International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 06 | June 2021 www.irjet.net p-ISSN: 2395-0072

An Advance Approach toward Sentiment Analysis using Quantum AI

Riddhi Mandal[1], Pragati Bhat[2]

[1-2]*Department of Artificial Intelligence, G H Raisoni College of Engineering, Nagpur, Maharashtra, India.*

Abstract: *Quantum Artificial Intelligence(AI) is a multidisciplinary field which mainly focuses on building quantum algorithms for improving computational tasks within artificial intelligence. Gender identification is one of the major problems in voice recognition. The recognition techniques suffer from the overfitting issues. To overcome this overfitting issue, this paper designs a novel ensemble-based quantum AI. The main objective of the paper is to identify whether the speaker is male or female by using quantum computing. Identifying the gender from the voice data set i.e., having pitch, median, frequency etc. kind of properties can be possible by using Quantum AI. The proposed system can be used to find the best algorithm among all quantum AI algorithms like Quantum K-nearest neighbors (Q-kNN), Quantum Random Forest, Quantum Logistic Regression, Quantum Decision Tree and Quantum support vector machine (QSVM) to detect the gender of the speaker with maximum possible efficiency and accuracy.*

Keywords: Sentiment Analysis, Voice Classification, Quantum AI Algorithms, Quantum Computation, Qubit.

1.INTRODUCTION

Recognition of a person by voice is one of the forms of biometric authentication, useful for identifying the person by a combination of unique voice characteristics. Gender Classification or Voice recognition of any person is very important nowadays because it can increase the efficiency of various applications like for marketing, voice identification and online advertisements involving voice. There are many genders with different voices around us[4]. The voice i.e

speech is a best way of communication which consists of language, age, emotional state as well as the person's gender. Any voice having sound waves that are different in all human beings where different frequencies carry different sound waves. Identifying a gender based on voice that deploys various applications like advertising, investigating in crime scenarios, marketing strategies and to enhance human computer interaction systems mainly to improve level of user satisfaction which depends on gender voice by customer services[5]

Identifying gender in any voice using pitch, aims to find the speaker which is made by analysing various sampledata of voice containing different parameters. The system which is made for finding the speaker is male or female is mainly based on pitch of the voice signal. There is slight variation between male and female pitch of voice which helps to identify the gender of the speaker.
Machine learning aims to generate simple classifiers which will be easily understood by human beings. Our main aim is to increase the computation power using various algorithms with quantum[6]. In this paper, we are using quantum with various machine learning algorithms like knn, random forest, decision tree, Support vector machine and logistic regression.

2. RELATED WORK

Gender-based recognition and voice segregation have been used for a long time. We have used quantum concepts in machine learning algorithms to improve the gender recognition system. According to gender-based research it shows that the word is first converted to different parameters based on different parameters. The main parameters are pitch and quantity. The classification is designed to differentiate between men and women. To create this program, it is first trained with the training data and test data presented and tested for the effectiveness of this information system. The results are different for different algorithms and practice to produce different results at different times.

Gender segregation using frequency and pitch for training and data testing shows that Logistic Regression is advanced and its accuracy is 92% and compared to other algorithms such as regression, Random Forest and voice data speaking the same language works better in Random Forest is also 93% accurate [8]. It therefore indicates that Random Forest will be suitable for speech recognition depending on the frequency and pitch to distinguish male and female [9]. Additional tuning is based on the capturing process to improve the efficiency of the produced results.

Voice-based word output proves that the algorithm works better with segregation that provides better results for vowel extraction in male samples [10]. As the samples are trained and tested, they produce the result accordingly.According to this paper it has been found that by increasing the unauthorized part of the speech, it increases the pitch in the event of male samples. Similarly, by increasing the volume of the spoken word, it reduces the volume of the voice. Therefore, it failed to identify it when the speaker spoke in two different tones.

The Gaussian Mixture Modeling (GMM) gender-based classification is analyzed by age, names, etc. A combination model is used and indicates an accuracy of up to 98%, one of the most effective methods of verbal gender analysis. The classification is based on integrated pitch parameters and spectral perceptual linear predictive (Rasta-Plp) coefficients modeling for male and female expression.

Speech recognition in adults means that they can automatically adjust the volume and volume and can sound like men and women. Therefore, it is difficult to distinguish between a man and a woman. Some women's voices are difficult to decipher based on the throat as a single aspect of women's vocabulary does not meet our needs. The different parameters were female users and varied from female to female which is why the database needs to be processed based on this before separating male and female. The basic frequency has a combination of grammar, metaphorical and non-verbal details and these three are related to the masculine and feminine and depend on high pitch and tone of voice [13]. This has succeeded in setting the frequency without general information and no member-external information. Speaker volume varies between high and low frequencies between speakers. Gender identification by the SVM shows that gender expression is analyzed by various forms of speech such as suppressed speech, telephone conversation and linguistic diversity and so on [14].

3. SYSTEM ARCHITECTURE

System Architecture illustrates the overall process of the proposed system. From Fig. 1 the voice dataset is loaded using a technique like normalization. Then the dataset is splitted into training and testing sets in the ratio of 80% and 20% respectively.Various Quantum Ml algorithms such as QKNN,QSVM,Q-Decision Tree etc. are trained on the training dataset. The trained model is now tested by using the testing dataset to calculate the evaluation metrics like accuracy, precision, Recall, F1 Score. Then Compare the the results of accuracy of various algorithms to find the best algorithm with the best accuracy.

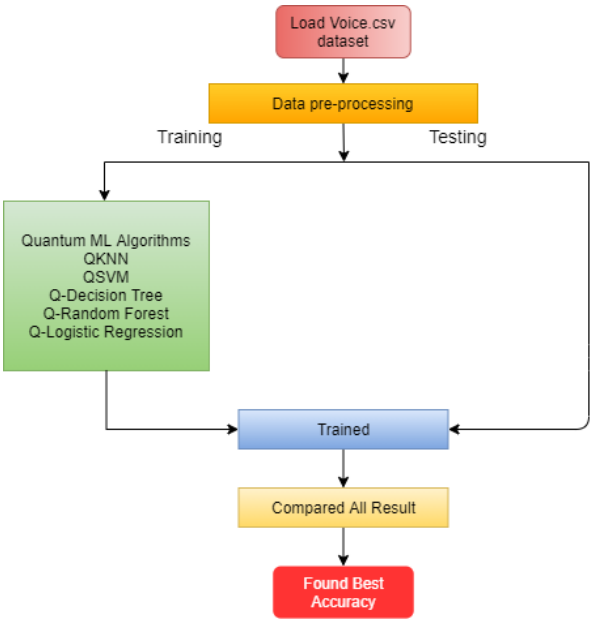

Fig.1 System Architecture

1.Methodology:

Dataset: The voice dataset having various attributes which makes it easy to identify the speaker's gender i.e. they are male or female. The dataset has 21 attributes and nearly 3164 samples of voice with all different parameters and they are collected from kaggle.com. The samples are pre-processed by acoustic analysis in R using the seewave and tuneR packages, with an analyzed frequency range of 0hz-280hz. The attributes are sd, meanfreq, median, Q25, Q75, IQR, skew, Kurt, sp.ent, sfm, mode, centroid, meanfun, minfun, maxfun, meandom, mindom, maxdom, dfrange, modindx, label. The target attribute is a label which has male and female values.

Data pre-processing: Data cleaning, data transformation and data reduction are the techniques used to pre-processed the dataset. As shown in Fig.2, after the preprocessing it was sent to quantum ML algorithms where it will train and give the predicted result (male or female). Data pre-processing is performed

before storing the data in the database. Using this process, we will find missing values and avoid noisy data (irrelevant).

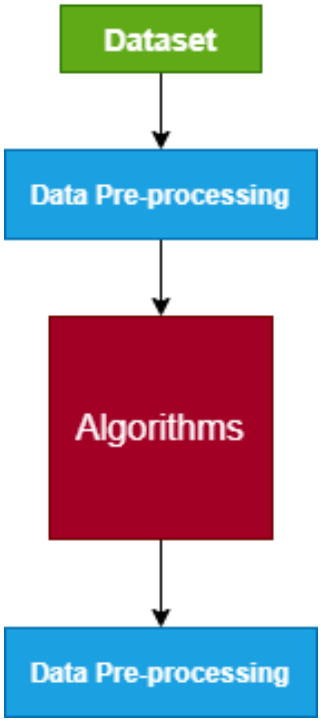

Fig.2 System model of Data Pre-processing

2. Implementation:

A. Quantum K-nearest Neighbour: K nearest Neighbour works on the principle of classifying the data based on the euclidean distance, which is the distance between two data points.It classifies the dataset into various different classes,when a new data point is inputted for testing the algorithm chooses the nearest K values which will be closest to the new data point.The value of K can be determined by finding the square root of the number of rows inputted in the dataset.Among the k closest values the algorithm will decide which class the new data resembles best by finding the maximum number of k points closest to the data.On classifical computers which means normal computers this algorithm takes quite a lot of memory which is generally stored in bits which takes more time to compute whereas on other hand if we use quantum computing(QKNN)[15]. where hamming distance,the number of bit positions in which the two bits are different, is used which is more easy for quantum based algorithms as in quantum computing the anomaly of classical knn gets solved as in quantum knn the qubits are used to store the data which when in entangled state can reduce time complexity by huge amount.When qubits are set to superposition they can be in multiple states at the same time which helps to save computation time and also when less qubits the algorithm becomes more power efficient.

B. Quantum Random Forest: Random forest works on the principle of building multiple decision trees and merging them to improve the precision and more accurate than individual decision trees.Random forest uses various methods like bootstrapping ,which means taking a dataset and creating many small samples from the dataset, random subsets of the features which are the independent variables of the dataset and average voting to make predictions. For our system we are using a bootstrapping method. We are constructing random forests using quantum computing[16].

Let A be a number of trees in a forest. It is used for predictions of each tree and getting a result. We get results from each tree. Random forest averages a result for a regression problem. A result of a random forest model is obtained by voting of all trees for classification problems.On other hand, the quantum version of random forest uses quantum amplitude amplification algorithm based on Grover's search quantum algorithm,which is a quantum search algorithm which can speed up an unstructured search problem quadraticallyThe quantum amplitude function uses trained trees of the random forest .The quantum procedure allows us to test in parallel an input object on each tree.

C. Quantum Decision Tree: Decision trees are popular data mining and supervised machine learning algorithms used for classification and regression.Decision tree is an inverted tree where each node of tree represents the features which are the independent variables of the dataset,the link or the branch of the tree represents a decision and the leaf nodes ,the last nodes of the tree containing no child nodes.[7] As we know, the decision tree classifier learns from a training dataset which contains observations about objects, which are either obtained analytically or acquired from experts.

In a quantum world, the training data consist of quantum objects instead of classical observations on classical data.In decision tree we traverse down the tree by calculating the probabilities of the nodes,but on complex dataset this task becomes tedious for classical computers making it time consuming ,on the other hand if we use quantum decision tree which uses qubits which are represented in vector form this calculation part becomes a lot faster as entangling the qubits helps to communicate easily and reversibility helps it to execute faster.

D. Quantum Logistic Regression:

Logistic Regression is used in machine learning mainly to solve classification problems i.e. the output will be divided or be put into a particular class or group like apple,mangoes.Logistic regression are used when the data categorical (can be put into categories) than being continuous which happens in the case of linear regression.In logistic regression a relationship between the independent and the dependent variables is found which is the slope and a graph gets plotted.The slope is formed in 's' shape in contrast to the straight line formed in the linear regression, the 's' shape in moves from 0 to 1 which means whether the data belongs to a group or not. Using quantum for solving heavy calculations makes less time complexity when the quantum algorithms perform on quantum computers[17]. Quantum Computers improve the computational efficiency of the linear regression model.

E. Quantum Support Vector Machine:

Quantum Support vector machine is a quantum version of SVM which uses a quantum rule for performing computation tasks[18]. Support Vector machines are supervised learning models which are associated with learning models that analyse the data used for classification and regression but mainly used for classification problems.Support vector machines takes the data points of a dataset and outputs the hyperplane, which in two dimensions is simply a line which separates the tags or the different classes/groups in the best and accurate form. For this model we have used quantum feature maps for translating the classical data into quantum states and then kernel matrix build using quantum states.

4. Result:

AI is transforming many industries. Quantum AI algorithms produce outputs which are superior in terms of sensitivity, the specificity and accuracy. The use of Quantum in the field of artificial intelligence or machine learning is open, and the development of the area to researchers will significantly improve the system performance of the customers and decision-makers to invest in this area. Improving the results can help the sellers to gain valuable feedback(as customer satisfaction with their products), and is able to release the government agencies' understanding of e-government and e-control. This indicates that voice classification or recognition is the most dynamic area of research, and to improve the performance in sentiment analysis, this area has much potential to be discovered further.

5. Conclusion:

In this paper, the model will easily classify the gender from voice i.e. into male and female accurately and efficiently. Using quantum computing with a machine learning algorithm increases the computation power. Q-kNN, Quantum Random Forest, Quantum Logistic Regression, Quantum Decision Tree, QSVM are the algorithms we will implement for speaker recognition. Models with good performance will help to use and develop voice-based gender recognition systems by classifying the voice more effectively in a wide range of aspects. The study of machine learning in the field of quantum is still in the building stage. Future work is to add more algorithms with quantum computing and compare the performance.

REFERENCES

1. A Raahul, R Sapthagiri, K Pankaj and V Vijayarajan."Voice based gender classification using machine learning" Volume 263, Issue 4
2. R. Shiva Shankar, J. Raghaveni, Pravallika Rudraraju, Y.Vineela Sravya. "Classification of Gender by Voice Recognition Using Machine Learning Algorithms" Vol. 29 No. 06 (2020): Vol. 29 No. 06 (2020)

3. Rupinderdeep Kaur, R. K. Sharma and Parteek Kumar. "An efficient speaker recognition using quantum neural network" https://doi.org/10.1142/S0217984918503840
4. GhazaalaYasmin, SuchibrotaDutt, ArijitGhosal (2017) "Discrimination of Male and Female Voice Using Occurrence Pattern of Spectral Flux", 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), pp.576-581.
5. Rami S. Alkhawaldeh (2019) "Gender Recognition of Human Speech Using One-Dimensional Conventional Neural Network", Hindawi Scientific Programming, Volume 2019, pp.1-11.
6. Vedant Bahel(2019) "Quantum Intelligent Systems: A Review "
7. Samuel L. Braunstein "Quantum decision tree classifier" 2013
8. Ericsdotter C and Ericsson A M 2001 Gender differences in vowel duration in read Swedish: Preliminary results Working Papers - Lund University Department of Linguistics
9. Whiteside S P 1996 Temporal-Based Acoustic-Phonetic Patterns in Read Speech: Some Evidence for Speaker Sex Differences,J International Phonetic Association
10. Byrd D 1992 Preliminary results on speaker-dependent variation in the TIMIT database J AcoustSoc Am
11. Zeng Y M, Wu Z Y, Falk T and Chan W Y 2006 Robust GMM based gender classification using pitch and RASTA-PLP parameters of speech Proceedings of the International Conference on Machine Learning and Cybernetics
12. Henton C G 1989 Fact and fiction in the description of male and female pitch Language and Communication
13. Bishop J and Keating P 2012 Perception of pitch location within a speaker's range: Fundamental Frequency, voice quality and speaker sex The Journal of the Acoustical Society of America
14. Gaikwad S, Gawali B, and Mehrotra S C 2012 Gender identification using SVM with combination of MFCC, Advances in Computational Research.
15. Yue Ruan, Xiling Xue, Heng Liu, Jianing Tan, Xi Li, "Quantum Algorithm for K-Nearest Neighbors Classification Based on the Metric of Hamming Distance" DOI 10.1007/s10773-017-3514-4
16. Kamil Khadiev, Liliya Safina, "The Quantum Version of Random Forest Model for Binary Classification Problem", Vol-2842
17. Kai Li, Hongyi Dai, Feng Jing, Mei Gao, Bin Xue, Pu Wang and Ming Zhang, "Quantum Algorithms for Solving Linear Regression Equation" doi:10.1088/1742-6596/1738/1/012063
18. Patrick Rebentrost, Masoud Mohseni and Seth Lloyd, "Quantum support vector machine for big data classification" Jul 2014

BIOGRAPHIES

	Miss. Riddhi Mandal, pursuing B.Tech in Artificial Intelligence from G H Raisoni College of Engineering, Nagpur.
	Miss. Pragati Bhat, an undergraduate student specializing in artificial intelligence from G H Raisoni College of Engineering, Nagpur.