https://doi.org/10.31449/inf.v49i18.9236 Informatica **49** (2025) 235–256 **235**

Multimodal Depression Detection from WhatsApp Statuses Using Hybrid Feature Selection with HMO-RTH-OOA and KPCA-CCA

Shaik Rasheeda Begum, Saad Yunus Sait*
Department of Computational Intelligence, School of Computing, SRM Institute of Science and Technology, Kattankulathur, Chennai - 603203, Tamil Nadu, India
E-mail: rashee22ameen@gmail.com, saady@srmist.edu.in
* Corresponding author

Keywords: BERT, ELECTRA, EfficientNetV2L, Depression detection, HMO-RTH-OOA, KPCA-CCA

Received: May 13, 2025

Depression is a common and serious condition characterized by persistent sadness. Early detection is critical to leading a healthy and fulfilling life. Due to the rapid development of social media, more people are using it to share their thoughts and feelings. Therefore, social media sites like Facebook, Twitter, and Instagram are now developing into a significant data source that can be used for identifying depression and mental illness. In this work, we demonstrate the detection of depression using WhatsApp data. A dataset of 3 months' worth of images, text, and behavior of 100 users was collected with consent, labeled by three psychological professionals, and finalized by majority vote. We examined WhatsApp-based depression from two perspectives: (a) message-based, which uses a single WhatsApp status image with corresponding extracted text, and (b) user-based detection, which utilizes all status images, texts, and WhatsApp behavioral data of each user. BERT and ELECTRA models were employed for text feature extraction, while EfficientNetV2L was used for image feature extraction. We have proposed two feature selection approaches: a hybrid multi-objective Red-Tailed Hawk - Osprey Optimization Algorithm (HMO-RTH-OOA), which attempts to select an optimal set of features, and Kernel Principal Component Analysis - Canonical Correlation Analysis (KPCA-CCA), which reduces the dimensionality of the multi-modal dataset. Additionally, we employed data augmentation to improve the dataset count for message-based depression detection to 5514 samples. Classification was performed using standard machine learning models, including XGBoost (XGB), AdaBoost (ADA), Support Vector Machine (SVM), and Random Forest (RF). User-based depression detection achieved 93.79% accuracy with KPCA-CCA, while message-based depression detection achieved 90.46% accuracy with HMO-RTH-OOA, demonstrating significant improvements over depression detection performed on Twitter and Instagram datasets.

Povzetek: Za zgodnje odkrivanje depresije iz WhatsApp aplikacije je predlagan dvostopenjski pristop z ekstrakcijo besedilnih (BERT, ELECTRA) in slikovnih značilk (EfficientNetV2L) ter dvema metodama izbire značilk: hibridnim optimizatorjem HMO-RTH-OOA in nelinearno KPCA-CCA fuzijo.

1 Introduction

A basic component of human well-being, mental health includes emotional, psychological, and social aspects. Mental illnesses like anxiety and depression have a substantial impact on daily life, compromising cognitive capacities, physical health, decreased productivity, absenteeism, difficulty making decisions, and emotional stability[1]. So, it is essential to diagnose depression early on. It is critical for enhancing mental health and avoiding unnecessary suffering. Those in the early stages of depression may choose to utilize platforms such as Facebook and WhatsApp because they feel more at ease online. Early identification is critical, as late-stage depression can lead to dangerous decisions such as suicide. In this paper, we focused on early-stage depression detection based on the user's WhatsApp status.

Generally speaking, speaking with others about their feelings can help many people feel less depressed. However, in today's digital age, many people choose social media as a platform to express their emotions publicly [2]. People express a wide spectrum of emotions, from joy and accomplishment to despair and frustration, with friends, family, and a larger internet audience. It provides a safe area for emotional support during difficult times, and the act of sharing frequently makes people feel less alienated. While it can be a great tool for emotional expression and support, it is critical to be careful with privacy and interaction quality. Support systems are still required when dealing with serious or persistent emotional problems. Numerous studies have discovered the potential of using social media platforms like Twitter [3], Facebook, and Instagram [4] to detect signs of depression. Researchers often analyze various factors, including how many posts are posted per day and user interactions, to gain insights into individuals' mental health. WhatsApp is widely recognized as one of the most user-friendly messaging applications [5] for sharing one's ideas, feelings, and emotions with others. Currently, WhatsApp-based sentimental analyses and personality analyses are being conducted. As far as we know, no notable studies have focused on utilizing WhatsApp status to detect mental health problems such as depression. The majority of studies on social media and mental health have focused on sites such as Twitter, Facebook, and Instagram. So, the main goal of our research is to identify depression based on WhatsApp status.

Based on dataset utilization, most of the previous work concentrated on detecting depression by using various data, such as electroencephalogram (EEG) data, speech signals, facial expressions, and social media data. As mentioned earlier, there has been a growing focus in recent years on using social media data for diagnosing depression. Many works in recent years have used text-only [6], image-only [4], text-image [7], and behavioral data to detect depression. Text-based depression detection focuses on sentiment analysis and extensively uses tweets as a dataset. Image-based depression detection uses Instagram and Facebook posts as datasets and uses more suitable methods to perform depression detection. When it comes to text-image-based depression detection, posts are considered as images, and post captions are considered text. A few more works utilize the user's behavioral data as a dataset. Data gathered about a person's behaviors, interactions, and habits is referred to as behavioral data. This information is typically collected via digital platforms, physical activities, or physiological reactions. In our proposed work, we have used behavioral data and WhatsApp status screenshot as dataset. This WhatsApp status can collect both image and text simultaneously by taking a screenshot; however, other methods, like Facebook and Instagram, require us to save images separately and captions separately during dataset collection. In Section 5.3, we discuss the details of the behavioral data. We have analyzed our datasets from two perspectives: message-based depression detection (utilized WhatsApp status only) and user-based depression detection (utilized both behavioral data and WhatsApp status). Both message and user-based analysis employ WhatsApp status screenshot images and texts derived from images via Optical Character Recognition (OCR).

The primary contributions of the study can be summed up as follows:

1. Novel data focus: The majority of recent research has concentrated on studying data from social media sites like Twitter, Facebook, and Instagram [4] for depression detection. However, due to its user-friendly interface, WhatsApp has become more popular than most other social media platforms [7]. So far, no work has utilized WhatsApp data for depression detection. To fill this research gap in depression detection, we collected users' WhatsApp statuses. Our work stands out since it examines WhatsApp status data and users' behavioral data. For the first time, a method to identify depression using WhatsApp status has been proposed.

2. Proposed features selection methods: To achieve optimal performance on WhatsApp-based depression detection, we proposed two feature selection algorithms and examined user depression through two analyses, such as message-based depression and user-based depression detection. The first feature selection approach, proposed HMO-RTH-OOA, was inspired by the recent optimization algorithms, Red-Tailed Hawk (RTH) [8] and Osprey Optimization Algorithm (OOA) [9], as well as the utilization of a proposed multi-objective fitness function. The RTH [8]optimizer excels at exploration but falls short at exploitation due to its rapid acceleration nature. On the other hand, the OOA optimizer excels at exploitation but struggles with exploration due to random solution selection. By combining these two strategies with multi-objective fitness functions, we hope to capitalize on their strengths while minimizing their weaknesses. The second approach is the KPCA-CCA feature selection methods, which combine the traditional KPCA and CCA approaches. CCA may have limitations because it is sensitive to non-linear data; however, by combining KPCA with CCA, it overcomes these limitations.

3. Integrated text and image analysis: Many earlier articles successfully analyzed text-image-based methods, but they treated image captions as text data. Contrarily, our approach treats WhatsApp statuses as image formats. During data collection, instead of saving each user's images directly, screenshots of each user's images, which contain both the image and mentioned text image were taken to prevent data loss, especially if the user had attached captions to their image.

4. Combined BERT and ELECTRA text features extraction: Combining the features of BERT [10] and ELECTRA [11]may allow for the possibility of maximizing the benefits of both models while resolving their unique drawbacks. This hybrid representation of features, which incorporates both contextual and token-level information, can provide a more thorough understanding of the text.

In this work, we aimed to address the following significant research questions (RQ):

RQ1: Is an instant messaging platform suitable for depression detection?

RQ2: In case of multimodal data, is it more suitable to perform feature fusion or feature concatenation, followed by feature selection?

The forthcoming portions of the paper are summarized as follows: A comprehensive discussion of the relevant works and scientific research on the topic was covered in Section 2. The background section of the proposed work is discussed in Section 3. The proposed strategy is covered in Section 4. Dataset details are discussed in Section 5. Sections 6 and 7 describe the feature extraction and proposed feature selection methods. The results of the experimental setup are analyzed and discussed in Section 8, and the work is concluded, and some future improvements are suggested in Section 9.

2 Related works

Depression detection can be addressed through multiple approaches, including facial expression identification, physiological signal analysis, acoustic-based processing, behavioral data tracking, and text analysis. Facial expression recognition catches up on subtle expressions and decreased emotional variability, while physiological cues, including heart rate variability (HRV) and EEG patterns, offer physiological indicators of depression. Acoustic-based analysis determines speech phrasing, tone, and rhythm for signs of boredom or emotional distress. Behavioral information on social interactions, as well as sleep habits and smartphone use, may bring insight into mood swings [12].

Text-based strategies use natural language processing (NLP) to examine patterns of language in diaries, texts, and social media posts to identify signs of depression [13]. Furthermore, multi-modal approaches that integrate these techniques improve precision and offer a thorough evaluation of depression [14]. In our study, we concentrate on multimodal-based depression detection by utilizing users' behavior data and WhatsApp status, which include both image & text. Here, we have discussed the literature on depression detection based on social media platforms, optimization-based feature selection, and a feature fusion method.

2.1. Depression detection using multiple data sources

According to [22], numerous works related to the detection of depression using social data from platforms like Facebook, Instagram, and Twitter have been conducted. In [23], tested a theoretical model that examines how using Instagram relates to feelings of depression. It focuses on the effect of negative social comparison as a moderator of this association, as well as the number of strangers one follows. A study [24] discovered that individuals who tweeted with a past emphasis were more likely to have an increase in cognitive sensitivity and depressive symptoms than those who did not tweet with a previous focus. A study [4] used an MTurk survey and users' Twitter histories, and it was discovered that, despite being applied to a different population, the model was effective at distinguishing between depressive and non-depressive content and performed better than the average success rates of general practitioners in identifying depression. According to [25], sudden increases in Facebook posting activity are positively correlated with symptoms of depression.

2.2. Depression detection based on text, image, multimodal data

Most prior research has concentrated on text-based depression detection rather than image-based and multimodal techniques. These are used on many social media platforms, such as Facebook, Twitter[15], Weibo[6], Instagram[4], Vkontakte[19] and Reddit[16] for data collection. Several studies utilized single data sources (text or picture only), while a few studies used multimodal data that included both text, images, and behavioral data. Table 1 shows a comparative summary of various studies on depression detection using text/image/behavioral data. From that table, most of the studies used different machine learning, ensemble learning, and deep learning methods for depression detection. However, only a few studies focus on feature selection and fusion to enhance the performance. Study [17] employed two feature selection methods on text-based depression detection: Information Gain (IG) and Most Frequent. The Most Frequent technique chooses the most popular terms based on their frequency in the dataset. Information Gain (IG) calculates the amount that a term influences category prediction based on its existence or absence in a document. Most other studies concatenated features rather than explicitly applying feature fusion methods. To address this research gap, we employed effective optimization-based feature selection and fusion methods for depression detection. The following section discusses literature on optimization-based feature selection and feature fusion.

2.3. Feature selection by using hybrid optimization methods

In detecting depression via social media, such as text, image, and both text-image cases, notable works have focused solely on feature selection. In [26], n-gram language models, LIWC dictionaries, automated image tagging, and a bag-of-visual-words are used for feature extraction. The efficacy of the approach is evaluated using correlation-based feature selection and nine different classifiers with standard evaluation metrics. Apart from social media data, predictor variables, EEG, face, and audio-based depression detection methods use feature selection. However, those studies used traditional methods like filter and wrapper-based methods. Only a few recent works have used optimization-based feature selection methods. When compared to filter-based methods, optimization-based feature selection [27] may achieve a better balance of effectiveness and flexibility, making it ideal for complex, high-dimensional, and small datasets. To combine the strengths of different approaches, studies often use hybrid optimization techniques instead of a single technique, resulting in improved feature selection and model performance.

Numerous hybrid metaheuristic methods have been created to identify the significant and optimal features from the raw dataset, specifically for the feature selection problem. [28] utilized the position update quality of the crow search algorithm (CSA) in the grey wolf optimizer (GWO) to achieve a good balance of exploration and exploitation. GWOCSA, a hybridized version of the algorithm, was used on twenty-one popular datasets from the UCI repository. Another study [29] suggested GWO-ABC, which combines the benefits of GWO and an artificial bee colony (ABC). To improve exploration ability, this algorithm combines ABC's information-sharing property with GWO's original hunting strategy.

Table 1: A comparative summary of various studies on depression detection using text/image/behavioral data from social media

Studies	Data	Source	Feature selection	Feature Fusion	Methodology	Metrics
[6]	Text	Weibo	None	None	DISVM	Accuracy = 86.15
[15]	Text	Twitter	None	None	Gradient Boosting Decision Tree	Accuracy = 91.1%
[3]	Text	Twitter	None	None	SVM, RF, XGB	F1 score = 82.05%
[16]	Text	Reddit	None	None	Ensemble of SVM and KNN	Accuracy = 98.5%
[17]	Text	Twitter	Information Gain, most frequent	None	Linear SVM	Accuracy = 82.5%
[18]	Text	Reddit	None	None	Convolutional neural network (CNN)	Accuracy = 75.13%
[13]	Text	Twitter	None	Modality Fusion	BERT + LSTM + Textual Topic Modeling	Accuracy = 82.3%
[4]	Image	Instagram	None	None	100-tree Random Forest classifier	F1 score = 64.70%
[19]	Image	Vkontakte	None	None	LR, SVM, MLP, RF, NB, KNN, CAT and RAND	F1 score = 65.50%
[13]	Image	Twitter	None	Modality Fusion	VGG + LSTM + Visual Topic Modeling	Accuracy = 66.7%
[7]	Image-Text	Twitter	None	Early Fusion	CNN +BERT	Accuracy = 88.4%
[20]	Image-Text	Twitter	None	Early Fusion	GRU +VGG - Net + COMMA	Accuracy = 90.0%
[13]	Image-Text	Twitter	None	Modality Fusion	MTAL	Accuracy = 84.2%
[21]	Image-Text-Behavioral	Instagram	None	Early Fusion	CNN	F1 Score = 82.3%
[14]	Image-Text-Behavioral	Twitter	None	Early Fusion	SVM	Accuracy =70%

Gated Recurrent Units (GRU), Visual Geometry Group (VGG) -Net, Cooperative misoperation multi-agent (COMMA), Multimodal Topic-enriched Auxiliary Learning (MTAL), Multimodal Topic-enriched Auxiliary Learning), Logistic Regression (LR), Multi-layer Perceptron (MLP), Naive Bayes (NB), CatBoost (CAT), and random-based classifier (RAND), K-Nearest Neighbors (KNN) , Deep integrated Support Vector Machine (DISVM) .

The Study [30] combines Artificial Bee Colony with Ant Colony Optimization to optimize feature subset selection. There is a literature gap in optimization-based feature selection in social media-based depression detection. For this purpose, we use the proposed hybrid HMO-RTH-OOA feature selection methods to improve feature selection efficiency, which were inspired by two recent optimization algorithms.

2.4. Feature fusion

Feature fusion may incorporate multiple datasets at once, as long as they contain supplementary modalities like text, images, audio, EEG signals, and so on. Various techniques, including feature-level fusion, decision-level fusion, and hybrid fusion, were utilized for data fusion. In this feature-level fusion method, feature text and image features are extracted separately and then combined or transformed with the help of various methods such as direct concatenation, Canonical Correlation Analysis (CCA), Principal Component Analysis (PCA), and Kernel Canonical Correlation Analysis (KCCA). Whereas in decision-level fusion, separate models are trained for text and images, and the results are combined at the decision level. Weighted sum fusion and voting methods are generally used for decision-level fusion. Hybrid fusion utilized a combination of deep learning techniques with feature-level and decision-level feature fusion. Hybrid and decision-level fusion-based studies are often used in recent research work. Feature-level fusion enables the model to capture joint representations of text and image data, resulting in a more comprehensive understanding.

Feature-level fusion method CCA outperforms PCA by capturing cross-modal relationships, making it ideal for multi-view learning. In contrast to PCA, which only considers variability within an individual dataset, CCA looks for correlations between multiple modalities (for example, text and image). It also maintains discriminative features, which improve classification performance. However, CCA can measure the global linear correlation between sample pairs and so tends to under-fit the data features in complex nonlinear settings. Kernel CCA, a typical nonlinear form of CCA, transfers all samples into a higher-dimensional space (called feature space), followed by traditional CCA is executed [31]. In our work, we suggest KPCA-based CCA for feature fusion to incorporate nonlinearity.

Among these existing works, the majority of the work used Facebook, Instagram, and Twitter for data collection, and many works utilized machine learning algorithms, with only a small percentage using deep learning algorithms. Only a small number of studies have focused on feature selection, and as far as we are aware, no optimization-based feature selection has been used for WhatsApp image-based depression detection. Additionally, WhatsApp status images and behavioral data were not used as a dataset in any earlier research, as per our knowledge. To address this research gap, we use WhatsApp status and behavior data and employ two different proposed feature selection methods.

3 Background section

In this study, we proposed two different feature selection algorithms for depression detection. The proposed HMO-RTH-OOA optimization algorithm was inspired by two recent optimization algorithms, such as RTH and OOA, that were proposed specifically for the optimization problem. These two optimizers are effective in optimization problems, although they have significant limitations when used for feature selection. To overcome this limitation, we hybridize these optimizers along with the proposed Multi-objective Fitness function. The second feature selection approach is based on fusion, namely KPCA-CCA-based fusion. In the following section, we discuss details of traditional methods and their limitations.

3.1. RTH optimization algorithm

The RTH algorithm [8] mimics this strategy of natural hunting to quickly and successfully address optimization issues in the real world. It is motivated by the shrewd predator, the red-tailed hawk, and the way it hunts. Three stages make up the hawk's hunting process: high soaring, low soaring, and swooping. The hawk searches the search space during the stage of high soaring to find the region where its prey is located. It moves about the chosen area during the low soaring stage to get the best location point for the hunt. Finally, after choosing the right position and time in the prior step, the hawk swooped its prey by stooping and increasing its acceleration speed (from 32-64 to 190 km/h) in a curved path. Because of its stooping and swooping nature, it may be useful for optimization-related problems, but it may have some difficulties when used for feature selection applications. In general, in feature selection applications, other agent solutions were updated based on the current best agent solution. In RTH, the hawk acceleration speed nature causes the current best position value to change dramatically while position updates, either reaching the maximum of the upper bound or the lower bound, resulting in the selection of more or no features. So, when utilizing RTH for feature selection, it works well for the exploration stage but struggles in the exploitation stage due to its nature of high-speed acceleration. Exploration in optimization involves discovering new and diverse areas of the solution space, whereas exploitation concentrates on refining and improving the best solutions discovered so far. The mathematical expression for RTH's optimization algorithm is discussed below:

3.1.1. High soaring

The hawk will soar far into the air in search of the area with the most readily available food. A mathematical model for this stage is represented by Eq. (1).

$$X^{new}{}_{i,j}(t) = A^* + \left(\hat{X} - X_{i,j}(t-1)\right) * L * T_f(t) \quad (1)$$

where $X_{i,j}(t)$ denotes the i^{th} (search agents) red-tailed hawk position of j^{th} feature at iteration of t, A^* denotes the global best position (best agents position) so far achieved, $\hat{X}$ denotes the mean of the search agents' positions, TF(t) stands for the transition factor function, which can be determined using Eq. (3), and Levy (L) denotes the levy flight distribution function, which can be derived using Eq. (2).

$$L = \delta \frac{u * \partial}{|v|^{\beta - 1}} \quad where \ \partial = \frac{\Gamma(1+\beta) * sinsin \frac{\pi\beta}{2}}{\Gamma\left(1+\frac{\beta}{2}\right) * \beta * 2^{\left(1-\frac{\beta}{2}\right)}} \quad (2)$$

$$T_f(t) = 1 + sin\ sin\ \left(2.5 + \frac{t}{MT}\right) \quad (3)$$

where u and v are random values [0 to 1], δ is a constant (0.01), MT is a max iteration, and β is a constant (1.5).

3.1.2. Low soaring

The red-tailed hawk surrounds its prey by soaring in a spiral manner, significantly nearer to the ground. The structure of its model can be described as follows:

$$X^{new}{}_{i,j}(t) = A^* + (a_t + b_t) * X^{new}{}_{i,j}(t) \quad (4)$$

The following formulas can be used to get the direction coordinates of a and b.

$$a_t = \Upsilon_t * sin\ sin\ \theta_t\ , b_t = \Upsilon_t * cos\ cos\ \theta_t \quad (4.1)$$

$$\Upsilon_t = \Upsilon_o * \left(R - \frac{t}{MT}\right) * rand()$$

$$\theta_t = A_G * \left(R - \frac{t}{MT}\right) * rand()$$

where A_G for the gain on the angle [5-15], Υ_o stands for the radius's initial value [0.5-3], R for a control gain [1, 2] and rand for a random gain [0-1].

3.1.3. Stooping & swooping

In this step, the hawk immediately stoops and attacks the target from the best position achieved during the low soaring stage. This phase can be formulated as follows

$$X^{new}{}_{i,j}(t) = \psi(t) * A^* + \mathbb{a}_t * Stepsize1(t) + \mathbb{b}_t * Stepsize2(t) \quad (5)$$

The formula below can be used to determine each step size:

$$Stepsize1(t) = X^{new}(t) - T_f(t) * \text{ mean } (X^{new})$$

$$Stepsize2(t) = \Phi(t) * X^{new}(t) - T_f(t) * A^*$$

where the acceleration factor ψ and the gravitational constant Φ, can be described as follows:

$$\psi(t) = sin^2\left(2.5 - \frac{t}{i}\right)$$

$$\Phi(t) = 2 * \left(1 - \frac{t}{i}\right)$$

where ψ(t) denotes the hawk's acceleration, which rises as t increases. As previously discussed, the hawk's high-speed acceleration causes the other position to change dramatically during position updates in the features selection application. It abruptly shifts each agent's location to the lower or upper boundaries, selecting either no features or all features. It is not suitable for feature selection, but it is beneficial for optimization tasks. To address the exploitation problem in RTH, we combined OOA exploitation with RTH high and low soaring.

3.2. Osprey optimization algorithm (OOA)

An innovative metaheuristic algorithm called the OOA [9] was developed in response to the natural Osprey hunting behavior. The original concept for OOA came from Ospreys, which are known for their exceptional fishing abilities. It mimics the osprey's hunting method, which entails locating the prey, capturing it, and then moving it to a location where it may be eaten. OOA is built with a mathematically modeled two-phase architecture that combines exploration and exploitation to mimic the behavior of ospreys naturally engaged in hunting. This special method effectively addresses actual optimization issues by utilizing osprey-inspired strategies. This OOA feature selection is poor at the exploration stage due to the solution being chosen randomly, but at the same time, it works well in the exploitation stage.

3.2.1. OOA

The osprey attacks one of these fish at random after detecting its position. Based on the modeling of the osprey's flight towards the fish, (6) is applied to determine a new position for the relevant osprey.

$$X^{new}{}_{i,j}(t) = X^{new}{}_{i,j} + R(t) + A^* - I(t) * X^{new}{}_{i,j} \quad (6)$$

Where $R(t)$, $I(t)$ random number interval between 0 to 1.

After this new position is determined by using Equation (7)

$$X^{new}{}_{i,j}(t) = X^{new}{}_{i,j}(t) + \frac{L_\rho + R(t) * (U_\rho - L_\rho)}{MT} \quad (7)$$

Where MT represents the maximum iteration count, U_ρ is an upper bound and L_ρ is a lower bound.

As far as we know, no work has been done on feature selection based on Red-Tailed Hawk and Osprey optimization. As discussed earlier, the RTH optimizer is good at the exploration stage, but poor at the exploitation stage due to its rapid acceleration speed. On the other side, the OOA optimizer is good at the exploitation process at the same time poor in the exploration process due to random solution selection. By combining these two strategies, we hope to capitalize on each's strengths while limiting their respective limitations.

3.3. Kernel principal component analysis (KPCA)

KPCA is an improved version of Principal Component Analysis that is designed to reduce nonlinear dimensionality. KPCA employs the kernel's operation to translate data into a higher-dimensional space where linear PCA can be performed. This method captures complex, nonlinear relationships in data, making it especially useful for pattern recognition, image processing, and data visualization. Polynomial, Gaussian (RBF), and linear kernels are among the most commonly used. KPCA expands PCA's capabilities, enabling more flexible and powerful data analysis.

3.4. Canonical correlation analysis (CCA)

The CCA is used to analyze the relationships between two sets of variables. To find linear combinations of variables from each feature set that are maximally correlated, CCA examines two datasets, in contrast to approaches that concentrate on a single dataset. To do this, vectors that maximize the correlation between these new variables referred to as canonical variates, must be found.

CCA is frequently used for data fusion, which allows the integration of multiple datasets to provide comprehensive insights. CCA facilitates the integration of information from various sources by identifying linear combinations of variables from each dataset that are maximally correlated. CCA may have limitations because it is sensitive to non-linear data; however, by combining KPCA with CCA, it overcomes these limitations.

4 Proposed methodology

Our proposed study focuses mostly on the analysis of WhatsApp status for the early diagnosis of depression. In this study, depression is analyzed by two different methods: message-based depression detection and user-based depression detection. Figure 1 represents a general block diagram of the message and used based depression detection workflow. A detailed description of the dataset collection was provided in Section 5, followed by the dataset collection, data preprocessing was performed. After preprocessing, the text was extracted from the image using the OCR method.

Following this, text features were extracted using BERT-ELECTRA, and image features were extracted from EfficientNetV2L. Following feature extraction, both text and image features undergo two feature selection processes, such as the proposed HMO-RTH-OOA and KPCA-CCA. In message-based depression detection, both image and text features were selected using two feature selection methods, and performance was evaluated using a variety of machine learning algorithms, including XGBoost (XGB), ADA boost (ADA), Random Forest (RF), and support vector machine (SVM). In contrast to message-based depression detection, we used another data modality termed behavioral data. The description of behavioral data collection was discussed in section 5.3. After text and image-based feature extraction, the means of each user's features were discovered before feature selection. For instance, if a user has 15 image samples, the extracted features (both image and text features) from these 15 images were averaged to obtain a representative feature vector. Once features were selected using HMO-RTH-OOA and KPCA-CCA, statistical features were applied to the results of the selected features, as well as behavioral data. Finally, these two statistical feature sets were combined to find user-based depression detection.

5 Dataset creation

The dataset was collected from the general public after obtaining their informed consent. We choose people mostly from the Asia region, particularly India. We contacted them through our close contacts and their contacts. These people mostly range in age from 18 to 45, including college students and other working and non-working professionals. We distributed a consent form via Google Forms to obtain consent before collecting their WhatsApp status and behavioral data information for research purposes.

5.1. Data collection

Our main objective is to create a system for analyzing WhatsApp statuses to identify signs of depression. To do this, we collected the 100 participants' contact numbers and stored them on our mobiles with their permission. We closely monitored their WhatsApp status for 3 months, collected their WhatsApp status image, and behavioral data of specific individuals. Approximately 10 to 20 screenshots were collected on average from each user. We labeled our dataset in collaboration with 3 psychological specialists, and the respective kappa scores are stated in Table 2. Out of the 100 users, 30 were determined to have depression, and the other 70 were deemed to be non-depressed. Figure 2 represents the sample image of the WhatsApp status of depressed and non-depressed users.

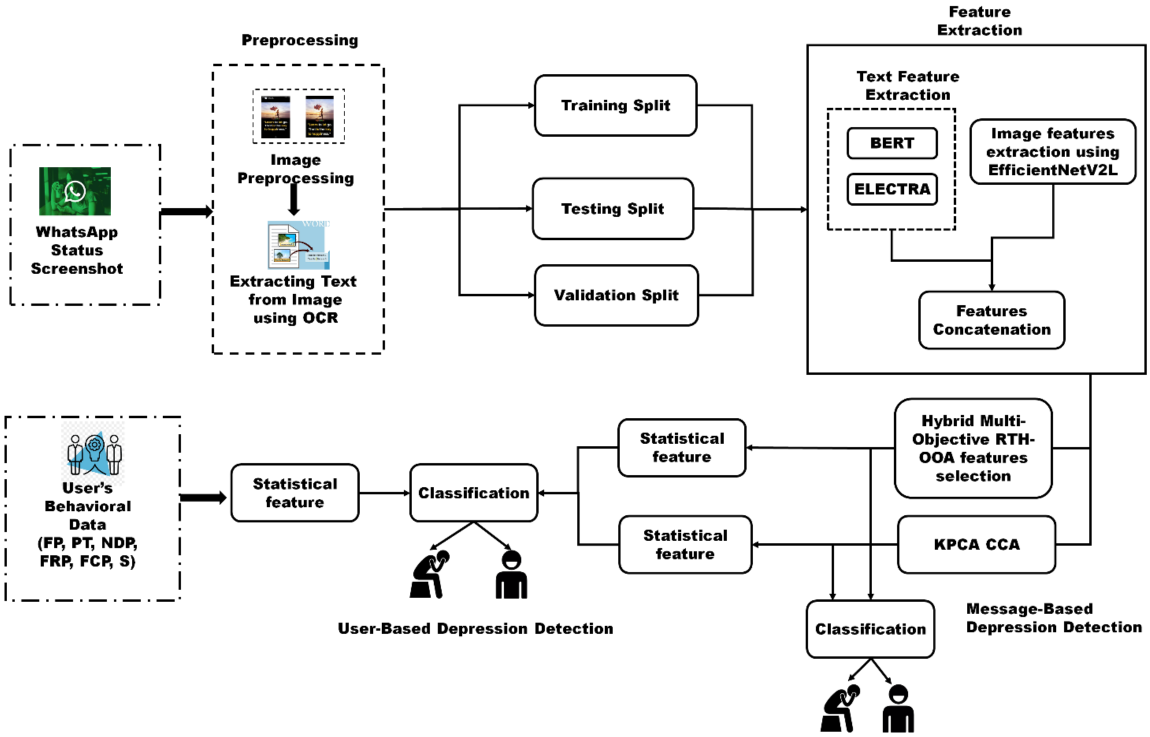


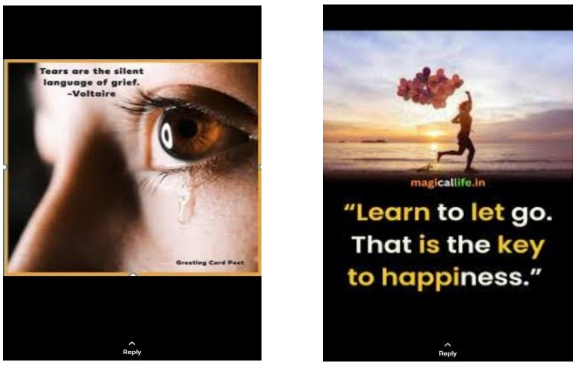
Figure 1: Block diagram of message and user-based depression detection

5.2. Inter-rater agreement

Cohen's kappa score is commonly used to evaluate the reliability and consistency of label annotators, especially in situations involving categorical data. Cohen's kappa is a measure of agreement between two annotators who classify items into mutually exclusive categories. It considers chance agreements, providing a more robust measure than simple percent agreement. The kappa score, which ranges from 0 to 1. At 1, there is a perfect agreement, whereas 0 indicates no agreement beyond chance. Negative values indicate poorer-than-expected agreement. To interpret the kappa values, assess the level of agreement: (<0.00) as indicating no agreement, (0.01-0.20) as none to slight, (0.21-0.40) as fair, (0.41 – 0.60) as moderate, (0.61-0.80) as substantial, and (0.81 to 1.00) as almost perfect agreement [32].

Table 2: Kappa score for different annotators

S No	Annotators	Cohen's Kappa Score
1	Doctor A	0.8820
2	Doctor B	0.8598
3	Doctor C	0.8623



(a) (b)

Figure 2: Samples of (a) Depressed and (b) Non depressed post

This scale allows us to comprehend the consistency and reliability of classifications in categorical data analysis. In our study, the final label was decided by the majority vote of two annotators. The computed Kappa score was then used to determine inter-rater reliability between each annotator and the final label. Table 2 clearly shows that all three annotators' kappa values exceed 0.85. It indicates all three doctors have kappa values in the "almost perfect" agreement category, indicating a very high level of agreement amongst them in their classifications.

5.3. Behavioral data

The behavioral data involves summarizing and interpreting patterns, trends, and characteristics observed in the behavior of individuals or groups. Our user-based depression detection system incorporates a multi-modal strategy that considers both behavioral data and WhatsApp status images. The following behavioral data attributes are taken into consideration: FP (Frequency of Posts), PT (Post Timing), ND (No Profile Picture), FRP (Frequently Removing a Profile), FCP (Frequency of Changing a Profile), and S (Sleeping Pattern). This complete approach combines behavioral data with visual information from WhatsApp status images to improve the accuracy and efficacy of our depression diagnosis procedure.

Table 3: Behavioral data analysis

Behavioral data	Depressed User	Non-depressed User
Frequency of Posts (FP)	More Frequent	Rare
Post Timing (PT)	Mid night	Moring
No Display Picture (NDP)	No	No
Frequently Removing a Profile (FRP)	Yes	No
Frequency of Changing a Profile (FCP)	Daily	Monthly once
Sleeping Pattern (S)	Rare	Normal

By utilizing these various data sources, we desire to obtain a greater understanding of the user's mental state, strengthening and improving the detection process. Table 3 shows sample behavioral data for both users, and Figure 3 represents the distribution of behavioral attributes for depressed and non-depressed users. This behavioral data is utilized in user-based depression detection methods. Behavioral data is converted to a label encoding format before being used. Followed by the feature extraction and selection of images and text, their statistical features were computed. These statistical features and label-coded features are combined and used for user-based depression detection.

5.4. Data preprocessing

Following the dataset's collection, the top and bottom portions of the images were removed in preprocessing. We collected a WhatsApp status screenshots dataset, as shown in Figure 4 (a). Screenshots of WhatsApp statuses that include both text and image information, as well as unnecessary information like their name and battery percentage.

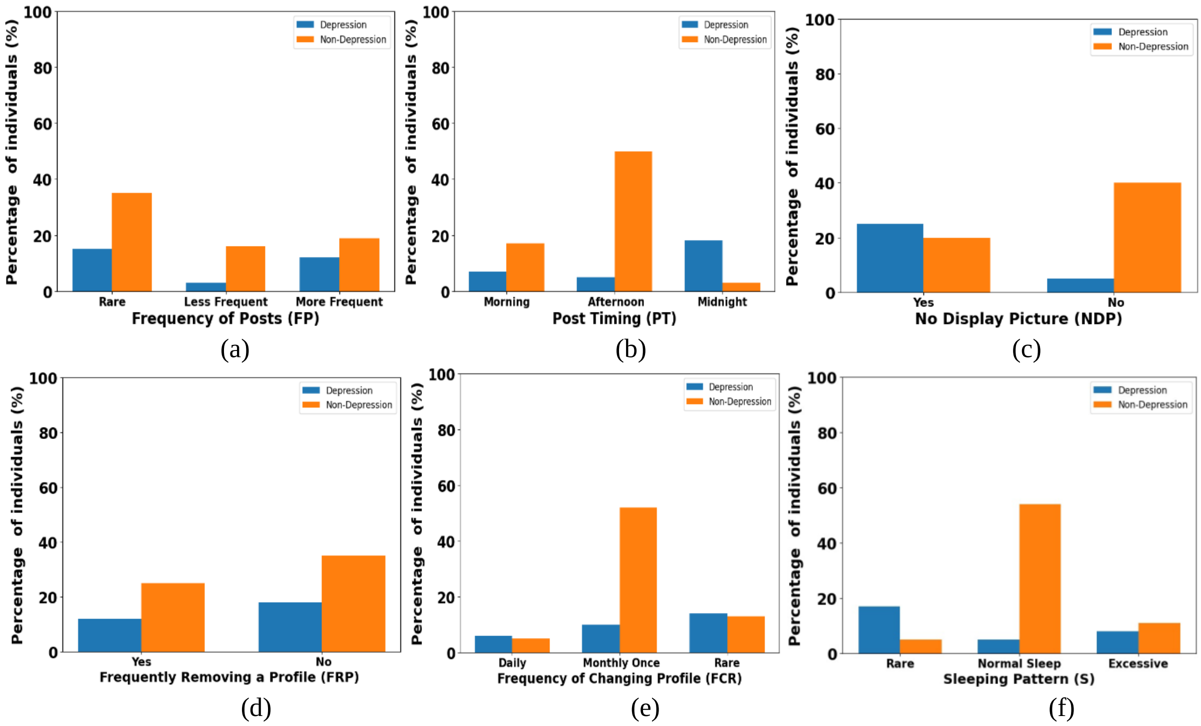


Figure 3: Distribution of behavioral attributes for depressed and non-depressed users: (a)Frequency of posts, (b) Post Timing, (c) No display picture, (d) Frequently removing a profile, (e) Frequency of changing profile, (f) Sleeping Pattern

Also, some users add a caption when posting. Therefore, we first removed the top and bottom parts of the gathered screenshot of WhatsApp statuses. The sample preprocessed image is given in Fig. 4 (b). Following this step, further text and image preprocessing were performed.

(a) (b)

Figure 4: (a) Before preprocessing (b) After Preprocessing

To ensure the best data quality and readiness for analysis, it is necessary to preprocess the text and image data separately before starting with feature extraction.

- **Image preprocessing:** As the EfficientNetV2L [33] architecture is used for feature extraction, the images were resized to a standard size of (224,224) pixels.
- **Text preprocessing:** The Optical Character Recognition (OCR) method is used to extract text from preprocessed WhatsApp status images. OCR-based text extraction achieved a 93.48% accuracy rate. After extracting the text, unnecessary symbols, URLs, and other components are removed from the text. It ensures that the text data is clean and ready for further processing. Following data preprocessing, data augmentation and feature extraction are performed on both image and text using different methods.

5.5. Data augmentation

As mentioned before, we collected a dataset from 100 users and collected 10 to 20 WhatsApp status images from each user. Due to dataset imbalance and a minimal dataset, we have used both image and text augmentation. We used data augmentation techniques only for message-based depression detection. Image augmentation techniques include image rotation, flipping, and noise addition. For text augmentation, synonyms, exchanging characters, randomly removing characters, and creating spelling problems are utilized. Before data augmentation, the dataset consisted of 491 samples in non-depression and 214 samples in depression. This class imbalance was resolved during the augmentation step to increase the model's robustness and generalization. After augmentation, we get 2946 non-depressed and 2568 depressed WhatsApp image statuses, which include both image and text. The class non-depression class comprises 53.4% of the total samples, while the class depression comprises 46.6%. The class ratio between non-depression and depression is around 1.15:1, indicating a well-balanced dataset suited for classification tasks. For further processing, these 5514 messages were divided into training, testing, and validation in the ratio of 70%, 15%, and 15%.

6 Feature extraction methods

Feature extraction is performed on preprocessed images. Text features are extracted using a combination of BERT and ELECTRA models, while image features are extracted using EfficientNetV2L.

6.1. Combined BERT and ELECTRA Text features extraction

Text extracted from an image was fed into BERT[10] and ELECTRA to extract contextual character-level and token-level understanding features. The state-of-the-art natural language processing (NLP) models BERT and ELECTRA were both created utilizing transformer architecture. They are made to pre-train on numerous text datasets, and then become fine-tuned for different downstream NLP jobs. Combining BERT and ELECTRA characteristics is a strategic technique in NLP. BERT was selected because of its deep bidirectional transformer design, which efficiently captures contextual dependencies in text. It is ideal for downstream language understanding tasks like classification and semantic analysis because of its large corpus pre-training. On the other hand, ELECTRA [11] was preferred because it provides a more sample-efficient pre-training technique delivering higher performance over various benchmarks while using fewer computational resources. In contrast to RoBERTa and ALBERT, the two selected pre-trained models, BERT and ELECTRA, achieved higher performance in depression detection. It is commendable that BERT can extract detailed contextual information from text, whereas ELECTRA focuses on token-level features. It is possible to combine the strengths of contextual comprehension with token-level discrimination by combining the features of both models. This hybrid representation provides a more complete knowledge of the text, making it versatile and robust across a wide range of NLP applications.

Furthermore, it handles each model's restrictions individually, providing a versatile and high-performance solution for a wide range of natural language comprehension challenges.

6.2. Efficient NetV2L for image feature extraction

The preprocessed image was fed to EfficientNetV2L [33] to extract image features, which is well-known for its effectiveness and efficiency in gathering rich visual features. EfficientNetV2L was employed for image score-based visual tasks because of its faster training time and compound scaling technique. Its scalable and optimized architecture enables it to outperform other algorithms like VGG and ResNet while preserving efficiency, making it a strong choice for image-based feature extraction in depression detection.

Generally, in computer vision problems, the EfficientNet family of neural network topologies strikes the ideal compromise between model performance and computing efficiency. It uses a technique known as compound scaling to start with the basic model, EfficientNetB0, and simultaneously change the model's depth, width, and picture resolution to produce larger or smaller versions of the network. The effective building blocks based on mobile inverted bottleneck designs are essential to its effectiveness. Before being fine-tuned for specific applications, Efficient Net models are often pretrained on big datasets like ImageNet to capture general picture properties. In the search for effective yet high-performing neural network topologies, the Efficient Net family has become a popular option. Features from BERT (1 x 768), Electra (1 x256), and efficientV2L (1 x 1280) are utilized for the next feature selection process.

7 Proposed features selection methods

In this section, we used two different feature selection mechanisms, such as the proposed HMO-RTH-OOA and KPCA-CCA, for both message and user-based depression detection and analyzed the performance of each feature selection mechanism in both cases.

The algorithmic pseudo-codes for feature extraction and selection are illustrated below.

Input: Preprocessed image and text
Output: selected features

1. Extract text features using BERT and ELECTRA
   - Load the pretrained model (BERT and ELECTRA)
   - For each text instance
     Feed the preprocessed text into the model (BERT and ELECTRA)
     Extract the hidden layer embeddings
     Store the extracted features vector
   - Combine all feature vectors
2. Extract image feature using EfficientNetV2L
   - Load the pretrained model EfficientNetV2L
   - For all images
     Feed the preprocessed image to the model
     Extract the image features
3. Concatenate both the image and text features
4. Perform the HMO-RTH-OOA feature selection as per Algorithm 1.
   - Initialize all the required parameters
   - Initialize the solution for all search agents and find the best solution
   - Update the search agents' positions as per algorithm 2.
   - Compute the best agent as per algorithms 3 and 4.
   - If the current best fitness > global best fitness, update the global best fitness and solution; otherwise, retain the previous global best.
   - It will be repeated until it reaches the maximum iteration
   - Finally, determine the selected index by binary the global best position

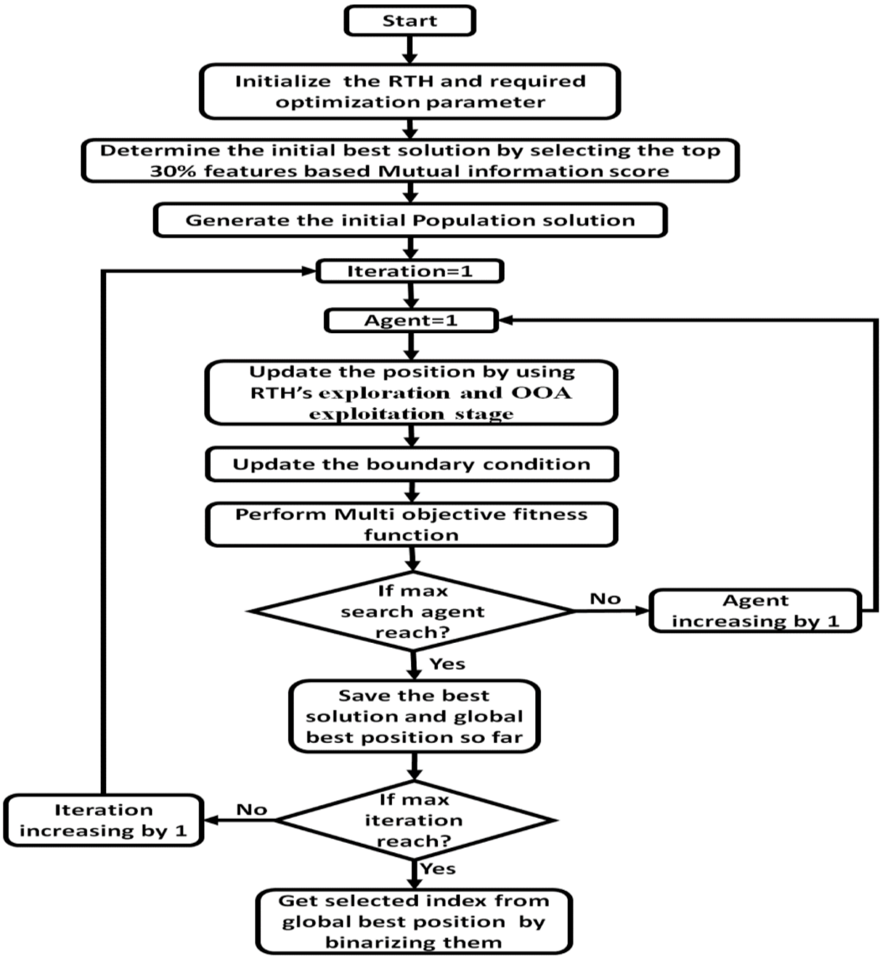


Figure 5: Proposed HMO-RTH-OOA optimization algorithm for feature selection

7.1. Proposed hybrid multi-objective HMO-RTH-OOA optimizer

In our proposed HMO-RTH-OOA feature selection method, we use RTH's exploration stage and OOA's exploitation stage along with the proposed multi-objective fitness function. We propose a hybrid strategy that addresses the limitations of each optimizer while combining the benefits of the RTH and OOA optimizers. RTH excels in exploration but struggles with exploitation due to its rapid acceleration nature. The detailed limitations of RTH's exploitation are discussed in section 3.1. Generally, during the RTH exploitation stage, the hawk grabs the prey and raises its acceleration to fly through the sky. In the feature selection application, the selected prey was considered the best agent, and other agent positions were updated accordingly. So, it changes dramatically as the position updates, either reaching the upper or lower bound, resulting in the selection of more or fewer features. On the other hand, OOA struggles with exploration because it selects solutions randomly. However, it does well in the exploitation stage.

By combining these two approaches, we can develop a comprehensive strategy that takes advantage of both RTH's exploration and OOA's exploitation stages. Along with these hybridizations, we used a proposed multi-objective fitness function to improve its performance. The flow chart of the proposed feature selection HMO-RTH-OOA is given in Figure 5. Initially, the required parameters were initialized. The initial best solution and fitness function were evaluated using the mutual information score. Afterward, the main loop executed. During the main loop search agents' positions were updated through RTH exploration stages, such as high soaring and low soaring, as well as OOA exploitation stages. Lastly, boundary conditions were updated, and the fitness was computed using the proposed multi-objective fitness function. Following the execution of the main loop, the global best solution was binarized, and the selected feature indexes were indicated by 1.

Algorithm 1, from lines 1 to 2, represents the determination of the initial best solution (global best solution) using a mutual information score. The mutual information (MI) score examines the dependence of two variables. It determines how much understanding one

ALGORITHM 1: General view of the Proposed hybrid multi-objective HMO-RTH-OOA

Input: General optimization parameters: Search agent M, Max iteration MT, upper bound U_ρ=1, lower boundL_ρ =0, total number of features n , Threshold value $l = 0.5$, Input Features F.

Output: $S -$ Set of features

1. Compute mutual information scores $I(F_i; C)$ between all features F_i and respective target C
   $\boldsymbol{m_{j_{[1\,x\,n]}}} = [\,MI(F_1; C), MI(F_2; C), \dots\dots\dots\dots.. MI(F_n; C)]$
2. Select top 30% of features F_i based on m_j
   $T = \{F_i \mid F_i \,\epsilon\, S \;\; and\; MI\;(F_i\,, C)\; is\; among\; top\; 30\;\%\}$
3. Compute $\boldsymbol{V} = \{v_j\}_{j=1}^{n} \quad where\; v_j = I_T\;(F_j)$
4. Generate the initial best solution based on V_j
   $\boldsymbol{A}^* = \left(v_j\;\, u_j\; + \;(1 - v_j)\; l_j\right)_{1\,x\,n}\;\; where\; u_j = random(0.5{,}1) and\; l_j = random\;(0{,}0.5)$
   Where u_j and l_j is a random value between 0.5 to 1, and 0 to 0.5, respectively.
5. Compute the initial fitness for the initial best solution A^*
   $f^* = \boldsymbol{FITNESS}\;(V, F, C)$
   Where f^* represents the best fitness function value
6. Initialize the Search agents randomly
   $\boldsymbol{X_{i,j}} = uniform(0{,}1)\; where\; i \in [1, M] and\; j \in [1, n]$
   Where $\boldsymbol{X_{i,j}}$ is a random value between 0 to 1, i is the search agent number M, and j is the features number n.
7. Repeat
   $\boldsymbol{X^{new}} \;=\; \boldsymbol{AGENT\_UPDATE}\;(\boldsymbol{X}, \boldsymbol{A}^*)$
   $(\,f, \boldsymbol{A}) \;= \boldsymbol{BEST\_AGENT}\;(\boldsymbol{X^{new}})$
   If $f \;>\; f^*$ then set
   $\boldsymbol{A}^* = \boldsymbol{A}, \boldsymbol{X} = \boldsymbol{X^{new}}, and\; f^* = f$
   Else
   Previous value of $\boldsymbol{A}^*, \boldsymbol{X}$ and f^* were maintained.
   Until it reaches the maximum number of iterations.
8. Determine the selected index S by binarizing the global best position

variable reduces uncertainty about the other. The mutual information between two independent variables F and C can be calculated as follows:

$$MI(F, C) = \sum_{c\in\mathcal{C}} \sum_{f\in\mathcal{F}} P_{(F,C)}(f, c) \log\left(\frac{P_{(F,C)}(f, c)}{P_F(f)\; P_C(c)}\right)$$

where $P_{(F,C)}$ is the joint probability mass function of F and C and P_F and P_C are the marginal probability mass functions of F and C respectively.

Mutual information scores were calculated for all features and their corresponding targets, and the top 30 percent of features were selected based on the mutual information score. These selected features index is treated as one, whereas the non-selected features index is treated as zero. From these binary format feature indexes, the initial best solution was generated as per lines 3 and 4. In line 5, the initial best fitness (global best fitness) was calculated by using algorithm 4. In line 6 the initial search agents' position was randomly assigned. After determining the initial best solution and fitness, as per line 7, the search agent's position is updated using RTH's high soaring, low soaring, and OOA's position updates, followed by updating the boundary condition. After updating all search agent's positions, the best agent and corresponding best fitness were computed. If the current best fitness exceeds the global best fitness, then the current best fitness is considered the global best fitness, and the corresponding solution is considered the global best solution; Otherwise, the previous value is maintained. It will be repeated until it reaches the maximum iteration. Finally, determine the selected index by binary the global best position A*.

ALGORITHM 2: AGENT_UPDATE($\mathbf{X},\; \boldsymbol{A}^*$)

1. Begin
2. $\boldsymbol{X^{new}}$ =HIGH_SOARING($\mathbf{X}$,$\boldsymbol{A}^*$)
3. $\boldsymbol{X^{new}}$ =LOW_SOARING ($\boldsymbol{X^{new}},\; \boldsymbol{A}^*$)
4. $\boldsymbol{X^{new}}$ =OSPREY ($\boldsymbol{X^{new}}$,$\boldsymbol{A}^*$)

5. For each search agent
6. For each dimension of search agent
7. $$\boldsymbol{X^{new}}_{\boldsymbol{i,j}} = \begin{cases} L_\rho & if \quad \boldsymbol{X^{new}}_{\boldsymbol{i,j}} < L_\rho \\ U_\rho & \text{if} \quad U_\rho > \boldsymbol{X^{new}}_{\boldsymbol{i,j}} \end{cases}$$
8. End for
9. End for
10. Return $\boldsymbol{X^{new}}$
11. End

Algorithms 2 and 3 represent the overview of agent update and best agent detection. In algorithm 2, lines 2 and 3 carry out the RTH exploration stage, whereas line 4 is the OOA exploitation stage. In these three steps, the position was updated using the corresponding equation, which is stated in Section 3. In lines 5 to 8, after the exploration and exploitation stages, the boundary position was updated and returned to the corresponding algorithm. If the current position is greater than the upper bound, the position is updated as 1; if the current position is less than the lower bound, the position is updated as 0; otherwise, the position remains unchanged. After updating the agents, algorithm 3 was used to determine the best agent position.

ALGORITHM 3: $BEST\_AGENT\ (\boldsymbol{X}, f^*)$

1. Begin
2. $fmax = 0$
3. For each row $\boldsymbol{X_i} = \{x_{i,j}\}_{i=1}^{n}\ of\ X$
4. Binarize the $\boldsymbol{X}_i$ by $\boldsymbol{B_i} = \left\{I\left(x_{i,j}\right)\{z \mid z \geq l\}\right\}_{j=1}^{n}$
5. $f_i = FITNESS(\boldsymbol{B_i}, F, C)$
6. if $f_i > fmax$
7. $fmax = \boldsymbol{f_i}$
8. $\boldsymbol{A} = \boldsymbol{X}_i$
9. End if
10. End For
11. Return **A**,$fmax$
12. End

In algorithm 3, agents' positions are initially binarized, as stated in line 4. In line 5 fitness values were estimated using Algorithm 4. In lines 6 to 8, if the current search performs well, it updates the maximum fitness function and search agent position; otherwise, the previous value is maintained. The process will continue until the maximum number of search agents is reached.

ALGORITHM 4: $FITNESS(B, F, C)$

The proposed multi-objective fitness function needs to satisfy three objectives. Algorithm 4 represents the pseudocode of the multi-objective fitness function. When calculating the fitness function, it should consider the satisfaction of all three objectives (line 4, line 5, and line 7). Where C is a binary label that contains 0's and 1's, where 0 indicate non-depression and 1 indicate depression.

1. Begin, Let $f = 0$
2. Let $\boldsymbol{B} = \{\boldsymbol{b_j}\}_{j=1}^{n}$
3. Compute $N' = \sum_{j=1}^{n} \boldsymbol{b_j}$
4. if $N' > 0.2 * n$ and
5. $\left(\sum_{i=1}^{n} I_{\{F_i \mid MI(F_i, C) \neq 0\}}(F_i) > 0.3\ N'\right)$
6. Then $S = \{F_i \mid \boldsymbol{b_i} \neq 0\}$
7. $f = KNN\ (S, C)$
8. return f
9. End

Line 2: The vector $\boldsymbol{b} = [b_1, b_2, \dots \dots .. b_n]$ is a binary vector associated with feature set F. Each b_i indicates whether the corresponding feature F_i is selected (1) or not selected (0). It has a selected feature set, and it will continue to explore Objective 1.

Objective 1: Lines 3 and 4 represent the first objective that the number of selected features should exceed 20% of the total count of original features. The total original feature size is n = 2304; Selection criteria were developed to ensure that the number of retained features exceeded 460, which is more than 20% of the original feature set. If Objective 1 is not met, the function returns f as 0 and stops the assessment of the objective function.

Objective 2: Once Objective 1 was met, the process proceeded to examine Objective 2, as represented in Line 5. In objective 2, we calculate the mutual information score between the selected feature set and the respective target variable. It provides a mutual score for each selected feature. A higher score signifies that the feature has a greater ability to predict and a stronger relationship with the target variable. In contrast, a lower score signifies that the feature is less efficient for predicting the target. In our case, we are considering features with non-zero scores only because it indicates the relevance of the features to the target. At the same time, the considered features should exceed 30% of the objective one selected feature count (i.e., non-zero Mi score 30% of >460 features). Failure to satisfy Objective 2 causes the function to return f as 0, stopping the objective function evaluation.

Objective 3: Line 7 indicates the third objective function, whereas line 6 selects the final selected feature set. Now, the accuracy (Acc) was calculated with the help of a KNN classifier for the selected feature set S and target variable C.

7.2. Kernel principal component analysis-canonical correlation analysis feature fusion

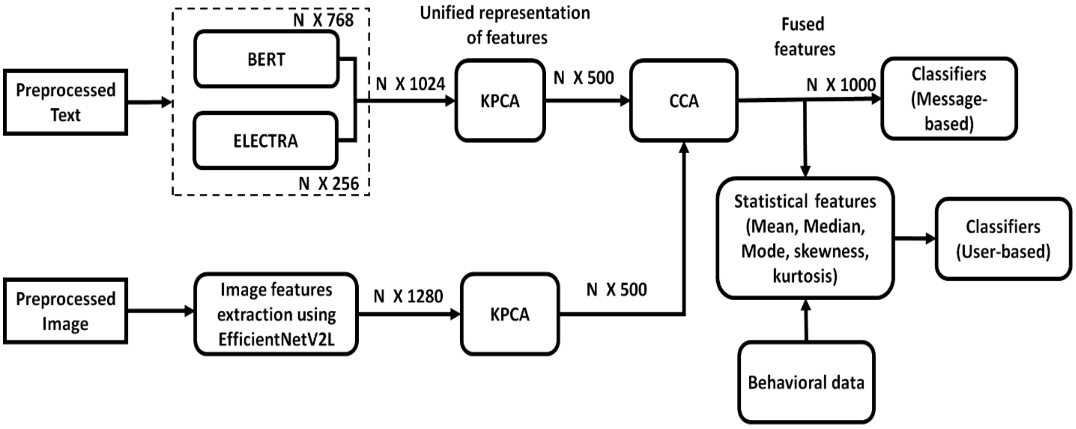


Figure 6: General block diagram of KPCA- CCA

In depression detection KPCA-CCA fusion method was utilized, and performance was analyzed with the help of a classifier. In KPCA-CCA, KPCA [34] converts image and text features into a unified representation, followed by the use of CCA [31] for fusion purposes.

CCA is used to detect patterns of relationships and dependencies between features, which include images and text. Figure 6 represents the general block diagram of KPCA-CCA. The feature fusion process comprises the following steps:

1. Concatenate the features extracted from BERT and ELECTRA together.
2. Convert nonlinear data to a unified representation each with help of Kernel PCA.

D1 ←Perform KPCA to text features with RBF kernel and gamma=1

D2 ←Perform KPCA to image features with RBF kernel and gamma=1

3. Reducing the dimensionality of the transformed data to 500 dimensions.

Utilize Canonical Correlation Analysis (CCA) to fuse the text features and image features. After fusing the features, the next step involves computing statistical features (Mean, Median, Mode, skewness, kurtosis) for the fused feature set. For user-based detection, we used a combination of behavioral and statistical features to analyze performance; for message-based detection, we only used fused features.

8 Results and discussion

In this section, we examine the performance of depression detection in 2 ways:

- Message-based depression detection
- User-based depression detection

For both message and user-based depression detection, the proposed feature selection algorithm, such as HMO-RTH-OOA and KPCA-CCA feature selection algorithm, was used, and classification was performed using machine learning algorithms such as XGB, ADA, SVM, and RF. We chose machine learning classifiers instead of deep learning approaches due to the relatively low dataset size. We selected SVM, ADA, RF, and XGB because of their effective performance on small to medium-sized datasets, particularly in previous research on depression detection and text/image classification. Ensemble approaches such as RF, ADA, and XGB are well-known for their ability to capture non-linear interactions while reducing overfitting via tree aggregation. Whereas SVM provides strong decision boundaries in sparse feature spaces. In this work, we prefer machine learning methods for classification over deep learning for a few reasons. ML models such as SVM, XGBoost, AdaBoost, random forests, and KNN perform well on small datasets, whereas models based on deep learning require large datasets to generalize effectively. Using a small dataset on a deep learning (DL) model can cause a number of issues, such as overfitting, poor generalization, and fluctuations in performance.

8.1. Message-based depression detection

In message-based depression detection, performance was analyzed using both feature selection methods, and XGB, ADA, SVM, and RF were utilized for classification. We employed Google Colab platform as the training setting for all model implementations. The training was performed using the NVIDIA Tesla T4 GPU, Python 3.10. Hyperparameter tuning was carried out utilizing the Bayesian search approach. A random seed of 42 was set to maintain reproducibility. To ensure a balanced class distribution in both the training, testing, and validation

Table 4: Performance analysis of message-based depression detection

Classifiers	Metrics	Without feature selection	With features Selection					
			Based on the optimization method			Based on the Fusion method		
			HMO-RTH-OOA	RTH	OOA	KPCA-CCA (no of components =500)	KPCA-CCA (no of components =700)	KPCA-CCA (no of components =1000)
XGBoost (XGB)	Mean Accuracy	85.78	90.46	82.8	83.35	71.93	74.09	78.89
	Precision	86.86	90.56	85.11	84.46	73.83	75.86	80.93
	Recall	86.68	90.12	83.21	84.21	73.37	75.56	78.94
	F1 score	86.99	89.51	81.7	83.7	73.57	75.70	79.91
	Std	0.063	0.02	0.031	0.03	0.01	0.01	0.007
Ada Boost (ADA)	Mean Accuracy	84.12	88	78.55	82.04	66.37	68.46	68.49
	Precision	83.12	88.61	80.54	85.54	67.62	70.35	72.32
	Recall	81.14	87.88	79.12	82.04	70.80	70.93	66.13
	F1 score	82.98	86.5	78.02	78.78	69.16	70.62	69.99
	Std	0.052	0.026	0.029	0.029	0.01	0.01	0.01
SVM	Mean Accuracy	79.45	87.15	70.39	73.21	53.66	53.73	53.95
	Precision	78.14	87.2	68	75.11	53.50	53.74	53.37
	Recall	78.85	87.15	73.59	74.04	69.65	69.74	70.98
	F1 score	79.25	87.15	79.19	73.21	60.34	69.88	69.51
	Std	0.051	0.022	0.06	0.068	0.09	0.01	0.01
Random Forest (RF)	Mean Accuracy	80.47	87.64	70.7	71.06	67.89	70.58	74.25
	Precision	79.98	88.44	73.21	72.41	68.73	71.94	79.39
	Recall	78.56	87.14	71.21	73.96	72.96	73.84	69.72
	F1 score	79.68	87.69	69.9	79.6	70.51	72.86	74.23
	Std	0.051	0.027	0.03	0.03	0.01	0.01	0.01

sets, we utilized a stratified train/test split with a 70:15:15 ratio. We used user-level separation to avoid data leakage, ensuring that there was no content overlap between the training, testing, and validation sets, and fully assigning all samples from a user to individual groups. We ran this experiment 10 times to ensure a robust and reliable evaluation of the model's performance.Table 4 represents the average validation performance analysis of with and without feature selection methods. We have analyzed our performance in terms of mean accuracy, precision, recall, F1 score, and standard deviation (std). Low Standard deviation represents the robustness of the model. Due to data augmentation and feature extraction scheme, it attains notable accuracy even without feature selection while using those machine learning algorithms. The dataset size increased to 5514 samples after amplification, which improved the diversity of the training data and reduced class imbalance. This enabled improved generalization and stable model performance. In this scenario, XGB attains a high performance of 85.78 % accuracy when compared to all other methods.

Our proposed feature selection method, HMO-RTH-OOA, attains 90.46% accuracy on XGB. The traditional RTH and OOA methodologies attain 82.8 % and 83.35 % accuracy, and it performs poorly when compared to those without feature selection. As discussed earlier, RTH is poor at the exploitation stage due to the hawk's high-speed acceleration nature. On the other hand, OOA demonstrates somewhat better performance compared to RTH, but less performance when compared to without feature selection. OOA is poor at the exploration stage because it randomly selects the solution (search agents).

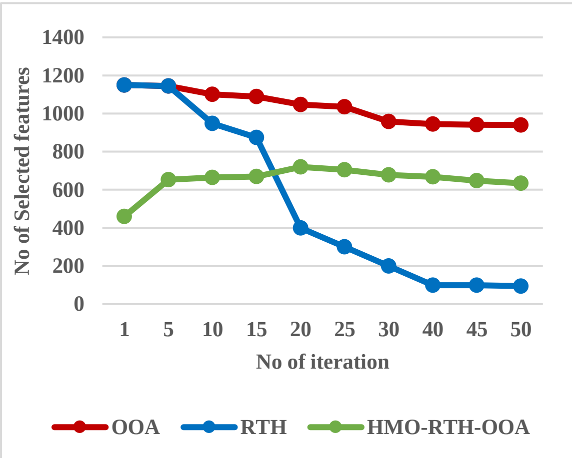


Figure 7: No of iterations vs no of selected features

To overcome each model's limitations, we have utilized the proposed features selection method HMO-RTH-OOA. Figure 7 represents the number of features selected per iteration for the 3-feature selection methods. Due to random population initialization, at the very first iteration, it selects approximately half of the original feature size in the case of RTH and OOA. In RTH, the iteration of 20 selects fewer than 400 features; subsequently, it decreases to 90 to 95 features due to its own limitations, such as rapid acceleration. Meanwhile, OOA maintains a maximum number of features selected scenario until it reaches max iteration. In the case of the proposed feature selection method, we calculated the initial population and fitness function by using the mutual information method instead of random initialization. So, at the very first iteration, it maintained the selected features as 460; afterward, it was improved to 630 to 650 features. With the help of the XGB classifier, HMO-RTH-OOA attains 5.45% increased accuracy, 9.25% increased accuracy, and 8.53 % increased accuracy when compared to without feature selection, with feature selection (RTH only), and with feature selection (OOA), respectively. XGB consistently performs better than SVM, RF, and ADA in various analyses.

XGB is based on gradient boosting, which builds trees sequentially, correcting errors made by previous trees. The model can concentrate on difficult situations, such as complex relationships between features and the target, noisy data, and eventually increase overall accuracy using this iterative methodology. Ada boosts it attain better accuracy when compared to SVM and RF. ADA develops one learner at a time, iteratively fixing previous learners' errors by modifying the weights of the training data. As a result, convergence may be slower than XGB. ADA may lose effectiveness and become more prone to overfitting in feature spaces with high dimensions. With its flexible learning rate and regularization, XGB is better able to handle high-dimensional data.

Additionally, we analyzed the performance of message-based depression detection by using the KPCA-CCA feature fusion method. The performance of KPCA-CCA depends upon the number of components selected in KPCA while employing KPCA-CCA features fusion for message-based dependency detection. Table 4 shows that KPCA-CCA accomplishes 71.93%, 66.37%, 53.66%, 67.89%; 74.09%, 68.46%, 53.75%, 70.58%; and 78.89%, 68.49%, 53.95%, 74.25% accuracy for all four classifiers XGB, ADA, SVM and RF when the number of components is 500,700 and 1000, respectively. It demonstrates that increasing the number of components while doing KPCA in KPCA-CCA increases accuracy. Message-based depression detection without feature selection and HMO-RTH-OOA feature selection achieved 85.78% and 90.46% accuracy, respectively, whereas the KPCA-CCA fusion method achieved 78.89% accuracy. It shows that using KPCA-CCA achieves a 14.66% decrease in accuracy compared to HMO-RTH-OOA in message-based depression detection when using the XGB classifier.

Optimization algorithms outperform KPCA in feature selection due to their direct optimization of task-specific objectives (e.g., accuracy, error), efficiently handling high-dimensional and noisy data using robust search strategies, and providing explicit feature selection rather than transforming features, resulting in more relevant and interpretable results. Utilizing the XGB classifier, without a feature selection method, achieved a precision of 86.86% and a recall of 86.68%, RTH obtained 85.11% precision and 83.21% recall, and OOA obtained 84.46% precision and 84.21% recall. KPCA-CCA recorded a precision of 80.93% and a recall of 78.94%, whereas the HMO-RTH-OOA method outperformed the rest with the highest recall of 90.56% and a precision of 90.12%. We evaluated the model using precision and recall to address class imbalance, as these metrics are more informative than precision in unbalanced datasets.

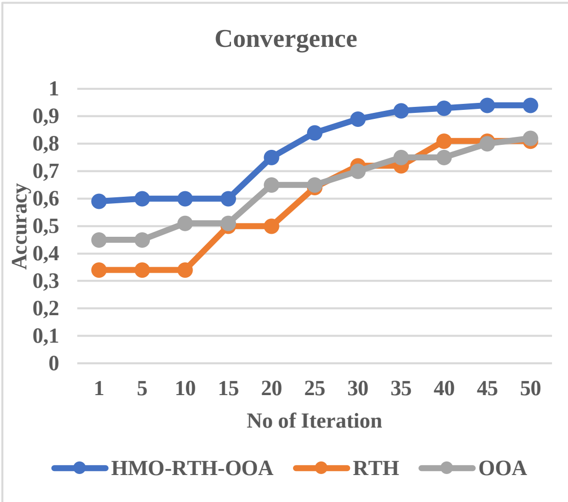


Figure 8: Convergence behavior of proposed HMO-RTH-OOA, RTH only, and OOA only

Recall focuses on accurately identifying depressed individuals, while precision helps ensure that non-depressed individuals are not mislabeled. These findings indicate that the hybrid feature selection HMO-RTH-OOA method was well-suited for sensitive tasks where minimizing false negatives is essential, such as depression detection. Figure 8 represents the convergence behavior of the HMO-RTH-OOA, RTH, and OOA algorithms. The method achieved rapid convergence within 30 iterations and stable performance with low oscillations, indicating reliable optimization. On average, this HMO-RTH-OOA method takes 1564 seconds to complete 50 iterations.

Table 5: Performance comparison of proposed HMO-RTH-OOA with other standard methods (Message-based)

Classifier	LASSO	RFE	ReliefF	HMO-RTH-OOA
XGB	82.04	81.42	83.99	90.46
ADA	75.65	72.13	72.05	88.00
SVM	80.40	76.78	74.43	87.15
RF	73.78	73.85	70.62	87.64

Table 5 represents the Performance comparison of the proposed HMO-RTH-OOA with other standard methods. Our method outperforms standard methods such as LASSO [35], RFE [35], and the ReliefF [36] feature selection method. Additionally, we employed paired t-tests to compare the performance of the HMO-RTH-OOA and other standard feature selection methods. The results show that HMO-RTH-OOA vs. LASSO (p = 0.021), HMO-RTH-OOA vs. RFE (p = 0.034), and HMO-RTH-OOA vs. ReliefF (p = 0.011) exhibit statistically significant differences in classifier performance. From these results, we determined that our method is significantly better than other standard feature selection methods.

In addition, our proposed WhatsApp-based depression detection method achieved an accuracy of 90.46%, outperforming several existing approaches, such as the CNN-based model by [21] with 82.3%, the SVM model by [14] with 70%, the MTAL approach by [13] with 84.2%, and the GRU + VGG-Net+COMMA model by [20] with 90.00%. Existing tasks have mostly been used without image-text-behavior integration to detect depression; however, some tasks have used this combination but failed to strengthen their fusion mechanism. Study [21] directly combines these features and sends them to the CNN model. In [13], the Modality Fusion method was used, which concatenates the image and text sequence, passes through an LSTM to get a new representation. KPCA-CCA-based fusion outperforms basic feature concatenation because it successfully captures complicated nonlinear interactions and optimizes the correlation between modalities. In our work, we used WhatsApp images with user behavior data, the robust feature fusion mechanism KPCA-CCA, and the HMO-RTH-OOA feature selection mechanism for depression detection.

8.2. User-based depression detection

In these analyses, we considered user-based message data along with behavioral data. In our dataset, we have 30 depressed and 70 non-depressed. To attain robust performance, we used 10-fold validation, which involved performing the experiment ten times with different training, testing, and validation sets. At each fold validation, we randomly select 30 non-depressed users from 70 users. At each N-fold, we considered 30 depressed and 30 non-depressed. The 10-fold validation average performance is given in Table 6. The incorporation of behavioral data alongside text and image data significantly improves the overall system performance, even in the absence of fusion methods. The accuracy without fusion approaches is noticeably less due to the intrinsic non-linear relationship between picture text features and labeled behavioral data.

While PCA-CCA obtains an accuracy of 81.93% when dealing with linear data, our suggested KPCA-CCA handles non-linear complexities, resulting in a significant accuracy boost to 93.79%. This demonstrates the efficacy of our proposed strategy in capturing complicated interactions in the multi-modal dataset. According to these findings, XGB outperformed ADA, SVM, and RF, particularly when using KPCA-CCA over PCA-CCA. KPCA recognizes complex, nonlinear relationships in data that PCA may miss, yielding more informative features. These improved features, when fed into a powerful classifier like XGB, can result in better overall model performance than linear PCA features. We analyzed the performance of HMO-RTH-OOA in user-based depression detection.

Table 6: Performance analysis of user-based depression detection

Classifiers	Metrics	Without Fusion	Features selection		
			Based on the fusion method		Based on the optimization method
			KPCA – CCA	PCA-CCA	HMO-RTH-OOA
XGB	Mean Accuracy	79.70	93.79	81.93	80.48
	Precision	81.21	94.92	80.85	78.35
	Recall	79.43	92.91	81.62	78.56
	F1 score	74.74	93.25	86.61	78.38
	Std	0.086	0.020	0.027	0.012
ADA	Mean Accuracy	72.81	80.51	77.80	75.71
	Precision	76.76	80.97	78.12	78.51
	Recall	74.31	81.25	84.70	77.81
	F1 score	73.10	81.74	79.56	78.65
	Std	0.058	0.087	0.048	0.075
SVM	Mean Accuracy	79.04	82.09	78.58	67.99
	Precision	79.67	84.88	79.03	75.99
	Recall	79.77	84.15	80.86	56.85
	F1 score	77.90	81.09	83.02	63.44
	Std	0.026	0.045	0.043	0.035
RF	Mean Accuracy	78.09	81.40	80.05	66.48
	Precision	81.47	84.33	83.94	76.41
	Recall	78.53	81.83	79.85	51.05
	F1 score	79.11	84.29	75.52	59.88
	Std	0.054	0.035	0.064	0.086

In user-based depression detection without fusion and using KPCA-CCA features fusion, we achieved approximately 79.70%, 72.81%, 79.04%, 78.09%, and 93.79%, 80.51%, 82.09%, 81.40% accuracy for all four classifiers, XGB, ADA, SVM, and RF, respectively. While we achieved 80.48% accuracy when using HMO-RTH-OOA with the help of XGB. When utilizing HMO-RTH-OOA for user-based depression, the XGB's accuracy decreased by 16.53% compared to KPCA-CCA.

This demonstrates that HMO-RTH-OOA performs worse even without feature fusion due to a limited number of samples in user-based depression detection; we use 30 depressed and 30 non-depressed users. The minimum number of samples produces a model that performs well on training data but poorly on test data when calculating fitness values for the HMO-RTH-OOA optimization algorithm.

In user-based depression detection without a feature selection method, a precision of 81.21% and a recall of 79.43% were achieved, whereas the PCA-CCA method attained 80.85% precision and 81.62% recall. The HMO-RTH-OOA accomplished 78.35% precision and 78.56% recall, while KPCA-CCA achieved the maximum recall of 94.92% and a precision of 92.91% when compared to other methods. When it comes to detecting depression, recall is more crucial than precision. In real-world applications, if a non-depressed individual is wrongly identified as depressed, they may receive unneeded medical consultation. However, false negatives constitute a greater risk because failure to identify a truly depressed person has more serious consequences and associated higher costs.

8.3. Analysis of message and user-based depression detection performance

In message-based depression detection, HMO-RTH-OOA accomplished a 5.45 % increase in accuracy when compared to the scenario without feature selection, a 9.25% increase in accuracy compared to the scenario with feature selection (RTH only), and an 8.53% increase in accuracy compared to the case with feature selection (OOA) only. This analysis shows that HMO-RTH-OOA performs the best when compared to the RTH and OOA categories alone. In user-based analysis, PCA-CCA attains less performance than KPCA- CCA due to CCA is not good at nonlinear data. Overall, KPCA-CCA accomplished a 14.47% increase in accuracy when compared to the PCA-CCA features fusion, and a 17.67% increase in accuracy compared to the without features fusion. Upon analyzing both message and user-based approaches, it becomes evident that user-based depression detection is more stable when compared to message-based depression detection. This higher performance is related to the integration of behavioral data in user-based analysis, demonstrating the need to incorporate multiple data modalities for a more comprehensive understanding of depressed tendencies.

8.4. Ablation study

In this section, we apply an ablation study to methodically analyze the impact of each component in our proposed methods. In particular, we investigate the influence of each key module, such as hybrid feature selection HMO-RTH-OOA and multimodal input features.

8.4.1. Ablation study on multi-model data

We investigate user-based depression detection methodologies to assess the efficiency of our proposed multimodal data framework. The performance of individual input sources (image-only and text-only) and their combined (image-text-behavioral) representation is shown in Table 7. Multimodal data achieved a mean accuracy of 93.79%, whereas image-only input yielded 71.66% and text-only input yielded 86.11% when using the XGB classifier and the KPCA-CCA method. Although these individual methods demonstrate reasonable performance, complementary and comprehensive information across visual, linguistic, and behavioral dimensions when used in isolation. The combination of image, text, and behavioral data leads to significantly better depression detection performance by recording visual expressions, verbal patterns, and digital behavior.

Table 7: Ablation study on multimodal data on user-based depression detection

Classifiers	Metrics	Image only	Text Only	Multi model data
XGB	Accuracy	71.66	86.11	93.79
	Precision	79.82	84.66	94.92
	F1 Score	74.53	86.09	92.91
	Recall	72.01	88.89	93.25
	Std	0.124	0.071	0.020
ADA	Accuracy	77.5	76.66	80.51
	Precision	82.61	80.76	80.97
	F1 Score	78.91	76.73	81.25
	Recall	77.26	78.96	81.74
	Std	0.112	0.092	0.087
SVM	Accuracy	70.00	79.44	82.09
	Precision	87.80	81.57	84.88
	F1 Score	69.44	81.04	84.15
	Recall	60.53	86.88	81.09
	Std	0.084	0.138	0.045
RF	Accuracy	67.5	78.33	81.40
	Precision	90.91	84.63	84.33
	F1 Score	63.95	76.73	81.83
	Recall	55.37	74.84	84.29
	Std	0.180	0.058	0.035

Table 8: Ablation study on HMO-RTH-OOA feature selection on message-based depression detection

Classifiers	Metrics	Without a multi-objective function			With a multi-objective function		
		RTH	OOA	HMO-RTH-OOA	RTH	OOA	HMO-RTH-OOA
XGB	Accuracy	81.97	78.82	85.46	82.8	83.35	90.46
	Precision	81.31	78.29	84.84	85.11	84.46	90.56
	F1 Score	82.43	79.24	85.97	83.21	84.21	90.12
	Recall	83.59	80.21	87.15	81.7	83.7	89.51
	Std	0.003	0.007	0.009	0.031	0.03	0.02
ADA	Accuracy	73.64	74.44	75.48	78.55	82.04	88
	Precision	74.94	75.40	76.14	80.54	85.54	88.61
	F1 Score	76.49	77.43	79.43	79.12	82.04	87.88
	Recall	78.25	79.88	83.13	78.02	78.78	86.5
	Std	0.008	0.008	0.006	0.029	0.029	0.026
SVM	Accuracy	67.92	70.52	82.10	70.39	73.21	87.15
	Precision	67.43	70.54	82.04	68	75.11	87.2
	F1 Score	68.62	71.11	82.86	73.59	74.04	87.15
	Recall	69.86	71.69	83.72	79.19	73.21	87.15
	Std	0.004	0.010	0.006	0.06	0.068	0.022
RF	Accuracy	68.88	67.02	74.06	70.7	71.06	87.64
	Precision	65.91	64.27	71.83	73.21	72.41	88.44
	F1 Score	73.10	70.52	77.86	71.21	73.96	87.14
	Recall	81.98	78.06	85.05	69.9	79.6	87.69
	Std	0.007	0.019	0.019	0.03	0.03	0.027

8.4.2. Ablation study on proposed hybrid feature selection

In this experiment, we assessed the impact of incorporating a multi-objective fitness function into three different optimization algorithms—RTH only, OOA only, and a hybrid of RTH-OOA.

Both the proposed multi-objective fitness function and a conventional single-objective fitness function were used to test each optimization-based feature selection algorithm. As stated earlier, two distinct modules were included in this HMO-RTH-OOA feature selection algorithm: the multi-objective fitness function module and the hybridization of RTH and OOA. The performance of individual and hybrid optimization techniques with and without a multi-objective fitness function is shown in Table 8.

From Table 8, it's clear that the inclusion of a multi-objective fitness function significantly improved the performance of individuals and hybrid optimization-based feature selection algorithms. When using XGB, RTH average accuracy improved from 81.97% to 82.8%, OOA from 78.82% to 83.35%, and the hybrid of RTH-OOA from 85.46% to 90.46%. Most notably, the hybrid of RTH-OOA with a multi-objective fitness function accomplished the highest performance among all. These findings show that the optimization methods can more effectively balance trade-offs between feature relevance and redundancy by using a multi-objective fitness function, providing feature subsets that are more generalizable and robust.

9 Conclusion

Depression detection is vital to society due to the significant risks that depressed individuals face; it should therefore be remedied at the earliest. The widespread use of social media offers an opportunity to detect depression in a non-intrusive manner. To this end, this work demonstrated how depression detection can be performed using WhatsApp data. We approached this task from two different perspectives: message-based depression detection and user-based depression detection. While the message-based approach attempts to detect depression using a single WhatsApp status image with corresponding text, user-based approach attempts to detect depression using all WhatsApp status images of a user with corresponding text and behavioral features. We first used the BERT and ELECTRA models to extract text features. By combining BERT with ELECTRA, we attempted to capture both contextual and token-level features, giving the features a more comprehensive representation. EfficientNetV2L was utilized for image feature extraction. We then proposed HMO-RTH-OOA and KPCA-CCA feature selection techniques to improve performance by reducing the features. As RTH optimizer is known to perform well in exploration stage, but not in exploitation stage, and vice versa for OOA optimizer, we leveraged the benefits of both optimizers and made their roles complementary with RTH being used for exploration and OOA used for exploitation. We utilized KPCA-CCA feature fusion to perform fusion of image and text data. Finally, we used standard machine learning methods, including SVM, RF, ADA, and XGB for classification.

From the results, user-based depression detection attained 93.79% accuracy using KPCA-CCA feature selection, and message-based depression detection attained 90.46% accuracy with HMO-RTH-OOA feature selection. Experiments were repeated 10 times, and we noted stable performance. The potential implication of our work is its ability to identify individuals who may be at risk of depression. Such identification can serve as a preliminary screening tool, with further confirmation obtained through clinical evaluation. The most significant implication is that our model demonstrates good accuracy while maintaining marginal error rates, making it a promising approach for practical applications. In future research, we plan to explore speech-based emotional and linguistic indicators, as voice cues are more prevalent. Tone, pitch, and hesitation patterns can enhance text and image-based analysis, resulting in greater robustness.

Acknowledgements: We thank Dr. R. Arul Saravanan and his colleagues from SRM Medical College Hospital and Research Centre, SRM Nagar, Potheri, Chengalpattu, Tamil Nadu 603203, for their assistance in verifying and labeling our dataset.

Data availability: Data supporting the findings of this study are not publicly available due to participant privacy.

References

[1] Taoussi, Chaimae, Soufiane Lyaqini, Abdelmoutalib Metrane, and Imad Hafidi. Enhancing machine learning and deep learning models for depression detection: A focus on smote, roberta, and cnn-lstm. *Informatica,* (49): 14 2025. https://doi.org/10.31449/inf.v49i14.7451 .

[2] Ríssola, Esteban A., David E. Losada, and Fabio Crestani. A survey of computational methods for online mental state assessment on social media. *ACM Transactions on Computing for Healthcare* 2(2):1-31,2021 https://doi.org/10.1145/3437259

[3] Liaw, Aik Seng, and Hui Na Chua. Depression detection on social media with user network and engagement features using machine learning methods.In *2022 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*, pp. 1-6. IEEE, 2022, https://doi: 10.1109/iicaiet55139.2022.9936814 .

[4] Reece, Andrew G., and Christopher M. Danforth. Instagram photos reveal predictive markers of depression." *EPJ Data Science* 6(1):15,2017. https://doi:10.1140/epjds/s13688-017-0110-z

[5] Takkar, Hardik, Krishna Sharma, Nagadamudi Vijay Karun Kumar, Manish Kumar Jha, Bal Krishna Saraswat, and Himani Tyagi. "MemoChat App: An Advancement in Existing WhatsApp." In *2024 1st International Conference on Advances in Computing, Communication and Networking (ICAC2N)*, pp. 22-28. IEEE, 2024, 10.1109/icac2n63387.2024.10895869 .

[6] Ding, Yan, Xuemei Chen, Qiming Fu, and Shan Zhong. A depression recognition method for college students using deep integrated support vector algorithm. *IEEE access,* 8: 75616-75629, 2020. https://doi: 10.1109/access.2020.2987523 .

[7] Lin, Chenhao, Pengwei Hu, Hui Su, Shaochun Li, Jing Mei, Jie Zhou, and Henry Leung. Sensemood: depression detection on social media." In *Proceedings of the 2020 international conference on multimedia retrieval*, pp. 407-411. 2020, https://doi:10.1145/3372278.3391932 .

[8] Ferahtia, Seydali, Azeddine Houari, Hegazy Rezk, Ali Djerioui, Mohamed Machmoum, Saad Motahhir, and Mourad Ait-Ahmed. Red-tailed hawk algorithm for numerical optimization and real-world problems.*Scientific Reports,* 13(1): 12950,2023.https://doi:10.1038/s41598-023-38778-3 .

[9] Dehghani, Mohammad, and Pavel Trojovský. Osprey optimization algorithm: A new bio-inspired metaheuristic algorithm for solving engineering optimization problems. *Frontiers in MechanicalEngineering* ,8,2023 https://doi.org/10.3389/fmech.2022.1126450.

[10] Cortiz, Diogo. Exploring transformers in emotion recognition: a comparison of bert, distillbert, roberta, xlnet and electra. *arXiv preprint arXiv:2104.02041* ,2021. https://doi:10.1145/3562007.3562051 .

[11] Zhang, Shunxiang, Hongbin Yu, and Guangli Zhu. An emotional classification method of Chinese short comment text based on ELECTRA.*Connection Science,* 34(1):254-273, 2022. https://doi: 10.1080/09540091.2021.1985968

[12] Xu, Xuhai, Prerna Chikersal, Janine M. Dutcher, Yasaman S. Sefidgar, Woosuk Seo, Michael J. Tumminia, Daniella K. Villalba et al. Leveraging collaborative-filtering for personalized behavior modeling: a case study of depression detection among college students. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies,* 5(1):1-27,2021. https://doi:10.1145/3448107 .

[13] An, Minghui, Jingjing Wang, Shoushan Li, and Guodong Zhou. Multimodal topic-enriched auxiliary learning for depression detection. In *proceedings of the 28th international conference on computational linguistics*, pp. 1078-1089. 2020, https://doi: 10.18653/v1/2020.coling-main.94 .

[14] De Choudhury, Munmun, Michael Gamon, Scott Counts, and Eric Horvitz. Predicting depression via social media. In *Proceedings of the international AAAI conference on web and social media*. 7(1): 128-137,2013. https://doi.org/10.5220/0009168602350240

[15] Skaik, Ruba, and Diana Inkpen. Using twitter social media for depression detection in the canadian population. In *Proceedings of the 2020 3rd Artificial Intelligence and Cloud Computing*

Conference, pp. 109-114. 2020. https://doi.org/10.1145/3442536.3442553

[16] Adarsh, V., P. Arun Kumar, V. Lavanya, and G. R. Gangadharan. Fair and explainable depression detection in social media. *Information Processing & Management,* 60(1): 103168, 2023. https://doi:10.1016/j.ipm.2022.103168 .

[17] AlSagri, Hatoon S., and Mourad Ykhlef. "Machine learning-based approach for depression detection in twitter using content and activity features. *IEICE Transactions on Information and Systems*, 103(8):1825-1832,2020. https://doi:10.1587/transinf.2020edp7023 .

[18] Kim, Jina, Jieon Lee, Eunil Park, and Jinyoung Han. A deep learning model for detecting mental illness from user content on social media. *Scientific reports*, 10(1): 11846, 2020. https://doi:10.1038/s41598-020-68764-y

[19] Maxim, Stankevich, Nikolay Ignatiev, and Ivan Smirnov. Predicting depression with social media images.In *Proc. 9th Int. Conf. Pattern Recognit. Appl. Methods*, 2: 128-138. 2020. https://doi.org/10.5220/0009168602350240

[20] Gui, Tao, Liang Zhu, Qi Zhang, Minlong Peng, Xu Zhou, Keyu Ding, and Zhigang Chen. Cooperative multimodal approach to depression detection in twitter. In *Proceedings of the AAAI conference on artificial intelligence*, 33(01):110-117, 2019. https://doi.org/10.1609/aaai.v33i01.3301110

[21] Huang, Yu-Ching, Chieh-Feng Chiang, and Arbee LP Chen. Predicting Depression Tendency based on Image, Text and Behavior Data from Instagram. *DATA*, 3(1): 32-40,2019. https://doi.org/10.5220/0007833600320040

[22] Kim, Jiin, Zara A. Uddin, Yena Lee, Flora Nasri, Hartej Gill, Mehala Subramanieapillai, Renna Lee et al. A systematic review of the validity of screening depression through Facebook, Twitter, Instagram, and Snapchat. *Journal of affective disorders* 286: 360-369, 2021. https://doi.org/10.1016/j.jad.2020.08.091

[23] Lup, Katerina, Leora Trub, and Lisa Rosenthal. Instagram# instasad? exploring associations among instagram use, depressive symptoms, negative social comparison, and strangers followed. *Cyberpsychology, Behavior, and Social Networking*, 18(5):247-252,2015. https://doi.org/10.1089/cyber.2014.0560

[24] Sasso, Maria P., Annaleis K. Giovanetti, Anastasia L. Schied, Hugh H. Burke, and Gerald J. Haeffel. "# Sad: Twitter content predicts changes in cognitive vulnerability and depressive symptoms." *Cognitive Therapy and Research,4*3(4):657-665,2019. https://doi:10.1007/s10608-019-10001-6 .

[25] Smith, Robert J., Patrick Crutchley, H. Andrew Schwartz, Lyle Ungar, Frances Shofer, Kevin A. Padrez, and Raina M. Merchant. "Variations in Facebook posting patterns across validated patient health conditions: a prospective cohort study. *Journal of medical Internet research*, 19(1), 2017. https://doi.org/10.2196/jmir.6486 .

[26] Safa, Ramin, Peyman Bayat, and Leila Moghtader. Automatic detection of depression symptoms in twitter using multimodal analysis. *The Journal of Supercomputing*, 78(4):4709-4744,2022. https://doi: 10.1007/s11227-021-04040-8

[27] Meesala, YV Nagesh, Ajaya Kumar Parida, and Anima Naik. Optimized feature selection using modified social group optimization. *Informatica* 48(11), 2024. https://doi: doi.org/10.31449/inf.v48i11.6160 .

[28] Arora, Sankalap, Harpreet Singh, Manik Sharma, Sanjeev Sharma, and Priyanka Anand. A new hybrid algorithm based on grey wolf optimization and crow search algorithm for unconstrained function optimization and feature selection.*Ieee Access* 7: 26343-26361, 2019. https://doi: 10.1109/access.2019.2897325 .

[29] Gaidhane, Prashant J., and Madhav J. Nigam. A hybrid grey wolf optimizer and artificial bee colony algorithm for enhancing the performance of complex systems. *Journal of computational science* 27:284-302,2018. https://doi: 10.1016/j.jocs.2018.06.008 .

[30] Comert, Serap Ercan, and Harun Resit Yazgan. A new approach based on hybrid ant colony optimization-artificial bee colony algorithm for multi-objective electric vehicle routing problems. *Engineering Applications of Artificial Intelligence* 123: 106375, 2023. https://doi.org/10.1016/j.engappai.2023.106375

[31] Zhang, Xiaowei, Jing Pan, Jian Shen, Zia ud Din, Junlei Li, Dawei Lu, Manxi Wu, and Bin Hu. Fusing of electroencephalogram and eye movement with group sparse canonical correlation analysis for anxiety detection. *IEEE Transactions on Affective Computing* , 13(2):958-971, 2020. https://doi: 10.1109/taffc.2020.2981440 .

[32] Cohen, Jacob. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological bulletin* , 7(4): 213, 1968. https://doi.org/10.1037/h0026256

[33] Devi, M. Shyamala, J. Arun Pandian, Mahesh Baburao Lonare, and Yeluri Praveen. Efficient net transfer learning based early prediction of monkey pox lesion. In *2022 3rd International Conference on Smart Electronics and Communication (ICOSEC)*, pp. 1297-1300. IEEE, 2022, https://doi:10.1109/icosec54921.2022.9952099 .

[34] Shen, Junxian, and Feiyun Xu. Method of fault feature selection and fusion based on poll mode and optimized weighted KPCA for bearings. *Measurement* , 194: 110950, 2022. https://doi:10.1016/j.measurement.2022.110950 .

[35] Dahr, Jasim Mohammed, and Alaa Sahl Gaafar. Performance Evaluation of the Filter, Wrapper, Mutual Information Theory, and Machine Learning Feature Selection Methods for XGBoost-Based Classification Tasks. *Informatica*, 49(24), 2025. https://doi:10.31449/inf.v49i24.8241 .

[36] Urbanowicz, Ryan J., Melissa Meeker, William La Cava, Randal S. Olson, and Jason H. Moore. Relief-based feature selection: Introduction and review. *Journal of biomedical informatics* ¸85 : 189-203, 2018 .https://doi : 10.1016/j.jbi.2018.07.014 .