max dimension of 300 so as to make equal representation. Due to different encoding representation of XLNet the L1 normalization is applied with XLNet sentence vector with (1)

$$S_{xlnorm} = \frac{S_{xl}}{|S_{xl}|} \quad (1)$$

The glove sentence vector as Sg and fasttext sentence vector as Sf combined with the following equation as mean vector so reduce make sure every word in sentence must be encoded that also reduce the problem of OOV along with higher weightage for the context vector with the help of (2) The combination is called as Combined Sentence embedding (CSemb).

$$CS_{emb} = Mean(S_g, S_f) \quad (2)$$

Now to create the ECE, static embedding obtained from (1) and (2) is is concatenated and form the final vector representation as

$$C_{emb=}S_{xlnorm} \oplus CS_{emb} \quad (3)$$

To reduce the sparcity the final vector length adjusted depends on the max length of sentences truncated to 300 vector length. So, to reduce the dimensionality technique of vector we applied the SVD to reduce the shape till n where n is required dimension as per length of sentences max up to 300 dimensions.

$$ECE = SVD(CS_{emb}, n) \quad (4)$$

The (4) Represents the final matrix called as Effective Composite Embedding (ECE) as making assured of keeping a context available in the sentence.

The created ECE is a vector representation at sentence level and further given to various Deep learning and state-of-art ML model.

These method used summarize the embedded vectors x0, x1, x2, xt,...xn into a single sequence representation, where o = e (x0, x1,..., xm). After ECE encoding, o is sent to a fully connected layer, represented by the letter f, to output the logits cross all labels: g = f (o).

The probability of sample s belonging to label li for binary classification is calculated as p (l_i | s) by

$$P(l_{i|}|s) = \frac{\exp{(g_i)}}{\sum_{j=0}^{T} \exp{(g_i)}} \quad (5)$$

where T refers to the number of labels. For binary sentiment classification, the probability is estimated by the sigmoid function

$$P(l_{i|}|s) = sigmoid(g_i) \quad (6)$$

Where the label l_i is predicted for the training example *s* if the estimated probability is larger than 0.5.

E Deep Learning Model :

To get the better performance the language model the basic CNN (convolutional neural network) and RNN (Recurrent Neural Network) described below and tested with different architectures.

a) Convolutional artificial neural network (CNN)

CNNs are multi-layered feed-forward neural network models used in deep learning [68],[76]. CNN uses spatial characteristics. CNN's image classification is improved by associated two-pixel values. This network topology is employed in text classification [68]. Convolution, pooling, and fully connected layers constitute CNN. A convolution layer automatically extracts features, while a pooling layer reduces them. a CNN representation that uses text as input to identify the text's class. Using filters, the convolution layer extracts picture or text features. Rectified Linear Unit activation (ReLU) function is used between the convolution and pooling layers to make features non-linear. In the pooling layer, feature map dimensions are lowered, which reduces computational cost in future layers and displays image or text features more

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7216
***Article Received**: 29 April 2023 **Revised:** 16 June 2023 **Accepted:** 29 June 2023*

efficiently. The convolutional neural network's last layer is a fully connected ANN. In this layer, artificial neurons represent image/text features and target class labels. The new text serves as an input to the CNN.CNN calculates class probability after training.

b) GRU: Gated Recurrent Unit

In particular, RNN can extract relevant features from data and may therefore be a popular option for capturing the semantics of lengthy texts. RNNs are biassed models because t hey place more weight on words that have appearedlately in a s equence, which can hinder their ability to accurately capture th e gist of a document's meaning [77]. To address these problems with long-term dependencies in RNN models, LSTM [78] was developed. BiLSTM, an extension of LSTM, consists of two LSTM cells: a forward LSTM cell and a backward LSTM cell [79]. When attempting to capture a text's emotion, LSTM looks at all the preceding words, while BiLSTM looks at both the preceding and subsequent words. GRU is smallest variation from LSTM that has two gates, an update gate r that combines forget and input gates, and a reset gate z. The related equation those need for computation using LSTM and GRU given below

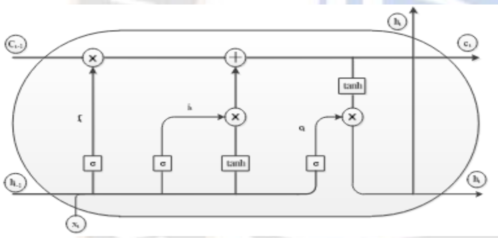

Figure 2 . A typical LSTM Unit

The LSTM model comprises repeating temporal modules with three gates: forget gate, input gate, output gate, and memory unit. The formula**.**
Forget gate:

$$f_t = \sigma(w_f . x_t + U_f . h_{t-1} + b_f) \quad (7)$$

Input gate:

$$i_t = \sigma(w_i . x_t + U_i . h_{t-1} + b_i) \quad (8)$$

Transformation: means that candidate memory cell status at the current time step

$$\hat{c}_t = tanh(w_c . x_t + U_c . h_{t-1} + b_c) \quad (9)$$

State update: the state value c_t of the current time in the memory cell.

$$c_t = i_t \odot \hat{c}_t + f_t \odot c_{t-1} \quad (10)$$

Output gate:

$$o_t = \sigma(w_o . x_o + U_o . h_{t-1} + b_o) \quad (11)$$

Hidden status:

$$h_t = o_t \odot \tanh(c_t) \quad (12)$$

Where U represents previous state , W_f , W_i , Wo , W_c are the weight of LSTM, b_f , b_i , b_o , b_c are the bias of LSTM, h_t is a hidden state of the t moment, σ is the activation function sigmoid, tanh is a hyperbolic tangent function. ⊙ denotes the elements calculated by multiplying point by point (Sukhbaatar et al. [70]).

Bidirectional LSTM improves regular LSTM and works well on NLP tasks. The bidirectional LSTM includes the forward and backward layers. Connect opposite-time-sequence LSTM networks to the same output. Forward LSTM may get sequence history. Backward LSTM can retrieve input sequence information. This framework considers past and future context. Stacking bidirectional LSTM layers. The first layer bidirectional LSTM output return sequence is the second layer bidirectional LSTM input, and the output is the concatenation of the forward and backward unit outputs. BiLSTM layers output h:

$$h = [h_{forward}, h_{backward}] \quad (13)$$

Based on the above equation of LSTM and CNN we proposed an ensemble model which is comprises of shallow LSTM, shallow GRU and mixed CNN and Bi-GRU language model. The MLP used as classifier. The following Figure 3. shows the proposed Ensemble model.

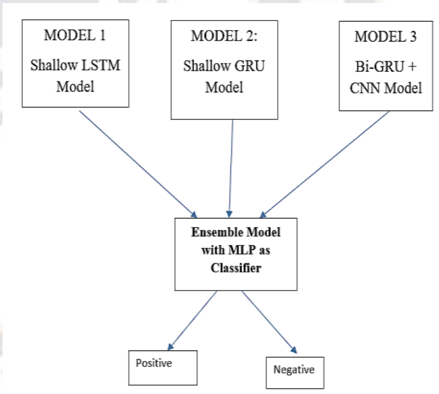

Figure 3. Proposed Ensemble Model

The Model 1 is implemented using single layer of LSTM along with dropout for controlling the overfitting of the network. The Model 2 is implemted with the help of smaller unit of LSTM family as GRU.It also implemented as shallow GRU model with dropout and Rmsprop as optimizer. The Model 3 is implemted with the help of combination of Bi-GRU and CNN layer. For CNN maxpooling layer is used to extract the important features. This further given to Bi-LSTM layer for further processing. The output of all these model has been stacked and given to Multilayer Perceptron(MLP) to decide the class of the sentence as positive or negative.The vector embedding ECE computed (4) further given to

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7216
***Article Received**: 29 April 2023 **Revised:** 16 June 2023 **Accepted:** 29 June 2023*

ensemble model. The multiple computation for CNN , BiLSTM, GRU done with the help of (7)-(13).The sentiment classiifcation with (5)-(6) .The different model parameter metioned in Table 4.

F. Experimental Specifications:

The datasets are selected, the downstream models are set up, and the pre-trained embeddings are chosen as part of the experiment setup. Below is a list of these components' specifications.

We chose three prominent classification benchmarking datasets to check generalization ability with ECE embedding. The size, number of classes, kinds, and categories of these datasets vary (topic or sentiment). Table 2 provides a summary of the datasets. The two are review dataset and other one is real time tweets on user complaint. The tweets have been labelled manually as positive or negative towards complaints.

MR: A benchmark dataset for sentiment classification [71]. Short movie review dataset with one sentence per review. Each review was labelled with their overall sentiment polarity (positive or negative). Overall, 5331 positive and 5331 negative reviews available. We used 10000 reviews.

Kindle: It consists of 982,619 product reviews and metadata obtained from Amazon regarding Kindle Store and used 3999 reviews on books for experimentation (He and McAuley) [72].

Realtime tweet Dataset: The tweet dataset corpus has been generated from twitter. The dataset has 10372 rows for balanced classes. The tweets have been recorded from the period of six months that is from June 2022 to Dec 2022. The tweets have been labelled as positive of negative complaints.

Out of all the available samples we chosen the following number of data sample for with consideration equal number of classes of each category representing balanced datasets. The training, testing ratio is kept as 80:20. Dataset representation as shown in Table 2.

TABLE 2. DATASET REPRESENTATION

Name of the Dataset	**Number of samples**	**Classes**	**Type**
MR	10000	binary	Movie review
Kindle	3999	binary	Books review
Real time tweet	10372	binary	Tweets

Pretrained embedding used:

Different variation of pretrained embedding available. From that following specific are used for experimentation. They listed below in Table 3. For XLNet does not need to much pre-processing so it can grasp the context with better way such as lower case etc. The tweet dataset has been preprocessed with removal of #, @, url along with text normalization followed by stemming and lemmatization is done.

TABLE 3. EMBEDDING SPECIFICATIONS

Name of Vector embedding	Specifications
Glove	glove.840B.300d.txt
FastText	wiki-news-300d-1M.vec
XLNet	XLNet-base-cased

a)Experimental setup and Hyperparameter Specification:

To do the ECE with SA experimental work Google Collaboratory is used as a platform with programming language as python. The different libraries used as Keras, Scikit-Learn, Matplotlib, NumPy. The data frame handling is done via Panda library. For the Deep net-based language model hyperparameter plays an important role such batch size, optimizer used, activation function used, loss computing function and to reduce the training parameters dropout strength. These hypermeters control the models in terms of overfitting or underfitting problems of deep net. Table 4 described same for experimental consideration. The regularization technique such as dropout rate [73] and Early stopping criteria [74] is used. The different model configuration given in Table 4.

TABLE 4. MODEL PARAMETERS

MODEL Configuration	**Model 1**	**Model 2**	**Model 3**
	LSTM (shallow LSTM)	**GRU (Shallow GRU)**	**Bi-GRU+CNN**
No of layers	1	1	2 BiGRU ,4 CNN
No of neurons	32	32	256,128,64,32
Embedding	Fasttext	Glove	ECE
Dense neuron	20	32,100	128,64
Droput	0.5	NA	0.2
Kernal size for CNN	NA	NA	2,3,5,7
Activation function	ReLU, Sigmoid	ReLU, Sigmoid	ReLU, Sigmoid
Optimizer	Adam	Rmsprop	Adam

b)Evaluation Matrix:

The following assessment measures are employed in this paper to assess the efficacy of the Deep learning model with ECE that is proposed. Table 5. represents way to evaluate SA model

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7216
***Article Received**: 29 April 2023 **Revised:** 16 June 2023 **Accepted:** 29 June 2023*

TABLE 5. EVALUATION MATRIX

	SA Model classification results	
	Positive sentiment	**Negative sentiment**
Positive sentiment	Correctly Predicted positive SA (CPS)	Mistakenly Negative SA samples (MNS)
Negative sentiment	Mistakenly Positive SA samples (MPS)	Correctly predicted Negative Samples(CNS)

The evaluation metrics are defined as:

precision (p) $$p = \frac{\text{CPS}}{\text{CPS} + \text{MPS}} \tag{13}$$

recall (r): $$r = \frac{\text{CNS}}{\text{CPS}+MNS} \tag{14}$$

F1-Score(f1): $$f1 = \frac{2}{\frac{1}{p} + \frac{1}{r}} = \frac{2*p*r}{p+r} \tag{15}$$

Accuracy (A): $$A = \frac{\text{CPS+CNS}}{\text{CCPS+MPS+CNS+MNS}} \tag{16}$$

IV. RESULTS AND DISCUSSIONS

The ECE embedding is tested across the very basic Deep Net architecture such as CNN, LSTM, Bi-LSTM and GRU, along with machine learning model [75],[76],[77]. The summarized details as given in Table 5. We compared our approach against Support Vector Machines (SVR), Random Forests (RF), and Multi-Layer Perceptron (MLP) regressor. The details on accuracy and F1-score shown in Table 6. The accuracy and f1 score are computed with the of (15) and (16).

TABLE 6. COMPARATIVE ANALYSIS WITH STATE-OF-ART METHODS

Models	**Kindle**		**MR**		**Real Time Tweets**	
	F1 Score	**Accuracy**	**F1 Score**	**Accuracy**	**F1 Score**	**Accuracy**
Linear Regression	0.52	0.52	0.55	0.51	0.75	0.75
SVM	0.40	0.49	0.48	0.48	0.69	0.69
Naive Bayesian	0.48	0.51	0.52	0.50	0.71	0.69
Random Forest	0.56	0.58	0.80	0.80	0.70	0.69
CNN+glove	0.88	0.88	0.87	0.85	0.75	0.76
CNN+fasttext	0.88	0.88	0.88	0.87	0.76	0.766
Multichannel +CNN+ECE	0.86	0.85	0.94	0.94	0.77	0.75
CNN+GRU+ECE	0.85	0.83	0.93	0.93	0.74	0.74
LSTM+ECE	0.85	0.85	0.94	0.94	0.75	0.75
CNN+ECE	0.87	0.87	0.94	0.93	0.74	0.75
Proposed Ensemble Model	**0.89**	**0.886**	**0.95**	**0.94**	**0.81**	**0.79**

The results computed on three different types of datasets. The two belong to review category while other belongs to tweet category. The results got varied because of language structure, use of words, presence of ambiguity and context in the sentences with the max length of the sentences. The model is applied with low-resource setting making the assured processing should be faster with good results. Looking at the result we found Kindle got highest accuracy about 88.6%, MR got 94% and real time complaint tweet got 79%. The corresponding F1-score results as 89%,95%,81% with CNN model for Kindle, MR and tweet dataset respectively. But the results are varied depends on the model used. Here in this due to ECE we could observer the F1 score is almost similar with simple model as well as with complicated model such as in Kindle where the F1 score by Multichannel+CNN+ECE 86%, CNN+GRU+ECE 85% and with LSTM 85%. Here we found the complexity of model is reduced with ECE embedding. Similar case found in tweet the performance is lowered due the complex nature of complaint related dataset. The accuracy remains same with respect to model Multichannel+CNN+ECE, LSTM+CNN and CNN+ECE. So, our motive of obtained the improvement in performance in low resource setting can be very well obtained. Without going with very complex structure model can give good results. The proposed ensembled model giving the quite promising results with real time complaint tweet as 81% as f1 score and 79% as accuracy.

With ECE we compared the existing star-of-art architecture proposed by researchers across other composite embedding strategy proposed by different researcher as below. The corresponding comparison is shown in Figure 4 and Figure 5

MCP-LSTM [80]: To learn semantic and sentiment feature representations, this model employs LSTM and CNN-multichannel network. It uses two types of word vectors; one is a static word vector as glove and the other is a non-static word vector as Multichannel word vector. The model uses features to combined with concatenation. The result given to fully connected layer to decide the class.

Hybrid CNN+GRU+glove [81]: The For the purpose of classifying sentiment polarity, this model has three convolutional layers, a max pooling layer, and a fully connected layer. For the sentiment classification in the proposed model the two branches were created. Both the active CNN character embedding and the LSTM FastText embedding can be used by one branch. The tweets are categorized by combining the features collected from both directions.

Hybrid Model-CNN+BiLSTM+Character+fastText+word2vec [82]: Proposed model combines different text representations and deep learning methods.

ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7216
***Article Received**: 29 April 2023 **Revised:** 16 June 2023 **Accepted:** 29 June 2023*

ABCDM [68]: Attention-based Bidirectional CNN-RNN Deep Model (ABCDM) uses two separate bidirectional LSTM and GRU layers to get context related information. In addition, ABCDM's bidirectional layer outputs are fed into the attention mechanism, for highlight particular phrases. ABCDM uses convolution and pooling layers reduce the dimensionality of data and extract position-invariant local features.

CNN+glove+fastText+BERT [83]: For the sentiment analysis tasks the static and contextual embedding combined. The resultant vector given to CNN model for classification tasks. The static embedding considered as glove whereas contextual embedding considered as BERT.

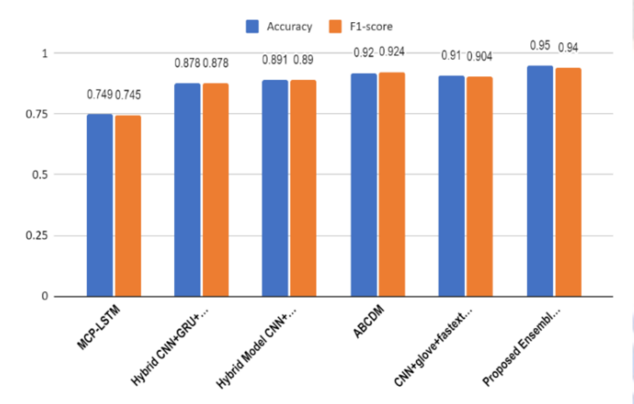

Figure 4. ECE performance, MR Dataset and other Models

The impact of composite embedding is higher as compared to state-of-art language models. With the ECE embedding for MR dataset the performance with reach up to 0.95 f1-score and similar with accuracy too. The other models also have the similar performance. But with ECE we able to get 3% improvement as compared to the CNN+glove+fastText+BERT combination implemented in [83] where as its good improvement with respect to MCP-LSTM, Hybrid embedding which is combination with CNN+ BiLSTM + character +FastText +word2vec [82] and also with ABCDM [68]. There is an 11% improvement with respect to [80].

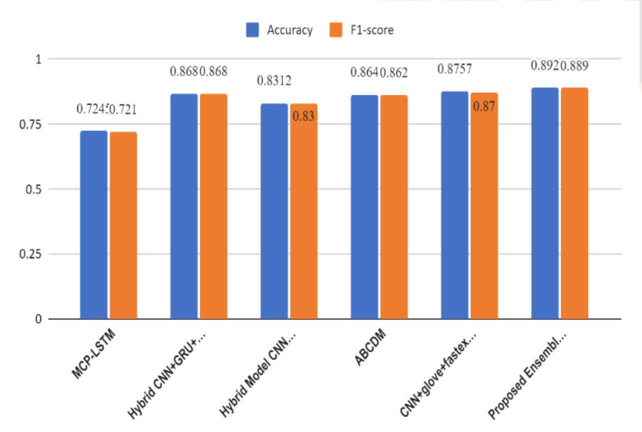

Figure 5. ECE performance, Kindle Dataset and other Models

When the accuracy obtained from is different language model compared based on the accuracy and F1-score. For the Kindle dataset the maximum improvement we reached up to 89% as accuracy and 0.889 as F1-score. When it compared to previous approaches and composite embedding strategy the improvement as quite comparable. It is about 6% as compared to [82]. When compared with MCP-LSTM the accuracy and F1-score is about 17%.

Explainability Test:

The explanation with NLP is discussed by many researchers provides the way to decipher the class decision at local and global level. The RQ3 addressed with LIME. LIME (Local Interpretable Model-agnostic Explanations) one kind of technique proposed in [37]. Here in our experimentation, we used it for understanding the decision. The model used for Kindle and Hate speech dataset is Random Forest. The model prediction given model is further verified with the explainability library such as LIME [37]. We deciphered the model decision with particular instance from the dataset.

The graphical representation of assigned probabilities by the model justified in Figure. 6, Figure.7, Figure 8 for MR, Kindle, Tweet dataset.

For MR Dataset: Local and global explanation designed with lime.

x_test1[52]“we thank our god for you every single day maga to kag “

y_test1[52]= 0

Probability(Negative) = 0.878

Probability(Positive) = 0.122

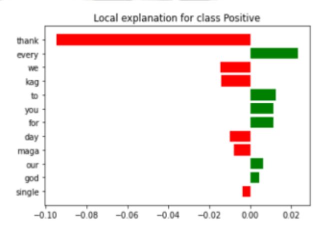

Figure 6. MR dataset Local Explanation using LIME form instance x_test1[52]

From the above graphical representation, we will come to know the important words that justifies the model decisions. Such as for Kindle dataset “Enjoy “got the highest probability score where as in MR the “thank” got highest probability weightage. From this we can get the global and local level information about how model would have assigned a particular polarity for a sentence with considering the probability at each word. Similarly, we found highest importance given for “slow” in case of tweet dataset. The LIME gives the explains the model decisions with the different types of explainers. We chose here the tree-explainer along with Random Forest.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7216
***Article Received**: 29 April 2023 **Revised:** 16 June 2023 **Accepted:** 29 June 2023*

x_test1[5]): "the book starts off at a good pace and only builds on it as the story progresses. one revelation after another followed by more questions and small mysteries keeps you engaged. although the plot wont win much for originality, the characters are interesting with good interactions between them and the action is plentiful"

(y_test1[5])=1

Probability(Negative) = 0.34

Probability(Positive) = 0.66

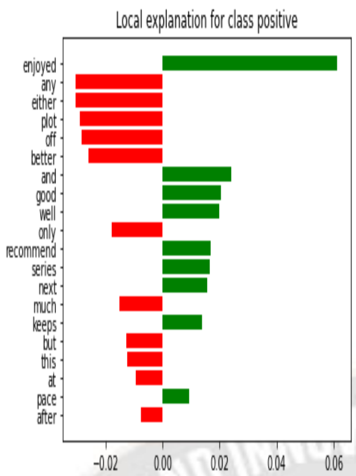

Figure 7. Kindle Dataset document id[5] Local Explanation using LIME

The same we applied with the Ensemble model to check the model decision with the help of LIME

x_test1[5]) over grnturtl slow today lone

(y_test1[5])= 1

Probability(Negative) = 0.89

Probability(Positive) = 0.11

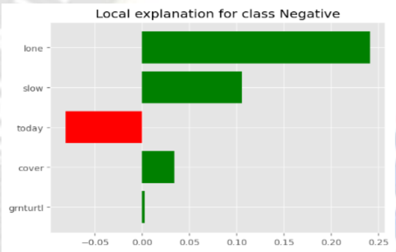

Figure 8. Tweet Dataset with Local Explanation using LIME

V. FUTURE SCOPE AND CONCLUSION

This study has various limitations that indicate the path of further investigation for future studies. First, this article's sentiment analysis approach is the binary classification method, which analyses the sentiment trend of brief consumer reviews and tweets. In the future, the approach of multi-class sentiment analysis can be utilised to refine the effect of sentiment analysis. In addition, advanced contextual embedding can be utilised to determine emotional strength. Second, the amount of data in this article can be increased to give a more credible resource for future large-scale data analyses. Lastly, the approach described in this article for combining each connection is changeable. With the constant growth of technology, it will be deemed obsolete in the near future and replaced with a technique that is more effective and precise.

Many practical situations need for better decision-making informed by people's opinions, making sentiment or emotion analysis an intriguing area of study. The sentiment analysis problem can be efficiently handled with the help of modern Deep learning architecture. The embedding which represents the feature representational layer plays an important role. ECE a pretrained embedding techniques can be efficiently solve sentiment analysis problem in low resource setting and avoiding the high computational resources. ECE with ensembled model is tested across three different dataset and method gives remarkable results along with preserving generalization ability. ECE embedding best fits with constraints and gives comparative performance.

REFERENCES

[1] Xu, Y., Ren, J., Wang, G., Zhang, C., Yang, J., & Zhang, Y. (2019). A blockchain-based nonrepudiation network computing service scheme for industrial IoT. IEEE Transactions on Industrial Informatics, 15(6), 3632-3641.

[2] O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010, May). From tweets to polls: Linking text sentiment to public opinion time series. In Fourth international AAAI conference on weblogs and social media.

[3] Liu, B. (2015). Sentiment analysis: Mining opinions, sentiments, and emotions. Cambridge: CUP. Go to original source..

[4] Pang, B., & Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. arXiv preprint cs/0409058. Elissa, "Title of paper if known," unpublished.

[5] Teng, Z., Vo, D. T., & Zhang, Y. (2016, November). Context-sensitive lexicon features for neural sentiment analysis. In Proceedings of the 2016 conference on empirical methods in natural language processing (pp. 1629-1638).

[6] Yan, K., Zhong, C., Ji, Z., & Huang, J. (2018). Semi-supervised learning for early detection and diagnosis of various air handling unit faults. Energy and Buildings, 181, 75-83.

[7] Lu, H., Yang, L., Yan, K., Xue, Y., & Gao, Z. (2017). A cost-sensitive rotation forest algorithm for gene expression data classification. Neurocomputing, 228, 270-276.

[8] Huang, Q., Chen, R., Zheng, X., & Dong, Z. (2017, August). Deep sentiment representation based on CNN and LSTM. In 2017 international conference on green informatics (ICGI) (pp. 30-33). IEEE.

[9] Neethu, M. S., & Rajasree, R. (2013, July). Sentiment analysis in twitter using machine learning techniques. In 2013 fourth international conference on computing, communications and networking technologies (ICCCNT) (pp. 1-5). IEEE.

[10] Xia, H., Yang, Y., Pan, X., Zhang, Z., & An, W. (2020). Sentiment analysis for online reviews using conditional random fields and support vector machines. Electronic Commerce Research, 20(2), 343-360.

[11] Qu, L., Ifrim, G., & Weikum, G. (2010, August). The bag-of-opinions method for review rating prediction from sparse text patterns. In Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010) (pp. 913-921).

[12] Campos, V., Jou, B., & Giro-i-Nieto, X. (2017). From pixels to sentiment: Fine-tuning CNNs for visual sentiment prediction. Image and Vision Computing, 65, 15-22.

[13] Schmidhuber, J. (2015). Deep learning in neural networks: An overview. Neural networks, 61, 85-117.

[14] Marasek, K. (2015). Deep belief neural networks and bidirectional long-short term memory hybrid for speech recognition. Archives of Acoustics, 40(2), 191-195.

[15] Jawale, S., & Sawarkar, S. D. (2020, December). Interpretable Sentiment Analysis based on Deep Learning: An overview. In 2020 IEEE Pune Section International Conference (PuneCon) (pp. 65-70). IEEE.

[16] Goldberg, Y. (2016). A primer on neural network models for natural language processing. Journal of Artificial Intelligence Research, 57, 345-420.
[17] Jawale, S., & Sawarkar, S. D. (2023). Sentiment Analysis and Vector Embedding: A Comparative Study. In Smart Trends in Computing and Communications (pp. 311-321). Springer, Singapore.
[18] Hameed, Z., & Garcia-Zapirain, B. (2020). Sentiment classification using a single-layered BiLSTM model. Ieee Access, 8, 73992-74001.
[19] Gupta, P., & Jaggi, M. (2021). Obtaining better static word embeddings using contextual embedding models. arXiv preprint arXiv:2106.04302.
[20] Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4), e1253.
[21] Al-Moslmi, T., Omar, N., Abdullah, S., & Albared, M. (2017). Approaches to cross-domain sentiment analysis: A systematic literature review. Ieee access, 5, 16173-16192.
[22] T. Mikolov, K. Chen, G. Corrado, J. Dean, “Efficient Estimation of Word Representations in Vector Space,” Proc. Workshop at ICLR, 2013.
[23] T. Mikolov, W.Yih, G. Zweig, “Linguistic Regularities in Continuous Space Word Representations,” Proc. NAACL HLT, 2013.
[24] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, J. Dean, “Distributed Representations of Words and Phrases and their Compositionality,” Proc.
[25] Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) (pp. 1532-1543).
[26] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017.Enriching word vectors with subword information. Transactions of the Association for Computational Linguistics 5 (2017), 135–146
[27] Matthew Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018a. Deep contextualized word representations. In NAACL
[28] Han, Xu, et al. "Pre-trained models: Past, present and future." AI Open 2 (2021): 225-250.
[29] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
[30] Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2019). Albert: A lite bert for self-supervised learning of language representations. arXiv preprint arXiv:1909.11942.
[31] Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., & Le, Q. V. (2019). Xlnet: Generalized autoregressive pretraining for language understanding. Advances in neural information processing systems, 32.
[32] Xlnet: Generalized autoregressive pretraining for language understanding. Advances in neural information processing systems, 32.
[33] Jawale, S. S., & Sawarker, S. D. (2022). Amalgamation of Embeddings With Model Explainability for Sentiment Analysis. International Journal of Applied Evolutionary Computation (IJAEC), 13(1), 1-24.
[34] Ghannay, S., Favre, B., Esteve, Y., & Camelin, N. (2016, May). Word embedding evaluation and combination. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16) (pp. 300-305).
[35] Khasanah, I. N. (2021). Sentiment Classification Using fastText Embedding and Deep Learning Model. Procedia Computer Science, 189, 343-350.
[36] Deep Learning Model. Procedia Computer Science, 189, 343-350.
[37] Tulio Ribeiro, M., Singh, S., & Guestrin, C. (2016). " Why Should I Trust You?": Explaining the Predictions of Any Classifier. ArXiv e-prints, arXiv-1602.
[38] Li, J., Chen, X., Hovy, E., & Jurafsky, D. (2015). Visualizing and understanding neural models in nlp. arXiv preprint arXiv:1506.01066.
[39] Liu, H., & Cocea, M. (2017). Fuzzy information granulation towards interpretable sentiment analysis. Granular Computing, 2(4), 289-302.
[40] Pota, M., Ventura, M., Catelli, R., & Esposito, M. (2020). An effective BERT-based pipeline for Twitter sentiment analysis: A case study in Italian. Sensors, 21(1), 133.
[41] Dorle, S., & Pise, N. N. (2017). Sentiment Analysis Methods and Approach: Survey. International Journal of Innovative Computer Science & Engineering, 4(6), 7-11.
[42] Ganaie, M. A., Hu, M., Malik, A. K., Tanveer, M., & Suganthan, P. N. (2022). Ensemble deep learning: A review. Engineering Applications of Artificial Intelligence, 115, 105151.
[43] S. B, S. S, G. D. T and V. R. N, "Sentiment Analysis on Movie Reviews: A Comparative Analysis," 2023 International Conference on Intelligent Systems for Communication, IoT and Security (ICISCoIS), Coimbatore, India, 2023, pp. 218-223. doi: 10.1109/ICISCoIS56541.2023.10100367
[44] Kamruzzaman, M., Hossain, M., Imran, M. R. I., & Bakchy, S. C. (2021, August). A comparative analysis of sentiment classification based on deep and traditional ensemble machine learning models. In 2021 International Conference on Science & Contemporary Technologies (ICSCT) (pp. 1-5). IEEE.
[45] .Hassan, A., & Mahmood, A. (2018). Convolutional recurrent deep learning model for sentence classification. Ieee Access, 6, 13949-13957.
[46] Joachims, T. (1998, April). Text categorization with support vector machines: Learning with many relevant features. In European conference on machine learning (pp. 137-142). Berlin, Heidelberg: Springer Berlin Heidelberg.
[47] Prabha, M. I., & Srikanth, G. U. (2019, April). Survey of sentiment analysis using deep learning techniques. In 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT) (pp. 1-9). IEEE.
[48] Z. Z. Wint, Y. Manabe and M. Aritsugi,2018 IEEEInternational Conference on Big Data, Cloud Computing, Data Science & Engineering (BCD), Yonago, 2018, pp. 91-96.
[49] Huy Nguyen and Minh-Le Nguyen,K. Hasida and W. P. Pa (Eds.): PACLING 2017, Communications in computer and information science (CCIS), vol 781, pp. 15–27, 2018.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7216
***Article Received**: 29 April 2023 **Revised:** 16 June 2023 **Accepted:** 29 June 2023*

[50] Jianqiang, Z., Xiaolin, G., & Xuejun, Z. (2018). Deep convolution neural networks for twitter sentiment analysis. IEEE access, 6, 23253-23260.
[51] Wei Xue, Tao Li in Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Long Papers), pages 2514– 2523 Melbourne, Australia, July 15 - 20, 2018
[52] Jianqiang, Z., Xiaolin, G., & Xuejun, Z. (2018). Deep convolution neural networks for twitter sentiment analysis. IEEE access, 6, 23253-23260.
[53] A. Onaciu and A. Nicoleta Marginean 2018 IEEE 14th International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, 2018, pp. 13-19
[54] Ganaie, M. A., Hu, M., Malik, A. K., Tanveer, M., & Suganthan, P. N. (2022). Ensemble deep learning: A review. Engineering Applications of Artificial Intelligence, 115, 105151.
[55] N. Aslam, F. Rustam, E. Lee, P. B. Washington and I. Ashraf, "Sentiment Analysis and Emotion Detection on Cryptocurrency Related Tweets Using Ensemble LSTM-GRU Model," in IEEE Access, vol. 10, pp. 39313-39324, 2022. doi: 10.1109/ ACCESS.2022. 3165621.
[56] Baroni, M., Dinu, G., & Kruszewski, G. (2014, June). Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 238-247).
[57] Roman, M., Shahid, A., Khan, S., Koubaa, A., & Yu, L. (2021). Citation intent classification using word embedding. Ieee Access, 9, 9982-9995.
[58] Jawale and Sawarkar,"Word Embedding and Intent Analysis: A Literature Survey (2022)",http://www.gisscience.net/VOLUME-9-ISSUE-2-2022/, DOI:20.18001.GSJ.2022.V9I2.22.38707
[59] Bollegala, D., & O'Neill, J. (2022). A Survey on Word Meta-Embedding Learning. arXiv preprint arXiv:2204.11660.
[60] Tan, K. L., Lee, C. P., Lim, K. M., & Anbananthen, K. S. M. (2022). Sentiment analysis with ensemble hybrid deep learning model. IEEE Access, 10, 103694-103704..
[61] Yesir, S., & Soğukpinar, İ. (2021, June). Malware Detection and Classification Using fastText and BERT. In 2021 9th International Symposium on Digital Forensics and Security (ISDFS) (pp. 1-6). IEEE.
[62] Sheng, D., & Yuan, J. (2021, May). An efficient long Chinese text sentiment analysis method using BERT-based models with BiGRU. In 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD) (pp. 192-197). IEEE.
[63] Wang, J., Zhang, Y., Yu, L. C., & Zhang, X. (2022). Contextual sentiment embeddings via bi-directional GRU language model. Knowledge-Based Systems, 235, 107663.
[64] Dedhia, C., & Ramteke, J. (2017, January). Ensemble model for Twitter sentiment analysis. In 2017 International Conference on Inventive Systems and Control (ICISC) (pp. 1-5). IEEE.
[65] Ghannay, S., Favre, B., Esteve, Y., & Camelin, N. (2016, May). Word embedding evaluation and combination. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16) (pp. 300-305).
[66] Alharbi, A. I., Smith, P., & Lee, M. (2021). Enhancing contextualised language models with static character and word embeddings for emotional intensity and sentiment strength detection in arabic tweets. Procedia Computer Science, 189, 258-265.
[67] Waheeb , M. Q. ., SANGEETHA, D., & Raj , R. . (2021). Detection of Various Plant Disease Stages and Its Prevention Method Based on Deep Learning Technique. Research Journal of Computer Systems and Engineering, 2(2), 33:37. Retrieved from https://technicaljournals.org/RJCSE/index.php/journal/article/view/30
[68] Da Silva, N.F.F., Hruschka, E.R., Hruschka, E.R.: Tweet sentiment analysis with classifier ensembles. Decis. Support Syst. 66, 170–179 (2014) (2014): 1093-1113
[69] Basiri, M. E., Nemati, S., Abdar, M., Cambria, E., & Acharya, U. R. (2021). ABCDM: An attention-based bidirectional CNN-RNN deep model for sentiment analysis. Future Generation Computer Systems, 115, 279-294
[70] Wall, M. E., Rechtsteiner, A., & Rocha, L. M. (2003). Singular value decomposition and principal component analysis. In A practical approach to microarray data analysis (pp. 91-109). Springer, Boston, MA
[71] Sukhbaatar, S., Weston, J., & Fergus, R. (2015). End-to-end memory networks. Advances in neural information processing systems, 28.
[72] Guo, B., Zhang, C., Liu, J., & Ma, X. (2019). Improving text classification with weighted word embeddings via a multi-channel TextCNN model. Neurocomputing, 363, 366-374.
[73] He, R., & McAuley, J. (2016, April). Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In proceedings of the 25th international conference on world wide web (pp. 507-517).
[74] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, andR. Salakhutdinov, ''Dropout: A simple way to prevent neural networksfrom overfitting,'' J. Mach. Learn. Res., vol. 15, no. 1, pp. 1958–7929,2014
[75] Prechelt, L. (2002). Early stopping-but when?. In Neural Networks: Tricks of the trade (pp. 55-69). Berlin, Heidelberg: Springer Berlin Heidelberg.
[76] Dessì, D., Dragoni, M., Fenu, G., Marras, M., & Recupero, D. R. (2019, April). Evaluating neural word embeddings created from online course reviews for sentiment analysis. In Proceedings of the 34th ACM/SIGAPP symposium on applied computing (pp. 2124-2127).
[77] Yoon, K. (2014). Convolutional Neural Networks for Sentence Classification [OL]. arXiv Preprint.
[78] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.
[79] Rao, G., Huang, W., Feng, Z., & Cong, Q. (2018). LSTM with sentence representations for document-level sentiment classification. Neurocomputing, 308, 49-57.
[80] Xu, G., Meng, Y., Qiu, X., Yu, Z., & Wu, X. (2019). Sentiment analysis of comment texts based on BiLSTM. Ieee Access, 7, 51522-51532.
[81] Long, Y., Li, Y., Luo, J., Miao, C., & Fu, J. (2019, November). MCP-LSTM Network for Sentence-Level Sentiment

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7216
***Article Received**: 29 April 2023 **Revised:** 16 June 2023 **Accepted:** 29 June 2023*

Classification. In 2019 International Conference on Virtual Reality and Visualization (ICVRV) (pp. 124-128). IEEE.

[82] Zouzou, A., & El Azami, I. (2021, October). Text sentiment analysis with CNN & GRU model using GloVe. In 2021 Fifth International Conference On Intelligent Computing in Data Sciences (ICDS) (pp. 1-5). IEEE.

[83] Salur, M. U., & Aydin, I. (2020). A novel hybrid deep learning model for sentiment classification. IEEE Access, 8, 58080-58093.

[84] Alghanmi, I., Espinosa-Anke, L., & Schockaert, S. (2020). Combining BERT with static word embeddings for categorizing social media.