Supervised Learning-Driven Emotion Recognition through Facial Landmark Trajectories Enhanced by SMOTE Methodology

Ishita Kar
Department of Computer Science and Engineering
St Thomas' College of Engineering and Technology
Kolkata, India
karishita025@gmail.com

Mehuli Chatterjee
Department of Computer Science and Engineering
St Thomas' College of Engineering and Technology
Kolkata, India
chatterjeemehuli@gmail.com

Sukanya Manna
Department of Computer Science and Engineering
St Thomas' College of Engineering and Technology
Kolkata, India
sukanya.manna.2002@gmail.com

Akash Das
Department of Computer Science and Engineering
St Thomas' College of Engineering and Technology
Kolkata, India
toosaddas@gmail.com

Stobak Dutta
Department of Computer Science and Engineering
St Thomas' College of Engineering and Technology
Kolkata, India
stobak.dutta@gmail.com

***Abstract*— The study highlights the importance of facial landmark trajectories in understanding mental health by investigating the application of supervised learning systems to identify emotions. Diagnoses and treatment plans are impacted by the difficulty in recognizing and controlling emotions, which are intrinsically tied to mental health. Although physiological signals like ECG and EEG are accurate, they have drawbacks including expense and user discomfort. Because they are non-invasive, facial landmarks present a viable substitute for practical uses. Using information from the ASCERTAIN dataset, this study divided emotions into four valence-arousal plane quadrants: High Arousal High Valence (HAHV), High Arousal Low Valence (HALV), Low Arousal High Valence (LAHV), and Low Arousal Low Valence (LALV). Algorithms including Random Forest, K-Nearest Neighbors (KNN), Decision Tree, and Support Vector Machine (SVM) were used to solve a 4-class classification problem. A common problem, class imbalance, was lessened by employing the Synthetic Minority Oversampling Technique (SMOTE). The study showed that SMOTE greatly enhanced predictions for minority classes, particularly for LALV, whereas classifiers strongly preferred the majority class in the absence of balance. Despite possible overfitting, Decision Tree has the greatest true positive rate among classifiers for LALV after SMOTE application. The findings demonstrate developments in emotional computing and provide information for applications in mental health support and human-computer interaction.**

Keywords— Emotion Recognition, Machine Learning, SVM, KNN, Decision Tree, Random Forest

I. INTRODUCTION

Understanding an individual's mental health has become a significant subject of study today. Mental health includes emotional, physiological, and social well-being. It affects the behaviour of a person. Mental well-being is important for healthy living and allows a human being to cope with the stresses faced in everyday life and to do his work effectively and efficiently. According to the World Health Organization, the burden of mental health problems in India is 2443 disability-adjusted life years (DALYs) per 10000 population; the age-adjusted suicide rate per 100000 population is 21.1[1]. Emotions are subjective psychological responses to any stimuli, feelings, or thoughts. Mental health and emotions are closely interlinked. Emotion regulation is essential for overall well-being. Proper management of emotions like happiness, sadness, anger, fear, and joy leads to positive mental health while unmanaged emotions are bound to cause problems like depression, anxiety, and mood disorders. Every person has a way of expressing emotions. Emotion recognition is important for accurate diagnosis of mental health conditions (e.g. depression, anxiety, etc.) and the development of efficient treatment strategies. Many individuals often struggle to identify emotions and thus it becomes difficult to regulate them. Emotions can be recognized from several sources including tone of voice, body language, facial expressions, or physiological signals like Electrocardiogram(ECG), Electroencephalogram(EEG), and galvanic skin response(GSR). Many researches have been carried out in recognizing emotions from physiological signals and they have given good results since physiological signals cannot be masked or suppressed. However, using physiological signals for emotion detection can be challenging since obtaining physiological signals from an individual requires using external devices like headsets or electrodes which can cause discomfort to the person. The person may feel nervous. This may interfere with the normal expression of their emotions. The setup required for collecting physiological signals is often expensive and unsuitable for everyday use. Recognizing emotions from facial landmark trajectories can be more comfortable and non-invasive than physiological signals. Users do not need to wear any external devices; thus, they will remain in their comfort zones which will increase the possibility of genuine

979-8-3315-3633-6/25/$31.00 ©2025 IEEE

emotional expressions. This method is also more feasible and suitable for real-world applications. Facial landmarks refer to particular points on the human face such as the corner of the eyes, the tip of the nose, the boundary of the eyebrows, etc that can be detected and tracked using Machine Learning algorithms from video or images. Trajectories involve tracking the location of these landmarks from one frame to another, showing changes in expression that indicate emotional states. There are various ways of representing emotions. Individual emotions like happy, sad, angry, calm, etc. can be generalized. In most cases, it is generalized into two dimensions valence and arousal. Valence refers to the pleasant or unpleasant quality of an emotion. Arousal refers to the intensity of the emotion. This work categorized the emotions into four quadrants of the valence arousal plane High Arousal High Valence(HAHV), High Arousal Low Valence(HALV), Low Arousal Low Valence(LALV), Low Arousal High Valence (LAHV). A comparative analysis of different classification algorithms has been shown to detect emotions from facial landmark trajectories. Initially, the algorithms were applied to the original dataset with class imbalance. Synthetic Minority Oversampling Technique(SMOTE) has been used to generate synthetic data to remove the class imbalance and the algorithms were applied again.

II. LITERATURE REVIEW

Emotion Recognition has become a crucial area of research in today's world. Many studies have been carried out in this field. Ayata et al. [2] developed an emotion recognition system using physiological signals from respiratory belts (RB), photoplethysmography (PPG), and fingertip temperature (FTT) sensors. DEAP dataset was used and their study involved classifying Arousal and Valence using Machine Learning Algorithms like random forest, logistic regression, and support vector machine. Separate classifiers were trained for each physiological signal followed by decision-level fusion of these classifiers. The fusion approach gave the highest accuracies, 73.08% for arousal and 72.18% for valence, the random forest classifier showed optimal performance. Girardi et al. [3] investigated emotion detection using non-invasive, low-cost sensors in a cross-subject classification framework. Their study used EEG, EMG, and GSR signals to classify valence and arousal into high or low categories. Music videos from the DEAP multimodal dataset were used for this purpose. Two binary classifiers were developed: one for arousal and another for valence, using Naive Bayes, SVM with a polynomial kernel, and decision tree Algorithm. The highest performance was achieved with SVM, showing an F1 score of 0.638 for arousal when combining EEG and GSR signals. Wiem and Lachiri [4] investigated emotion recognition from electrocardiogram (ECG), galvanic skin response (GSR), skin temperature, and respiration volume using the MAHNOB-HCI dataset. They classified emotions into three regions within the valence-arousal model: calm, medium arousal, and excited, as well as unpleasant, neutral, and pleasant valence. The study assessed each signal independently to identify the most relevant features for emotion detection. Feature-level fusion was then applied to combine the signals. Support vector machines (SVM) Algorithm was applied with different kernels. The radial basis function (RBF) kernel achieved the highest accuracy, with classification rates of 54.73% for arousal and 56.83% for valence. Dutta et al. [5] used Galvanic Skin Response (GSR) signals for emotion detection. Emotions were classified into four quadrants in the valence-arousal plane: High-Arousal High-Valence (HAHV), Low-Arousal High-Valence (LAHV), High-Arousal Low-Valence (HALV), and Low-Arousal Low-Valence (LALV). They used three machine learning algorithms: K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Logistic Regression (LR) on the ASCERTAIN dataset for the experiment. The classifier performance was evaluated using the F1 score. Similar work was performed by them using ECG and EEG signals [6]. KNN, SVM, Decision Tree, Logistic Regression, and Linear Discriminant Analysis Algorithms were used. The KNN algorithm correctly identified all four classes from the ASCERTAIN dataset, but its overall F1 score was 48%. When the dataset was made class-balanced, the Logistic Regression algorithm outperformed the other algorithms with an F1 score of 53.5%. The different algorithms either failed to identify certain classes or had lower overall F1 scores. Sourina and Liu [7] used EEG signals for emotion recognition using a fractal dimension (FD) algorithm based on the Arousal-Valence model. Music and sound stimuli from the IADS database were used. They classified emotions using SVM obtaining 58% accuracy for distinguishing three arousal levels. The studies reviewed primarily classify valence and arousal as separate binary targets, which can simplify emotional states and fail to capture the wide range of emotions across the valence-arousal plane. The emotional states that do not fit completely into high or low categories may be neglected. The use of external devices to collect physiological signals may cause discomfort to participants that may affect the performance of the model. Hassouneh et al. [8] developed a real-time emotion recognition system for physically disabled individuals and children with autism, using facial landmarks with EEG signals. They used convolutional neural networks (CNN) and long short-term memory (LSTM) classifiers to identify six basic facial expressions: sadness, anger, happiness, fear, disgust, and surprise. The system achieved a peak accuracy of 99.81% for facial expression recognition with CNN and 87.25% for emotion detection with EEG signals. Munasinghe [9] developed a facial emotion recognition system utilizing facial landmarks and a random forest classifier trained on the Extended Cohn-Kanade database. It identified four discrete emotions: anger, happiness, sadness, and surprise. This model achieved an average classification accuracy of 90%. Aneta Kartali et. al. [10] compared five methods for real-time emotion recognition of four basic emotions: happiness, sadness, anger, and fear, using facial images. The methods include three deep-learning approaches (AlexNet CNN, commercial Affdex CNN, and custom FER-CNN) and two conventional approaches (Support Vector Machine and Multilayer Perceptron using HOG features). It was seen that SVM and MLP showed lower performance and Affdex CNN achieved a higher accuracy of 85.05%. Kondaveeti and Goud [11] utilized deep learning architectures in Keras with transfer learning from models like VGG-16, ResNet152V2, InceptionV3, and Xception. The models were evaluated on a dataset combining Cohn-Kanade (CK+) and JAFFE, achieving accuracies of 83.16%, 82.15%, 77.1%, and

78.11% respectively. Deshmukh, Jagtap, and Paygude [12] highlighted that facial features are crucial for emotion detection. Their methodology involved extracting meaningful information from facial features to classify emotions effectively. The study involves image acquisition, preprocessing, face detection, feature extraction, and classification using machine learning. They classified various movements of facial parts like "open eyes, open mouth, lip corner pulled, cheeks raised" as the emotion "happy" and so on, which helped classification of Facial Landmark Trajectories detection. In the above studies, a fixed set of specific emotions like happiness, sadness, anger, fear, etc. has been considered for emotion recognition which may fail to capture the wide range of emotions.

III. METHODOLOGY

The ASCERTAIN [13] dataset, obtained from IEEE Transactions on Affective Computing, is chosen for the initial exploration of our study because of its state-of-the-art works. The dataset has been preprocessed to obtain our desired dataset. The columns label_valence and label_arousal have been combined to obtain the target value. For every instance of the data where both label_valence and label_arousal give positive value, it was denoted as HAHV, where label_arousal is positive, label_valence is negative, it was denoted as HALV, where label_arousal is negative, label_valence is positive, it was denoted as LAHV, and where both label_valence and label_arousal give negative value, it was denoted as LALV. Then the features that correspond to the measure of facial landmark trajectories have been extracted from the total set of features. So now the dataset is of a 4-class classification problem. The features were standardized such that each feature had a standard deviation of 1 and a mean of 0. The resultant dataset is then used with four supervised learning algorithms - Decision Tree, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Random Forest. In each of these algorithms, 20% of the data has been used for testing while the rest has been used for training the model. A huge amount of class imbalance was observed. So synthetic observations were generated from the existing samples of the dataset for the training data using Synthetic Minority Oversampling Technique (SMOTE). The resultant class-balanced dataset was then used with the above four algorithms. Fig. 1 shows the system overview.

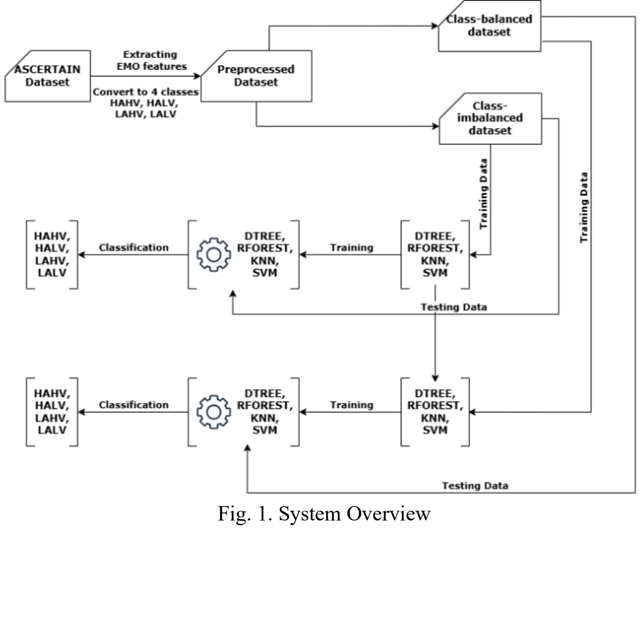

Fig. 1. System Overview

IV. SYNTHETIC DATA GENERATION

Class imbalance is a problem in machine learning that occurs when one class has a much higher number of samples than the other classes. Due to this the model gives good result for majority classes but does not perform well for minority class causing inaccurate predictions for the minority classes. To solve this problem a technique called oversampling is used. Oversampling increases the number of data points in the minority class to balance the dataset before model training. This can be done either by duplicating the existing data points or generating synthetic data points. Traditional oversampling methods include Random oversampling in which data points from the minority classes are selected randomly and duplicated without any change until the classes become balanced. The Synthetic Minority Oversampling Technique (SMOTE) [14] is another widely used method for removing class imbalance. SMOTE operates by generating synthetic samples along the line segment joining the existing minority class samples and their K-Nearest Neighbors. The steps are:

1. Identifying Nearest Neighbors: For each data point of the minority class, k-nearest neighbors are identified within the same class. Normally, the value of k is taken as 5.
2. Random Selection of Neighbors: Depending on the desired oversampling percentage, a subset of neighbors is selected.
3. Generating Synthetic Points: For each selected neighbor, the difference between the sample and its nearest neighbor is calculated. This difference is scaled by a random number between 0 and 1 and added to the original sample, generating a new synthetic sample.

V. RESULT AND DISCUSSION

A. *Class Imbalanced Dataset*

After applying the Decision Tree algorithm to the dataset, it is seen that 74% of the HAHV class samples have been correctly classified as HAHV. 29% of the HALV class samples have been correctly classified as HALV. It is observed that the classifier majorly failed to classify samples from the LALV and LAHV classes correctly. PCA is applied to the dataset to reduce dimensionality, and the Decision Tree algorithm is applied again. It is seen that 91% of the HAHV class samples have been correctly classified as HAHV. 8% of the HALV class samples have been correctly classified as HALV. It is observed that the classifier fails to classify samples from the LAHV and LALV class completely. The Confusion Matrix is shown in Fig. 2.

For the SVM algorithm, it was seen that 99% of the HAHV class samples have been correctly classified as HAHV. At the same time, the classifier failed to correctly classify the samples from HALV, LALV, and LAHV classes. Most of the data points belonging to these three classes were misclassified as HAHV. After applying PCA on the dataset to reduce its dimensionality and then applying the SVM algorithm, it can be seen from Fig. 3 that 99% of the HAHV class samples have been correctly classified as HAHV whereas only 3% of the HALV class samples have been correctly classified as HALV. The classifier fails to classify samples from the LAHV and LALV classes correctly.

Fig. 4 shows the Confusion Matrix Obtained after applying the Random Forest Algorithm to the dataset. From the confusion matrix, it is seen that 88% of the HAHV class samples have been correctly classified as HAHV. 11% of the HALV class samples have been correctly classified as HALV. It is observed that the classifier majorly fails to classify samples from the LALV and LAHV classes correctly. After applying PCA to reduce its dimensionality, the Random Forest Algorithm is reapplied to the dataset. It is seen that 86% of the HAHV class samples have been correctly classified as HAHV. 15% of the HALV class samples have been correctly classified as HALV. The classifier fails to classify samples from the LAHV and LALV classes correctly.

Fig. 5 shows the confusion matrix obtained after applying the KNN Algorithm to the dataset. From the confusion matrix, it can be seen that 78% of the HAHV class samples have been correctly classified as HAHV. 21% of the HALV class samples have been correctly classified as HALV. It is observed that the classifier majorly fails to classify samples from the LALV and LAHV classes correctly. PCA is applied to the dataset to reduce its dimensionality and then the KNN algorithm is applied again. It has been observed that 77% of the HAHV class samples have been correctly classified as HAHV. 22% of the HALV class samples have been correctly classified as HALV. 5% of the LAHV class samples are correctly classified as well. The classifier fails to classify samples from the LALV class correctly in both cases.

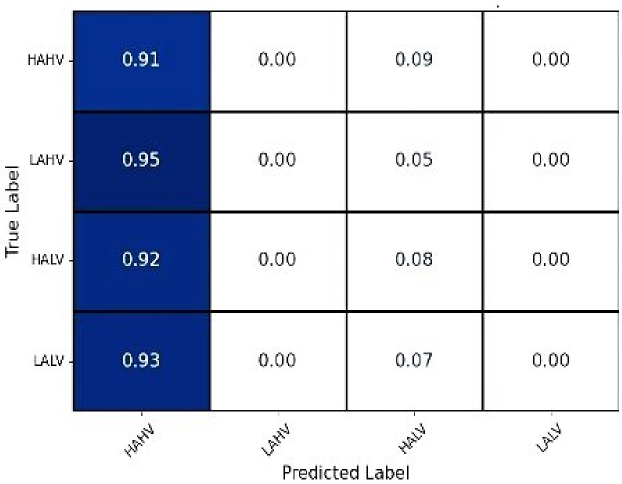

Fig. 2. Confusion Matrix after applying Decision tree

Fig. 3. Confusion Matrix after applying SVM

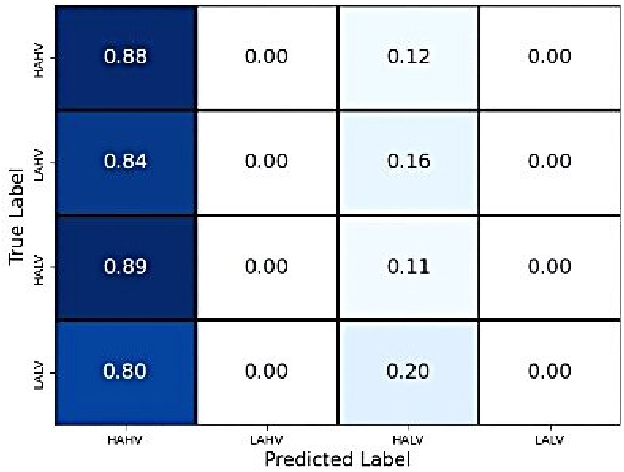

Fig. 4. Confusion Matrix after applying Random Forest

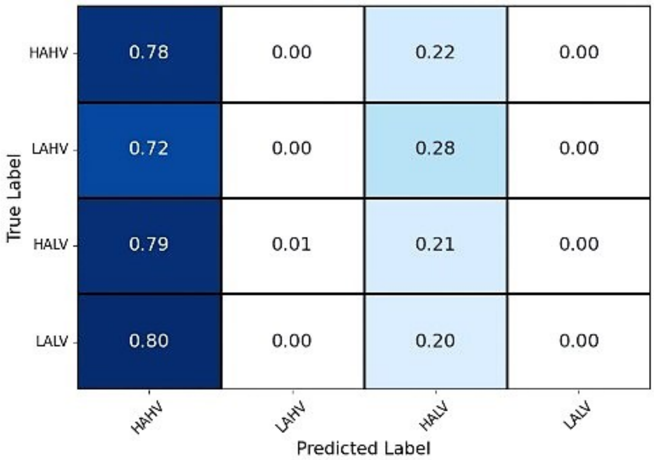

Fig. 5. Confusion Matrix after applying KNN

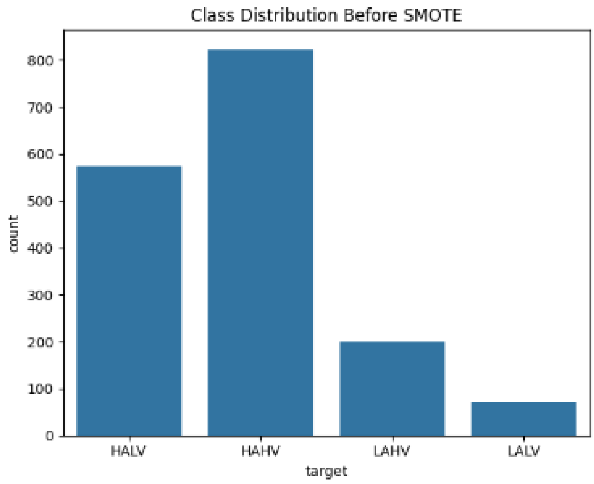

Fig. 6(a). Class distribution before applying SMOTE

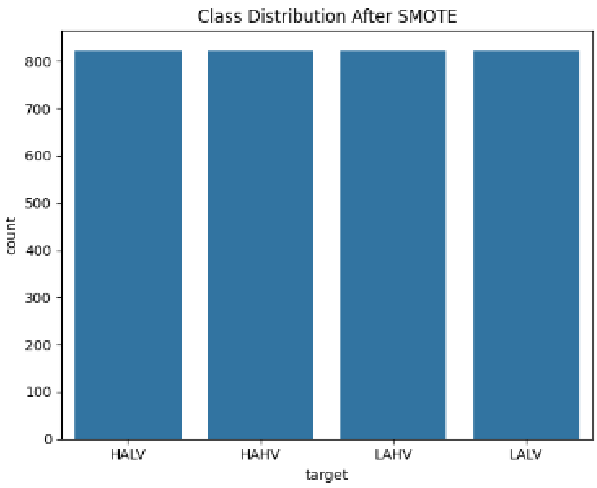

Fig. 6(b). Class distribution after applying SMOTE

B. Class Balanced Dataset

From the above Confusion Matrices, it can be seen that the Classifiers fail to classify samples from the LALV and LAHV classes correctly and a large percentage of the data points are misclassified as HAHV class. It is evident that the HAHV class contains more data points than the other classes and is the majority class. It indicates that there is a class imbalance in the dataset. To rectify the class imbalance SMOTE has been used to generate synthetic data in the minority classes so that all four classes have the same number of data points. Now all four Algorithms are again applied to the dataset. Fig. 6(a) shows the class distribution before applying SMOTE to the dataset. Fig. 6(b) shows the class distribution after applying SMOTE to the dataset.

After applying the Decision Tree Algorithm to the dataset, it can be seen that 28% of the HAHV class samples are correctly classified as HAHV. 24% of the HALV class samples have been correctly classified as HALV. 53% of the LALV class samples have been correctly classified as LALV.7% of the LAHV class samples have been correctly classified as LAHV. PCA is applied to reduce dimensionality and the Decision Tree Algorithm is used on the dataset again. It can be seen that 27% of the HAHV class samples are correctly classified as HAHV. 22% of the HALV class samples have been correctly classified as HALV. 73% of the LALV class samples have been correctly classified as LALV. 19% of the LAHV class samples have been correctly classified as LAHV. The Confusion Matrix is shown in Fig. 7.

For the SVM Algorithm, it can be seen that 28% of the HAHV class samples are correctly classified as HAHV, 21% of the HALV class samples have been correctly classified as HALV, 60% of the LALV class samples have been correctly classified as LALV and only 9% of the LAHV class samples have been correctly classified as LAHV. Changes were observed after applying PCA to reduce dimensionality as shown in Fig 9. The algorithm identified 67% of LALV samples correctly whereas the identification of the HAHV class reduced to 23%, the HALV class reduced to 20% and LAHV increased to 16%. The Confusion Matrix is shown in Fig. 8.

Fig. 9 shows the confusion matrix obtained after applying the Random Forest Algorithm. It can be seen that 41% of the HAHV class samples are correctly classified as HAHV. 22% of the HALV class samples have been correctly classified as HALV. 25% of the LALV class samples have been correctly classified as LALV. 9% of the LAHV class samples have been correctly classified as LAHV. A confusion Matrix was obtained after applying PCA to reduce dimensionality and then the Random Forest Algorithm again. It was seen that 45% of the HAHV class samples are correctly classified as HAHV. 24% of the HALV class samples have been correctly classified as HALV. 25% of the LALV class samples have been correctly classified as LALV. LAHV class samples are not detected correctly at all.

Fig. 10 shows the confusion matrix obtained after applying the KNN algorithm. It can be seen that 27% of the HAHV class samples are correctly classified as HAHV. 51% of the HALV class samples have been correctly classified as HALV. 6% of the LALV class samples have been correctly classified as LALV. 22% of the LAHV class samples have been correctly classified as LAHV. After applying PCA to reduce dimensionality, the KNN algorithm was applied again. It was seen that 33% of the HAHV class samples are correctly classified as HAHV. 51% of the HALV class samples have been correctly classified as HALV. 18% of the LALV class samples have been correctly classified as LALV. 18% of the LAHV class samples have been correctly classified as LAHV.

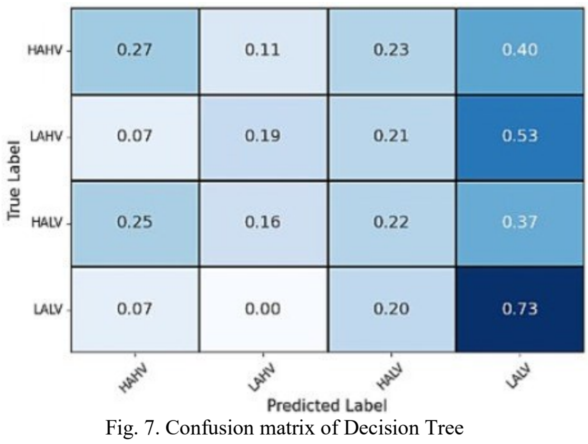

Fig. 7. Confusion matrix of Decision Tree

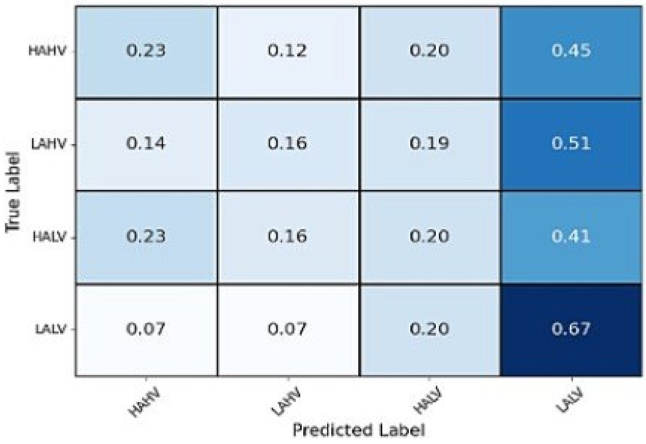

Fig. 8. Confusion matrix of SVM

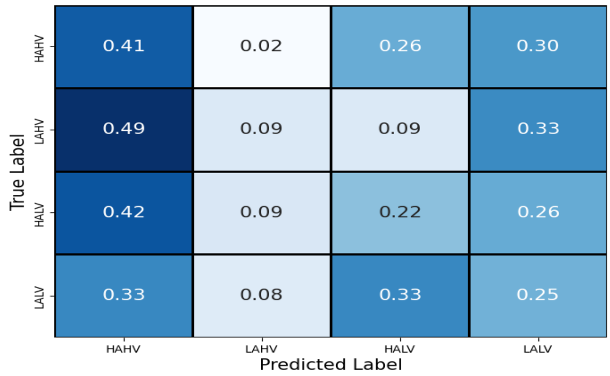

Fig. 9. Confusion matrix of Random Forest

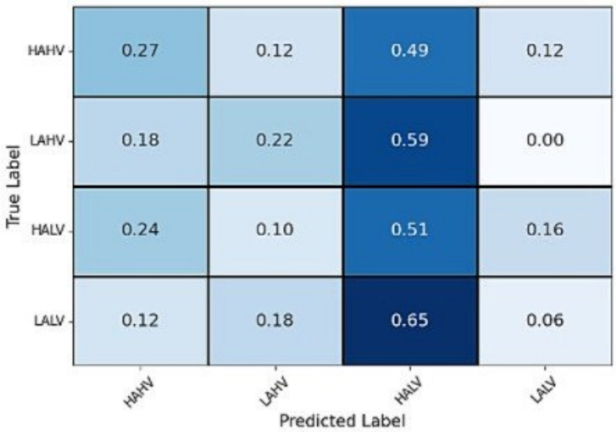

Fig. 10. Confusion Matrix of KNN

VI. CONCLUSION

The study successfully demonstrated how accurately supervised learning algorithms can recognize diverse emotions by categorizing them into 4 classes in accordance to their similarity in valence and arousal value. A significant amount of class imbalance was found in the dataset which was addressed by using SMOTE, increasing the model's performance in detecting minority classes. Supervised learning algorithms like Decision Tree, Support Vector Machine, Random Forest and KNN were used on the dataset, both before and after applying SMOTE. We saw with the class imbalanced dataset, that the classifiers correctly predicted the majority class, i.e. HAHV, while they failed to predict the minority class, i.e. LALV. Most of the classifiers wrongly predicted other classes as HAHV showing a high bias towards the majority class. SMOTE technology boosted performance, particularly for minority classes. We saw that with a class-balanced dataset, the classifiers were able to predict the other classes, Now the classifiers were able to detect the minority class LALV. Among the four, the Decision Tree classifier detected the class LALV with the highest true positive. The study highlighted a limitation. With increased SMOTE sampling, there arises a chance of potential overfitting. This study opens the door for further exploration of alternative oversampling techniques, investing transfer learning and multimodal emotion recognition.

REFERENCES

[1] Mental health [Internet]. www.who.int.

[2] Ayata D, Yaslan Y, Kamasak ME. Emotion recognition from multimodal physiological signals for emotion aware healthcare systems. Journal of Medical and Biological Engineering. 2020 Apr;40:149-57.

[3] Girardi D, Lanubile F, Novielli N. Emotion detection using noninvasive low cost sensors. In2017 seventh international conference on affective computing and intelligent interaction (ACII) 2017 Oct 23 (pp. 125-130). IEEE.

[4] Wiem MB, Lachiri Z. Emotion sensing from physiological signals using three defined areas in arousal-valence model. In2017 International Conference on Control, Automation and Diagnosis (ICCAD) 2017 Jan 19 (pp. 219-223). IEEE.

[5] Dutta S, Mishra BK, Mitra A, Chakraborty A. An analysis of emotion recognition based on GSR signal. ECS Transactions. 2022 Apr 24;107(1):12535.

[6] Dutta S, Mishra BK, Mitra A, Chakraborty A. A Multi-modal Approach for Emotion Recognition Through the Quadrants of Valence–Arousal Plane. SN Computer Science. 2023 Jun 22;4(5):460.

[7] Sourina O, Liu Y. A fractal-based algorithm of emotion recognition from EEG using arousal-valence model. InBiosignals 2011 Jan 26 (pp. 209-214).

[8] Hassouneh A, Mutawa AM, Murugappan M. Development of a real-time emotion recognition system using facial expressions and EEG based on machine learning and deep neural network methods. Informatics in Medicine Unlocked. 2020 Jan 1;20:100372.

[9] Munasinghe MI. Facial expression recognition using facial landmarks and random forest classifier. In2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS) 2018 Jun 6 (pp. 423-427). IEEE.

[10] Kartali A, Roglić M, Barjaktarović M, Đurić-Jovičić M, Janković MM. Real-time algorithms for facial emotion recognition: a comparison of different approaches. In2018 14th Symposium on Neural Networks and Applications (NEUREL) 2018 Nov 20 (pp. 1-4). IEEE.

[11] Kondaveeti HK, Goud MV. Emotion detection using deep facial features. In2020 IEEE international conference on advent trends in multidisciplinary research and innovation (ICATMRI) 2020 Dec 30 (pp. 1-8). IEEE.

[12] Deshmukh RS, Jagtap V, Paygude S. Facial emotion recognition system through machine learning approach. In2017 international conference on intelligent computing and control systems (iciccs) 2017 Jun 15 (pp. 272-277). IEEE.

[13] Subramanian R, Wache J, Abadi MK, Vieriu RL, Winkler S, Sebe N. ASCERTAIN: Emotion and personality recognition using commercial sensors. IEEE Transactions on Affective Computing. 2016 Nov 4;9(2):147-60.

[14] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research. 2002 Jun 1;16:321-57.