Devarakaggalahalli, Harohalli, Kanakapura Road, Dt. Ramanagara, Karnataka 562112.

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

(DATA SCIENCE)

PROJECT REPORT

ON

“Emotion Recognition using Facial and Galvanic Skin Response (GSR) Data”

2024-2025

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE & ENGINEERING (DATA SCIENCE)

Submitted By

Greesha N Malwade -ENG23DS0059

Kanchi Joshitha -ENG23DS0064

Kondasani Sarayu-ENG23DS0066

Under the Supervision of:

Prof. Sindhu A

Assistant Professor

COMPUTER SCIENCE &ENGINEERING (DATA SCIENCE)

DAYANAND SAGAR UNIVERSITY

SCHOOL OF ENGINEERING

Devarakaggalahalli, Harohalli, Kanakapura Road,

Dt. Ramanagara, Karnataka -562112.

CERTIFICATE

It has been carried out at Dayananda Sagar University, Devarakaggalahalli, by **Greesha N Malwade -ENG23DS0059, Kanchi Joshitha -ENG23DS0064, Kondasani Sarayu-ENG23DS0066**, Bonafide students of fourth semester, B.Tech in partial fulfilment for the award of degree in Bachelor Of Technology in Computer Science and Engineering (Data Science) during academic year 2024-25. It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the report deposited in departmental library. The project report has been approved as it satisfies the academic requirements in respect of project work for the said degree.

Signature of the Guide **Signature of the Chairperson**

ACKNOWLEDGEMENT

The completion of project brings with and sense of satisfaction, but it is never completed without thanking the person who are all responsible for its successful completion. We wish to express our profound feelings of gratitude to this great institution of our DAYANANDA SAGAR UNIVERSITY for providing the excellent facilities.

I am especially thankful to our Chairperson, **Dr. Shaila S G**, for providing necessary departmental facilities, moral support and encouragement. The largest measure of our acknowledgment is reserved for **Prof. Sindhu A** whose guidance and support made it possible to complete the project work in a timely manner.

I have received a great deal of guidance and co-operation from the staff and I wish to thank all that have directly or indirectly helped me in the successful completion of this project work.

Greesha N Malwade -ENG23DS0059

Kanchi Joshitha -ENG23DS0064

Kondasani Sarayu-ENG23DS0066

DECLARATION

I hereby declare that the MATLAB Project entitled **"Emotion Recognition using Facial and Galvanic Skin Response (GSR) Data",** submitted to Dayananda Sagar University, Bengaluru, is a Bonafide record of the work carried out by the team under the guidance of Prof. Shivamma D, Assistant Professor – Department of Computer Science – Data Science, School of Engineering, Dayanand Sagar University, and this work is submitted in partial fulfilment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science and Engineering (Data Science).

Greesha N Malwade -ENG23DS0059

Kanchi Joshitha -ENG23DS0064

Kondasani Sarayu-ENG23DS0066

ABSTRACT

This project focuses on improving emotion recognition from facial features and galvanic skin response (GSR) signals using robust data analysis and preprocessing techniques. The dataset, enhanced with synthetic noise to mimic real-world conditions, underwent thorough cleaning, including missing value imputation and outlier handling through statistical methods.

Initial analysis involved exploring feature distributions, label balance, and correlations using visual tools such as histograms, boxplots, and heatmaps. These insights guided the modeling phase, where classifiers like Support Vector Machines, Decision Trees, Random Forests, and Ensemble methods were applied.

The results showed that effective preprocessing significantly boosted model accuracy, with the best classifiers achieving over 80% accuracy. This demonstrates the critical role of early data handling in building reliable emotion recognition systems.

TABLE OF CONTENTS

INTRODUCTION:

Emotion recognition has gained attention in recent years due to its wide applications in healthcare, human-computer interaction, and behavioral analysis. Accurately identifying emotional states from physiological and facial data can enhance personalized technologies and improve user experiences. This project explores the task of emotion recognition by using a dataset that includes facial features and galvanic skin response signals, aiming to build reliable classification models.

The dataset used in this study was intentionally made noisy to simulate real-world conditions where data often contains imperfections. It included 21 distinct emotions, allowing for a broad and nuanced understanding of human affect. Preparing such data before applying machine learning techniques is essential for meaningful outcomes. The project began with inspecting and refining the dataset to address inconsistencies and outliers that could interfere with learning patterns. Statistical summaries and visualizations helped understand the structure of the data and shaped the decisions made during modeling.

Various classification algorithms were then trained and tested to evaluate their effectiveness in recognizing emotional states. The approach highlights how careful data handling paired with strong modeling techniques can lead to accurate and practical emotion recognition systems.

OBJECTIVE:

The main objective of this project is to develop an accurate emotion recognition system using facial and physiological data. To achieve this, the project focuses on preparing and cleaning a noisy dataset to improve the quality of input data for modeling. Different machine learning classifiers are applied to identify 21 distinct emotions from the dataset. The goal is to evaluate the performance of these classifiers and analyse the impact of data preprocessing on the overall accuracy. This study aims to demonstrate how careful data handling and model selection contribute to effective emotion recognition.

SCOPE OF WORK:

This project focuses on developing an emotion recognition system using a dataset with 21 different emotion categories. It includes preprocessing steps such as cleaning noisy data, handling missing values, and managing outliers to improve data quality. Exploratory data analysis is conducted to better understand the features and their relationships. Various machine learning models, including SVM, decision trees, random forest, and ensemble classifiers, are applied and evaluated. Additionally, visualization techniques like principal component analysis and feature importance charts are used to interpret the data and model results. The outcome serves as a foundation for further studies in emotion classification and related applications.

DESCRIPTION OF WORK:

The project starts by loading a dataset that contains 21 different emotions. The initial step involves cleaning the data to handle noise, missing values, and outliers. This process ensures the dataset is of higher quality and suitable for analysis.

Next, exploratory data analysis is conducted to understand the characteristics of the data. Key statistics are calculated, and visualizations are created to show the distribution of features and the relationships between them. This step helps in identifying patterns and potential issues within the data.

After cleaning and analysis, the features are normalized, and the dataset is divided into training and testing subsets. Several machine learning models, including support vector machines, decision trees, random forests, and ensemble methods, are then trained using the cleaned data.

Finally, the models' performance is evaluated based on accuracy and confusion matrices. Additional visualizations like principal component analysis and feature importance plots offer deeper insight into the data and the factors affecting the classification. This methodical approach aims to develop an effective emotion recognition system and lays the foundation for further enhancements.

MODEL DESCRIPTION:

System Overview

The system is designed to recognize 21 distinct emotional states using machine learning techniques. It processes numerical feature data extracted from facial and physiological signals to classify emotions accurately.

The process begins with data cleaning and normalization to prepare the dataset. Next, the data is split into training and testing sets to evaluate model performance effectively.

Various classification models are trained and tested to predict emotional categories. This framework provides a comprehensive approach to emotion recognition from the given data.

Model Details

The core models used in the system include Support Vector Machines (SVM), Decision Trees, Random Forests, and Ensemble learning methods.

SVM builds optimal decision boundaries to separate different emotion classes effectively. Decision Trees partition the data by selecting feature thresholds that best split the dataset into distinct classes.

Random Forests create multiple decision trees and combine their results to reduce overfitting and improve prediction stability. Ensemble learning combines predictions from several classifiers to further enhance accuracy.

Principal Component Analysis (PCA) is applied for dimensionality reduction and visualization of the data in lower-dimensional space. Additionally, feature importance analysis helps identify key variables that influence emotion classification, supporting both accuracy and interpretability.

REQUIREMENTS:

Software Requirements

The project requires MATLAB software with Statistics and Machine Learning Toolbox for data processing, visualization, and model training. A compatible version of MATLAB that supports advanced functions such as fitcecoc, TreeBagger, and fitcensemble is necessary.

Additional tools include data visualization features like confusion charts, heatmaps, and plotting utilities to aid in exploratory data analysis and result interpretation.

The system also relies on CSV file reading capabilities to import the emotion dataset efficiently for preprocessing and modeling.

Methodology

The methodology follows a structured approach starting with data cleaning to handle missing values and outliers, ensuring the dataset is reliable for analysis.

Exploratory data analysis is performed to understand the data distribution, feature relationships, and label balance. Visualization techniques such as histograms, boxplots, and correlation heatmaps help in this process.

After preprocessing, the data is normalized and split into training and testing subsets to evaluate model performance.

Multiple machine learning models including SVM, Decision Trees, Random Forests, and Ensembles are trained and tested to identify the best classifier for emotion recognition.

Dimensionality reduction with PCA and feature importance analysis further support model understanding and refinement.

EXPECTED RESULT:

The project aims to develop an emotion recognition system that can accurately classify 21 different emotional states from the dataset. Through effective data cleaning and analysis, the models are expected to perform well even in the presence of noisy data.

The system should achieve high classification accuracy using techniques such as Support Vector Machines, Decision Trees, Random Forests, and Ensemble methods. Additionally, the project will identify the most important features influencing emotion recognition, which can guide future enhancements.

Visualizations and statistical summaries will help validate the quality of the cleaned data, and confusion matrices will provide detailed insights into model performance across different emotion categories.

DATA SAMPLE:

	A	B	C	D	E	F	G	H	I	J
1	emotion_label	mouth_width	eye_distance	eyebrow_raise	smile_intensity	frown_intensity	eye_openness	mean_GSR	GSR_peaks	GSR_slope
2	Embarrassed_Nervous	32.61183151	43.8940144	0.615225981	0	0	0.594217682	4.066821495	3	-0.220134981
3	Nervous	30.41015261	45.68544301	0.530324943	0	0.740948327	0.549559667	4.118765078	2	-0.031219515
4	Happy_Nervous	33.46900756	46.66537903	0	0.789948449	0.345829323	0.523889952	4.293096229	4	0.077128786
5	Angry	33.695494	43.04472488	0.034667848	0.028427635	0.947257325	0.498387978	4.856981037	2	-0.085465921
6	Embarrassed_Nervous	30.80809487	44.04062676	0.78278777	0	0	0.579300968	4.005178173	6	0.141641915
7	Happy	32.56291426	44.98709564	0.217407583	0.922274574	0.092963197	0.568652619	4.532930956	6	-0.034387671
8	Happy_Nervous	32.73909052	43.1138286	0	0.708230084	0.306432229	0.591719938	4.277679903	3	-0.100537395
9	Embarrassed	30.01101911	42.75504087	0.751457076	0	0	0.447381018	3.848221399	5	-0.06941111
10	Sad	30.00806053	43.61336634	0.003478992	0.061537198	0.854630707	0.322068285	1.737050668	5	0.020345211
11	Sad_Angry	31.51724932	44.45950434	0	0	0.816608209	0.403054807	4.297354824	6	0.055517012
12	Angry	30.43355588	45.15290932	0.076349811	0.025885342	0.814499133	0.491571433	4.925796249	4	-0.030687451
13	Happy_Relaxed	30.37958934	46.79258784	0	0.834818894	0	0.504189473	2.515630099	3	0.080396095
14	Frustrated	30.37650261	44.9976835	0	0	0.77062082	0.463427695	4.713789333	1	0.059938168
15	Surprise	31.24487449	44.06029118	0.815603581	0.342398095	0	0.808685415	4.95958854	4	-0.095721465
16	Bored	31.51486522	45.46497434	0	0.018780969	0.009183123	0.452109639	1.298819422	2	0.046097255
17	Disgust_Angry	33.527361	45.26348105	0	0	0.803013445	0.489015664	4.429228122	4	-0.001867705
18	Surprise	34.63927425	44.4067042	0.823062061	0.354590223	0	0.869678061	4.997336266	3	0.086061324
19	Fear	33.87783251	46.58945206	0	0	0.779059396	0.443314914	4.738909272	2	0.012627874
20	Disgust_Angry	33.89360146	46.86052842	0	0	0.914539108	0.400045667	4.413284462	5	0.02934013
21	Disgust_Angry	30.98914337	46.62436163	0	0	0.782535823	0.422305475	4.426193499	4	-0.100010929
22	Surprise	30.43911307	43.01495862	0.803472856	0.354273551	0	0.826693446	4.993689032	5	0.009271899

emotion_dataset_noisy

Figure: Image showing structured dataset containing facial and GSR features used for emotion classification

The dataset used in this project consists of multiple numeric features representing emotional signals, along with 21 distinct emotion labels. The data includes some noise and missing values, which are addressed during the cleaning process. Features are normalized to ensure uniform scaling before training the models.

Exploratory analysis includes feature histograms, boxplots, and correlation heatmaps to understand the distribution and relationships within the data. This helps confirm the data's readiness for machine learning tasks and supports the selection of suitable classification algorithms.

Source Code:

```
%% Load Emotion Dataset (Noisy Version)
data = readtable('emotion_dataset_noisy.csv');

%% Data Cleaning (SFDS Step 1)


% Check for missing values
missing_counts = sum(ismissing(data));
fprintf('Missing values per column:\n');
```

```
Missing values per column:
```

```
disp(missing_counts);
```

```
     0     0     0     0     0     0     0     0     0     0
```

```

% Impute missing numeric data with median
numeric_vars = varfun(@isnumeric, data, 'OutputFormat', 'uniform');
for i = 1:width(data)
    if missing_counts(i) > 0 && numeric_vars(i)
        col = data{:,i};
        col(ismissing(col)) = median(col(~ismissing(col)));
        data{:,i} = col;
    end
end

% Handle outliers using IQR-based winsorization
features_raw = data{:, 2:end};
for col = 1:size(features_raw, 2)
    Q1 = quantile(features_raw(:,col), 0.25);
    Q3 = quantile(features_raw(:,col), 0.75);
    IQR = Q3 - Q1;
    lower_bound = Q1 - 1.5 * IQR;
    upper_bound = Q3 + 1.5 * IQR;

    features_raw(features_raw(:,col) < lower_bound, col) = lower_bound;
    features_raw(features_raw(:,col) > upper_bound, col) = upper_bound;
end
data{:, 2:end} = features_raw;

%% Exploratory Data Analysis (SFDS Step 2)

% 1. Basic statistics
```

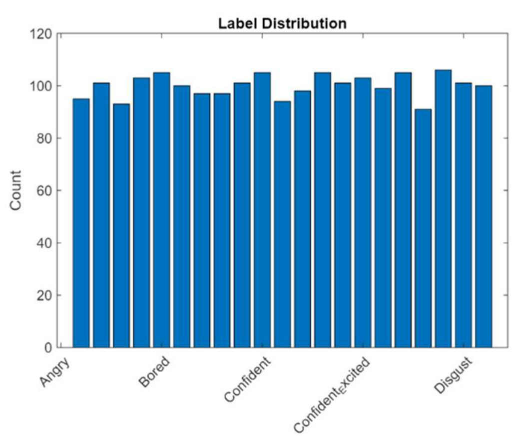
```
stats = array2table([mean(features_raw); median(features_raw);
std(features_raw)]', ...
     'VariableNames', {'Mean', 'Median', 'StdDev'}, ...
     'RowNames', data.Properties.VariableNames(2:end));
disp('Basic statistics for features:');
```

```
Basic statistics for features:
```

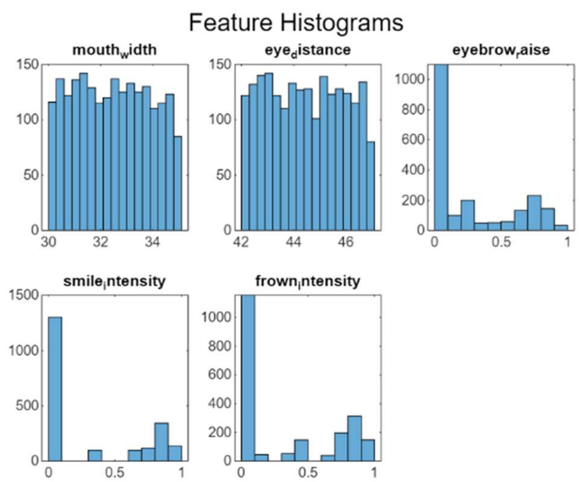
```
disp(stats);
```

	Mean	Median	StdDev
mouth_width	32.473	32.453	1.4358
eye_distance	44.476	44.449	1.4526
eyebrow_raise	0.26811	0.06943	0.32355
smile_intensity	0.30051	0.034833	0.38082
frown_intensity	0.32416	0.036961	0.38086
eye_openness	0.54599	0.53275	0.12925
mean_GSR	3.902	4.2571	1.1149
GSR_peaks	3.5319	3	1.7129
GSR_slope	0.00084208	0.001698	0.098804

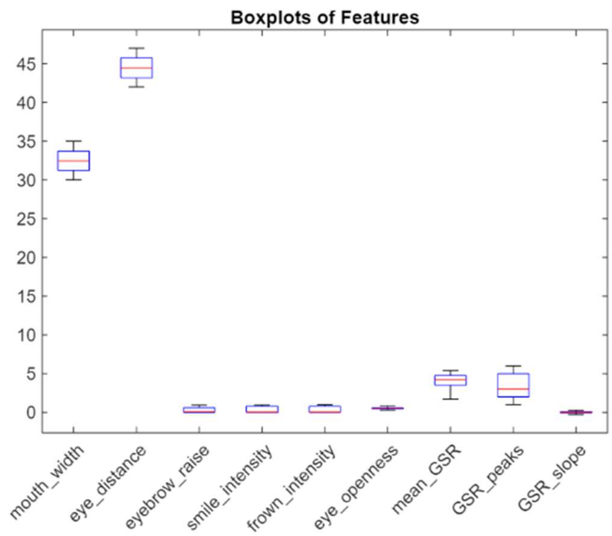
```
% 2. Label distribution
figure;
label_counts = countcats(categorical(data{:,1}));
bar(label_counts);
xticklabels(categories(categorical(data{:,1})));
xtickangle(45);
ylabel('Count');
title('Label Distribution');
```



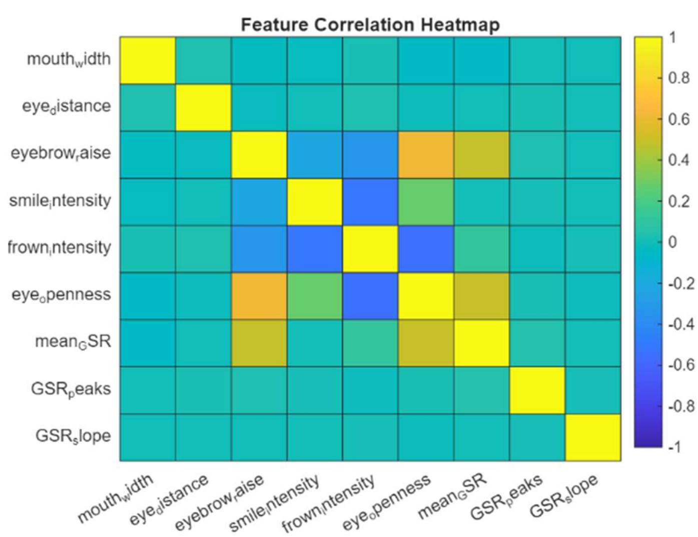
```
% 3. Feature histograms (first 5 features)
figure;
for i = 1:min(5, size(features_raw,2))

    subplot(2,3,i);
    histogram(features_raw(:,i));
    title(data.Properties.VariableNames{i+1});

end
sgtitle('Feature Histograms');
```




```
% 4. Boxplots
figure;
boxplot(features_raw, 'Labels', data.Properties.VariableNames(2:end));
xtickangle(45);
title('Boxplots of Features');
```

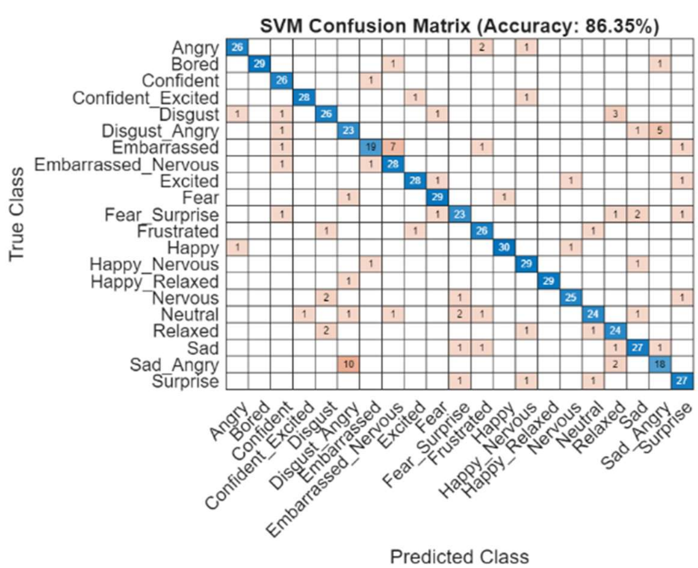



```
% 5. Correlation heatmap
corr_mat = corr(features_raw);
figure;
heatmap(data.Properties.VariableNames(2:end),
data.Properties.VariableNames(2:end), corr_mat, ...
     'Colormap', parula, 'ColorLimits', [-1 1], 'Title', 'Feature Correlation
Heatmap');
```

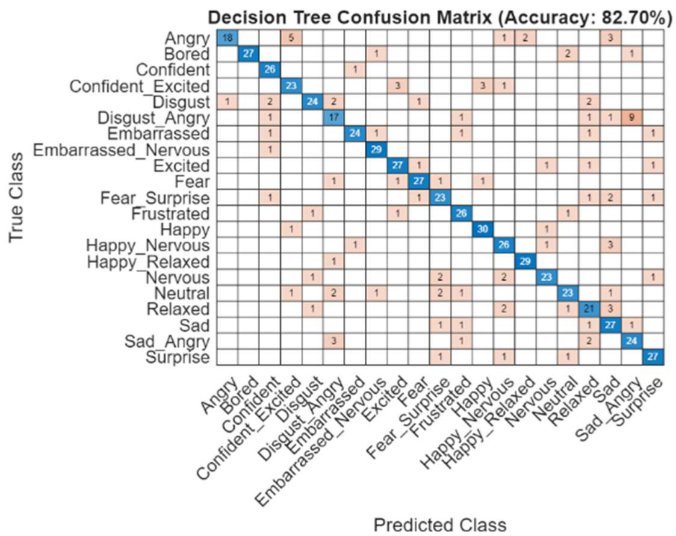



```
%% Feature Extraction and Normalization

features = data{:,2:end};
labels = categorical(data{:,1});
features = normalize(features);
```

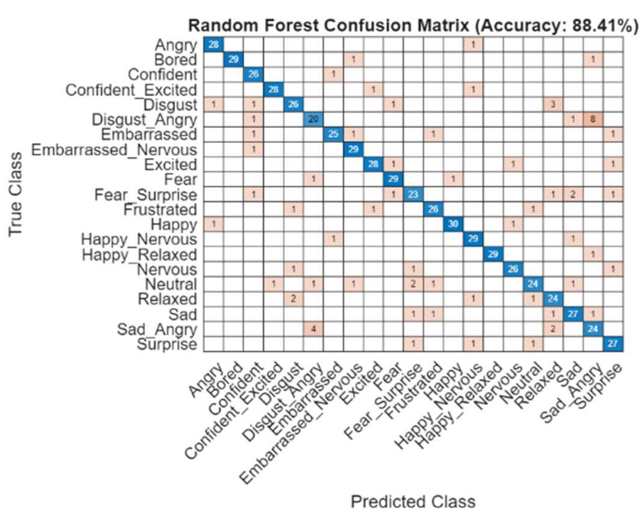
```
%% Train-Test Split

cv = cvpartition(labels, 'HoldOut', 0.3);
trainIdx = training(cv);
testIdx = test(cv);

X_train = features(trainIdx, :);
X_test = features(testIdx, :);
y_train = labels(trainIdx);
y_test = labels(testIdx);

%% SVM Classifier

svm = fitcecoc(X_train, y_train);
pred_svm = predict(svm, X_test);
acc_svm = mean(pred_svm == y_test) * 100;

figure;
confusionchart(y_test, pred_svm);
title(sprintf('SVM Confusion Matrix (Accuracy: %.2f%%)', acc_svm));
```




```
%% Decision Tree Classifier

tree = fitctree(X_train, y_train);
pred_tree = predict(tree, X_test);
acc_tree = mean(pred_tree == y_test) * 100;

figure;
confusionchart(y_test, pred_tree);
title(sprintf('Decision Tree Confusion Matrix (Accuracy: %.2f%%)',
acc_tree));
```

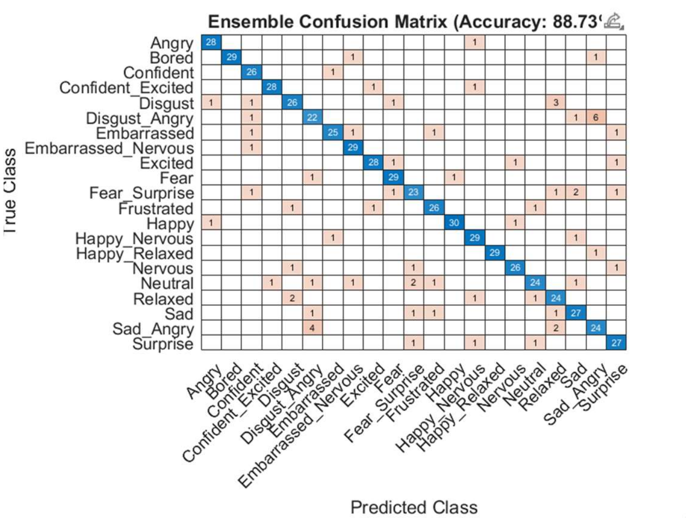



```
%% Random Forest Classifier

rf = TreeBagger(100, X_train, y_train, 'OOBPrediction', 'On', 'Method',
'classification');
pred_rf = predict(rf, X_test);
pred_rf = categorical(pred_rf);
acc_rf = mean(pred_rf == y_test) * 100;

figure;
confusionchart(y_test, pred_rf);
title(sprintf('Random Forest Confusion Matrix (Accuracy: %.2f%%)', acc_rf));
```

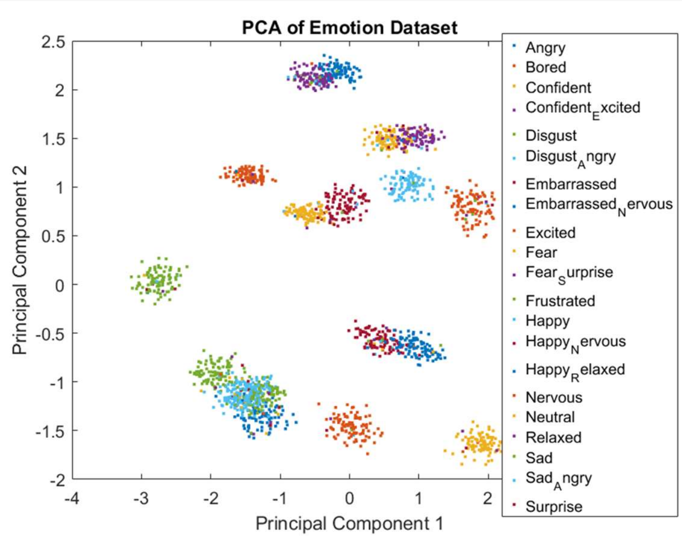



```
%% Ensemble Classifier

ens = fitcensemble(X_train, y_train, 'Method', 'Bag');
```

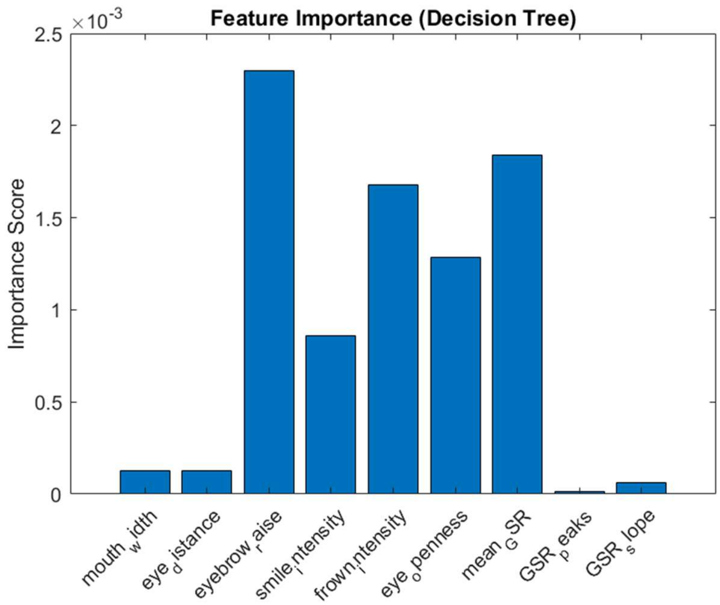
```
pred_ens = predict(ens, X_test);
acc_ens = mean(pred_ens == y_test) * 100;

figure;
confusionchart(y_test, pred_ens);
title(sprintf('Ensemble Confusion Matrix (Accuracy: %.2f%%)', acc_ens));
```




```
%% PCA Visualization

[coeff, score] = pca(features);
figure;
gscatter(score(:,1), score(:,2), labels);
title('PCA of Emotion Dataset');
xlabel('Principal Component 1');
ylabel('Principal Component 2');
```



```
%% Feature Importance (Decision Tree)

figure;
bar(tree.predictorImportance);
xticklabels(data.Properties.VariableNames(2:end));
xtickangle(45);
ylabel('Importance Score');
title('Feature Importance (Decision Tree)');
```




```
%% Final Accuracy Report

fprintf('SVM Accuracy: %.2f%%\nDecision Tree Accuracy: %.2f%%\nRandom Forest
Accuracy: %.2f%%\nEnsemble Accuracy: %.2f%%\n', ...
    acc_svm, acc_tree, acc_rf, acc_ens);
```

```
SVM Accuracy: 86.35%

Decision Tree Accuracy: 82.70%

Random Forest Accuracy: 88.41%

Ensemble Accuracy: 88.73%
```

FUTURE ENHANCEMENT:

Although the current system achieves promising accuracy across several machine learning models, there are multiple avenues for further improvement. One major enhancement could be the integration of deep learning models, such as Convolutional Neural Networks (CNNs), which are particularly effective for high-dimensional data like facial features. These models can automatically extract complex patterns and may outperform traditional classifiers in recognizing subtle emotional expressions.

Another potential improvement is expanding the dataset with real-world samples collected from diverse individuals under varied lighting and environmental conditions. This would increase the generalizability of the model and improve its performance in real-life scenarios. Additionally, combining facial features with other physiological signals, such as EEG or heart rate data, could lead to a more holistic and accurate emotion recognition system.

Real-time implementation is another promising direction. With optimization, the model could be embedded into mobile applications or web platforms to detect emotions on the fly. This would require further work in latency reduction, model compression, and system integration but holds immense practical value in areas such as mental health monitoring, interactive education, and adaptive user interfaces.

CONCLUSION:

This project successfully implemented an emotion recognition system by combining facial and galvanic skin response (GSR) data. Beginning with a noisy dataset, the process involved structured data cleaning, statistical analysis, and the application of various machine learning algorithms. Each step was carefully designed to ensure the data quality and model reliability were maintained throughout.

The final system demonstrated strong classification performance across models such as Support Vector Machines, Decision Trees, Random Forests, and Ensemble methods. Among these, the Ensemble model achieved the highest accuracy, indicating its effectiveness in handling complex data with varied emotional labels. The analysis also provided insights into feature importance and relationships through visualization and dimensionality reduction techniques like PCA.

Overall, the project highlights the potential of integrating statistical thinking with modern machine learning approaches to solve real-world classification problems. It sets a strong foundation for future work and enhancements, particularly in areas like real-time emotion detection and deployment in interactive systems.

REFERENCES:

1. Picard, R. W. (1997). *Affective Computing*. MIT Press. https://mitpress.mit.edu/9780262661157/affective-computing
2. Ko, B. C. (2018). A brief review of facial emotion recognition based on visual information. *Sensors*, 18(2), 401. https://doi.org/10.3390/s18020401
3. Soleymani, M., Lichtenauer, J., Pun, T., & Pantic, M. (2012). A multimodal database for affect recognition and implicit tagging. *IEEE Transactions on Affective Computing*, 3(1), 42–55. https://doi.org/10.1109/T-AFFC.2011.25
4. Custom Dataset. Synthetic dataset for emotion recognition using facial and GSR features. Generated using MATLAB, 2025.
5. MATLAB Documentation. (2023). *fitcecoc – Train multiclass models for support vector machines or other classifiers*. https://www.mathworks.com/help/stats/fitcecoc.html