International Journal on Advanced Science Engineering Information Technology

Vol.13 (2023) No. 6
ISSN: 2088-5334

Regression-based Analytical Approach for Speech Emotion Prediction based on Multivariate Additive Regression Spline (MARS)

Budi Triandi [a,*], Syahril Efendi [b], Herman Mawengkang [c], Sawaluddin [c]

[a] *Doctoral Program, Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Medan, 20155, Indonesia*
[b] *Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Medan, 20155, Indonesia*
[c] *Faculty of Mathematics and Sciences, Universitas Sumatera Utara, Medan, 20155, Indonesia*

Corresponding author: [*]buditriandi@students.usu.ac.id

***Abstract*—Using regression analysis techniques for speech-emotion recognition (SER) is an excellent method of resource efficiency. The labeled speech emotion data has high emotional complexity and ambiguity, making this research difficult. The maximum average difference is used to consider the marginal agreement between the source and target domains without focusing on the distribution of the previous classes in the two domains. To address this issue, we propose emotion recognition in speech using a regression analysis technique based on local domain adaptation. The results of this study show that the model's generalization ability with the function of the local additive method is very good for improving speech emotion recognition performance. Even though it provides excellent benefits in resource efficiency, regression analytical techniques are rarely used in the SER field; however, we believe this method is the best solution for SER problems. Using the Multivariate Additive Regression Spline, this study developed a predictive model for the existence of angry and non-angry emotions (MARS). Using probability analysis of error values, this approach can overcome regression on data that is not typically distributed. This method yields an ideal basis function that significantly impacts changes in emotional form. This study generates a prediction model with a Mean Square Error (MSE) of 0.0130, a Generalized Cross Validation (GCV) value of 0.0062, and a R Square (RSQ) value of 0.9721, yielding test results with a 97% accuracy rate.**

***Keywords*— Speech emotion; multivariate additive regression spline; regression analytic; predictive model; generalized cross validation.**

Manuscript received 23 Jan. 2023; revised 29 Apr. 2023; accepted 21 Aug. 2023. Date of publication 31 Dec. 2023.

I. INTRODUCTION

Speech emotion recognition (SER) has grown in popularity as a research topic in recent years, and it is critical for various applications, including the development of intelligent devices and tools for measuring mental states in clinical settings [1]. Emotion prediction systems typically comprise a frontend that extracts features—emotion-related material extracted from utterances—and a backend that records the relationship between features and emotion labels. Many different frontends for emotion prediction systems have been developed over the years [2]–[4], with the eGeMAPS acoustic feature set [5]–[7] being one of the most widely used. Emotional labels, typically one of three major representations of human emotional states supported by psychologists, significantly impact backend design. This is a categorical, dimensional, and assessment-based representation [8]. The main focus of the work is the representation of the emotional dimension, which tends to capture the time dynamics of emotional interaction more accurately [9]. To use an acceptable regression model as a backend, emotions are represented by a bipolar circumplex model comprised of arousal (activation level) and valence (pleasure level).

Although many other regression modeling techniques have been used to predict emotions, Support Vector Regression [10] and Long-Short Memory Recurrent Neural Network (LSTM-RNN) [11] are the most popular models because they capture the temporal dynamics of emotion. Furthermore, relevance vector machine (RVM) [12] and Gaussian mixed regression (GMR) [13] both successfully predict emotions. However, these models: 1) involve complicated non-linear models with numerous model parameters that need to be estimated, such as LSTM-RNN and GMR; 2) have an appreciable increase in computation time with a comparatively small increase in the size of the training data, such as SVR; or 3) ignore temporal dependencies in emotion labels, like the majority of RVM and SVR based systems. The non-interpretability of this non-linear model further hinders the development of an emotion prediction system [14]. A linear model on the back will make interpretation more accessible, but it is insufficient to capture the connection between speech and emotion. It is challenging to predict the presence of potential emotions given the finding from earlier studies that voice signals have complex voice feature data with a large amount of data and contain uncertain parameters that cause problems [15]. Research in the SER field is not typically distributed, and it can be challenging to predict the voice feature data for a particular type of emotion [16].

Meanwhile, linear models for speech emotion recognition, such as multivariate linear regression [17], focus only on the emotional content of the current time step, whereas linear models that account for temporal dynamics [18] focus on the emotional content of the entire speech. The data used in SER research does not have a distribution, so a nonparametric regression approach, such as the Multivariate Adaptive Regression Spline (MARS) nonparametric regression method, is required to determine a pattern of relationship between speech and emotion [19]. MARS is a function in a nonparametric regression model that can determine patterns of relationships or relationships between unknown response variables and predictor variables.

Statistical methods have also been used to classify various emotions in speech [20]. Statistical features have an extraordinary ability [21] to distinguish emotions using many degrees of standard deviation, but accuracy in SER systems is affected by many factors [22], such as the platform used (data set), the type of feature, the feature selection algorithm used, the classifier, and prediction [23]. In 2018, a study aimed to evaluate the effect of tone-related features on the detection of various emotions from specific children's speech signals using multivariate analysis. Decision trees were used for feature reduction and producing a good level of accuracy [24]. The results show that the selected tone-based features have relatively great power in fear recognition, with the highest accuracy reaching 78.7% using the KNN algorithm [25], but the complexity of the features causes high data dimensions that allow overfitting to occur in learning. Several studies in speech emotion recognition (SER) use sensing techniques and predictions of audio signal features to figure out how well emotions can be guessed from speech [26].

II. MATERIALS AND METHOD

This section will introduce the work of developing a predictive model to recognize emotional states (anger and neutral) based on MARS model development, using a multivariate regression spline analysis approach as the foundation for establishing SER prediction models and producing new methods of speech emotion recognition.

A. Developing the SER Model with a Regression-based Analytical Approach

The proposed emotion recognition model was formed using a nonparametric regression analysis method. The regression method is rarely used to predict the capture of sound features and emotion labels [27]. Consistency between sound characteristics and studied emotion labels can result in a more discriminative model for predicting emotion labels [28]. Lastly, the regression analytic approach is utilized for model prediction, and algorithm optimization will be employed to determine the optimal function basis [29]. There are numerous methods for determining the relationship pattern between predictor variables and responses when attempting to predict speech emotions for which the regression curve is unknown [30]. The spline model has unique and highly effective statistical and visual interpretations. The steps for establishing the MARS model must first determine the maximum Functional Basis (FB), Maximum Interaction (MI), and Minimum Observation (MO) between nodes to obtain the optimal model with the least amount of general cross-validation (GCV).

B. MARS-Based Model Development

MARS uses a sequential algorithm [31] to determine the relationship between the response variable and the predictor variable, as explained by the basics function (BF) [32]. Typically, the selected basis function is a continuous polynomial derivative at each vertex, whereas the fundamental function in each zone is parametric. Polynomial slices in MARS are splined and have segmented properties; they are more flexible than ordinary polynomials and can adapt to local characteristics to make a function or data set more efficacious [33]. In general, the MARS model can be described as follows:

Basis function equations:

$$f(x) = a_o + \sum_{m=1}^{M} a_m B_m(x) \quad (1)$$

This can be simplified to:

$$\hat{f}(x) = a_o + a_1 BF_1 + a_2 BF_2 + \cdots + a_m BF_m \quad (2)$$

$$= a_o + \sum_{m=1}^{M} a_m B_{mi}(x, t) \quad (3)$$

Where:

$$B_{mi}(x,t) = \prod_{k=1}^{K_m} \left[S_{K_m} \cdot \left(x_{vN_{(k,m)}} - t_{km} \right) \right]_+ \quad (4)$$

With:

$x_{v(k.m)} \in \{x_j\}_{j=1}^{p}$ and $t_{km} \in \{x_{v(k.m)i}\}_{i=1}^{n}$

Modification of the basis function equation:

$$f(x_{1i}, x_{2i}, \ldots, x_{pi}) = a_0 + \sum_{m=1}^{M} a_m \prod_{k=1}^{K_m} \left[S_{km} \cdot \left(x_{vN_{(k,m)}} - t_{km} \right) \right]_+ ; \quad (5)$$

$$i = 1, 2, \ldots, n$$

MARS employs piecewise linear Basis Functions (BFs) expansion on a dataset-by-dataset basis to offer a flexible method for high-dimensional nonparametric regression. For BFs $(y - \tau)_+$ and $(\tau - y)_-$

$$(y - \tau)_+ = \begin{cases} x - \tau, & \text{if } x > \sigma \\ 0 & \text{,otherwise} \end{cases}$$

$$(\tau - y)_- = \begin{cases} \tau - x, & \text{if } x < \sigma \\ 0 & \text{,otherwise} \end{cases}$$

So that:

$$[+(x - \tau)]_+, [-(x - \tau)]_+ \quad (6)$$

The MARS method focuses on solving the problem of the truncated spline's weakness; in deep learning, the MARS work scheme belongs to the feed-forward category [34]. MARS determines essential parameters based on covariance and correlation between parameters. Using the elimination of a subset of all features, an intelligent algorithm can determine the attribute's weight based on its contribution, making this method more flexible for prediction. Therefore, it is improbable that the learning data process will fail because MARS uses the selected attribute on the associated basis function. Furthermore, because MARS employs filtered important attributes to reduce processing time and significantly increase processing speed, the learning model on MARS can accurately and effectively analyze each feature [35], [36].

C. Optimal Base Function Selection in MARS Models

For optimal basis selection, both forward and backward stepwise procedures are possible [37]. The purpose of the forward stepwise method is to construct a model by adding truncated spline basis functions (knots and interactions), which will result in a model containing the greatest possible number of basis functions [38]. The forward stepwise process is a hidden layer in MARS, and the value obtained from forward stepwise will be used to determine if backward stepwise modeling is feasible.

1) Forward Stepwise: Forward stepwise is enabled on MARS to select the maximum number of basic functions. $\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{f}(x_i, \alpha))^2$ is a function used as a basis function selection criterion to obtain the minimum average sum of square residuals. To obtain the minimum ASR value, divide the number of parameters for the C(M) model by n observations minus 1 of this function, which is the ASR dividing value. The following are the forward stepwise steps in the process of selecting the basis function:

Initialization and constant-coefficient conjecture

$$B_0 = 1$$

Assuming there are $K+1$ basis functions $B_0, B_1(x), \dots, B_k(x)$, add two new basis functions:

$$B_{k+1}(x) = B_1(x)[+(x_{1(k,l)} - x^*_{1(k,l)})]^m_+$$

$$B_{k+1}(x) = B_1(x)[-(x_{1(k,l)} - x^*_{1(k,l)})]^m_+$$

Where $B_k(x)$ is the initial function basis (parent), $x_{1(k,l)}$ is a variable not included in the basis function $B_k(x)$, and $x*_{1(k,l)}$ is the knot point.

$$x^*_{i(k,l)} \in \{x_{i(k,l)}\}, i = 1, \dots, n \tag{7}$$

Minimum determinants according to GCV standards

$$GCV = \frac{1}{n}\frac{\sum_{i=1}^{n}\left(y_i - \hat{f}\alpha(x_i,\alpha)\right)^2}{1-(\tilde{C}(\alpha)/n)^2} \tag{8}$$

The addition of the basis function is still being done in order to ensure that the maximum value of M_{max} represents the basis for the K function.

2) Backward Stepwise: The backward stepwise step in MARS is used to simplify models with overlapping regions by removing the basis function and producing a model with continuous derivatives. Using the recursion partition regression algorithm (RPR) on a set of basis functions $J^* = \{1,2,\dots,M_{max-1}\}$ the sum of the basis functions from the previous step forward stepwise, removing the basis function on MARS does not result in a hole in the predictor variable space (as long as the constant basis is not removed). The following is a list of the backward stepwise steps involved in getting rid of the basis function:

Initials of $J^* = \{1,2,\dots,M_{max-1}\}$; $K^* \leftarrow K$. Lack of fit (LOF) will be assigned to non-conforming functional bases.

$$lof \leftarrow min_{\{a_j|j\in J*\}}\, LOF\left[\sum_{j\in J^*}\alpha_k B_k\,(x)\right] \tag{9}$$

Each iteration outside the for loop causes the elimination of one basis function, while each iteration inside the for loop selects a basis function. The base constant $B_k\,(x) =$ will not be eliminated during the backward stepwise elimination process. This algorithm can produce up to $B_k\,(x) =$ model lines. Each succeeding sequence has one fewer basis function than its predecessor.

D. MARS Model Estimation And GCV

The MARS model will select a sub-model to minimize the prediction of residual estimation (chosen based on optimal function basis) according to the generalized cross-validation (GCV) criteria in equation (8). $\hat{\alpha}$ is the parameter to be estimated via the penalized residual sum of squares problem (PRSS). The following is a PRSS model with M_{max} Basis Functions (BFs) utilizing the MARS method:

$$PRSS \approx \left(y_i - B(\tilde{d})\theta\right)_2^2 + \sum_{m=1}^{M_{max}\Sigma\sum_{i=1}^{(N+1)^{Km}} L^2_{im}\theta^2_{im}} \alpha_m \tag{10}$$

Where,

The $(N \times (M_{max}+1))$- matrix is denoted by $B(\tilde{d}) = \left(B(\tilde{d}_i), \dots B(\tilde{d}_N)\right)^T$, the Euclidean form is indicated by ||.||2, and the number L_{im} is what defines the roots.

For the probability model based on the equation below:

$$P(Y=1|X=x) = \pi(x) = \frac{e^{g(x)}}{1+e^{g(x)}} \tag{11}$$

probability transformation of the MARS model

$$in\left[\frac{\pi(x)}{1-\pi(x)}\right] = a_0 + \sum_{m=1}^{M} a_m \prod_{k=1}^{K_m}\left[S_{km}\cdot\left(x_{vN_{(k,m)}} - t_{km}\right)\right]_+ \varepsilon_i \tag{12}$$

III. Results and Discussion

This research was carried out to develop a predictive model by employing a regression analytic strategy with the MARS model and applying it to a collection of uncertain sound feature data. In addition to allowing for greater flexibility in data analysis regression problems, splines are also developed to accommodate local data and exhibit distinct behavior in each sub-interval. Because spline is a piecewise polynomial, i.e., polynomials with segmented properties, this indicates that the spline regression approach that we propose in the field of speech emotion recognition with high data dimensions is ideally suited for the problem of predicting specific forms of emotion. This segmented property offers greater flexibility than conventional polynomials, allowing for more effective adaptation to the characteristics of a function or database.

A descriptive analysis of the initial phases of data exploration to gain an overview of the employed data. This study employs the RAVDESS dataset consisting of 120 WAV-format sound files along with sound feature extraction and selection techniques employing Mel Frequency Cepstral Coefficients (MFCC) to generate 180 sound features $\{x_1, x_2, \ldots, x_n; x_n = x_{180}\}$. This leads to two forms of emotion, 'angry' = 1 and not angry 'neutral' = 2, the sound feature data produced is not normally distributed, so a nonparametric regression analysis approach is highly appropriate.

The next step is to determine the Basis Function (BF), Maximum Interaction (MI), and Minimum Observation (MO) values after determining the shape of the data distribution through data exploration. For values of BF = 36, MI, and MO in this study using values 1, 2, and 3, more than three results in a model that is extremely complex and less accurate [36]. The stages involved in determining the value of BF, MI, and MO are shown in the table below.

TABLE I
MI AND MO DETERMINATION

BF	MI	MO	GCV	MSE	R2
36	1	None	0.0172	0.0079	0.9643
36	1	0	0.0164	0.0062	0.9720
36	1	1	0.0173	0.0061	0.9723
36	1	2	0.0130	0.0062	0.9721
36	1	3	0.0161	0.0065	0.9706
36	2	None	0.0172	0.0079	0.9643
36	2	0	0.0164	0.0062	0.9720
36	2	1	0.0173	0.0061	0.9723
36	2	2	0.0130	0.0062	0.9721
36	2	3	0.0161	0.0065	0.9706
36	3	None	0.0172	0.0079	0.9643
36	3	0	0.0164	0.0062	0.9720
36	3	1	0.0173	0.0061	0.9723
36	3	2	0.0130	0.0062	0.9721
36	3	3	0.0161	0.0065	0.9706

Table 1 demonstrates that the minimum observation significantly influences the 36 basis functions in determining the minimum GCV value. In contrast, the maximum value of interaction does not affect the formation of speech emotion prediction models, so MI = 1, MO = 2 with BF = 36 function basis can be used.

MARS modeling's development prediction model employs Google Colab and the Python 3.8 programming language. As shown in Fig. 1, the MARS regression of the training data, with the intercept value and the maximum variable coefficient for each predictor variable for the response variable and the maximum variable coefficient for each predictor variable.

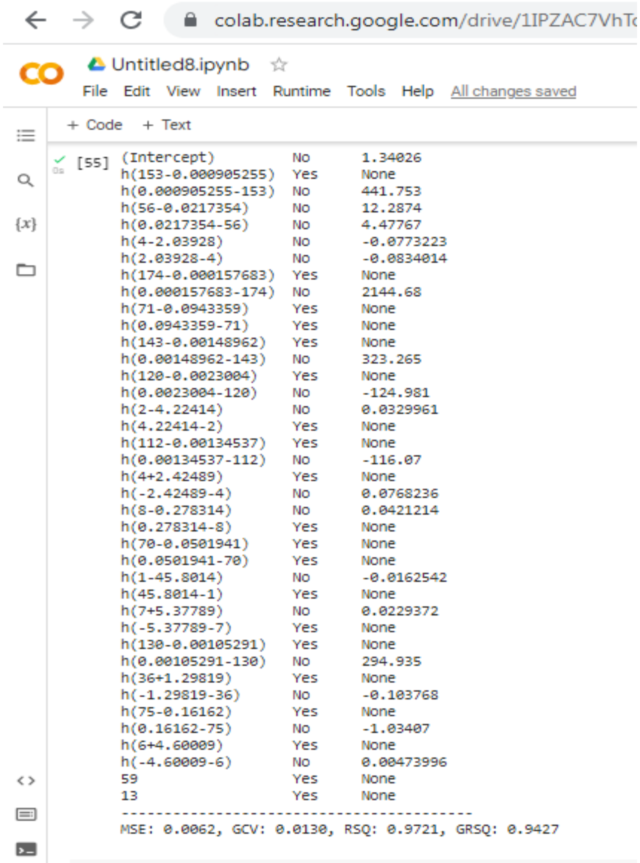

Fig. 1 MARS regression training data

Based on the forward and backward stepwise, the MSE value of 0.0062 is obtained for the sought optimal model formation function. With a minimum GCV of 0.0130, RSQ of 0.9721, and GRSQ of 0.9427, the estimation results are as follows:

$$
\begin{aligned}
f(\underline{x}) = 1.34026 &+ 2144.68(0.00157683 - X174) \\
&+ 441.753(0.000905255 - X153) \\
&+ 323.265(0.00148962 - X143) \\
&+ 294.935(0.00105291 - X130) \\
&+ 12.2874(X56 - 0.0217354) \\
&+ 4.47767(0.0217354 - X56) \\
&+ 0.0768236(2.03928 - X4) \\
&+ 0.0421214(X8 - 0.278314) \\
&+ 0.0329961(-2.42489 - X4) \\
&+ 0.0229372(X7 + 5.37789) \\
&+ 0.00473996(-4.60009 - X6) \\
&- 0.0162542(X1 - 45.8014) \\
&- 0.0773223(X4 - 2.03928) \\
&- 0.0834014(X2 - 4.22414) \\
&- 0.103768(-1.29819 - X36) \\
&- 1.03407(0.16162 - X75) \\
&- 116.07(0.00134537 - X112) \\
&- 124.981(0.0023004 - X120)
\end{aligned}
$$

Then, the results of the optimal estimation of the accommodating weight can be interpreted using equation (11), an example of estimation/interpretation of the optimal estimation of weight:

Approximate example:
Maximum coefficient value = 2144.68 with a base function B8 = (0.000157683-X174)

$$\pi(x) = \frac{\exp\,(2144.68 + 1.34026)}{1 + \exp\,(2144.68 + 1.34026)} = 0.9995$$

Assuming that the other independent variables remain constant, the emotion of anger is determined if the added value of the feature less than 1 (X174 < 1) has the potential to affect the emotional form of 0.9995, and the coefficient B8 (function 8th basis) increases the effect on changes in the emotional form of anger by 2144.86. In contrast, if the value of the added feature is greater than 1 (X174 > 1), the B8 coefficient (8th function basis) is not significant, and its value is 0. The optimal function basis estimate results for estimating the probability of emotional shape variables are presented in the table below.

TABLE II
FUNCTIONAL BASIS FOR ESTIMATING THE PROBABILITY OF VARIABLES OF EMOTIONAL SHAPE

Basis Function	**Anger Emotional Possibility $\pi(x)$**	**Not Anger Emotional Possibility (neutral) $1 + \pi(x)$**
B2 = (0.000905255-X153)	0.9977	0.0023
B3 = (X56-0.0217354)	0.9316	0.0684
B4 = (0.0217354-X56)	0.8533	0.1467
B5 = (X4-2.03928)	0.5581	0.4419
B6 = (2.03928-X4)	0.5569	0.4431
B8 = (0.000157683-X174)	0.9995	0.0005
B12 = (0.00148962-X143)	0.9969	0.0031
B14 = (0.0023004-X120)	1.0089	-0.0089
B15 = (X2-4.22414)	0.5786	0.4214
B18 = (0.00134537-X112)	1.0088	-0.0088
B20 = (-2.42489-X4)	0.5863	0.4137
B21 = (X8-0.278314)	0.5803	0.4197
B25 = (X1-45.8014)	0.5697	0.4303
B27 = (X7+5.37789)	0.5768	0.4232
B30 = (0.00105291-X130)	0.9966	0.0034
B32 = (-1.29819-X36)	0.5529	0.4471
B34 = (0.16162-X75)	0.2344	0.7656
B36 = (-4.60009-X6)	0.5718	0.4282

From the estimated results in Table 2 the learning process or learning of the input variables or variable xi. This study produces a speech emotion recognition prediction model with a number of variables that are recognized as variables that influence changes in emotional form as much as 15 of 180 voice features Mel Frequency Cepstral Coefficient (MFCC) with an accuracy rate of 97.21%. The outcomes of learning test data and training data from predictive models were constructed utilizing the multivariate additive regression spline (MARS) regression approach, acquiring four prediction errors and 116 accurate predictions, as shown in Table 3.

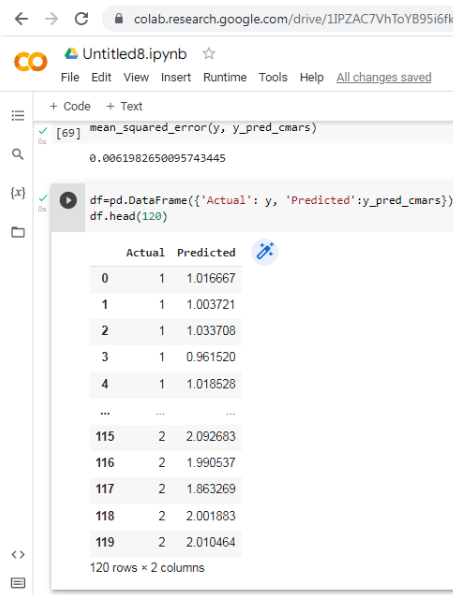

Fig. 2 Prediction Results of Actual Emotional Data

TABLE III
ACCURACY LEVEL MEASUREMENT

		Actual Value	
		TRUE	FALSE
Predicted Value	POSITIVE	70	1
	NEGATIVE	46	3

$$pression = \frac{TP}{TP+FP} \times 100 \qquad \frac{70}{70+1} \times 100 = 99\%$$

$$recall = \frac{TP}{TP+FN} \times 100 \qquad \frac{70}{70+3} \times 100 = 96\%$$

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad \frac{70+46}{70+46+1+3} \times 100 = 97\%$$

The accuracy test for the predictive model yields 97% with 99% precision and 96% recall. The prediction model was created for speech emotion recognition using a nonparametric regression analysis approach, and the best model accurately predicts the presence of potential angry, non-angry , or neutral emotions.

IV. CONCLUSION

In this study, we devised a method using nonparametric regression analysis to develop a predictive model for emotion recognition in speech. Generally, sound feature data are coefficient values and lack a data distribution; therefore, nonparametric regression is the appropriate method. Developing a prediction model using the MARS model to examine the relationship between predictors and responses through the basis function is one method that can be used to solve problems in SER. By determining the optimal estimation of the basis function, it is possible to reduce sound features that influence emotional form changes. For future studies, we can extend the nonparametric regression analytic approach to sound feature data types containing noise or spontaneous sound data under various conditions.

ACKNOWLEDGMENT

The first author expresses his deepest gratitude to his supervisors and thanks the Faculty of Computer Science and Information Technology at the University of North Sumatra. Finally, all authors thank all anonymous reviewers for providing valuable and constructive feedback.

REFERENCES

[1] S. Latif, J. Qadir, and M. Bilal, "Unsupervised Adversarial Domain Adaptation for Cross-Lingual Speech Emotion Recognition," in *2019 8th International Conference on Affective Computing and Intelligent Interaction, ACII 2019*, 2019.

[2] E. Lieskovská, M. Jakubec, R. Jarina, and M. Chmulík, "A review on speech emotion recognition using deep learning and attention mechanism," *Electronics (Switzerland)*, vol. 10, no. 10. 2021.

[3] S. Dutta and S. Ganapathy, "Multimodal Transformer with Learnable Frontend and Self Attention for Emotion Recognition," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 6917–6921.

[4] R. Orjesek, R. Jarina, and M. Chmulik, "End-to-end music emotion variation detection using iteratively reconstructed deep features," *Multimed. Tools Appl.*, vol. 81, no. 4, pp. 5017–5031, 2022.

[5] L. Kerkeni, Y. Serrestou, M. Mbarki, K. Raoof, and M. A. Mahjoub, "Speech emotion recognition: Methods and cases study," in *ICAART 2018 - Proceedings of the 10th International Conference on Agents and Artificial Intelligence*, 2018, vol. 2.

[6] P. Baki, H. Kaya, E. Çiftçi, H. Güleç, and A. A. Salah, "A Multimodal Approach for Mania Level Prediction in Bipolar Disorder," *IEEE Trans. Affect. Comput.*, vol. 13, no. 4, pp. 2119–2131, 2022.

[7] R. Li, J. Zhao, J. Hu, S. Guo, and Q. Jin, "Multi-modal fusion for video sentiment analysis," in *Proceedings of the 1st International on Multimodal Sentiment Analysis in Real-life Media Challenge and Workshop*, 2020, pp. 19–25.

[8] V. Vielzeuf, C. Kervadec, S. Pateux, and F. Jurie, "The many variations of emotion," in *Proceedings - 14th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2019*, 2019.

[9] E. Avots, T. Sapiński, M. Bachmann, and D. Kamińska, "Audiovisual emotion recognition in wild," in *Machine Vision and Applications*, 2019, vol. 30, no. 5.

[10] J. Han, Z. Zhang, F. Ringeval, and B. Schuller, "Prediction-based learning for continuous emotion recognition in speech," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2017.

[11] D. Le, Z. Aldeneh, and E. M. Provost, "Discretized continuous speech emotion recognition with multi-task deep recurrent neural network," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2017, vol. 2017-Augus.

[12] Y. Liu, J. She, H. Duan, and N. Qi, "Hybrid Model Based on Maxwell-Slip Model and Relevance Vector Machine," *IEEE Trans. Ind. Electron.*, vol. 68, no. 10, 2021.

[13] D. Wu, P. Yan, Y. Guo, H. Zhou, and J. Chen, "A gear machining error prediction method based on adaptive Gaussian mixture regression considering stochastic disturbance," *J. Intell. Manuf.*, 2021.

[14] Y. H. H. Tsai, M. Q. Ma, M. Yang, R. Salakhutdinov, and L. P. Morency, "Multimodal routing: Improving local and global interpretability of multimodal language analysis," in *EMNLP 2020 - 2020 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, 2020.

[15] R. A. Khalil, E. Jones, M. I. Babar, T. Jan, M. H. Zafar, and T. Alhussain, "Speech Emotion Recognition Using Deep Learning Techniques: A Review," *IEEE Access*, vol. 7, 2019.

[16] L. Tan *et al.*, "Speech emotion recognition enhanced traffic efficiency solution for autonomous vehicles in a 5G-enabled space–air–ground integrated intelligent transportation system," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 3, pp. 2830–2842, 2021.

[17] L. Kerkeni, Y. Serrestou, M. Mbarki, K. Raoof, M. Ali Mahjoub, and C. Cleder, "Automatic Speech Emotion Recognition Using Machine Learning," in *Social Media and Machine Learning*, 2020.

[18] M. B. Akçay and K. Oğuz, "Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers," *Speech Communication*, vol. 116. 2020.

[19] A. Çevik, G. W. Weber, B. M. Eyüboğlu, and K. K. Oğuz, "Voxel-MARS: a method for early detection of Alzheimer's disease by classification of structural brain MRI," *Ann. Oper. Res.*, vol. 258, no. 1, 2017.

[20] D. Li, Y. Zhou, Z. Wang, and D. Gao, "Exploiting the potentialities of features for speech emotion recognition," *Inf. Sci. (Ny).*, vol. 548, 2021.

[21] D. Luo, Y. Zou, and D. Huang, "Investigation on joint representation learning for robust feature extraction in speech emotion recognition," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2018, vol. 2018-Septe.

[22] H. A. Abdulmohsin, H. B. Abdul wahab, and A. M. J. Abdul hossen, "A new proposed statistical feature extraction method in speech emotion recognition," *Comput. Electr. Eng.*, vol. 93, 2021.

[23] M. A. Siddiqui, S. A. Ali, and N. G. Haider, "Reduced Feature Set for Emotion Based Spoken Utterances of Normal and Special Children Using Multivariate Analysis and Decision Trees," *Eng. Technol. Appl. Sci. Res.*, vol. 8, no. 4, 2018.

[24] S. Chebbi and S. Ben Jebara, "On the use of pitch-based features for fear emotion detection from speech," in *2018 4th International Conference on Advanced Technologies for Signal and Image Processing, ATSIP 2018*, 2018.

[25] Mustaqeem, M. Sajjad, and S. Kwon, "Clustering-Based Speech Emotion Recognition by Incorporating Learned Features and Deep BiLSTM," *IEEE Access*, vol. 8, 2020.

[26] A. Bose, C. H. Hsu, S. S. Roy, K. C. Lee, B. Mohammadi-ivatloo, and S. Abimannan, "Forecasting stock price by hybrid model of cascading Multivariate Adaptive Regression Splines and Deep Neural Network," *Comput. Electr. Eng.*, vol. 95, 2021.

[27] Y. Gu *et al.*, "Mutual correlation attentive factors in dyadic fusion networks for speech emotion recognition," in *MM 2019 - Proceedings of the 27th ACM International Conference on Multimedia*, 2019.

[28] S. E. Kahou *et al.*, "EmoNets: Multimodal deep learning approaches for emotion recognition in video," *J. Multimodal User Interfaces*, vol. 10, no. 2, 2016.

[29] S. H. Samareh Moosavi and V. K. Bardsiri, "Poor and rich optimization algorithm: A new human-based and multi populations algorithm," *Eng. Appl. Artif. Intell.*, vol. 86, 2019.

[30] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Computer Science*, vol. 2, no. 3. 2021.

[31] M. Shahbaz, N. Khraief, and M. K. Mahalik, "Investigating the environmental Kuznets's curve for Sweden: evidence from multivariate adaptive regression splines (MARS)," *Empir. Econ.*, vol. 59, no. 4, 2020.

[32] J. R. Leathwick, J. Elith, and T. Hastie, "Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions," *Ecol. Modell.*, vol. 199, no. 2, 2006.

[33] M. H. Ahmadi, B. Mohseni-Gharyehsafa, M. Farzaneh-Gord, R. D. Jilte, R. Kumar, and K. wing Chau, "Applicability of connectionist methods to predict dynamic viscosity of silver/water nanofluid by using ANN-MLP, MARS and MPR algorithms," *Eng. Appl. Comput. Fluid Mech.*, vol. 13, no. 1, 2019.

[34] S. Dargan, M. Kumar, M. R. Ayyagari, and G. Kumar, "A Survey of Deep Learning and Its Applications: A New Paradigm to Machine Learning," *Arch. Comput. Methods Eng.*, vol. 27, no. 4, 2020.

[35] M. Besharati Fard, D. Hamidi, J. Alavi, R. Jamshidian, A. Pendashteh, and S. A. Mirbagheri, "Saline oily wastewater treatment using Lallemantia mucilage as a natural coagulant: Kinetic study, process optimization, and modeling," *Ind. Crops Prod.*, vol. 163, 2021.

[36] J. H. Friedman, "Multivariate adaptive regression splines," *Ann. Stat.*, vol. 19, no. 1, pp. 1–67, 1991.

[37] T. Hastie, R. Tibshirani, and R. Tibshirani, "Best Subset, Forward Stepwise or Lasso? Analysis and Recommendations Based on Extensive Comparisons," *Stat. Sci.*, vol. 35, no. 4, 2020.

[38] X. Ju, J. M. Rosenberger, V. C. P. Chen, and F. Liu, "Global optimization on non-convex two-way interaction truncated linear multivariate adaptive regression splines using mixed integer quadratic programming," *Inf. Sci. (Ny).*, vol. 597, 2022.