

Decoupling Affective Intelligence from Sensing Platforms: A Generalizable Framework for Real-Time Emotion Recognition across Multiple Modalities

Bhavya Sri Sanku

Independent Researcher, USA

Abstract - The probability of real-time emotional state recognition on affective computing systems have shown great potential, however most of the current methods still exhibit a very high reliance on particular sensing environments like VR systems as well as fixed physiological sensor arrangements. This study suggests an affect inference framework that is modality-agnostic with the machine learning layer decoupled from the sensing modality. The proposed structure combines heterogeneous physiological and behavioral characteristics with an integrated feature representation layer that can be used with the same classification model with all sensing platforms. Its methodology involves processing features, outlier manipulation, feature selection, modality adaptation and affect classification using machine learning with 3500 data. Various classifiers such as Random Forest, Support Vector machine and Neural Network models are tested on the basis of standard performance measures. Experimental research proves that the suggested architecture has significant effectiveness in terms of predicting emotional states as well as enhancing the model portability and scalability. The framework promotes the impending implementation in VR, wearable, mobile apps, and smart human-computer interaction.

Keywords - Affective Computing, Modality-Agnostic Learning, Emotion Recognition, Physiological Computing, Sensor-Independent Framework, Feature Representation, Stress Detection

I. INTRODUCTION

Background and Research Motivation

Affective computing has become the significant field of research that is aimed at providing machines with the ability to read, comprehend, and react to the emotional and mental conditions of humans. In modern physiological computing, recent innovations have made it possible to achieve real-time emotion and stress recognition with bio signals like electroencephalography (EEG), electrocardiography (ECG), galvanic skin response (GSR), eye tracking, and facial expressions [1]. The Virtual Reality (VR) environment is progressively becoming part of the study of affective computation due to its controlled and immersive environment in which physiological responses can be obtained [2]. However, most of the current models of affect recognition are highly contextualized with fixed sensing contexts, where the machine learning could be trained in respect to a given collection of devices and input modalities. This reliance constrains the

scalability and practice of affective AI systems in various platforms. A model trained on VR-based EEGs cannot work in mobile applications, wearable hardware or web interface which has other sensor properties [3]. Hence, it is desirable that an inference framework of generalized affect is necessary to isolate a layer of intelligence and the sensing environment.

Research Problem

Existing affective computing solutions tend to use a single pipeline of data acquisition, feature extraction and classification and do not rely on an abstract model that can address diverse sensors and application settings. Such inflexibility makes the models difficult to reuse and cross-platform using [4]. The research problem that is tackled in the work is that it should develop a modality-agnostic affect inference model that can sustain a reliable emotional state Hypothesis prediction as well as ignoring inputs of different sensing modalities.

Research Aim and Objectives

Aim:

The research aims to create a generalized machine learning model of real-time emotion state predication which is no longer dependent upon particular sensing modalities and application platforms.

Objectives:

- To create sensor-independent architecture to decouple affect classification and modality-based feature extraction.
- To assess a single feature representation scheme that can be able to combine heterogeneous physiological and behavioral inputs.
- To evaluate the ability of the proposed model to be generalized in various sensing conditions and tasks.
- To investigate the effectiveness of the framework in terms of the classification performance, flexibility, and real-time deployment performance.

Research Contributions

This study presents a modality-independent affective computing model that separates the emotional inference model and the VR-based sensorial worlds. The main contribution can be seen in the creation of a reusable AI architecture in which various sensor inputs can be changed without the fundamental part of the classifier having to be changed [5]. The proposed solution enhances the flexibility, scalability, and practical relevance of the

affect recognition systems with the ability to deploy in VR platforms, mobile apps, wearable, and web interfaces. Further, the present work explores cross-modal generalization and how a single affect model can be used to facilitate a wide variety of sensing configurations and retain the ability to reliably predict emotional states in real-time.

II. RELATED WORK

General Affective Computing and Emotion Recognition

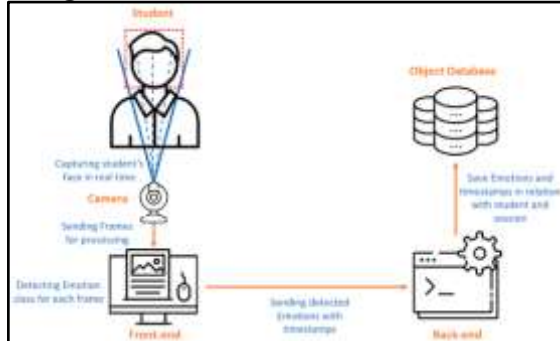


Fig. 1. General Affective Computing and Emotion Recognition

Affective computing is concerned with creation of intelligent systems that understand, recognize and react to human feelings [6]. The initial studies are mainly based on psychological theories like the valence-arousal model and specific categories of emotions [7]. Due to the developments in artificial intelligence, machine learning and deep learning methods have become commonly used in automatic emotion recognition [8]. Scholars have been used to research several sources of data such as facial expressions, speech patterns and physiological cues to enhance the prediction of emotions states [9]. Most current solutions however are specific to particular datasets and environments, which hinders their capacity to extrapolate to a range of actual applications and sensing situations [10].

Emotion Recognition using Physiological Signals

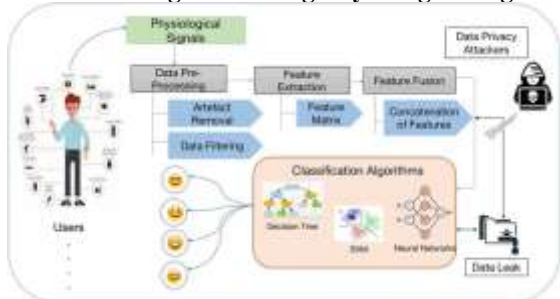


Fig. 2. Emotion Recognition using Physiological Signals

The promising physiological indicators related to emotional reactions have attracted intense interest to physiological computing [11]. EEG, ECG, and heart rate variability (HRV) as well as electro dermal activity (EDA) signals are well-researched determinants of stress, attention, or emotional state [12]. Support vector machine and random forest machine learning models, as well as neural

networks, have exhibited potential in the classification of physiological emotions [13]. Nevertheless, these models tend to rely on a particular sensor configuration and feature extraction methods [14]. The high reliance on specific biosensors poses a problem in the transfer of trained models to new platforms or other sensing platforms [15].

Multimodal Learning to Learn Affective States

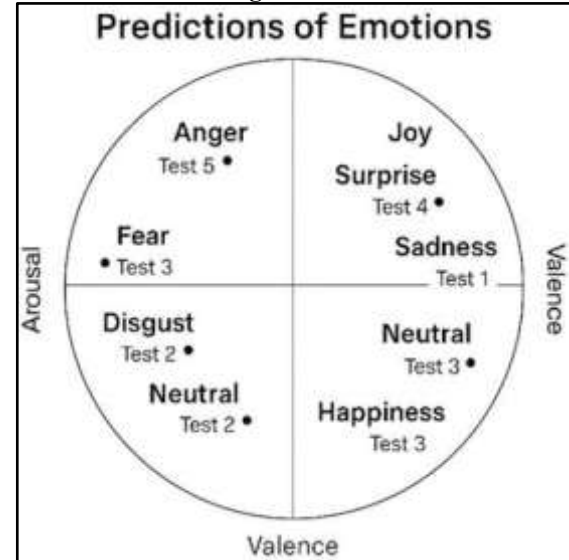


Fig. 3. Multimodal Learning to Learn Affective States

Multimodal affective computing involves a combination of data gathered in various modes, including physiological data, facial muscles, voice and movement, to enhance effective emotion recognition [17]. This has been made possible by deep learning techniques, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs) and transformer-based models, which have facilitated successful data fusion across various types of data [18]. Multimodal research has also been advanced by Virtual Reality environments in terms of emotional stimulation in a controlled setting, and real-time collection of data [19]. Nevertheless, most multimodal systems despite the better performance are highly integrated with fixed sets of sensors and they are hard to configure to a new situation when new modalities or other application conditions come into play [20].

ML Generalization and Transfer learning in Affective computing

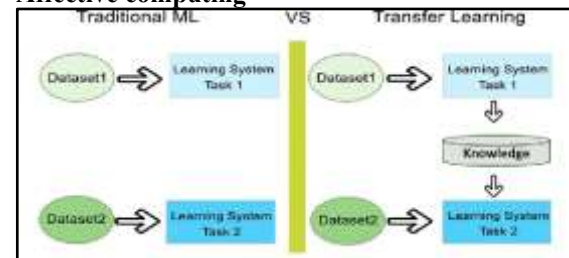


Fig. 4. ML Generalization and Transfer learning

Recent research has explored the use of transfer learning, domain adaptation, and representation learning methods to enhance the resiliency of affect recognition models in various users and settings [21]. The aim of these methods is to minimize variations induced by sensor differences in addition to personal physiological peculiarities and data distribution discrepancies [22]. Even though these techniques enhance flexibility of models, the majority of current models need partial retraining, or domain-specific adaptation in case of sensing modality adaptation [23]. Architecture of entirely decoupled affect inference, in which a trained model can be run off the outputs of a variety of replaceable input sources, is a novel research direction [24].

Research Gap

Current systems on affect recognition have been capable of high performance in dedicated sensing environments; however, it tends to be not so adaptive in other modalities [25]. The modern methods tend to integrate sensor grabbing and classification in a single pipeline, decreasing the portability and scalability. The development of an actual modality-agnostic affect inference framework in which the essence machine learning model is not dependent on the sensing platform is under-researched [26]. This study helps to fill this void by suggesting a reusable architecture that has the potential to support many input modalities without compromising real-time emotional state prediction.

III. PROPOSED MODALITY-AGNOSTIC FRAMEWORK

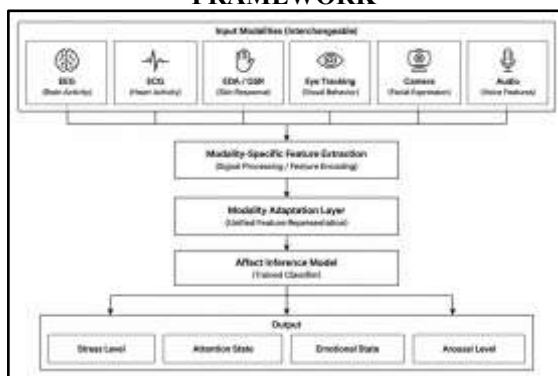


Fig. 5. Proposed Modality Agnostic Framework

The suggested modality-independent system provides a loose framework of real-time affective state inference through decoupling of the model of emotional intelligence and the idea of sensing active platform [27]. In contrast to more traditional affect recognition systems that are established based on a pre-determined sensor setup, this structure allows the same machine learning application to run with various interchangeable inputs, which can be physiological, behavioral or environmental cues. The model comprises four large parts, namely: modality-specific feature extraction, modality adaptation layer, affect inference model, and emotional state output layer. During the first stage, raw sensor data of various sources, including EEG,

ECG, EDA/GSR, eye tracking, camera-based facial expression or audio signals, are separately handled [28]. Each modality is processed through the suitable preprocessing and feature extraction measures to get significant representations.

The features extracted are then fed into the modality adaptation layer which maps heterogeneous sensor representations to a single feature space. The layer limits the reliance on a single sensor, and different modalities can give complementary consumptions to the central inference model. Modality alignment can be realized by using techniques like feature normalization, embedding generation and representation learning.

The last step is the affect inference model which is the central component of AI that predicts emotional and cognitive states [29]. Depending on the needs of the application, machine learning models like neural networks, support vector machines, random forests, or transformer-based models may be combined. Examples of output generated by the model are stress level, attention state, emotional category and arousal level.

IV. METHODOLOGY

Research Design

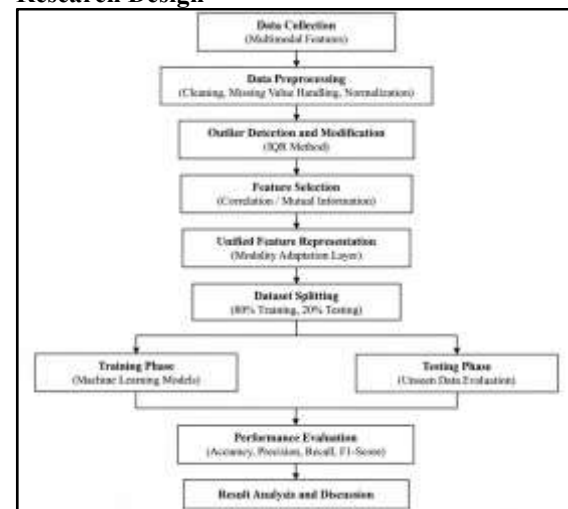


Fig. 6. Methodological Flow Diagram

The intended study employs an explicit machine learning approach that can build and test a modality-indifferent inference system that can forecast emotional conditions based on various modalities of sensing [30]. As the goal of this study is to confirm the generalizability of the AI architecture and not be limited by a certain hardware platform, a synthetic multimodal physiological dataset is created to recreate different sensor environments.

Dataset Generation

A synthetic dataset is generated to profile a variety of sensing modalities associated with effects such as EEG, ECG, EDA/GSR, eye-tracking, facial expression features and audio-based behavioral features. The dataset is formulated as per physiological and behavioral features that are usually linked to emotional conditions like stress,

attention, arousal and relaxation with 3500 entries and 21 attributes.

Individual generated samples consist of feature representations that are modality independent and affective labels. The phenomenon of realistic variations used in the synthetic data generation process includes individual variations, sensor noise, and signal variations used to mimic the real-world conditions. The data is composed of various stress groups, such as low stress, medium stress and high stress.

The resulting dataset can be formed as:

Target Labels: Stress_Level

- 0 = Low Stress
- 1 = Medium Stress
- 2 = High Stress

Data Preprocessing

The data generated are subjected to preprocessing before the model is trained, ensuring higher quality and consistency in the data. Handling of missing values and inconsistencies is detected and dealt with by means of statistical replacement. Normalization measures are done to perform feature scaling, to avoid features in different modalities having different ranges. The Gaussian noise injection and random sensor feature dropout are implemented to test the model robustness in the conditions of realistic noisy or incomplete sensing.

The max-min normalization techniques are defined as:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

where X represents the original feature value, X_{min} and X_{max} represent the minimum and maximum values of the feature, and X_{norm} represents the normalized feature value. As various modalities result in dissimilar feature distributions, normalization is used to remake all the features to a shared representation space. This is done to enhance compatibility with the modality adaptation layer and less biasness towards high-scale features.

The preprocessing pipeline consists of:

- Validation and data cleaning.
- Missing value correction.
- Feature normalization.
- Reduction of noise and smoothing of signals.
- Transformation to a single feature representation.

Outlier Detection and Modification

Outlier analysis is conducted to check the presence of deviant values of the features, which can adversely impact the model. Extreme values are not eliminated; physiological signals naturally have some variation which can then be analyzed carefully.

IQR method is used to find outliers:

$$IQR = Q3 - Q1$$

The acceptable feature range is calculated as:

$$LowerBound = Q1 - 1.5(IQR)$$

$$UpperBound = Q3 + 1.5(IQR)$$

Rather than the deletion of these samples, an outlier modification approach is implemented with the help of value transformation and boundary clipping. This retains significant differences in physiological responses and does not allow extreme values to affect model training.

Feature Selection

The process of feature selection is carried out to determine the most informative features regarding the prediction of the affective state and decrease the dimension that has no meaning. Seeing that the framework permits several sensing modalities, the selection of features per modality is performed using each modality and a unified representation of the features which are generated.

The features selected are:

EEG Features:

- Power of alpha, beta, and theta frequencies.
- Signal entropy.
- Brain activity variations.

ECG Features:

- Heart rate.
- RR interval.
- The variability of the heart rate (HRV).

EDA/GSR Features:

- Skin conductance level.
- Response amplitude.
- Signal variation rate.

Eye and Facial Features:

- Eye movement patterns.
- Blink frequency.
- Facial expression embedding.

Audio Features:

- Voice intensity.
- Frequency characteristics.
- Speech-related emotional indicators.

Machine learning models are also used in the extraction of redundant features, correlation analysis, mutual information, and feature importance ranking to eliminate redundant features and achieve a substantial amount of affect-related information.

Label Generation

The target variable Stress_Level is generated by mapping the simulated stress intensity score into three classes:

- Low Stress (0)
- Medium Stress (1)
- High Stress (2)

The labeling process is based on the combination of physiological and behavioral feature variations to represent realistic stress-response patterns.

Adaptation to modality and representation of features

Upon the feature selection, the modality-specific features are converted to a single representation through a modality adaptation layer. The step

guarantees that the various sensor inputs can be consumed by the identical affect inference model.

Adaptation process includes:

- Feature embedding generation.
- Dimensionality alignment.
- Feature normalization.
- Creation of a common latent feature space.

This enables this model to be given various input combinations of modalities without having to make changes to the underlying classifier.

Model Training and Testing Strategy.

The trained data is separated into testing and training data to measure the performance and power to generalize the proposed framework.

The data is divided with the ratio of 80:20 and:

- 80% of the data is used for model training.
- 20% of the data is reserved for independent testing.

The machine learning model parameters are optimized and associations between the feature representations, which is unified, and the affective states are learned using the training dataset. Various classification models such as the Artificial Neural Networks (ANN), Support Vector Machines (SVM), random forest, and deep learning can be compared. Model performance accuracy is computed as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

where:

- TP = True Positive samples
- TN = True Negative samples
- FP = False Positive samples
- FN = False Negative samples

Testing is only applied after the model is trained to assess the capability of the model to classify unseen samples. Accuracy, precision, recall, F1-score and inference latency are used to perform a performance evaluation.

Pseudo Code

```
1. Load dataset
2. Preprocess data and normalize features
3. Detect and modify outliers using IQR method
4. Select important features using feature selection
5. Transform features into unified representation
6. Split data into 80% training and 20% testing
7. Train ML models (RF, SVM, ANN)
8. Predict affective states using test data
9. Evaluate using Accuracy, Precision, Recall, and F1-score
```

```
10. Compare model performance
11. Generate analysis results
```

The algorithm describes the multimodal affective features, performs optimization steps, trains machine learning classifiers, evaluates emotional state prediction performance, and identifies the most effective modality-independent inference model in a proper way.

Cross-Modality Generalization Evaluation

One of the experiments that are conducted to confirm the modality-independent nature of the suggested framework is conducted through cross-modality experiments. This model is trained with a single combination of sensor representations, and evaluated with different modality configurations.

For example:

- Training: EEG + ECG aspects.
- Testing: ECG + EDA features

This test identifies whether or not the proposed architecture is capable of preserving the performance of affect prediction, when the sensing modalities vary. The findings showcase the potential of the framework to deliver reusable affect intelligence in varied application platforms like VR, mobile devices and wearable systems.

A modality adaptation layer is implemented using independent neural projection networks for each sensing modality. Each modality encoder transforms the original feature vector into a shared 64-dimensional latent representation.

The common dimension Latent is represented as:

$$\mathbf{Z} \in \mathbb{R}^{64}$$

The modality embeddings are combined using latent averaging, producing a sensor-independent representation that is passed to the stress classification network.

A classic feature-fusion baseline was introduced which involved direct stacking of all the features of the modalities into a single vector and then proceeding to classify without modality adaptation. In order to determine the strength of the framework, modalities were tested by testing the structure with and without the removal of individual modalities to determine the performance under conditions of incomplete sensor availability.

The model is trained and evaluated on cross-modality test results by training on one sensor combination and testing on new modality combinations to assess generalization ability.

V. RESULTS AND DISCUSSION



Fig. 7. Multiple event table summary statistics analysis

The statistical analysis demonstrates that synthetic dataset has 3500 samples with different physiological/behavioral characteristics. The feature representation indicates realistic variations in EEG,

ECG, EDA, eye-tracking and audio modalities. The data set distribution ensures a fit in the training and testing of multimodal affect recognition model.

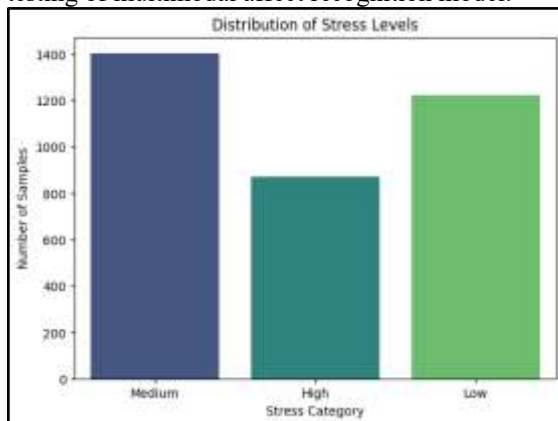


Fig. 8. Stress Level Distribution Level

The level distribution of stress reveals that there are three levels of emotional intensity, namely Low, Medium, and High stress. The highest is medium stress samples and then the Low and High stress conditions. The equal distribution aids in the fair classification of machine learning and minimizes the risk of the inclination to a certain category of stress.

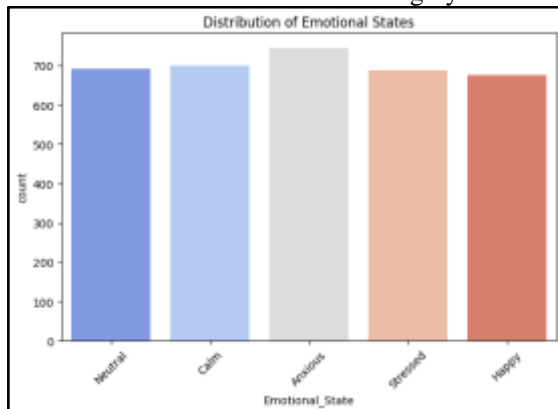
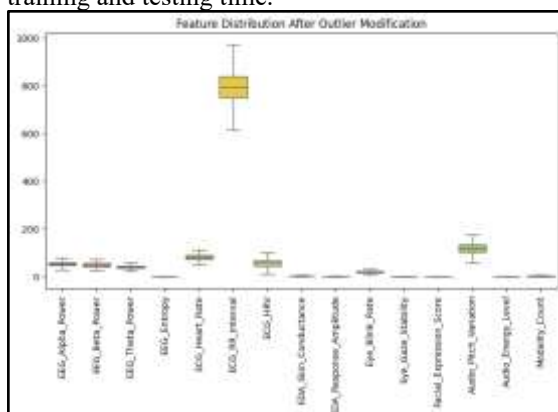


Fig. 9. Distribution Analysis of Emotions

Visualization of emotional states depicts a comparatively equal spread between Calm, Happy, Neutral, Stressed and Anxious. This equalized representation makes sure that each emotional category has enough examples to the classifier, which enhances its ability to generalize during both training and testing time.



10.48047/jocaaa.2024.33.04.56

Fig. 10. Distribution of features when there is an outlier modification analysis

The boxplot diagram (outliers) validates the effective handling of outliers through the IQR-based modification strategy. Cases that give extreme values of features have been clamped but unnatural fluctuations exist in physiological signals. This preprocessing stage enhances the stability of the data and makes sure that data abnormality does not have adverse impacts on the performance of machine learning models.

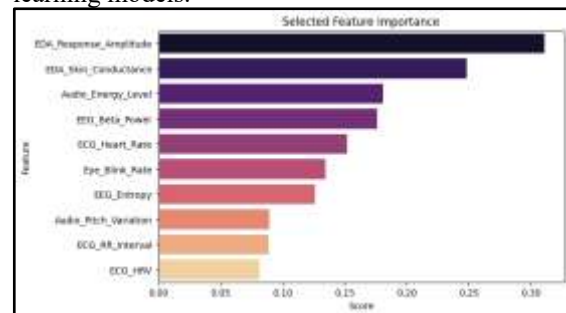


Fig. 11. Selected Feature Importance Analysis

According to the feature importance graph, EDA response amplitude, data, skin conductance, the EEG beta power, and the ECG heart rate are very influential features. The above features yield important information to predict the affect, indicating the effectiveness of the feature selection process in minimizing the redundant information.

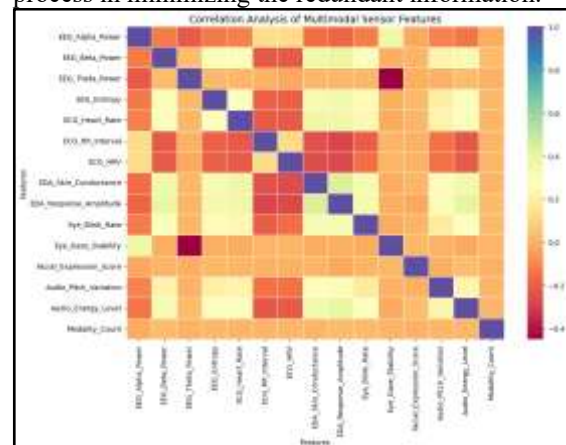


Fig. 12. Multimodal Feature Summative Correlation Analysis

Correlation heat map shows the effectiveness of various physiological and behavior features related to each other. The visualization indicates the different levels of EEG, ECG, EDA, visual, and audio features correlations, demonstrating that multiple modalities are capable of delivering complementary information on emotions. This promotes the need to establish a single modality adaptation framework.

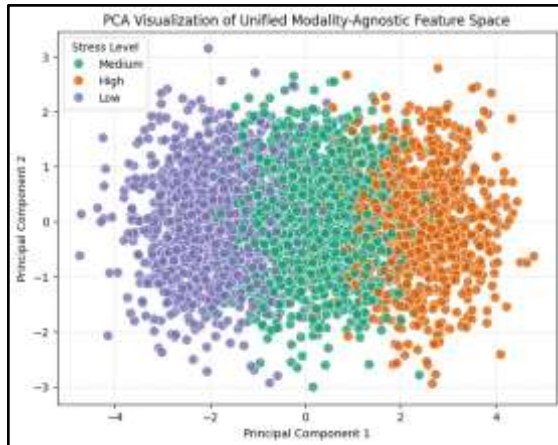


Fig. 13. Unified Feature Space Analysis based on PCA

The figure Feature Space Analysis based on PCA indicates how transformed features without modality are dispersed in a two-dimensional reduced space. The clustering behavior of the various stress groups shows that the suggested feature representation shows significant variation of emotions in them involving various sensing modalities.

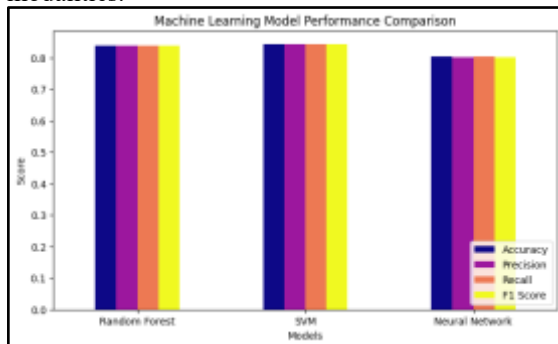


Fig. 14. Comparison of the performance of machine learning models

The comparative study of the two models suggests that the performance of the two random forest and SVM models are higher in terms of classification than the neural network model. The findings confirm that the unified modality-independent feature representation can serve well as a learner in standard methods of machine learning, and the offerings in the proposed framework are effective.

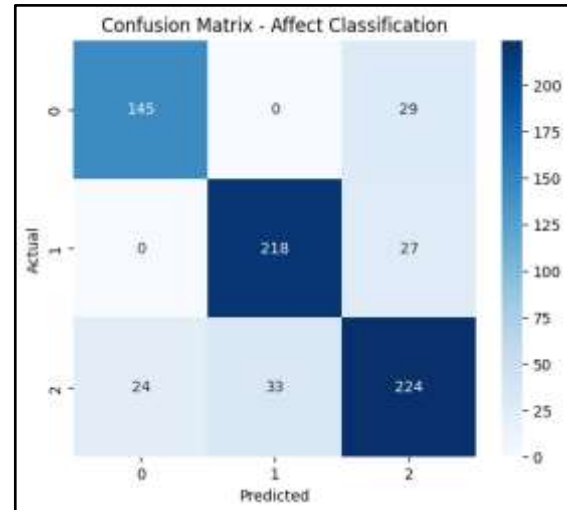


Fig. 15. Confusion Matrix Analysis

The confusion matrix illustrates the ability of the trained affect inference model to classify. The value of the diagonal is increased which shows correct prediction of various types of stress. Misclassification does occur to some extent between adjacent emotional states, which again makes sense, considering the resemblances in physiological reactions.

5-Fold Macro-F1 Scores:
[0.82513875 0.83832667 0.85842734 0.84100175 0.82092937]
Mean Macro-F1: 0.8367647773466622
Standard Deviation: 0.013233160568869197

Fig. 16. 5-Fold Cross-Validation

The dataset is evaluated using stratified 5-fold cross-validation to ensure reliable performance estimation. The data was divided into five equal subsets, where each fold was used once as a testing set while the remaining four folds are used for training. The final performance was reported as the average and standard deviation of all five folds.

Discussion

TABLE I. Summary of Model Comparison

Evaluati on Metric	Random Forest	SVM	Neural Network
Accurac y	0.838571	0.842857	0.802857
Precision	0.838398	0.842456	0.801726
Recall	0.838571	0.842857	0.802857
F1 Score	0.838394	0.842605	0.802108

The comparison of the performance reveals that the Support Vector Machine (SVM) has had the best classification compared to all other considered models, having a classification score of 84.28% and an F1-score of 84.25. Random Forest is also found

to compete as it yielded about 83.86% accuracy demonstrating good learning potential through the modality-independent feature representation. The Neural Network model also achieved relatively low results with a 80.28% performance, which could be explained by the size of the synthetic data and the complexity of features. In general, the findings confirm that the proposed framework of modality-agnostic features can actually work to achieve affect classification on unified multimodal-based features, where SVM shows the most stable predictive ability. Individual modality models (EEG, ECG, EDA, facial and audio) are tested individually in order to quantify the independent contribution to the performance of stress-level classification.

VII. LIMITATIONS AND FUTURE WORK

Limitations

Sensor Variability and Real World Noise:

Though the framework is implemented in such a way that it can accommodate various sensing modalities, differences in the quality of the sensor, the environmental phenomena, and other physiological variations might influence the model in terms of reliability in the real world.

Limited Personalization Capability:

The emotive reactions of various people to situations differ greatly. The existing framework is generalized on the prediction of affect however it needs additional study on individualized adaptation methods to enhance individualized emotional state of users.

Future Direction

Further studies will be directed towards improving the flexibility and intelligence of the suggested modality-agnostic framework with sophisticated deep learning technologies like transformers, self-supervised learning, and domain adaptation algorithms. Inclusion of continuous learning processes can empower the model to accommodate the characteristics of different users and the emerging environments. The use of wearable technology, mobile platforms, and immiscible systems to deploy in real-time will be investigated further [30]. To facilitate secure affect recognition, privacy-preserving methods like federated learning may be included as well. Further, testing of the framework using different populations and application settings will aid in enhancing robustness, scalability, and realistic application.

VII. CONCLUSION

This study introduces a modality-independent affect inference system, which separates emotional state forecasting to the sensing modality. The proposed solution enhances the versatility and reuse of affective computing systems by adding the representation of unified features and the layer of affect classification that are rendered independent. Experimental assessment proves that machine learning models can utilize multimodal features to learn the stress and emotional state and are

dependable enough to classify those. The framework tackles sensor dependencies on affect models by supporting cross-platform implementation in VR, mobile, and wearable and web-based systems. The proposed architecture lays the groundwork to come up with scalable and adaptable emotional intelligence solutions in future human-computer interaction apps.

REFERENCE

- [1] Comas, J., Aspandi, D. and Binefa, X., 2020, November. End-to-end facial and physiological model for affective computing and applications. In *2020 15th IEEE international conference on Automatic Face and Gesture Recognition (FG 2020)* (pp. 93-100). IEEE.
- [2] Poria, S., Hazarika, D., Majumder, N., Naik, G., Cambria, E. and Mihalcea, R., 2019, July. Meld: A multimodal multi-party dataset for emotion recognition in conversations. In *Proceedings of the 57th annual meeting of the association for computational linguistics* (pp. 527-536).
- [3] Bulagang, A.F., Weng, N.G., Mountstephens, J. and Teo, J., 2020. A review of recent approaches for emotion classification using electrocardiography and electrodermography signals. *Informatics in Medicine Unlocked*, 20, p.100363.
- [4] Goh, G., Cammarata, N., Voss, C., Carter, S., Petrov, M., Schubert, L., Radford, A. and Olah, C., 2021. Multimodal neurons in artificial neural networks. *Distill*, 6(3), p.e30.
- [5] Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J. and Krueger, G., 2021, July. Learning transferable visual models from natural language supervision. In *International conference on machine learning* (pp. 8748-8763). PmLR.
- [6] Hu, Z., Lou, S., Xing, Y., Wang, X., Cao, D. and Lv, C., 2022. Review and perspectives on driver digital twin and its enabling technologies for intelligent vehicles. *IEEE Transactions on Intelligent Vehicles*, 7(3), pp.417-440.
- [7] Zucco, C., Calabrese, B. and Cannataro, M., 2019, July. Emotion mining: From unimodal to multimodal approaches. In *international workshop on brain-inspired computing* (pp. 143-158). Cham: Springer International Publishing.
- [8] Roth, D., Bente, G., Kullmann, P., Mal, D., Purps, C.F., Vogeley, K. and Latoschik, M.E., 2019, November. Technologies for social augmentations in user-embodied virtual reality. In *Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology* (pp. 1-12).
- [9] Toyoguchi, S., Coronado, E. and Venture, G., 2023. A Human-Centered and Adaptive Robotic System Using Deep Learning and Adaptive Predictive Controllers. *Journal of Robotics and Mechatronics*, 35(3), pp.834-843.
- [10] Lian, H., Lu, C., Li, S., Zhao, Y., Tang, C. and Zong, Y., 2023. A survey of deep learning-based

multimodal emotion recognition: Speech, text, and face. *Entropy*, 25(10), p.1440.

- [11] Westera, W., Prada, R., Mascarenhas, S., Santos, P.A., Dias, J., Guimarães, M., Georgiadis, K., Nyamsuren, E., Bahreini, K., Yumak, Z. and Christyowidiasmoro, C., 2020. Artificial intelligence moving serious gaming: Presenting reusable game AI components. *Education and Information Technologies*, 25(1), pp.351-380.
- [12] Vaiani, L., La Quatra, M., Cagliero, L. and Garza, P., 2022, October. Viper: Video-based perceiver for emotion recognition. In *Proceedings of the 3rd International on Multimodal Sentiment Analysis Workshop and Challenge* (pp. 67-73).
- [13] Liang, P.P., Lyu, Y., Fan, X., Wu, Z., Cheng, Y., Wu, J., Chen, L., Wu, P., Lee, M.A., Zhu, Y. and Salakhutdinov, R., 2021. Multibench: Multiscale benchmarks for multimodal representation learning. *Advances in neural information processing systems*, 2021(DB1), p.1.
- [14] Smoleń, M. and Augustyniak, P., 2020. Assisted living system with adaptive sensor's contribution. *Sensors*, 20(18), p.5278.
- [15] Abbasi, N., Nouri, H., Didan, K., Barreto-Muñoz, A., Chavoshi Borujeni, S., Salemi, H., Opp, C., Siebert, S. and Nagler, P., 2021. Estimating actual evapotranspiration over croplands using vegetation index methods and dynamic harvested area. *Remote Sensing*, 13(24), p.5167.
- [16] Pranga, J., Borra-Serrano, I., Aper, J., De Swaef, T., Ghesquiere, A., Quataert, P., Roldán-Ruiz, I., Janssens, I.A., Ruyschaert, G. and Lootens, P., 2021. Improving accuracy of herbage yield predictions in perennial ryegrass with UAV-based structural and spectral data fusion and machine learning. *Remote Sensing*, 13(17), p.3459.
- [17] Lang, C.E., Barth, J., Holleran, C.L., Konrad, J.D. and Bland, M.D., 2020. Implementation of wearable sensing technology for movement: pushing forward into the routine physical rehabilitation care field. *Sensors*, 20(20), p.5744.
- [18] Stramska, M. and Aniskiewicz, P., 2019. Recent large scale environmental changes in the Mediterranean sea and their potential impacts on *Posidonia oceanica*. *Remote Sensing*, 11(2), p.110.
- [19] Hickey, B.A., Chalmers, T., Newton, P., Lin, C.T., Sibbritt, D., McLachlan, C.S., Clifton-Bligh, R., Morley, J. and Lal, S., 2021. Smart devices and wearable technologies to detect and monitor mental health conditions and stress: A systematic review. *Sensors*, 21(10), p.3461.
- [20] Persiani, S.G., Kobas, B., Koth, S.C. and Auer, T., 2021. Biometric data as real-time measure of

physiological reactions to environmental stimuli in the built environment. *Energies*, 14(1), p.232.

- [21] He, Z., Li, Z., Yang, F., Wang, L., Li, J., Zhou, C. and Pan, J., 2020. Advances in multimodal emotion recognition based on brain-computer interfaces. *Brain sciences*, 10(10), p.687.
- [22] Yang, K., Wang, C., Gu, Y., Sarsenbayeva, Z., Tag, B., Dingler, T., Wadley, G. and Goncalves, J., 2021. Behavioral and physiological signals-based deep multimodal approach for mobile emotion recognition. *IEEE Transactions on Affective Computing*, 14(2), pp.1082-1097.
- [23] Bota, P.J., Wang, C., Fred, A.L. and Da Silva, H.P., 2019. A review, current challenges, and future possibilities on emotion recognition using machine learning and physiological signals. *IEEE access*, 7(99), pp.140990-141020.
- [24] Egger, M., Ley, M. and Hanke, S., 2019. Emotion recognition from physiological signal analysis: A review. *Electronic Notes in Theoretical Computer Science*, 343, pp.35-55.
- [25] Tan, C., Ceballos, G., Kasabov, N. and Puthanmadam Subramaniyam, N., 2020. Fusionsense: Emotion classification using feature fusion of multimodal data and deep learning in a brain-inspired spiking neural network. *Sensors*, 20(18), p.5328.
- [26] Gao, Y., Wang, L., Liu, J., Dang, J. and Okada, S., 2023. Adversarial domain generalized transformer for cross-corpus speech emotion recognition. *IEEE Transactions on Affective Computing*, 15(2), pp.697-708.
- [27] Rescigno, M., Spezialetti, M. and Rossi, S., 2020. Personalized models for facial emotion recognition through transfer learning. *Multimedia Tools and Applications*, 79(47), pp.35811-35828.
- [28] Dresvyanskiy, D., Ryumina, E., Kaya, H., Markitantov, M., Karpov, A. and Minker, W., 2022. End-to-end modeling and transfer learning for audiovisual emotion recognition in-the-wild. *Multimodal Technologies and Interaction*, 6(2), p.11.
- [29] Qiu, S., Cui, X., Ping, Z., Shan, N., Li, Z., Bao, X. and Xu, X., 2023. Deep learning techniques in intelligent fault diagnosis and prognosis for industrial systems: A review. *Sensors*, 23(3), p.1305.
- [30] Marín-Morales, J., Llinares, C., Guixeres, J. and Alcañiz, M., 2020. Emotion recognition in immersive virtual reality: From statistics to affective computing. *Sensors*, 20(18), p.5163.