Chapter 5

Ensemble Learning

In this chapter, we propose and present the results obtained for ensemble learning models for speech recognition of continuous Gujarati sentences. To increase the accuracy and have a better generalisation for the said problem of speech recognition for sentences spoken in the Gujarati language, we propose two ensemble models: first, an ensemble of multilayered perceptron models, and later, an ensemble of hidden Markov models. Researchers across the world are utilizing ensemble learning models to enhance the performance of speech recognition [84] [85]. For the multilayered perceptron approach discussed in the section 4.4.2, we achieved 84.62% accuracy, and for the hidden Markov model, we achieved 84.38% accuracy. We aim to use ensemble learning methods to improve the accuracy further for the test sentence. To achieve this, we will first discuss the ensemble learning methods in the next section and then give a comparison of the results obtained for the two models of ensemble learning in the subsequent sections.

5.1 Ensemble Learning

Ensemble learning is a method of combining several trained models to achieve better classification accuracy as compared to the individual models. Some of the trained models can be weak in classification, while others can be strong in classification. The idea of ensemble learning is to combine such strong models with weak models [86]. The generalisation and overall performance of the combined models are reasonably better. There are many methods of ensemble learning, like bagging, boosting, stacking, etc.

We have used the bagging method for ensemble learning models. Bagging is also known as bootstrap aggregation. In this method of ensemble learning, various models are trained over different training subsets of the original dataset [87]. The test set is used on every model to determine the prediction accuracy of various models. The prediction for the

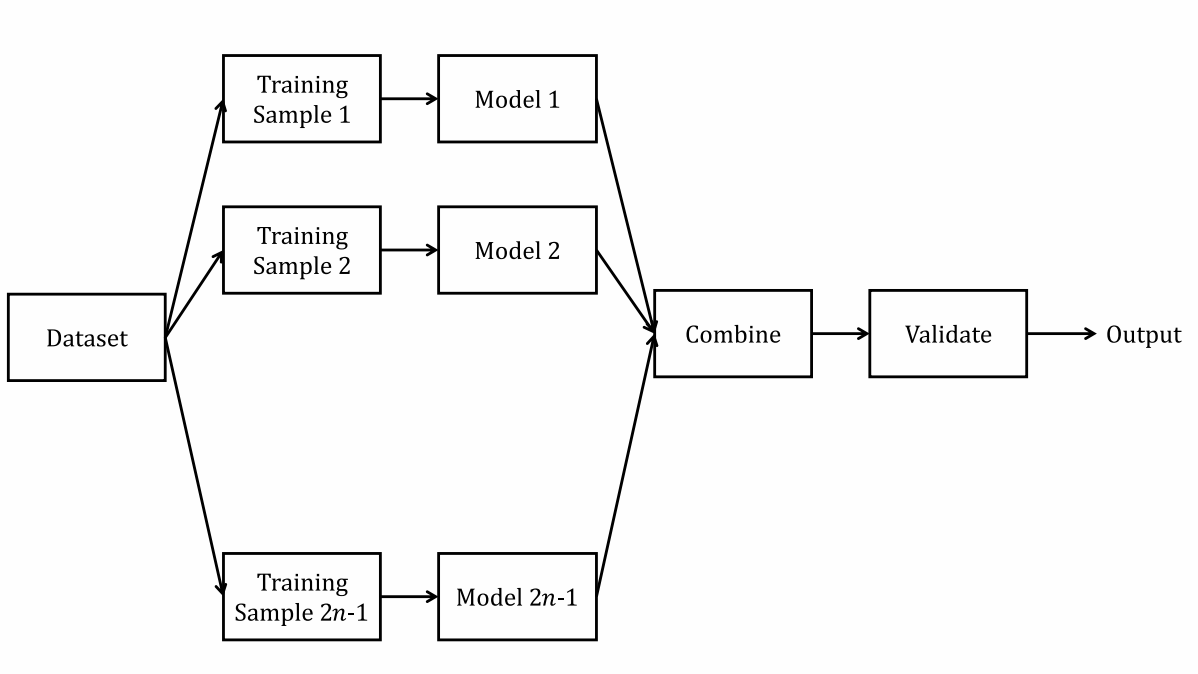

Figure 5.1: Block diagram to understand Bagging ensemble learning method

test sample is chosen by the max-vote technique in the case of classification. That is, a test sample is said to be predicted correctly if it is classified correctly by more than half of the models. Usually, an odd number of ensemble learning models are prepared.

The bagging method of ensemble learning can be easily understood by the block diagram shown in the Figure 5.1. In this figure, n is a positive integer. Hence, $2n-1$ is an odd positive integer. The output of the ensemble is obtained by using the max-vote technique for classification problems. That is, the maximum of all the class labels given as an output by each model in the ensemble is taken as an output of the ensemble. This method is useful in cases of over fitting because it reduces the variance. In the next section, we will discuss bagging ensemble learning of models based on multilayered perceptrons.

5.2 Approach 1: Ensemble Learning models for Multilayered Perceptron (ELMP)

In this approach, the dataset used for models, presented in section 4.4.3 is considered. That is, the one that consists of the recording of five sentences and 32 words by 10 speakers. These sentences are shown in the Figure 5.2. The recording, pre-processing, and feature extraction steps are as mentioned in the section 4.4.3.

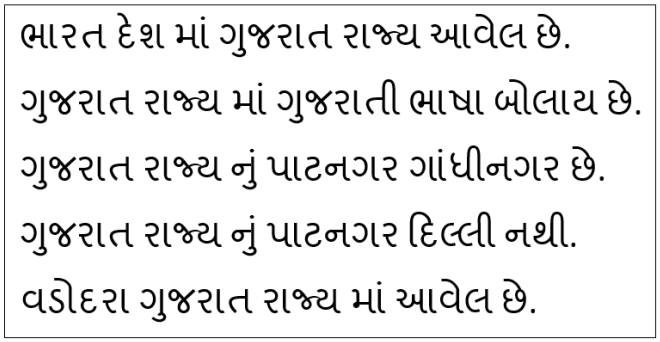

Figure 5.2: 5 sentences having 32 words used for the ensemble learning models

The multilayered perceptron, with architecture $N^2_{1027,100,1}$ is constructed. So, it has 1027 neurons in the input layer, 100 neurons in the hidden layer, and 1 neuron in the output layer. It is trained using the Adam algorithm and the ReLU activation function. Out of all the 320 feature vectors, 280 are considered for training and 40 are considered for testing.

A multilayered perceptron is trained using five different training subsets, and their test accuracy is noted. Then, two new sentences are recorded for the validation of these five models. These sentences are different from the trained sentences, but they have the same vocabulary as the trained words. Sentence 1 contained 8 words, and Sentence 2 contained 7 words, as shown in the Figure 5.3.

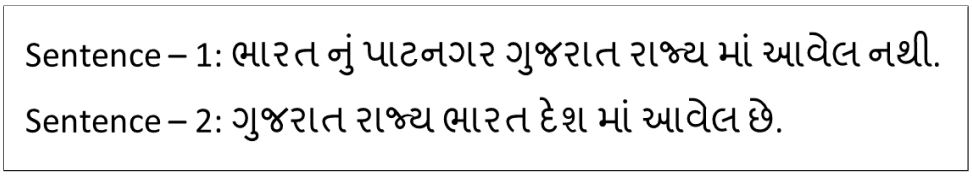

Figure 5.3: Sentences chosen for the validation of models

The performance of all five models is validated using these two sentences, and then the overall accuracy of the model for recognising these two sentences is determined. We define that a word is said to be recognised if it is recognised by three or more models out of five. So, for sentence 1, we achieved 100% recognition accuracy using the bagging ensemble. The summary of the results for sentence 1 is shown in the Figure 5.4.

Similarly, for sentence 2, we achieved 85.71% recognition accuracy using the bagging ensemble because, overall, 6 out of 7 words are recognised. It is summarised in the Figure 5.5. In the next section, we will discuss bagging ensemble learning models based on hidden Markov models.

Sentence 1	Model 1	Model 2	Model 3	Model 4	Model 5	Number of Models recognising word	Recognised correctly?
Word 1	Yes	Yes	Yes	No	No	3	Yes
Word 2	Yes	Yes	Yes	Yes	Yes	5	Yes
Word 3	Yes	Yes	Yes	Yes	Yes	5	Yes
Word 4	Yes	Yes	Yes	Yes	Yes	5	Yes
Word 5	Yes	Yes	Yes	Yes	Yes	5	Yes
Word 6	Yes	Yes	Yes	Yes	Yes	5	Yes
Word 7	Yes	Yes	Yes	Yes	Yes	5	Yes
Word 8	Yes	Yes	Yes	Yes	Yes	5	Yes

Figure 5.4: Validation of 5 ANN models on sentence-1

5.3 Approach 2: Ensemble Learning models for Hidden Markov Model (ELHMM)

In the second approach of ensemble learning, the bagging ensemble method is used, as discussed in the previous section. But this time, the base model selected is a hidden Markov model. In this approach, we consider the same dataset as that in section 5.2. The recording, pre-processing, and feature extraction steps remain similar. The two sentences selected for testing are also the same as shown in the Figure 5.3.

In this approach, a hidden Markov model is built with five states. Following the procedure of bagging ensemble, this model is trained with nine different training subsets of the dataset. In the training set, there are 280 feature vectors, and for testing, there are 40 feature vectors. Then, these nine trained models are validated using two test sentences.

For each test sentence, nine confusion matrices are prepared, representing the validation performance of different words in each sentence by different models. One such confusion matrix for validation of model-1 on sentence-1 is shown in the Figure 5.6. Similarly, Figure 5.7 represents the validation performance of model-1 on sentences 2. Then, the word of the sentence is defined as being recognised if more than half of the models are able to recognise that word.

As shown in the Figure 5.8, we conclude that 8 out of 9 words in sentence 1 are recognised using the bagging ensemble model. In this table, the letter "Y" (for yes) in the row i and

Sentence 2	Model 1	Model 2	Model 3	Model 4	Model 5	Number of Models recognising word	Recognised correctly?
Word 1	Yes	Yes	Yes	Yes	Yes	5	Yes
Word 2	No	Yes	No	Yes	No	2	No
Word 3	Yes	No	Yes	No	Yes	3	Yes
Word 4	Yes	No	Yes	Yes	No	3	Yes
Word 5	Yes	Yes	Yes	Yes	No	4	Yes
Word 6	Yes	Yes	Yes	Yes	Yes	5	Yes
Word 7	No	Yes	Yes	Yes	No	3	Yes

Figure 5.5: Validation of 5 ANN models on sentence-2

	Model 1								
Sentence 1	6/8	ભારત	નું	પાટનગર	ગુજરાત	રાજ્ય	માં	આવેલ	નથી
	ભારત	1	0	0	0	0	0	0	0
	નું	0	1	0	0	0	0	0	0
	પાટનગર	0	0	0	1	0	0	0	0
	ગુજરાત	0	0	0	1	0	0	0	0
	રાજ્ય	0	0	0	1	0	0	0	0
	માં	0	0	0	0	0	1	0	0
	આવેલ	0	0	0	0	0		1	0
	નથી	0	0	0	0	0	0	0	1

Figure 5.6: Confusion matrix of validation of sentence-1 on first hidden Markov model

column j represent that the model j is able to recognise the word i, otherwise letter "N" (for no) is used. is recognising Based on the overall validation performance of all models on sentence 1, we can conclude 87.50% accuracy, is achieved for the validation of sentence 1.

Similarly, as shown in the Figure 5.9, we conclude that 7 of 8 words in sentence 2 are recognised using the bagging ensemble model. Based on the overall validation performance of all models on sentence 2, we can conclude 85.71% accuracy is achieved for the validation of sentence 2.

		Model 1						
	6/7	ગુજરાત	રાજ્ય	ભારત	દેશ	માં	આવેલ	છે
Sentence 2	ગુજરાત	1	0	0	0	0	0	0
	રાજ્ય	0	0	0	0	0	1	0
	ભારત	0	0	1	0	0	0	0
	દેશ	0	0	0	1	0	0	0
	માં	0	0	0	0	1	0	0
	આવેલ	0	0	0	0	0	1	0
	છે	0	0	0	0	0	0	1

Figure 5.7: Confusion matrix of validation of sentence-2 on first hidden Markov model

		Models											
	Words	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Total**	**Word recognised?**	**Accuracy**
Sentence 1	*Word 1*	Y	Y	Y	Y	Y	Y	Y	Y	Y	9	Yes	87.50%
	Word 2	Y	Y	Y	Y	Y	Y	Y	Y	Y	9	Yes	
	Word 3	N	N	N	Y	Y	Y	N	N	N	3	No	
	Word 4	Y	Y	Y	Y	Y	Y	Y	N	Y	8	Yes	
	Word 5	N	N	Y	N	Y	Y	N	Y	Y	5	Yes	
	Word 6	Y	Y	Y	Y	Y	Y	Y	Y	Y	9	Yes	
	Word 7	Y	Y	Y	Y	N	Y	N	N	N	5	Yes	
	Word 8	Y	Y	Y	Y	Y	Y	Y	Y	N	8	Yes	

Figure 5.8: Validation of 9 HMM models on sentence-1

5.4 Conclusion

In this chapter, we reviewed ensemble learning and discussed the implementation using multilayered perceptrons and hidden Markov models for speech recognition of Gujarati sentences. The approach for ensemble learning used, is the bagging ensemble. For the multilayered perceptron approach, 5 models are trained using 5 different subsets of the dataset, and for the hidden Markov model approach, 9 models are trained using 9 different subsets of the dataset. The testing is done on two unknown sentences. The accuracy obtained is summarised in the Table 5.1. This accuracy is quite reasonable because the ensemble learning models perform better than the individual models discussed in sections 4.4.2 and 4.4.3. Moreover, we observe that multilayered perceptron-based ensemble learning models perform well with a smaller number of models as compared to hidden Markov models, which require a larger number of models. In the next chapter, we will

		Models											
	Words	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Total**	**Word recognised?**	**Accuracy**
Sentence 2	*Word 1*	Y	Y	Y	Y	Y	N	Y	Y	Y	8	Yes	
	Word 2	N	Y	Y	Y	N	N	N	N	N	3	No	
	Word 3	Y	Y	Y	Y	N	Y	Y	Y	Y	8	Yes	
	Word 4	Y	Y	Y	Y	Y	Y	Y	Y	Y	9	Yes	85.71%
	Word 5	Y	Y	Y	Y	Y	Y	Y	Y	Y	9	Yes	
	Word 6	Y	N	N	N	Y	Y	Y	Y	Y	6	Yes	
	Word 7	Y	Y	Y	Y	Y	Y	Y	Y	Y	9	Yes	

Figure 5.9: Validation of 9 HMM models on sentence-2

	Accuracy in percentage	
	ELMP	ELHMM
Sentence 1	100.00	87.50
Sentence 2	85.71	85.71

Table 5.1: Summary of models of ensemble learning

use trained models in the back-end and include the graphical user interface for speech recognition with recording and recognising options.