# ENHANCING THE ACCURACY IN CLASSIFYING HUMAN EMOTION VIA SPEECH RECOGNITION USING NOVEL MULTILAYER PERCEPTRON COMPARED WITH LOGISTIC REGRESSION

**S.Devakar[1], D. Beulah David[2*]**

**Article History: Received:** 12.12.2022 **Revised:** 29.01.2023 **Accepted:** 15.03.2023

**Abstract**

**Aim:** To enhance accuracy of human emotion classification via speech recognition using Novel Multilayer Perceptron compared with Logistic Regression.
**Materials and Methods:** This study contains 2 groups i.e Novel Multilayer Perceptron (MLP) and Logistic Regression (LR). Each group consists of a sample size of 6. Gpower software is used to determine sample size with pretest power value 0.8 and alpha is 0.05 with the p = 0.138.
**Result:** The Novel Multilayer Perceptron 70.37% more accurate than Logistic Regression 60.99% in classifying human emotion.
**Conclusion:** The Novel Multilayer Perceptron is significantly better than Logistic Regression in classifying human emotion via speech recognition.

**Keywords:** Novel Multilayer Perceptron, Logistic Regression, Speech Emotion Recognition, Emotion Classification, Machine Learning, Accuracy.

[1]Research Scholar, Department of Information Technology, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India. Pincode: 602105.
[2*]Department of Information Technology, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India. Pincode: 602105.

## 1. Introduction

Speech emotion recognition is one of the fastest growing fields in the world. The key to human existence is communication (Hansen et al. 2022). Speech emotion technology was a technology that was used to extract the emotions through their features by the computers. Machine learning techniques play a major role in speech emotion recognition. The extracted feature is used to analyze what kind of emotion it is (Huang et al. 2014). Researchers have raised the importance of feeling in multidisciplinary applications. Analyzing human emotion is very much important in many research areas, which request accurate recognition in uncontrolled situations (Shahsavarani 2018). Analysts have broadly concentrated because of enthusiastic variables, on dynamic. The significance of perceiving feelings for requirements of this present reality use is becoming unavoidable ("Speech Emotion Recognition in Noisy Environment" n.d.). Speech is constantly connected with some type of feelings. For example, anger, happiness, fear and sadness etc are effective ("Recognition of Fear from Speech Using Adaptive Algorithm with MLP Classifier" n.d.). The various applications of this system are in the field of Medical diagnosis, lie detection, learning environments, and human emotional state recognition in call centers.

There are about 475 articles in google scholar and 20 articles in science directly related to this study. Speech signals were analyzed to categorize emotions by studying prosodic features of speech [formant frequencies, entropy, variance, minima, median, linear prediction corrector (LPC)], which were extracted by spectral analysis and feature extraction algorithms (Jaso 2006). Machine learning plays a vital role in speech emotion recognition. The person needs to know the other person's emotion through their speech. For example "This is fantastic" can be categorized as happy or excited (Arena et al. 1998).Emotions are mental activities that are expressed through a variety of manifestations such as speech, facial emotions, body expressions, and so on. Speech emotion recognition is also used in call centers for complaints or marketing in the voice based virtual assistant (Liu et al. 2021). Emotion classification is an important task in speech emotion recognition. The best or worst performing feature is chosen via recursive feature elimination, and a pruned set is executed periodically. The algorithm can be used to make human perception more precise and effective. To proliferate feelings, humans use a variety of data assets in human social cooperation. Emotion classification just goes with traits of the emotion to be specific; non-verbal communication, acoustic data like pitch and sound, eye staring and look complete one another to guarantee that the data is passed on to the audience (Kamaruddin and Wahab 2013). Speech emotion identification is a new topic of interest that is drawing the attention of the research community. To normalize the situation of Speech emotion classification, automatic speech recognition can be put to use. Furthermore, emotion and information influence each other were derived from speech and used to improve human-computer interaction (Satapathy, Bhateja, and Das 2017). These investigations have utilized various indicators for feeling acknowledgment and agreement for what indicators impact feeling acknowledgment is deficient. Past examinations show that prosodic boundaries including power, major recurrence and talking rate are conceivably compelling feeling indicators for sound documents. Voice quality and present moment unearthly indicators have likewise been accounted for as successful indicators. There is additionally significant vulnerability concerning the best calculation for feeling arrangement and regardless of whether one calculation can adequately anticipate feelings from consolidated sound documents of an alternate vocal channel; for instance speech and song (Javaheri, n.d.).

Our institution is passionate about high quality evidence based research and has excelled in various domains (Vickram et al. 2022; Bharathiraja et al. 2022; Kale et al. 2022; Sumathy et al. 2022; Thanigaivel et al. 2022; Ram et al. 2022; Jothi et al. 2022; Anupong et al. 2022; Yaashikaa, Keerthana Devi, and Senthil Kumar 2022; Palanisamy et al. 2022).The study is focusing on enhancing the accuracy of classifying human emotion through speech recognition. The existing system has less accuracy and the dataset used by the researchers have made their own feature extraction. The drawback of the existing systems is extracting less number of emotions at a time. The goal is to get accuracy by extracting more emotion at a single time using a good dataset.

## 2. Materials and Methods

This work is carried out in the Data Analytics Lab, Department of Information Technology at Saveetha School of Engineering. This study consists of two groups i.e where each group consists of 6 samples. The input dataset for the proposed work RAVDESS dataset is collected from Kaggle.com (https://www.kaggle.com/search). The test was done with Gpower of 80%, confidence interval at 95% and enrollment ratio as 1. The RAVDESS was a dataset which consisted of 8 different emotions recorded by 24 actors. Table 1 represents the audio files and recognized emotions from the dataset. The significance value for the algorithm is 0.138.

The dataset consists of two columns. The first column consists (Wav filename) of an audio file in wav format. The second column consists (Recognised emotions) of the number of emotions recognised from the dataset. The second column consists (Extracted emotion) of the emotions which are extracted from the dataset. The dataset consists of 1414 Instances. The independent variables in this method are audio wav files and the dependent variables are the recognised emotions from the dataset. Extracting emotion is the first step in the process which extracts all the emotions from the dataset. After this the emotion classification is done based on the emotion recorded in the audio file.

The dataset was split into training and testing parts accordingly using a test size of 0.3. For training of the Novel Multilayer Perceptron and Logistic Regression the test size is about 30% of the total dataset and the remaining 70% is used for training sets. The whole dataset is fitted for training the Novel Multilayer Perceptron and Logistic Regression model.

**Novel Multilayer Perceptron**

Novel Multilayer Perceptron is one of the techniques in machine learning. Novel Multi layer perceptron is non-deterministic and universal in the sense that It can effectively estimate any continuous non-linear function on a compact interval, and it is used for a variety of applications such as classification and regression. Neurons in the MLP are organized in the layers starting with the input layer and the progression through hidden layers to the output layer. The network layer is feedforward because two layers can join together MLP is made up of multiple highly interconnected processing neurons that can be employed in parallel to find a solution to the given problem. The following equation (1) is formula for MLP

$$y = \varphi(\sum_{i=1}^{n} \omega_i x_i + b) = \varphi(W^T X + b) \quad (1)$$

where w denotes the vector of weights, x is the vector of inputs, b is the bias and φ is the non-linear activation function. One of the MLP disadvantages is that it may under perform while using a Nonlinear optimization problem. MLP are useful in research because they can solve issues Stochastically allowing for approximate solutions to exceeding complications such as fitness approximation. MLP are demonstrated by cybenko's theorem which is a universal approximator that can be utilized to generate mathematical models through regression analysis. MLP is a good classifier algorithm because classification is a special case of regression when the response variable is categorical. Extracting emotion is one of the basic and important steps for extracting the emotion in the dataset. Pseudocode for the algorithm is shown in Table 2. Table 4 represents the accuracy of Novel Multilayer Perceptron recorded by testing the algorithm with 6 different sample sizes.

**Logistic Regression**

Logistic Regression is one of the algorithms in machine learning**.** The logistic model (or logit model) is used in statistics to model the probability of a specific class or event, such as pass/fail, win/lose, alive/dead, healthy/sick, true/false. This can be used to represent a variety of occurrences, such as determining whether an image contains a cat, dog, lion, or other animal. Each detected object in the image would be assigned a probability ranging from 0 to 1, with a total of one. Logistic regression is a statistical model that uses a logistic function to represent a binary dependent variable in its most basic form, though there are many more advanced variants. Logistic regression (or logit regression) is a method of estimating the parameters of a logistic model in regression analysis. The following equation (2) is formula of logistic curve

$$P = \frac{e^{a+bX}}{1+e^{a+bX}} \quad (2)$$

where P is the probability of 1, e is the base of natural logarithm, a and b are the parameters of the model. Logistic regression makes no assumption about disseminations of classes in highlight space. Logistic Regression can undoubtedly reach out to numerous classes (multinomial relapse) and a characteristic probabilistic perspective on class forecasts. It is exceptionally quick at arranging obscure records. Assuming the quantity of perceptions is lesser than the quantity of elements, Logistic Regression ought not be utilized, if not, it might prompt overfitting. Logistic regression develops straight limits. Pseudocode for the algorithm is shown in Table 3. Table 5 represents the accuracy of Logistic Regression recorded by testing the algorithm with 6 different sample sizes.

**STATISTICAL ANALYSIS**

Statistical Package for the Social Sciences Version 23 software tool was used for statistical analysis. An independent sample T-test was conducted for accuracy. Standard deviation, standard mean errors were also calculated using the SPSS Software tool.

The minimum requirements to run the softwares used here are intel core 13 dual core cpu@3.2 GHZ, 4GB RAM, 64 bit OS, 1TB Hard disk space Personal Computer and Software specification includes Windows 8 , 10 , 11 , python 3.8 and MS-Office.

The significance values of proposed and existing algorithms contain group statistical values of proposed and existing algorithms. The dependent variables are the emotions extracted from the dataset and the independent variables are audio files which are in wav format.

## 3. Results

The group statistical analysis on the two groups shows Novel Multilayer Prceptron has more mean accuracy than Logistic Regression. The Novel Multilayer Perceptron scored an accuracy of 70.37% and Logistic Regression scored an accuracy of 60.99%. The graphical representation of Novel Multilayer Perceptron and Logistic Regression is figured out below in Fig. 1. Group id is given as a grouping variable and emotions is given as the testing variable. Group id is given as 1 for Novel Multilayer Pereptron and 2 for Logistic Regression. Descriptive Statistics is applied for the dataset in SPSS and shown in Table 6. Table 7 consists of values of the independent sample T-test of both the algorithms.

## 4. Discussion

The results of this study shows that the Novel Multilayer Perceptron is proved to be having better accuracy than the Logistic Regression. MLP has an accuracy of 70.37% whereas LR has an accuracy of 60.99%. The statistical analysis of the two groups shows the Novel Multi layer perceptron has more mean accuracy than the Logistic Regression and the standard error mean including standard deviation mean is slightly less than the Novel Multi layer perceptron.

In the study by Anjani Reddy the Novel Multilayer Perceptron gained the highest accuracy of 55.11% ("[No Title]" n.d.)**.** The study of Iliou, Theodoros, and Christos-Nikolaos Anagnostopoulos the accuracy gained for the MLP is 89.1% (Iliou and Anagnostopoulos 2010). The study of Rumagit RY, Alexander G and Saputra the accuracy gained by the MLP classifier is 83% whereas the accuracy mentioned is gained for extracting only one emotion ("Model Comparison in Speech Emotion Recognition for Indonesian Language" 2021). The common factor in all these were they have extracted the only emotion at a time. The disadvantage in this model is that the precision of MLP might get influenced because of the conflicting information and trouble in getting the right datasets for examination. The majority of the information is reenacted from nature which is a long way from the real world. The accessibility of more cross-language discourse related datasets of feeling, viable information preprocessing methods, and the mix of MLP with other AI calculations. Emotion recognition has a great impact on the technologies (Zhang et al. 2021).

The limitations in the model is that the accuracy may get affected due to inconsistent data and difficulty in getting the right dataset for analysis. The key is getting high accuracy in extracting the emotion from the dataset. Future study is to compare the MLP with other algorithms and to check the accuracy gained from that.

## 5. Conclusion

Based on the results from the experiments done, the Novel Multilayer Perceptron has got more accuracy and proved to recognize speech emotion more significantly than the Logistic Regression. So, it can be used for the emotion recognition system which are used in the robots for recognizing the command given by the humans.

**DECLARATIONS**

**Conflicts of Interest**

No conflict of interest in this manuscript.

**Authors Contribution**

Author SD was involved in data collection, analyzing data, extracting data, manuscript writing. Author DBD was involved in conceptualization, data validation, and critical review of the manuscript.

**Acknowledgement**

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Science (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

**Funding**

We thank the following organizations for providing financial support that enabled us to complete the study.

1. Vee Eee Technologies Solution Pvt. Ltd., Chennai.
2. Saveetha University.
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha School of Engineering.

## 6. References

Anupong, Wongchai, Lin Yi-Chia, Mukta Jagdish, Ravi Kumar, P. D. Selvam, R. Saravanakumar, and Dharmesh Dhabliya. 2022. "Hybrid Distributed Energy Sources Providing Climate Security to the Agriculture Environment and Enhancing the Yield." Sustainable Energy Technologies and Assessments. https://doi.org/10.1016/j.seta.2022.102142.

Arena, Paolo, Luigi Fortuna, Giovanni Muscato, and Maria Gabriella Xibilia. 1998. Neural Networks in Multidimensional Domains: Fundamentals and New Trends in Modelling and Control. Springer.

Bharathiraja, B., J. Jayamuthunagai, R. Sreejith, J. Iyyappan, and R. Praveenkumar. 2022. "Techno Economic Analysis of Malic Acid Production Using Crude Glycerol Derived from Waste Cooking Oil." Bioresource Technology 351 (May): 126956.

Hansen, Lasse, Yan-Ping Zhang, Detlef Wolf, Konstantinos Sechidis, Nicolai Ladegaard, and Riccardo Fusaroli. 2022. "A Generalizable Speech Emotion Recognition Model Reveals Depression and Remission." Acta Psychiatrica Scandinavica 145 (2): 186–99.

Huang, Chenchen, Wei Gong, Wenlong Fu, and Dongyu Feng. 2014. "A Research of Speech Emotion Recognition Based on Deep Belief Network and SVM." Mathematical Problems in Engineering 2014 (August). https://doi.org/10.1155/2014/749604.

Iliou, Theodoros, and Christos-Nikolaos Anagnostopoulos. 2010. "SVM-MLP-PNN Classifiers on Speech Emotion Recognition Field - A Comparative Study." 2010 Fifth International Conference on Digital Telecommunications. https://doi.org/10.1109/icdt.2010.8.

Jaso, Danel. 2006. Speech Enhancement, Modeling and Recognition.

Javaheri, Behzad. n.d. "Speech & Song Emotion Recognition Using Multilayer Perceptron and Standard Vector Machine." https://doi.org/10.20944/preprints202105.0441.v1.

Jothi, K. Jeeva, K. Jeeva Jothi, S. Balachandran, K. Mohanraj, N. Prakash, A. Subhasri, P. Santhana Gopala Krishnan, and K. Palanivelu. 2022. "Fabrications of Hybrid Polyurethane-Pd Doped ZrO2 Smart Carriers for Self-Healing High Corrosion Protective Coatings." Environmental Research. https://doi.org/10.1016/j.envres.2022.113095.

Kale, Vaibhav Namdev, J. Rajesh, T. Maiyalagan, Chang Woo Lee, and R. M. Gnanamuthu. 2022. "Fabrication of Ni–Mg–Ag Alloy Electrodeposited Material on the Aluminium Surface Using Anodizing Technique and Their Enhanced Corrosion Resistance for Engineering Application." Materials Chemistry and Physics. https://doi.org/10.1016/j.matchemphys.2022.125900.

Kamaruddin, N., and A. Wahab. 2013. "Emulating Human Cognitive Approach for Speech Emotion Using MLP and GenSofNN." 2013 5th International Conference on Information and Communication Technology for the Muslim World (ICT4M). https://doi.org/10.1109/ict4m.2013.6518885.

Liu, Dong, Zhiyong Wang, Lifeng Wang, and Longxi Chen. 2021. "Multi-Modal Fusion Emotion Recognition Method of Speech Expression Based on Deep Learning." Frontiers in Neurorobotics 15 (July): 697634.

"Model Comparison in Speech Emotion Recognition for Indonesian Language." 2021. Procedia Computer Science 179 (January): 789–97.

"[No Title]." n.d. Accessed September 30, 2021. https://www.ijera.com/papers/vol11no6/Series-2/G106023438.pdf.

Palanisamy, Rajkumar, Diwakar Karuppiah, Subadevi Rengapillai, Mozaffar Abdollahifar, Gnanamuthu Ramasamy, Fu-Ming Wang, Wei-Ren Liu, Kumar Ponnuchamy, Joongpyo Shim, and Sivakumar Marimuthu. 2022. "A Reign of Bio-Mass Derived Carbon with the Synergy of Energy Storage and Biomedical Applications." Journal of Energy Storage. https://doi.org/10.1016/j.est.2022.104422.

Ram, G. Dinesh, G. Dinesh Ram, S. Praveen Kumar, T. Yuvaraj, Thanikanti Sudhakar Babu, and Karthik Balasubramanian. 2022. "Simulation and Investigation of MEMS Bilayer Solar Energy Harvester for Smart Wireless Sensor Applications." Sustainable Energy Technologies and Assessments. https://doi.org/10.1016/j.seta.2022.102102.

"Recognition of Fear from Speech Using Adaptive Algorithm with MLP Classifier." n.d. Accessed September 29, 2021. https://doi.org/10.1109/ICCPCT.2016.7530149.

Satapathy, Suresh Chandra, Vikrant Bhateja, and Swagatam Das. 2017. Smart Computing and Informatics: Proceedings of the First International Conference on SCI 2016. Springer.

Shahsavarani, Somayeh. 2018. Speech Emotion Recognition Using Convolutional Neural Networks.

"Speech Emotion Recognition in Noisy Environment." n.d. Accessed September 29, 2021. https://doi.org/10.1109/ATSIP.2016.7523189.

Sumathy, B., Anand Kumar, D. Sungeetha, Arshad Hashmi, Ankur Saxena, Piyush Kumar Shukla, and Stephen Jeswinde Nuagah. 2022. "Machine Learning Technique to Detect and Classify Mental Illness on Social Media Using Lexicon-Based Recommender System." Computational Intelligence and Neuroscience 2022 (February): 5906797.

Thanigaivel, Sundaram, Sundaram Vickram, Nibedita Dey, Govindarajan Gulothungan, Ramasamy Subbaiya, Muthusamy Govarthanan, Natchimuthu Karmegam, and Woong Kim. 2022. “The Urge of Algal Biomass-Based Fuels for Environmental Sustainability against a Steady Tide of Biofuel Conflict Analysis: Is Third-Generation Algal Biorefinery a Boon?” Fuel. https://doi.org/10.1016/j.fuel.2022.123494.

Vickram, Sundaram, Karunakaran Rohini, Krishnan Anbarasu, Nibedita Dey, Palanivelu Jeyanthi, Sundaram Thanigaivel, Praveen Kumar Issac, and Jesu Arockiaraj. 2022. “Semenogelin, a Coagulum Macromolecule Monitoring Factor Involved in the First Step of Fertilization: A Prospective Review.” International Journal of Biological Macromolecules 209 (Pt A): 951–62.

Yaashikaa, P. R., M. Keerthana Devi, and P. Senthil Kumar. 2022. “Algal Biofuels: Technological Perspective on Cultivation, Fuel Extraction and Engineering Genetic Pathway for Enhancing Productivity.” Fuel. https://doi.org/10.1016/j.fuel.2022.123814.

Zhang, Shiqing, Ruixin Liu, Xin Tao, and Xiaoming Zhao. 2021. “Deep Cross-Corpus Speech Emotion Recognition: Recent Advances and Perspectives.” Frontiers in Neurorobotics 15 (November): 784514.

**Tables And Figures**

Table 1. Audio files and Recognized emotion collected from Ravdess dataset.

| **Audio file** | **Recognized Emotion** |
|---|---|
| Actor_01 | (22,6) |
| Actor_02 | (24,7) |
| Actor_03 | (33,8) |
| Actor_04 | (42,11) |
| Actor_05 | (37,16) |
| Actor_06 | (41,26) |
| Actor_07 | (36,22) |
| Actor_08 | (19,21) |
| Actor_09 | (34,20) |
| Actor_10 | (29,8) |

Table 2. Pseudocode for Novel Multilayer Perceptron.

| // **I : Input dataset records** |
|---|
| Import the required packages |
| Extracting the emotion from the dataset |
| Assigning the data a_train, b_train, a_test, b_test variables |
| Using train_test_split function, pass the train and test variables |
| Splitting the data |
| Importing the Multilayer Perceptron classifier |
| Evaluate the result using a_test and y_test function |

| |
|---|
| Get the accuracy of the model. |
| // **Output Accuracy** |

Table 3. Pseudocode for Logistic Regression.

| // **I : Input dataset records** |
|---|
| Import the required packages |
| Extracting the emotion from the dataset |
| Assigning the data a_train, b_train, a_test, b_test variables |
| Using train_test_split function, pass the train and test variables |
| Splitting the data |
| Importing the Logistic Regression classifier |
| Evaluate the result using a_test and y_test function |
| Get the accuracy of the model. |
| // **Output**<br>**Accuracy** |

Table 4. Accuracy of speech emotion recognition using Novel Multilayer Perceptron.

| **Test size** | **Accuracy** |
|---|---|
| Test 1 | 66.56 |
| Test 2 | 70.06 |
| Test 3 | 70.37 |
| Test 4 | 68.24 |
| Test 5 | 66.29 |
| Test 6 | 67.86 |

Table 5. Accuracy of speech emotion recognition using Logistic Regression.

| **Test size** | **Accuracy** |
|---|---|
| Test 1 | 59.41 |
| Test 2 | 58.46 |
| Test 3 | 59.26 |
| Test 4 | 59.86 |
| Test 5 | 58.86 |

| | |
|---|---|
| Test 6 | 60.99 |

Table 6. Group Statistics analyzing the Novel Multilayer Perceptron and Logistic Regression.

| Algorithm | | N | Mean | Std.Deviation | Std.Error mean |
|---|---|---|---|---|---|
| **Accuracy** | **NMLP** | 6 | 68.2300 | 1.70985 | 0.69804 |
| **Accuracy** | **LR** | 6 | 59.4633 | 0.86294 | 0.35229 |

Table 7. Independent Sample Test with confidence interval as 95% and the level of significance as 0.05. (Novel Multilayer Perceptron works significantly better than Logistic Regression with the significance value p = 0.139).

| **Accuracy** | **Levene's Test for Equality of Variances** | | **T-test for Equality of Means** | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **F** | **Sig.** | **t** | **df** | **Sig. (2-tail)** | **Mean Difference** | **Std. Error Difference** | **95% Conf. Interval Lower** | **95% Conf. Interval Upper** |
| **Equal Variances assumed** | 2.592 | 0.138 | 11.145 | 10 | 0.000 | 8.75667 | 0.78571 | 7.00599 | 10.50734 |
| **Equal Variances not assumed** | 2.592 | 0.138 | 11.145 | 7.492 | 0.000 | 8.75667 | 0.78571 | 6.92320 | 10.59013 |

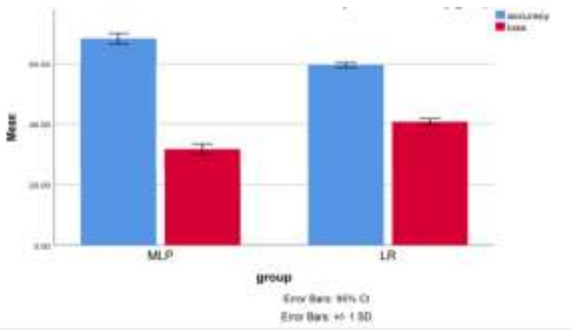


Fig. 1. Comparison of Novel Multi layer perceptron and Logistic regression in terms of accuracy. The mean accuracy is Novel Multi layer perceptron greater than Logistic Regression and the standard deviation is also slightly higher than Logistic Regression. X-axis: Novel Multilayer Perceptron vs Logistic Regression. Y-axis: Mean accuracy of detection + 1 SD.