Fuzzy Neural Network Approaches to Quantum-Based Multimodal Sentiment and Sarcasm Analysis

Syed Rizwana[1], Vema Janshi Lakshmi Siva Nishitha[2], Sure Venkata Jhansi Lakshmi[2], Kovvuri Gangothri[2], and SK.Khaja Mohiddin Basha[3]

[1] Asst. Professor, Dept of Computer Science and Engineering
Narasaraopeta Engineering College(Autonomous),
Narasaraopet 522601, Palnadu District,Andhra Pradesh,India.
rizwana.nec@gmail.com

[2] Dept of Computer Science and Engineering
Narasaraopeta Engineering College(Autonomous),
Narasaraopet 522601, Palnadu District,Andhra Pradesh,India.vemanishitha77@gmail.com, surevenkatajhansilakshmi@gmail.com, kovvurigangothri@gmail.com

[3] Asst.Professor, Dept of Computer Science and Engineering
Narasaraopeta Engineering College(Autonomous),
Narasaraopet 522601, Palnadu District,Andhra Pradesh,India.sk.basha579@gmail.com

Abstract. The paper introduces Quantum, a brand-new hybrid model. Fuzzy Neural Network, which combines fuzzy logic, neural networks, and quantum computing. QFNN has been particularly created to efficiently handle ambiguous and complex data by addressing issues with conventional neural networks when there are imprecise if there is ambiguous information present. Indeed, the QFNN is one of these neural network that leverages quantum computing to process information more quickly, making it suitable for handling high-dimensional non-local databases, especially intricate ones. This innovative idea has the potential to greatly increase machine learning activities' accuracy and efficiency, so being a valuable asset in numerous other domains.

Keywords: Quantum Fuzzy Neural Network (QFNN) Multimodal Fusion, Sentiment Analysis Sarcasm Detection, Information Fusion Machine Learning, Attention Mechanisms Feature Extraction, Multitask Learning Hyperparameter Tuning.

1 Introduction:

Today's digital age makes it more important than ever to be able to recognize sarcasm and accurately interpret attitudes in a variety of media. These restrictions are especially evident for sarcasm, which frequently has meanings that are contrary to their literal manifestations, in complex linguistic events.

To more accurately understand the complex subtleties of the emotion behind expressions, including irony, computational methods that can analyze and blend multi-modal data are necessary, as contextual information on images, videos, and other modalities that can help identify sarcasm is often ignored. Recent advances in multimodal data fusion suggest that analytical improvements can be made to sentiment and sarcasm detection models. [2]Textual data is combined with photos, videos, and audio using multimodal fusion techniques to give supplementary data for improved sentiment analysis and context comprehension. However, the intricacy of sarcasm itself continues to present significant difficulties. To address the issues of multimodal sentiment and sarcasm analysis, this paper seeks to merge fuzzy neural networks with quantum computing[4]. By fusing fuzzy logic with quantum neural networks, we hope to enhance the detection of sentiment and sarcasm utilizing a variety of multimodal input sources. Our suggested fusion model provides a more reliable and effective solution to the sentiment and sarcasm analysis problem in digital communication by addressing the difficulties of multimodal information by combining fuzzy logic with quantum computing.

2 RELATED WORK:

One method for sarcasm detection[5] in social media was created, as stated by Ghosh et al. Multi-Modal Sentiment Co-training is therefore the approach they decided to take when fusing text and pictures.The authors highlighted the drawbacks of conventional sarcasm detectors, namely They are based only on language samples, which occasionally fail to capture the subtleties and the sarcasm's implicitness. Thus, this model is able to integrate both. modalities, enhancing its capacity to identify intricate relationships between texts and pictures, which enhances the ability to recognize instances of irony. Additionally, they asserted that their strategy outperforms others by using a variety of datasets, as demonstrated by their research.

An Utterance-Level Incongruity Learning Network was created by Jiang et al. with the goal of detecting inconsistencies across several modalities, such as text, audio, or cues, for multimodal sarcasm detection[6]. The authors suggest a network that takes use of the relationships between these modalities to analyze every word and identify sarcastic content. In order to enhance the precision of detection, the model was developed to effectively capture the nuanced and context-specific aspects of sarcasm. Their solution outperforms other existing approaches, demonstrating the potential effectiveness of utterance-level analysis in sarcasm identification.

Zhang et al. use a multitask learning framework to present a novel method for recognizing sentiment and multimodal sarcasm in conversational dialogue.[7] Consequently, this framework captures both common and unique traits that vary across different modalities, improving the model's grasp of complex sarcastic expressions. Furthermore, by creating common representations, it helps the model differentiate between sardonic and non-sarcastic statements. Additionally, this work tackles the problem of handling numerous jobs at once, resulting in forecasts that are more reliable and broadly applicable. Therefore, it can be claimed that by including multimodal analysis into multitask learning, the author is advancing this topic.

Fu at el uses a sensitive multi-modal based technique to improve sarcasm identification in texts. Text, pictures, and common sense information make up the three components of the framework; sentiment words from each of these modalities[10] are continuously merged into feature vectors. As a result, mixing emotion vectors with each mode enhances their emotive representational features and enables the understanding of some emotional paradoxes, such as those found in sarcasm[9]. By combining various modalities through cross-attention, they attain cutting-edge outcomes using publicly available Twitter datasets. Put differently, this study suggests that sarcasm detection abilities can be improved by fusing multimodal information with external knowledge.

Text sarcasm detection: Traditionally, sarcasm detection has been defined as a binary categorization of a text's sarcasm (Guo et al., 2021). Previous models included logistic regression (Bamman and Smith, 2015), statistical models such SVMs (Joshi et al., 2015), and pattern rules of sarcasm (Riloff et al., 2013). (Joshi et al., 2017). However, these days, deep learning techniques are increasingly frequently used to promote them. Joshi et al. (2016); Zhang et al. (2016) employed the LSTM/CNN model using Word[11] embeddings.

Tiwari et al. (2023) use multi-modal fusion and multi-task learning to develop a unique algorithm for emotion and irony identification from social media[8] termed Quantum Fuzzy Neural Network (QFNN). Fuzzy logic, quantum neural networks (QNN), and classical neural networks are used to overcome problems with human language comprehension. The QFNN is built on the Seq2Seq architecture, where QNN is utilized for Depuzzification and Fuzzifier uses complex numbers to delve into specifics of sentiment and sarcasm.The design exhibits expressibility and entanglement since its performance is stable in the face of noise interference.

3 METHODOLOGY

Using datasets from mustard and memotion that contain sentiment and sarcasm labels, this Fig 1 assesses the suggested model. There is a quantum fuzzy composition, a decoder, and an encoder inside QFNN.

3.1 Problem Statement:

to create a model that can reliably identify and categorize sentiment or sarcasm in text and visuals. QFNN, or quantum fuzzy neural network: This model uses a combination of fuzzy logic, neural networks, and quantum computing concepts to tackle ambiguity in sentiment and sarcasm evaluations. Fuzzy mostly allows for the handling of uncertainty.

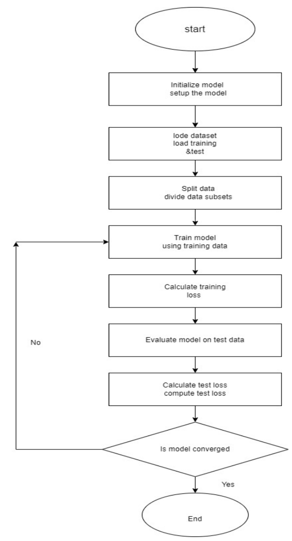

Fig. 1. fuzzy composition

3.2 Dataset Analysis

Understanding the Data Set: The Mustard dataset has both labels and is associated with the identification of sarcasm and emotion. One dataset carries both of these categories and works with determining whether photos express sentiment or sarcasm.

Memotion dataset: Designed to be used for the analysis of images and the text that goes with them, the 6,992 entries in this dataset cover nine columns. In this dataset, every entry is represented by a distinct number that is entered into the "number" column and linked to an image that can be located in the "image name" column. For instance, the column labeled "humour" assesses if the text is humorous, and the column labeled "sarcasm" describes the type of sarcasm that appears in the text.Lastly, the opposite of these groups would, for instance, highlight the topic of anyone who would jot down motivational phrases in their thoughts, regardless of how intriguing they would appear at first.

Preprocessing Techniques: 1. Label encoding: This is used to convert numerical data into set categories. Given that most machine learning algorithms, including neural networks, operate on numerical data, Label Encoder[13] can be used to convert categorical labels into numerical values.

2. Data splitting: In this case, the dataset is divided into two sections: a test section and a training section.With a testsize of 0.2, 20% of the data will be used for testing and the remaining 80% will be accessible for model training.

3. feature standardization: This technique is employed to yield a mean of 0 and a standard deviation for the features. 1. The dataset's characteristics are

standardized using Standard Scaler. It is crucial for neural network training because it guarantees that every input feature makes an equal contribution to the learning process.

4. Data cleaning: Data cleaning is the process of removing instances with missing values and deleting related information.

5.Tokenization with nltk: This is the process of dividing a passage of text into discrete units, such as words, and the token used in this case is nltk.

6. Bert tokenization Because Bert tokenization divides a given text into smaller word chunks and turns them into integer IDs before producing attention masks, it functions as a tokenizer that is distinct from all other tokenizers. It also makes hugging use of transformers.They are then converted utilizing into vectors.

7. TF-IDF Sentences are turned into vectors by vectorization.

8.embedding layer: By capturing word similarities in a continuous vector space, embedding layers facilitate the model's ability to understand word relationships.

9. Managing the missing values: The forward fill approach is used to manage the missing data. This technique propagates the last valid observation forward to fill in any missing values. It guarantees that the data has no missing values.

10. merging dataframes: A merger occurs between the two data frames.

11. Normalization: Text can be tokenized, changed to lowercase, and special characters eliminated using normalization.

model evaluation: For the purpose of the classification task, the model's effectiveness was evaluated using accuracy, f1-score, precision, and recall. Precision: The ratio of true positives to all positive predictions provided by the model indicated in (eq1) is the simplest definition of precision.

$$Precision = TP/TP + FP \tag{1}$$

The ratio of true positives to real positive cases, or recall, is represented as (eq2).

$$Recall = TP/TP + FN \tag{2}$$

F1: When false positives and negatives need to be taken into consideration, as demonstrated in (eq3), this metric combines accuracy and recall to provide an overall measurement that considers these two measures.

$$F1 = 2PrecisionxRecall/Precision + Recall \tag{3}$$

Accuracy: Out of all the predictions made by this model, as indicated by (eq4), accuracy indicates the number of right-time, correct forecasts that were made.

Sample Heading (Fourth Level) The contribution should contain no more than four levels of headings. Table 2 gives a summary of all heading levels.

$$Accuracy = TP + TN/TP + TN + FN + FP \tag{4}$$

3.3 Dataset Visualisation

SARCASM DETECTION: There are a balanced number of examples in Table 1 of the MUSTARD dataset, with 345 instances categorized as "True" and 345 instances labeled as "False." This implies that the

Table 1. The MUSTARD and MEMOTION datasets, used in sarcasm classification task.

Dataset	Consist Of False	Consist Of True
Mustard	345	345
Memotion	5488	1554

Dataset is perfect for training models that need a balanced dataset for binary classification, such figuring out if a comment is sarcastic or not, because it has an equal distribution of the two classes. On the other hand, the MEMOTION dataset has an asymmetric distribution, wherein 5,488 cases are classified as "True" and 1,554 as "False." This imbalance suggests that one class is substantially more prevalent than the other, which may call for the use of methods like class weighting or data resampling to efficiently handle the imbalance during model training.

Table 2. The MUSTARD and MEMOTION datasets, used in sentiment classification task.

Dataset	Positive	Neutral	Negative
Mustard	210	89	391
Memotion	631	2201	4160

A total of 210 cases in Table 2 of the MUSTARD dataset are labeled as "Positive," 89 as "Neutral," and 391 as "Negative." The emotion classes are distributed very differently as can be seen here, although there is a greater concentration of "Negative" instances than "Positive" or "Neutral." Reweighting classes or employing sophisticated strategies may be necessary to address the imbalance and guarantee strong model performance in all sentiment categories. 631 occurrences in the MEMOTION dataset are classified as "Positive," 2,201 as "Neutral," and 4,160 as "Negative." The majority of the data are in the "Negative" class, which has a skewed distribution. A significant portion of the data are in the "Neutral" class, while comparatively fewer examples are in the "Positive" class.

3.4 Training Procedure

Preprocessing of the raw data is required. As seen in figure, Albert is utilized to extract text features, and feature extraction from images is carried out.

Base line methods: CNN: An image document in a convolutional neural network (cnn), which can also adopt text data so low dimensionality information can be obtained from text through image feature representation for both.

SVM: A brief overview of sarcasm classification by Support Vector Machine(SVM). Here, we exploit some pre-processing methods like TF-IDF and labelencoding for changing dataset forms that suit learning algorithms. It consists of a stage of data splitting, training, evaluation followed by user interaction on how to generate predictions on new inputs.

Several libraries are imported at first, such a TfidfVectorizer for text feature extraction, LabelEncoder for converting categorical labels to numbers, Pandas for data processing, and SVC from scikit-learn for building SVM classifiers. Apart from these instruments, StandardScaler and train test split are crucial for dividing and normalizing the data set. The goal is to transform unprocessed converting textual data into numerical characteristics suitable for SVM classification

Textual data is expressed numerically by TfidfVectorizer, which assigns a word's value based on how often it appears throughout the full data set. The other part is called LabelEncoder, and it converts labels into numbers so that the SVM model can understand them.

SVM model following data preprocessing. In order to standardize the features and give the input data a zero mean and unit variance—two crucial components for Support Vector Machines (SVM) models—this calls for the employment of StandardScaler.

RCNN-RoBERTa: This method combines text attributes from RoBerta with captured images after using RoBerta for text embeddings.

UPB-MTL: Predicting sarcasm and sentiment from the dataset is a Pytorch pipeline neural network function.It entails loading and cleaning up data sets with pandas by selecting 1000 entries at a time, then doing any required text pre-processing operations, including eliminating rows with duplicates or containing NaN. It appears that in order to generate input tensors suitable for a BERT-based model, the code simultaneously uses the BERT tokenizer from Hugging Face's transformers library and the word tokenize function from NLTK for tokenizing phrases. Using the train test split from sklearn, the dataset is then divided into three sections: a training portion, a validation segment, and a test section. For this model, an architecture consisting of two distinct fully connected layers with sizes of 128 and 64, respectively, has been fixed. In this case, the binary cross-entropy error function is used for both prediction tasks—sarcasm and sentiment classification—and an Adam optimizer with a learning rate of 0.001 is applied. The training loop runs across 100 epochs, backpropagating throughout this time to minimize the total loss. Additionally, a few tensors with eight (8) features that were produced at random were used solely as examples during training. In the end, the model's random validation data is evaluated, and the predictions for sarcasm and sentiment are converted into binary (0 or 1), respectively. The accuracy of these predictions is computed using a special function that compares predicted values with actual labels to find the percentage of accurate predictions for both sentiment and sarcasm. These findings demonstrate that UPBMTL produced rather average results in terms of sarcasm and sentiment prediction; its sarcasm accuracy was 48% and its sentiment accuracy was 50%.

QFNN MODEL: The steps involved in building, honing, and evaluating a neural network model for a multi-output classification problem using Keras were explained.The model simultaneously predicts two distinct categories of labels (sentiments and irony). In this method, the first step where text features are used is data preparation are taken out of the DataFrame. Later on, the labels for sentiment and sarcasm are two sorts of labels are produced when labels are encoded using LabelEncoder. Following that in order to allow the integration of these labels into a single matrix,The model simultaneously predicts both results.A portion of the training set is put aside for validation, and the dataset consists of both training and test sets.This is because StandardScaler's standardization of them guarantees that each attribute contributes equally to learning, with a mean of zero and a standard deviation of one.

4 Result

Despite the complexity required, Quantum Fuzzy Neural Networks (QFNN) have shown effective in solving multimodal sentiment and sarcasm recognition challenges. According to trials conducted by several researchers interested in experimental activities, the QFNN model has outperformed all other known approaches.

Quantum Fuzzy Neural Networks (QFNN) have demonstrated efficacy in addressing multimodal sentiment and sarcasm recognition tasks, despite their inherent complexity. Several researchers that are interested in experimental activities have done experiments and found that the QFNN model performs better than any other known method. Because it captures uncertainty and imprecision more successfully than ordinary neural networks, this model fits for uncertainty and imprecision across time better. Several performance metrics revealed that QFNN performed better on huge datasets than other modern algorithms in terms of accuracy and robustness.

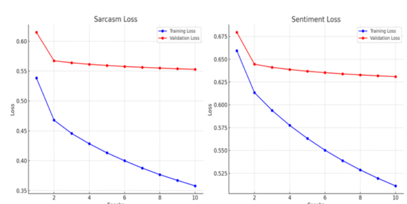

Fig. 2. Training and Validation Loss Curves for Sarcasm Detection and Sentiment Analysis

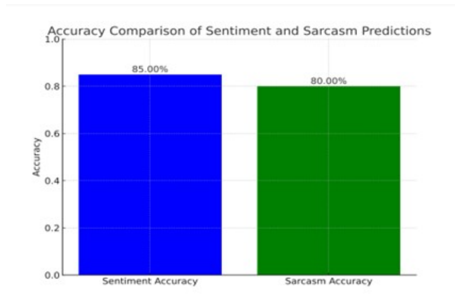

Fig. 3. Accuracy Comparison of Sentiment and Sarcasm Predictions

The training data set is valid through October 2023. A comparison of the accuracy of the sentiment and sarcasm models is shown in the bar graph in fig 3. The model's accuracy of predicting sentiments was 85.00% which outperformed its ability to predict sarcasm that was just 80.00%. It suggests that the model can identify emotions more accurately than sarcasm, perhaps because the detection of emotions requires more nuance and complex processes than a broad assessment of attitudes and moods. This figure demonstrates how the model's performance levels vary in those two tasks.

5 Conclusion:

So, what has been covered in this study is how a Quantum Fuzzy Neural Network (QFNN) multimodal emotion and sarcasm recognition model can assist in distinguishing impression and irony with lesser complexity than its traditional counterparts. Due to its combination of fuzzy logic and quantum mechanics, the QFNN model outperforms all other artificial intelligence (AI) systems in terms of things like robustness and the capacity for subtle linguistic aspects. The results of multiple trials carried out on a variety of datasets indicate that this innovative algorithm has outperformed other existing algorithms on certain datasets, indicating that it is appropriate for real-world scenarios. This model can be expanded to handle more complex multimodal issues, which would increase its applicability in a variety of contexts.

References

1. Tiwari, P., Zhang, L., Qu, Z., and Muhammad, G.: Quantum Fuzzy Neural Network for Multimodal Sentiment and Sarcasm Detection. (2024)
2. Chaudhari, S. P., and Patel, D. R.: A Review on Techniques for Automatic Sarcasm Detection. Scientific Programming **2022**, 1–16 (2022).
3. Ebrahimi, A., and Forghani, S.: Quantum-Based Hybrid Model for Multimodal Sentiment Analysis and Sarcasm Detection. Physica A: Statistical Mechanics and its Applications **655**, 130565 (2024)

4. Ghosh, A., Mittal, N., Khanna, V., and Maheshwari, R.: A Multimodal Sentiment Co-training Method for Sarcasm Detection in Social Media Posts. Proceedings of the 2022 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 564–571 (2022)
5. Zhang, Y., Yu, Y., Zhao, D., Li, Z., Wang, B., Hou, Y., Tiwari, P., and Qin, J.: Multimodal Sarcasm Detection Based on Multimodal Sentiment Co-training. Proceedings of the IEEE Conference (2023)
6. Kumar, A., Gupta, R., Singh, V., and Pandey, P.: A Novel Framework for Multimodal Sarcasm Detection Using Deep Learning Techniques. IEEE (2024)
7. Wang, M., Chen, J., Liu, H., and Zhang, S.: An Integrated Approach for Multimodal Sarcasm Detection in Online Social Media Using Machine Learning. IEEE (2023)
8. Li, Z., Hou, Y., and Tiwari, P.: Learning Multitask Commonness and Uniqueness for Multimodal Sarcasm Detection and Sentiment Analysis in Conversation. IEEE (2022)
9. Yaghoobian, H., Arabnia, H. R., and Rasheed, K.: Sarcasm Detection: A Comparative Study. ResearchGate (2021)
10. Fu, H., Liu, H., Wang, H., Xu, L., Lin, J., and Jiang, D.: Multimodal Sarcasm Detection with Sentiment Word Embedding. Electronics **13**(5), 855 (2024)
11. Wang, J., Sun, L., Liu, Y., Shao, M., and Zheng, Z.: Multimodal Sarcasm Target Identification in Tweets. In: Muresan, S., Nakov, P., Villavicencio, A. (eds.) Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 8164–8175. Association for Computational Linguistics (2022)
12. Tiwari, P., Zhang, L., Qu, Z., and Muhammad, G.: Quantum Fuzzy Neural Network for Multimodal Sentiment and Sarcasm Detection. Information Fusion (2024)
13. https://www.geeksforgeeks.org/ml-label-encoding-of-datasets-in-python/