

Depression Detection Using Stacked Autoencoder from Facial Features and NLP

Ravi Kumar
M.Tech Scholar
Department of CSE

National Institute for Interdisciplinary
Science and Technology
Bhopal, M.P, India

kumarravi47@gmail.com

Santosh Kumar Nagar
Assistant Professor
Department of CSE
National Institute for
Interdisciplinary Science and
Technology
Bhopal, M.P, India
santoshnagar9@gmail.com

Anurag Shrivastava
Associate Professor
Department of CSE
National Institute for Interdisciplinary
Science and Technology
Bhopal, M.P, India
anurag.shri08@gmail.com

Abstract— Depression has become one of the most common mental illnesses in the past decade, affecting millions of patients and their families. However, the methods of diagnosing depression almost exclusively rely on questionnaire-based interviews and clinical judgments of symptom severity, which are highly dependent on doctors' experience and makes it a labor-intensive work. This research work aims to develop an objective and convenient method to assist depression detection using facial features as well as textual features. Most of the people conceal their depression from everyone. So, an automated system is required that will pick out them who are dealing with depression. In this research, different research work focused for detecting depression are discussed and a hybrid approach is developed for detecting depression using facial as well as textual features. The main purpose of this research work is to design and propose a hybrid system of combining the effect of three effective models: Natural Language Processing, Stacked Deep Auto Encoder with Random forest (RF) classifier and fuzzy logic based on multi-feature depression detection system. According to literature several fingerprint as well as fingervein recognition system are designed that uses various techniques in order to reduce false detection rate and to enhance the performance of the system. A comparative study of different recognition technique along with their limitations is also summarized and optimum approach is proposed which may enhance the performance of the system. The result analysis shows that the developed technique significantly advantages over existing methods.

Keywords— Depression Detection, Facial Features, Natural Language Processing, Stacked Autoencoder, Fuzzy Logic.

I. INTRODUCTION

Depression is a common mental disorder that already affects more than 350 million people worldwide [1]. It will not only make a bad influence on the patients but also on their families. The World Health Organization said that depression will become the second leading cause of illness by the year 2020 [2]. However, the assessment methods of diagnosing depression almost exclusively rely on the patient-reported and clinical judgments of the symptom severity [3]. Current diagnostic techniques of depression have obvious disadvantages, which are associated with the patient denial, poor sensitivity, subjective biases and inaccuracy [4]. Finding

an objective, accurate and practical method for depression detection still remains a challenge.

Early detection of an individual at risk of depression in initial stages and in mild form is beneficial both for the individual and society. Our study contributes to screening of individuals at risk of this disease. The study provides the original way of depression screening based on an analysis of how the writer uses the language to enable the automatic detection of writer's risk of depression. Although depression may occur only once during your life, people typically have multiple episodes [6]-[10]. During these episodes, symptoms occur most of the day, nearly every day and may include:

- Feelings of sadness, tearfulness, emptiness or hopelessness
- Angry outbursts, irritability or frustration, even over small matters
- Loss of interest or pleasure in most or all normal activities, such as sex, hobbies or sports
- Sleep disturbances, including insomnia or sleeping too much
- Tiredness and lack of energy, so even small tasks take extra effort
- Reduced appetite and weight loss or increased cravings for food and weight gain
- Anxiety, agitation or restlessness
- Slowed thinking, speaking or body movements
- Feelings of worthlessness or guilt, fixating on past failures or self-blame
- Trouble thinking, concentrating, making decisions and remembering things
- Frequent or recurrent thoughts of death, suicidal thoughts, suicide attempts or suicide
- Unexplained physical problems, such as back pain or headaches

Ideally, machine learning tools for detecting depression should have access to the same information flow used by a doctor during diagnosis [11]-[15]. Consequently, the

functionalities used by these classifiers must represent any type of communication: face and gesture, voice, language and language.

A. Visual Indicators

Visual indicators for the analysis of depression, including body movements, gestures, subtle expressions and periodic muscle movements, have been widely studied.

B. Speech Indicators

Recent research has shown that the use of language as a diagnostic and monitoring tool for depression is promising. A person's vocal production system is very complex, and therefore small cognitive or physiological changes can cause changes in acoustic language. This idea led research to use language as an objective indicator for depression. Depressive language has been systematically associated with a variety of prosodic, origin, training and spectral indicators.

C. Textual Indicators

Early detection of an individual at risk of depression in initial stages and in mild form is beneficial both for the individual and society. Many studies are presented that focuses on discovery of the relationship between linguistic characteristics of a written text and the level of the emotional state of depression. The focus is on non-content (non-semantic) computational linguistic markers of a written text.

II. RELATED WORK

Depression is all about the emotions. Plutchik [5] listed the basic eight types of emotions as mentioned below:

1. Fear is the feeling of being afraid or a feeling of insecurity. It is usually induced by perceived threat or danger. It can cause a change in behavior. It is an unpleasant feeling.
2. Anger is intense negative emotion that leads to a hostile response to perceived provocation, threat or hurt [6]-[9]. It is usually accompanied by the feeling of doing something violent and taking revenge.
3. Sadness is the feeling of despair, grief, disappointment, and sorrow. It changes the behavior of the person to a state where a person feels very bad and sometimes feel like crying.
4. Joy is a positive feeling of happiness and great pleasure. It enables the person to enjoy the moment and try to find good things in everything. It promotes good behavior in a person.
5. Disgust is the feeling of regret and disapproval usually caused by something unpleasant or offensive. It is usually followed by anger.
6. Surprise is a feeling of astonishment when something unexpected happens. It leaves the person unable to believe what he/ she perceives.

7. Trust is the feeling of reliability, truth or confidence in someone.
8. Anticipation is the feeling of looking forward positively on something that is about to happen. There are other emotions like acceptance, rage, ecstasy etc [16]-[20]. However, all these emotions are a combination of two or more emotions mentioned above. Affective Computing has been an active area of research in the field of machine learning, it is a broad area.

The authors are well motivated by the work described above and have decided to introduce a depression detection model that correctly and quickly recognizes the person's mental state [21]-[23]. In the given system, we try to determine a person's mental state, i.e. whether the person is depressed or normal. Table I gives some existing techniques with their result analysis.

Table I: Comparative Analysis of Existing Works

Author	Technique Used	Conclusion
Kulkarni and Patil [1]	Fisher vector algorithm and LTrP	Accuracy = 87.67
Shen et al. [3] (2018)	Deep Neural Network model with Feature Adaptive Transformation & Combination strategy	Accuracy=77%
Sharma [5] (2018)	Deep Learning	Discussed the advantages of deep learning in depression detection
Vonikakis et al. [9] (2016)	Geometric features Partial Least Squares regression	RMSE = 0.8316
Wang et al. [12] (2013)	Vocabulary rules to calculate the depression inclination	Accuracy=80%
Xinyu Wang et al. [13] (2013)	Calculate the probability of a user being depressed	Accuracy=95%
Alghowinem et al. [14] (2013)	Active Appearance Models	Accuracy=70% Accuracy=75%

	Gaussian Mixture Models Support Vector Machines	
Lu-Shih et al. [16] (2011)	Teager energy operator (TEO)	Accuracy=80.5%
Lu-Shih et al. [17] (2009)	Gaussian mixture models (GMM)	Accuracy=59.55%
Cohn [18] (2009)	Active Appearance Modeling (AAM) and Pitch Extraction	Accuracy=79%

III. METHODOLOGY

In this research work an approach is proposed to develop a depression level detection system using facial features as well as textual features. The algorithm is designed to match with given training dataset that gives matching result in probability.

Figure 1 shows the different stages of the proposed biometric recognition system that will be analyzed in the following step by step.

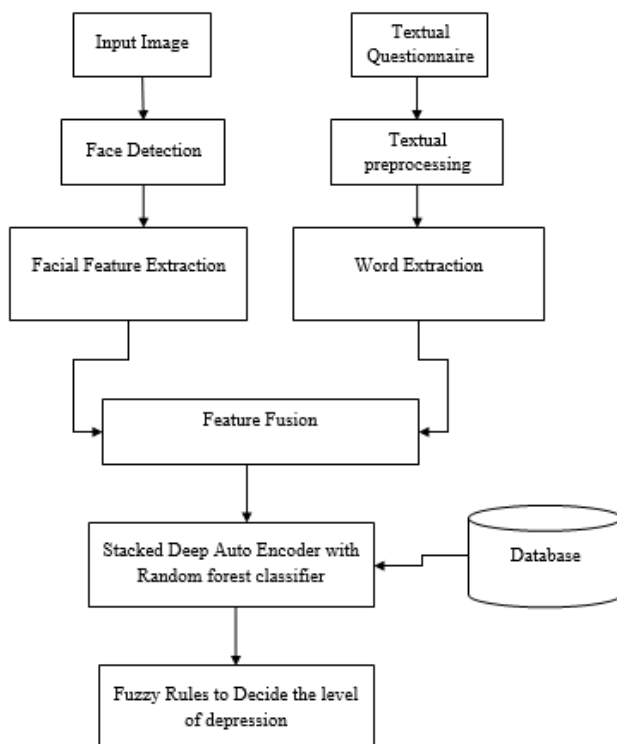


Figure 1: Proposed Hybrid Biometric Identification System

The flowchart of the proposed methodology is discussed as below:

Step 1: Input face image and Textual Questionnaire
 Step 2: Facial features are extracted and Textual features are extracted from text.
 Step 3: Feature fusion and learning using autoencoder with RF classifier for further processing.
 Step 4: Training the proposed network and identify the result.
 Step 5: Designing the fuzzy rules for deciding the level of depression.

A. Facial Feature Extraction

Each subject had a neutral expression and the six basic expressions. The feature points of these facial images of the participations were artificial marked with 68 points which were defined in Xm2vts [22] frontal face data.

Facial Feature Points Tracking in Video

In this step, each frames of the videos are analyzed. First, the Viola-Jones face detector was used to detect the human face region in the frame. Then, the current positions of the feature points were obtained by using the AAM on each frame.

Facial Feature Extraction

Facial expressions and eyes feature points are selected to detect the depression, including eye pupil movement, blinking frequency, and movement changes of bilateral eyebrows and corners of mouth [24][25].

- **Eyes:** The distance between the left pupil (31, the position of each points). Left inner eye corner point (29). The distance between the right pupil (36). The right inner eye corner point (34).
- **Eyebrows:** The distances between the medial three feature points (23, 24, 25) on left eyebrow and the left inner eye corner point (29) which is invariant relative to the face. The distances of the medial three feature points (17, 18, 19) on the right eyebrow and the right inner eye corner point (34). The distance between the feature point on the left side of the eyebrow (21) and left inner eye corner point (29). The distance between the feature point on the right side of the eyebrow (15) and right inner eye corner point (34).
- **Corners of mouth:** The distances between the nose tip point 67 and six characteristic points (48, 49, 59, 53, 54, and 55) around the corners of mouth. The maximum, minimum and standard deviation of these distances mentioned above are used to measure the degree of facial expression changes.

At the same time, in order to eliminate the difference between different faces and the projection difference caused by the distance from the camera, we divide those maximum, minimum and standard deviation distance values by the mean distance.

A total of 49 statistical features were extracted, which are:

- i. Standard deviation/mean, maximum/mean, minimum/mean of the 8 points on eyebrows (3 × 8)
- ii. Standard deviation/mean, maximum/mean and minimum/mean of 2 eye pupils (3 × 2).
- iii. Standard deviation/mean, maximum/mean and minimum/mean of the 6 points on eyebrows (3 × 6)
- iv. Blink frequency of eyes (1 × 1)

B. Textual Feature Extraction

Data Gathering

The very first step of proposed research methodology is dataset preparation in which data gathering is first of all performed. Figure 2 represents this step in which manually analysed questionnaire are collected and saved in a dataset format.

Text Pre-processing

As it is necessary to clean the dataset. This process is termed as data preprocessing. This step is necessary to remove unnecessary terms used by an individuals in questionnaire such as comma, full-stop, colons or any special symbols or characters. These terms doesn't associate any sentiment values. So, to reduce further complexity it is needed to removes such terms. In this research work this step is performed in two steps as stated below:

- Some special symbols or characters are removed.
- Comma, full stops, colons and semi-colons are removed.

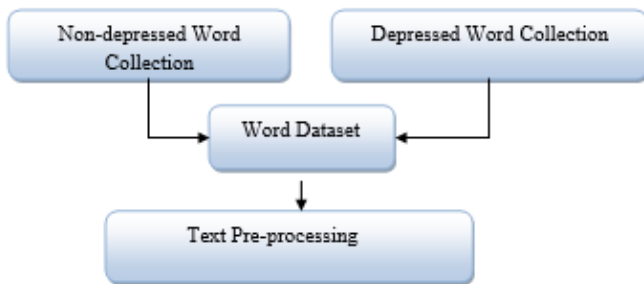


Figure 2: Data Gathering for Proposed Work

Textual Feature Extraction

Extraction of some useful information out of these questionnaires is known as feature extraction. This information is a set of features that is very useful for classifiers. Figure 3 illustrates the process of feature extraction for proposed methodology.

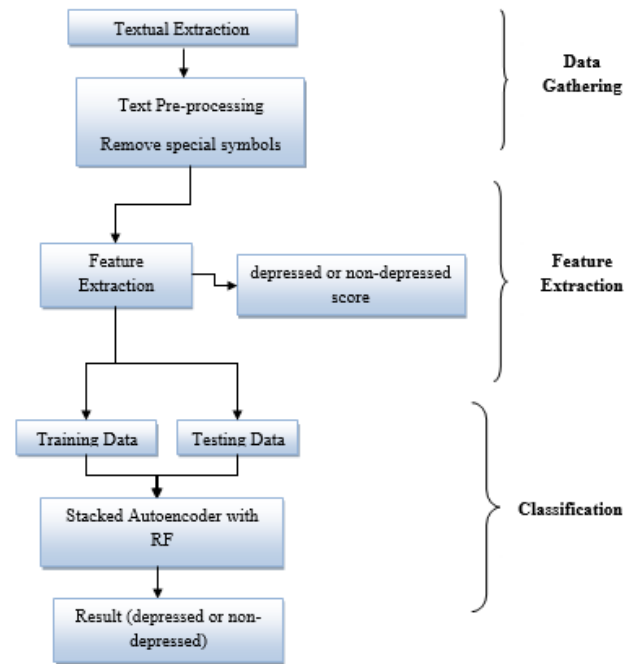


Figure 3: Proposed Flow Diagram of Textual Depression Analysis

For generation of feature vector associated with depressed or non-depressed score of the questionnaire sentence. A depression dictionary is referred. In this research work sentiment is calculated by summarizing the depressed or non-depressed score of each word in entire sentence. If the non-depressed score of entire sentences is greater than depressed score of entire sentences then the overall sentiment score of that sentence/review is non-depressed.

Similarly, if the non-depressed score of entire sentences is less than depressed score of entire sentences then the overall sentiment score of that sentence is depressed. The calculation of depressed or non-depressed score of the sentence is determined as follows:

$$\text{Depressed_Score_Sentence} = \frac{\text{Sum of Depressed Score of Each Word}}{\text{Total Number of words in sentence}} \tag{1}$$

$$\text{Non - depressed_Score_Sentence} = \frac{\text{Sum of Non - depressed Score of Each Word}}{\text{Total Number of words in sentence}} \tag{2}$$

Stacked Deep Auto Encoder

With the rapid development of unsupervised learning in recent years, the use of untagged data to extract functions with Autoencoder has become an appropriate medium. The Autoencoder model is essentially a multilayer neural network. A deep stacked autocoder is constructed by combining a stacked autocoder that includes a desired number of cascading automatic encoding layers. In Autoencoder networks, the learning phase of the functionalities is not monitored since no labeled data is used. The basic architecture of an unattended

auto encoder is a step forward with an input level, often a hidden level and an output level.

An automatic encoder can be used for pre-training or to reduce dimensionality if the architecture has the shape of a bottleneck. For simplicity, consider a car encoder with a hidden layer. The automatic encoder can then learn different display levels by stacking the hidden levels. It is a feature extraction algorithm. Helps find a representation of the data. The functionality generated by the automatic encoders represents the data point better than the points themselves.

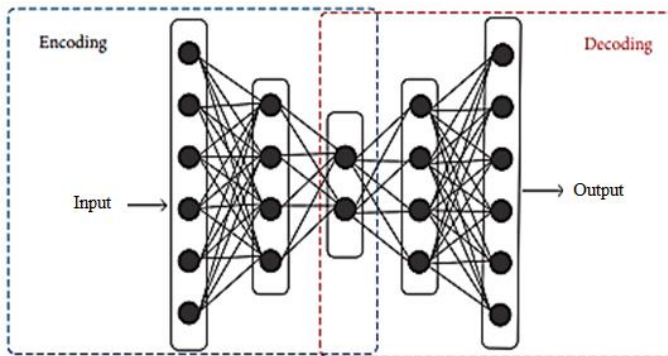


Figure 4: Stacked Deep Auto Encoder

The main difference between ordinary forward neural network and autoencoder is that an autoencoder's output is always the same as or similar to its input. The basic formula can be expressed as follows:

$$F_v = h(x) = \sum W_e * X_i + B_i \tag{3}$$

$$F_v' = h'(x) = \sum W_d * X'_i + B'_i \tag{4}$$

Where

W_e = weight matrices of encoder

W_d = weight matrices of decoder

An automatic encoder can be considered as a combination of encoder and decoder. The encoder contains an input layer and a hidden layer which converts an input image I into a characteristic vector F_v . The decoder includes a hidden layer and an output layer that transform feature F_v to output feature F_v' . tanh activation functions, which is used to activate the unit in each layer.

And transfer function is calculated as :

$$f(x) = \frac{1}{1 + e^{-y}} \tag{5}$$

Where, e=error value

The stacked deep autoencoder neural network involves multiple layer of autoencoders neural network and the loss function that is to be minimized as :

$$loss_{min} = |X - (W_1 \theta (W_2 \theta \dots \dots (W_l (f(x)))))| \tag{6}$$

Where, W_1, W_2, \dots, W_l = weight function of all autoencoders

θ = Decoding function of autoencoders

$f(x)$ = function to calculate data values at each layer

Fundamentally, this involves the proposed shift from the encoder-decoder paradigm (symmetric) and towards utilizing just the encoder phase (non-symmetric). The reasoning behind this is that given the correct learning structure, it is possible to reduce both computational and time overheads, with minimal impact on accuracy and efficiency.

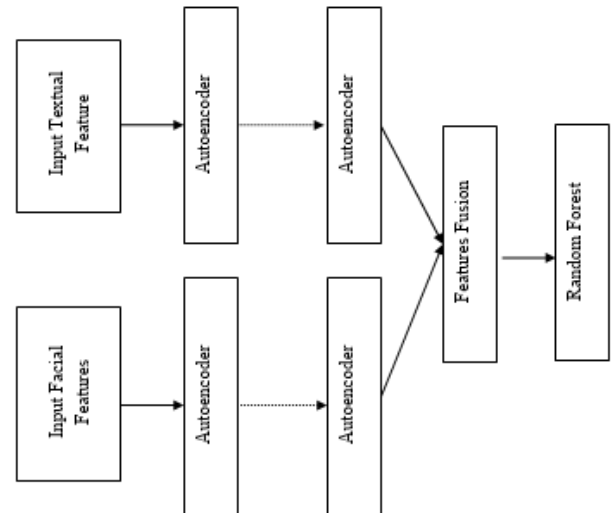


Figure 5: Stacked Deep Auto Encoder with RF Classifier

And finally, all features are combined to form the fusion feature of each facial as well as textual. These features of image as well as textual are fused together for extracting global features and these features are further used for classification of depression. Random forest is used for classification of extracted fused features for personal identification either depressed as well as non-depressed.

C. Fuzzy Rules for Deciding Depression Level

After classification of person either depressed as well as non-depressed, fuzzy rules are designed the level of depression by designing rules. The depression level is designed in three basic level.

- Highly Depressed (high)
- Mildly Depressed (average)
- Non-Depressed (min)

Facetedetect (facial classification) have two conditions : min (non-depressed) and max (depressed)

Textualdetect (text classification) have five conditions : min (depression level < 0.3), min_2 (depression level > 0.3 and < 0.5), average (depression level > 0.5 and < 0.7), high_2 (depression level > 0.7 and < 0.9) and high (depression level > 0.9).

Output is determined in three conditions : Highly Depressed (high), Mildly Depressed (average), Non-Depressed (min)

IV. RESULT ANALYSIS

For result analysis simulation is performed using MATLAB platform. In this section, screen shots of simulated features are represented below. The GUI of home page contains two functions, one for facial expression analysis and another for textual feature analysis.

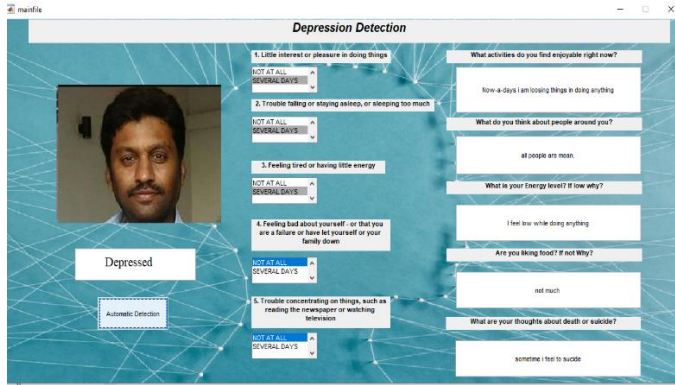


Figure 6: Depression Testing Example 1

In figure 6, a test image is processed while giving a depressed expression and the proposed methodology gives output as depressed.

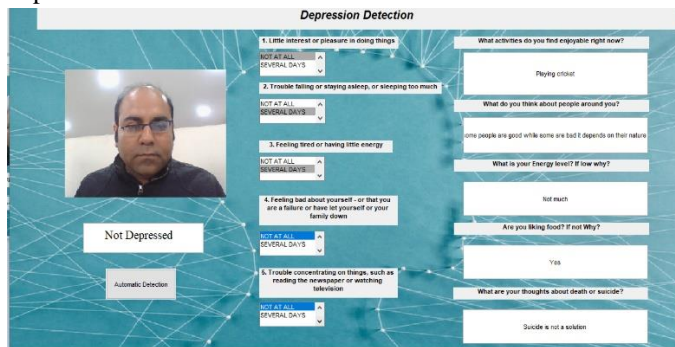


Figure 7: Depression Testing Example 2

In figure 7, a test image is processed while giving a non-depressed expression and the proposed methodology gives output as non-depressed.

The detection result is shown which processing the hybrid features of depression detection. If the detection level is greater than 90% then it is considered to be in high level of depression, while it is less than 30% then it is considered to be as not-depressed and between then the patient is considered to be as mild depressed.

Table II represents the performance evaluation of proposed stacked auto-encoder based on random forest classifier for hybrid depression detection system in terms of accuracy precision rate, recall rate, f_measure and Equal Error Rate (EER).

Table II: Performance Evaluation of Proposed Algorithm

Parameters	Values
Accuracy	94

Precision	96
Recall/ Sensitivity	98
Specificity	91
F_Measure	97
EER	0.12

Table III represents the comparative performance evaluation with respect to existing work.

Table III: Comparative Performance Evaluation

Techniques	Kulkarni et al. [1] (in %)	Proposed (in %)
Accuracy	88	94
Recall/ Sensitivity	94	98
Specificity	88	91

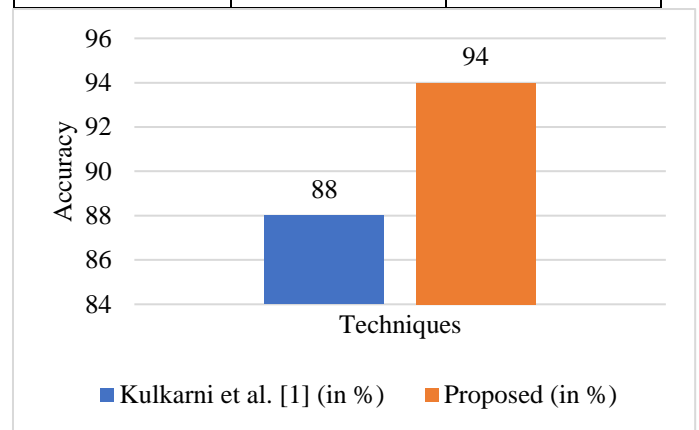


Figure 4.4: Comparative Accuracy Evaluation

Figure 4.4 represents the comparative accuracy evaluation of proposed stacked autoencoder based random forest classifier with respect to existing work.

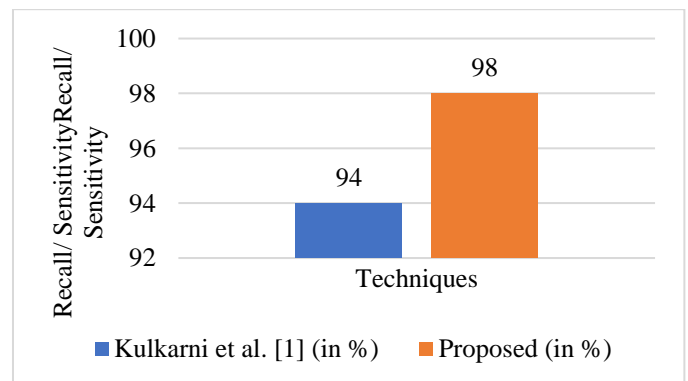


Figure 4.5: Comparative Recall/ Sensitivity Evaluation

Figure 4.5 represents the comparative recall/sensitivity evaluation of proposed stacked autoencoder based random forest classifier with respect to existing work.

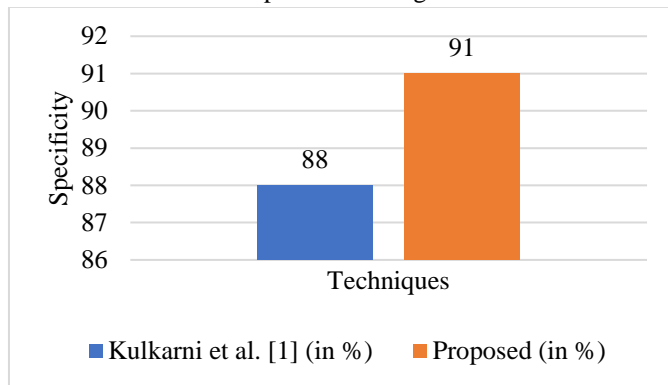


Figure 4.6: Comparative Specificity Evaluation

Figure 4.6 represents the comparative specificity evaluation of proposed stacked autoencoder based random forest classifier with respect to existing work.

V. CONCLUSION

Depression is a serious mental illness, and the current diagnosis process still needs to be conducted by a specially trained psychiatrist or psychologist, usually using a scale and careful observation in communication, which depends on the doctor's experience. And it's hard for non-psychiatrists to diagnose and treat depression.

Previous research has shown that context can have an impact on speed and accuracy when identifying facial expressions of emotion. Many research work was conducted with facial expression. But this system gives a bias result while differentiating between depressed, happy and sad features. So, in this research work, a hybrid system is developed with facial as well as textual which reduces the drawbacks of facial expression system. The main purpose of this research work is to design and propose a hybrid system of combining the effect of three effective models: Natural Language Processing, Stacked Deep Auto Encoder with Random forest (RF) classifier and fuzzy logic based on multi-feature depression detection system. According to literature several fingerprint as well as finger vein recognition system are designed that uses various techniques in order to reduce false detection rate and to enhance the performance of the system. The result analysis shows approx. 6% accuracy, 4% recall rate as well as 3% specificity that shows enhancement over existing work. Depression and anxiety disorders are critical problems in modern society. In future study, other multi-modal features and multi-modal fusion strategies will be explored such as EEG signals in order to obtain promising performance.

REFERENCES

- [1] P. B. Kulkarni and M. M. Patil, "Clinical Depression Detection in Adolescent by Face," International Conference on Smart City and Emerging Technology (ICSCET), Mumbai, 2018, pp. 1-4.
- [2] Q. Cong, Z. Feng, F. Li, Y. Xiang, G. Rao and C. Tao, "X-A-BiLSTM: A Deep Learning Approach for Depression Detection in Imbalanced Data," IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Madrid, Spain, 2018, pp. 1624-1627.
- [3] Shen, Tiancheng, "Cross Domain Depression Detection via Harvesting Social Media." Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. 2018.
- [4] Singh, Harman, Neeti Dhanak, Haroon Ansari, and Krishan Kumar. "HDML: Habit Detection with Machine Learning." In Proceedings of the 7th International Conference on Computer and Communication Technology, pp. 29-33. ACM, 2017.
- [5] Sharma, Shikhar, et al., "D-FES: Deep facial expression recognition system." Information and Communication Technology (ICT), 2017 Conference on. IEEE, 2017.
- [6] S. Shikhar, et al., "LEXER: LEXicon Based Emotion AnalyzeR." International Conference on Pattern Recognition and Machine Intelligence. Springer, Cham, 2017.
- [7] Pampouchidou, Anastasia, "Depression Assessment by Fusing High and Low Level Features from Audio, Video, and Text." Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge. ACM, 2016.
- [8] T. Baltrusaitis et al., "OpenFace: An Open Source Facial Behavior Analysis Toolkit," in IEEE-WACV, 2016, pp. 1-10.
- [9] V. Vonikakis et al., "Group happiness assessment using geometric features and dataset balancing," in 18th ACM-ICMI. ACM, 2016, pp. 479-486.
- [10] M. Valstar et al., "AVEC 2014: 3D Dimensional Affect and Depression Recognition Challenge," in 4th ACM-AVEC, 2014, pp. 3-10.
- [11] Michel F. Valstar, Björn Schuller, Kirsty Smith, Florian Eyben, Timur Almaev, Jarek Krajewski, Roddy Cowie, Jeff Cohn and Maja Pantic, 'AVEC 2014 – The 3D Dimensional Affect and Depression Severity Estimation Challenge', proc. Int'l Conf. ACM Multimedia, 2014.
- [12] Wang, Xinyu, "A depression detection model based on sentiment analysis in micro-blog social network." Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer, Berlin, Heidelberg, 2013.
- [13] Wang, Xinyu, Chunhong Zhang, and Li Sun. "An improved model for depression detection in micro-blog social network." 2013 IEEE 13th International Conference on Data Mining Workshops (ICDMW). IEEE, 2013.
- [14] S. Alghowinem, "Eye movement analysis for depression detection," in 20th IEEE-ICIP, 2013, pp. 4220-4224.
- [15] Videbeck, Sheila, and Sheila Videbeck. Psychiatric-mental health nursing. Lippincott Williams & Wilkins, 2013.
- [16] Low, Lu-Shih Alex, et al. "Detection of clinical depression in adolescents speech during family interactions." IEEE Transactions on Biomedical Engineering 58.3, 2011.
- [17] Low, Lu-Shih Alex, "Content based clinical depression detection in adolescents." Signal Processing Conference, 2009 17th European. IEEE, 2009.
- [18] Cohn, Jeffrey F., "Detecting depression from facial actions and vocal prosody." Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on. IEEE, 2009.
- [19] Gilson, Mark, and Arthur Freeman. Overcoming depression: A cognitive therapy approach therapist guide. Oxford University Press, 2009.
- [20] Clarke, Greg, "Overcoming Depression on the Internet (ODIN)(2): a randomized trial of a self-help depression skills program with reminders." Journal of medical Internet research 7.2, 2005.
- [21] K. Messer, J. Matas, J. Kittler, K. Jonsson, XM2VTS: the extended M2VTS database. In: Second International Conference on Audio- and Video-Based Biometric Person Authentication, 2000, pp. 72-77.
- [22] Persons, Jacqueline B. "Psychotherapy outcome studies do not accurately represent current models of psychotherapy: A proposed remedy." American psychologist 46.2, 1991.
- [23] Plutchik, Robert. The emotions. University Press of America, 1991.
- [24] P. H. Waxer, "Therapist training in nonverbal communication. I: Nonverbal cues for depression," Journal of Clinical Psychology, vol. 30, no. 2, p. 215, 1974.
- [25] P. Ekman and W. V. Friesen, The Psychology of Depression: Contemporary Theory and Research. Oxford, England: John Wiley & Sons, 1974, ch. Nonverbal behavior and psychopathology, p. 318.