

Interpretable Deep Convolutional Neural Network Framework for Online Reputation Management Using Customer Reviews and Social Media Feedback

An Interpretable Deep Learning Approach for Online Reputation Management

Abdullah Saad AL-Malaise^{1*}, Bandar A. Alrami Alghamdi², Mahmoud Ragab³

Faculty of Computer Studies, Arab Open University, Riyadh, Saudi Arabia^{1,2}

Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia³

Abstract—Social media has become a significant channel for communication and interaction, significantly influencing how organizations are perceived in the digital era. Online reputation management (ORM), as an addition to public relations, concentrates on shaping and maintaining an organization's image through continuous surveillance of customer reviews and online feedback. Traditional methodologies, such as machine learning and deep learning-based reputation management, depend upon learned patterns and predetermined features from historical information; recent ORM approaches exploit real-world data from social platforms to extract dynamic public sentiment. Meanwhile, online opinions can transform rapidly because of human-generated text, and preserving a reliable reputation has become progressively difficult for organizations. To improve interpretability and transparency, explainable AI algorithms are frequently incorporated into recent ORM architectures. This study presents an Interpretable Deep Convolutional Neural Network Framework for Online Reputation Management (IDCNN-ORM). The proposed model integrates an improved mRMR approach to identify highly informative and non-redundant features. Moreover, the deep convolutional neural network is applied to learn hierarchical feature representations from input data. The Adabelief optimizer is integrated into the proposed framework to enhance training stability by adaptively refining the learning process. Furthermore, the LIME-based explainability artificial intelligence technique is utilized to enhance transparency in the model's decision-making process. The experimental analysis of the proposed IDCNN-ORM model is carried out using the Social Media Sentiments Analysis Dataset. The comparative analysis demonstrates promising performance with an accuracy of 94.85% compared with existing methodologies.

Keywords—Online reputation management; social media feedback; adabelief optimizer; explainable artificial intelligence; deep convolutional neural network

I. INTRODUCTION

Social media networks have developed into an important platform where customers can express their sentiments, engage with brands, and share their experiences in the modern world [1]. The velocity and volume of user-generated content (UGC) on applications like LinkedIn, Facebook, YouTube, Twitter, and Instagram offer a valuable data resource that can be used for

*Corresponding author. several marketing purposes.

Among these, Online reputation management (ORM) is the process of improving, tracking, observing, and handling an individual's or organization's reputation in the digital era [2]. Nowadays, UGC plays an essential part in influencing public perception through blog comments, feedback, ratings, posts, forums, and reviews. Although UGC has enhanced data sources, it is difficult to extract useful and meaningful insights from the volume and the considerable difficulty [3]. Reading and analyzing the sentiment of all relevant online texts requires huge effort and time, and it is frequently quite subjective. Experts struggle with a huge amount of information, unreliable language, and complex emotional interpretation [4]. This type of interpretation is ineffective and restricts a firm's capability to gather real-world insights, respond quickly to novel issues, and evaluate opinions through large-scale databases or prolonged periods. Inefficient processing of data streams can result in a failure to understand consumer requirements, enhance promotional measures, and prevent damage to brand reputation [5].

Over the past few decades, sentiment analysis (SA) has evolved as a vital domain within machine learning (ML) and natural language processing (NLP), driven by the explosive growth of digital information [6]. Given the applications of social media networks, customer opinion, online reviews, and feedback, comprehending the sentiments extracted in a text is important [7]. SA, termed opinion mining, is the computing operation of classifying, mining, and evaluating subjective data from a text or content to determine the complete attitude or sentiment expressed toward specific products, entities, events, or topics [8]. This method includes identifying text into pre-defined classes like neutral, positive, and negative sentiments, or more refined opinions and emotions. Usually, ML-driven SA methods are extended, and novel gaps are classified by SA identification. There is a move toward using more intricate methods to solve this limitation [9]. Deep learning (DL) models are a branch of the ML framework that provide effective solutions to these tasks with structures that involve recurrent neural networks (RNNs) and Long Short-Term Memory (LSTM), indicating efficiency in managing intricate SA challenges. These methods have gained attention for their capability of capturing the contextual interaction within text and have exhibited better performance compared with other DL models in sentiment analysis tasks [10].

Online Reputation Management (ORM) has become important owing to the huge increase in social media feedback and customer comments on digital platforms. Organizations need an automatic strategy to examine a huge amount of noisy and unstructured textual data in the real world. The Kaggle social media sentiment analysis dataset signifies real-time sentiment demonstrations from distinct online sources. Conventional ML systems frequently fail to extract semantic relationships and deep contextual information in certain cases. Therefore, there is a need for innovative DL-based techniques to enhance feature representation and sentiment classifier outcomes in ORM models. This motivates the proposed Interpretable Deep Convolutional Neural Network Framework for Online Reputation Management (IDCNN-ORM) using social media data.

- Utilize an improved mRMR approach to effectively select the most relevant and non-redundant features from the dataset.
- Design a deep convolutional neural network (DCNN) approach to learn hierarchical feature representations from input data by continuously capturing complex patterns at multiple levels to improve classification performance.
- Present an Adabelief optimizer for adaptively modifying learning rates, thereby ensuring stable and efficient training.
- Integrate a LIME-based explainable artificial intelligence (XAI) model for providing interpretability and improving transparency in the decision-making process.
- Evaluate the performance of the proposed model using the Social Media Sentiments Analysis Dataset, demonstrating its effectiveness under various evaluation measures.

II. LITERATURE REVIEW

This section concisely reviews the existing research on online reputation management. Semenda et al. [11] discussed the objective of matrix-based examination and game theory in improving social media marketing algorithms for e-commerce industries. To integrate these mathematical techniques with social media analytics, the goal is to deliver a comprehensive architecture that can predict customer action. This system allows sellers to organize more adaptive, targeted, and efficient marketing campaigns, drive growth, and improve customer fulfillment in the digital marketplace. Istanbuluoglu and Sakman [12] examined the associations among business responses to customer repurchase intentions and social media complaints. The connection between purchase intention and the five scopes of handling complaints, as led by customer trust in the corporation, was surveyed, and users' propensity to trust was hypothesized as a mediator in this correlation. Saheb et al. [13]

examined the connection between social media and AI. The study employs a mixed-model strategy to examine 1540 scholar records, incorporating systematic literature review and bibliometric approaches. The major objective of this study is to detect the most significant themes and trends, and also the impact of company values and suggestions, in the AI Social Media field. The initial level of the study encompassed a quantitative keyword analysis, which led to the identification of ten key themes.

Alsemaree et al. [14] offer the requirement for a nuanced consideration of consumer sentiments under several coffee goods, allowing corporations to generate informed decisions concerning improvement, discontinuation, and product promotion. Antczak [15] discusses the major goals for exploring the connections among social media marketing, digital marketing, and customers' online purchasing actions. The study will investigate how marketing actions in social media and digital media, namely social media advertising, product reviews, influencer marketing, and e-mail marketing, impact user behavior and purchase choices. The authors [16] proposed a systematic literature review (SLR) investigating the role of SMIs in the e-commerce market. This study examines recent knowledge on how influencer authority, demarcated by expertise, trustworthiness, and reliability, impacts user behavior, purchase intention, and brand engagement. Odionu et al. [17] introduced the use of big data analytics for enhancing CRM engagement and retention approaches. The main goal is to examine how big data analytics can be exploited to gain insights into consumer actions, interactions, and preferences, thus enhancing CRM proposals. This study contains a comprehensive literature review and case study examination of organizations that have effectively incorporated big data analytics into their CRM methodology.

Existing sentiment analysis systems for ORM are frequently limited in interpretability, making their forecasting complex to trust in real applications. Several strategies fail to efficiently manage noisy and informal social media content, namely, emojis, abbreviations, and slang. The usage of CNN-based frameworks to capture hierarchical textual attributes is still unexplored in comparison with DL algorithms. Furthermore, more research focuses on accuracy without combining explainability methodologies. There is a limited number of unique architecture that integrates higher results with method transparency.

III. METHODOLOGICAL FRAMEWORK

The proposed IDCNN-ORM model follows a comprehensive pipeline for improving a reputation system using social media data. Fig. 1 shows the overall architecture of the proposed model. As depicted in Fig. 1, the methodology involves five major stages: data preprocessing, feature selection, classification, parameter tuning, and explainability analysis. These processes are integrated to ensure effective model performance and accurate sentiment-based reputation analysis.

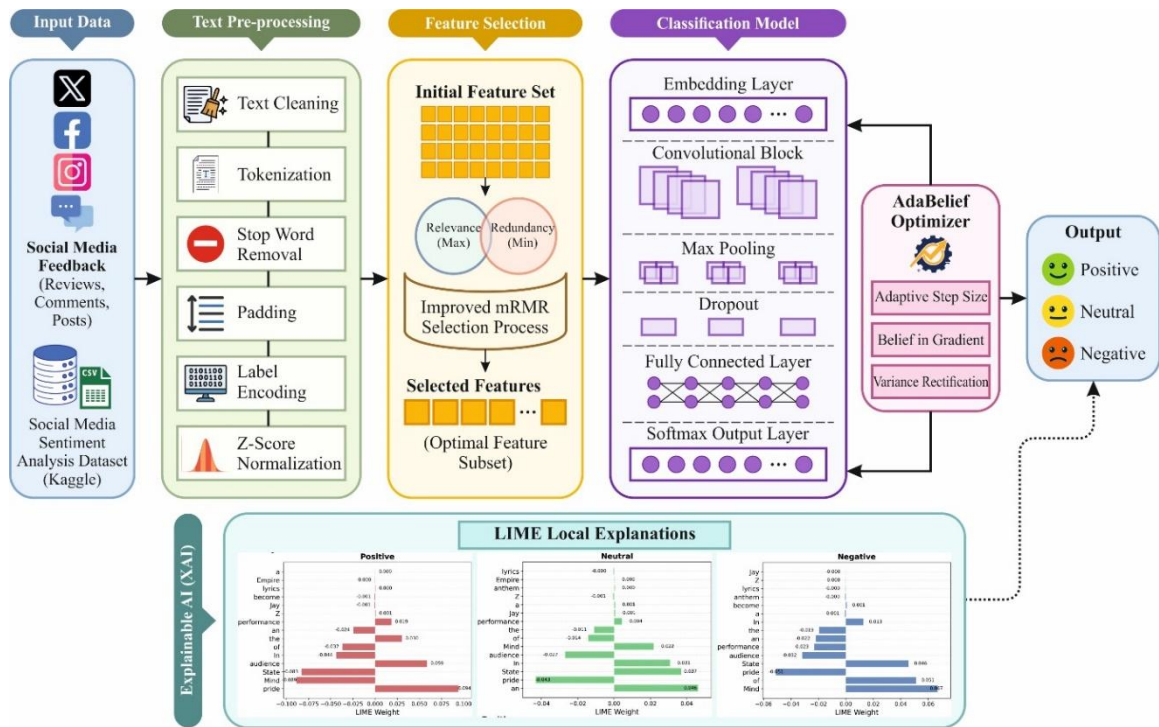


Fig. 1. Overall architecture of the IDCNN-ORM model.

A. Data Preprocessing

A multi-stage preprocessing stage includes text cleaning, tokenization, stop-word removal, and padding to convert raw data into a structured format [18].

1) *Text cleaning*: It eliminates hashtags symbols (#), URLs, punctuation, emojis, and user mentions from the "Text" column, and special characters. It ensures the preservation of significant expressive words from social media posts.

2) *Tokenization*: It divides content into smaller units like subwords or words referred to as tokens, while handling emojis, hashtags, special characters, and URLs to extract meaningful phrases for examination.

3) *Stop-word removal*: It removes common words like "the", "and", and "is" that do not enhance meaning. This helps to lessen data dimensionality and increases processing efficiency.

4) *Padding*: It is used to pad token-based post sequences to guarantee a unified input length for every instance. This is necessary for the DL strategy; later posts in the dataset are different in structure and length.

Before model training, the proposed IDCNN-ORM framework performs text cleaning, tokenization, stop-word removal, padding, label encoding, and Z-score normalization to preprocess the social media text data.

5) *Label encoding*: This encoding approach is leveraged to convert categorical variables into numerical form. It allocates a unique integer value to every category, allowing effective learning from categorical data.

Z-score normalization (standardization) converts information; thus, it has a standard deviation of 1 and a mean of 0. During training, it is utilized to scale each attribute to a standard scale, which improves system performance, faster convergence, and stability.

$$X_{norm} = \frac{X - \mu}{\sigma} \quad (1)$$

Here, μ means the mean of the feature, X signifies the actual feature value, X_{norm} indicates the scaled value, and σ implies the standard deviation of the feature.

B. Improved mRMR-Based Feature Selection

An improved mRMR-based feature selection (FS) strategy is applied to select the most relevant and non-redundant attributes from the dataset. It reduces feature redundancy and computational complexity while improving the effectiveness of the classification process [19]. The standard mRMR is an effective FS strategy that repeatedly increases the number of attributes in the feature subset by a gradual search that balances the dual terms: minimal redundancy to the gathered feature subset and maximum relevance to the label. The dual relevance is computed.

$$\max A(S, C) = \frac{1}{n} \sum_{f_i \in S} I(f_i, C) \quad (2)$$

$$\min R(S) = \frac{1}{n^2} \sum_{f_i, f_j \in S} I(f_i, f_j) \quad (3)$$

here, C signifies the data label, S denotes the chosen feature subset. $A(S, C)$ refers to the correlation between the chosen subset and the label, f_i and f_j indicate the dual attributes

features in S , $I(*)$ represents the mutual information, and $R(S)$ implies the redundancy among the chosen subset.

$$I(X, Y) = \int_X \int_Y P(X, Y) \log \frac{P(X, Y)}{P(X)P(Y)} \quad (4)$$

As an instance, $P(X, Y)$ refers to their marginal probability distributions. Conversely, its mutual data-driven importance assessment is more appropriate for diverse information. These dual behaviors can be examined to manage continuous data, and it has been established that they can deliver information. Pearson correlation and T -test coefficient equations are defined.

$$t(f_i) = \frac{|\overline{f_{i_{pos}}} - \overline{f_{i_{neg}}}|}{\sqrt{\frac{S_{i_{pos}}^2}{n_{pos}} + \frac{S_{i_{neg}}^2}{n_{neg}}}} \quad (5)$$

$$\rho(f_i, f_j) = \frac{\sum (f_i - \overline{f_i})(f_j - \overline{f_j})}{\sqrt{\sum (f_i - \overline{f_i})^2 \sum (f_j - \overline{f_j})^2}} \quad (6)$$

Here, $\overline{f_{i_{neg}}}$ implies the mean value of f_i in the negative instance; $\overline{f_{i_{pos}}}$ represents the mean value of the positive instance, n_{pos} and n_{neg} indicate the sample counts; $S_{i_{pos}}^2$ and $S_{i_{neg}}^2$ refer to the variance of f_i .

$$f_i \max \left[t(f_i) - \frac{1}{S} \sum f_j \in F_s \rho(f_i, f_j) \right] \quad (7)$$

Additionally, the stopping principle of the standard mRMR methodology for increasing the count of chosen feature subgroups, which needs to be set analytically, generates limitations to ensure the outcome of the last chosen feature subgroups.

C. Online Reputation Classification via DCNN

A deep convolutional neural network is used to analyze sentiment patterns in social media data for effective reputation assessment. The model utilizes deep convolutional layers to learn hierarchical feature representations and improve classification [20]. A neural network (NN) is an effective method comprising several interrelated nodes; here, every node uses an activation function to produce an output. The relationships among nodes carry weights that verify how intensely data is distributed among neurons and serve as the network memory. The output of the network, based on these weights, is learned, and correlations are structured. During Online Reputation Management (ORM), the major features impacting reputation include opinion strength, sentiment polarity, variations in customer reviews, and user engagement and social media comments. Conversely, these components cannot exactly signify the last reputation category, namely neutral, positive, or negative, creating a classification problem utilizing raw information.

The actual reputation grade is derived from consumer feedback and social media text, which precisely imitates public opinion. Hence, these data sources are employed as key inputs for classifying reputation into multiple classes. The DCNN

progress comprises capturing attributes from the data and upgrading system hyperparameters through learned extractions. These learned hyperparameters are assigned to different DCNN sub-methods to reduce computation, while preserving results. The n DCNN sub-models for learning hyperparameters are represented by.

$$\theta^S = \{\theta_1^S, \theta_2^S, \dots, \theta_n^S\} \quad (8)$$

where, each θ_n^S corresponds to one sub-model, and θ^S denotes the combined parameters. All sub-models encompass three fully connected layers and five convolution layers to study patterns from social media text and consumer feedback. This transfer learning system assists in transferring knowledge from a source dataset to the targeted ORM classifier study, minimizing the requirement for huge labeled data while sustaining superior precision. Subsequently, transfer learning, n DCNN-TL sub-models are designed, and an ensemble system is created utilizing these sub-models together with one classification layer and one fully connected layer.

The convolution layer captures significant features by employing filters across the input text extraction. The convolution output at position (i, j) for the k_{th} filter.

$$\begin{cases} Z_{i,j,k}^{l_{conv}} = C(X, K)_{i,j,k} = \sum_{r=1}^{k_h} \sum_{s=1}^{k_w} \sum_{t=1}^{k_c} x_{i',j',t'} k_{r,s,t',k} + b_k \\ i' = (i-1)S_h + r \\ j' = (j-1)S_w + s \end{cases} \quad (9)$$

In Eq. (9), b_k signifies the bias term, and k denotes the filter weights. The convolution layer helps to capture suitable patterns from text, including contextual meaning and sentiment words. These features are applied to the classifier for reputation as neutral, negative, or positive.

The method repeatedly updates the feature extraction depending on observed information. During training, it improves forecasting by comparing predicted and actual class labels and reducing errors.

A major benefit of DCNN is that it can manage variable-length text information from comments and social media without strict input size challenges. It acquires valuable patterns from noisy and unorganized text data and transforms them into essential features for classification.

$$q(\chi_k | x_{0:k}(i), y_{1:k}) = p(\chi_k | \chi_{k-1}(i)) \quad (10)$$

During learning, this supports the model in concentrating on the most essential instances. If the count of efficient records becomes too small, then the algorithm is restarted to maintain results. Afterward, resampled, a novel group of attributes is formed and employed for the final classification.

$$w_t^{(i)} = (w_{t-1}^{(i)})^\alpha \frac{p(z_t | x_t^{(i)}) p(x_t^{(i)} | x_{t-1}^{(i)})}{q(x_t^{(i)} | x_{t-1}^{(i)}, z_{1:t})} \quad (11)$$

where, α influences how strongly weights are upgraded. Finally, the system is evaluated on the basis of detection precision and computational effectiveness. An added balancing parameter is applied to control imbalance among sentiment

categories. The proposed techniques efficiently classify online reputation from user feedback and social media information, utilizing ensemble transfer learning and DCNN-based feature learning. Fig. 2 denotes the overview of the DCNN model for reputation classification.

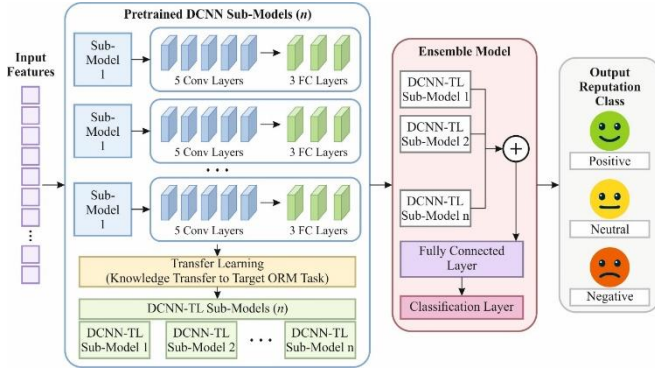


Fig. 2. Overview of the DCNN model for reputation classification.

D. Hyperparameter Optimization

AdaBelief optimizer is an adaptive optimization algorithm that adjusts learning rates based on the difference between predicted and observed gradients [21]. Adaptive optimizers, namely Adam, converge more quickly than SGD; however, their consistency and outcome might not be sufficient for larger datasets. This optimizer proposes an alteration to the Adam optimization for improving model generalization and training stability. During Adabelief's dissimilar Adam, a novel variable s_t is defined as the exponential moving average of $(g_t - m_t)^2$, while $g_t - m_t$ denotes the deviation. A smaller variance amongst g_t and m_t refers to a robust belief, while in the case of a weaker belief, the transformation is smaller. According to this methodology, the Adabelief measures s_t and biased correction $\hat{s}_t$ than v_t and $\hat{v}_t$,

$$s_t = \beta_2 s_{t-1} + (1 - \beta_2)(g_t - m_t)^2 + \varepsilon, \quad (12)$$

with its bias-corrected form represented by:

$$\hat{s}_t = \frac{s_t}{1 - \beta_2^t} \quad (13)$$

and the AdaBelief upgrade principle has been illustrated by

$$w_{t+1} = w_t - \frac{\eta}{\sqrt{\hat{s}_t} + \epsilon} \hat{m}_t \quad (14)$$

AdaBelief leveraged the same parameter settings as Adam, with identical default values.

E. Model Interpretability

LIME is an extensively used technique in the field of XAI, which intends to deliver transparent explanations for independent forecasting of intricate ML approaches [22]. It works on the principle that it can be approximated more easily by explainable strategies. Therefore, a model-agnostic system, LIME, is developed across a wide range of classifications and regressions without demanding access to the internal pattern of the algorithm. The major goal is to produce local approximators of a system's action in the specified occurrence to construct an interpretable method, usually a linear regression strategy.

$$\mathcal{L}(f, g, \pi_x) + \Omega(g) \quad (15)$$

Here, g represents a group of simpler approaches, $\pi_x(z)$ refers to a proximity among z and the χ , $\mathcal{L}(f, g, \pi_x)$ denotes the fidelity of g .

$$\pi_x(z) = \exp\left(-\frac{D(x, z)^2}{\sigma^2}\right) \quad (16)$$

where, σ regulates the width of the neighborhood, and $D(x, z)$ implies a distance metric, $\Omega(g)$ represents a complexity penalty.

$$\mathcal{L}(f, g, \pi_x) = \sum_{z \in Z} \pi_x(z) \cdot (f(z) - g(z))^2 \quad (17)$$

Here, Z signifies a group of perturbed instances in χ .

IV. PERFORMANCE ANALYSIS AND DISCUSSION

The experimental assessment of the IDCNN-ORM technique is achieved using the Social Media Sentiments Analysis Dataset [23]. This dataset captures a vibrant tapestry of emotions, tendencies, and relationships across diverse social media platforms. This dataset offers images of user-generated content, hashtags, combined text, timestamps, retweets, likes, and nations. Every entry reveals unique stories, namely moments of admiration, excitement, surprise, thrill, contentment, and many more exchanged by individuals globally. The dataset comprises three classes: positive, negative, and neutral, as denoted in Table I and Table II, which demonstrate the sample social media content used for sentiment analysis.

TABLE I. DETAILS OF DATASET

Class Label	Count
Positive	455
Neutral	112
Negative	165
Total	732

TABLE II. SAMPLE CONTENT

S.no	Text	Sentiment
1	Reflecting on the past and looking ahead.	Positive
2	Trying out a new dinner recipe tonight.	Neutral
3	Traffic was terrible this morning.	Negative

TABLE III. SENTIMENT ANALYSIS OF THE IDCNN-ORM METHOD WITH 70% OF TRAIP

Classes	Accur _y	Preci _n	Recal _i	F1	ROC _{AUC}	MCC
TRAIP (70%)						
Positive	91.21	91.34	95.03	93.15	89.88	81.03
Neutral	96.68	91.04	84.72	87.77	91.68	85.93
Negative	91.02	82.73	77.12	79.82	86.15	74.13
Average	92.97	88.37	85.62	86.92	89.24	80.36

Table III and Fig. 3 present the sentiment analysis of the IDCNN-ORM approach with three classes for 70% of TRAIP. The proposed IDCNN-ORM method achieved an average value with *accuracy* of 92.97% and MCC of 80.36%, implying its effectiveness as a social media sentiment classifier model.

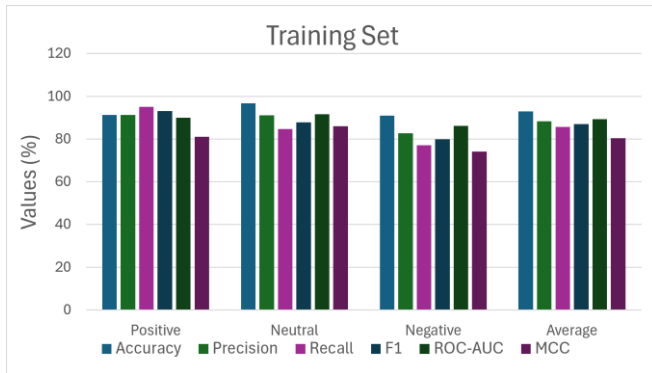


Fig. 3. Performance of the IDCNN-ORM approach with 70% of the training set.

Table IV and Fig. 4 display the sentiment analysis of the IDCNN-ORM method with three classes for 30% of TESTP. The proposed IDCNN-ORM approach achieved an average value of *accuracy* of 94.85% and an MCC of 86.09%, denoting its efficacy in the social media sentiment classification task.

TABLE IV. SENTIMENT ANALYSIS OF THE IDCNN-ORM METHOD WITH 30% OF TESTP

Classes	Accur _y	Preci _n	Recal _i	F1	ROC _{AUC}	MCC
TESTP (30%)						
Positive	93.18	93.38	95.49	94.42	92.57	85.69
Neutral	96.36	92.11	87.50	89.74	92.92	87.58
Negative	95.00	89.13	87.23	88.17	92.17	85.01
Average	94.85	91.54	90.07	90.78	92.55	86.09

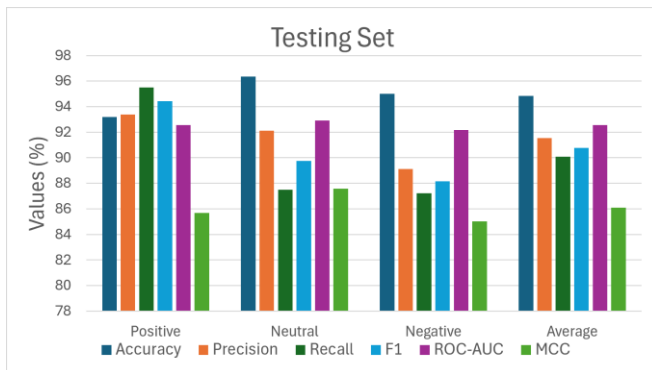


Fig. 4. Performance of the IDCNN-ORM approach with 30% of the testing set.

Fig. 5 depicts the training (TRAINING) and validation (VALIDN) *accuracy* curves of the IDCNN-ORM technique. The TRAINING *accuracy* continually increases, indicating effective learning and correct convergence of the system. The VALIDN *accuracy* follows the training curve with small differences,

suggesting higher generalization performance and lower overfitting.

Fig. 6 exemplifies the TRAINING and VALIDN loss curves of the IDCNN-ORM method. The losses decline quickly in the earlier epochs and then decrease smoothly, representing effective learning and appropriate convergence. The minor differences amongst the TRAINING and VALIDN curves exhibit less overfitting and superior generalization outcomes.

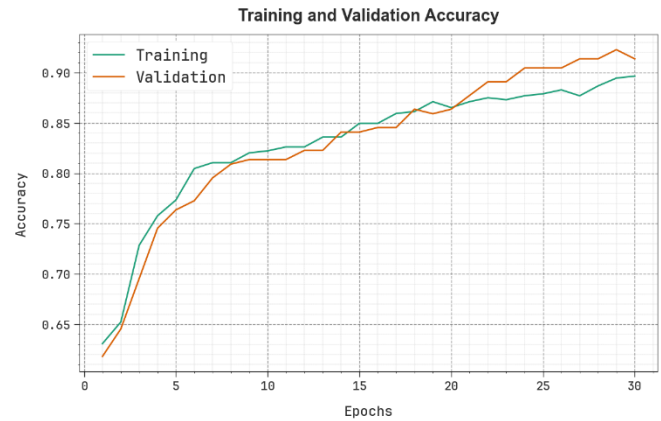


Fig. 5. *Accuracy* curve of the IDCNN-ORM method.

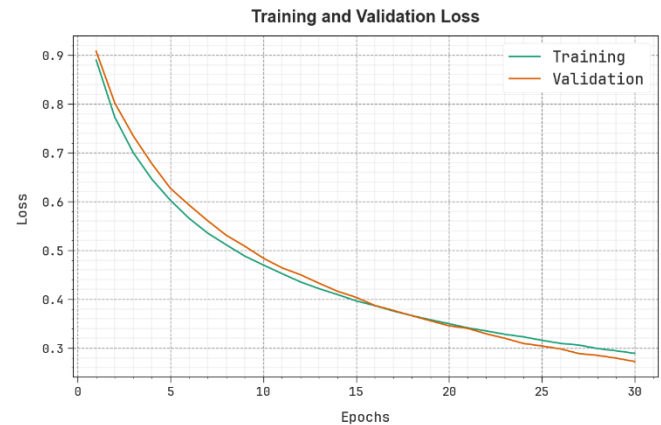


Fig. 6. Loss curve of the IDCNN-ORM approach.

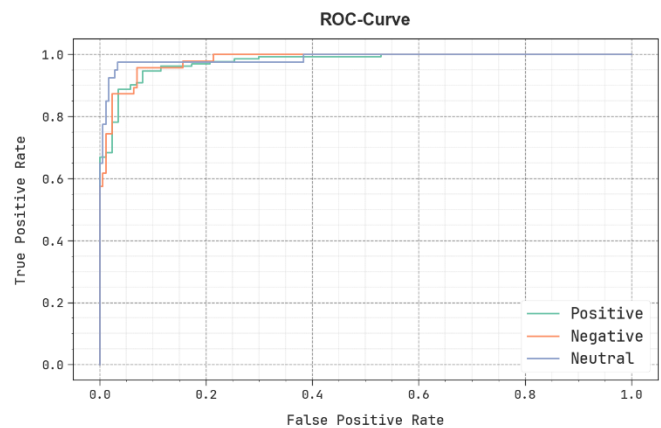


Fig. 7. ROC curve of the IDCNN-ORM method.

Fig. 7 portrays the ROC curve evaluation of the IDCNN-ORM system, representing boosted performance across all classes, signifying its strong discriminative ability. The consistently higher ROC values indicate the efficiency of the model in differentiating among class boundaries, and this steady pattern across numerous categories assures its consistency and robustness in the classification process.

Table V and Fig. 8 depict the comparison outcome of the IDCNN-ORM approach with existing methodologies using evaluation metrics [24-26]. The results imply that the SVM, LSTM, and DistilBERT approaches have lower performance, whereas the RoBERTa, CNN, and RNN methods have slightly greater values. Moreover, the CNN-LSTMHybrid has obtained moderate performance. At the same time, the proposed IDCNN-ORM approach has achieved superior performance with $accu_r$ of 94.85%, $preci_n$ of 91.54%, $recal_l$ of 90.07%, and F1 of 90.78%, emphasizing its robustness of the classifier tasks.

TABLE V. COMPARATIVE RESULTS OF THE IDCNN-ORM METHOD WITH OTHER APPROACHES

Methods	Accur _y	Preci _n	Recal _l	F1
SVM	70.04	71.52	70.4	70.14
CNN	87.62	85.5	88.5	88.7
RNN	87.02	88.1	88.3	87
LSTM	70.66	70.34	69.92	68.45
RoBERTa	80.61	80.54	80.01	80.92
DistilBERT	79.34	79.93	79.14	79.63
CNN-LSTMHybrid	92.14	89.64	87.53	88.32
IDCNN-ORM	94.85	91.54	90.07	90.78

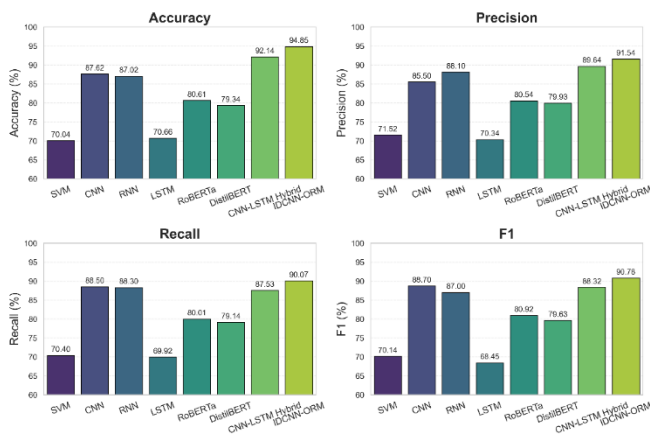


Fig. 8. Comparative results of the IDCNN-ORM method with various metrics.

Fig. 9, LIME-based local explanation of the proposed model's predictions for negative, neutral, and positive reputation classes. This figure indicates the contribution of individual features, where positive LIME weights support the predicted class and negative weights oppose it. Higher absolute weights indicate greater influence on the model's decision.

The proposed IDCNN-ORM architecture exhibits robust solutions in sentiment analysis for the ORM study. The improved mRMR feature selection successfully minimizes

repetition and enhances the quality of the input features. The deep convolutional framework extracts rich hierarchical and contextual data from social media content. The AdaBelief optimizer accelerates convergence and improves training stability, enhancing learning efficacy. The experimental outcome demonstrates that the method attains promising outcomes over conventional systems, validating its efficiency.

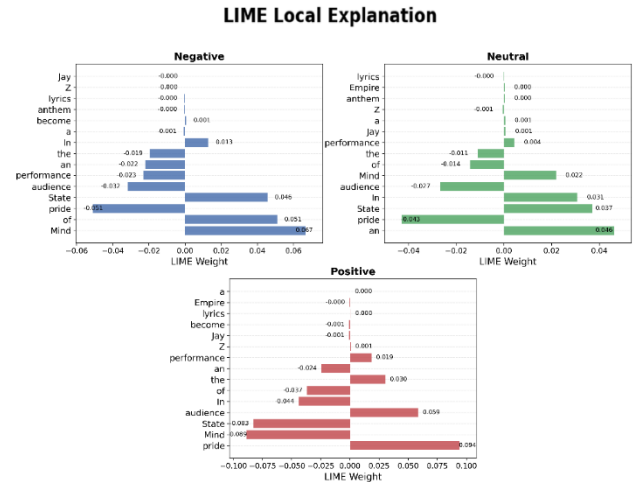


Fig. 9. LIME local explanation of feature contributions for negative, neutral, and positive reputation classifications.

V. CONCLUSION AND FUTURE WORK

This research introduced an IDCNN-ORM approach for enhancing reputation analysis using social media data. The proposed model incorporates an improved mRMR approach to select highly informative and non-redundant features. Moreover, the DCNN is employed to learn hierarchical feature representations from input data. The Adabelief optimizer is applied to enhance training stability by adaptively refining the learning process. Additionally, the LIME-based XAI technique is utilized to enhance transparency in the model's decision-making process. The experimental result analysis of the proposed IDCNN-ORM model is performed using the Social Media Sentiments Analysis Dataset. The comparative analysis demonstrates promising performance with an accuracy of 94.85% compared with existing methodologies. Despite the strong performance of the IDCNN-ORM approach, it still has challenges in handling extremely ambiguous and contextual sentiments in social media content. In the future, it can concentrate on combining transformer-based frameworks to extract long-range contextual dependencies in text information. An innovative feature selection algorithm can be investigated to minimize repetition and enhance adaptability. This technique can be applied to sentiment analysis for wider applicability across global social media platforms. Moreover, integrating progressive explainable AI methodology can improve transparency and customer trust in real-time ORM organizations.

Data Availability Statement: The data that support the findings of this study are openly available in the Kaggle repository at <https://www.kaggle.com/datasets/kashishparmar02/social->

media-sentiments-analysis-dataset?select=sentimentdataset.csv, reference number [23].

ACKNOWLEDGMENT

The authors extend their appreciation to the Arab Open University for funding this work through research fund No. (AOUKSA-524008).

REFERENCES

- [1] Akpuokwe, C.U., Chikwe, C.F. and Eneh, N.E., 2024. Innovating business practices: The impact of social media on fostering gender equality and empowering women entrepreneurs. *Magna Scientia Advanced Research and Reviews*, 10(2), pp.032-043.
- [2] Vale, L. and Fernandes, T., 2025. Social media and sports: driving fan engagement with football clubs on Facebook. In *Contemporary Issues in Sports Marketing* (pp. 37-55). Routledge.
- [3] da Silva Pinho, A., Céspedes Izquierdo, V., Lindström, B. and van den Bos, W., 2024. Youths' sensitivity to social media feedback: A computational account. *Science advances*, 10(43), p. eadp8775.
- [4] Banerji, R. and Singh, A., 2024. Do social media marketing activities promote customer loyalty? A study on the e-commerce industry. *LBS Journal of Management & Research*, 22(1), pp.93-109.
- [5] Muftah, M., 2024. Impact of social media on learning the English language during the COVID-19 pandemic. *PSU Research Review*, 8(1), pp.211-226.
- [6] Shafiezzad, O. and Mostofi, H., 2024. Sentiment analysis of Berlin tourists' food quality perception through artificial intelligence. *Tourism and Hospitality*, 5(4), pp.1396-1417.
- [7] Drossos, D., Coursaris, C. and Kagiouli, E., 2024. Social media marketing content strategy: A comprehensive framework and empirically supported guidelines for brand posts on Facebook pages. *Journal of Consumer Behaviour*, 23(3), pp.1175-1192.
- [8] Okeke, N.I., Alabi, O.A., Igwe, A.N., Ofodile, O.C. and Ewim, C.P.M., 2024. Customer journey mapping framework for SMEs: Enhancing customer satisfaction and business growth. *World Journal of Advanced Research and Reviews*, 24(1), pp.2004-2018.
- [9] Inavolu, S.M., 2024. Exploring AI-driven customer service: Evolution, architectures, opportunities, challenges and future directions. *International Journal of Engineering and Advanced Technology*, 13(3), pp.156-163.
- [10] Rawat, A., Kumar, S. and Samant, S.S., 2024. Hate speech detection in social media: Techniques, recent trends, and future challenges. *Wiley Interdisciplinary Reviews: Computational Statistics*, 16(2), p.e1648.
- [11] Arslan, Y., 2025. Online Reputation Management and User-Generated Content. *Online Retail Marketing Practice: Technology, Operations, and Globalization*, p.115.
- [12] Semenda, O., Sokolova, Y., Korovina, O., Bratko, O. and Polishchuk, I., 2024. Using social media analysis to improve E-commerce marketing strategies. *International Review of Management and Marketing*, 14(4), p.61.
- [13] Istanbuluoglu, D. and Sakman, E., 2024. Successful complaint handling on social media predicts increased repurchase intention: The roles of trust in company and propensity to trust. *European Management Journal*, 42(1), pp.11-22.
- [14] Saheb, T., Sidaoui, M. and Schmarzo, B., 2024. Convergence of artificial intelligence with social media: A bibliometric & qualitative analysis. *Telematics and Informatics Reports*, 14, p.100146.
- [15] Alsemaree, O., Alam, A.S., Gill, S.S. and Uhlig, S., 2024. Sentiment analysis of Arabic social media texts: A machine learning approach to deciphering customer perceptions. *Heliyon*, 10(9).
- [16] Antczak, B., 2024. The influence of digital marketing and social media marketing on consumer buying behavior. *Journal of Modern Science*, 56(2), pp.310-335.
- [17] Putri, L.W.B. and Setiawan, B.L.T., 2025. Analyzing the strategic contribution of social media influencers to e-commerce marketing effectiveness. *JUMDER: Jurnal Bisnis Digital dan Ekonomi Kreatif*, 1(2), pp.12-21.
- [18] Odionu, C.S., Bristol-Alagbariya, B. and Okon, R., 2024. Big data analytics for customer relationship management: Enhancing engagement and retention strategies. *International Journal of Scholarly Research in Science and Technology*, 5(2), pp.050-067.
- [19] Pinheiro, J.M.H., de Oliveira, S.V.B., Silva, T.H.S., Saraiva, P.A.R., de Souza, E.F., Godoy, R.V., Ambrosio, L.A. and Becker, M., 2025. The impact of feature scaling in machine learning: Effects on regression and classification tasks. *IEEE Access*, 13, pp.199903-199931.
- [20] Xie, W., Wang, L., Yu, K., Shi, T. and Li, W., 2023. Improved multi-layer binary firefly algorithm for optimizing feature selection and classification of microarray data. *Biomedical Signal Processing and Control*, 79, p.104080.
- [21] Wang, S., Ren, P., Takyi-Aninakwa, P., Jin, S. and Fernandez, C., 2022. A critical review of improved deep convolutional neural networks for multi-timescale state prediction of lithium-ion batteries. *Energies*, 15(14), p.5053.
- [22] Altinel, D., 2025. Development of Deep Learning Optimizers: Approaches, Concepts, and Update Rules. *arXiv preprint arXiv:2509.18396*.
- [23] Givisis, I., Kalatzis, D., Christakis, C. and Kiouvrekis, Y., 2025. Comparing explainable AI models: SHAP, LIME, and their role in electric field strength prediction over urban areas. *Electronics*, 14(23), p.4766.
- [24] <https://www.kaggle.com/datasets/kashishparmar02/social-media-sentiments-analysis-dataset?select=sentimentdataset.csv>
- [25] Gothane, S., Reddy, G.V., Kumar, K.P., Baswaraj, D., Devi, G.P., Thanugundala, S. and Changala, R., 2024. Sentiment analysis in social media using deep learning techniques. *Int J Intell Syst Appl Eng*, 12(3), pp.1588-1597.
- [26] Nurlanuly, A.T. (2025) 'Sentiment analysis of texts from social networks based on machine learning methods for monitoring public sentiment,' *ArXiv.org* [Preprint].
- [27] Rahamana, S.M.U. and Sudhe, S. (2025) 'Sentiment analysis in social media platforms using deep learning and data mining for business intelligence,' *International Journal of Research Publication and Reviews*, 6(9), pp. 4534-4541.