International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Qubits and Sentiments: Unveiling New Perspectives in Hindi Textual Data

Vaibhav Prakash Vasani[1]* Asha Ambhaikar[2]

Submitted:10/03/2024 **Revised**: 25/04/2024 **Accepted**: 02/05/2024

Abstract. Sentiment analysis (SA) is a critical component of Natural Language Processing (NLP), particularly for automatic text classification. However, previous approaches have shortcomings in capturing nuances such as negation, word pairings, and contextual understanding, particularly in languages like Hindi. To solve these issues, this research offers a novel Quantum Neural Network (QNN) technique designed specifically for sentiment analysis in Hindi text data, which employs quantum computing concepts to capture linguistic nuances and context better. This study uses a Quantum Variational Auto Encoder to encode classical data into a quantum form, capturing diverse sentiments like sarcasm and colloquial expressions. A Tuned Quantum Convolutional Neural Network architecture is introduced to capture complex linguistic syntax. A novel Sequential-based hyperband optimization technique is used to enhance model performance. The hybrid approach significantly improves accuracy and efficiency in handling quantum data, contributing to the advancement of SA, particularly in Hindi Movie Reviews. The findings demonstrate that the proposed strategy performs best with accuracy 97.64 %, precision 85.93%, recall 99.17%, F-1 score 92.20%, than other accepted strategies.

Keywords. *Quantum Neural Network (QNN), Quantum Variational Autoencoder (Q-VAE), Tuned Quantum CNN, Sequential-based hyperband optimization technique, Hindi movie reviews.*

1. Introduction:

SA, a prominent branch of NLP, involves identifying and extracting sentiments or opinions expressed in textual data. SA has become essential for understanding public perception, customer feedback, and social trends as digital content on social media, product reviews, and online forums has exploded [1]. This process involves determining whether a text conveys a positive, negative, or neutral sentiment.

In recent years, [2] machine learning (ML) and deep learning (DL) techniques have significantly advanced sentiment analysis methodologies. ML approaches typically employ algorithms that learn patterns and relationships in labelled training data to predict unseen text. DL, particularly through neural networks, has gained prominence for its ability to capture intricate features and hierarchical representations in text data, resulting in more nuanced sentiment predictions [3], [4]. Extending sentiment analysis to languages such as Hindi reflects the global need to comprehend various linguistic contexts [5]. Due to the rich morphology, varying sentence structures, and the presence of code-switching, analyzing sentiment in Hindi text presents unique challenges [6], [7]. Creating sentiment analysis models for Hindi entails adapting and fine-tuning existing approaches to account for the complexities of the language.

ML-based sentiment analysis frequently employs techniques such as Support Vector Machines (SVM), Naive Bayes, and Random Forests [8]-[11]. These models predict sentiment classes by extracting features from text data such as word frequencies or n-grams. While these approaches are effective, they may struggle to capture semantic nuances and context-dependent sentiments [12]. DL models, particularly recurrent neural networks (RNNs) and transformer architectures such as BERT (Bidirectional Encoder Representations from Transformers), have demonstrated outstanding performance in SA tasks [13], [14]. These models capture dependent context and long term dependencies, improving their ability to detect subtle emotions. DL methods, on the other hand, may necessitate large amounts of labelled data and computational resources.

Researchers have investigated techniques such as machine translation for preprocessing, domain-specific lexicons, and sentiment lexicons tailored for the language in the context of sentiment analysis in Hindi. Despite these advances, challenges remain, such as the scarcity of labelled datasets for Hindi sentiment analysis and the need for additional research to address linguistic variations [15].

The limitations of different approaches in predicting sentiment classes include potential bias in training data, difficulties in dealing with sarcasm or irony, and the model's sensitivity to context changes [16]. Furthermore, when

[1]**Department of Computer Science and Engineering*
Kalinga University, Kotni, Near Mantralaya, Naya Raipur Chhattisgarh – 492101
https://orcid.org/0000-0001-6498-553X
**Corresponding Author Email: vaibhav.prakash@kalingauniversity.ac.in*
[2]*Department of Computer Science and Engineering*
Kalinga University, Kotni, Near Mantralaya, Naya Raipur Chhattisgarh – 492101
https://orcid.org/0000-0002-6814-5949
Co author Email: asha.ambhaikar@kalingauniversity.ac.in

confronted with colloquial expressions, domain-specific jargon, or changing language trends, sentiment analysis models may struggle. As the field evolves, researchers and practitioners are actively working to overcome these challenges, aiming to enhance efficiency and robustness of SA models across languages and domains. Thus, the study requires a new framework for sentiment analysis in Hindi text. The most significant aspect of the proposed work is as follows:

- Introducing an innovative method for sentiment analysis in Hindi text data that makes use of quantum neural networks.
- The proposed Quantum Variational Auto Encoder encodes classical data into a quantum form, allowing for the representation of various feelings such as sarcasm and colloquial expressions.
- Additionally, the introduction of a Tuned Quantum Convolutional Neural Network (Q-CNN) architecture enables the accurate capture of complex linguistic syntax in Hindi text, including sarcasm and colloquial expressions.
- Furthermore, the utilization of a Sequential-based hyperband optimization technique enhances model performance while considering computational complexity.
- The Hindi movie review dataset outperforms previous methodologies in terms of performance metrics.

The study proposes a unique and successful method that significantly improves accuracy and efficiency in handling quantum data, making notable strides in advancing sentiment analysis, particularly in the domain of Hindi Movie Reviews.

The next section describes the related works, Section 3 describes the approach for the suggested model, and following 4th part reveals the system evaluation, and discusses the result. Finally, section 5 discusses the conclusion and future work of predictive work in Hindi Sentiment Analysis.

2. Literature Survey

The Sentiment Analysis using Hindi Text in the literature is examined in-depth in this section.

To analyse the opinion of reviews in Hindi Textual, Dupakuntla et al. [17] used the NB classifier to categorize Hindi user analyses into positive or negative sentiments. The performance of this model using movie reviews and achieved better accuracy in SA, highlighting the need for more research in this area. However, there may be challenges in handling complicated sentiments, context-specific language, and evolving linguistic trends in Hindi reviews.

In India, there was a lack of Hindi language interfaces, necessitating the development of a system using sentiment analysis for Hindi news articles. A study conducted by Yadav et al. [18] used applications of natural language processing in news articles and computational linguistics to standardize the information was positive, negative, or neutral. The study aimed to help identify an individual's attitude and emotional state, and it also contained a dictionary of negative and positive words for precise analysis. However, challenges in accurately capturing the diverse and contextual nature of sentiments in Hindi, including sarcasm and colloquial expressions, remained a concern.

Influenced by the increasing acceptance of DL models, Rani et al. [19] conducted experiments using convolutional neural networks (CNN) to perform SA on Hindi movie reviews from online newspapers and websites. The dataset was manually annotated by native Hindi speakers. The CNN model outperformed both traditional ML algorithms and cutting-edge methods, achieving 95% accuracy demonstrating DL models have the potential to improve human language understanding. When compared to traditional ML algorithms and cutting-edge methods, DL models demonstrate their effectiveness in capturing intricate features and nuances in the sentiment of Hindi textual data.

The research by Shrivastava et al. [20] utilized a DL-based approach for SA of Hindi movie reviews, combining a GRU network with a Hindi word embedding model. This model achieved superior performance compared to traditional methods. However, it is essential to consider that its results in longer training times and increased resource requirements due to the utilization of a GA to mechanically build the GRU network architecture might introduce additional computational complexity, especially during the optimization process.

Focuses on bipolar sentiment classification of Hindi movie reviews using two publicly available IIT-P datasets. Sharma et al. [21] developed a robust stacked ensemble-based architecture to effectively classify Hindi reviews, combining the strengths of both techniques. The results suggested that SEBA outperforms individual baselines and displays superior performance with unigrams and TF-ISF characteristics, making it suitable for online deployment in binary review classification tasks. However, despite SEBA's success, the stacked ensemble approach may pose computational complexity and resource requirements challenges, potentially limiting its applicability in resource-constrained environments. Furthermore, the quality and representativeness of training datasets may have an impact on the model's performance, necessitating careful consideration in a variety of real-world scenarios.

Jain et al. [22] used Hindi tweets on COVID-19 as input data for SA. NLP is used for feature extraction, and the optimal characteristics are selected using GWO. A hybrid of CNN and a LSTM model is used to categorize sentiments as positive, negative, or neutral. The model outperforms other ML techniques, achieving the highest accuracy. However, it is crucial to highlight that the performance of the hybrid model may be affected by the quality and representativeness of the input data, and biases in the training data may harm the model's generalisation to new settings. Punetha et al. [23] developed a new MCDM (Multi-Criteria Decision Making) to provide systematic approaches for evaluating and ranking alternatives and a game theoretic mathematical framework for analyzing Hindi review sentiment. They collected ratings and feedback from Hindi speaking reviewers using a game-theoretic approach and a Nash equilibrium. The model classified reviews as good, negative, or neutral, assigned star ratings and polarity scores and classified unstructured sentiment. However, that the complexity of the game-theoretic approach and its reliance on Nash equilibrium may pose challenges in real-world implementation, potentially necessitating careful consideration of computational resources and adaptability to varying linguistic expressions in Hindi reviews.

Based on a comprehensive literature review, the purpose of this study is to investigate and address existing constraints in sentiment analysis of Hindi textual data. It specifically aims to address issues such as dealing with complex sentiments, context-specific language, emerging linguistic trends, and capturing the various and contextual characters of sentiments, including sarcasm and colloquial expressions.

3. Proposed Methodology

SA is essential for organizations to comprehend customers' opinions and input, enabling informed decision-making and personalized interactions. ML and DL algorithms have been used for Hindi sentiment analysis, but challenges exist in the field, especially in languages like Hindi. Limited large and diverse labelled datasets and models may struggle to generalize well to underrepresented classes. Additionally, fine-tuning hyperparameters can lead to overfitting and underfitting, reducing performance and generalization of new data. To address these issues, improved algorithms, specialized datasets, and novel neural frameworks for sentiment analysis using the Hindi textual language are needed.

Building upon the insights from prior studies, this research proposes the development of QNN model tailored for SA of Hindi textual information. The goal of incorporating quantum computing principles in this process is to increase the efficiency of sentiment classification, notably in the field of Hindi movie reviews. This methodology includes several critical processes, such as text preparation, quantum data encoding, architectural design, optimisation, and evaluation. Each stage is intended to solve unique sentiment analysis difficulties while also leveraging quantum computing capabilities to improve accuracy and efficiency. The overview architecture of the innovative study is shown in Figure 1.

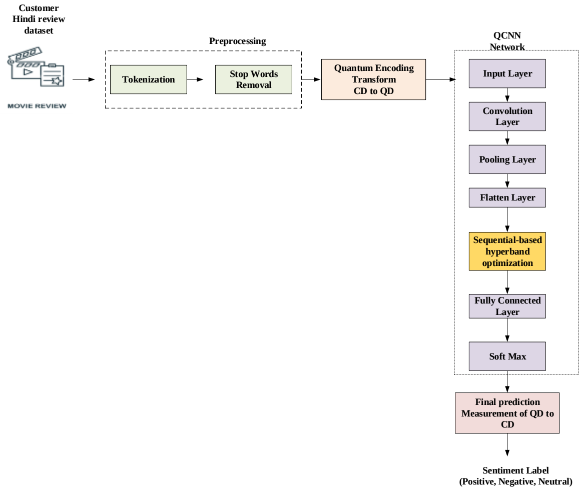

Fig. 1. Overflow of Proposed Work

The study preprocesses input Hindi textual data using tokenization and word removal. It then uses a Q-VAE to encode classical data into a quantum form, integrating quantum computing principles into sentiment analysis. This process captures diverse sentiments, including nuanced expressions like sarcasm and colloquialisms. The architecture design includes a Tuned Q-CNN to capture complex linguistic syntax in Hindi text. The model's hyperparameters are optimized using a Sequential-based Hyperband Optimization (SBHO) technique, balancing exploration and exploitation to enhance performance while considering computational complexity. Each step is intended to solve unique sentiment analysis difficulties while also leveraging quantum computing capabilities to improve accuracy and efficiency.

3.1 Text Preprocessing

Previous methods may have had difficulties in successfully segmenting Hindi text into relevant units for analysis. In some circumstances, simple tokenization algorithms may fail to capture the intricacies found in the Hindi language, resulting in inaccurate sentiment analysis. In this study, text preprocessing is the initial step in preparing raw Hindi textual data for sentiment analysis. To begin, tokenization breaks down the input text into separate tokens, which can be words or subwords. This segmentation enables the model to do granular text analysis, making it easier to extract significant features for sentiment classification. Furthermore, stop word removal is used to filter out common words that do not communicate substantial sentiment information, such as articles, prepositions, and other frequently used phrases. By removing these unnecessary words, the preprocessing strategy tries to reduce noise and streamline the sentiment analysis process permitting the algorithm to concentrate to focus on meaningful content that contributes substantially to sentiment classification. Overall, the combination of tokenization and stop word removal improves input data quality. This improves the precision and efficacy of SA in Hindi written content.

3.2 Quantum Data Encoding

Q-VAE integrates quantum computing principles into sentiment analysis, making it feasible to capture an extensive range of sentiments, including nuanced expressions such as sarcasm and colloquialisms. By exploiting quantum principles, Q-VAE provides a novel approach to encoding textual data, allowing the model to capture complex patterns and correlations that would be difficult to extract using conventional approaches alone. This capability is well-suited to addressing the difficulties faced in prior sentiment analysis efforts, particularly when dealing with the diverse and contextual nature of sentiments in Hindi textual data.

Previous methods to encode textual data into quantum form may have been limited by a lack of algorithms designed expressly for sentiment analysis tasks. Traditional encoding approaches may not adequately capture the complexities of sentiment, resulting in poor performance in recognizing and categorizing subtle sentiments. However, by including Q-VAE in proposed methods, which is to address these constraints by providing a framework for properly encoding textual data in a quantum space. Q-VAE can catch subtle differences in sentiment, such as sarcasm and colloquial idioms, which improves the model's capacity to reliably classify sentiments in Hindi textual data.

QVAE consists of two basic components such as an encoder and a decoder. The encoder converts the input data to a lower-dimensional latent space, while the decoder reconstructs the original data from the underlying representation. Both the encoder and decoder are built utilizing quantum circuits, allowing the manipulation of quantum states. Figure 2 describes the quantum data encoding process.

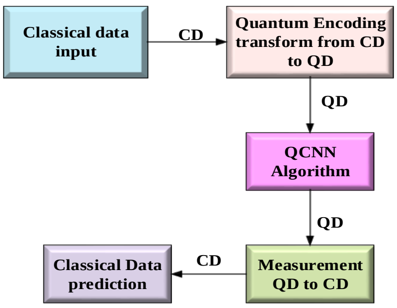

Fig. 2. Processing technique of Quantum DL

3.2.1 Quantum State Space (Qss)

The study explores using quantum embeddings to represent linguistic features in Hindi text, potentially capturing intricate word relationships in a Qss. In classical computing, a bit represents two distinct states, such as 0 and 1. In quantum computing, a qubit is a fundamental unit, and the Dirac symbol is commonly used to describe one, such as $|0\rangle$ and $|1\rangle$. Furthermore, a qubit can exist in a superposition of $|0\rangle$ and $|1\rangle$. A total of 2^n information can be stored in n qubits, which has the advantage of parallel computing and potentially results in exponential improvements over traditional approaches. A qubit may be a continuous representation of ground states $|0\rangle$ and $|1\rangle$, such as

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \quad (1)$$

Where $|\psi\rangle$ is called superposition state, and α and β are complex-valued numbers. Furthermore, it is incredible to obtain an unknown qubit. Measuring can provide $|0\rangle$ and $|1\rangle$, where $|\alpha|^2$ is the probability of obtaining $|0\rangle$ and $|\beta|^2$ is the probability of obtaining $|1\rangle$, which gets $|\alpha|^2 + |\beta|^2 = 1$. For example, $\frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle$, has a 50% chance of yielding $|0\rangle$ and 50% of $|1\rangle$. So, the equation can also be represented as

$$|\psi\rangle = e^{i\gamma}\left(\cos\frac{\theta}{2}|0\rangle + e^{i\varphi}\sin\frac{\theta}{2}|1\rangle\right) \quad (2)$$

Where θ, γ, φ are all real numbers, additionally $e^{i\gamma}$ can be eliminated as it has no noticeable effect.

3.2.2 Quantum System

Two qubits are in a superposition of their four ground states, as shown below:

$$|\psi\rangle = \alpha_{00}|00\rangle + \alpha_{01}|01\rangle + \alpha_{10}|10\rangle + \alpha_{11}|11\rangle \quad (3)$$

Where $\alpha_{ij}(i, j\epsilon\{0,1\})$ is known as the amplitude. In a two-qubit system, only one of the qubits, say the second bit, can be measured. Assuming a measurement result of 1, the measured state $|\psi\rangle$ collapses to

$$|\psi'\rangle = \frac{\alpha_{01}|01\rangle + \alpha_{11}|11\rangle}{\sqrt{|\alpha_{01}|^2 + |\alpha_{11}|^2}} \quad (4)$$

Normalisation is achieved through the application of factor $\sqrt{|\alpha_{01}|^2 + |\alpha_{11}|^2}$. The Bell state $\frac{|00\rangle + |11\rangle}{\sqrt{2}}$ is a crucial two-quantum state as it ensures that measurements of two qubits always yield the same result. It is a critical component of quantum teleportation and ultra-dense coding. Consider an n-qubit system with the ground state $|x_1\, x_2 \ldots . x_n\rangle$ and amplitudes of 2^n. Compared to traditional systems, n qubit systems feature an exponential improvement in storage and computing.

3.2.3 Evaluation of Qs

A quantum computer is composed of quantum circuits gates that are utilised to process quantum data. A feature map is a function that converts data into quantum states, transforming it into a linear space called the Hilbert Space in linear algebra [24]. It is described as a function that maps input data to the space of features which is shown in Figure 3. A quantum circuit is being designed to analyze sentiment in Hindi text, using quantum-encoded input and gates to process quantum states in a manner that reflects sentiment analysis computations.

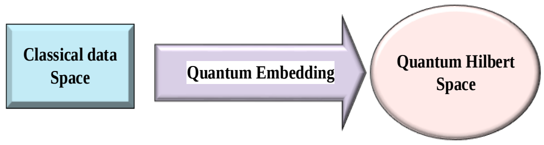

Fig. 3. QNN shows data embedding in Hilbert Feature Space

The input text is encoded into quantum states (qubits), representing linguistic features or words in Hindi text. The Hadamard gate (H) creates superposition, allowing multiple states to exist simultaneously. It is a fundamental quantum gate in QC that transforms basis states $|0\rangle$ and $|1\rangle$ into superposition states, enabling quantum bits to exist in multiple states simultaneously. It is represented by a matrix.

$$H = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{5}$$

When applied to a qubit, the Hadamard gate uses matrix multiplication to modify its state. The Hadamard gate's impact on a single qubit in the $|0\rangle$ the state is examined.

$$H\left|0\right\rangle = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{|0\rangle+|1\rangle}{\sqrt{2}} \tag{6}$$

Similarly, when applied to the state $|1\rangle$

$$H\left|1\right\rangle = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \frac{|0\rangle-|1\rangle}{\sqrt{2}} \tag{7}$$

the Hadamard Gate superimposes the $|0\rangle$ and $|1\rangle$ states, creating a balanced qubit state. This superposition is a fundamental aspect of quantum mechanics that is used in quantum algorithms to conduct computations in parallel across different potential states.

3.2.4 Quantum Operation for Sentiment Analysis

Quantum operations ($U(\theta)$) are used in sentiment analysis computations, focusing on capturing sentiment-related information from quantum-encoded text, with the specific operations and parameters (θ) varying based on the sentiment analysis algorithm and linguistic features being analyzed. Quantum states are used to measure sentiment outputs, with different outcomes indicating positive or negative sentiment categories, and the probabilities of these outcomes provide insight into sentiment strength.

The selection of Q-VAE as the quantum data encoding technique in proposed methodology is motivated by its capacity to incorporate quantum computing concepts into sentiment analysis, as well as its potential to overcome restrictions discovered in earlier research. By utilising Q-VAE, this proposed work improves the precision and efficacy of SA in Hindi textual data by capturing a wider range of sentiments and contextual nuances.

3.3 Proposed Q-Cnn Network

After encoding the classical input into quantum form with the Q-VAE, the output is fed into the proposed Q-CNN network. This Q-CNN stems from its unique capabilities in capturing complex linguistic syntax in Hindi text while leveraging principles of quantum computing. Previous sentiment analysis algorithms may have had limitations in capturing the numerous features and subtleties prevalent in Hindi textual data, particularly in terms of language grammar. Traditional neural network topologies may struggle to understand the contextual and subtle nature of sentiments, resulting in inferior performance in sentiment classification tasks. However, by incorporating QCNN into proposed methods, this work aims to overcome these constraints by leveraging the power of quantum computing to improve the model's capacity to accurately capture language syntax. This QCNN model preserves fundamental aspects of standard CNNs, such as nonlinearity, locality, and extensibility to deep structures, making it ideal for sentiment analysis tasks. It also introduces a novel aspect by utilising a parametric quantum circuit, which allows for the exploration of data point correlations across enormous linear regions. The architecture of the QCNN network is shown in Figure 4. This novel approach enables the model to detect subtle fluctuations and correlations in textual input, increasing its performance in sentiment analysis tasks.

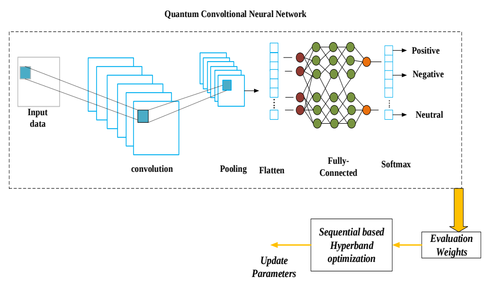

Fig. 4. Architecture of Proposed Network

The system combines a QCNN and a sequential hyperband optimisation model, with hyper-tuning for optimal parameters to improve classification accuracy. The major purpose of these hyperparameter optimisations in DL is to investigate and discover the parameters that provide the maximum performance, as measured on a validation set. The SBHO optimisation and utility function are useful in determining the prediction method's accuracy.

Q conv is considered the first layer on QCNN. This layer extracts features from the input data via convolution processes. In the context of sentiment analysis, this layer filters the input quantum data to capture patterns and relationships between words or subwords. Each filter recognises specific features, such as word pairings or linguistic structures, which aids the model in determining significant information for sentiment classification. The convolutional layer can operate on the quantum-encoded text. Traditional convolutional layers in classical CNNs use filters to capture local patterns in data. Hence Q-convolutional layer would need to be adapted to operate on quantum states. The convolutional layer applies filters (kernels) to the input quantum data X via convolution operations.

$$Z_i = f(\sum_j W_j * X_{i+j} + b \quad (8)$$

Where Z_i is the feature maps output, W_j is denoted as filter weights, X_{i+j} represent the input data within the accessible field, b is the bias term and f is the activation function such as ReLU.

Following the conv layer, the pooling layer down samples the feature maps generated by the convolutional processes. This sampling strategy reduces the level of detail of the feature maps while maintaining the most useful data. The two most prevalent pooling processes are max pooling, which selects the largest value from each zone, and average pooling, which calculates the average value. It saves computation costs and eliminates overfitting by emphasising the most significant elements of the dataset.

Quantum fully connected layers capture global relationships in quantum-encoded data, performing computations similar to classical fully connected layers but adapted for quantum states. This layer combines high-level features from previous layers, consisting of universal single-qubit quantum gates and CNOT gates. This FC layer, composed of densely connected neurons, performs high-level feature extraction and classification based on information from previous layers. In sentiment analysis, it aggregates extracted features and learns complex patterns to predict input text sentiment by predicting complex patterns. It computes the output by multiplying the flattened input by the weights, adding biases, and then applying an activation function.

$$\hat{Y} = g(W_{FC}.Y_{flatten} + b_{FC}) \quad (9)$$

Where, $\hat{Y}$ is the output prediction, W_{FC} represents the weights, b_{FC} is the bias term, and g is activation function.

The softmax is the final layer in the Q-CNN design which is commonly employed for multi-class classification applications. It turns the previous layer's raw output scores into probability distributions for multiple sentiment classes, including positive, neutral, and negative. The softmax function ensures that the total of the probabilities for all classes equals one, allowing the model to generate a probability distribution reflecting the likelihood of the input falling into each sentiment class.

3.3.1 Model Creation And Training

The proposed work addresses previous study challenges in model creation and training, including longer training times, increased resource requirements, and computational complexity, particularly in dealing with quantum computing constraints, in a comprehensive approach. To mitigate these challenges, this study proposed an SBHO technique. This method combines the strengths of Hyperband optimisation (HBO) and Bayesian optimisation (BO) to establish a balance between exploration and exploitation, accelerating the hyperparameter search process and identifying high-performing configurations more effectively. The proposed approach aims to improve the model training process by employing SBHO within Q-CNN architectures. Furthermore, this technique is designed to adapt to quantum limitations, ensuring compliance with the unique properties of quantum computing.

The model was created, the forecast model was used, and outcome was calculated in attribute selected in the dataset. The data was divided into two categories: training and testing, with training data accounting for 70% and testing data accounting for 30% of the pre-processed data. This study proposed an optimization technique SBHO which combines HBO and BO are achieved by using a model with uncertainty that measures the density of excellent topologies in the input space and sampling from it rather than sampling uniformly at random. This is because, while HB has high efficacy for discovering combinations that yield satisfactory outcomes, it uses arbitrary search and so does not quickly locate the optimal configurations.

3.3.1.1 hyperband optimization

It is a system that directs energy towards more viable hyperparameter combinations. It creates a series of n trial points, each representing one hyperparameter setting. Each trial point is allocated a budget and evaluated for performance. Challenge points with inadequate performance are removed, and the process repeats until only one trial point remains in the set. This method ensures that the most promising configurations receive the most resources. However, the more resources each trial point receives, the more trustworthy this indicator becomes. For a limited budget, each trial point is allocated $B/_n$ resources. A large n may result in fewer resources for each trial point, affecting its performance. Conversely, a small n may result in a pool of fewer candidates. Hyperband addresses this dilemma by experimenting with different n values and returning the best hyperparameter configuration.

The HBO is presented in Algorithm 1, which uses the function $get\_hyperparameter\_configuration(n)$ to return $n\ i.i.d.$ trial points from a specified space for setting up. The function $run\_then\_return\_obj\_val(x, r)$ first trains the trial point x with r resources and returns the evaluated objective function value f(x). The function $top\_k(trials, obj_{vals}, k)$ ranks all trials based on their objective function obj_vals value.

Algorithm 1 HBO algorithm
Input: Maximum resource allocation for a single hyperparameter configuration (R) and percentage controller (η). Output: A single hyperparameter configuration 1. initialization: $S_{max} = \lfloor log_{\eta}(R) \rfloor, B = (S_{max}+1)R$ 2. for $S \in \{S_{max}, S_{max} - 1, \ldots ,0\}$ do 3. $\eta = \left\lceil \frac{B}{R} \frac{\eta^2}{(S+1)} \right\rceil, r = R\eta^{-S}$ 4. $X = get_{hyperparameter_{configuration(n)}}$ 5. for $i \in 0, \ldots S$ do 6. $n_i = \lfloor n\eta^{-i} \rfloor$ 7. $r_i = r\eta^i$ 8. $F = \{run\_then\_return\_obj\_val(x, r_i): x \in X$ 9. $X = top\_K(X, F, \lfloor ^{n_i}/_{\eta} \rfloor)$ return setup with the optimal objective function value

3.3.1.2 Bayesian Optimization

The primary notion behind BO is as follows. Because examining the unbiased function f for a experimental point x is exceedingly affluent, it approximates f using a probabilistic substitute model, which is significantly cheaper to evaluate, and is an iterative process. The algorithm uses BO to sample trial points sequentially, utilising all the details in historical context reflected by the constructed alternative model. It evaluates $f(x_{t+1})$, updates the surrogate model based on the new data point, and goes back to step eq (1), ensuring that the next trial point is determined by all previously sampled points.

Algorithm 2 shows BO, where the subsequent sample point is collected at the point of acquisition phase$\mu(x|\mathrm{Dt})$. The anticipated improvement in Equation 10 is used as the acquisition function, where x^+ is the trial point with the best objective function value in the first t steps. This function balances the exploration of underexplored regions with the exploitation of visited places in the arrangement of space.

$$\mu(x|Dt) = \mathbb{Z}(max\{0, f_{t+1}(x) - f(x^+)\}|Dt) \quad (10)$$

Two probabilistic models, discriminative Gaussian Process (GP) and generative Tree-structured Parzen Estimator (TPE), have been proposed to compute the expectation in Equation 10, with TPE achieving better results than GP in terms of $p(x|f(x))$ and $p(f(x))$.

Algorithm 2 Bayesian Optimization
1. initialization: $D_0 = \emptyset$ 2. for $t \in \{1,2, \ldots\}$ do 3. $x_{t+1} = argmax_x \mu(x|D_t)$ 4. Evaluate $f(x_{t+1})$ 5. $D_{t+1} = D_t \cup \{(x_{t+1}, f(x_{t+1}))\}$ 6. Upgrade the stochastic proxy model with D_{t+1}

3.3.1.3 Sbh Optimization

SBH is the combination of both HBO and BO are powerful approaches with their strengths and weaknesses. They complement each other, with the strength of one being the strength of the other. Combining them can lead to an enhanced hyperparameter optimization algorithm called sequential-based hyperband optimization.

Algorithm 3 SBH Optimization
Input: Maximum resource allocation for an individual extreme parameter setup, R, and percentage controller η. Output: A single hyperparameter configuration 1. initialization: $S_{max} = \lfloor log_n(R) \rfloor$, $B = (S_{max}+1)R$ 2. for $S \in \{S_{max}, S_{max} - 1, \ldots ,0\}$ do 3. $n = \left\lceil \frac{B}{R} \frac{\eta^2}{(S+1)} \right\rceil, r = R\eta^{-s}$ 4. $for\ i \in 0, \ldots , s$ do 5. $n_i = \lfloor n\eta^{-i} \rfloor$ 6. $n_i = r\eta^i$ 7. if $i == 0$ then 8. $X = \emptyset, D_0 = \emptyset$ 9. for $t \in \{1,2, \ldots , n_i\}$ do 10. $x_{t+1} = argmax_x \mu(x|D_t)$ 11. $f(x_{t+1}) = run\_then\_return\_obj\_val(x, r_i)$ 12. $X = X\{x_{t+1}\}$ 13. $D_{t+1} = D_t \cup \{(x_{t+1}, f(x_{t+1}))\}$ 14. The probabilistic surrogate model can be updated by adding a new term, D_{t+1} 15. else 16. $F = \{run\_then\_return\_obj\_val(x, r_i): x \in X\}$ 17. $X = top\_k(X, F, \lfloor ^{n_i}/_{\eta} \rfloor)$ Return setup with the optimal objective function value

Hyperband samples all trial points independently, but once a trial point is evaluated, it provides insight into the objective function's behaviour in the configuration space. If the objective function is poor, it's advisable to avoid sampling near x, or consider sampling more. However, this is only sensible if the objective function is smooth over the arrangement space. This approach may not utilize lessons learned from previous trials, potentially wasting time on repeating errors.

Bayesian optimization and Hyperband are two approaches to learning ML. Bayesian optimisation allocates resources to each trial point x to obtain the desired function value $f(x)$, while Hyperband samples trials sequentially based on previous experience. BO uses an early stopping mechanism, discarding trials inferior to others throughout the initial stages.

To combine both approaches, the study proposes a combination approach called SBHO. They follow the Hyperband algorithm but use Bayesian optimization to sample trial points one by one. In each first round of Hyperband, the intermediate performance of a trial point x_{t+1} is evaluated, contributing to the updating of the surrogate probabilistic model in Bayesian optimization. This new data point helps to improving the substituting statistical model, which is used to sample the following trial point. This approach allows for more efficient and accurate training of DNN, reducing the time and effort required for each trial point.

As a result, SBHO not only improves the Q-CNN model's performance against computational expenses, but it also improves the sentiment analysis process, resulting in more efficient and effective sentiment classification in Hindi textual data.

4. Evaluation Analysis

This approach presents the findings of proposed methodology for SA of Hindi content data, as well as analyse the implications of them. The tuned quantum CNN model in this study was trained on the dataset presented in Section 3. The IIT Patna dataset was divided into three subsets: testing, validation, and training.

4.1 Dataset Description

In [25], the study established the IITPatna Hindi Reviews dataset will make to easier study in the field of aspect category detection. Hindi sentences are extracted from IIT-Patna's freely available Review Sentiments Dataset. The dataset is in the XML (Extensible Markup Language) format which cover 12 various domains which include laptops, mobiles, tablets, cameras, headphones, home appliances, speakers, televisions, smartwatches, mobile apps, travel, and movies. The dataset contains product reviews classified as good, negative, or neutral. The dataset is divided into 70% for training, 10 % for validation, and 20% for testing parts with each portion kept individually in a CSV file. Each record in the CSV file has the following format: |label, review| and sentiment classification for Indian languages, as there was no existing dataset available. They carefully marked customer reviews from Hindi websites based on predefined criteria. The dataset contains product reviews classified as good, negative, or neutral. Each record in the CSV file has the following format: |label, review|. Table 1 shows some of the IIT-Patna dataset's basic metrics.

Table 1. Basic Statistics of dataset

Metrics	IIT Patna Dataset
Sentences	5417
Aspect terms	4509
Tokens	96140
Average of aspect terms per sentence	0.81
% of Bs (Begin)	4.899
% of Is (Inside)	3.496
% of Os (Outside)	91.603
Number of domains	12
Maximum number of words in any sentence	76

4.2 Experiment Setup

The proposed model was implemented in Python and built with the Keras, Scikit-learn, and TensorFlow libraries. The Google Colab virtual platform served as the development environment. The hardware specification included 12 GB of RAM. The Matplotlib library was used to visualise and interpret the results. An early-stopping method was implemented to optimise training and minimise overfitting. The model might cease learning if the confirmation loss remains unchanged by 1.0×10^{-2} for 15 consecutive epochs. It protected the model from training for a longer period when additional performance increases were unlikely. Table 2 summarises the optimised hyperparameters used to create the model.

Table 2. Hyperparameter details of the proposed model

Parameter	Description
The maximum length of the input	128
batch size	32
Learning rate	1.0×10^{-2}
Training epochs	15
Activation Function	Softmax

For each domain in the dataset, a separate tuned Quantum CNN algorithm is trained. For this experiment, the dataset is split into training, testing and validation.

4.3 Evaluation Metrics

The proposed work evaluated the performances using common metrics such as accuracy, precision, recall, and F1-score. Furthermore, the study conducted a thorough comparison with existing methods for SA of Hindi textual data to assess efficacy of this methodology. The exact calculating procedures are as follows:

Accuracy: Ratio of accurately predicted terms to total terms

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (11)$$

- F1-score: It is the harmonic mean of precision and recall

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (12)$$

- Precision: Ratio between the number of successfully classified aspect terms and the total number of positive terms classified by the classifier.

$$Precision = \frac{TP}{TP+FP} \quad (13)$$

- Recall: ratio of correctly identified aspect terms to the total number of aspect terms present.

$$Recall = \frac{TP}{TP+FN} \quad (14)$$

4.4. Experimental results and evaluation

The experiment with sentiment classification to validate proposed strategy. This article introduces the models used for comparison, analyses the experimental data, and discusses the benefits of Tuned QCNN over classical neural networks. Table 3 results on IIT Patna Hindi movie review sentiment classification dataset and the evaluation graph is shown in Figure 5.

Table 3. Results obtained for the proposed model

Performance metrics	Results Obtained (%)
Accuracy	97.64
Precision	85.93
Recall	99.17
F1-score	92.20

Figure 6 demonstrates that after 15 training epoch repetitions, the proposed model-tuned Q-CNN performed better. This accomplishment highlights the model's effectiveness in sentiment analysis designed especially for evaluations of Hindi films. Using meticulous tuning and enhancement, the Q-CNN model proved to be more capable than traditional techniques in identifying and interpreting the subtle emotions contained in the evaluations. Because the training procedure was iterative, continual optimization was possible, which helped the model gradually improve both its predicted accuracy and robustness. Notably, the framework's adaptability and capacity to extract information from the underlying data patterns is demonstrated by the steady progress seen throughout the training epochs. This finding has important ramifications for sentiment analysis uses in Hindi movie reviews and provides a reliable framework for assessing viewer emotions and feelings. The effectiveness of the modified Q-CNN model also highlights the possibility of applying cutting-edge ML methods to efficiently assess and understand sentiment in languages other than English, accommodating a variety of linguistic settings and cultural nuances.

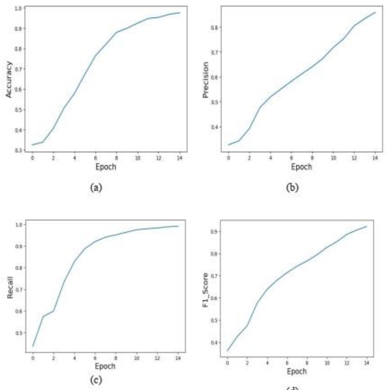

Fig. 5. Performance Evaluation of Proposed Model

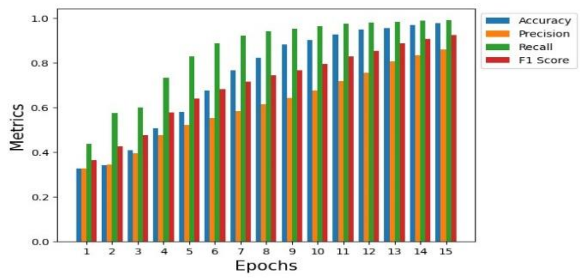

Fig. 6. Results obtained for various epochs

This experiments demonstrated the methodology and achieved significant improvements in sentiment analysis accuracy compared to baseline methods. The Q-CNN architecture, combined with Q-VAE for data encoding and Sequential-based hyperband optimization for model optimization, consistently outperformed traditional ML and DL models in capturing diverse sentiments and contextual nuances in Hindi movie reviews.

This proposed model's performance was closely compared to a number of currently used methods that are frequently used for sentiment analysis of Hindi film reviews, such as state-of-the-art transformer-based models like mBERT NLI-M, mBERT-E-MV, and mBERT-E-AS, as well as NB, DT, and SMO. After a thorough assessment and comparison, this works strategy continuously outperformed these other approaches in terms of effectiveness. Proposed model demonstrated its capacity to accurately identify and comprehend the subtle sentiments seen in movie reviews by outperforming conventional techniques like NB and DT, which are frequently employed but frequently find it difficult to capture the nuances of sentiment in languages like Hindi. Moreover, the proposed model outperformed sophisticated transformer-based models that make use of multilingual embedding and sophisticated topologies, such as mBERT NLI-M, mBERT-E-MV, and mBERT-E-AS. This relative advantage highlights the effectiveness of proposed method, which is specially designed for SA of Hindi movie reviews. As a result of outperforming both traditional and cutting-edge methods, proposed model not only demonstrates its efficacy but also it’s potential to be a standard in sentiment analysis, especially when applied to multilingual movie reviews. These findings emphasise the importance of suggested approach in developing SA techniques by providing more precise and complex insights into the opinions of the public regarding Hindi movies. The comparison graph is shown in Figure 7 whose values are summarised in Table 4.

Table 4. values are summarised

Method	Acc (%)	Prec (%)	Recall (%)	F1 (%)
NB [25]	87.78	56.66	63.32	59.81
DT [25]	91.62	64.16	64.38	64.27
SMO [25]	91.62	48.60	63.26	54.97
mBERT-QA-M [26]	73.03	77.09	77.52	77.31
mBERT-E-MV [26]	78.09	80.70	77.53	79.08
mBERT-E-AS [26]	79.77	80.95	76.40	78.61
Proposed model	97.64	85.93	99.17	92.20

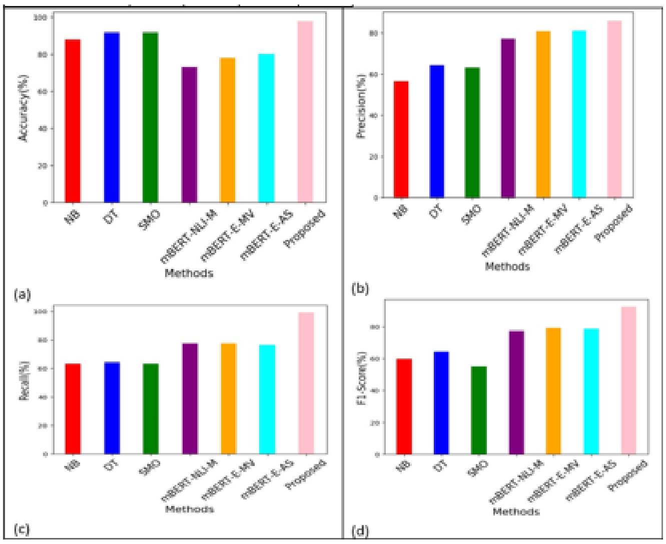

Fig. 7. Performance Comparison chart

There are various reasons why the proposed approach performs better than others. First off, the model's ability to comprehend intricate linguistic syntax and minute changes in sentiment, such as sarcasm and colloquial idioms, is made possible by the incorporation of quantum computing concepts into sentiment analysis. Sequential-based hyperband optimization also makes it possible to explore the hyperparameter space effectively, which finds high-performing model configurations while consuming the least number of computational resources. Furthermore, this findings also demonstrate the promise of quantum computing in improving sentiment analysis methods, especially in languages like Hindi where conventional methods could have trouble correctly capturing subtleties and contextual knowledge. It is imperative to recognize potential constraints, though, like the accessibility of superior training data and the computer power needed for activities using quantum computing. Future directions for investigation could encompass researching innovative methods for preparing data, examining substitute quantum architectures, and broadening the utilization of quantum computing concepts to encompass additional languages and domains.

5. Conclusion

This study proposed methodology for SA of Hindi content that takes advantage of quantum computing and DL concepts and demonstrated significant improvements in sentiment analysis accuracy over traditional methods by combining Q-VAE for data encoding, Q-CNN for feature extraction, and Sequential-based hyperband optimisation for model optimisation. The experiments on a dataset of Hindi movie reviews show that proposed approach exceeds existing approaches in capturing diverse sentiments, such as sarcasm and colloquial language. By leveraging the capability of quantum computing, proposed approach provides a more nuanced comprehension of Hindi text, allowing for more accurate sentiment classification of achieving a high accuracy of 97.64%. However, it is critical to recognise the proposed methodology's limitations, which include dependence on high-quality training data, computational resource requirements, and challenges in handling unusual linguistic expressions. Despite these limitations, this research represents a major advancement in sentiment analysis methodologies for Hindi textual data. Future research should address these limitations and identify areas for improvement. Efforts can be focused towards improving the methodology's robustness and scalability, especially when dealing with unusual or unconventional linguistic expressions in Hindi text.

DECLARATIONS

Ethics approval and consent to participate

This article does not contain any studies with human participants or animals performed by any of the authors

Consent for publication

All contributors agreed and given consent to Publication.

Availability of data and material

Data that has been used is confidential

Competing interests

On behalf of all authors, the corresponding author states that they have no competing interest.

Funding

No fund was received for this work

Authors' contributions

The authors confirm contribution to the paper as follows and all authors reviewed the results and approved the final version of the manuscript.

First author (Corresponding author): Vaibhav Prakash Vasani: Writing original draft, Methodology, study conception and design, analysis and interpretation of results, Reviewing and editing

Second author: Asha Ambhaikar

Conceptualization, data collection, Reviewing and editing

References

[1] S. Mishra, M. Aggarwal, S. Yadav, and Y. Sharma, “An Automated Model for Sentimental Analysis Using Long Short-Term Memory-based Deep Learning Model,” *International Journal of Engineering and Manufacturing*, vol. 13, no. 5, pp. 11-20, 2023.

[2] A. P. Pandian, “Performance evaluation and comparison using deep learning techniques in sentiment analysis,” *Journal of Soft Computing Paradigm (JSCP)*, vol. 3, no. 02, pp. 123-134, 2021.

[3] M. M. Hasan, R. B. Hossain, M. S. Hossain, K. Hasan, A. R. Palash, F. Hasan, and H. Mengdan, “Synergizing Convolutional Neural Networks and Pre-processing for Precision Sentiment Analysis,” *Networks*, vol. 6, no. 9, pp. 53-71, 2023.

[4] C. Suhaeni, and H. S. Yong, “Mitigating Class Imbalance in Sentiment Analysis through GPT-3-Generated Synthetic Sentences,” *Applied Sciences*, vol. 13, no. 17, pp. 9766, 2023.

[5] K. Shrivastava, and S. Kumar, S, “A sentiment analysis system for the hindi language by integrating gated recurrent unit with genetic algorithm,” *Int. Arab J. Inf. Technol*, vol. 17, no. 6, pp. 954-964, 2020.

[6] B. Samanta, N. Ganguly, and S. Chakrabarti, “Improved sentiment detection via label transfer from monolingual to synthetic code-switched text,” arXiv preprint arXiv:1906.05725, 2019.

[7] S. Thara, and P. Poornachandran, “Social media text analytics of Malayalam–English code-mixed using deep learning,” *Journal of big Data*, vol. 9, no. 1, pp. 45, 2022.

[8] M. S. Başarslan, and F. Kayaalp, “Sentiment analysis using a deep ensemble learning model,” *Multimedia Tools and Applications*, pp. 1-25, 2023.

[9] A. Madasu, and S. Elango, “Efficient feature selection techniques for sentiment analysis,” *Multimedia Tools and Applications*, vol. 79, no. 9, pp. 6313-6335, 2020.

[10] M. Archana, and T. Velmurugan, “IMPACT OF CUSTOMER REVIEWS ON PURCHASE BASED DATA USING SENTIMENT ANALYSIS WITH MACHINE LEARNING ALGORITHM.,”

[11] M. Islam, A. Anjum, T. Ahsan and L. Wang, “Dimensionality reduction for sentiment classification using machine learning classifiers,”*In 2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 3097-3103, 2019.

[12] Kennedy, B., Ashokkumar, A., Boyd, R. L., & Dehghani, M. (2021). Text analysis for psychology: Methods, principles, and practices.

[13] S. T. Kokab, S. Asghar, and S. Naz, “Transformer-based deep learning models for the sentiment analysis of social media data,” *Array*, vol. 14, pp. 100157, 2022.

[14] A. Topbaş, A. Jamil, A. A. Hameed, S. M. Ali, S. Bazai, and S. A. Shah, “Sentiment analysis for covid-19 tweets using recurrent neural network (rnn) and bidirectional encoder representations (bert) models,” *In 2021 International Conference on Computing,*

Electronic and Electrical Engineering (ICE Cube), pp. 1-6, 2021.

[15] R. Pradhan, and D. K. Sharma, “RETRACTED ARTICLE: An ensemble deep learning classifier for sentiment analysis on code-mix Hindi–English data,” *Soft Computing*, vol. 27, no. 15, pp. 11053-11053, 2023.

[16] C. I.. Eke, A. A. Norman, and L. Shuib, “Context-based feature technique for sarcasm identification in benchmark datasets using deep learning and BERT model,” *IEEE Access*, vol. 9, pp. 48501-48518, 2021.

[17] V. P. Dupakuntla, H. Veeraboina, M. V. K Reddy, M. M. Satyanarayana, and Y. Sameer, “Learning based approach for Hindi text sentiment analysis using Naive Bayes classifier” *LEARNING*, vol. 7, no. 8, 2020.

[18] O. Yadav, R. Patel, Y. Shah, and S. Talim, “Sentiment analysis on Hindi news articles,” *International Research Journal of Engineering and Technology (IRJET)*, vol. 7, 05, 2020.

[19] S. Rani, and P. Kumar, “Deep learning based sentiment analysis using convolution neural network,”*Arabian Journal for Science and Engineering*, vol. 44, pp. 3305-3314, 2019.

[20] K. Shrivastava, and S. Kumar, “A sentiment analysis system for the hindi language by integrating gated recurrent unit with genetic algorithm,”. *Int. Arab J. Inf. Technol*, vol. 17, no. 6, pp. 954-964, (2020).

[21] A. Sharma, and U. Ghose, “Toward Machine Learning Based Binary Sentiment Classification of Movie Reviews for Resource Restraint Language (RRL)–Hindi” *IEEE Access*, 2023.

[22] V. Jain, and K. L. Kashyap, “Ensemble hybrid model for Hindi COVID-19 text classification with metaheuristic optimization algorithm,” *Multimedia Tools and Applications*, vol. 82, no. 11, pp. 16839-16859, 2023.

[23] H. Kwon, H. Lee, and J. Bae, “Feature Map for Quantum Data: Probabilistic Manipulation,”. arXiv preprint arXiv:2303.15665, 2023.

[24] M. S. Akhtar, A. Ekbal, and P. Bhattacharyya, “Aspect based sentiment analysis: category detection and sentiment classification for Hindi,” *In Computational Linguistics and Intelligent Text Processing: 17th International Conference, CICLing 2016, Konya, Turkey, April 3–9, 2016, Revised Selected Papers, Part II*, vol. 17 , pp. 246-257, , 2018.

[25] A. Pathak, S. Kumar, P. P. Roy, and B. G. Kim, “Aspect-based sentiment analysis in Hindi language by ensembling pre-trained mBERT models,” *Electronics*, vol. 10, no. 21, pp. 2641, 2021.