A Hybrid Feature Selection Model for Emotion Recognition using Shuffled Frog Leaping Algorithm (SFLA)-Incremental Wrapper- Based Subset Feature Selection (IWSS)

Sri Raman Kothuri*

Assistant Professor, Department of Computer Science and Engineering, Vel Tech Ranagrajan Dr Sagunthala R&D Institute of Science and Technology,
Avadi, Chennai, 600062, India
sriramankothuri@veltech.edu.in

Dr N R Rajalakshmi

Professor, Department of Computer Science and Engineering, Vel Tech Ranagrajan Dr Sagunthala R&D Institute of Science and Technology,
Avadi, Chennai, 600062, India
drnrrajalakshmi@veltech.edu.in

Abstract

Emotion recognition method is required for therapy to recognize the emotions of patient and helps in treatment. Many computer science based emotion recognition works focused on facial expression, speech, body gesture and multi-modal based machine learning techniques. Existing methods have limitations of poor convergence and easily trap into local optima. In this research, the Shuffled Frog Leaping Algorithm (SFLA)- Incremental Wrapper-based Subset Selection (IWSS) hybrid method is proposed to improve the emotion recognition. The proposed method involves in analysis the emotion of user through video, audio, and text features and recommends the music to the users. The analysis shows that hybrid modality shows the higher performance in emotion recognition. AlexNet model is applied for the feature extraction in video data and Latent Dirichlet Allocation (LDA) is applied for text feature extraction. Multi-Class Support Vector Machine (MC-SVM) model is used for the classification. The proposed SFLA-IWSS method has 97.05 % accuracy and existing gSpan method has 90 % accuracy.

Keywords: AlexNet; Incremental Wrapper-based Subset Selection; Latent Dirichlet Allocation; Multi-Class Support Vector Machine; Shuffled Frog Leaping Algorithm.

1. Introduction

In the field of human-computer interaction and artificial intelligence, emotion recognition plays a promising role. Various techniques like heartbeat, blood pressure, body movements, speech recognition, facial expressions and textual information were used to detect emotions of the users (Batbaatar et al., 2019). Individual's mental state related with behavior, feelings, thoughts are often defined as an emotion. Emotion recognition is one of the popular research in Artificial Intelligent and its ability to mine opinions in social media data such as Twitter, Reddit, YouTube, and Facebook, and others (Poria et al., 2019). Speech is considered as natural way to express ourselves and this is used for emotion recognition. Text is used to way of communication in emails, messages and this is used to recognize the importance of the emotion. Speech Emotion Recognition (SER) is often used for the emotion recognition [Akcay et al. (2020)]. Emotion recognition embedded in a healthcare system to monitor the patient physical and mental state and prescribe suitable medicine or therapy [Hossain et al. (2019)]. Another important module in emotion recognition is facial expression. Facial expression in video is applied to extract the facial features for emotion recognition [Jain et al. (2019)].

Multi-modal emotion recognition is interesting field of research for effective performance of sentiment analysis and computing process. Emotion recognition system is more accurate for different nature of signal carried out for exploiting the information (Nemati et al., 2019). Existing methods treated the features at each time step as

independent samples and ignored emotions property of temporal dependency (Tang et al., 2017). Recent, deep learning models are applied with great success for emotion detection. Convolution Neural Network (CNN) model based features are extracted from visual and audio signal for emotion recognition (Tzirakis et al., 2017). Effective fusion of multimodal representation is required in video and audio domains and existing methods have lower efficiency in challenging task [Sajjad et al., 2020 & Kwon.S., 2020]. In this research, the SFLA-IWSS is proposed to improve the performance of the emotion recognition. The YouTube and SAVEE datasets are used to test the performance of the proposed method. These results shows higher performance compared to existing works in emotion recognition.

The paper is formulated as section 2 has review of recent methods in emotion recognition, the proposed SFLA-IWSS method explanation is given in section 3, simulation setup is given in section 4, results is in section 5 and conclusion is in section 6.

2. Literature Review

Users share their feelings in social networks such as YouTube, Facebook, Twitter, etc in picture, post and videotapes. Sentiment analysis on video or audio data helps to recognize the emotion of users. Some of the recent researches in sentiment analysis in audio or video data were reviewed in this section.

A method of audio-video-textual based multimodal sentiment analysis is a feature level fusion method that is applied in extracted feature from different modalities. Oppositional Grass Bee Optimization (OGBO) method is applied to select the optimal features to train the classifier. The 12 benchmark functions are applied to validate the effectiveness and numerical efficiency of the developed method. The selected features are applied to Multi-Layer Perceptron (MLP) for sentiment classifiers. The developed OGBO method has higher performance in classifier compared to existing methods. The convergence of OGBO method is low and this feature selection efficiency is less (Bairavel et al., 2020).

Works with applying two modalities of facial expression and affective speech for multimodal emotional recognition system had been done with the common low-level descriptors of spectral audio features and prosodic are extracted for affective speech. Temporal variation of each landmark time series is applied individually for extracting primary visual features. Discrete Wavelet Transform (DWT) is applied to analysis the signal. A variety of dimensionality reduction scheme are applied to reduces the complexity of derived model and improves the efficiency. Several audio feature level fusion and proposed visual features are applied to exploit the advantages of multimodal emotional recognition. The eNTERFACE05, RML and SAVEE datasets were used to test the performance of the developed method (Rahdari,et al. 2019).

Another work is with hybrid fuzzy evolutionary computation to perform dimensional reduction and learning features. This work had included the time, frequency, and time-frequency of multiple features are extracted from the EEG signal. The hybrid fuzzy c means, genetic algorithm and neural network is applied for classification of unimodal data of either EEG or speech. A separate model is used for each modality and integrate with posterior probabilities. The developed method has higher performance compared to existing methods in emotional recognition. The SAVEE and MAHNOB datasets were used to test the performance of developed method. The genetic algorithm method has easily trap into local optima and has lower performance in feature selection (Ghoniem et al., 2019).

The other work representing the face region as graph with nodes and edges for facial emotional recognition. The gSpan frequent sub-graphs mining method is applied to find the frequent sub-section in the graph of each emotions. Once the final sub-graph is encoded, input facial queries of six level classification is applied for binary classification. Binary cat swarm intelligent method is applied to each level of classification to select proper sub-graph for improving performance. The result shows that developed method has higher performance in classification than existing methods (Hassan et al., 2020).

Using Tunable Q wavelet transform (TQWT), twine shuffle pattern, and discriminative features had applied in another work for emotion recognition. The TQWT is multi-level wavelet transform that is used to generate low-level, medium level and high level transform. The four publicly available datasets were used to test the performance of TQWT method. The TQWT method has higher performance in speech emotion recognition than existing method. The feature efficiency of the model is low and deep learning method is required for efficient feature analysis (Tuncer et al., 2021).

3. Proposed Method

In this research, the SFLA-IWSS method is proposed to improve the performance of emotion recognition. The Video, audio and text features were used for emotion recognition in proposed method. AlexNet model is applied to extract video features and LDA is applied for text feature extraction. The 12 feature extraction is applied for audio analysis and improves the performance. The overall block diagram of the SFLA-IWSS method is shown in Figure 1.

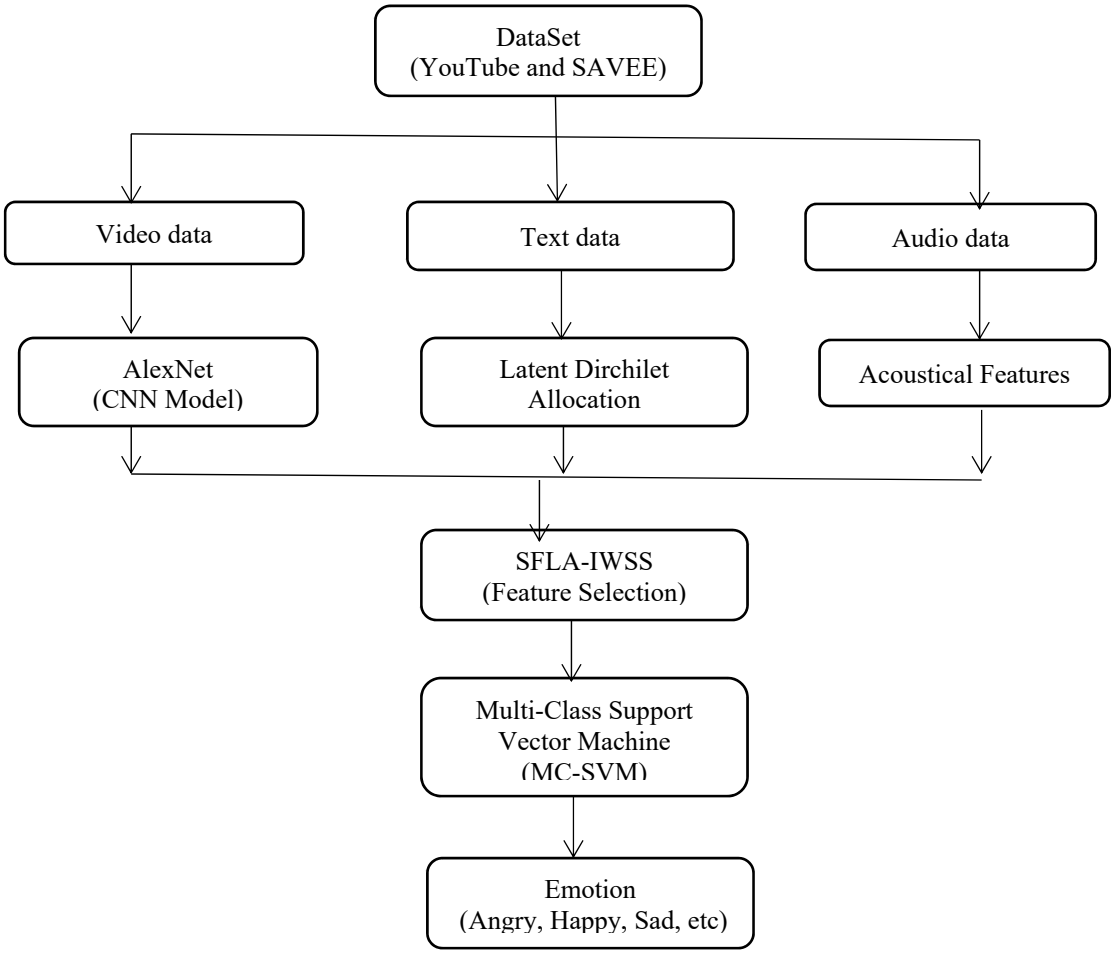

Figure 1. The block diagram of SFLA-IWSS method

3.1 Feature extraction in audio signal

For the audio signal, 12 features are fixed and worked prominently and they are mean duration to pauses, pause-to-word ratio, total duration of speech, long pause count(raw), short pause count(raw), skewness, kurtosis, zero crossing rate, mean, variance, auto correlation Linear Prediction Coefficients (LPC), and Cross correlation (XCORR).

3.2 AlexNet – Convolutional Neural Network

Image classification performance is superior to other tradition CNN models. The researchers make it efficient in deep learning model. AlexNet is large network structure consists of 650,000 neurons and 60 million parameters.

Activation function is first improvement and activation function is applied in neural networks to analysis non-linearity (Han et al., 2017 & Lu et al., 2019). Traditional activation functions are tanh, logistic function etc. These functions are tending to run in vanishing gradient problem and gradient is a large value for input is around

a small range of 0. In order to overcome this problem, Rectified Linear Unit (ReLU) was used. The equation of ReLU is given in equation 1.

$$ReLU(x) = \max(x, 0) \quad (1)$$

The ReLU gradient is one if input is not less than 0. The deep networks of ReLU activation function has faster converge than tanh unit. This acceleration is highly improved in training process.

Secondly, dropout was applied to eliminate the overfitting problem. This is usually applied in fully connected layers and in every iteration, only a part of neurons in dropout was trained. For instance, if ratio is set as 50 %, only half of parameters is trained in every iteration. Dropout forces a neuron to cooperate with others to reduces neurons in joint adaptation and improves the generalization.

For automatic feature reduction and extraction, convolution and pooling layers were applied. Signal analysis is used in convolution technique. Consider an image M in size of (m, n), the convolution is expressed in *C(m,n).*

Before applying in next layers, feature maps are normalized. Several adjacent maps in same position generates sum of normalization of cross channel. The real neurons are used for this process.

Classification is performed in fully connected layer. Adjacent fully connected layers of neurons are directly linked. Softmax is applied in layers for activation function. Softmax output is in range of $(0,1)$ that ensures activation of neurons.

3.3 Latent dirichlet allocation

Latent Dirichlet allocation (LDA) is a hierarchical Bayesian model that tries to map a text document into a latent low dimensional space spanned by a set of automatically learned topical bases. The model assumes that a document consists K topics, and the generative probability distribution for all the documents is given in Equation 2. (Jelodar et al., 2019).

$$p(v, z, \theta|\alpha, \pi) = p(\theta|\alpha) \prod_{m=1}^{M} p(z_m|\theta) p(v_m|z_m, \pi) \quad (2)$$

where $\theta \sim Dirichlet(\alpha)$, which is the topic proportion of a K-dimensional vector and $\sum_{k=1}^{K} \theta_k = 1$. z_m is a K-dimensional indicator vector (only one element is 1, and all others are 0) referring to topic label of the mth word; π is a matrix consisting of K rows for those K topics, with each row representing a multinomial distribution over words in a given vocabulary. The graphical model representation of LDA model is depicted. The model is often used to find latent high-level features, i.e., topics of a document.

3.4 Hybrid Feature Selection Method

In this proposed method, as a first step, it has been used the SFLA method (Jelodar et al., 2019) to find optimal response and further extend the optimality over Symmetrical Uncertainty(SU) in proposed IWSS method as explained in section 3.4.2. This IWSS method evaluate for *BestData* subsets by wrapping up incrementally, while selection of frogs in the former SFLA method.

3.4.1 Shuffled Frog Leaping Algorithm

The SFLA is population based metaheuristic optimization method that mimic a group of frog evolution when looking for a place with more food available. The random and definite strategies are applied in SFLA method to find optimal response (Lu et al., 2015 & Abiodun et al., 2021).

Frogs primitive population is denoted as $sfla_p$ and this is randomly generated from possible answers. Frog situation or position is possible solution to the problem. Vectors are used to represent the frogs and variables or problem solutions are used to structures the solution. Initial population is partition into $sfla_m$ groups called memplex. Memplexes have bunch of frogs $sfla_n$ that individually search for a solution in search space. A sub-memplex is applied in each memplex to avoid falling in local optima. Each sub-memplex has $sfla_q$ frogs and randomly selected based on probability function, as given in equation 3.

$$P_j = \frac{2(sfla_n + 1 - j)}{sfla_n(sfla_n + a)}, \quad j = 1,2, \ldots, sfla_n \quad (3)$$

Where probability of selecting j^{th} frog is P_j and memplex has number of frogs of $sfla_n$. According to a descending fitness order and decreasing the fitness value, frogs in each memplex are sorted and lowered the probability of selecting frogs. In sub-memplex, search space of a better positioned frog has greater chance of choosing as a member. The worst frog (P_w) in each sub-memplex, performs leaping based on position and experiences of best frog in memplex (P_b). The worst frog is selected from sub-memplex and the leaping step size of frog P_w is denoted as in equation 4.

$$S_B = \begin{cases} \min\{int(rand.[P_b - P_w]).S_{max}\} & for\ a\ positive\ step \\ \max\{int(rand.[P_b - P_w].-S_{max})\} & for\ a\ negative\ step \end{cases} \quad (4)$$

Where $rand$ denotes random number is in range of [0, 1]. And maximum leap length is denoted as S_{max}. The worst frog position in next step is in equation 5.

$$P'_w = P_w + S_B \quad (5)$$

If new frog (P'_w) is better than original frog, this frog replaces the original frog, otherwise edited P_w based on best frog in total population (P_G), as in equations (6 & 7).

$$S_G = \begin{cases} \min\{int(rand.[P_G - P_w].S_{max})\} & for\ a\ positive\ step \\ \max\{int(rand.[P_G - P_w].-S_{max})\} & for\ a\ negative\ step \end{cases} \quad (6)$$

$$P''_w = P_w + S_G \quad (7)$$

If P''_w is better than original frog (P_w), this frog replaces P''_w frog and new random frog is replace the worst frog of sub-memplex. After these steps in dividing memplex in sub-memplexes, all frogs are combined and re-divided into $sfla_m$ memplexes. This process is continuing until the end of iteration. The worst frog is leap towards the best frog. Average fitness of frog gradually increases by repeating the process in evolution step and converge with certain degree. Based on this process, P_G and P_w are changed in each iteration and fitness value is converge to desired response.

3.4.2 Incremental Wrapper-based Subset Selection

Ranking is computed in step 1 to 4 and filter evolution of $O(n)$ is required in this stage. Predictive attributes are used to evaluate the Symmetrical Uncertainty (SU) and SU is a non-linear information to interpreted mutual information normalized in interval of [0, 1], as in equation 8.

$$SU(A_i, C) = 2\left(\frac{H(C) - H(C|A_i)}{H(C) + H(A_i)}\right), \quad (8)$$

The class is denoted as C and Shannon entropy is denoted as $H()$. In increasing order of SU, attributes are ranked and more informative attributes are applied first.

The initialization of S is carried out in steps 5 and 6 based on ranking first variable. The data from evaluating subset stored in $BestData$. The function $evaluate(C, Saux, T)$ learns and validates the classifier C based on 5 fold cross-validation in training set T over subset $S \cup \{C\}$. Thus, $BestData$ will contain an array $BestData.f[1..5]$ with the accuracy obtained for each fold and a real value $BestData.av$ with the averaged accuracy over the 5 folds. The Pseudo code of IWSS method is given below.

Algorithm: *Incremental Wrapper-based Subset Selection*

1. *List* $R = \{\}$
2. *For each attribute* $A_i \in T$
   a. $Score = M_T\ (A_i, class)$
   b. *Insert* A_i *in R according to* $Score$
3. $S = \{R[1]\}$
4. $BestData = evaluate(C, S_{aux}, T)$
5. *For* $i = 2$ *to* n
   a. $S_{aux} = S \cup \{R[i]\}$
   b. $AuxData = evaluate(C, S_{aux}, T)$
   c. *If* $(AuxData \rightarrow BestData)$
      i. $S = S_{aux}$
   d. $BestData = AuxData$

3.5 Multi-Class Support Vector Machine

Binary classifiers $f_1, f_2, \ldots f_N$ is constructed for $1 \ldots N$ classes, each trained to be different from one class to the others (Kaur et al., 2019). Multi-class category is obtained based on the maximal output before applying the activation (sgn) function.

Where $argmax\ g^k(x)$

Where $g^k(x) = \sum_{i=1}^{n} y_i \alpha_i^k k(x, x_i) + b^k$

Where $k = 1, \ldots, N$

Where hyper plane distance to the point x of a signed real value is denoted as $g^k(x)$ which is referred as the confidence value. The higher value increases the confidence where x belongs to positive class. The highest confidence value is assigned with x.

The input data is denoted as $X = \{x_1, x_2, \dots, x_m\} \in R^d$, the hypersphere radius is denoted as r, and the centre is denoted as $c \in R^d$. The minimum hypersphere which encloses the optimization problem is given in equation 9.

Minimize r^2

Subject to $\left|\left|\Phi(x_j) - c\right|\right|^2 \leq r^2,\ j = 1, \dots, m$

$$L(c, r, \alpha) = r^2 + \sum_{j=1}^{m} \alpha_j \{\left|\left|\Phi(x_i) - c\right|\right|^2 - r^2\} \quad (9)$$

Derive $\frac{\partial L(c,r,\alpha)}{\partial c} = 2\sum_{j=1}^{n} \alpha_j (\Phi(x_j) - c) = 0$

Equation 10 is obtained.

$$\sum_{j=1}^{m} \alpha_j = 1 \text{ and } \sum_{j=1}^{m} \alpha_j \Phi(x_j) \quad (10)$$

Hence, the equation 9 becomes equation (11).

$$L(c, \gamma, \alpha) = \sum_{j=1}^{m} \alpha_j k(x_j, x_j) - \sum_{i,j=1}^{m} \alpha_i \alpha_j k(x_i, x_j) \quad (11)$$

The optimization problem is solved based on dual form of α, as given in equation 12.

Maximizing,

$$W(\alpha) = \sum_{i=1}^{m} \alpha_i k(x_i, x_i) - \sum_{i,j=1}^{m} \alpha_i \alpha_j k(x_i, x_j) \quad (12)$$

Subject to $\sum_{i=1}^{m} \alpha_i = 1$ and $\alpha_i \geq 0$, $i = 1tom$.

Lagrange multiplier possibilities of non-zero if the inequality constraints are a solution equality.

Optimal solutions complementarity conditions for $\alpha, (c, \gamma)$ is given in equation 13.

$$\alpha_i \left\{\left|\left|\Phi(x_i) - c\right|\right|^2 - r^2\right\},\ i = 1, \dots, m \quad (13)$$

Training samples x_i lie on the surface of the optimal hypersphere related to $\alpha_i > 0$.

Equation 14 provides the decision function solution.

$$f(x) = sgn\left(r^2 - \left|\left|\Phi(x) - c\right|\right|^2\right) \quad (14)$$

An equation 15 and 16 is provided.

$$f(x) = sgn(r^2 - \{\Phi(x).\Phi(x) - 2\sum_{i=1}^{m} \alpha_i \Phi(x).\Phi(x_i) + \sum_{i,j=1}^{m} \alpha_i \alpha_j (\Phi(x_i).\Phi(x_j)\}) \quad (15)$$

$$f(x) = sgn(r^2 - \{k(x, x) - 2\sum_{i=1}^{m} \alpha_i k(x, x_i) + \sum_{i,j=1}^{m} \alpha_i \alpha_j k(x_i, x_j)\}) \quad (16)$$

The method aims to obtain minimum enclosing hyper sphere consists of satisfy all training samples.

4. Simulation Setup

The hybrid feature selection with deep learning feature extraction method is proposed for emotional recognition and music recommendation. The datasets, system requirement, Metrics used, and parameter settings were discussed in this section.

Datasets: The SAVEE dataset (Jackson et al., 2014) has recordings of four male speakers (identified as DC, JE, JK, KL), researcher and postgraduate at University of Surrey aged from 27 to 31 years. Discrete categories of emotions of surprise, sadness, happiness, fear, disgust and anger are present in dataset. There are 7 emotions in the dataset with neutral. The audio files are present in 'wav' format and sampled at 44.1 kHz.

System Configuration: The proposed SFLA-IWSS method is tested on Intel i7 processor, 6 GB graphics card, and 16 GB RAM. The MATLAB R2018b tool was used to implement and test the performance of SFLA-IWSS method in emotion recognition.

Metrics: Metrics such as Accuracy, Sensitivity, Specificity, PPV and NPV were used to measure the performance of proposed SFLA-IWSS method in emotion recognition.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request. Further, the MATLAB code of the new model/method used in this analysis will be made available through https://in.mathworks.com/matlabcentral/profile/authors/11633819 once the research article is published.

5. Results

Emotion recognition is required in therapy to treat the patients with stress and negative emotions. Emotion recognition with music recommendation helps to lighten up the mood of the users. Emotion recognition is challenging task and some researchers were carried out emotional recognition based on machine learning

models. Hybrid modality has better performance in emotion recognition compared to single modality. The YouTube and SAVEE datasets were used to test the performance of the proposed SFLA-IWSS method.

Data Types	Accuracy	Sensitivity	Specificity	PPV	NPV
Text	90.48	90.64	82.14	84.89	78.18
Audio	90.95	89.64	81.43	88.24	70.02
Video	95.43	92.21	96.43	95.42	94.78
Video + Text	94.05	94.64	96.43	93.33	89.87
Audio + Text	86.90	93.43	78.57	84.44	78.26
Video + Audio	95.24	94.64	98.10	94.94	92.50
Hybrid features (Text, Audio, Video)	**96.10**	**95.11**	**97.97**	**95.76**	**96.83**

Table 1. Proposed SFLA-IWSS method on various data types in YouTube dataset

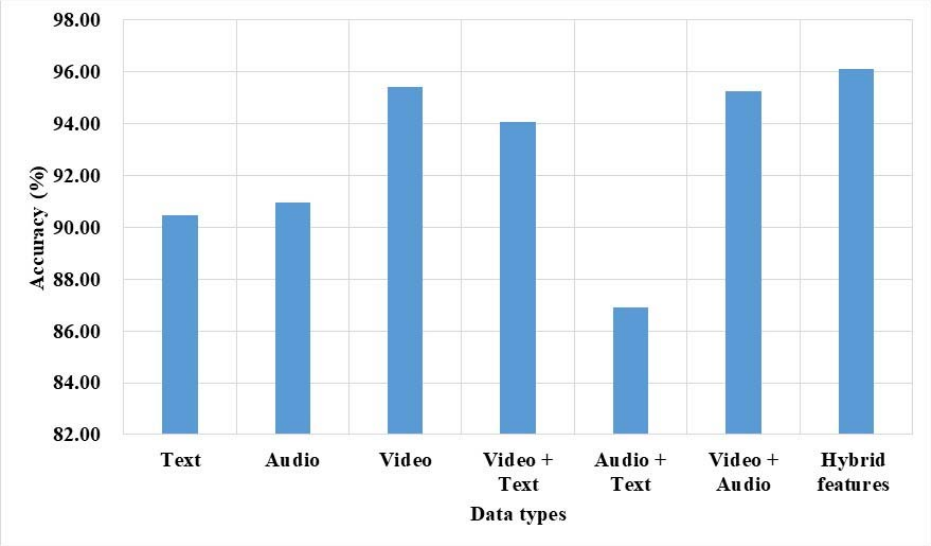

Figure 2. Proposed SFLA-IWSS method on various data types in YouTube dataset

The proposed SFLA-IWSS method is tested on various data types such as text, audio, video and combinations of data types in YouTube dataset, as shown in Figure 2 and Table 1. This shows that hybrid modality has higher performance due to many relevant features for recognition. Video features has second higher performance due to extraction of deep learning features from facial features. Video with other types of audio and text has higher performance and as it seems, text features degrades the video performance. Audio and text features combination has lower performance in the recognition due to irrelevant analysis in text features.

Data types	Accuracy	Sensitivity	Specificity	PPV	NPV
Text	92.62	90.28	88.33	90.09	88.66
Audio	95.83	92.92	94.30	94.44	93.78
Video	94.43	95.10	94.70	95.39	96.28
Video + Text	95.24	94.31	91.67	93.66	92.68
Audio + Text	92.31	99.31	79.17	86.36	84.79
Video + Audio	93.45	93.61	95.83	93.98	93.19
Hybrid features (Text, Audio, Video)	**96.40**	**95.70**	**95.45**	**95.68**	**96.50**

Table 2. Proposed SFLA-IWSS method in SAVEE dataset

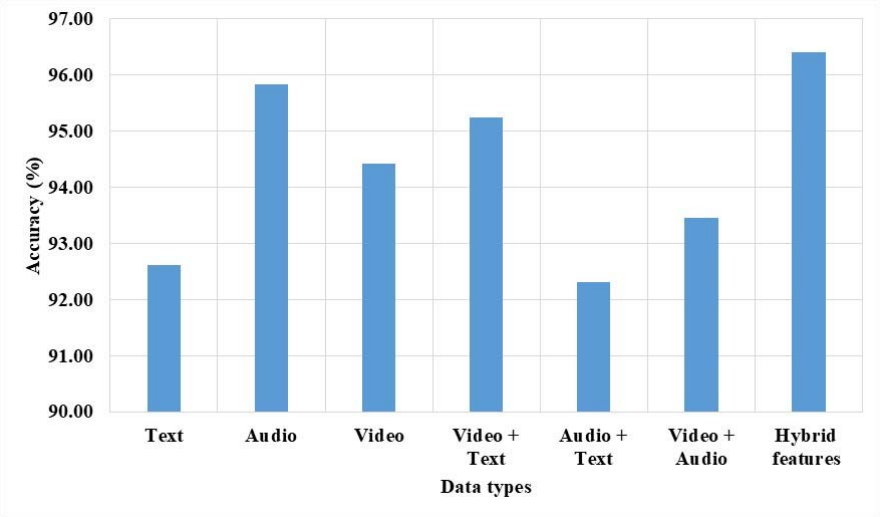

Figure 3. Accuracy of SFLA-IWSS method in SAVEE dataset

The emotion recognition performance of SFLA-IWSS method for various data types in SAVEE dataset, as shown in Figure 3 and Table 2. The SFLA-IWSS method has higher performance in hybrid features due to its capacity to converge the features. Similar to YouTube dataset, text data has lower performance and combination of text-audio has poor performance in recognition. Video individual feature has second higher performance due to its deep learning feature extraction. Alex Net model has efficient feature extraction from facial features in the video.

Feature Selection Methods	Accuracy	Sensitivity	Specificity	PPV	NPV
Relieff	94.05	93.43	92.86	92.86	89.29
Relieff with SFLA	92.86	92.90	89.29	94.00	92.12
UDFS with SFLA	86.90	92.86	89.29	87.56	81.36
Proposed	**97.81**	**96.21**	**95.57**	**96.28**	**97.43**

Table 3. SFLA-IWSS comparison with feature selection method in YouTube dataset

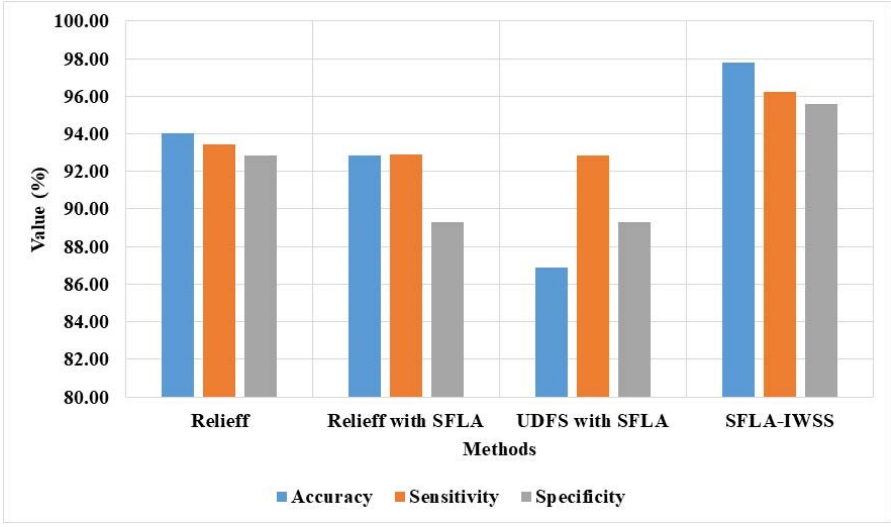

Figure 4. Feature selection Method Comparison in YouTube dataset

The proposed SFLA-IWSS method is compared with feature selection methods of Relief-F, SFLA-Relief-F, and UDFS-SFLA features, as shown in Figure 4 and Table 3. Features selection methods are commonly tested with Multi-Class SVM in emotion recognition. The SFLA-IWSS method has higher performance than existing feature selection in terms of accuracy, sensitivity, specificity, PPV and NPV metrics. The proposed SFLA-IWSS method has advantage of selects the features in high dimension and with high convergence. The Relief-F method selects the features in the filtering manner and not suitable to handle high dimension features.

SFLA-Relief F features has poor convergence that affects the performance of recognition. The SFLA-UDFS method has lower performance in recognition due to its trap into local optima. The SFLA-IWSS method has higher performance due to its escape local optima and maintain high convergence.

Feature Selection Methods	Accuracy	Sensitivity	Specificity	PPV	NPV
Relieff	95.81	99.31	95.67	95.00	96.70
Relieff with SFLA	92.86	91.95	95.83	95.83	97.54
UDFS with SFLA	89.88	99.31	91.67	93.48	92.50
Proposed	**97.05**	**97.86**	**97.25**	**96.89**	**98.59**

Table 4. SFLA-IWSS comparative analysis with feature selection methods in SAVEE dataset

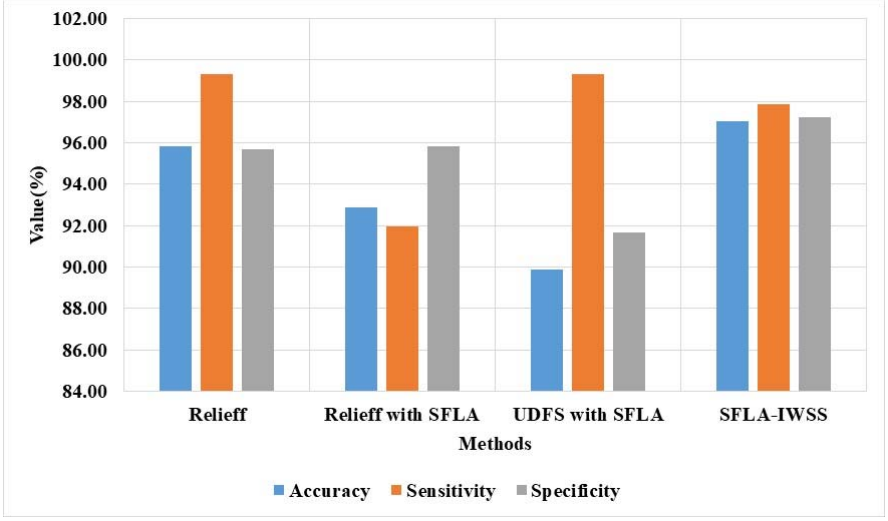

Figure 5. SFLA-IWSS comparative analysis with existing feature selection in SAVEE dataset

The proposed SFLA-IWSS method is tested in SAVEE dataset and compared with feature selection methods in Figure 5 and Table 4. The SFLA-IWSS method has higher performance due to its capacity to handle high dimension features and maintain convergence in feature selection. The Relief-F features has higher sensitivity and lower accuracy due to selection is based on class that also has lower performance in high dimensional data. The UDFS-SFLA method has higher sensitivity and lower accuracy due to this method trap into local optima. The SFLA-Relief F method has lower performance due to its lower convergence. The SFLA-IWSS method easily escape local optimal due to its fitness update in the feature selection method.

5.1 Comparative Analysis

The proposed SFLA-IWSS method is tested with existing methods of emotion recognition, as shown in Table 5.

Method	Dataset	Accuracy (%)
OGBEE [11]	YouTube	95.2
gSpan [14]	SAVEE	90
Twin shuffle [15]	SAVEE	80.05
SFLA-IWSS	YouTube	97.81
	SAVEE	97.05

Table 5. Comparative analysis of proposed method

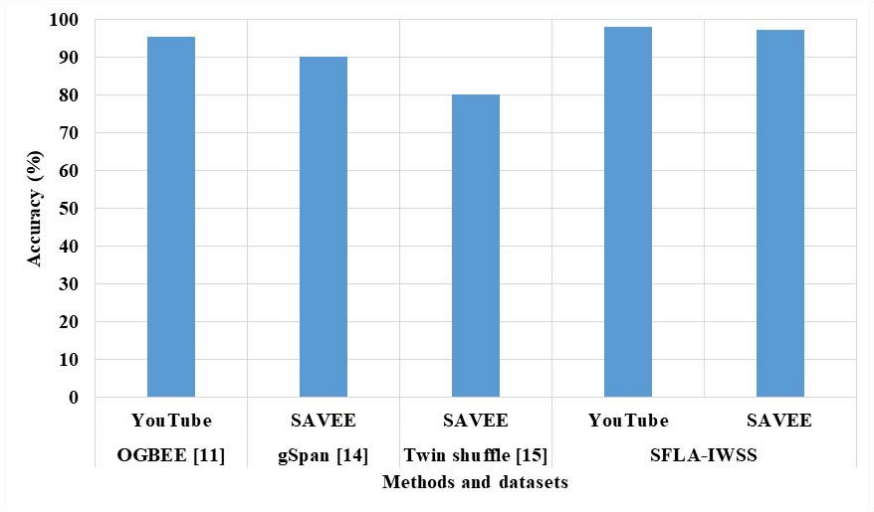

Figure 6. Proposed method comparison

The proposed SFLA-IWSS method is compared with existing methods in emotion recognition in YouTube and SAVEE dataset, as shown in Figure 6 and Table 5. The SFLA-IWSS method has higher performance in emotion recognition compared to existing methods. The SFLA-IWSS method has advantage of good convergence, handle high dimensional data and easily escape from local optima. The OGBEE method has lower convergence and easily trap into local optima. The gSpan method has lower search performance in graph and Twin shuffle method has lower feature efficiency.

6. Conclusion

Music Recommendation based on emotion of user is interesting research and this requires efficient emotion recognition model. Existing methods applies various feature selection method to improve the efficiency of emotion recognition. Existing methods have limitations of poor convergence and easily trap into local optima. In this research, the SFLA-IWSS method is proposed to improve the efficiency of emotion recognition. The SFLA-IWSS method has advantage of effectively handle high dimension features, good convergence and escape from local optima. The YouTube and SAVEE datasets were used to test the performance of SFLA-IWSS method. The SFLA-IWSS method has 97.05 % accuracy in emotion recognition than existing gSpan method has 90 % accuracy. The future direction of the method involves in applying Natural Language Processing (NLP) feature selection method to improve text features efficiency.

7. References

[1] Abiodun, E.O., Alabdulatif, A., Abiodun, O.I., Alawida, M., Alabdulatif, A. and Alkhawaldeh, R.S., 2021. A systematic review of emerging feature selection optimization methods for optimal text classification: the present state and prospective opportunities. Neural Computing and Applications, pp.1-28.
[2] Akçay, M.B. and Oğuz, K., 2020. Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers. Speech Communication, 116:56-76.
[3] Batbaatar, E., Li, M. and Ryu, K.H., 2019. Semantic-emotion neural network for emotion recognition from text. IEEE Access, 7:111866-111878.
[4] Bairavel, S. and Krishnamurthy, M., 2020. Novel OGBEE-based feature selection and feature-level fusion with MLP neural network for social media multimodal sentiment analysis. Soft Computing, 24:18431-18445.
[5] Ghoniem, R.M., Algarni, A.D. and Shaalan, K., 2019. Multi-modal emotion aware system based on fusion of speech and brain information. Information, 10(7):239.
[6] Hassan, A.K. and Mohammed, S.N., 2020. A novel facial emotion recognition scheme based on graph mining. Defence Technology, 16(5):1062-1072.
[7] Han, X., Zhong, Y., Cao, L. and Zhang, L., 2017. Pre-trained alexnet architecture with pyramid pooling and supervision for high spatial resolution remote sensing image scene classification. Remote Sensing, 9(8):848.
[8] Hossain, M.S. and Muhammad, G., 2019. Emotion recognition using deep learning approach from audio visual emotional big data. Information Fusion, 49:69-78.
[9] Jackson P, Haq S, 2014. Surrey audio-visual expressed emotion(SAVEE) database. University of Surrey, Guildford.
[10] Jain, D.K., Shamsolmoali, P. and Sehdev, P., 2019. Extended deep neural network for facial emotion recognition. Pattern Recognition Letters, 120:69-74.
[11] Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y. and Zhao, L., 2019. Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. Multimedia Tools and Applications, 78(11):15169-15211.
[12] Kaur, P., Singh, G. and Kaur, P., 2019. Intellectual detection and validation of automated mammogram breast cancer images by multi-class SVM using deep learning classification. Informatics in Medicine Unlocked, 16:100151.

[13] Kwon, S., 2020. A CNN-assisted enhanced audio signal processing for speech emotion recognition. Sensors, 20(1):183.
[14] Lu K, Ting L, Keming W, Hanbing Z, Makoto T, Bin Y., 2015. An Improved Shuffled Frog-Leaping Algorithm for Flexible Job Shop Scheduling Problem. *Algorithms*. 8(1):19-31. https://doi.org/10.3390/a8010019.
[15] Lu, S., Lu, Z. and Zhang, Y.D., 2019. Pathological brain detection based on AlexNet and transfer learning. Journal of computational science, 30:41-47.
[16] Nemati, S., Rohani, R., Basiri, M.E., Abdar, M., Yen, N.Y. and Makarenkov, V., 2019. A hybrid latent space data fusion method for multimodal emotion recognition. IEEE Access, 7:172948-172964.
[17] Poria, S., Majumder, N., Mihalcea, R. and Hovy, E., 2019. Emotion recognition in conversation: Research challenges, datasets, and recent advances. IEEE Access, 7:100943-100953.
[18] Rahdari, F., Rashedi, E. and Eftekhari, M., 2019. A multimodal emotion recognition system using facial landmark analysis. Iranian Journal of Science and Technology, Transactions of Electrical Engineering, 43(1):171-189.
[19] Sajjad, M. and Kwon, S., 2020. Clustering-based speech emotion recognition by incorporating learned features and deep BiLSTM. IEEE Access, 8:79861-79875.
[20] Tang, H., Liu, W., Zheng, W.L. and Lu, B.L., 2017,. Multimodal emotion recognition using deep neural networks. In International Conference on Neural Information Processing,Springer, Cham:811-819.
[21] Tuncer, T., Dogan, S. and Acharya, U.R., 2021. Automated accurate speech emotion recognition system using twine shuffle pattern and iterative neighborhood component analysis techniques. Knowledge-Based Systems, 211:106547.
[22] Tzirakis, P., Trigeorgis, G., Nicolaou, M.A., Schuller, B.W. and Zafeiriou, S., 2017. End-to-end multimodal emotion recognition using deep neural networks. IEEE Journal of Selected Topics in Signal Processing, 11(8):1301-1309.

Authors Profile

Sri Raman Kothuri He is currently a research scholar and working as Assistant Professor in Vel Tech Rangarajan Dr Sagunthala R&D Institute of Science and Technology, Avadi, Chennai. His research intersts are Artificial Intelligence, Machine Learning, Deep Learning, Music Information Retrieval, Sentimental Analysis and soft computing. He completed B.E., M.Tech and currently pursuing Ph.D. He has 17 years of teaching experience.
https://orcid.org/0000-0002-2103-5283

Dr. N R. Rajalakshmi, Professor in Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, India from 2018. She is having vast teaching experience and her research interest includes Cloud Computing, IoT, Machine Learning, Sentimental Analysis, Deep Learning and Data Science.