

CHAPTER 7

RESULTS AND DISCUSSION

Methods that effectively retrieves text documents from large databases, using two languages is an area of research that is still active, as the various challenges required by information retrieval systems have not yet been fully explored. With increasing research in the area of CLIR, new methods that retrieve documents automatically need to be developed. In order to meet the objectives formulated in Chapter 1, this research work proposes CLIR systems that are designed in three phases. Phase I of the study focuses on algorithms that perform speech recognition to convert Tamil speech query to text. Phase II involved the methods used for translating the Tamil Query to English Query. The actual retrieval of documents is performed in Phases III.

The performance of the algorithms proposed in each phase was tested vigorously and the results obtained are presented in this chapter. The results of the proposed algorithms are compared with the corresponding conventional / existing counterpart algorithms using appropriate performance metrics. The experiments were planned in two stages.

Stage 1 : The first stage of experiments was designed to evaluate the performance of the speech-based CLTR (TECLTR-S).

Stage 2 : The second stage of experiments was designed to evaluate the performance of the proposed text-based CLTR (TECLTR-T).

Each stage of the experiment is designed to evaluate the effect of the proposed enhanced algorithms on document retrieval performance. The experiments in each stage evaluates the pre-processing, feature extraction, feature selection and classification algorithms enhanced in each step of TECLTR-S and TECLTR-T and analyze the advantage obtained by the enhancement algorithms introduced. This chapter presents the results of the experiments conducted.

7.1. EXPERIMENTAL SETUP

This section presents the details regarding the test environment, and describes the databases used, along with the performance metrics that were used to evaluate the proposed algorithms.

7.1.1. Datasets Used

All the experiments were conducted on the news corpora provided by FIRE 2011 dataset (<http://www.isical.ac.in/~clia/2011>). FIRE follows the footsteps of TREC, CLEF and NTCIR with the aim of encouraging research in South Asian Language Information Access technologies by providing reusable large-scale test collections for Indian Language Information Retrieval (ILIR) experiments. It explores new Information Retrieval / Access tasks that arise as information needs evolve, and new needs emerge. They provide a common evaluation infrastructure for comparing the performance of different IR systems. They investigate evaluation methods for information access techniques and methods for constructing a reusable large-scale data set for CLIR experiments. FIRE contains lot of different multilingual data sets supporting Indian languages. This research is concerned with Tamil and English languages.

The corpora consist of English news articles taken from the magazine “The Telegraph”. The articles are grouped into different categories like Business, Foreign, Nation, Opinion, Sports, Job and Entertainment. The FIRE dataset 2011 consists of 50 queries in Tamil language divided into training and testing. The training set consists of 2250 sentences uttered by 15 persons with 3 repetitions, while the testing set consists of 1000 sentences uttered by 10 persons with 2 repetitions.

Document format used in FIRE collection follows the standard representation of TREC collection. Each document in the corpora, irrespective of the language, consists of a unique id mentioned within the <DOCNO> </DOCNO> tags, and the document content (actual news content) enclosed within the <TEXT> </TEXT> tags. An example of a text file is shown in Figure 7.1a.

Topics file contain some pre-fixed queries for the data set. These queries almost cover every document within the data set. The queries provided follow the

Cross-Language Evaluation Forum (CLEF) format. Each query topic has a unique number identifying the topic, a title, a description and a narrative. A title is typically of a few words in length and is characteristic of a real world IR query. The description of a topic contains more detailed description of what the user is looking for, as a natural language statement. A narrative contains a little more information than the description, in the sense that it also gives additional information of what is relevant and what is not relevant. A sample is shown in Figure 7.1b.

Qrels file format describes the presence and absence of every query terms in the document. Format of qrels is shown in Figure 7.1c, where the first place shows iteration, second column shows the query ID that is according to the topic file, third place the document ID that is mentioned in document format and the last column shows the presence and absence of that query in document by 0 or 1.

7.1.2. Performance Metrics

Performance evaluation of the enhancement methods developed is the most important step in any research. Different researchers use different parameters for analysis. All the algorithms were developed using Matlab 2009 and Java. This section presents the various performance metrics used during experiments.

- **Speech Recognition**

The evaluation of the noise removal and silence removal algorithms were evaluated using two performance metrics, namely, Signal to Noise Ratio, Mean Square Error and speed of pre-processing. The analysis of the feature extraction algorithm and the performance of speech recognition algorithm were conducted, using the recognition accuracy and speed of recognition performance metrics.

<pre> <DOC> <DOCNO>doc_03/0003</DOCNO> <TEXT> Americans used more health services and spent more on prescription drugs in 2013, reversing a recent trend, though greater use of cheaper generic drugs helped control spending, according to a report issued on Tuesday by a leading healthcare information company. </TEXT> </DOC> </pre>
(a) Document Format
<pre> <top lang="en"> <num>3</num> <title>World Cup Cricket 2007</title> <desc>Find documents reporting 2007 ICC Cricket World Cup in West Indies.</desc> <narr>The documents should contain information re- garding the ICC Cricket World Cup which took place in West Indies. Reports on the teams which qualified for the semi-finals, final and eventually the team which won the cup and India's performance as a team are of interest.</narr> </top> </pre>
(b) Sample Query/Topic in English
<pre> 1 Q0 doc_13/0013 1 1 Q0 doc_14/0014 0 1 Q0 doc_15/0015 0 1 Q0 doc_16/0016 0 1 Q0 doc_17/0017 0 1 Q0 doc_18/0018 0 1 Q0 doc_19/0019 1 1 Q0 doc_20/0020 0 </pre>
(c) Qrel File

Figure 7.1 : FIRE 2011

The first metric used during evaluation of pre-processing algorithms is Signal to Noise Ratio. SNR is used to quantify how much a signal has been corrupted by noise. It is defined as the ratio of signal power to the noise power corrupting the signal. The SNR is calculated in two ways. One is Pre SNR and the other is Post SNR which is obtained before and after applying the pre-processing operation. Denoising is successful if Post SNR is higher than Pre SNR. Equation (7.1) presents the formula used to estimate SNR.

$$\text{SNR}_{\text{db}} = 10 \log_{10} \left(\frac{P_{\text{signal,db}}}{P_{\text{noise,db}}} \right) = P_{\text{signal, db}} - P_{\text{noise, db}} \quad (7.1)$$

Mean Square Error is used to quantify the difference between values implied and the true being estimated. The input MSE is defined using Equation (7.2) and output MSE is defined using Equation (7.3), where N is the length of the signal, x_i is the original signal, y_i is the noisy signal and \bar{x}_i is estimated x_i (noisy signal y passed through denoising algorithm (Figure 7.2)). Lower MSE indicates a closer match between the two signals.

$$\text{Input MSE} = \frac{1}{N} \sum_i (X_i - Y_i)^2 \quad (7.2)$$

$$\text{Output MSE} = \frac{1}{N} \sum_i (X_i - \bar{X}_i)^2 \quad (7.3)$$

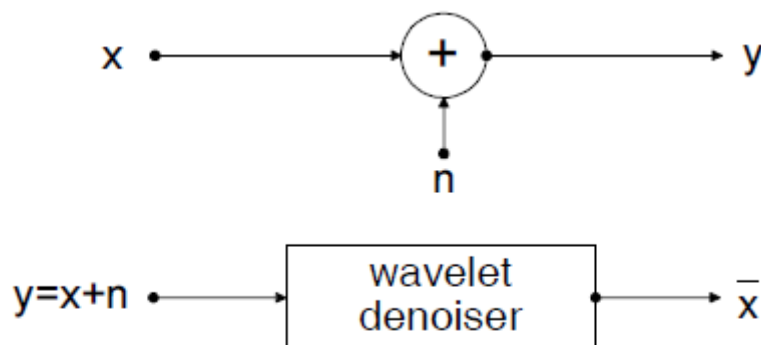


Figure 7.2 : Signal Used for Calculating MSE

The next performance metric, accuracy is estimated using Equation (7.4). A high accuracy value indicates maximized speech recognition performance.

$$\text{Accuracy}(\%) = \frac{\text{No. of Words Correctly Reconnize d}}{\text{Total No. of Words}} \times 100 \quad (7.4)$$

The final performance metric is the speed measured in seconds. This indicates the time taken by the algorithm to perform pre-processing or recognition tasks.

- **Query Translation**

The query translation tasks were evaluated using metrics like performance metrics, namely, Precision (Equation 7.5), Recall (Equation 7.6), F-Measure (Equation 7.7), Accuracy (Equation 7.8), Error rate (Equation 7.9) and speed of retrieval. Each of these metrics is explained below.

$$\text{Precision}(\%) = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (7.5)$$

$$\text{Recall}(\%) = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (7.6)$$

$$\text{F}(\%) = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7.7)$$

$$\text{Accuracy}(\%) = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (7.8)$$

$$\text{Error Rate}(\%) = \frac{\text{FP} + \text{FN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (7.9)$$

In above equations TP, TN, FP, FN represents true positive, true negative, false positive and false negative respectively. Speed is the time taken by the CLTR systems to successfully complete the query translation task. The speed is measured in seconds.

- **Document Retrieval**

In order to evaluate results, it is possible to use several measures related to information retrieval. This research uses the most commonly used metrics to evaluate the algorithms presented in Phase III of the research work. They are F1 Measure,

Macro F1 Measure, Accuracy, Mean Average Precision, R-precision and speed. The F-measure and accuracy performance metrics were calculated using equations described in the previous section (Equations 7.7 & 7.8) respectively.

For systems that return a ranked sequence of documents, it is preferable to consider the order in which the returned documents are presented. This measure averages the precision values from the rank positions where a relevant document is retrieved (Equation 7.10).

$$\text{Average Precision (AP)} = \frac{\sum_{r=1}^N (P(r) \times \text{rel}(r))}{C_r} \quad (7.10)$$

where r is the rank, N is the number of documents retrieved, $\text{rel}(r)$ is a function on the relevance of a given rank and $P(r)$ is the precision at a given cut-off rank (top 20 documents in this work).

Using this measure, the Mean Average Precision can be estimated as in Equation (7.11) and is used to summarize rankings by averaging average precision.

$$\text{MAP} = \frac{\sum_{q=1}^Q \text{AP}(q)}{|Q|} \quad (7.11)$$

The next metric, R-precision, is the precision after R documents have been retrieved, where R is the number of documents for the topic.

The macro F1 Measure is the average of F1 Measure of all topics (Equation 7.11).

$$\text{Macro } F_1 = \text{Average}(F_1) \quad (7.12)$$

Finally, the speed (seconds) measures the time taken for the algorithm to retrieve documents relevant to the query.

7.2. SPEECH RECOGNITION

This section presents experimental results that analyze the noise removal, silence removal, feature extraction and speech recognition algorithms proposed in Phase I of the research work. The coding scheme presented in Table 7.1 is used during discussion.

TABLE 7.1
CODING SCHEME USED IN PHASE I

Code	Description
ATSQR	Automatic Tamil Speech Query Recognition System
Pre-processing	
WHT	Wavelet Denoising based on Hard Thresholding
WST	Wavelet Denoising based on Soft Thresholding
W2SHT	Wavelet Denoising based on Switching Soft and Hard Thresholding
SRDSTQ	Silence Removal from Denoised Speech Tamil Query
Feature Extraction	
MFCC	Mel-Frequency Cepstral Co-efficient
WPMFCC	Wavelet Packet Based MFCC
LPCC	Linear Prediction Cepstral Co-efficient
WPLPCC	Wavelet Packet Based LPCC
Tamil Speech Recognition	
S	Conventional SVM Classifier
SS	SVM classifier enhanced with SOM
SES	Ensemble SVM Classifier enhanced with SOM

7.2.1. Pre-processing

The pre-processing step of ATSQR performs noise removal and silence removal. Three types of noises were used during evaluation. They are white noise, babble noise and external noise. Figures 7.3 to 7.5 show the average SNR analysis of the proposed W2SHT denoising and SRDSTQ silence removal algorithms. The results are compared with the conventional WHT and WST based denoising and silence removal algorithms.

While considering white noise removal, on an average, the W2SHT algorithm showed an efficiency gain of 32.06% and 15.08% over WHT and WST algorithms respectively. Similarly, the SRDSTQ algorithm showed 45.09% and 31.36% over the conventional algorithms respectively, after removing white noise.

The removal of babble noise, using W2SHT and SRDSTQ algorithms, showed a positive increase in efficiency by 19.71% (WHT), 10.76% (SHT) and 27.33% (WHT), 19.22% (WST) respectively. The efficiency gain obtained by incorporating the silence removal algorithm in W2SHT algorithm on an average was 4.77%.

Removal of external noise using W2SHT algorithm, on average, showed an efficiency gain of 11.47%, 5.94% over WHT and WST algorithms respectively. The same was 16.00% and 10.43% while using SRDSTQ algorithm.

The results thus prove that with respect to SNR analysis results, both the noise removal and silence removal algorithms are successful in improving the quality of the Tamil speech query.

Figures 7.6 to 7.8 present the performance of the proposed and conventional pre-processing algorithms in terms of Mean Square Error. In these figures, IM denotes Input MSE and OM represents output MSE.

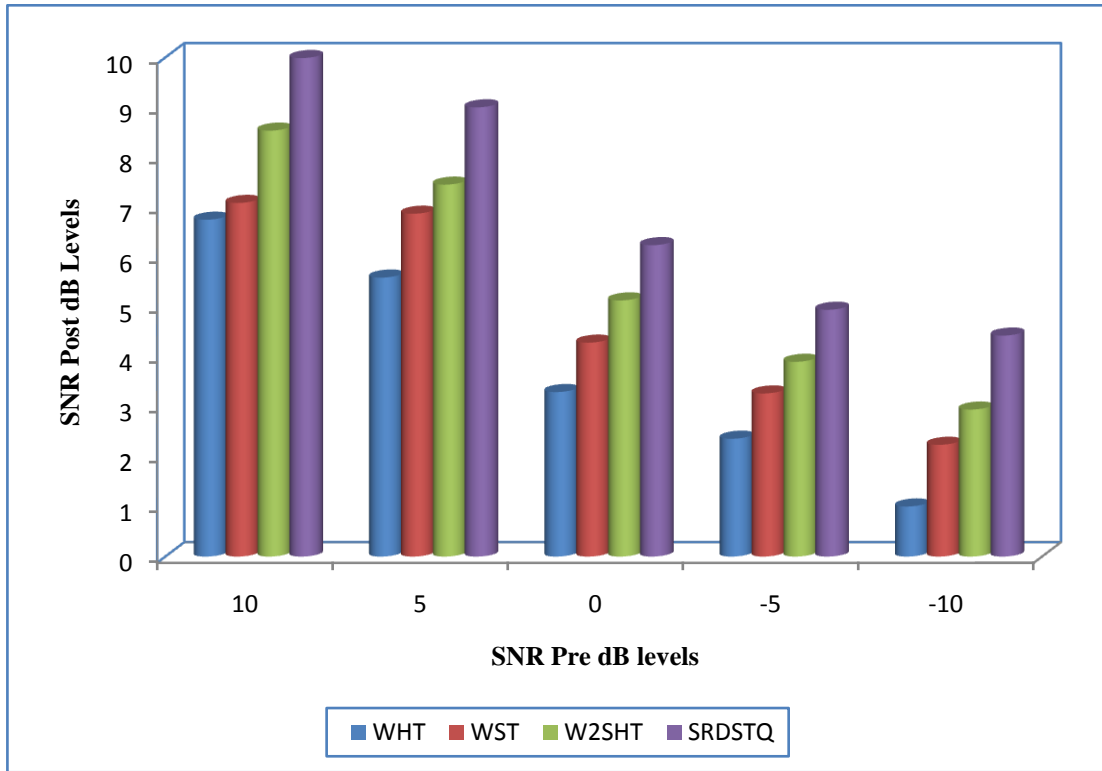


Figure 7.3 : Average SNR Analysis of White Noise Removal

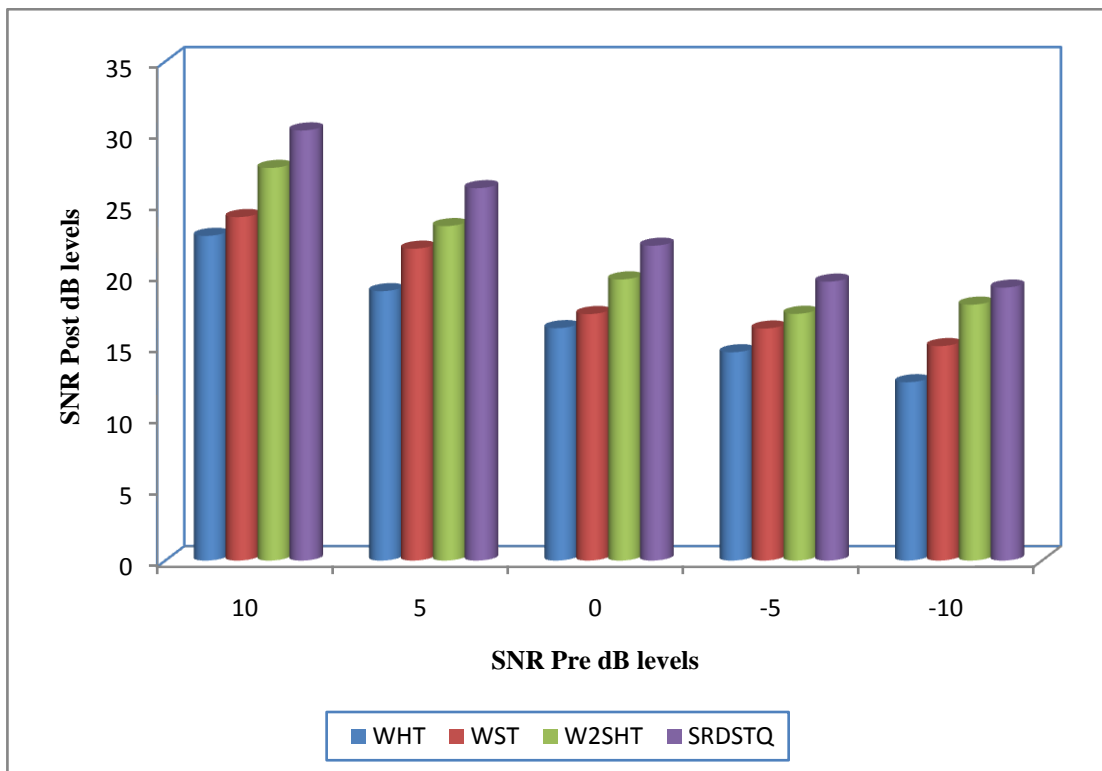


Figure 7.4 : Average SNR Analysis of Babble Noise Removal

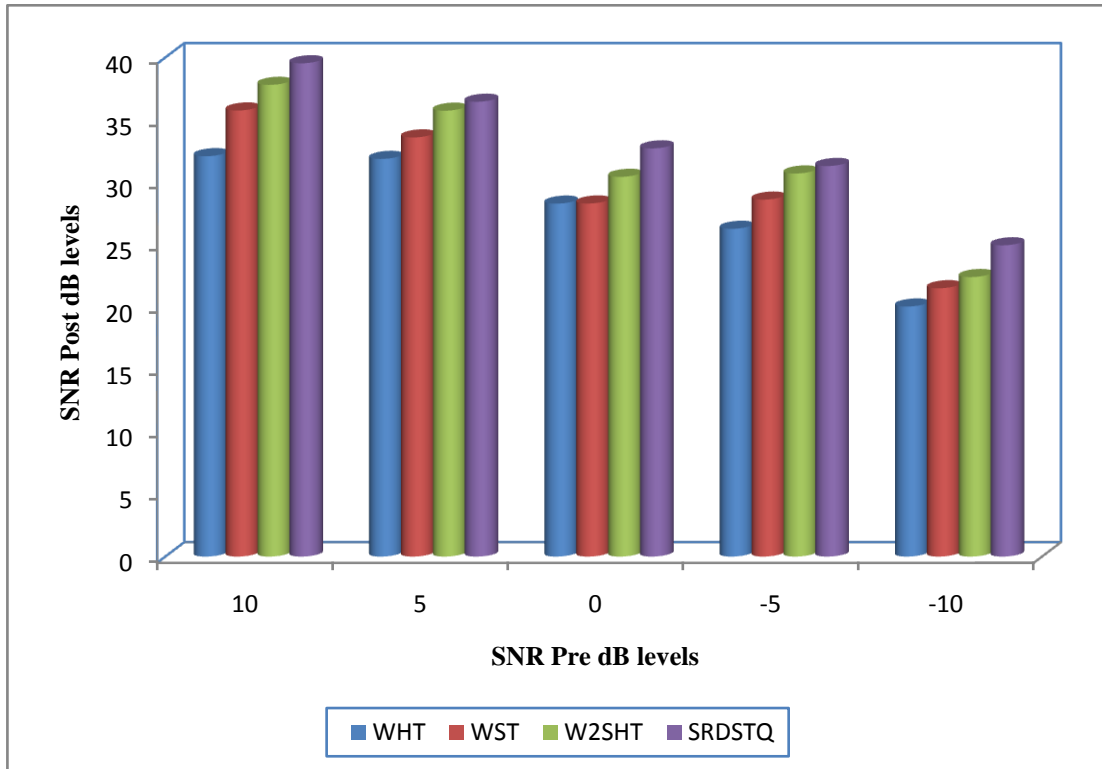


Figure 7.5 : Average SNR Analysis of External Noise Removal

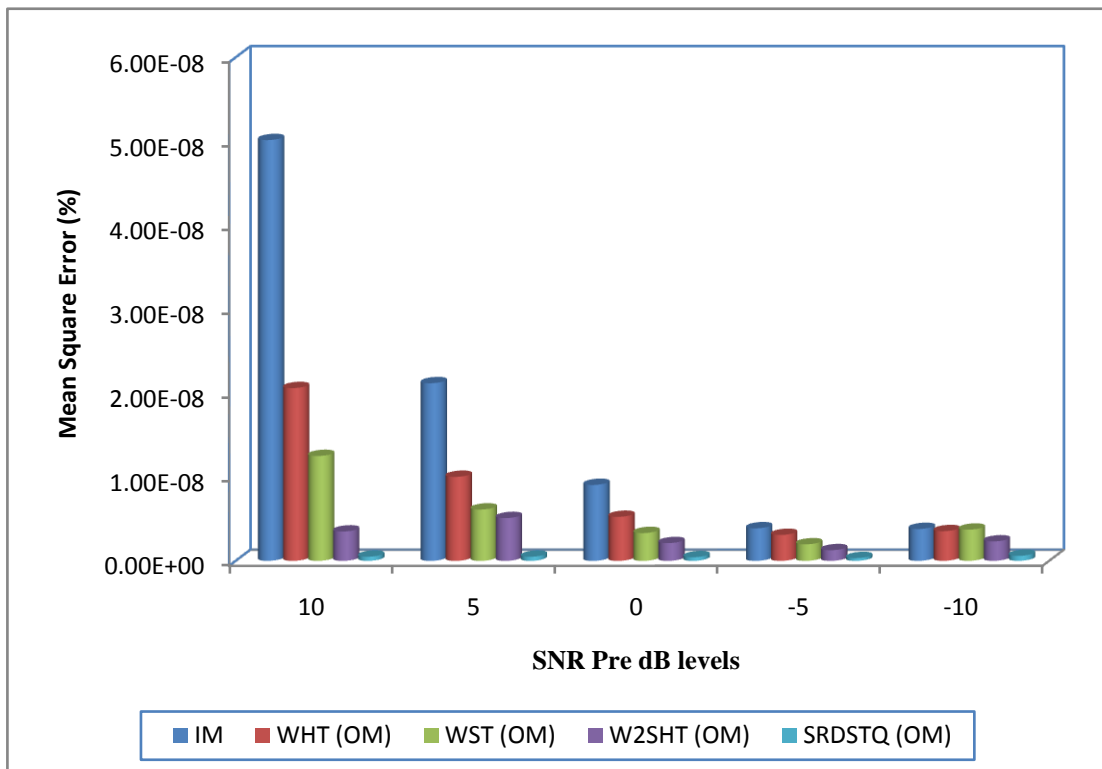


Figure 7.6 : Average MSE Analysis of White Noise Removal

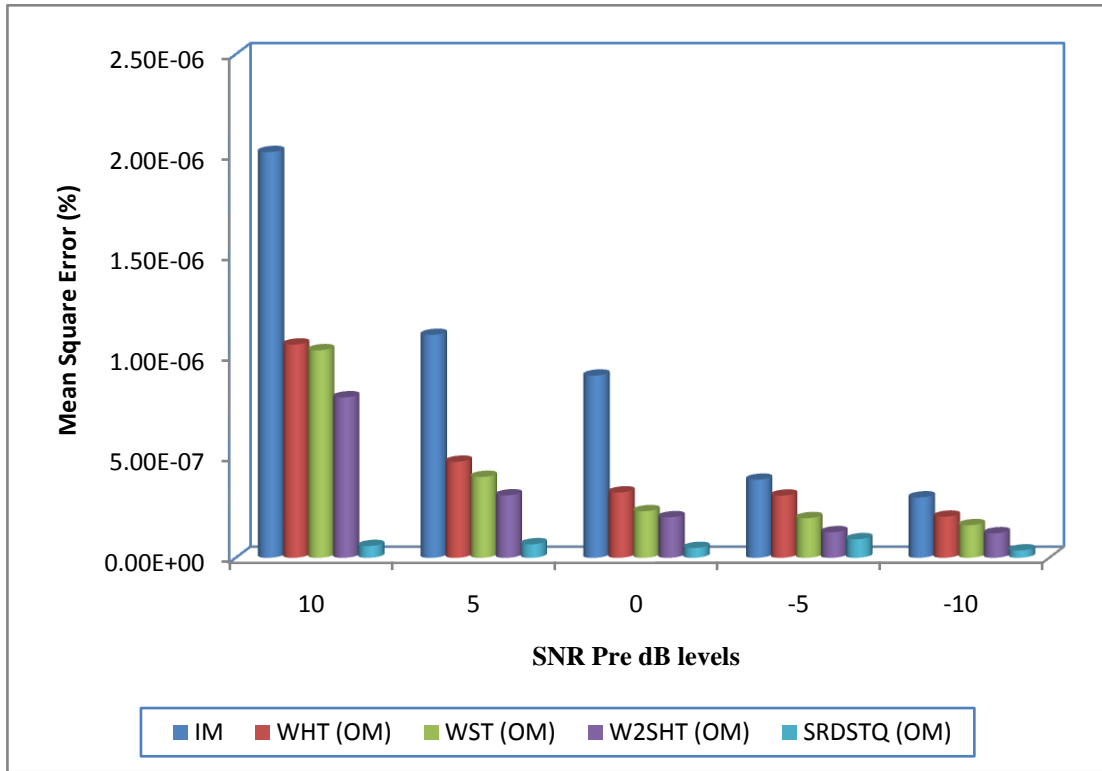


Figure 7.7 : Average MSE Analysis of Babble Noise Removal

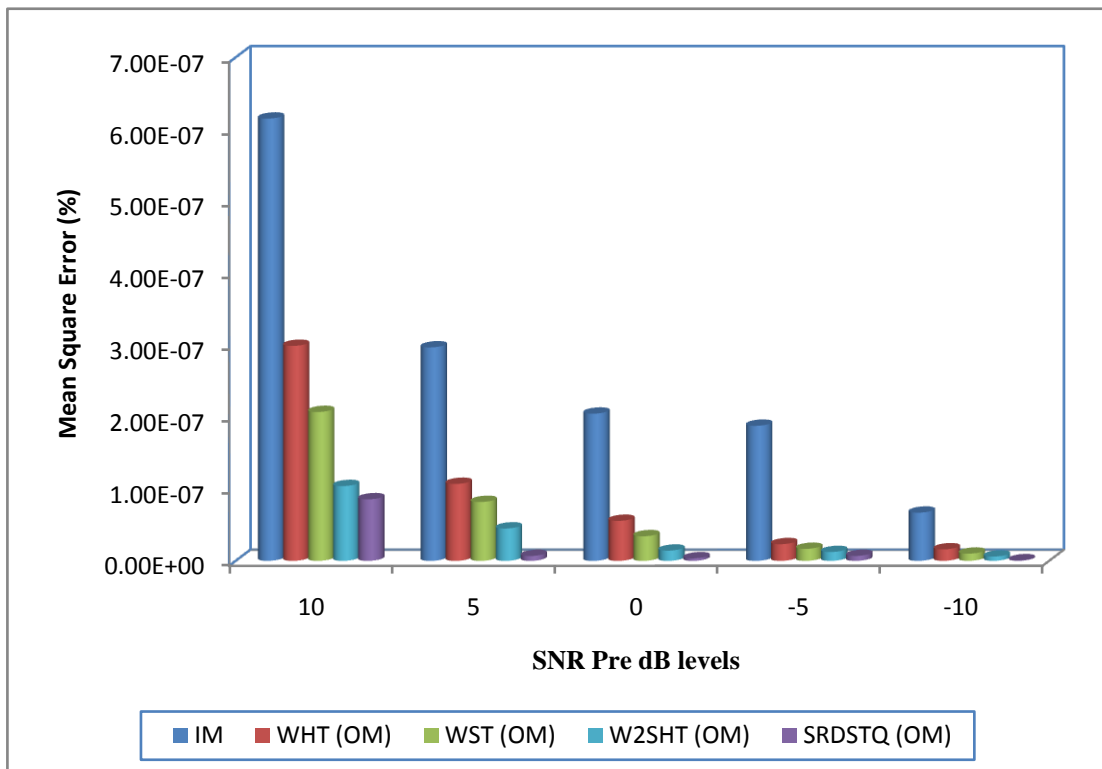


Figure 7.8 : Average MSE Analysis of External Noise Removal

Analysis of white noise removal results revealed that the average input MSE was $1.76083\text{E-}08$ for WHT and WST algorithms produced an average output MSE of $8.49584\text{E-}09$ and $5.52364\text{E-}09$, which decreased to $2.86\text{E-}09$ and $4.69\text{E-}10$ with W2SHT and SRDSTQ output MSE. Similarly, babble noise removal produced on average $9.42828\text{E-}07$ input MSE, which decreased to $4.74704\text{E-}07$ and $4.04564\text{E-}07$ with conventional WHT output MSE and WST MSE output MSE. The average results obtained by the proposed W2SHT and SRDSTQ algorithms were $3.11452\text{E-}07$ and $6.076\text{E-}08$ output MSEs. Removal of external noise showed an average input MSE of $2.74228\text{E-}07$ and the WHT and WST algorithms showed $1.0014\text{E-}07$ and $6.9952\text{E-}08$ average output MSEs. This result was improved while using the proposed algorithms, which showed an average output MSE of $3.63\text{E-}08$ and $2.07218\text{E-}08$ respectively.

According to the criterion of MSE, the pre-processing operation is successful if output MSE is lower than input MSE. Adhering to this, the output MSE results pertaining to all the three selected noises and silence removal is less than the input MSE, indicating that the proposed algorithms are improved version of the conventional algorithms.

Figure 7.9 shows the average speed measured in seconds that the conventional and proposed algorithm, taken to finish the pre-processing tasks. The performance with respect to speed shows that the SRDSTQ algorithm is slightly slow when compared to the conventional algorithms. This is due to the fact that the SRDSTQ algorithm performs two operations, namely, noise removal and silence removal. Analysis shows that the difference, when compared to W2SHT, is only 0.5 seconds, and hence, it can be concluded that the proposed SRDSTQ algorithm is enhanced with respect to time efficiency also.

Thus, the results show that the proposed SRDSTQ algorithm has improved the quality of the input speech signal, and can be used safely to improve the process of feature extraction and speech recognition.

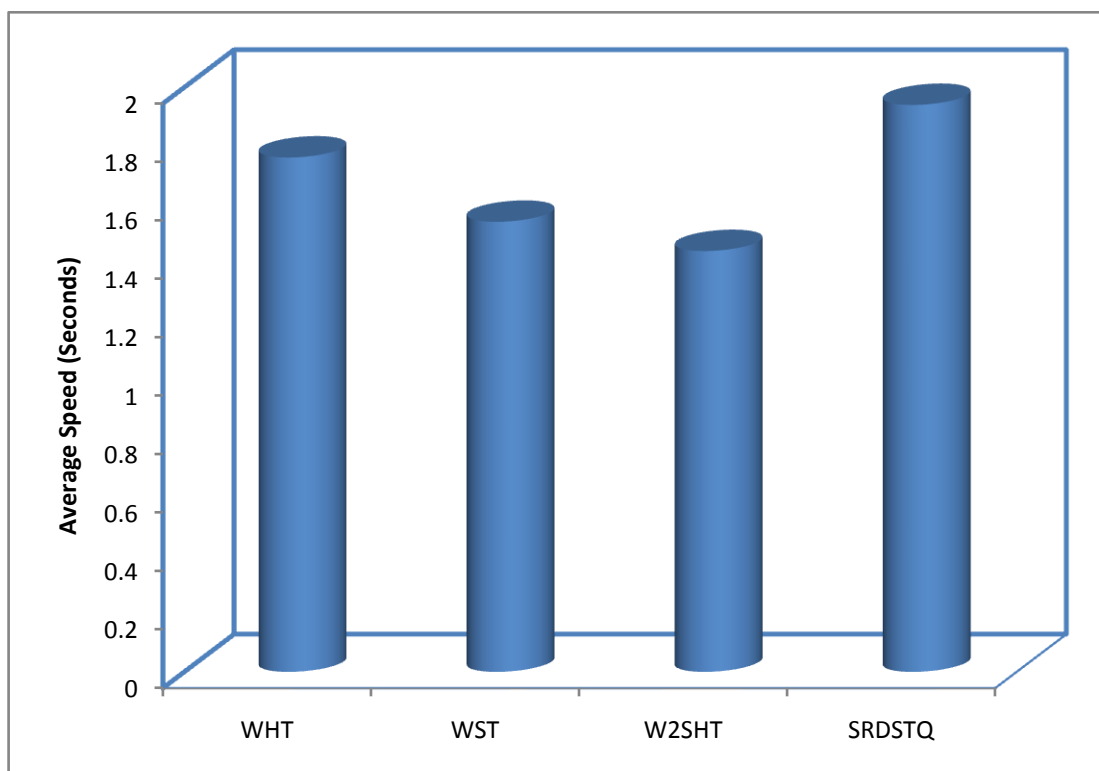


Figure 7.9: Average Speed (Seconds) of Pre-processing Algorithms

7.2.2. Feature Extraction and Recognition

This section presents the effect of the conventional and proposed feature extraction techniques on recognition accuracy. The recognition accuracy is analyzed for speech recognizers that use the conventional SVM classifier, SVM with SOM classifier and proposed ensemble SVM with SOM classifier. Table 7.2 shows the recognition accuracy at sentence level, while using the sample of 20 Tamil title queries from FIRE 2011 dataset.

The overall efficiency of the feature extraction and recognition classifiers were analyzed by average accuracy while testing with all 50 queries from FIRE dataset. The results obtained are shown in Figure 7.10. This figure also analyzes the effect of pre-processing on recognition. In this Figure, WPP (With Pre-Processing) and WOPP (Without Pre-Processing) indicate the inclusion and exclusion of noise and silence removal algorithms respectively. .

TABLE 7.2**SENTENCE LEVEL ACCURACY (%) OF ATSQR SYSTEM WHILE USING THE TOP 20
TAMIL QUERIES IN FIRE DATASET**

	QUERIES	LPCC			MFCC			WPLPCC			WPMFCC		
		S	SS	SES	S	SS	SES	S	SS	SES	S	SS	SES
1	பன்றிக்காய்ச்சல் தடுப்பு மருந்து	80.20	83.86	85.41	83.35	87.28	87.72	87.55	90.29	90.85	90.00	93.33	95.56
2	அரிய கதிர்வீச்சு நிகழ்வுகள்	75.41	80.94	83.43	79.19	86.92	86.74	80.14	86.74	89.05	86.79	89.01	90.14
3	கோத்ரா ரயில் தாக்குதல்	77.59	83.68	86.85	78.96	89.33	90.45	83.08	90.78	90.49	88.91	90.75	93.65
4	மைக்கேல்ஜாக்சனின் அகால மரணம்	78.25	80.89	80.44	80.12	83.43	85.24	82.91	88.09	87.84	85.49	89.06	90.79
5	பெட்ரோலிய பொருட்களின் விலை உயர்வு	78.37	78.79	80.96	80.81	85.42	87.54	80.65	87.18	87.80	82.92	88.49	92.90
6	பத்திரிகையாளர்கள் கொலை மற்றும் ஆள் கடத்தல்	76.88	80.82	83.20	85.79	87.07	90.85	87.23	87.61	91.81	90.90	90.82	93.19
7	பராக்ஓபாமாவின் வெற்றி	75.05	83.31	83.42	83.51	85.84	87.97	78.47	85.77	85.37	82.57	92.72	92.85
8	அபூகாரேப் சிறைச்சாலையில் அத்துமீறல்கள்	73.34	77.54	80.91	78.91	86.10	87.54	80.23	90.39	85.98	83.73	93.12	88.34
9	சிமி மீது தடை விதிக்கப்பட்டது	78.57	82.01	82.04	80.13	88.81	88.56	80.36	88.75	90.33	84.81	90.60	92.06
10	இந்தியாவின் வேளாண்மை சார்ந்த மத்திய பட்ஜெட்	78.67	82.91	81.76	84.97	87.13	87.72	83.42	87.99	88.91	87.42	88.99	92.57

	QUERIES	LPCC			MFCC			WPLPCC			WPMFCC		
		S	SS	SES	S	SS	SES	S	SS	SES	S	SS	SES
11	பொழுதுபோக்கு உலகின் போலிகள்	78.90	81.87	82.10	83.33	82.51	87.18	78.82	82.99	82.76	85.70	87.94	92.65
12	எல்.டி.டி.ஈ தலைவரின் இறப்பு	77.03	83.11	83.71	83.12	85.62	87.26	81.85	87.50	86.77	86.94	87.95	93.54
13	இந்தியாவில் பெண்களுக்கான இடஒதுக்கீடு மசோதா	75.45	80.98	83.40	80.37	86.36	89.29	84.29	89.05	88.25	85.12	90.28	93.37
14	சோமாலி போலிகள் தோற்கடிப்பு	72.89	79.59	80.38	86.89	80.30	83.60	85.06	87.46	87.22	89.20	89.81	93.12
15	விண்வெளியில் உயிரினம் மற்றும் தண்ணீர் தேடுதல்	78.08	80.30	82.74	83.60	85.35	88.90	84.83	87.41	87.45	87.42	88.45	90.30
16	குளோனிங் முறையில் மனித குழந்தைகள் பிறப்பு	78.39	79.58	79.39	83.35	88.74	86.53	84.30	89.16	88.71	87.68	89.54	94.02
17	சட்டவிரோதமாக வெட்டப்பட்ட மரங்கள்	77.56	83.31	82.06	80.35	90.20	88.60	82.98	88.29	90.68	89.60	90.14	92.32
18	டாடாவின் சிறிய கார்	75.30	79.04	80.66	78.47	85.32	83.84	83.36	85.65	85.78	85.25	89.14	94.46
19	போபால் விஷ வாயு சம்பவம்	80.81	82.54	82.99	86.48	87.37	84.25	88.57	89.11	85.16	89.02	90.02	90.52
20	பெனாசீர் பூட்டோ படுகொலை	77.47	78.21	80.71	83.80	83.81	85.98	85.69	87.75	87.62	88.96	90.01	90.35

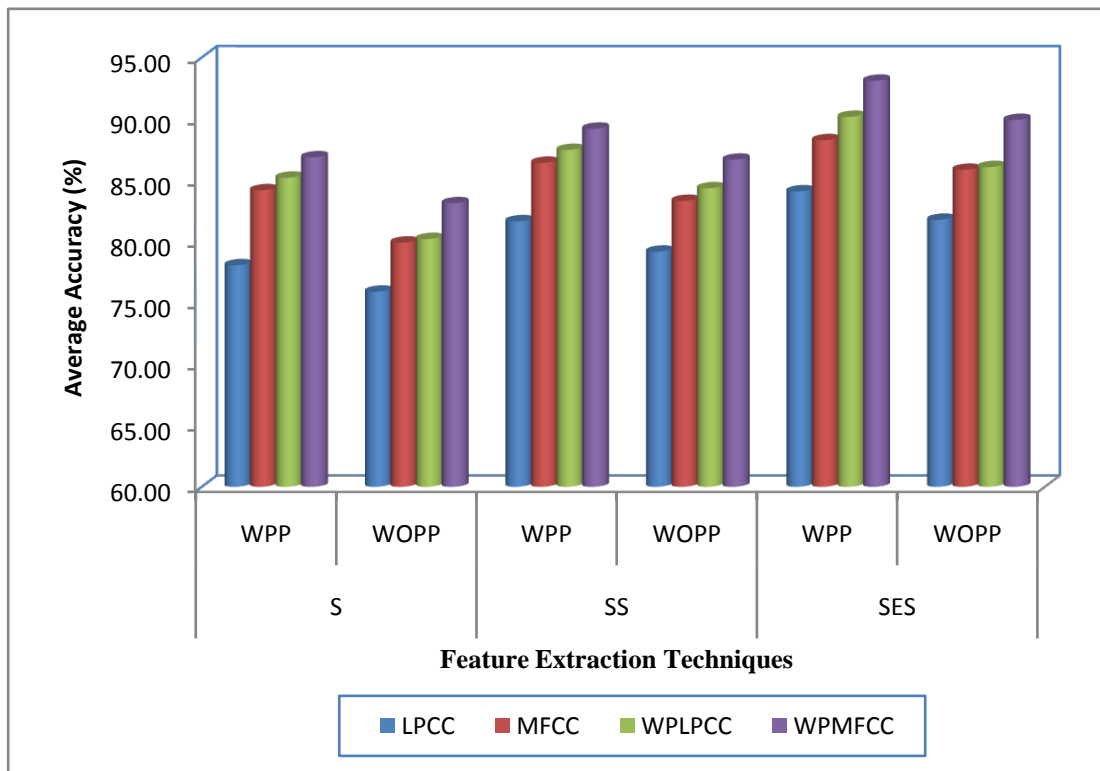


Figure 7.10: Effect of Pre-processing and Feature Extraction Techniques with different classifiers on ATSQR with respect to Average Accuracy (%)

From Table 7.2 and Figure 7.10, it is evident that both the enhanced feature extraction techniques, WPLPCC and WPMFCC, perform better than their conventional counterparts, LPCC and MFCC. Comparison between WPLPCC and WPMFCC shows that the WPMFCC technique produces higher accuracy than WPLPCC. It is also evident that the inclusion of pre-processing algorithm has increased the efficiency of speech recognition. The proposed SES classifier shows an efficiency gain of 3.39% while using pre-processing. On average, the efficiency gain obtained by SES algorithm after application of pre-processing over the convention single SVM classifier and SVM enhanced with SOM classifier is 5.99% and 3.06% respectively.

Figure 7.11 shows the performance of the feature extraction and recognizers with respect to execution time.

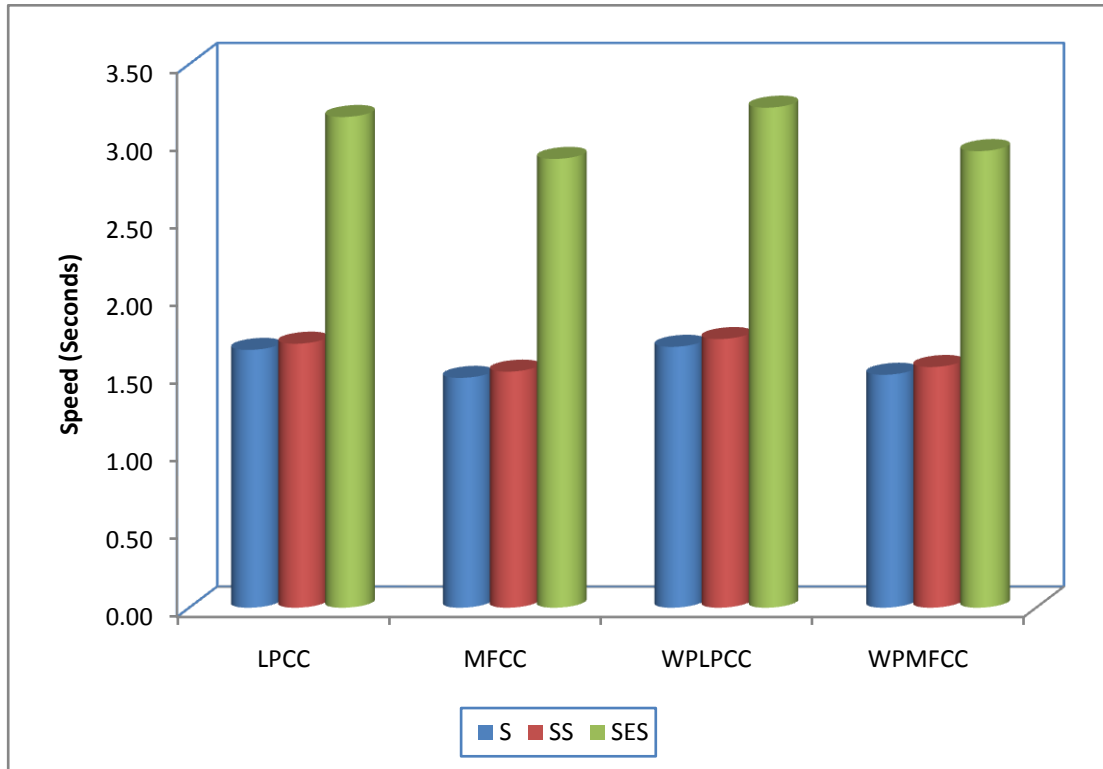


Figure 7.11 : Speed (Seconds) of Feature Extraction and Recognition

Results pertaining to speed shows that the MFCC-based feature extraction algorithms are faster than LPCC-based algorithms. The speed complexity of the proposed recognition system is high because of the use of ensemble SVM classifier. However, as the accuracy of the speech recognized is more important to increase the performance of the TECLTR-S system, this speed complexity is ignored in this research.

Thus, it can be concluded that the proposed ATSQR system that combines the Wavelet Denoising based on Switching Soft and Hard Thresholding algorithm (W2SHT), Silence Removal from Denoised Speech Tamil Query (SRDSTQ), Wavelet Packet Based Mel-Frequency Cepstral Co-efficient features (WPMFCC) and Ensemble SVM Classifier enhanced with SOM (SES) produces improved speech recognition results, and therefore, can be used in TECLTR-T for efficient retrieval of relevant documents.

7.3. QUERY TRANSLATION

Among the various steps in Phase II, this research work proposes algorithms for POS tagging, Morphological Analyzer, Word Sense Disambiguation and Translation. This section presents the experimental results of these algorithms. Table 7.3 presents the coding scheme used in this section.

TABLE 7.3
CODING SCHEME USED IN PHASE II

Code	Description
TQTS	Tamil Query Translation System
POSSA	POS Tagging using SVM classifier with Simulated Annealing and AMOSA
POSHES	POS Tagging using Hybrid SVM-WNN Ensemble Classifier with Ensemble Feature Selection
SMA	SVM-based Morphological Analyzer
ESMA	Ensemble SVM-based Morphological Analyzer
CWSD	Conventional Word Sense Disambiguation Algorithm
WSDPCS	Word Sense Disambiguation Algorithm enhanced with POS and Clustering based Sense-collocation
WWQT	Word by Word Query Translation
HQT	Hybrid Query Translation (Combining Rule Based and Knowledge Based Systems)
HQTP	Hybrid Query Translation with POS tagging
HQTM	Hybrid Query Translation with Morphological analysis
HQTW	Hybrid Query Translation with Word sense disambiguation
HQTPMW-S	Hybrid Query Translation With POS, MA, WSD-Speech query
HQTPMW-T	Hybrid Query Translation With POS, MA, WSD-Text query

7.3.1. POS Tagging

The first step during query translation is POS tagging. The performance evaluation of the proposed POSSA and POSHES algorithms were evaluated using five performance metrics, namely, precision, recall, F-measure, accuracy and speed. The experiments aim to evaluate the proposed feature selection algorithm, and the classification algorithm used to enhance the process of POS tagging. As mentioned in the previous chapter, the algorithm proposes a fusion feature vector that combines features selected by three different feature selection algorithms, namely, F-Score, Decision Tree (DT) and Linear Discriminant Analysis (LDA). The fused feature vector is termed as EFS. The SVM classifier is enhanced to combine the merits of WNN and an ensemble SVM-WNN classifier is proposed. The results compare the proposed algorithm with the existing algorithm.

Table 7.4 presents the precision, recall and F-measure of the existing (POSSA) and proposed (POSHES) algorithms while using the four different feature selection algorithms.

TABLE 7.4

PRECISION, RECALL AND F-MEASURE OF POSSA AND POSHES ALGORITHMS

Feature Selection Algorithm	Precision (%)		Recall (%)		F-Measure (%)	
	POSSA	POSHES	POSSA	POSHES	POSSA	POSHES
No Feature Selection	81.90	91.21	63.61	71.64	71.61	80.25
F-Score	83.26	92.15	64.10	74.94	72.43	82.66
LDA	83.91	92.77	65.76	75.43	73.73	83.21
DT	84.55	93.02	68.29	78.42	75.56	85.10
EFS	91.06	95.71	74.24	86.69	81.79	90.98

From the above table results, it is evident that the proposed POSHES algorithm using fused feature vector (EFS) and ensemble SVM-WNN algorithm produces better tagging result. The F-Measure metric, combining precision and recall,

is used as a measure for overall performance, and the POSHES algorithm showed an efficiency gain of 10.09% when compared with POSSA algorithm. The results also show that the inclusion of feature selection algorithm increases the efficiency of the proposed tagging system. This is evident from the vast performance difference in precision, recall and F-measure when no feature selection algorithm is used. The EFS algorithm produces maximum benefits to accurate tagging, which is efficient from the efficiency gain of 9.14%, 8.54% and 6.46% obtained over F-Score, LDA and DT feature selection algorithm. This shows that the EFS algorithm that fuses the results from the three feature selection algorithm, works better than single feature selection algorithm.

Figure 7.12 shows the accuracy (%) of the existing and proposed POS tagging algorithms. Figure 7.13 presents the speed of tagging while using POSHES and POSSA with the selected feature selection algorithms.

The accuracy results again prove that the proposed POSHES (97.84%) produces more accurate tagging results than POSSA (89.46%) while using EFS feature selection algorithm. While comparing the algorithms in terms of speed, it can be seen that the proposed POSHES algorithm is slow when compared to POSHES. This is due to the fact that an ensemble classifier has more computational demands and implementation complexities than single classifier system (https://en.wikipedia.org/wiki/Ensembles_of_classifiers#Weaknesses). Despite this weakness, the usage of ensemble classifiers results in more accurate results. The results are more stable and outperform single classifier system in all cases (Utami *et al.*, 2014; Guo *et al.*, 2015). Due to these advantages, it can be concluded that the proposed POSHES produces maximum improvement in the process of POS tagging.

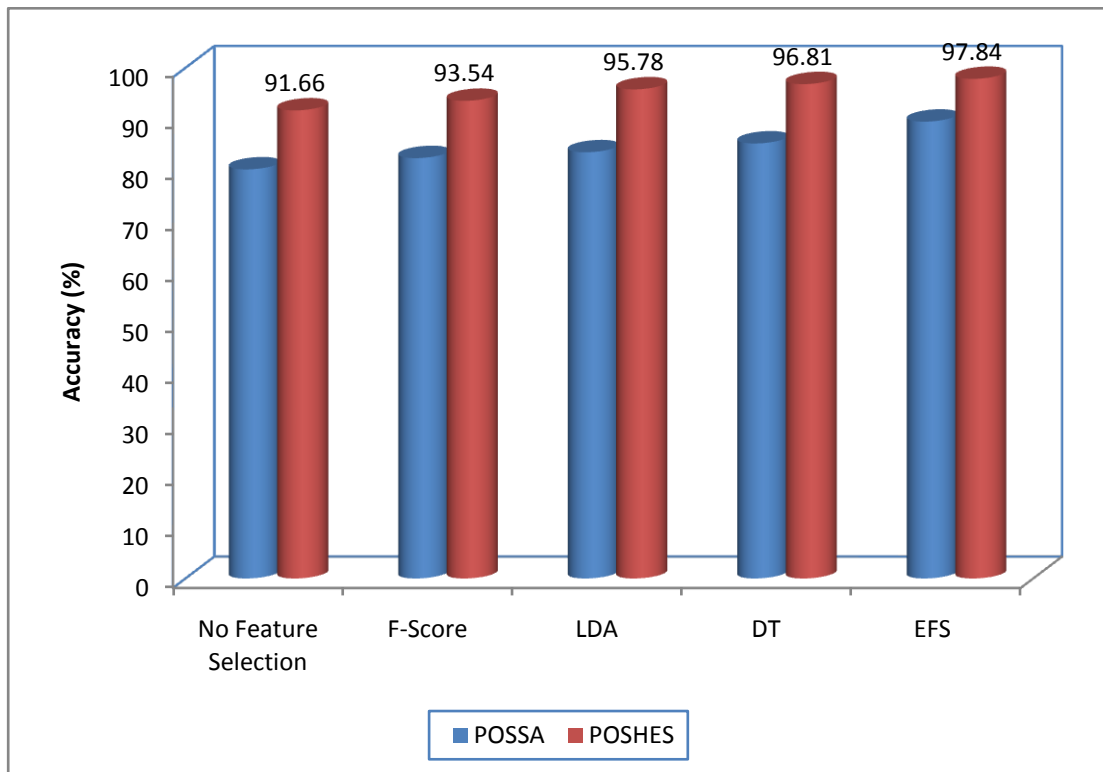


Figure 7.12 : Accuracy of POS Tagging Algorithms

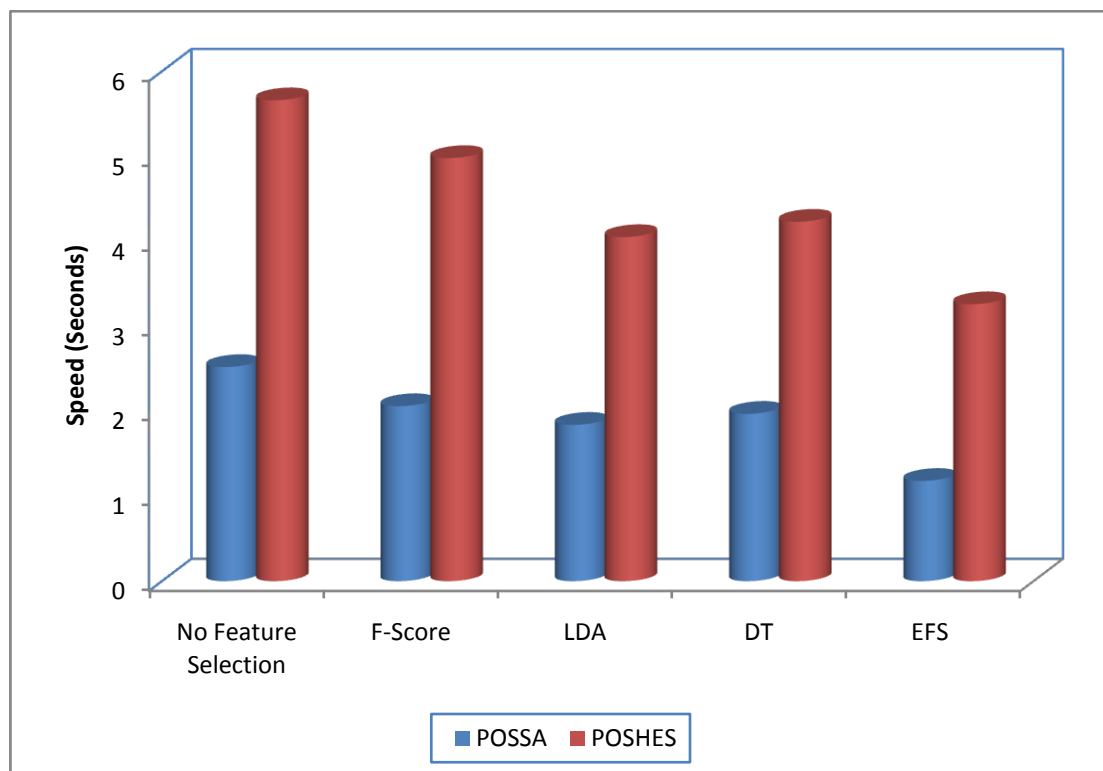


Figure 7.13 : Speed of POS Tagging Algorithms

Detailed error analyzes was also performed to identify the misclassified tags. Table 7.5 presents this analysis, where the first column is the tag and the second and third column represents the trails and hits, while the last column represents the error. For example, in the first line, the system failed to identify the tag <NN> 23 times. From the results, it is again clear that the error rate of the proposed system is less, which is an added advantage while improving the query translation process.

TABLE 7.5
POS TAGGING TRIALS AND ERRORS

Tags	Trials	Hits	Errors	Tags	Trials	Hits	Errors
<NN>	4260	4237	23	<QF>	32	25	7
<NST>	1386	1365	21	<QC>	642	632	10
<NNP>	1890	1872	18	<QO>	30	25	5
<PRP>	975	953	22	<CL>	575	562	13
<DEM>	102	92	10	<INTF>	84	77	7
<VM>	486	459	27	<INJ>	102	89	13
<VAUX>	277	265	12	<NEG>	25	22	3
<JJ>	356	332	24	<UT>	15	12	3
<RB>	288	277	11	<SYM>	6	6	0
<PSP>	30	27	3	<C>	32	31	1
<RP>	45	40	5	<RDP>	2	0	2
<CC>	68	61	7	<ECH>	13	12	1
<VQ>	212	201	11				

Thus, from the various results, it can be concluded that the proposed POSHES tagging algorithm that combined fused feature selection algorithm with SVM-WNN ensemble classifier, to tag Tamil words, is successful, and hence can be used during query translation.

7.3.2. Morphological Analysis

The main goal of using a morphological analyzer is to obtain the root word along with grammatical information, about all the possible word classes for a given word. For this purpose, the use of an ensemble SVM classifier was analyzed. This

section presents the results of the analysis. The analysis is performed using five performance metrics, namely, precision (Table 7.6), recall (Table 7.7), F Measure (Table 7.8), Accuracy (Table 7.9) and speed (Table 7.10) is presented in this section.

TABLE 7.6

**PERFORMANCE OF MORPHOLOGICAL ANALYZER
PRECISION (%)**

Analyzer	Noun (N)	Verb (V)	Pronoun (P)	PN	Others (O)
Tagged SMA	80.95	78.76	82.59	79.81	81.21
Tagged ESMA	81.77	80.86	83.59	81.31	82.62

TABLE 7.7

**PERFORMANCE OF MORPHOLOGICAL ANALYZER
RECALL (%)**

Analyzer	Noun (N)	Verb (V)	Pronoun (P)	PN	Others (O)
Tagged SMA	76.55	73.79	80.47	77.46	79.01
Tagged ESMA	79.24	75.85	82.34	79.09	81.34

TABLE 7.8

**PERFORMANCE OF MORPHOLOGICAL ANALYZER
F-MEASURE (%)**

Analyzer	Noun (N)	Verb (V)	Pronoun (P)	PN	Others (O)
Tagged SMA	78.69	76.19	81.52	78.62	80.09
Tagged ESMA	80.48	78.27	82.96	80.18	81.97

TABLE 7.9
PERFORMANCE OF MORPHOLOGICAL ANALYZER
ACCURACY (%)

Analyzer	Noun (N)	Verb (V)	Pronoun (P)	PN	Others (O)
Tagged SMA	92.78	94.55	96.47	93.87	93.16
Tagged ESMA	94.32	97.13	97.23	95.83	94.22

TABLE 7.10
PERFORMANCE OF MORPHOLOGICAL ANALYZER
SPEED (SECONDS)

Analyzer	Noun (N)	Verb (V)	Pronoun (P)	PN	Others (O)
Tagged SMA	5.12	5.44	5.89	5.16	4.63
Tagged ESMA	9.14	9.35	10.88	9.2	8.87

From the results, it is clear that, the implementation of the ensemble SVM classifier for morphological analysis is successful and improves the process in terms of all selected performance metrics, except for speed. As mentioned in the previous section, this is due to the usage of ensemble concept.

7.3.3. Word Sense Disambiguation

The results, evaluating the performance of the existing and proposed word sense disambiguation algorithms, are presented in this section. The parameters used in previous sections are used. Table 7.11 shows the average results with respect to the selected parameters produced by the existing (CWSD) and proposed (WSDPCS) algorithms. The average performance was estimated by considering the whole database.

From the average results, it is evident that the word sense disambiguation algorithm enhanced with clustering and sense-collocation dictionary perform better than the conventional algorithm. The overall performance indicated by F-Measure and accuracy parameters show that the proposed WSDPCS algorithm has improved the

process of word sense disambiguation by 4.11% and 1.94% when compared with CWSD algorithm.

TABLE 7.11
AVERAGE PERFORMANCE OF WSD ALGORITHM

	CWSD	WSDPCS
Average Precision (%)	84.45	87.76
Average Recall (%)	80.87	84.62
Average F-Measure (%)	82.62	86.16
Accuracy (%)	93.80	95.66
Speed (Seconds)	0.28	0.22

Table 7.12 shows the results of CWSD and WSDPCS for 12 randomly selected ambiguous words for the selected metrics. For each ambiguous word, only two major senses are considered.

In the table, the first and second columns show the ambiguous word and its corresponding two major senses. Both CWSD and WSDPCS were trained using 80% of the corpus data, and the remaining 20% were used as test data. The rest of the columns in the table present the performance of the algorithms in disambiguating the unseen raw occurrences from the 20% test corpus, with respect to the selected performance metrics.

It can be observed from the table, that the algorithm works well for the words such as katir, Kadai and ariah, but words such as tanti and vilangu produced lower results. This is probably due to the data sparseness in the corpus for these words. In all the cases, the proposed algorithm produces improved results than the existing algorithm, indicating that the optimization procedures included are successful in improving the task of word sense disambiguation.

TABLE 7.12
PERFORMANCE OF WSD ALGORITHMS

Sample Words	Senses	Precision (%)		Recall (%)		F-Measure(%)		Accuracy (%)		Speed (sec)	
		CWSD	WSDPCS	CWSD	WSDPCS	CWSD	WSDPCS	CWSD	WSDPCS	CWSD	WSDPCS
Malai	Evening / garland	84.16	86.23	82.66	84.65	83.40	85.43	91.87	92.86	0.28	0.21
Vilangu	Animal / hand-cuff	76.90	79.25	75.43	77.35	76.16	78.29	89.04	89.64	0.28	0.21
Katir	Rays / grain	80.49	82.08	79.16	79.76	79.82	80.90	93.14	94.15	0.28	0.20
Nul	Book / thread	84.11	87.13	82.36	84.76	83.23	85.93	91.40	93.12	0.28	0.22
Tanti	Telegram / string	75.89	78.18	74.65	75.94	75.27	77.04	89.16	90.22	0.26	0.21
Varisai	Queue / offerings in marriage	80.26	82.91	79.12	81.65	79.69	82.27	93.22	94.20	0.27	0.21
Kadai	Shop / ancient weapon	87.64	90.00	86.15	88.85	86.89	89.42	93.77	95.27	0.27	0.22
Tundu	Towel / piece	78.92	81.25	77.09	78.89	77.99	80.05	90.56	91.84	0.28	0.21
Adu	Goat / dance	80.91	82.99	78.92	81.07	79.90	82.02	92.32	92.97	0.29	0.20
Ariah	Cut/ Rare	83.45	86.28	81.59	84.25	82.51	85.25	93.23	94.53	0.26	0.20
Poli	Sweet/piracy	85.25	87.26	82.91	85.94	84.06	86.59	92.15	93.00	0.27	0.21
Tata	Car/bye	81.42	83.50	79.68	81.47	80.54	82.47	92.89	94.17	0.27	0.19

7.3.4. Translation and Transliteration

The final step of Phase II, after performing all the above steps, is the query translation. This section evaluates the proposed algorithm (HQT) used to translate the Tamil query to English language and compares the results with the existing algorithm (WWQT). The effect of using POS tagging, morphological analysis and word sense disambiguation algorithms are also analyzed.

The FIRE dataset consists of three types of queries, namely, Short Title Topic Queries (STTP), Descriptive Topic Queries (DTQ) and Narrative Topic Queries (NTQ). Three performance metrics, namely, F-Measure, Accuracy and Speed are used during evaluation of the translation algorithms. In the tables and figures, the letter P indicates the inclusion of POSHES algorithm, the letter M indicates the inclusion of ESMA, the letter W indicates the inclusion of WSDPCS algorithm with WWQT. The letters S and T indicate that the query translated is obtained as speech data and text data respectively. As speech query is used only with STTP, the F-Measure for DTQ and NTQ are left blank for HQT_{PMW-S}. Table 7.13 shows the F-Measure values obtained by the HQT and WWQT algorithms before and after using the optimization algorithms.

TABLE 7.13

F-MEASURE OF THE QUERY TRANSLATION ALGORITHMS

Translation Algorithm	STTQ	DTQ	NTQ
WWQT	0.79	0.74	0.70
HQT	0.85	0.76	0.73
HQTP	0.87	0.79	0.76
HQTM	0.88	0.82	0.80
HQTW	0.90	0.84	0.81
HQT _{PMW-S}	0.92	-	-
HQT _{PMW-T}	0.94	0.90	0.87

From the above table, it can be understood that the proposed HQT algorithm performs better than the WWQT algorithm. The HQT algorithm improved the translation process by 6.70%. The inclusion of tagging, morphological analysis and word sense disambiguation has a positive impact on the performance of query translation algorithm. The algorithm produces maximum result with STTQ, followed by DTQ and NTQ algorithms. The HQTPMW-S algorithm showed an efficiency gain of 14.19% when compared with WWQT algorithm when using speech-based Tamil query. On the other hand, HQTPMW-T algorithm showed an efficiency gain of 16.02%, 18.23% and 19.10% while using text-based STTQ, DTQ and NTQ respectively. The performance of the translation system was high when used with text-based query than speech based query.

Figures 7.14, 7.15 and 7.16 show the accuracy of the query translation algorithms when used with STTQ, DTQ and NTQ respectively.

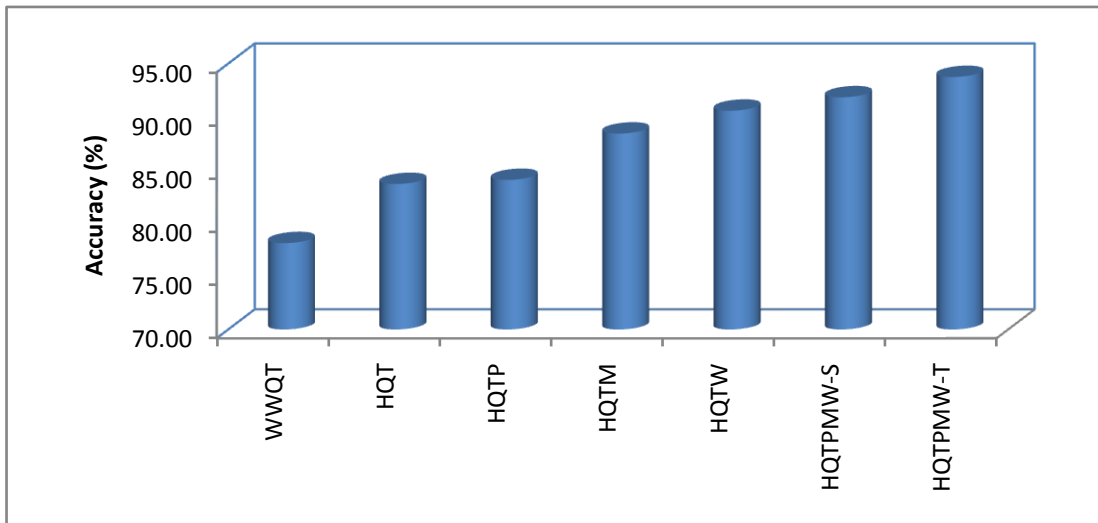


Figure 7.14 : Query Translation using STTQ

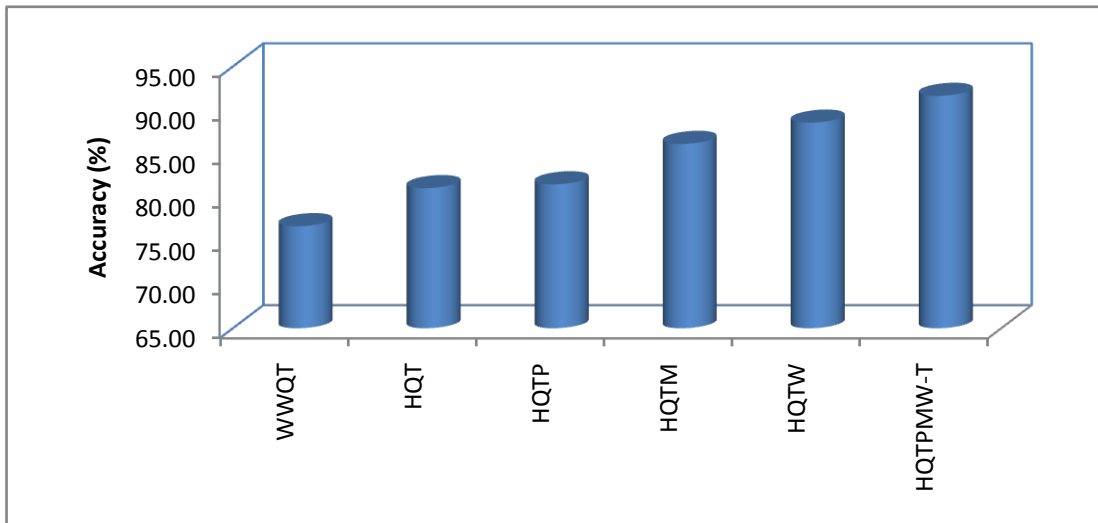


Figure 7.15 : Query Translation using DTQ

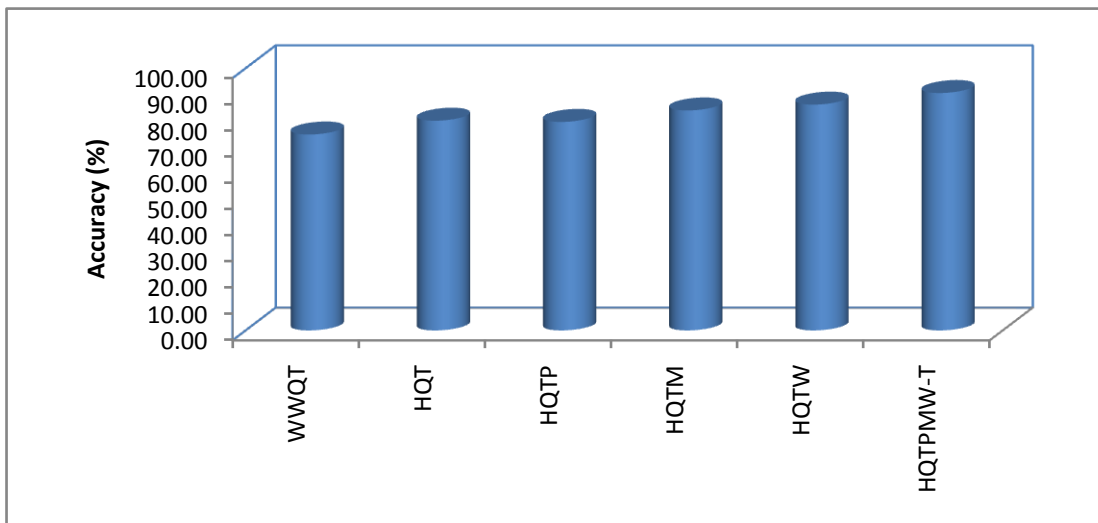


Figure 7.16 : Query Translation using NTQ

The trend envisaged with accuracy performance metric is the same as that of F-Measure. The proposed HQT algorithm performed better when compared to the existing algorithm, and the inclusion of POSHES, ESMA and WSDPCS algorithm has improved the accuracy of query translation.

Figure 7.17 shows the speed of the proposed and existing algorithms without the inclusion of the algorithms presented in the previous sections. The time displayed is the average time taken by a single query to be translated.

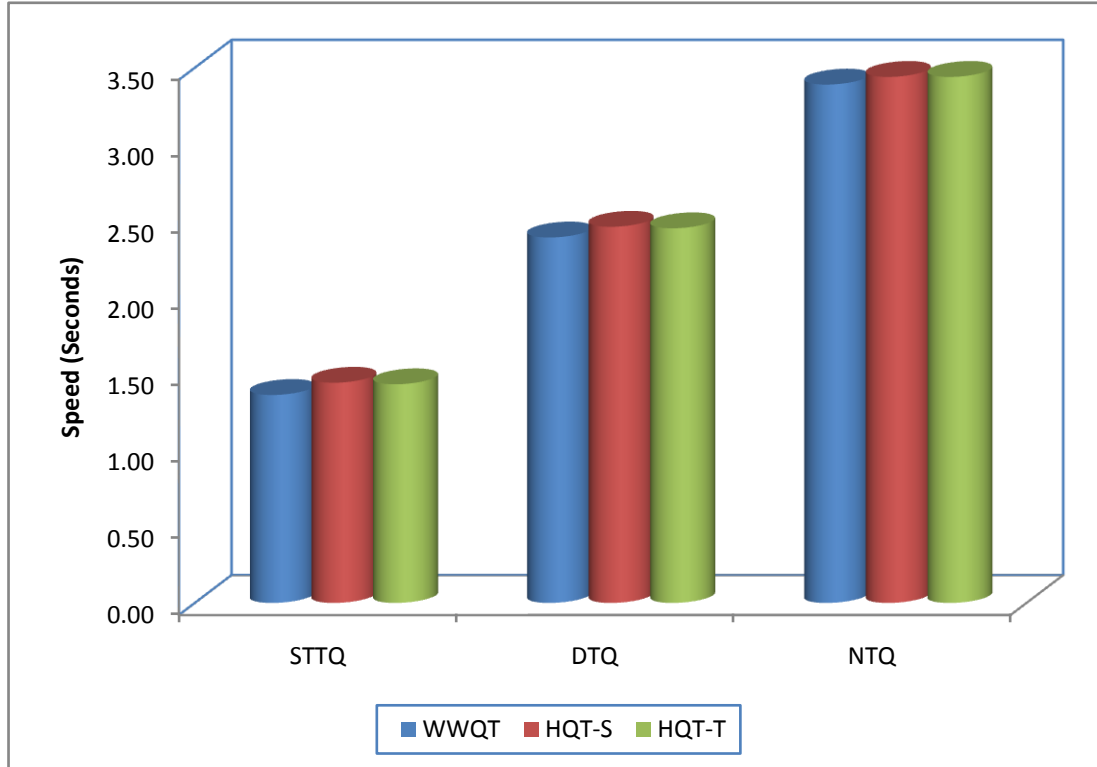


Figure 7.17 : Speed (Seconds) of the Query Translation Algorithms

From the results, it can be seen that the time taken to transfer the query irrespective of its origin (speech or text) is similar. Comparison between existing and proposed algorithm shows that the HQT algorithm is slightly slower than the existing approach. But the difference is only 0.07seconds, 0.06 seconds and 0.05 seconds and does not make much difference during translation. Thus, in conclusion, the proposed HQT algorithm produces maximum efficiency during translation when combined with tagging, morphological analysis and word sense disambiguation.

7.4. TEXT RETRIEVAL

Details regarding the experimental results, conducted to evaluate the performance of the text retrieval algorithms proposed in Phase III of the research work, are presented in this section.

This section uses F Measure, Macro F Measure, MAP, precision, R-precision and speed performance metrics during evaluation. Table 7.14 presents the coding scheme used in this section.

TABLE 7.14
CODING SCHEME USED IN PHASE III

Code	Description
KNN	K-Nearest Neighbour Classifier
ARC	Associative Rule Classifier
IKNN	Improved KNN
IARC	Improved ARC
HAKNN	Hybrid Association Rule Mining and IKNN
EHAKNN	Ensemble Hybrid ARC and IKNN Classifier

Tables 7.15, 7.16 and 7.17 show the performance of the text retrieval algorithms with respect to F-Measure while using STTQ, DTQ and NTQ for the ten categories in FIRE dataset. Figures 7.18, 7.19 and 7.20 present the macro F Measure while using STTQ, DTQ and NTQ in FIRE dataset.

TABLE 7.15**F-MEASURE OF TEXT RETRIEVAL SYSTEMS WITH SPEECH-BASED STTQ**

CATEGORIES	KNN	ARC	IKNN	IARC	HAKNN	EHAkNN
Bengal	0.7413	0.7128	0.7682	0.7532	0.7962	0.8125
Business	0.8121	0.7862	0.8365	0.8273	0.8547	0.8679
Calcutta	0.7089	0.6845	0.7289	0.7099	0.7475	0.7625
Foreign	0.7452	0.7254	0.758	0.7585	0.7789	0.7923
Front page	0.7186	0.7055	0.7362	0.7245	0.7559	0.7712
Nation	0.8064	0.7832	0.8231	0.8125	0.8435	0.8658
Opinion	0.7312	0.7122	0.7525	0.7466	0.778	0.8032
Sports	0.7444	0.731	0.7659	0.7589	0.795	0.8257
Jobs	0.8624	0.8415	0.8742	0.8695	0.8912	0.9122
Entertainment	0.7295	0.6989	0.7462	0.7312	0.7708	0.7962

TABLE 7.16**F-MEASURE OF TEXT RETRIEVAL SYSTEMS WITH TEXT-BASED STTQ**

CATEGORIES	KNN	ARC	IKNN	IARC	HAKNN	EHAkNN
Bengal	0.7728	0.7518	0.8161	0.8003	0.8351	0.8427
Business	0.8458	0.8210	0.8723	0.8679	0.9010	0.9110
Calcutta	0.7539	0.7179	0.7653	0.7550	0.7817	0.7957
Foreign	0.7923	0.7704	0.8015	0.8015	0.8129	0.8325
Front page	0.7608	0.7523	0.7798	0.7614	0.7912	0.8028
Nation	0.8475	0.8329	0.8602	0.8501	0.8810	0.9057
Opinion	0.7774	0.7458	0.7825	0.7946	0.8216	0.8495
Sports	0.7770	0.7619	0.8037	0.8001	0.8275	0.8654
Jobs	0.9007	0.8745	0.9124	0.9049	0.9282	0.9464
Entertainment	0.7741	0.7318	0.7800	0.7787	0.8041	0.8370

TABLE 7.17**F-MEASURE OF TEXT RETRIEVAL SYSTEMS WITH TEXT-BASED DTQ**

CATEGORIES	KNN	ARC	IKNN	IARC	HAKNN	EHAkNN
Bengal	0.7852	0.7625	0.8156	0.7933	0.8362	0.8582
Business	0.8492	0.8294	0.8752	0.8645	0.8925	0.9214
Calcutta	0.7567	0.7326	0.7897	0.7713	0.8066	0.8345
Foreign	0.7745	0.7528	0.796	0.7795	0.8258	0.856
Front page	0.7525	0.7312	0.7945	0.7768	0.8195	0.8466
Nation	0.8355	0.821	0.8625	0.8468	0.8854	0.9087
Opinion	0.7762	0.7532	0.8035	0.788	0.8252	0.8518
Sports	0.7842	0.7705	0.8158	0.7945	0.8375	0.8635
Jobs	0.8933	0.8764	0.9146	0.8988	0.9315	0.9517
Entertainment	0.7515	0.7382	0.7835	0.7558	0.8062	0.8292

TABLE 7.18**F-MEASURE OF TEXT RETRIEVAL SYSTEMS WITH TEXT-BASED NTQ**

CATEGORIES	KNN	ARC	IKNN	IARC	HAKNN	EHAkNN
Bengal	0.8262	0.8058	0.8468	0.8298	0.8643	0.8816
Business	0.8760	0.8524	0.8915	0.8845	0.9254	0.9470
Calcutta	0.8058	0.7815	0.8356	0.8147	0.8562	0.8835
Foreign	0.8185	0.7938	0.8462	0.8298	0.8695	0.8948
Front page	0.7962	0.7712	0.8587	0.8263	0.8740	0.9047
Nation	0.8794	0.8550	0.9072	0.8861	0.9267	0.9522
Opinion	0.8065	0.7862	0.8558	0.8326	0.8853	0.9166
Sports	0.8417	0.8165	0.8759	0.8554	0.9078	0.9344
Jobs	0.9225	0.9018	0.9336	0.9298	0.9522	0.9747
Entertainment	0.8364	0.8145	0.8643	0.8500	0.8945	0.9234

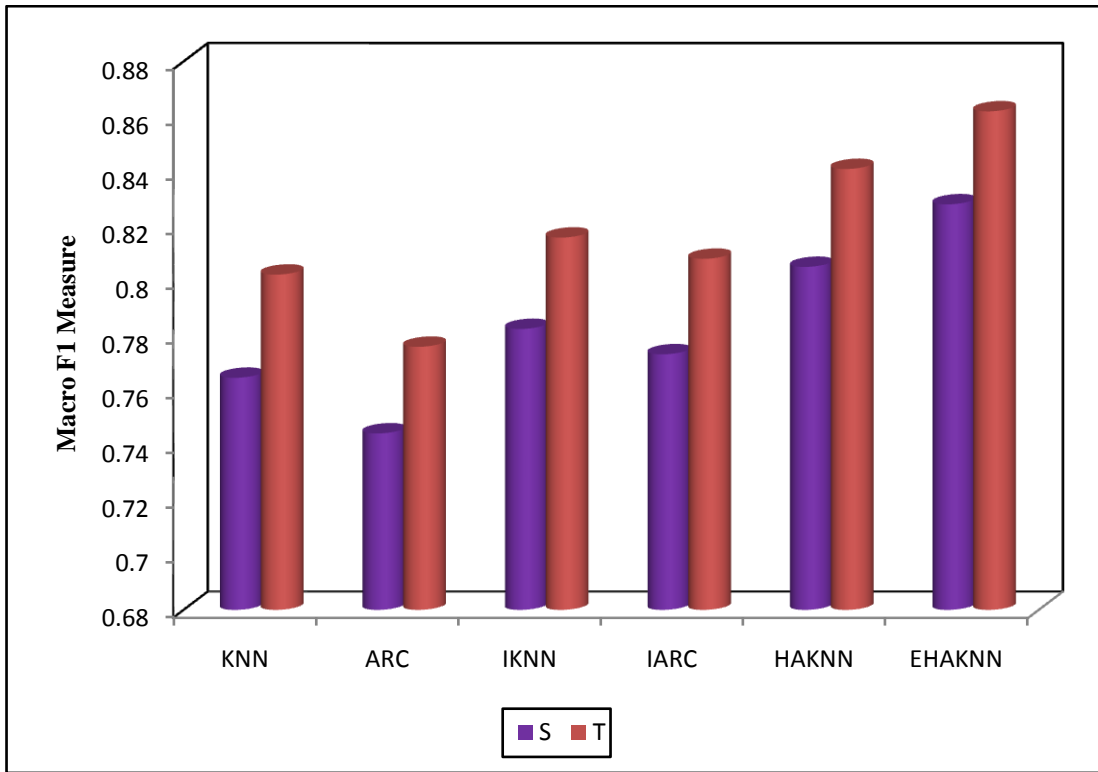


Figure 7.18 : Macro F-Measure of Text Retrieval Systems with Speech-Based and Text-Based STTQ

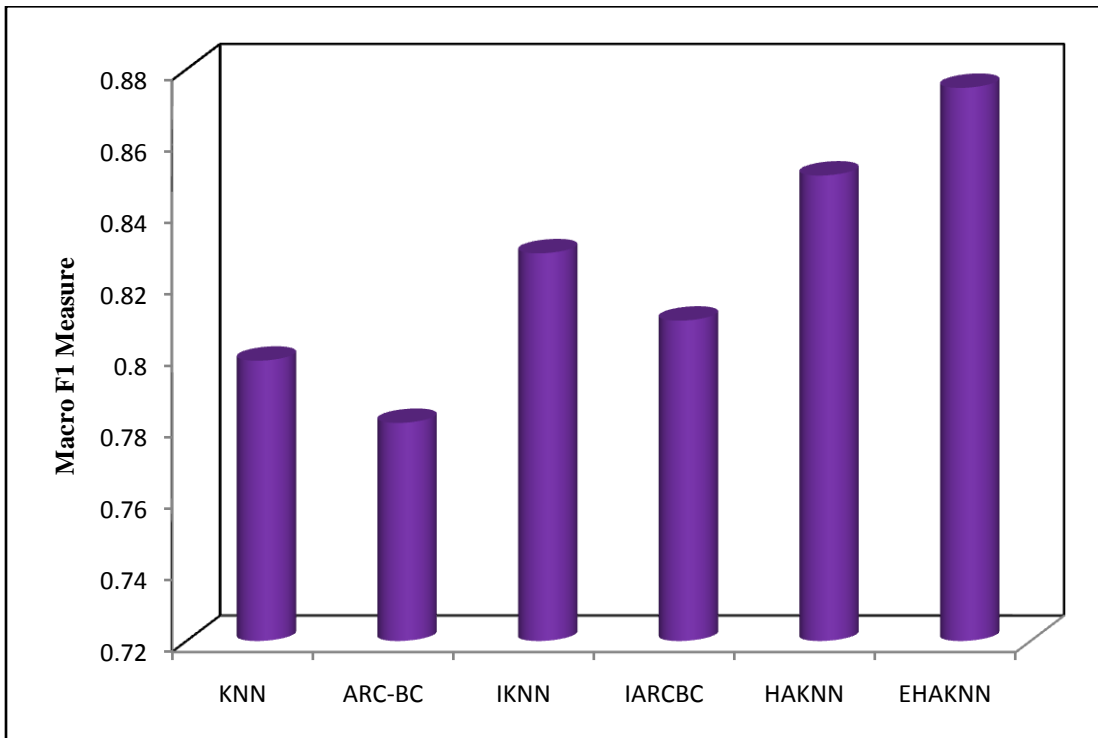


Figure 7.19 : Macro F-Measure of Text Retrieval Systems with Text-Based with DTQ

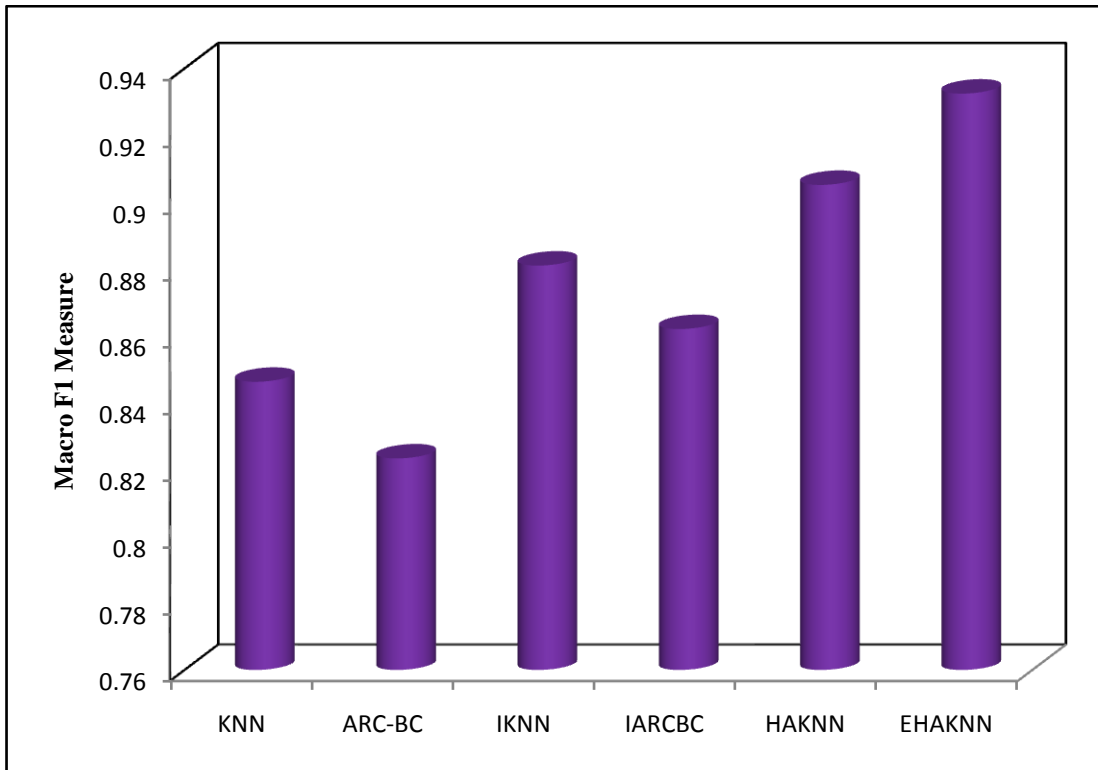


Figure 7.20: Macro F-Measure of Text Retrieval Systems with Text-Based NTQ

Analysis of macro F Measure of STTQ showed that the EHAKNN algorithm improved the text retrieval process by 6.90% and 9.96% when compared with the conventional KNN and ARC classifiers respectively, while using speech-based queries. The text retrieval system using text-based queries, on the other hand, showed an efficiency gain of 7.63% and 10.09% over KNN and ARC classifiers respectively. The efficiency gain achieved by EHAKNN was 8.72% and 10.72% with DTQ and 9.25% and 11.71% with NTQ over KNN and ARC classifiers respectively, when used with text based queries (TECLTR-T).

The retrieval results with respect to F Measure and Macro F Measure reveal that while all the proposed classifiers perform better than the existing KNN and ARC classifier, the proposed ensemble hybrid associative rule mining and KNN classifier (EHAKNN) produced maximum improvement during text retrieval with all the three types of queries. This held true when applied with both speech-based query and text-based query.

Figures 7.21, 7.22 and 7.23 show the average accuracy of the proposed TECLTR system while using the enhanced classifiers proposed in Phase III when provided with the queries in STTQ, DTQ and NTQ from FIRE dataset.

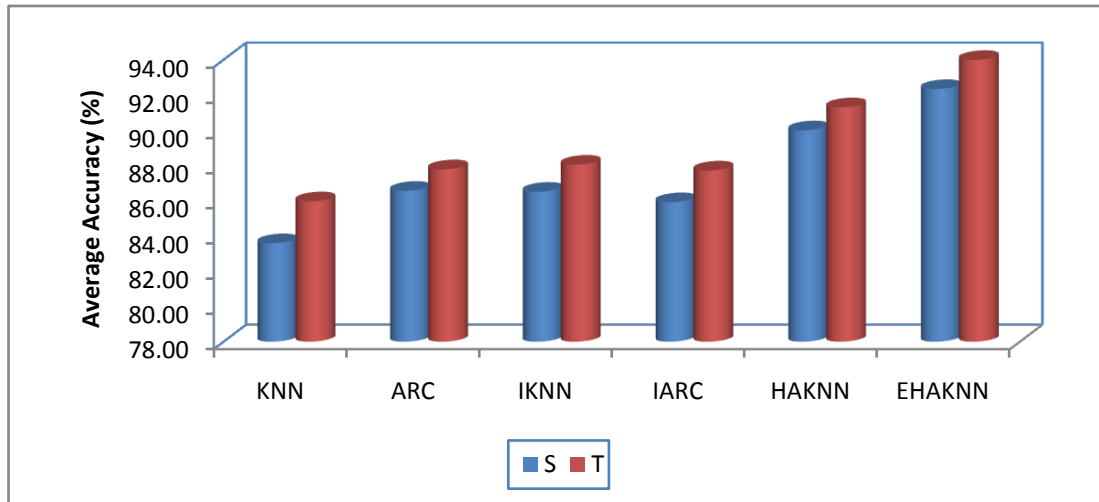


Figure 7.21 : Average Accuracy of the Text Retrieval Systems with STTQ

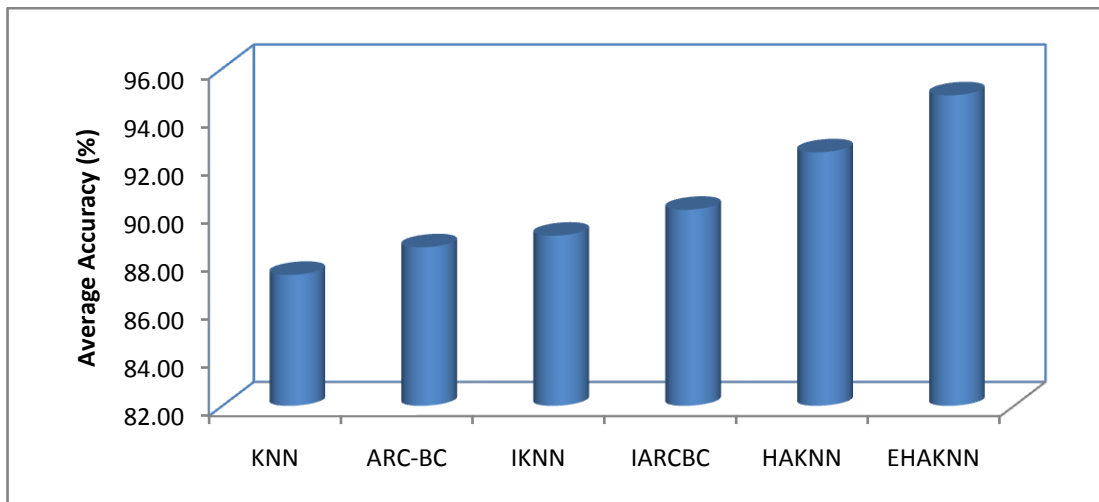


Figure 7.22 : Average Accuracy of the Text Retrieval Systems with DTQ

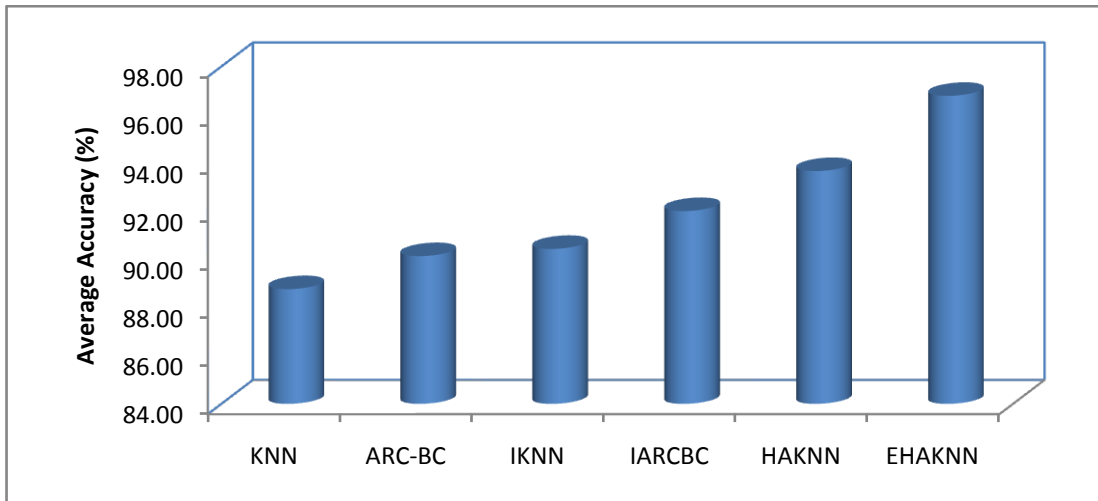


Figure 7.23 : Average Accuracy of the Text Retrieval Systems with NTQ

Examination of results from these figures showed that a high accuracy was achieved by EHAKNN classifier with both TECLTR-S and TECLTR-T with all types of queries. Thus, the results concerning accuracy again reveal that the proposed ensemble classifier produces better CLTR results, when compared with the existing KNN and ARC classifiers.

The results projected in this section prove that the EHAKNN algorithm produces maximum efficiency while retrieving text documents from FIRE dataset. The following paragraphs compare the performance of this system with monolingual system in terms of MAP, precision, recall and R-precision and analyze the effect of using query expansion with STTQ in CLTR. Mono-Lingual Text Retrieval System (MLTR) refers to Information Retrieval System that can find relevant documents in the same language as the query was expressed and the Existing Cross-Language Text Retrieval System (ECLTR) uses word-by-word translation approach. In the discussion, QE represents the application of Query Expansion algorithm, which is applied only to queries belonging to STTQ and as with earlier sections the speech query is used only with STTQ.

Table 7.19 presents the MAP, precision and R-precision values obtained by the CLTR systems using STTQ from FIRE dataset.

The precision and recall curve of the CLTR systems are shown in Figures 7.24, 7.25 and 7.26 for queries from STTQ, DTQ and STQ respectively.

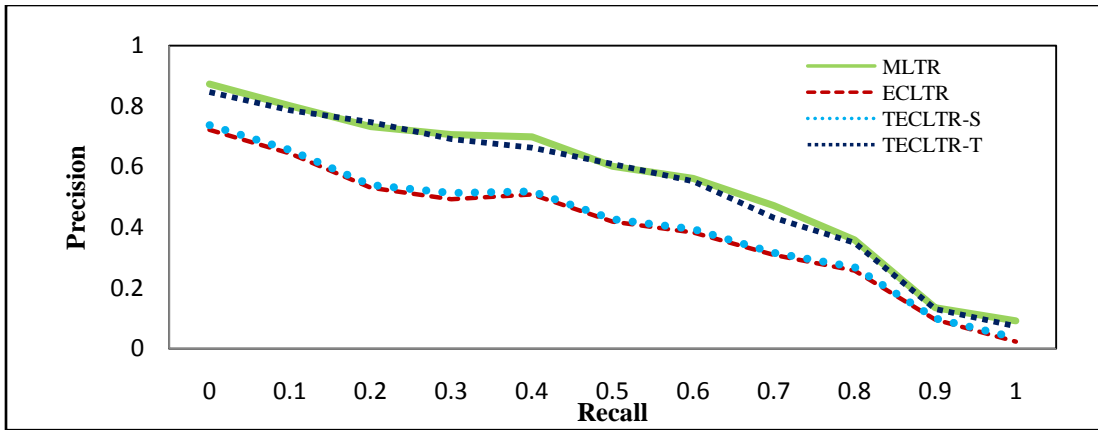


Figure 7.24 : P-R Curve for CLTR Systems using STTQ Queries

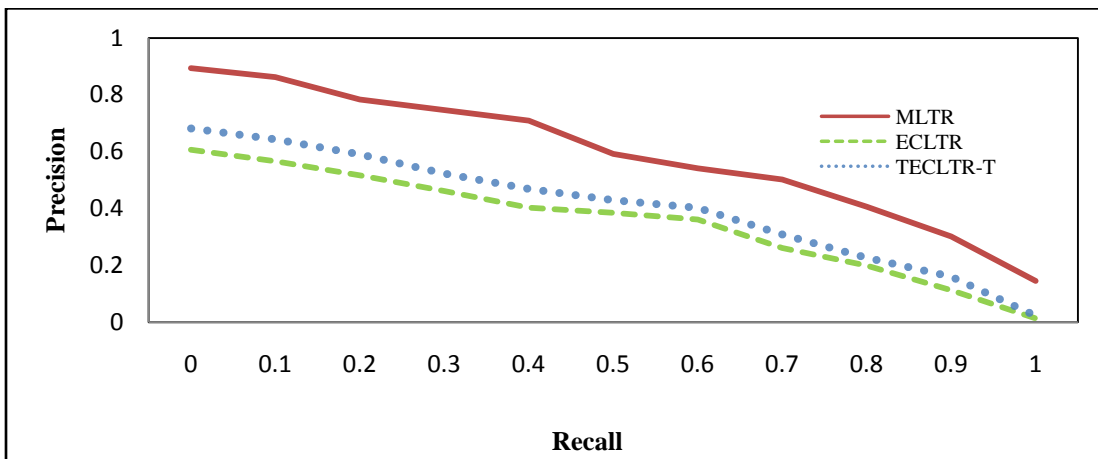


Figure 7.25 : P-R Curve for CLTR Systems using DTQ Queries

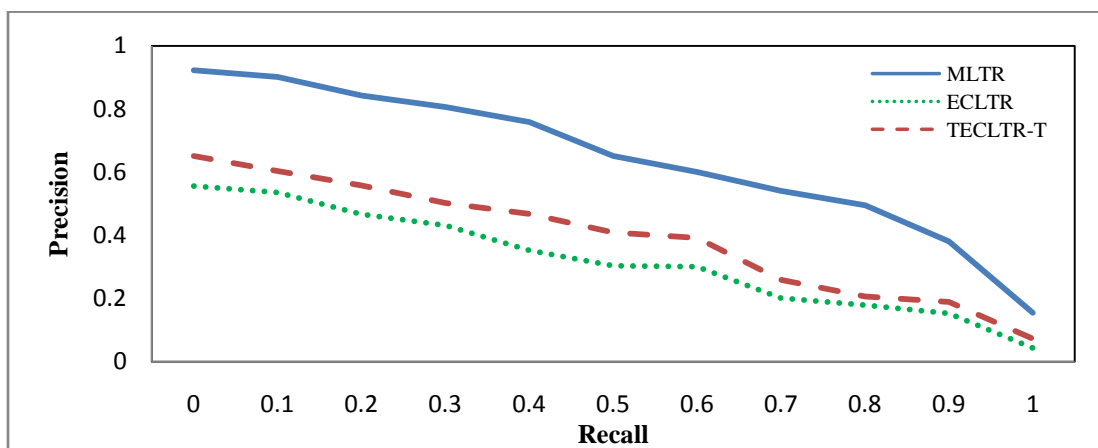


Figure 7.26 : P-R Curve for CLTR Systems using NTQ Queries

TABLE 7.19

EVALUATION OF THE MONO-LINGUAL AND CROSS-LINGUAL RUNS

Queries	CLTR	MAP	P@10	P@20	R-Precision
STTQ	MLTR	0.5845	0.7140	0.6542	0.5228
	ECLTR	0.4962	0.5460	0.5015	0.4162
	TECLTR-S	0.5004	0.5509	0.5061	0.4204
	TECLTR-T	0.5684	0.6763	0.6077	0.5123
DTQ	MLTR	0.6176	0.7528	0.7022	0.5615
	ECLTR	0.5275	0.6327	0.5925	0.4724
	TECLTR-T	0.5521	0.6805	0.6348	0.5000
NTQ	MLTR	0.6312	0.7730	0.7045	0.5812
	ECLTR	0.5525	0.6385	0.6124	0.4835
	TECLTR-T	0.5722	0.6972	0.6460	0.5175

From the results, it is clear that the speech recognition algorithm included with the proposed TECLTR-S system has shown positive improvement and its performance is quite near to that of the existing word-to-word CLTR system (ECLTR). The proposed TECLTR-T produced best results when compared with both monolingual, ECLTR and TECLTR-S. This trend was the same with all the selected performance metrics and query types.

The execution time of the CLTR systems is presented in Table 7.20. This is estimated as the time duration of the algorithm from the input state to retrieval state and involves the tasks of all the three phases of the research work.

TABLE 7.20**SPEED (SECONDS) OF CLTR SYSTEMS**

Queries	CLTR	Speed
STTQ	ECLTR	73.96
	TECLTR-S	116.11
	TECLTR-T	95.82
DTQ	ECLTR	94.23
	TECLTR-T	122.3
NTQ	ECLTR	101.89
	TECLTR-T	136.94

The execution time results show that the proposed algorithms are slower than the monolingual and existing CLTR systems. This is due to the computational complexities introduced by the ensemble classifier used in various steps of the proposed TECLTR systems. Increased computations required for classifying an input query, indicates that all component classifiers (instead of a single classifier) must be processed and thus it requires more execution time (https://en.wikipedia.org/wiki/Ensembles_of_classifiers #Weaknesses; Anasyali and Kusman, 2011).

As producing more accurate information retrieval is the main aim of any CLTR system, this research work concludes that the proposed CLTR systems are successful and perform well when supplied with speech-based and text-based short title topic queries, descriptive topic queries and narrative topic queries.

7.5. CHAPTER SUMMARY

This chapter presented the performance evaluation of the various proposed techniques used during the design and implementation of the proposed CLTR systems. From the various results, the entire proposed algorithms have contributed significantly to the improvement of the

performance of CLTR, both in text-based and speech-based. The work is summarized with future research direction in the next chapter, Chapter 8, Summary and Conclusion.