Chapter 8

Proposed Two Stage method for emotion classification

8.1 Introduction and Problem formulation

‘Affective computing’ is a growing research area used to train the devices in such a way to detect and respond to human emotions in a more appropriate and empathic manner [126]. It is mainly used to enhance the communication between human and the machine by capturing and processing information effectively [20] [76]. With this a machine can respond to the user in a natural way like a human being. Even though an extensive development and usage of speech emotion recognition is performed in certain applications like education, entertainment, multi media contents management, text to speech synthesis and medical diagnosis, at there is lack of development in speech understanding and recognition applications. For instance, in real time driving scenario application, detecting driver’s emotion and alerting him from causing an accident is a difficult task because the driver’s behavior changes with the emotion which occurs due to the communication with co-passengers, entertainment system and mobile devices as shown in Fig. 8.1.

The features extracted from the pre processed speech samples carry more emotional information. In the first stage feature selection Energy, Pitch and MFCC features are selected and fused together. Even though the performance of the system is improved with feature fusion than with individual features, it does n’t reach to an optimal state because of curse of dimensionality i.e., the
number of features extracted are more when compared with the number of speech samples. So to improve the performance further, the innovative step is to use second stage feature selection technique followed by feature fusion. In this stage of feature selection, an optimal feature subset selection methods like Sequential Forward Selection (SFS) and Sequential Floating Forward Selection (SFFS) are applied to these fused features followed by an application of various classification techniques like Linear Discriminant Analysis (LDA), Regularized Discriminant Analysis (RDA), Support Vector Machine (SVM) and k Nearest Neighbor (kNN). This chapter focuses more on experimental and analysis of results with feature fusion and two stage feature selection.

The performance of overall emotion recognition system is validated over Berlin and Spanish databases. An optimal uncorrelated feature set is obtained by using SFS and SFFS individually. Results could reveal that SFFS is a better choice for feature subset selection because SFS suffers from nesting problem i.e it is difficult to discard a feature after it is retained into the set. SFFS overcomes this nesting problem by making the set not to be fixed at any stage but floating up and down during the selection based on the objective function. Here the objective function used is Un weighted Average Recall. Experimental results showed that the efficiency of the classifier was improved by 15%-20% with two stage feature selection method when compared with feature fusion.
8.2 Methodology

The Block diagram of the proposed ‘Two Stage Feature Selection’ method is shown in Fig. 8.2. It shows the overall process, how the speech samples are processed in different stages to recognize the emotional state of the person. Initially speech samples are preprocessed before the feature extraction. In the first stage of feature selection, the appropriate features which best classify the emotion are selected and fused together to enhance the performance of the system. For further improvement, we propose a second stage feature selection based on SFS and SFFS (optimal feature subset selection methods) are used to reduce the dimensionality of the fused feature vector by collecting a set of uncorrelated features. This optimal feature set is given as input to several classifiers for correct emotional classification.

Figure 8.2: Block diagram for speech emotion recognition system using two stage feature selection. stage1-Feature Fusion Set stage2- optimal feature set selection

Filtering, framing and windowing comes under the task of preprocessing of speech samples. A high pass filter is used to reduce the environmental noise while recording the speech sample. In framing, speech signal is split into several frames with 256 speech samples for each frame, an overlapping of 100 samples is done with a hamming window and a feature vector is extracted from each frame. This feature vector is used to classify the emotion of the speech sample based on
some simple statistics like mean, variance, minimum, range, skewness and kurtosis because they are less sensitive to the linguistic data. The detailed analysis of framing and windowing are given in the first stage feature selection.

The dimensionality of the feature vector is very high after feature fusion, so to reduce the dimensionality of the fused vector, an optimal feature subset selection techniques such as SFS and SFFS are also used to enhance the performance of the classifier.

8.2.1 First Stage Feature Selection

An important module in the speech emotion recognition system is the selection of the best features because there is no theoretical basis about the features which could classify the emotion. So the work is based on the features obtained from direct comparison of speech samples portraying different emotions. This comparison is useful for identifying the best features which are useful for emotion identification. In this section the features like Energy, Pitch, and Mel Frequency Cepstral Coefficients are used which contain most of the emotion specific information based on the literature.

Feature extraction includes extraction of feature vectors from the speech sample. The feature vectors are generated for each frame. Energy and Pitch are the most emotion specific prosodic features which are extracted from the speech sample. The information provided by the first and second order derivatives of these features are also considered. The amplitude variations of the speech signal are used to calculated the energy and the auto correlation method is used to calculate the pitch. The short time energy and pitch features for each frame are represented by Eq. (8.1) and (8.2)

$$E = \sum_{i=1}^{N} x_i^2$$  \hspace{1cm} (8.1)
\[ R_n(k) = \sum_{i=1}^{N} x(i)x(i+k) \] (8.2)

where \( x_i \) is the speech sample for the \( i^{th} \) frame, \( k \) is the time lag and \( N \) is the total number of frames. In total, 18 values are extracted from each prosodic feature which includes the corresponding feature and their first, second derivatives by using the six statistics mean, variance, minimum, range, skewness and kurtosis. Finally 36 values are extracted from both Energy and Pitch features in total.

In order to extract the correct emotional state of the speech sample the most efficient spectral representation of the speech sample is Mel Frequency Cepstral Coefficients (MFCC) [67]. The Mel-frequency scale is given in Eq 8.3

\[ \text{mel}(f) = 2595 \cdot \log_{10}(1 + \frac{f}{100}) \] (8.3)

where \( f \) is the normal frequency and \( \text{mel}(f) \) is the mel frequency for a given \( f \). The procedure for implementing the MFCC is shown in detail in [103] [131] [129]. 18 MFCC Coefficients are estimated from each frame along with their first and second derivatives which give a total of 54 spectral features. All these features are estimated over simple six statistics, making 324 spectral features in total which are extracted from each speech sample.

The 36 prosodic feature values and 324 spectral feature values are fused together with feature fusion and obtain 360 feature values in total. The fused vector is given to a classifier for recognition of emotion of speech samples. This will give better results than when working with individual features [60]. Even though it could give better results, it should not reach to an optimal state because of high dimensional feature vector.

8.2.2 Second Stage Feature Selection

To reduce the curse of dimensionality, which occurs in first stage feature selection and also to obtain better performance of speech emotion recognition system,
second stage feature selection is necessitated. This stage includes extraction of optimal feature subset from the fused feature set. Basically, feature selection is the process of finding a subset of \( n \) features from a given set of \( N \) features i.e \( n < N \) without significantly degrading the performance of the classifier. In feature subset selection, each feature is assigned a value to reflect its usefulness.

The fused feature vector is given as input to the feature subset selection technique. Here we used two types of feature subset selection techniques viz., Sequential Forward Selection (SFS) and Sequential Floating Forward Selection (SFFS) individually.

8.3 Feature Subset Selection Techniques

Optimal feature selection techniques can be distinguished as filters or wrappers based on the criterion function employed. Feature selection can be performed by using the properties such as orthogonality, mutual information, correlation etc. In filter approach, all the features are given a ranking by using some statistical criteria. Highest ranking features are selected and lowest ranking features are ignored. But the disadvantage of this method is, they ignore the interdependency of the features and also ignore the interaction with the classifier\cite{116}. Due to this, the performance of the classifier decreases. Even though the wrappers are slower than the filters, the selected features are more discriminative for specific classifier as wrappers train a classifier using the selected features and estimate the classification error using the validation set\cite{51} \cite{56}. The most promising methods for wrappers are feature subset selection techniques \cite{93} \cite{138}. If the total number of features is \( n \) then the possible number of feature subset is \( 2^n \) in the search space. Now our task is to search for the space of possible feature subsets and to find the best optimal feature subset which will classify the emotional states with low classification error rate\cite{137} \cite{29}. 
Each feature subset selection algorithm concentrates on search strategy and evaluation method. Search strategy is used to select the feature subsets and evaluation method is used to test their fairness and suitability based on some criterion function. Search strategies can be classified into exhaustive, sequential and random search. In exhaustive research the number of possible feature subsets grow exponentially, making this method impractical even for small feature sets. In Sequential search algorithms, insertion or deletion of features is done sequentially. Sequential Forward Selection (SFS) and Sequential Backward Selection (SBS) are some examples of sequential algorithms which are simple to implement and their execution time is very low. These algorithms are proposed by Whitney in 1971. SFS starts with an empty set and adds the best selected feature to the feature set in each iteration and in a similar way SBS starts with entire feature set and removes the worst performing feature from the entire set. SFS and SBS suffer from nesting effect and to prevent this nesting of feature subsets, another method called ‘plus-l-minus-r’ is developed by Stearns in 1976. The drawback in this method is that there is no procedure to predict the values of ‘l’ and ‘r’ to achieve the best feature set. Instead of fixing these values keep them to float i.e to change the values flexibly to approximate the optimal solution. These are called floating search methods which include and exclude features based on the direction of the search. The Sequential Floating Forward Selection (SFFS) is the floating method in the forward direction and Sequential Floating Backward Selection (SFBS) is the search method in the opposite direction. In random search algorithms, it randomly selects subset and insert or delete feature sets. Evolutionary algorithms come under random search algorithms which are used for feature selection.

In this thesis the Sequential Forward Selection and Sequential Floating Forward Selection are applied for feature selection in order to maximize the emotion classification performance with a low dimensionality feature vector.
8.3.1 Sequential Forward Selection (SFS)

SFS starts with an empty set, and the feature set is iteratively updated by adding the most significant feature by using a criterion function in each step. The criterion function used here is the Unweighted Average Recall (UAR). If the selected feature satisfies the criterion function, then it should be included into the updated set else it should search for the next best feature from the fused feature set. The functionality of the criterion function is, to check the performance of the classifier with the newly updated feature set and with the old one. The efficiency of the classifier is assessed by using predicted labels and actual labels of the speech samples. If we get the better performance, then only the most significant feature should be included into the feature set or else it should not be added into the feature set. In this way to add most significant feature into the updated set this process is repeated with each feature continuously till we get the optimal feature subset. But the drawback in this method is, once a feature is added into the feature set, it should not be possible to remove that feature from the set. This is called ‘nesting effect’. The problem with this nesting effect is overcome by using Sequential Floating Forward Selection.

8.3.2 Sequential Floating Forward Selection (SFFS)

SFFS includes new features by using basic SFS procedure followed by successive conditional deletion of least significant feature in the updated set which provides a better feature subset. The procedure for the selection of subset of features from the fused feature set using SFFS is shown in Fig. 8.3 which also includes SFS. From this Figure, initially a fused feature vector $X$ generated from the first stage is used. In order to get the best useful features from this feature vector, 3 steps are used mainly. First it starts with an empty set $Y = \phi$ and the SFS is used to select the most significant feature from the feature vector and includes it into the set $Y$. If the newly added set satisfies the criterion function, then we
include that feature into the set or else we select the next best feature from the feature vector and add into the set. The nesting effect occur in SFS is eliminated in this SFFS using the selection least significant feature from the newly added set. If the deletion of the least significant feature satisfies the criterion function, then we exclude that feature from the set or else we continue with the forward selection. If the deletion does n’t satisfy the criterion function, then we continue with conditional exclusion which is the final step in this procedure. The criterion function ‘Unweighted Average Recall’ is calculated by using the formula given in
CHAPTER 8. PROPOSED TWO STAGE METHOD FOR EMOTION CLASSIFICATION

Eq. 8.4

\[ \text{Accuracy} = \left( \frac{\text{sum}(N\text{correct samples})}{\text{sum}(N\text{instances})} \right) \times 100 \]  

(8.4)

8.4 Experimental results

In this work various classification algorithms are used for classification, namely Linear Discriminant Analysis (LDA), Regularized discriminant Analysis (RDA), Support Vector Machine (SVM) and k-Nearest Neighbour (kNN) along with Sequential Forward Selection (SFS) and Sequential Floating Forward Selection (SFFS) feature selection methods. The implementation of these methods is carried out using MATLAB programming. The description of each algorithm is given in detail in [60]. The emotional classes used in this work are happy, neutral, anger, sad, fear and disgust. The way the class label assigned to each test speech sample depends on the minimum of the Euclidian distance of the training samples. Each classifier is given an enough training data for better classification of test speech samples. LDA suffers from singularity problem because of high dimensional and low sample size. So it is difficult to get the accurate results with LDA. This singularity problem is overcome by RDA with a regularization technique, with which the performance of the classifier improved. kNN basically, does not deal with feature relevance but SVM and RDA can better deal with high dimensionality and irrelevant features. These things justify that RDA and SVM better classify the speech samples. The concept of two stage feature selection would play a major role here, which improves the performance of each classifier effectively. The performance of the classification method mainly depends here on the quality of the feature set.
CHAPTER 8. PROPOSED TWO STAGE METHOD FOR EMOTION CLASSIFICATION

8.4.1 Material used

In general, research with emotional speech samples deal with acted, induced and completely spontaneous database of emotions [61][139]. More number of emotional speech databases are designed. Some of them are Emo-DB (Berlin emotional speech database), DES (Danish database of emotional speech), SES (Spanish emotional Speech database), Chinese and English emotional speech databases. In this work, the experiments are conducted over Berlin and Spanish emotional databases which come under acted emotional speech databases. The details of these databases are given in the Table 8.1.

Berlin database is an open source, simulated emotional speech database which contains 7 basic emotions anger, boredom, disgust, fear, happiness, sadness and neutral etc. There are totally 535 different German emotional speech samples which are simulated by 10 professional native German actors (5 actors and 5 actresses) [13].

Spanish database contains seven emotions anger, sadness, joy, fear, disgust surprise and neutral. There are 184 sentences for each emotion which include isolated words, sentences and a text passage. There are 1288 Spanish emotional speech samples in total. These emotional speech samples are recorded by one professionally male speaker and one professionally female speaker [41][62].

<table>
<thead>
<tr>
<th>Type of files</th>
<th>No.of files in database</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Berlin</td>
</tr>
<tr>
<td>Training samples</td>
<td>357</td>
</tr>
<tr>
<td>Test samples</td>
<td>178</td>
</tr>
<tr>
<td>Total samples</td>
<td>535</td>
</tr>
<tr>
<td>Emotions</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 8.1: Description of Type of files, No.of files and corresponding number of emotions in Berlin and Spanish databases.
8.4.2 Parameter settings

A two stage feature selection for speech emotion classification is applied individually for Berlin and Spanish databases with individual feature selection techniques like Sequential Forward selection (SFS) and Sequential Floating Forward selection (SFFS), and the results are compared effectively. Databases are divided into training and testing set, in which 2/3 rd of the whole data samples are used for training and 1/3 rd of the samples are used for testing. Initially each classifier is trained with the data provided in training set. Test speech sample is classified by using a classifier and the information provided by the training speech samples.

In order to validate the results of different emotional speech samples on various classifiers like Linear Discriminant Analysis (LDA), Regularized Discriminant Analysis (RDA), Support Vector Machine (SVM) and k Nearest Neighbour (kNN) the experiments have been conducted in two phases. First phase consists of Base line results with feature fusion and the second phase comprises results with optimal feature subset selection techniques.

8.4.3 Base line results

The task of feature fusion is to combine the Pitch, Energy and MFCC features extracted from emotional speech samples. These fused features are classified by several classification techniques and the results are shown in Table 8.2. According to these results, the performance of LDA and kNN are between the range 50%-60%. Even though the performance of the RDA and SVM classifiers is between 60%-70%, it does not reach to an optimal state because of high dimensional feature vector. So, to have further improvement in the performance of all the classifiers, a two stage feature selection technique need to be adopted.
### Table 8.2: Emotion recognition percentage accuracy of various classifiers (LDA, RDA, SVM and kNN) over Berlin and Spanish databases using feature fusion.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Feature Fusion Set (FFS)</th>
<th>Berlin (%)</th>
<th>Spanish (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDA</td>
<td>57.2</td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>RDA</td>
<td>73.8</td>
<td>69.4</td>
<td></td>
</tr>
<tr>
<td>SVM</td>
<td>70.5</td>
<td>67.6</td>
<td></td>
</tr>
<tr>
<td>kNN</td>
<td>65</td>
<td>60.8</td>
<td></td>
</tr>
</tbody>
</table>

8.4.4 **Results with optimal two stage feature selection techniques**

Initially, a set of fused features are collected by combining the Energy, Pitch and Mel Frequency Cepstral Coefficients were extracted from the speech sample. In order to get the best useful features from this fused features a feature subset selection algorithm is applied. This optimal feature set leads to an effective improvement in the performance of the classifiers when compared with base line results and are shown in Table 8.3. From this table it is observed that SFFS gives better emotional recognition performance results than with SFS. It is also observed that RDA and SVM performances considerably better when compared with the remaining classifiers.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Berlin</th>
<th>Spanish</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SFS (%)</td>
<td>SFFS (%)</td>
</tr>
<tr>
<td>LDA</td>
<td>65</td>
<td>72.5</td>
</tr>
<tr>
<td>RDA</td>
<td>85.2</td>
<td>92.6</td>
</tr>
<tr>
<td>SVM</td>
<td>80.4</td>
<td>88.1</td>
</tr>
<tr>
<td>kNN</td>
<td>73.8</td>
<td>81.1</td>
</tr>
</tbody>
</table>

Table 8.3: Emotion recognition percentage accuracy of various classifiers (LDA, RDA, SVM and kNN) over Berlin and Spanish databases using SFS and SFFS feature selection methods.
The overall recognition performance of an RDA classifier with SFS and SFFS when compared with Feature Fusion Set (FFS) is shown in Fig. 8.4. The horizontal axis represents the name of the feature selection method and the vertical axis represents the performance of the classifier. From the graph it is observed that the performance of the classifier is effectively improved by 20% approximately with SFFS when compared with baseline results. This proves that two stage feature selection is a better technique, for improving the performance of the classifier.

![Comparison of emotion recognition performances of feature subset selection algorithms using Berlin and Spanish databases](image)

Figure 8.4: Comparison of emotion recognition performances of feature subset selection algorithms using Berlin and Spanish databases

### 8.4.4.1 Analysis of results with each emotion using various classifiers

The emotions and the classifiers considered here are happy, neutral, anger, sad, fear, disgust and LDA, RDA, SVM and kNN. The results are analysed with each emotion using various classifiers individually with both Berlin and Spanish databases. The recognition accuracy of each classifiers on each emotion using SFS and SFFS are shown in Tables 8.4 and 8.5. The left column of both the tables represent the name of the classifier and the title of the row represents the name of the feature selection method and the name of the emotion. Each cell represents the emotion
recognition accuracy of the corresponding classifier. Even though LDA suffers with singularity problem (i.e the number of speech samples are less than that of the dimension of the feature set), it’s performance is not so poor and its recognition accuracy nearly reaches to 70%. This problem is eliminated by RDA by using a regularization technique by which it can reach to an accuracy of 90%. In a similar way the next highest emotion recognition accuracy is obtained with SVM. kNN also gives a very good emotion recognition performance with an accuracy of nearly 80%.

It is also observed from the tables 8.4 and 8.5 that SFFS shows the higher recognition accuracy than that of the SFS because it suffers from nesting problem. i.e once the feature is recovered, it can’t be eliminated. This nesting problem is overcome with SFFS by removing the unwanted features thereby it leads to an effective improvement in the performance of the classifier.

\[
\begin{array}{|c|c|c|c|c|c|c|}
\hline
\text{Algorithm} & \text{H} & \text{N} & \text{A} & \text{S} & \text{F} & \text{D} \\
\hline
\text{LDA} & 65 & 63 & 72 & 63 & 65 & 62 \\
\hline
\text{RDA} & 87 & 80 & 88 & 92 & 81 & 83 \\
\hline
\text{SVM} & 82 & 79 & 83 & 78 & 80 & 88 \\
\hline
\text{kNN} & 77 & 84 & 67 & 77 & 73 & 67 \\
\hline
\end{array}
\]

(a)

\[
\begin{array}{|c|c|c|c|c|c|c|}
\hline
\text{Algorithm} & \text{H} & \text{N} & \text{A} & \text{S} & \text{F} & \text{D} \\
\hline
\text{LDA} & 72 & 82 & 67 & 75 & 71 & 68 \\
\hline
\text{RDA} & 95 & 93 & 96 & 92 & 89 & 91 \\
\hline
\text{SVM} & 90 & 88 & 91 & 90 & 84 & 86 \\
\hline
\text{kNN} & 90 & 87 & 83 & 70 & 80 & 77 \\
\hline
\end{array}
\]

(b)

Table 8.4: Recognition accuracy percentage for emotions (H-happy, N-neutral, A-anger, S-sad, F-Fear and D-disgust) with various classifiers in Berlin database using feature selection algorithms (a) SFS-Sequential Forward Selection (b) SFFS-Sequential Floating Forward selection.

We obtain an optimal feature set with SFS and SFFS which are chosen
CHAPTER 8. PROPOSED TWO STAGE METHOD FOR EMOTION CLASSIFICATION

Table 8.5: Recognition accuracy percentage for emotions (H-happy, N-neutral, A-anger, S-sad, F-Fear and D-disgust) with various classifiers in Spanish database using feature selection algorithms (a) SFS-Sequential Forward Selection (b) SFFS-Sequential Floating Forward selection.

<table>
<thead>
<tr>
<th>Spanish Algorithm</th>
<th>SFS</th>
<th>SFFS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>H</td>
<td>N</td>
</tr>
<tr>
<td>LDA</td>
<td>66</td>
<td>60</td>
</tr>
<tr>
<td>RDA</td>
<td>85</td>
<td>80</td>
</tr>
<tr>
<td>SVM</td>
<td>80</td>
<td>75</td>
</tr>
<tr>
<td>kNN</td>
<td>70</td>
<td>70</td>
</tr>
</tbody>
</table>

Table 8.6: Optimal feature subset extracted by using SFS and SFFS for both the databases

<table>
<thead>
<tr>
<th>SFFS</th>
<th>Best Feature Combination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Berlin(12 features)</td>
<td>12 23 50 84 30 326 34 281 17 156 189 4</td>
</tr>
<tr>
<td>Spanish(18 features)</td>
<td>21 9 78 92 58 24 43 22 98 301 252 120 301 15 260 19 4 56</td>
</tr>
</tbody>
</table>

Table 8.6: Optimal feature subset extracted by using SFS and SFFS for both the databases

The graphical representation of efficiency of each classifier with each and every feature selection technique is shown in Fig. 8.5.
bars represent the results with Full Feature Set (FFS), Sequential Forward Selection (SFS) and Sequential Floating Forward Selection (SFFS) respectively. Among all the bars, the green bars belong to the classifiers RDA and SVM shows the efficient performance by using SFFS in both the databases.

![Bar Graphs](image)

Figure 8.5: Comparison of emotion recognition performances of various feature subset selection algorithms using (a) Berlin database (b) Spanish database

### 8.4.4.2 Analysis of results with Receiver Operating Characteristic (ROC) Curves

The Experimental results are analysed with Receiver Operating Characteristic Curves. The results, thus obtained with these ROC curves show that, the performance of the classifier with feature subset selection techniques is approximately equal to that of the results obtained using ROC Curves. The accuracy, sensitivity, specificity and Area Under Curve are the elements extracted from the ROC curves. The detailed analysis of these ROC curves are given in [33] [60]. The shape of the curve determines the power of the feature selection method. The extent of ‘Area Under Curve’ should lie between 0 and 1. The more the extent of the ‘Area Under Curve’, the more the performance of the method. The ROC curves are drawn for each feature selection method including both the databases for all the classifiers. Here we mention a set of ROC curves which are drawn for the classifier RDA and the corresponding graphs are shown in Fig. 8.6. The ‘Area Under Curve’ is
more for both the databases in all the classifiers for Sequential Floating Forward Selection algorithm.

![Graphs](attachment:image.png)

Figure 8.6: Comparison of emotion recognition performances of various feature subset selection algorithms using (a) Berlin database (b) Spanish database

<table>
<thead>
<tr>
<th>Area under curve</th>
<th>Diagnostic accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9-1.0</td>
<td>excellent</td>
</tr>
<tr>
<td>0.8-0.9</td>
<td>Very good</td>
</tr>
<tr>
<td>0.7-0.8</td>
<td>Good</td>
</tr>
<tr>
<td>0.6-0.7</td>
<td>Sufficient</td>
</tr>
<tr>
<td>0.5-0.6</td>
<td>Bad</td>
</tr>
<tr>
<td>&lt;0.5</td>
<td>Test not useful</td>
</tr>
</tbody>
</table>

Table 8.7: The relationship between area under curve and its diagnostic accuracy. Table 2 in [120]

From the literature survey [120] the ‘Area Under Curve’ and its diagnostic accuracy are shown in Table 8.7. The values extracted from the ROC curves are given in Table 8.8. The ‘Area Under Curve’ in SFFS is 0.937 for Berlin database and 0.908 for Spanish database which shows that the performance of the classifier is excellent, if we use SFFS as feature selection algorithm.
### 8.5 Comparison of results

Based on the literature survey [122], ‘Four Stage feature selection to recognize emotion from speech signals’ is existed. In this method in the first stage, whole feature set is divided into groups in such a way that the dimension of the each group is small. In further stages these features are regrouped to reduce the number of groups and increase the feature dimension. The disadvantage with this method is grouping and regrouping and the number of stages is not fixed.

So in this work ‘Two stage feature selection method’ is proposed which is very new technique and is not proposed in any other work before. It includes the combination of ‘feature fusion’ and ‘optimal feature selection method’. Berlin database is common in ‘Four stage feature selection’ and ‘Two stage feature selection’ in which the accuracy of the system is 81% in Four Stage feature selection and 92.7% in our proposed ‘Two stage feature selection’.

<table>
<thead>
<tr>
<th>Berlin</th>
<th>Accu(%)</th>
<th>Sens(%)</th>
<th>Spec(%)</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>FFS</td>
<td>73.8</td>
<td>74.2</td>
<td>73.4</td>
<td>0.734</td>
</tr>
<tr>
<td>SFS</td>
<td>85.2</td>
<td>85.0</td>
<td>85.3</td>
<td>0.845</td>
</tr>
<tr>
<td>SFFS</td>
<td>92.7</td>
<td>94.4</td>
<td>91.0</td>
<td>0.937</td>
</tr>
</tbody>
</table>

(a)

<table>
<thead>
<tr>
<th>Spanish</th>
<th>Accu(%)</th>
<th>Sens(%)</th>
<th>Spec(%)</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>FFS</td>
<td>69.5</td>
<td>70.1</td>
<td>68.9</td>
<td>0.692</td>
</tr>
<tr>
<td>SFS</td>
<td>83.7</td>
<td>83.9</td>
<td>83.4</td>
<td>0.831</td>
</tr>
<tr>
<td>SFFS</td>
<td>90.5</td>
<td>90.1</td>
<td>89.8</td>
<td>0.908</td>
</tr>
</tbody>
</table>

(b)

Table 8.8: Shows the values (Accu: Accuracy, Sens:Sensitivity, Spec:Specificity, AUC: Area Under Curve) extracted from ROC plot for different feature subset selection algorithms for (a) Berlin and (b) Spanish
8.6 Summary

The principal objective of the proposed ‘Two Stage Feature Selection’ method is to reduce the dimension of the fused feature vector and improve the performance of the classifier. The high dimensional fused feature vector contains some irrelevant features, having a very less emotion specific information. The two stage feature selection method overcomes such type of features and generates a new feature vector with better features, which are less in number and possessing more emotion specific content. The emotion recognition accuracy of the classifier is effectively improved with this optimal feature vector. Various experiments are conducted over emotional speech samples of Berlin and Spanish databases and are systematically evaluated by using various feature selection techniques and several classification methods. The experimental results infer that the classifiers RDA and SVM with SFFS yield best emotion recognition performance, besides producing improvised recognition accuracy in kNN and LDA classifiers when compared with base line results. The results also show that the recognition accuracy is improved by 15%-20% approximately with each classifier and they also reveal that SFFS is a better choice as a feature subset selection method as it eliminates the nesting problem that occurred in Sequential Forward Selection algorithm.