HOF-CNN: The Hybrid Optimized Feature-based CNN For Sentiment Review Classification Model

Sandeep Kumar, Nand Kumar

Department of Computer Science and Engineering, Lingaya's Vidyapeeth, Faridabad (Haryana), India
Email:- sandeepnimbal@gmail.com

Abstract—

Nowadays, opinion mining tools enable consumers to process many reviews to determine the essential views proficiently.In the proposed research, Kernel Principal Component Analysis (KPCA) method is used for feature extraction of semantic and context features of words. The research model uses CNN (Convolutional Neural Network) to understand complex designs in-text information. The hybrid A-PSOmethod is used to enhance the selection of the best metric to solve the double optimization issue. Wide simulations calculated on the IMDb movie database proved that our research method Hybrid Optimized Feature-based CNN (HOF-CNN), secure precision over the traditional techniques like SVM and Naïve Bayesthat outcome in maximum classification accuracy rate. According to the experiment analysis, the research method outperforms previous methods in different parameters and enhances the accuracy to 97.0% on average.

Keywords—*Opinion Mining;HOF-CNN (Hybrid Optimized Feature-based CNN) model; Convolutional Neural Network; KPCA (Kernel Principal Component Analysis); A-PSO (Adaptive-Particle Swarm Optimizer.*

Number: 10.14704/nq.2022.20.9.NQ440013 **Neuro Quantology 2022; 20(9):110-120**

I. INTRODUCTION

When deciding what to buy, consumers frequently seek out the advice of others. To determine whether to buy a product, support a civilapplicant, or pick a movie, it is customary to ask friends, favorite bloggers, and reviewers for their advice. On websites, weblogs, and social media platforms, opinions frequently in the form of sentiments are available more regularly. Businesses spend a lot of money gathering and analyzing web data to understand the underlying patterns of consumer sentiment and utilize that data to better their goods and services and drive their production plans and marketing strategies.

Because opinions stated online are primarily written in natural language, opinion mining technologies are needed to make it easier to extract and analyze statements from unstructured material.These technologies frequently use algorithms from the fields of ML, information retrieval, and NLP.

Opinion mining is generally proposed by removing contents for a particular area like a movie, product, music, etc. [1] and evaluating opinion mining at several stages of text granularity such as; document, sentence,and feature-level [2]. At the document and sentence level, opinion mining objectives categorize the complete sentiment orientation, i.e., defined in a document [4,5] or sentence [3]. Opinion mining at the featurelevel is not easy because it needs profound consideration of sentence organization and knowledge of the issue area to accurately classify sentences based on their polarity.

Opinion Mining research at the feature level works different methods like association rule mining [6], ML, and SK (semantic knowledgebase)techniques to mainlyenhance the result of the feature extraction job, thus refining the performance of the opinion arrangement job. These methodsprimarily rely on NLP to verify nouns and noun phrases as domain feature sets. In contrast, ML methods rely on vast training data to study domain feature sets from sentiments. SK [7] procedures have been organized for domain feature extraction with capable accomplishment. These methods use domain knowledgebase to fetch domain feature sets from sentiments that

comprise a hypothesized knowledge background of the field.

Currently, the SK and ML approaches have occurred as an operative method for enhancing the procedure of opinion mining at the feature-level, specifically, for improving the sentiment analysis job [8]. Moreover, the described determinations have primarily been absorbed on numerous classification tasks verifying whether the content has a positive or negative sentiment. The research work defined in this article is mainly inspired by the essential to growing a HOF-CNN method to enhance the presentation of opinion mining at the feature-based and enhance the opinion classification ofthe CNN classification model.

Theplanned hybrid optimized feature-based CNN deep learning method can be summarized as follows:

- KPCA technique enhances the accuracy and precision of the extracted feature sets. The KPCA method uses complete information of the selected area and public-connected open data sources like the IMDb movie database.
- The hybrid optimized feature-based CNN method reduces false positive sentiments or reviews.
- To propose and implement optimized feature extraction with KPCA to advance the proposed presentation of the predictionmethod.

The article is managed as tails: prior work is defined in section 2. Section 3 discusses the problem formulation in this work, and existing methods are described in section 4. Section 5 describedthe research Hybrid optimized feature-based CNN deep learning method for opinion mining. Section 6 describes the experiment result and analysis. Lastly,conclusion and advanced improvement are defined in section 7.

II. Related Work

This section addresses relevant opinion mining literature, emphasizing methods for extracting the right emotions and meaning behind tweets using deep learning techniques and machine learning algorithms.**Stefanos Zervoudakis et al. (2021) [9]**presented opinionMine, a Bayesian-based outline for opinion mining, developing a Twitter dataset. Firstly, the research model imports Tweets enormously by utilizing Twitter API. Subsequent, the introduced Tweets were more managed spontaneously for creating a set of un-trained principles and random variables. So, the bayesian network was presented using the un-trained principles. The random variables and a significantcollection. Afterward, the proficient method may be utilized to calculate novel Tweets. Lastly, the built process may be re-trained incrementally, so fetchingwas more vigorous. **Diaa Salama Abdelminaamet al. (2021) [10]**described this work was to studied ATs (Arabic Tweets) in terms of positive, negative, or neutral. In learning the reviews of the AL (Arabic language), a limitation lies in utilizing various dialects. They overviewed two main features: Initial works of six ML techniques, adding DT, LR, KNN, RF, SVM, and NB, by TF-IDF technique temporary as the FE (feature extraction). The 2nd section analyzes 3 alternatives of DL-based mLSTM, LSTM, and GRU by relating WE (word embedding) as the initial vector.

Mohammadreza Keyvanpour et al. (2020) [11] developed a model for OMLML, a suitabletechnique based on vocabulary and ML, which was planned. Itstop features and advantages over other methods include addressing testslike; high-dimensions of feature space, uncertainty in acknowledgment of expressive concepts of some words, and disbursingconsiderationto only the weightbased on wordfrequency in opinion mining.The excellent training cost wassimultaneously based on time or memory utilized and usability in any specific area. The polarity of sentimentsnear a target word was first identified using a techniquebased on the suggested method's linguistic and textual properties of words and sentences.**Prayag Tiwari et al. (2019) [12]** analyzed and proposed an ML method to predict the customer's Tweets relating to the aircraft services to appreciate the design of sentiments. They adopt RF and LR methods to predict each Tweet into +ive and –ive sentiment.The calculation of the gathered basic information validates that these binary techniques were able to attain an accuracy of 80 percent. **Noor Afiza Mat Razali et al. (2021) [13]** used two machine-learning models to handle the sentiment categorization problem. They attained an accuracy of 82 percent based on the collected data. This study influences the aviation sector since it gives an effective and efficient technique

for airline businesses to monitor customers' attitudes to enhance their service. In the future, they hope to examine the information deeper and extract more helpful information to provide suggestions to various airline organizations and customers. It would also employ a larger dataset than the one now being used since a more extensive dataset may yield better results than the one currently being used. **Dubey, K. P., et al. (2019) [14]**described the viewpoints of 'Random Forest' that directly affect the comprehensive correctness of the interpretation or analysis. For categorization, the plannedmethod has an accuracy of 81.69 percent. As a result of the comparison between various approaches and the suggested way, the proposed techniques were preferred in the critical analyses of quantifiable aspects of successfully categorizing out of all instances, SP, F-score, and AUC. In future study, the authors are planning tocombine the increased precision of the built model with various ensemble approaches to categorize miners presenting positive, highly optimistic, neutral, negative, or strongly negative attitudes. Table 1comprehensively studies several authors' works on opinion mining regarding their comparative methodologies, datasets utilized, and different parameters investigated.

TABLE 1 COMPARATIVE ANALYSIS

Ref. no.	**Comparative methods**	**Dataset**	**Parameters**	**Further Work**
[9]	Test cases based on tweets	Twitter	MSE MAE RMSE	The Bayesian framework can be used in multiple models from the same training on different topics.
[10]	Arabic Social Media	TEAD ATSAD ASTD	NB KNN LR SVM DT RF	Swarm algorithms will be used to choose the optimal hyper parameters.
[11]	High dimension of feature, space, and ambiguity of words	TeS TrS	Accuracy Precision Recall F-measure	Increase the efficiency of the OMLML method
[12]	Passenger's tweets regarding the airplane services	US Airlines 2016, which contains 14500 passenger tweets	LR and RF methods	To identify sarcasm in sentences more efficiently.
[13]	Tweets and IPL hashtags	Twitter's API	Accuracy Sensitivity Specificity Prevalence F-score	To include more accurateness of the build model with different ensemble methods.

Table 2 summarizes the approaches to Opinion Mining of many authors, the recommended methodologies they have presented via their study and models, the obstacles they encountered and the gap that remains to be filled, as well as the tools being utilized and the study's results.

TABLE 2 ANALYSES WITH EXISTING METHODS

Author's name and Year	**Proposed method**	**Simulation Tools**	**Outcome**
Stefanos Zervoudakis et al., (2021) [9]	Bayesian-based Framework	VADER tool &ProbLog tool	The resultant model proved to be progressively retrained with additional training sets to become more robust.
Diaa Salama Abdelminaamet al., (2021) [10]	Optimized Deep Neural Networks, mLSTM LSTM and GRU	-	Associated to the state-of-the-art, demonstrateda precisepresentation for DL approaches based on a 10-folds strategy.
Mohammadreza Keyvanpour et al., (2020) [11]	Lexicon and ML (machine learning) are known as OMLML.	Lexicon-based SA tool	Quantitative and qualitative trials reveal that mapping facts into a novel space reduces training costs and that the suggested method's performance is satisfactory, especially in terms of accuracy.
Prayag Tiwari et al., (2019) [12]	RF and LR methods	-	The examination of the acquired real data shows that these two approaches can attain an accuracy of 80%.
Noor Afiza Mat Razali et al., (2021) [13]	ML and lexicon-based method, hybrid method, and Kanseimethod in mining the opinion and emotion-based on text are used	Review and opinion mining, attitude examination, or appraisal extraction	The Kansei method should be considered a complementing aspect in constructing an emotion-focused vocabulary in the national security area.
Dubey, K. P., et al., (2019) [14]	As favorable or negative sentiment on tweets provides the efficacy of a machine learning model. The collecting of tweet data and its processing using the finest of the permitted IPL hashtags.	Accuracy of Prediction, Random forest technique & sentiment classification	Individuals' positive and negative thoughts are displayed in the practical result. This form of opinion analysis may provide significant input to the company and assist them in identifying a bad Twitter switch in audience comprehension.

Abbreviations:MSE(Mean Square Error Rate), MAE (Mean Absolute Error), RMSE(Root Mean Squared Error), TEAD(Tweets Emoji Arabic Dataset), ATSAD(Arabic Tweets Sentiment Analysis Dataset), ASTD(Arabic Sentiment Tweets Dataset) NB(Naive Bayes), KNN(K-

Nearest Neighbor), LR(Logistic Regression), SVM(Support Vector Machine), DT(Decision Tree), RF(Random Forest), SP(Specificity),mLSTM (multiplicative Long Short-Term Memory), GRU (Gated Recurrent Unit), NB(Naïve Bayes).

III. Problem Formulation

The thoughts and feelings of others can influence human choices (opinions). Humans are frequently affected by perceptions of someone else to make judgments. Opinions are subjective statements that outline the user's thoughts or opinions regarding a specific topic. Opinion mining is an essential and ever-evolving discipline that dynamically brings a slew of new obstacles [15] [16]. For instance, determining which section of text is the opinion, choosing the polarization of the opinion, determining the sentiment power, determining the truthfulness or legitimacy of the opinion, spam filtering, phrase behind the opinion, identifying worthless opinion, identifying opinion falsely, identifying uninformed viewpoint, uncertain knowledge, etc. are all outstanding issues further for research and innovation. Opinions are commonly produced spontaneously and incredibly different, posing many obstacles. There are various existing opinion mining techniques but still suffer from multiple challenges. Several challenges in opinion mining include authorization, non-expert opinion, spam opinion detection, opinion credibility, relevancy of the opinions, NLPoverhead, typographical errors, etc.

- **Authorization:** The acceptability of the viewpoint is critical. The information source should be generally recognized in society. In other terms, the approval of the opinion holder may be determined, for instance, if the opinion is delivered by a particular area or a compelling factor.
- **Non-expert Opinion:** Non-expert viewpoints wreak havoc on websites and open opportunities. Exceptionally dedicated websites, such as programmer blog posts, in which an expert's opinion is relevant, impact opinion mining since they do not correctly review or opinion.
- **NLP (Natural Language Processing):** The overhead of NLs, such as uncertainty, co-reference, indirectness, interpretation, and so on, restricted sentiment analysis tools.
- **Typographical Errors:** Humans can comprehend words such as knowlege for knowledge, a school for school, and other grammatical mistakes. However, extracting and evaluating typo mistakes is difficult for an opinion miner.

Sometimes, it is not easy to analyze social site sentiment reviews. Because duplicate data (good, not bad, etc.) is presented in the dataset, it is the main reason for achieving the minimum accuracy rate, maximum error rate, and time consumption. The literature review has studied the various machine learning and deep learning algorithms. But in some cases, the dataset is not filtered and preprocessed accurately. It is the second main reason to degrade the accuracy and high error rates.

IV. Existing work

This section describes the existing work with semantic knowledgebase ML and sentiment analysis using an optimized LSTM classifier.

A. Novel Hybrid-Semantic Knowledgebase ML method [17]

This learning technique combines the semantic knowledgebase with the MLmethodto increase the performance of the opinion miningprocedure. The proposed model has improved the core tasks of opinion mining, such as obtaining domain characteristics, linking them with their related attitudes, and opinion categorization. This technique analyses unstructured textual input, extracts domain characteristics then fuse those elements with appropriate feelings. The novel technique then computes the polarity for each related feature-sentiment combination and enters the resulting data into a modeled domain knowledgebase. The new approach uses the modeled domain knowledgebase to generate a semantic feature database, which was then combined with a statistical database and utilized as input to anMLmethod, which provides a multi-point scale rating for the managedsubstances. The researchnovelhybrid SK-ML method framework consists of the following components [18]:

- Knowledge-based Population
- NLP
- Emotion extraction
- Domain-based feature emotion association
- Polarity
- Knowledgebase enhancement

- Sentiment classification.

The created domain emotion vocabularies were used to exclude designswithexpressive opinions. The semanticallyorganized domain knowledgebaseused to bootstrap thedomainfeature extractionprocedure was enriched in the fourth stage with new semantic information related to the reviewed review and the conformingsemantically explained domain features, sentiments, and polarities. The resultant domain knowledgebase was an excellent resource for anticipating overall opinion about a domain and for refinedrecovery of views related to a given attribute. This method strategy for identifying the generalidea of the reviews on a multi-point scale was created in the fifth step—recent techniques for sentiment analysis and opinion mining that use ML and DL algorithms primarily concerned with categorization tasks. To the authors' knowledge, no practices suggest using a semantic knowledgebase method in conjunction with DL methods for classification tasks.

B. Sentiment Analysis Using Optimized LSTM

The author [19] proposed an efficient sentimental analysis framework. By combining OBL with the PSO method, the optimization problem was addressed. For word embedding, the skip grammar model-based feature extraction was applied. Compared to alternative Word to Vector representations, the Skip-gram word-to-vector form takes minimum memory space and consistently yields superior precision. Because of its high generalization capabilities, LSTM [18] may produce good results even without optimization. The APSO method is used to optimize the weight parameters of the LSTM to boost its performance even more. Using APSO to set weight metrics for LSTM- NN (neural networks) improves accuracy while decreasing computing complexity. Four types of datasets were employed for the experimental analysis. The suggested method's performance was evaluated using parameters such as accuracy, F-measure, etc. The recommended technique achieved extreme accuracy of 96.8 percent for the Amazon database. The suggested Methodology outperforms the competition in terms of empirical findings.

V. RESEARCH METHODOLOGY

Fig.1 defines the research model structure. The research model will show the various steps such as dataset collection (reviews), data preprocessing, and feature extraction of the different types of review analysis.

A. Data Collection and Pre-processing

Typically, it will collect the reviews with different categories such as positive, negative, etc. Most of the analyzers can collect the dataset from online websites or universities. After the collection step, it will upload the input review data and identify the feasibility of the dataset when the dataset will search. Before classifying the opinions, the subsequent phases are functional to the database:

- Tokenization is a procedure reviews are split into symbols, tokens, words, etc.
- Stop words like I, th/e, for, would, etc. These stopwords are removed from the NLP toolkit stop word list.
- The stemming process overcomes the word length to its base forms.

B. Feature Extraction, Selection, and Classification

After the preprocessing phase, the model will move to the next step, feature extraction. It will extract the semantic features and reliable optimized feature sets with the help of the feature extractionmethod using optimized KPCA. The extracted feature vector data is then processed with the feature optimization process. Here the optimized KPCA and adaptive PSO optimization process take place. This module processes the dataset, provides the optimized feature vectors for classification models, and reduces errors. The optimization process only improves the classified accuracy rate, precision, etc.

This feature extraction model extracts each FC (feature component) from a dataset; it should evaluate various KF (kernel function) as there are standard designs. It is more computational and practical to extract the feature patterns. It mainly works on space and time complexities—the adaptive optimized PSO with the CNN layers procedure with several processing metrics and output. The labels are the name of reviews where the feature sets take place. The training model will be utilized to classify the outcome for the testing phase. So, it gets saved inside the

current execution directory of the proposed model.

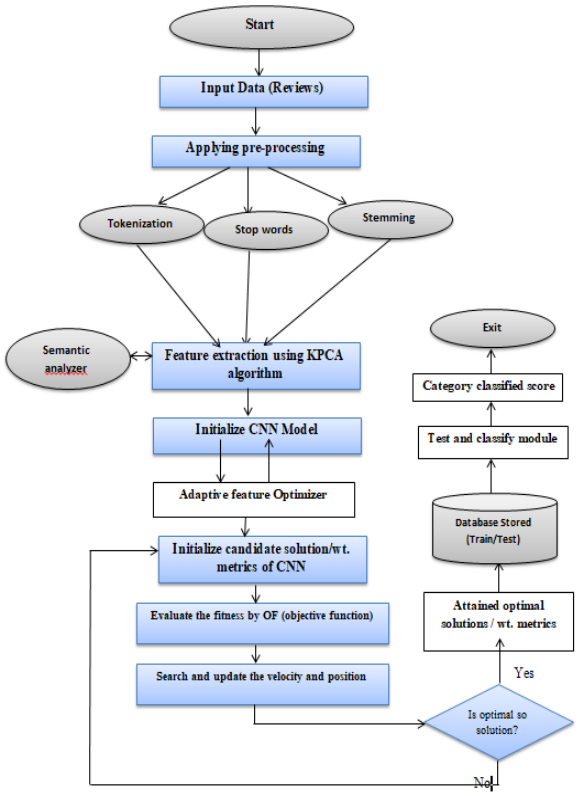

Fig. 1Proposed Hybrid Optimized Feature-based CNN AlgorithmFlowchart

Generally, the best output attained during optimization is measured to be the optimal solution to the issue search by the proposed model is called an optimized, featured PSO with the CNN model. This architecture is automatically analyzed the significant feature sets without human interaction. This model obtains the advantages of giving maximum efficiency and performance. At the end of the research system, several performance metrics are calculated and associated with the existing methods to search the improvements in the research model.

TABLE III. HOF-CNN PROPOSED ALGORITHM

Pseudo Code: HOF-CNN model
Input: Movie review parameters
Output: Classify the opinion (Positive, Negative). Start

Upload the database as DB in string format

Pre-process Step:

Tokenization process: Tk

D=Tk (str);

Stopwords Process: Sp

While (more Tk exit in token table) do

T=next_Tk;

If T in Sp

Then delete the entry of T from Tk;

End if

End while

Stemming Process: Stp

S1 = stem (T);

Feature Extraction Process:

Input

Sample S = S1, S2, S3…. Sn

No. of PCs (principal components) P

Evaluate normalize Kmatrix

C = top_V (Kmatrix,P)

H = Top_E (Kmatrix,P)

for i=1:P

for j = 1: P

$[L]_{I,j} = \frac{1}{\sqrt{H_j}} \sum_{k=1}^{M} Kmatrix\ (S1, Sn)\ C_{n.Kmatrix})$

end for

end for

return matrix of co-ordinates L

APSO Process:

do

start()

evaluate FFn (); // fitness function

while (stop_criteriais not satisfied)

do

if FFn>0

a_star();

else

evaluate best_pso_vel;

evaluate next_pos;

```
  end while
end if
 Needs: Train_set, Test_set,label
 Classify →classify or predict opinion
 Create network()
Initialize_net (network)
For err>=0
Do
Err_net_Train(network,train,label)
End for
Opt_feat= APSO(Train_set,label)
T_Train→HiddenLayer(network, Train_set)
Train_com←T_Train +Opt_feat
Model_hfo-cnn←HFO-CNN_Train(Train_com)
T_Test→HiddenLayer(network, Test)
Test_com←T_Test +opt_feat
Classify ← HFOCNN(Model_hfo-cnn,Test_com)
Acc ← calculate (final_test,classify)
Return( classify, acc)
```

Table 3 discusses the working of a proposed HOF-CNN algorithm. The proposed method defines several steps like dataset collection (movie reviews), data preprocessing, feature extraction, selection, and classification or prediction of the different types of opinions. This research work initially is designed to enhance the accuracy of the CNN classifier by discovering the part of reliable feature sets and evaluating the steadfast rules for kernel features for KPCA and Layers for the CNN method. To experience this, APSO-based improved structure is utilized. A HOF-CNN algorithm includes DL methods by enhancing the performance metrics of CNN using APSO.

VI. Experimental Analysis

The research opinion mining on movie database is implemented in MATLAB 2018 andWindows 2010 with 64-bit and 8 GB RAM. This experiment used a movie database. The training process used 80% with HOF-CNN (Hybrid Optimized Feature-based CNN), and 20% of the database was usedto test the model. Data pre-processing, feature extraction, and selection are utilized in the training database, and the CNN classifier is trained. Generally, data pre-processing and feature extraction are also completed on the test database. Lastly, the feature extraction from the test database is defined as the input to the trained model. During the presentation of the trained model, the feeling score of the effort movie reviews isexamined. After the prediction model (Positive, Negative, and Neutral) of input reviews, based onprecision, accuracy, etc. The subsequent section describes the opinion mining system performance parameters.

A. Dataset

The research has used the IMDb database [20] with 50,000 movie reviews for MLP or TAs (text analytics). IMDb database for binary sentiment classification comprises significantly more data than existing benchmark databases. The research work has provided a set of 25k highly polar movie reviews for the train and 25k for the test. Then, classify the num_of_reviews, such as positive or negative, using either DL method prediction.

B. Mathematical Parameters

The mathematical parameters like precision and accuracy are shown with the t_p, t_n, f_p, and f_n. Where t_p signifies true_positive, t_n signifies true_negative, f_p represents false_positive, and f_nsignifiesfase_negative. The mathematical performance parameters are shown as follows:

$$precision = \frac{t\_n}{f\_p + t\_p} \quad \text{..... (i)}$$

$$accuracy = \frac{(t_n + t\_p)}{(t\_n + t\_p + f\_n + f\_p)} \quad \text{..... (ii)}$$

C. Result Analysis of the Sentiment Analysis Classification

In this simulation, the research HOF-CNN method was calculated for the CNN method of movie reviews. The main objective is to control whether extra semantic and texture features can outcome in a high-performance classification accuracy rate. The semantic and texture features were created from the IMDb movie sentiment knowledge domain after data pre-processing the reviews by the HOF-CNN deep learning method.

IMDb movie database was feedback into the SVM[21], and NB [22] methods and a hybrid optimized feature-CNN modelpresentation were calculated. Both ways were regulated using different layers.

To represent the research method's efficiency, various comparison analyses were undertaken. The better selection of FE (feature extraction) and FS (feature selection) methods pays well for better opinion mining classification tasks.

TABLE IV. METRICS UTILIZED IN THIS RESEARCH WORK

Hyper metrics	**Values**
Method	CNN
Optimization	A-PSO (Adaptive-Particle Swarm Optimization)
No. of Layers	3
Loss_method	MSE (mean square error)
Learning_rate	0.001
Hidden_unit	20
Training set	80
Testing set	20
Performance metrics	Precision Recall

The metric situation for the opinion mining model is defined above in Table 3, which shows the values of hyper metrics utilized in this experiment analysis. The CNN model hidden unit is set to 20. The learning_rate is 0.001, training_set used 80%, and testing_set used 20%.

TABLEV. PROPOSED PARAMETER ANALYSIS OF HYBRID FEATURE OPTIMIZED-CNN CLASSIFIER AT IMDB MOVIE DATABASE

Parameters	**Values**
Precision (%)	90.1
Accuracy (%)	97.0

The simulation outcomes using the IMDB movie database, as in table 4. The HOF-CNN method chooses the optimal metrics to improve the kernel's proposed performance. As defined in table 6, the research approach obtained a high accuracy rate of 97.0 percent and a precision of 90.12 percent. It is due to optimized selection using the A-PSO method. The consistent figure representation is defined in fig 2.

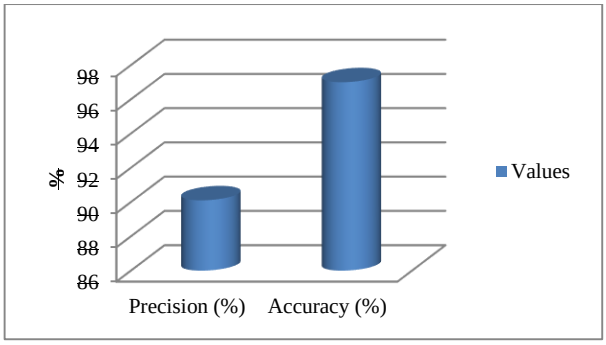

Fig 2. Simulation result analysis using IMDB movie dataset

TABLE VI. ACCURACY AND PRECISION OF DIFFERENT CLASSIFIER MODELS

Approaches	**Accuracy (%)**	**Precision (%)**
Hybrid optimized feature-based CNN model	97.0	90.1
SVM [21]	91.8	81.0
NB [22]	-	70

Due to SVM [21], NB[22], and HOF-CNN models, better outcomes may be attained. Fig 3 defines the result achieved using the IMDb movie database. The research model got a maximum accuracy rate of 97.0 percent.

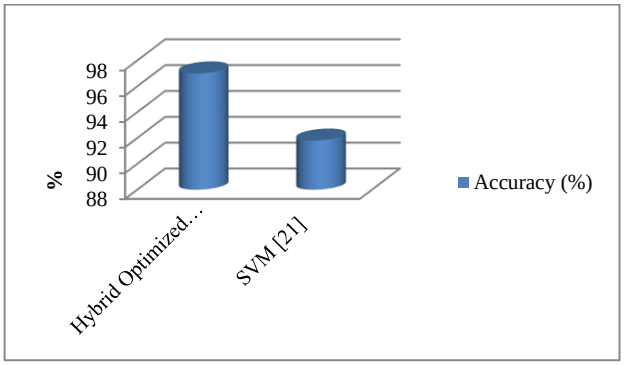

Fig 3. Comparative Analysis: Accuracy Rate (%)

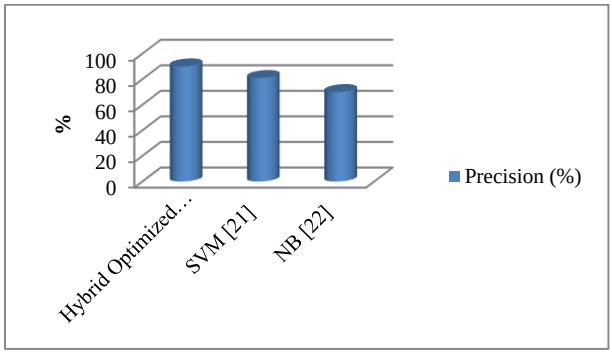

Fig 4. Comparative Analysis: Precision (%)

Also, the research method obtained a more precision rate of 90.1 percent. Comparison analysis using the IMDb movie database is defined in fig 3 and 4. The research approach receives the maximum accuracy rate for examining the IMDb movie database compared to models such as SVM and NB. The graphical demonstration of the IMDb movie review database is defined in fig 3 and 4.

VII. Conclusion and Future Scope

Reviews play a significant role in supporting user decisions about services or purchasing products, etc. Opinion mining simulation tools are required to give consumers an effective procedure of many online opinions or reviews. In this research article, the hybrid optimized feature-based CNN model has been implemented where a KPCA has been used as a feature extraction approach. It uses informationabout the selected domain and reliable community-connected open data (IMDb movie database) to develop the accuracy of the domain feature extraction and selection job. The proposed method also enhances the precision of calculating the movie opinions by developing a novel hybrid optimized feature-based CNN model that relies on a created sentiment lexicon for the unique feature. Lastly, our novel method improves opinion classification's precision and accuracy rate.

An effective OM outline is implemented in this proposed work. The optimization issue is resolved by addingthe A-PSO method. The KPCA reduction dimensionality-based feature extraction is utilizedto extract the reliable feature sets. KPCA method needs minimum memory space and high accuracy as compared with others. The proposed work improves the HOF-CNN method's performance and is used to extract the features, optimizing the weight metrics of the CNN model. The main objective of A-PSO is to choose weight metrics for the CNN model, increase the accuracy rate and minimize the complexities. For simulation analysis, the IMDb movie database has been utilized. The research technique has been assessed using accuracy, precisionetc. The research methodology obtained the maximum accuracy rate of 97.0 percent for the movie database.

Further work will prpose and develop the SK (semantic knowledgebase), Bi-LSTM, and RNN methods on massive databases to facilitate the requirement for DL methods and to calculate the impact of the proposed method when crossbred with DL instead Bi-LSTM method. The foremost objective of this study is to enhance the superiority of the training database for classification jobs utilizing an SK method.

References

1. Deng, Zhi-Hong, Kun-Hu Luo, and Hong-Liang Yu. "A study of supervised term weighting scheme for sentiment analysis." *Expert Systems with Applications* 41, no. 7 (2014): 3506-3513.
2. Martín-Valdivia, María-Teresa, Eugenio Martínez-Cámara, Jose-M. Perea-Ortega, and L. Alfonso Ureña-López. "Sentiment polarity detection in Spanish reviews combining supervised and unsupervised approaches." *Expert Systems with Applications* 40, no. 10 (2013): 3934-3942.
3. Min, Hye-Jin, and Jong C. Park. "Identifying helpful reviews based on customer's mentions about experiences." *Expert Systems with Applications* 39, no. 15 (2012): 11830-11838.
4. Eirinaki, Magdalini, ShamitaPisal, and Japinder Singh. "Feature-based opinion mining and ranking." *Journal of Computer and System Sciences* 78, no. 4 (2012): 1175-1184.
5. Kontopoulos, Efstratios, Christos Berberidis, TheologosDergiades, and Nick Bassiliades. "Ontology-based sentiment analysis of Twitter posts." *Expert systems with applications* 40, no. 10 (2013): 4065-4074.
6. Dalvi, Bhavana, EinatMinkov, Partha P. Talukdar, and William W. Cohen.

"Automatic gloss finding for a knowledge base using ontological constraints." In *Proceedings of the eighth ACM international conference on web search and data mining*, pp. 369-378. 2015.

7. Polpinij, Jantima, and Aditya K. Ghose. "An ontology-based sentiment classification methodology for online consumer reviews." In *2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, vol. 1, pp. 518-524. IEEE, 2008.
8. Sulthana, A. Razia, and Ramasamy Subburaj. "An improvised ontology-based K-means clustering approach for classification of customer reviews." *Indian Journal of Science and Technology* 9, no. 15 (2016): 1-6.
9. Zervoudakis, Stefanos, EmmanouilMarakakis, HaridimosKondylakis, and Stefanos Goumas. "OpinionMine: A Bayesian-based framework for opinion mining using Twitter Data." *Machine Learning with Applications* 3 (2021): 100018.
10. Abdelminaam, Diaa Salama, Nabil Neggaz, Ibrahim Abd Elatif Gomaa, Fatma Helmy Ismail, and Ahmed A. Elsawy. "ArabicDialects: An efficient framework for Arabic dialects opinion mining on Twitter using optimized deep neural networks." *Ieee Access* 9 (2021): 97079-97099.
11. Keyvanpour, Mohammadreza, Zahra Karimi Zandian, and Maryam Heidarypanah. "OMLML: a helpful opinion mining method based on lexicon and machine learning in social networks." *Social Network Analysis and Mining* 10, no. 1 (2020): 1-17.
12. PANDEY, HARI MOHAN, Prayag Tiwari, Aditya Khamparia, and Sachin Kumar. "Twitter-based opinion mining for flight service utilizing machine learning." *Informatica (Slovenia)* 43, no. 3 (2019): 381-386.
13. Razali, Noor Afiza Mat, Nur AtiqahMalizan, Nor AsiakinHasbullah, MuslihahWook, NorulzahrahMohd Zainuddin, Khairul Khalil Ishak, Suzaimah Ramli, and SazaliSukardi. "Opinion mining for national security: techniques, domain applications, challenges and research opportunities." *Journal of big data* 8, no. 1 (2021): 1-46.
14. Dubey, Kamanksha Prasad, and Sanjy Agrawal. "An opinion mining for Indian premier league using machine learning techniques." In *2019 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU)*, pp. 1-4. IEEE, 2019.
15. Shobana, J., and M. Murali. "An efficient sentiment analysis methodology based on long short-term memory networks." *Complex & Intelligent Systems* 7, no. 5 (2021): 2485-2501.
16. Alfrjani, Rowida, Taha Osman, and Georgina Cosma. "A hybrid semantic knowledgebase-machine learning approach for opinion mining." *Data & Knowledge Engineering* 121 (2019): 88-108.
17. Osimo, David, and Francesco Mureddu. "Research challenge on opinion mining and sentiment analysis." *Universite de Paris-Sud, Laboratoire LIMSI-CNRS, Bâtiment* 508 (2012).
18. Shobana, J., and M. Murali. "An efficient sentiment analysis methodology based on long short-term memory networks." *Complex & Intelligent Systems* 7, no. 5 (2021): 2485-2501.
19. Sagnika, Santwana, Bhabani Shankar Prasad Mishra, and Saroj K. Meher. "An attention-based CNN-LSTM model for subjectivity detection in opinion-mining." *Neural Computing and Applications* 33, no. 24 (2021): 17425-17438.
20. IMDB Dataset of 50K Movie Reviews. (2022). Retrieved 22 June 2022, from https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews.
21. Tehrany, MahyatShafapour, Biswajeet Pradhan, ShattriMansor, and Noordin Ahmad. "Flood susceptibility assessment using GIS-based support vector machine model with different kernel types." *Catena* 125 (2015): 91-101.
22. Domingos, Pedro, and Michael Pazzani. "On the optimality of the simple Bayesian classifier under zero-one loss." *Machine learning* 29, no. 2 (1997): 103-130.