Neural Processing Letters (2025) 57:41
https://doi.org/10.1007/s11063-025-11760-y

# Predicting Depression in Older Adults: A Novel Feature Selection and Neural Network Framework

**Ashir Javeed[1] · Peter Anderberg[2] · Ahmad Nauman Ghazi[2] · Muhammad Asim Saleem[3] · Johan Sanmartin Berglund[2]**

Accepted: 21 March 2025


**Abstract**
Depression in older adults is a significant public health issue with broad impacts on both individuals and society. The multifaceted nature of depression underscores the complexity of identifying and predicting risk factors, necessitating a sophisticated and accurate approach based on new emerging technologies. Compared to traditional statistical methods, machine learning provides a more detailed and individualized understanding of risk variables by analyzing large datasets, identifying patterns, and building predictive models. This study presented a novel feature selection method based on the relief and lasso algorithms. The proposed feature selection method selected the ten most significant features from the dataset. A neural network (NN) with hyperparameters optimized by a grid search technique was used to categorize depression. The feature selection and classification modules work together as a single unit, namely as (Relief_Lasso_NN). Data from the Swedish National Study on Aging and Care (SNAC) was used for this study. The collected dataset consists of 726 samples with 75 features per sample. Four experiments were conducted to validate the performance of the proposed (Relief_Lasso_NN) framework. The proposed model achieved an accuracy of 90.34% in predicting depression using only ten features from the dataset. The top 10 features identified by the proposed feature selection method significantly impact depression in older adults. Furthermore, the performance of seven other state-of-the-art machine learning models was also compared with the proposed framework.

✉ Ashir Javeed
ashir.javeed@bth.se

Peter Anderberg
peter.anderberg@bth.se

Ahmad Nauman Ghazi
nauman.ghazi@bth.se

Muhammad Asim Saleem
Muhammadasim.s@uestc.edu.cn

Johan Sanmartin Berglund
sanmartin.berglund@bth.se

[1] Department of Computer Science, Blekinge Institute of Technology, Karlskrona 37179, Sweden

[2] Department of Health, Blekinge Institute of Technology, Karlskrona 37179, Sweden

[3] Center of Excellence in Artificial Intelligence, Machine Learning and Smart Grid Technology, Department of Electrical Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok 10330, Thailand

Published online: 16 April 2025

**Keywords** Neural networks · Depression · Feature selection · Optimization · Lasso · Relief

## 1 Introduction

Depression in older adults has become a major public health issue with significant effects on both individual and societal well-being. With the growing senior population, addressing mental health issues is crucial. Depression reduces older adults' quality of life and places significant burdens on healthcare systems, caregivers, and communities [1].

The multifaceted nature of depression underscores the complexity of identifying and predicting risk factors, which requires a nuanced and comprehensive approach. Recent advances in health research have shown that incorporating machine learning techniques improves our understanding of the risk factors associated with depression in older [2]. Machine learning allows for analyzing large datasets, identifying trends, and creating predictive models that provide a more personalized and accurate understanding of depression risk factors than traditional statistical methods [3]. Growing concern about depression in older adults is rising. The World Health Organization (WHO) estimates that 7% of people over 65 globally suffer from mental or neurological disorders, with depression being one of the most common. As the prevalence varies from region to region, it is important to conduct research relevant to each area to provide targeted solutions [4]. In addition, depression in older people is often undiagnosed and untreated, highlighting the need for further research to improve our ability to identify and mitigate the risk factors associated with depression [5].

Identifying risk factors for depression in older people is not easy due to the interplay of biological, psychological and social influences [6]. Biological factors, like genetic predispositions and age-related neurological changes, interact with psychosocial factors such as social isolation, loss of loved ones, and chronic health conditions [7]. Untangling these interrelated variables requires sophisticated analytical methods that can cope with the ageing population's heterogeneity and complexity.

Machine learning offers various approaches to investigate the complex risk factors linked to depression in older adults [8]. While traditional approaches have advantages, they often fail to account for the dynamic and nonlinear interactions in datasets with multiple variables. Machine learning methods like decision trees, support vector machines, and neural networks can analyze large datasets, detect complex patterns, and build predictive models tailored to each case's specific characteristics.

In this research, machine learning techniques are used to predict depression and associated risk variables in the elderly population. We aim to develop predictive models that identify potential risk factors and clarify the complex relationships between these variables by analyzing a dataset containing demographic information, medical history, lifestyle factors, and social influences. For this reason, we have developed a hybrid feature selection technique based on the Lasso and Relief algorithms. Neural networks were used for the classification task. The hyperparameters of the neural network were optimized using a grid search approach. Depression is a result of a mental health disorder based on complex interactions between multiple risk factors such as psychological factors, genetics, and lifestyle, where the relief method also considers the interaction and relationships between features, which eventually help to predict depression with improved accuracy. On the other hand, the most important feature from the dataset is retained based on feature selection and regularization through Lasso. As a result, the hybrid approach (Relief_Lasso) effectively deals with problems of imbalanced data, captures feature interactions, and assures sparsity for interpretability. Therefore, the hybrid approach is a more reliable option for depression prediction than standalone filters, wrappers, and embedding methods. The aim is to facilitate the development of tailored interventions and individualized methods for the mental health of older people and thus improve the overall mental well-being of this population group.

This article focuses on the literature review, methodology, dataset, and results obtained by machine learning approaches to detect and predict risk factors for depression in the elderly. The findings of this study will improve the understanding of depression in older people and suggest strategies for early intervention and support.

## 2 Literature Review

Handling et al. presented a machine learning method based on random forest analysis (RF) for predicting risk factors for depression in European adults. They used 56 risk and protective indicators for depression in a large representative sample of older people in Europe. Their study found that among the 56 predictors, the two most important risk factors for depression in middle-aged and older people were self-rated poor health and self-perceived social isolation. There appears to be a gender difference in how difficulties with everyday living (for men) and increased family burden (for women) affect the risk of depression [9]. Hatton [10] used a machine learning method called extreme gradient boosting to analyze baseline demographic and psychometric data from 284 patients, predicting the likelihood of persistent depressive symptoms in older adults after one year. The effectiveness of their proposed prediction model was evaluated in comparison to conventional statistical method (logistic regression).

Montorsi et al. [11] utilized data from the Survey of Health, Aging and Retirement in Europe (SHARE) to evaluate the performance of six different machine learning algorithms. They examine the algorithms' performance with various configurations of life-course data. However, they obtained comparable predictive performance across algorithms and obtained the best predictive performance when sequence data was used to create semi-structured representations of life courses. They used the additive Shapley explanations approach to identify the most salient predictive patterns. The most important predictors of depression risk in later life are age, health status, childhood conditions, and low educational attainment. However, new predictive patterns have emerged concerning evidence of instability in the life course and inadequate dental care. Yang and Bath [12] developed models for predicting depression in older adults using five machine-learning techniques. Several model ensemble strategies were put forth to combine the output from distinct predictive models to enhance prediction performance further. In each domain area, notable protective or risk factors for depression symptoms in the elderly were identified separately. Their proposed approach attained the maximum area under the curve (AUC) of 89.20%. Zhou et al. [13] devised a machine-learning model via cross-sectional observational research to detect apathy, anxiety, and depression in older persons with minor cognitive impairment by analyzing their speech patterns and facial expressions. Using the random forest method, a multiclass emotion classification model was created. With a weighted-average F1 score of 96.6%, the model did well in classifying emotions. High recall, precision, and accuracy (87.4%, 86.6%, and 87.6%, respectively) were also demonstrated by the model. In their study, Zihan et al. [14] used information from the National Health and Nutrition Examination Survey (2005–2018). Five conventional machine learning methods and tenfold cross-validation led to the discovery of a dataset consisting of 2,546 veterans. The model's performance was evaluated using the following metrics: F1 score, specificity, accuracy, recall, area under the receiver characteristic curve (AUC), and precision. With an AUC of 0.891 (95% CI 0.869−0.914) and a specificity of 0.906, deep learning outperformed other methods in diagnosing depression in veterans. Subsequent depression studies in veterans of different ages show that the AUC values for deep learning are 0.929 (95% CI 0.904−0.955) for the middle-aged cohort and 0.924 (95% CI 0.900−0.948) for the older cohort.

Ah-Ram et al. [15] examined information on depression, quality of life, lifestyle, cognitive ability and demographic data. Cognitive function was assessed using the Korean Mini-Mental Status Examination, while depression was evaluated with the Center for Epidemiological Studies Depression Scale. No correlation was found between depression and smoking, but there was a correlation with age, gender, residential area, education level, alcohol consumption, regular exercise, life satisfaction, and cognitive function. In addition, a machine learning analysis yielded a prediction accuracy of 80.26% for depression, which was significantly influenced by life satisfaction, age, residential area, regular exercise, and cognitive level.

Lin et al. [16] presented a deep learning model for rapid binary depression classification in older adults. Demographic and acoustic data were collected from 56 Mandarin-speaking older adults diagnosed with major depressive disorder (MDD) using the Mini-International Neuropsychiatric Interview (MINI) and DSM-5 criteria, along with 47 control subjects. A deep learning model was developed and evaluated on an independent validation group to interpret audio data recorded with different smartphones. The effectiveness of the model is validated using the ROC curve. The caliber of the speech that was gathered impacted the model's accuracy. The model's initial 95% confidence intervals for sensitivity and specificity were 82.14% and 80.85%, respectively. Recently, Javeed et al. [17] investigated the efficacy of several feature extraction algorithms integrated with classification techniques, applying them to the Swedish National Study on Aging and Care (SNAC) dataset. Their work showed that principal component analysis (PCA) and logistic regression (LR) had better classification accuracy than the other methods studied.

Building on prior research that applied machine learning to predict depression in older adults, this study aims to enhance predictive accuracy by integrating novel feature selection methods with neural networks. By leveraging insights from past studies on key risk factors and data-driven approaches, our framework aspires to provide a more precise and interpretable model for identifying depression in aging populations.

## 3 Methods

### 3.1 Dataset

This study analyzed electronic health records (EHR) data from SNAC, which examines and collects information on the health problems of older adults and the risk factors associated with various diseases. SNAC was developed to study the healthcare and social needs of the aging population, compiling an electronic database of physical exams, psychological and social assessments, lifestyle factors, and medical conditions [18]. The SNAC database variables were chosen for their relevance to aging (health, social/emotional support, lifestyle factors, economic status, and personal assets). Seventy-five variables were extracted from the SNAC-Blekinge cohort (2000 to 2003), comprising seven main categories: demographic data, social factors, lifestyle, medical history, biochemical assessments, physical examinations, psychological assessments, and other health instruments. The description of selected variables along with their category is given in supplementary materials.

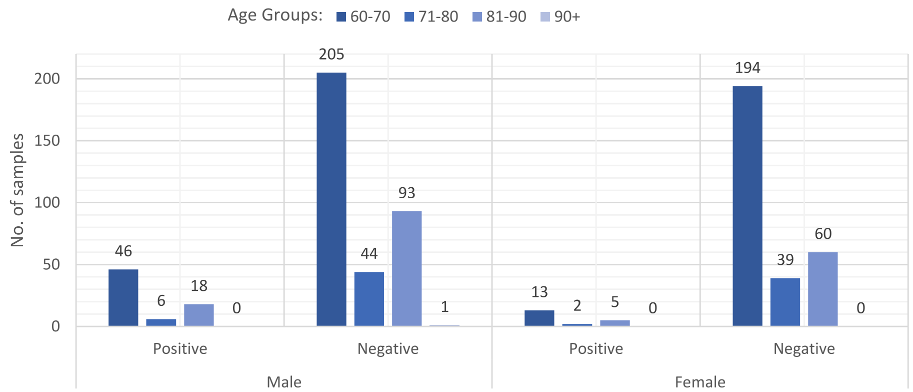



**Fig. 1** Data distribution in term of age group

In addition, 726 people were observed in the study (313 men and 413 women), of whom 90 (12.4%) were suffering from depression and 635 (87.5%) were in good health. The number of male and female cases of positive and negative depression and their age groups are given in Fig. 1.

### 3.2 Proposed Method

In order to find the best features for classification, feature selection approaches are crucial for machine learning algorithms [19]. This helps to mitigate the model's overfitting problem as well as reduce the execution period of the developed machine learning model. This fact served as our inspiration for developing a novel feature selection method based on the least absolute shrinkage and selection operator (Lasso) and Relief algorithms. An overview of the Relief and Lasso strategies' function is provided below.

#### 3.2.1 Relief Feature Selection Method

Relief is an algorithm for selecting features that assigns a weight to each feature in the dataset. After then, these weights can be progressively changed [20]. It is important to ensure that features with high weights for significance and low weights for insignificance are identified. For calculating feature weights, relief employs methods that are similar to those used in k-nearest neighbors (KNN). Kira and Rennell had demonstrated this widely recognized algorithm for feature selection [21]. $\mathbb{R}_i$ stands for the randomly selected sample from the dataset. Relief looks for the two neighbors closest to it: one from the opposing class, known as closest miss $\alpha$, and one from the same class, known as closest hit $\beta$. It improves the computation of consistency $\omega[F]$ for the feature $F$ with respect to the values of the $\mathbb{R}_i$, $\alpha$, and $\beta$. If there's a significant disparity between $\mathbb{R}_i$ and $\alpha$, then the performance of $\omega[F]$ will be significantly reduced. However, if there is a significant disparity between $\mathbb{R}_i$ and $\beta$ for the feature F, then F might be utilized to differentiate between various classes. Hence weight $\omega[F]$ is increased. This procedure will be repeated for $n$ times where $\beta$ is an adjustable parameter.

#### 3.2.2 Least Absolute Shrinkage and Selection Operator Method

The operator's ability to minimize selection and shrinkage depends on the change in the coefficient of the absolute values of the features. A subset of features can exclude features with negative coefficients or features with zero coefficient values. When it comes to feature values with small coefficients, Lasso performs exceptionally well. Large coefficient features will be present in the selected feature subsets. Lasso has some extraneous features [22]. The reliability of the given feature can also be improved by iteratively going through the aforementioned process and ultimately selecting the features that are considered the most important. One of Lasso's limitations is that it requires a powerful machine to run because of its parallel programming. This also shows the way it is used in the current application, where the vector of the associated $i^{th}$ sub-region keys is denoted by $\gamma^i$.

The proposed feature selection technique is based on the Relief and Lasso methods that give us the top significant features from the dataset. Each feature selection method selected the most appropriate features from the feature space (dataset). After selecting the feature, we take the intersection among the selected features; this process gives us common features. The common features are tested for the best accuracy. In order to identify the best features from the dataset, this process is repeated until a predetermined set of features is examined. The Algorithm 1 outlines the proposed feature selection method's operation as follows:

**Algorithm 1** Proposed feature selection method (Relief_Lasso)

**Input:** [$F_S$: Feature space (dataset), $M_F$: No. of Max. features selected]
**Output:** [$O_F$: Optimal set of features]

```
1: Best_Acc = 0
2: M_F = 15
3: while M_F ≠ 1 do
4:     V_Relief ⇐ Relief(M_F, F_S)
5:     V_Lasso ⇐ Lasso(M_F, F_S)
6:     R_F ⇐ F_Lasso ∩ F_Relief
7:     Test_Acc = LR(R_F)
8:     if Test_Acc > Best_Acc then
9:         O_F = R_F
10:        Best_Acc = Test_Acc
11:    end if
12:    M_F ⇐ M_F − 1
13: end while
14: return O_F
```

Notably, neural networks (NNs) are also capable of extracting features on their own. For example, the model of [23] uses feature extraction instead of feature selection to derive features or rules from the data. For feature selection, we concentrate exclusively on the most important features from the dataset and eliminate all superfluous ones. During the feature extraction process, each and every feature is considered as a newly constructed feature.

Neural networks use many nonlinear components, or neurons, to capture highly complex relationships or functions. Likely, irrelevant features are suitably represented in the feature space as well. Noise is produced when irrelevant features are considered [24]. As a result, learning the noise from insignificant features has a negative impact on understanding the distribution of the data overall. If there are non-meaningful features in the feature space, there is also the risk of the network becoming overfitted on the training set [25, 26]. This is when the network receives unnecessary details from the training data. It does well on

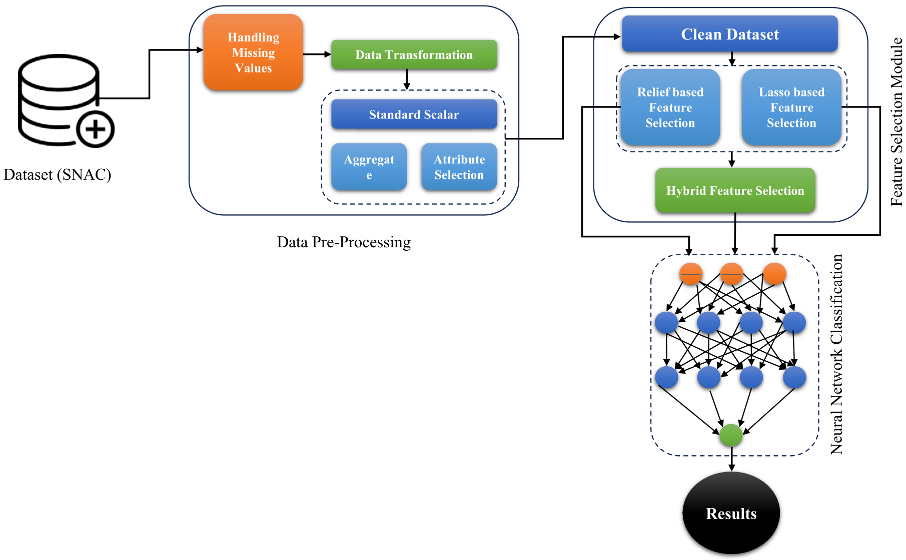



**Fig. 2** Working of proposed framework

the training set as it becomes increasingly biased toward the previously observed data [27]. Hence, it cannot be applied to the unknown testing or validation data.

We employ a hybrid feature selection approach, relying on the Relief and Lasso algorithms to eliminate unnecessary features from the feature space before applying the feature vector to NNs in order to address the issues caused by irrelevant features. The hybrid feature selection method (Relief_Lasso) removes irrelevant features from the dataset. Figure 2 presents the workings of the proposed framework. The neural network can use the optimal subset of features, but inadequate architecture leads to underperformance. Insufficient adaptation occurs when the neural network architecture does not have sufficient capacity for the classification task, which is the main cause of suboptimal performance [28]. In such cases, the neural network shows suboptimal performance on both the training and test datasets. Conversely, an architecture of the neural network that is too large leads to overfitting of the training set, which improves performance on the training data while degrading performance on the test data. We must go deeper to find the optimal neural network architecture that performs well on the training and test data. To understand the relationship between the architecture of a neural network and its performance, it is important to understand the formulation of neural networks first. The specification of NN is presented as follows:

The computational system constructs neural networks using mathematical models that mimic the structure of the human brain. A node, also known as a perception, is the central component of the neural network model. Layers refer to the formation of nodes into groups. Like biological neurons, artificial neurons operate on the same principle. An artificial neuron processes the information as it comes in from nearby neurons, sending the output to the following perceptron. It does this by receiving one or more inputs. The link between artificial neurons is referred to as a weight. In the output computation, the input data, $\nu_i$ is assigned either a positive or negative weight. For the problem under consideration, weights and an internal threshold value $\psi$ are assigned. The related weight $\varpi_n$ which is adjusted by the threshold value $\psi$ and the input values $\nu_n$ are multiplied to determine the result on each node.

After that, the output is calculated by using a transfer function ($\mu$) or activation function, as is given in the equation below:

$$\mho i = \mu \left( \sum \varpi_n . \nu_i - \psi \right) \tag{1}$$

Both linear and nonlinear transfers are possible. The hyperbolic tangent function is commonly used as a nonlinear activation function. Equation 2 represents the process of calculating the sigmoid function $\mu(\mho i)$, at the subsequent layer. $\theta$ and the sigmoid function's shape are connected. The sigmoid function became more nonlinear as the parameter $\theta$ value increased.

$$\mu(\mho i) = \frac{1}{[1 + exp(-\theta \mho i)]} \tag{2}$$

By connecting artificial neurons, an optimized neural network is formed [29].

## 4 Validation and Evaluation Metrics

Various validation techniques are used in machine learning (ML) to assess the efficacy of developed models. However, holdout validation and cross-validation are the common approaches employed by the researchers to evaluate the effectiveness of predictive models [30, 31]. The holdout validation technique halts the dataset to train and test the ML model. Mezzadri et al. [32] argue that a limitation of holdout validation is its vulnerability to the arbitrary selection of training and test datasets, which can impact the evaluation of an ML model's performance. In such cases, cross-validation is beneficial. K-fold cross-validation, the most common method, splits the dataset into k equal parts. The model is trained k times, using one fold for testing and the others for training. The evaluation score is calculated as the average of the performance metrics after all iterations. Seraj et al. [33] claim that cross-validation improves the estimation of a model's generalization performance by reducing the dependence of performance evaluation on the specific selection of training and test sets. Therefore, to test the effectiveness of the proposed methodology, we used the cross-validation technique.

This study uses accuracy, sensitivity, specificity, and the Matthews correlation coefficient (MCC) as evaluation metrics to assess the effectiveness of the proposed model. Accuracy is the proportion of observations that were correctly classified (both depressed and non-depressed) out of the total observations. Sensitivity evaluates how well the ML model identifies observations with depression. A high sensitivity ensures that most actual depression observations are detected, which is crucial in medical applications to minimize false negatives. Sensitivity measures how effectively the model identifies observations without depression. High specificity reduces the risk of false positives, ensuring that observations are not incorrectly classified as depressed. Matthews Correlation Coefficient (MCC) presents a balanced evaluation by considering all four components of the confusion matrix (true positives, false positives, true negatives, and false negatives). MCC is particularly valuable for imbalanced datasets, where one class (e.g., non-depressed individuals) may be more prevalent, such as in the given study, where the observation with true depression cases is much smaller than false depression cases. Moreover, MCC is a robust statistical measure used to evaluate the quality of binary classifications that ranges from -1 (total disagreement) to +1 (perfect prediction), making it a robust metric for assessing model performance in depression prediction [34]. The aforementioned evaluation criteria are formulated quantitatively as follows:

$$Sensitivity = \frac{TP}{TP + FN} \tag{3}$$

$$Specificity = \frac{TN}{TN + FP} \tag{4}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

Where *TP* denotes the number of true positives, *FP* represents false positives, *TN* true negatives, and *FN* false negatives.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{6}$$

## 5 Results and Discussion

Four distinct experiments were conducted using the SNAC dataset to evaluate the efficacy of the proposed technique thoroughly. In the first experiment, we assessed how well the neural network and features selected via the Lasso approach perform in terms of depression prediction. A grid search technique is employed to determine the best hyperparameters of a neural network. In the second experiment, we assessed the performance of a neural network and features selected through the Relief algorithm. In the third experiment, a hybrid feature selection method based on the Lasso and Relief technique was tested using a neural network to predict depression accurately. The performance of baseline machine-learning models was assessed using the same SNAC dataset in the fourth experiment.

The entire computational task was done using a 64-bit version of Windows 10 running on an Intel (R) Core (TM) i5-8250U CPU running at 1.60 GHz and 1800 MHz. A software program called Python was used to carry out each experiment.

### 5.1 Experiment 01: Relief Based Feature Selection

In this experiment, we developed only a relief-based feature selection algorithm along with a neural network for the prediction of depression. The best parameters of the Relief algorithm through which the top ten features from the dataset were selected are given as follows: The number of iterations of the Relief algorithm was set to the total sample size of the dataset (N = 726), the number of neighbors was set to ten (K = 10), and the distance metric was set as Euclidean. Figure 3 presents the ranking of all the features in the SNAC dataset.

After evaluating the features, we applied a threshold value to select the ten most optimal features from the dataset. Figure 4 illustrates the association between the top 10 features identified by the relief method. The selected top ten features were then entered into the neural network for depression prediction. A grid search algorithm determined the hyperparameters of the neural network. Table 1 refers to the effectiveness of the Relief and neural network-based solution. The performance of the Relief_NN model was assessed using various metrics, including accuracy (Acc.), sensitivity (Sens.), specificity (Spec.), and Matthew's correlation coefficient (MCC). Table 1 presents the performance of the constructed (Relief_NN) system, including the activation function ($A_F$), the optimizer, and the number of neurons (N) in the neural network, all optimized by a grid search technique. The Relief_NN system achieved a maximum accuracy of 88.97% with the ReLU activation function, the Adam optimizer, and 64 neurons.

The receiver operating characteristic (ROC) curve was used to assess the model's effectiveness. A graph with a larger area under the curve (AUC) is considered more accurate [35].

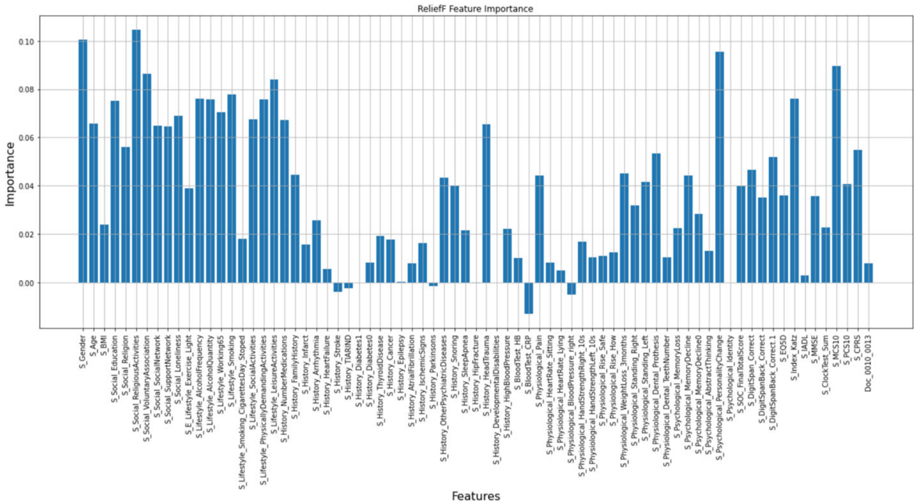



**Fig. 3** Feature ranking based on relief algorithm

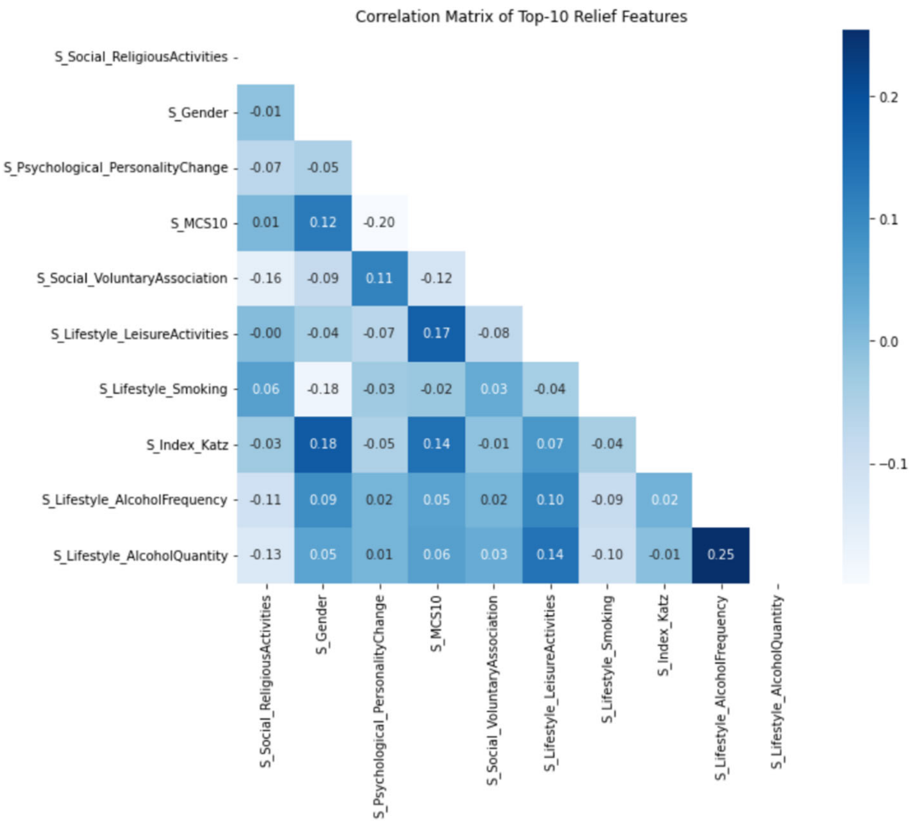



**Fig. 4** Correlation among top 10 selected features by the relief method

**Table 1** Performance evaluation based on Relief_NN

| $A_F$ | Optimizer | $N$ | Acc | Sens | Spec | MCC |
|---|---|---|---|---|---|---|
| softmax | rmsprop | 8 | 85.63 | 88.00 | 92.85 | 0.7335 |
| tanh | rmsprop | 64 | 88.77 | 75.26 | 98.50 | 0.7856 |
| relu | adam | 16 | 88.58 | 80.00 | 95.45 | 0.7700 |
| sigmoid | adam | 32 | 88.38 | 85.86 | 93.89 | 0.7622 |
| relu | adam | 64 | 88.97 | 90.50 | 98.00 | 0.7875 |
| softmax | adam | 8 | 87.99 | 88.95 | 96.72 | 0.7695 |
| sigmoid | rmsprop | 32 | 88.18 | 84.87 | 90.00 | 0.7685 |
| softmax | rmsprop | 16 | 87.00 | 85.26 | 95.45 | 0.7595 |

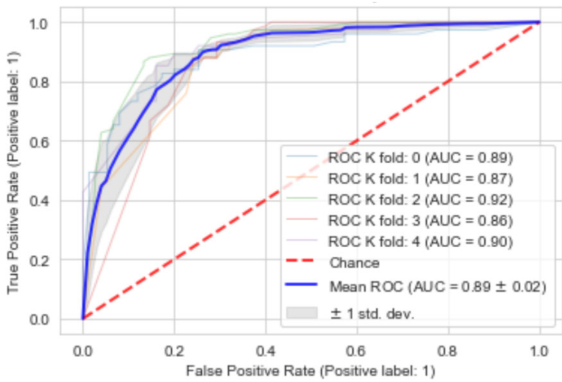



**Fig. 5** Performance analysis of Relief_NN model based on ROC

We applied a five-fold (k = 5) cross-validation to assess the effectiveness of the developed model (Relief_NN). The performance of the Relief_NN model was evaluated using the AUC metric. Figure 5 presented that the developed model (Relief_NN) achieved an average of 89.00% AUC.

### 5.2 Experiment 02: Lasso Based Feature Selection

In this experiment, we construct a neural network and apply a lasso-based feature selection method to predict the occurrence of depression. The Lasso approach ranks all features in the SNAC dataset based on relevance, as illustrated in Fig. 6. After ranking the features, a threshold was applied to select the top 10 features using the lasso algorithm. The best parameters of the Lasso algorithm through which the top ten features from the dataset were selected are given as follows: The number of maximum iterations of the Lasso was set to 'max_iter' = 5000. Alpha controls the strength of the L1 penalty of Lasso, which was set to 'alpha' = 0.01. Selection defines how the features are selected during coefficient updates in Lasso, which was set to random. Tolerance sets the stop criteria for optimization of Lasso set to 'tol' = 0.01.

The correlation among the top ten features determined by the Lasso algorithm can be depicted from Fig. 7. To predict depression in older people, the neural network utilized the top ten features determined by the Lasso algorithm. With a grid search algorithm, the neural network's hyperparameters were optimized. The Lasso and neural network-based systems (Lasso_NN) performance is given in Table 2. Several evaluation metrics, including

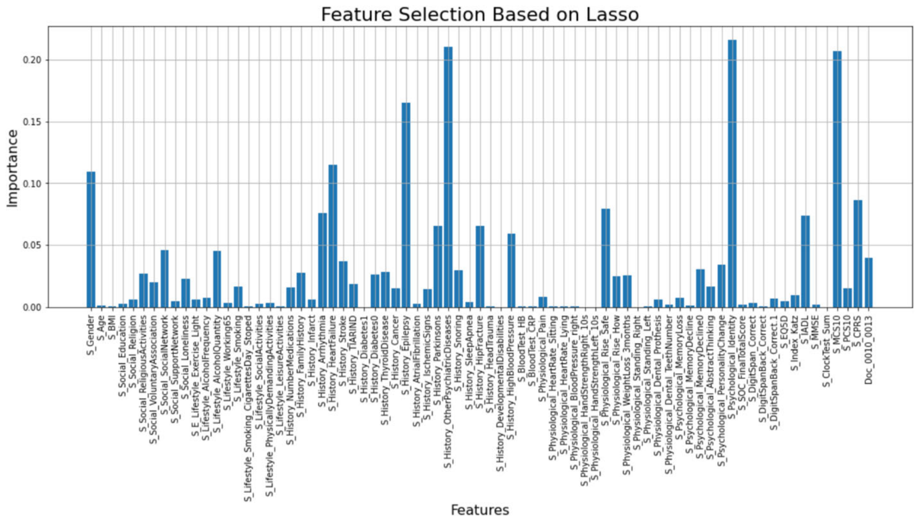



**Fig. 6** Feature ranking based on Lasso algorithm

**Table 2** Performance evaluation based on Lasso_NN

| $A_F$ | Optimizer | $N$ | Acc | Sens | Spec | MCC |
|---|---|---|---|---|---|---|
| softmax | adam | 8 | 85.24 | 80.45 | 90.65 | 0.7300 |
| sigmoid | rmsprop | 16 | 86.14 | 85.00 | 88.50 | 0.7395 |
| relu | adam | 16 | 86.85 | 91.25 | 90.75 | 0.7400 |
| softmax | rmsprop | 32 | 88.58 | 89.65 | 94.00 | 0.7700 |
| tanh | adam | 64 | 87.38 | 91.20 | 89.88 | 0.7675 |
| softmax | rmsprop | 16 | 88.38 | 95.00 | 90.65 | 0.7689 |
| relu | adam | 32 | 88.00 | 89.00 | 96.00 | 0.7600 |
| sigmoid | adam | 64 | 88.18 | 95.15 | 91.48 | 0.7615 |
| tanh | rmsprop | 64 | 87.79 | 90.88 | 93.00 | 0.7585 |

Matthews' correlation coefficient (MCC), sensitivity (Sens.), specificity (Spec.), and accuracy (Acc.), were used to evaluate the performance of the constructed model (Lasso_NN). The performance of the constructed system (Lasso_NN) as well as the activation function ($A_F$), optimizer, and number of neurons (N) of the neural network that was optimized using a grid search algorithm are given in Table 2. The Lasso_NN system with 32 neurons, the Adam optimizer, and the ReLU activation function achieves a maximum accuracy of 88.58%.

Figure 8 shows the performance evaluation of the Lasso_NN model using the AUC metric, depicted through the ROC curve. From Fig. 8, it is evident that the model that was constructed, the Lasso_NN, has the greatest AUC, at 88.00%).

## 5.3 Performance of Proposed Hybrid Feature Selection Method

This subsection details the design and development of the proposed hybrid feature selection approach, combining the lasso and relief methods. The dataset was first handed to a recently constructed feature selection algorithm that identified the most significant features from the

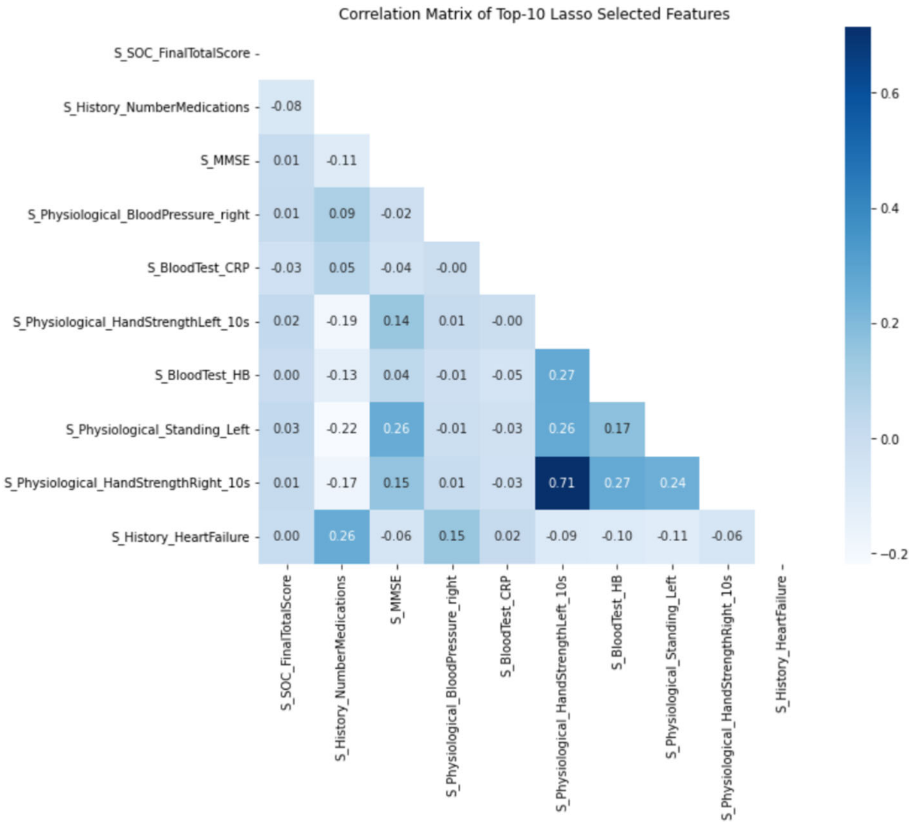
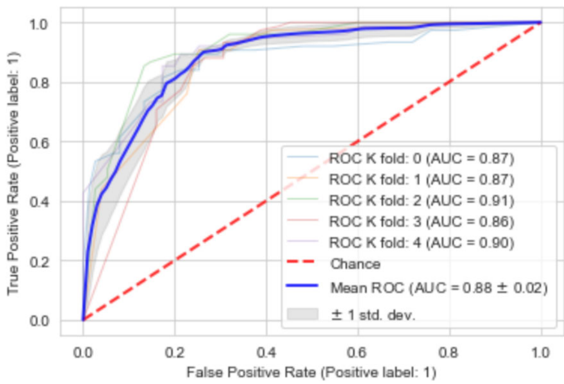



Fig. 7 Correlation among top 10 selected features by the Lasso method

Fig. 8 Performance analysis of Lasso_NN model based on ROC




dataset. Figure 9 shows the features that were scored highest in terms of feature importance. Once the features are ranked, we set a threshold to choose the top ten features.

Figure 10 presents the correlation among the top ten features, founded by the proposed hybrid feature selection method. The neural network employed the top 10 features, ascertained by the hybrid feature selection (Relief_Lasso) algorithm, to predict depression in the elderly

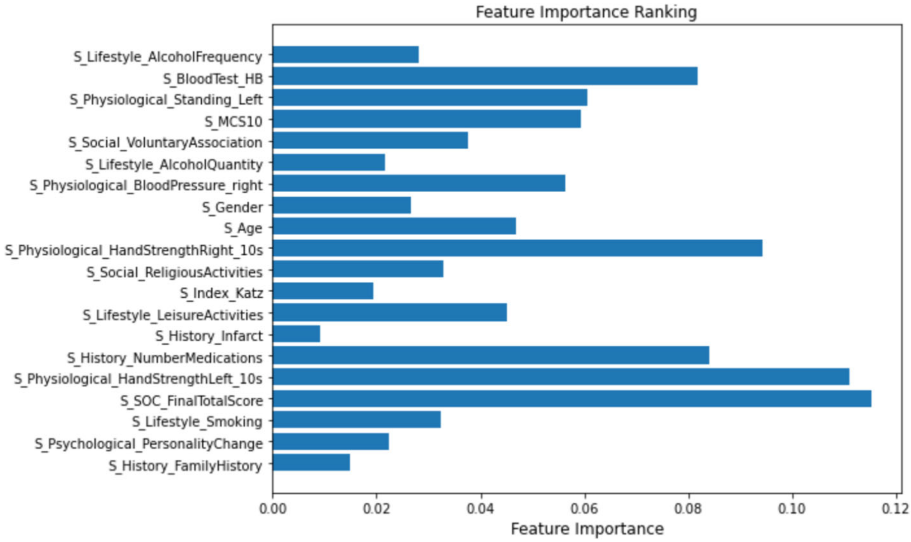



**Fig. 9** Feature ranking based on proposed feature selection method (Lasso_Relief)

**Table 3** Performance of proposed (Relief_Lasso_NN) model

| $A_F$ | Optimizer | $N$ | Acc | Sens | Spec | MCC |
|---|---|---|---|---|---|---|
| relu | rmsprop | 32 | 88.97 | 90.38 | 96.00 | 0.7685 |
| relu | adam | 64 | 89.15 | 89.00 | 98.00 | 0.7700 |
| sigmoid | rmsprop | 16 | 88.58 | 91.75 | 93.80 | 0.7650 |
| sigmoid | adam | 64 | 88.18 | 88.50 | 94.00 | 0.7600 |
| tanh | rmsprop | 32 | 89.36 | 94.14 | 97.50 | 0.7738 |
| tanh | adam | 64 | 89.76 | 95.00 | 98.00 | 0.7800 |
| sigmoid | adam | 32 | 90.34 | 94.20 | 99.00 | 0.7825 |
| softmax | adam | 32 | 89.80 | 93.84 | 96.50 | 0.7810 |
| sigmoid | rmsprop | 32 | 90.14 | 95.00 | 98.50 | 0.7815 |

population. The neural network's hyperparameters were optimized using a grid search algorithm. Table 3 presents the performance of the proposed model (Relief_Lasso_NN) based on various metrics such as Matthews' correlation coefficient (MCC), sensitivity (Sens.), specificity (Spec.), and accuracy (Acc.). Results of the proposed model (Relief_Lasso_NN) are given in Table 3, which consists of the number of neurons (N), optimizer, and activation function ($A_F$). With 32 neurons, the Adam optimizer, and the sigmoid activation function, the proposed system (Relief_Lasso_NN) obtained the highest level of accuracy of 90.34%.

The performance evaluation of the (Relief_Lasso_NN) model based on AUC is shown in Fig. 11 by using the ROC curve. It is clear from Fig. 11 that the proposed model (Relief_Lasso_NN model) has the highest AUC of 90.00%.

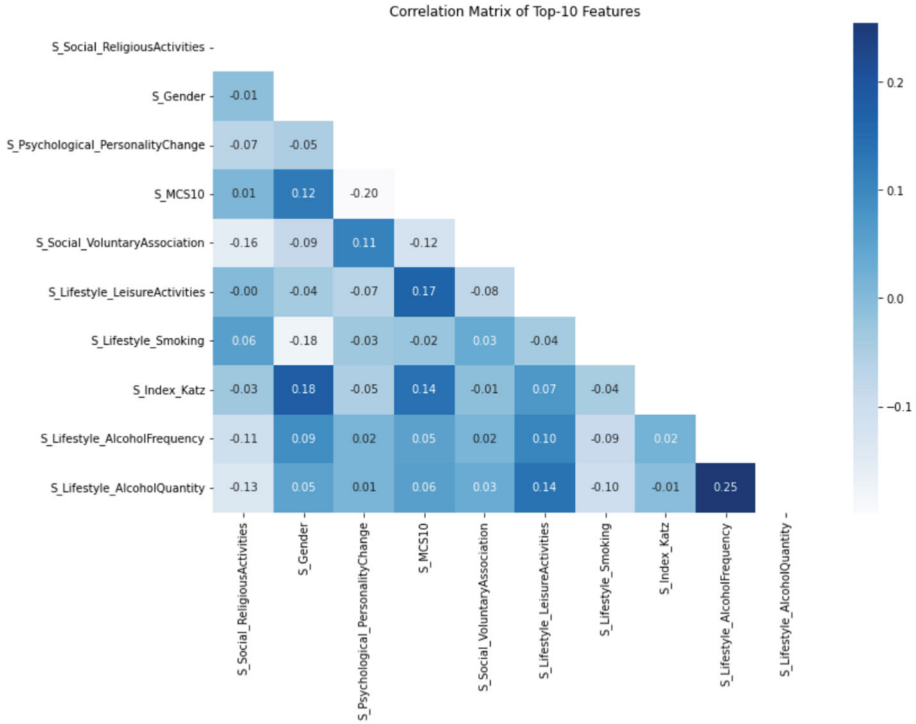
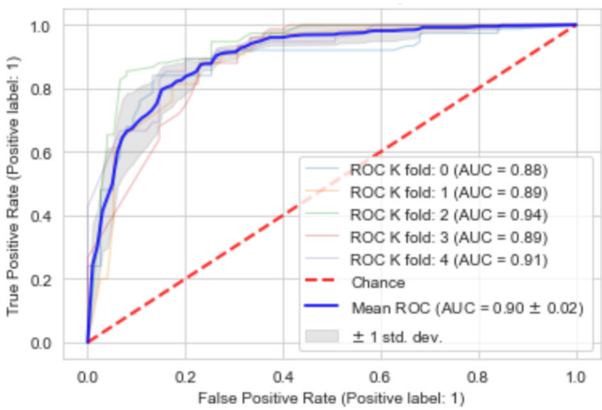


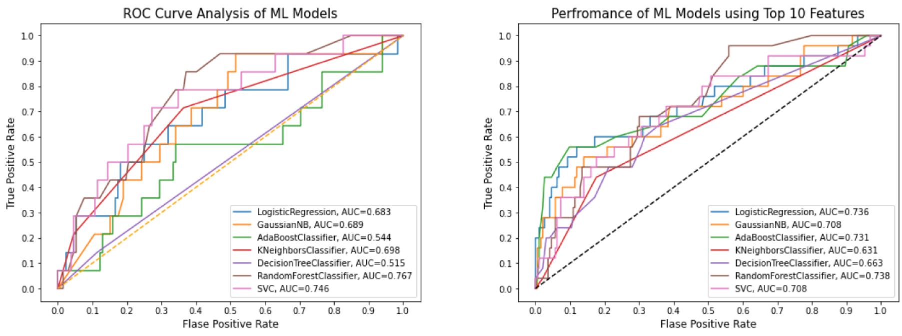

**Fig. 10** Correlation among top 10 selected features by the hybrid Lasso_Relief method

**Fig. 11** Performance analysis of proposed model (Relief_Lasso_NN) based on ROC




## 5.4 Performance of Baseline ML Models

In this experiment, the effectiveness of seven baseline ML models was evaluated, such as logistic regression (LR), Gaussian naive Bayes (GNB), AdaBoost, K-nearest neighbors (KNN), decision tree (DT), random forest (RF), and support vector classification (SVC). We employed a holdout validation scheme to test the performance of the above-mentioned ML models. The dataset was divided into a ratio of 30% to 70% for the testing and training purposes of the ML models. We employed a receiver operating characteristic (ROC) curve,




(a) ROC analysis based on all features (b) ROC analysis based on top ten features

**Fig. 12** AUC-based performance comparison

focusing on the area under the curve (AUC) as the evaluation metric. A larger AUC signifies a more accurate ROC curve. The performance of the advanced machine learning models was assessed using all dataset features and the top 10 features selected through the proposed hybrid feature selection strategy. The ML model results are presented in Fig. 12. The Fig. 12a illustrates the performance of ML models utilizing the complete set of features from the dataset, whereas Fig. 12b illustrates the performance of ML models based on the top ten features selected by the proposed hybrid feature selection method (Relief_Lasso).

## 6 Discussion

This study demonstrates the use of machine learning to predict depression and associated risk factors in the elderly population. The dataset is sourced from the SNAC. The compiled dataset has seventy-five features and a total of 726 observations. The proposed machine learning model comprises two modules: the first selects the most significant features from the dataset, and the second employs a neural network to classify individuals as healthy or depressed. The research involved two feature selection algorithms: Relief and Lasso. In addition, we developed a hybrid feature selection method using the Relief and Lasso algorithms. After identifying the top 10 features from the dataset by three feature selection methods, a neural network was able to predict sadness successfully. The hyperparameters of the neural network are crucial for optimizing performance. Therefore, we used a grid search strategy to improve the neural network's architecture.

Furthermore, we have conducted several experiments to evaluate the performance of the ML models and the newly developed (Relief_Lasso_NN) model based on various evaluation metrics. We selected accuracy, sensitivity, specificity, MCC, and area under the curve (AUC) through the receiver operating characteristic curve (ROC). All of these metrics were evaluated on the built ML models using a cross-validation scheme. The cross-validation scheme has several advantages over the holdout scheme. Thus, we selected the cross-validation scheme to rigorously validate the efficiency of the constructed ML models. Furthermore, we conducted four different types of experiments in this study. In our first experiment, we utilized the Relief algorithm to select the most significant features from the dataset. After feature selection, the optimized neural network was deployed for the prediction of depression. Table 1 provides a performance overview of the developed model (Relief_NN) with different values of hyperparameters. In this experiment, the highest accuracy of 88.97% was obtained by the developed

**Table 4** Performance analysis and limitations of ML models

| Models | AUC | Performance Analysis | Limitations |
|---|---|---|---|
| LR | 0.736 | Moderate; works well with linear features | Struggles with class imbalance and non-linearity |
| GNB | 0.708 | Simple and fast | Assumes feature independence, often unrealistic |
| k-NN | 0.631 | Works well with small datasets | Sensitive to imbalanced data, poor with high dimensions |
| DT | 0.663 | Captures non-linearity | Easily overfits, biased towards majority class |
| RF | 0.738 | Handles feature interactions well | Computationally expensive, may still favor majority class |
| SVC | 0.708 | Effective with kernel tricks | Sensitive to imbalanced data unless properly tuned |
| AdaBoost | 0.731 | Improves weak models | Sensitive to noisy data and class imbalance |
| Relief_Lasso_NN | 0.900 | High accuracy, robust | Computationally intensive |

model (Relief_NN). In the second experiment, the top ten features from the dataset were determined through the Lasso algorithm and the neural network employed for the classification. Table 2 provides a performance summary of the constructed model (Lasso_NN) with different hyperparameter values. The highest accuracy of 88.18% was achieved by the developed model (Lasso_NN). Our third experiment proposed a hybrid feature selection method combining the Relief and Lasso algorithms. This approach identified the 10 most important features, which were input into an enhanced neural network for classification. The proposed model (Relief_Lasso_NN) achieved the maximum accuracy of 90.14% as shown in Table 3.

In the fourth experiment, we employed baseline ML models such as logistic regression (LR), Gaussian naive Bayes (GNB), AdaBoost, K-nearest neighbors (KNN), decision tree (DT), random forest (RF), and support vector classification (SVC). The efficiency of the machine learning models was evaluated using the Area Under the Curve (AUC) metric. Figure 12 illustrates the efficiency of the machine learning models using the full collection of dataset features and the top 10 features identified by the proposed hybrid feature selection method (Relief_Lasso). In terms of performance analysis and limitation of each baseline model, the proposed (Relief_Lasso_NN) was also assessed. The breakdown performance analysis and limitation of each model are given as follows in Table 4 which consists of AUC, performance analysis, and limitation of each ML model using the top 10 features selected by the newly developed hybrid feature selection method (Relief_Lasso).

It is also important to discuss the top ten features selected by the proposed hybrid feature selection (Relief_Lasso) method for depression prediction. Table 5 provides an overview of the features that are selected by the proposed hybrid feature selection (Relief_Lasso) method for accurate prediction of depression. The majority of the variables that the proposed algorithm predicted are associated with the variables' lifestyle category. Therefore, we can prevent depression by having a better lifestyle.

From the literature, it is evident that the top ten identified features for depression prediction were also studied by several researchers, such as E. Bergmann et al. employed linear models to examine the relationship between depressive symptoms and cognitive performance. Their study reported that older individuals had more depressive symptoms [36]. The factor of

**Table 5** Top ten selected features

| S.No | Features |
|---|---|
| 01 | Age |
| 02 | Social religious activities |
| 03 | Social voluntary association |
| 04 | Lifestyle alcohol frequency |
| 05 | Lifestyle alcohol quantity |
| 06 | Lifestyle smoking |
| 07 | Lifestyle leisure activities |
| 08 | Psychological personality change |
| 09 | Index_Katz[1] |
| 10 | MCS[2] |

Index_Katz: assessing functional ability in people with disabilities - MCS: Mental component score

social religious activities associated with the depression had been widely explored by the researchers. Few studies suggest that involvement in social religious activities at a later stage of life helps to reduce depression [37–39]. Another factor that was identified by the proposed (Relief_Lasso_NN) model was the social volunteering association of older adults. In the literature, many studies mentioned the factor of social volunteering association of older adults that significantly contributes to reducing depression [40–42]. Lifestyle alcohol frequency and lifestyle alcohol quantity are the two more features that were identified by the proposed model that cause depression in older adults. These two factors were also highlighted by the other researchers for the development of depression in older adults [43, 44]. Smoking was also studied by the researchers, who acted as a major factor of depression in older adults, such as Wu et al. [45] conducted a cross-sectional study on US adults to identify the relationship between smoking and depression. Their study suggests that smoking increases the risk of depression. Y. Cui et al. also study the association between leisure activity and depression in Chinese older adults [46]. Their finding suggested that leisure activity in older adults helps to eliminate the risk of depression in older adults. Many researchers in the literature also discuss the association of psychological personality change with depression [47–49]. The last two factors that were reported by the proposed model were Index_Katz (assessing functional ability in people with disabilities) and mental component score (MCS). The association of these two factors with the depression is also studied by the researcher in the past. These factors align with existing research, further validating their significance in understanding and predicting depression.

## 7 Conclusion

This study presented a machine learning-based approach for predicting depression and its risk factors in older adults. The proposed machine learning-based framework consists of two modules, such as feature selection and classification. For feature selection, a hybrid method based on relief and lasso algorithms was developed, while for classification, a neural network was deployed. Through the integration of novel feature selection techniques (Relief_Lasso) and neural network classification, the study introduces Relief_Lasso_NN as an effective system for depression prediction. The neural network was optimized by employing a grid search algorithm to find the optimal hyperparameters of the neural network. The study shows that the proposed approach is effective in predicting depression based on data from the Swedish National Study on Aging and Care (SNAC). It achieves a noteworthy accuracy of 90.34% while using only ten features that were determined by the hybrid feature selection method (Relief_Lasso). The features that the system found provide vital details on the factors affecting depression. Additionally, the thorough assessment of the proposed (Relief_Lasso_NN) system in comparison to other cutting-edge machine learning models highlights its superiority in depression prediction. The results demonstrate that cutting-edge technology can improve our knowledge and ability to treat depression in older adults.

For future research directions in the field of depression prediction, several potential extensions can be explored, such as including a dataset consisting of diverse geographical regions and cultural backgrounds, which would help to validate the proposed model's generalizability. Furthermore, experimenting with deep neural networks or integrating attention mechanisms could improve predictive performance. Lastly, incorporating explainable artificial intelligence such as SHAP or LIME would enhance the interpretability of the proposed (Relief_Lasso_NN) model for clinical applications. These potential extensions could contribute to advancing the field of machine learning-based mental health prediction and improving real-world applicability.

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s11063-025-11760-y.

**Author Contributions** Conceptualization and Data Curation by Peter Anderber, Formal Analysis by Ahmad Nauman Ghazi, Writeup and Methodology by Ashir Javeed, Evaluation and Validation by Muhammad Asim Saleem, Supervised by Johan Sanmartin Berglund.

**Funding** Open access funding provided by Blekinge Institute of Technology. This research received no external funding.

**Availability of data and materials** Data used in this study can be requested from the SNAC-Blekinge center.

## Declarations

**Conflict of interest** The authors declare no conflict of interest.



## References

1. Pien L-C, Cheng W-J, Chang W-P, Chen S-R, Chou K-R, Wang C-H (2023) Relationships between stroke prevalence, health literacy, and oral health-related quality of life in middle-aged and older adults: a national survey study. BMC Geriatr 23(1):1–8
2. Gomes S, Schantz M, Leocadio-Miguel M (2023) Predicting depressive symptoms in middle-aged and elderly adults using sleep data and clinical health markers: a machine learning approach. Sleep Med 102:123–131

3. Chen ZS, Galatzer-Levy IR, Bigio B, Nasca C, Zhang Y, et al (2022) Modern views of machine learning for precision psychiatry. Patterns 3(11)
4. Wasay M, Younis S, Charway-Felli A, Basri H (2023) World health organization (who)'s intersectoral global action plan (igap); implications for asian and african countries. J Neurol Sci 444
5. Tutun S, Johnson ME, Ahmed A, Albizri A, Irgil S, Yesilkaya I, Ucar EN, Sengun T, Harfouche A (2023) An ai-based decision support system for predicting mental health disorders. Inf Syst Front 25(3):1261–1276
6. Maier A, Riedel-Heller S.G, Pabst A, Luppa M (2021) Risk factors and protective factors of depression in older people 65+. A systematic review. PloS One 16(5), 0251326
7. Tiemeier H (2003) Biological risk factors for late life depression. Eur J Epidemiol 18:745–750
8. Murri MB, Cattelani L, Chesani F, Palumbo P, Triolo F, Alexopoulos GS (2022) Risk prediction models for depression in community-dwelling older adults. Am J Geriatr Psychiatry 30(9):949–960
9. Handing EP, Strobl C, Jiao Y, Feliciano L, Aichele S (2022) Predictors of depression among middle-aged and older men and women in Europe: a machine learning approach. Lancet Reg Health Europe 18
10. Hatton CM, Paton LW, McMillan D, Cussens J, Gilbody S, Tiffin PA (2019) Predicting persistent depressive symptoms in older adults: a machine learning approach to personalised mental healthcare. J Affect Disord 246:857–860
11. Montorsi C, Fusco A, Van Kerm P, Bordas SP (2024) Predicting depression in old age: combining life course data with machine learning. Econ Hum Biol 52:101331
12. Yang H, Bath PA (2019) Automatic prediction of depression in older age. In: Proceedings of the 3rd international conference on medical and health informatics, pp 36–44
13. Zhou Y, Han W, Yao X, Xue J, Li Z, Li Y (2023) Developing a machine learning model for detecting depression, anxiety, and apathy in older adults with mild cognitive impairment using speech and facial expressions: A cross-sectional observational study. Int J Nurs Stud 146:104562
14. Qu Z, Wang Y, Guo D, He G, Sui C, Duan Y, Zhang X, Lan L, Meng H, Wang Y et al (2023) Identifying depression in the united states veterans using deep learning algorithms, nhanes 2005–2018. BMC Psychiatry 23(1):620
15. Kim A-R, Park J-H, Park HY (2021) Analysis of factors affecting depression in older adults in south Korea. Int J Environ Res Public Health 18(18):9887
16. Lin Y, Liyanage BN, Sun Y, Lu T, Zhu Z, Liao Y, Wang Q, Shi C, Yue W (2022) A deep learning-based model for detecting depression in senior population. Front Psych 13:1016676
17. Javeed A, Anderberg P, Ghazi AN, Javeed A, Dallora AL, Berglund JS (2024) Optimizing depression prediction in older adults: a comparative study of feature extraction and machine learning models. In: 2024 international conference on control, automation and diagnosis (ICCAD), pp 1–6. IEEE
18. Lagergren M, Fratiglioni L, Hallberg I.R, Berglund J, Elmståhl S, Hagberg B, Holst G, Rennemark M, Sjolund B-M, Thorslund M, et al (2004) A longitudinal study integrating population, care and social services data. The Swedish national study on aging and care (snac). Aging Clin Exp Res 16(2), 158–168
19. Ali L, Zhu C, Golilarz NA, Javeed A, Zhou M, Liu Y (2019) Reliable parkinson's disease detection by analyzing handwritten drawings: construction of an unbiased cascaded learning system based on feature selection and adaptive boosting model. Ieee Access 7:116480–116489
20. Urbanowicz RJ, Meeker M, La Cava W, Olson RS, Moore JH (2018) Relief-based feature selection: introduction and review. J Biomed Inform 85:189–203
21. Chikhi S, Benhammada S (2009) Reliefmss: a variation on a feature ranking relieff algorithm. Int J Bus Intell Data Min 4(3–4):375–390
22. Tibshirani R (1996) Regression shrinkage and selection via the lasso. J R Stat Soc Ser B Stat Methodol 58(1):267–288
23. Taherkhani A, Cosma G, McGinnity TM (2018) Deep-fs: a feature selection algorithm for deep Boltzmann machines. Neurocomputing 322:22–37
24. Ali L, Javeed A, Noor A, Rauf HT, Kadry S, Gandomi AH (2024) Parkinson's disease detection based on features refinement through l1 regularized svm and deep neural network. Sci Rep 14(1) https://doi.org/10.1038/s41598-024-51600-y
25. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 15(1):1929–1958
26. Javeed A, Rizvi SS, Zhou S, Riaz R, Khan SU, Kwon SJ (2020) Heart risk failure prediction using a novel feature selection method for feature refinement and neural network for classification. Mob Inf Syst 2020:1–11
27. Javed K, Babri HA, Saeed M (2010) Feature selection based on class-dependent densities for high-dimensional binary data. IEEE Trans Knowl Data Eng 24(3):465–477

28. Ali L, Wajahat I, Amiri Golilarz N, Keshtkar F, Bukhari S.A.C (2021) Lda–ga–svm: improved hepatocellular carcinoma prediction through dimensionality reduction and genetically optimized support vector machine. Neural Comput Appl 33, 2783–2792
29. Ali L, Rahman A, Khan A, Zhou M, Javeed A, Khan JA (2019) An automated diagnostic system for heart disease prediction based on chi sequare statistical model and optimally configured deep neural network. Ieee Access 7:34938–34945
30. Maillard G, Arlot S, Lerasle M (2021) Aggregated hold-out. J Mach Learn Res 22(20):1–55
31. Yadav S, Shukla S (2016) Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification. In: 2016 IEEE 6th international conference on advanced computing (IACC), pp 78–83. IEEE
32. Mezzadri G, Laloë T, Mathy F, Reynaud-Bouret P (2022) Hold-out strategy for selecting learning models: application to categorization subjected to presentation orders. J Math Psychol 109:102691
33. Seraj A, Mohammadi-Khanaposhtani M, Daneshfar R, Naseri M, Esmaeili M, Baghban A, Habibzadeh S, Eslamian S (2023) Cross-validation. In: Handbook of hydroinformatics, pp 89–105. Elsevier, Amsterdam
34. Chicco D, Jurman G (2023) The matthews correlation coefficient (mcc) should replace the roc auc as the standard metric for assessing binary classification. BioData Mining 16(1):4
35. Yang S, Berdine G (2017) The receiver operating characteristic (roc) curve. Southwest Respirat Crit Care Chronic 5(19):34–36
36. Bergmann E, Harlev D, Wolpe N (2025) Depressive symptoms are linked to age-specific neuroanatomical and cognitive variations. J Affect Disord 369:1013–1020
37. Judijanto L, Siminto S, Rahman R (2024) The influence of religious beliefs and religious practices on social cohesion in modern society in Indonesia. Eastasouth J Soc Sci Hum 1(03):139–150
38. Zhang H, Hao X, Qin Y, Yang Y, Zhao X, Wu S, Li K (2025) Social participation classification and activities in association with health outcomes among older adults: results from a scoping review. J Adv Nurs 81(2):661–678
39. Li Y-P, Fan S-Y, Hu F-W, Liu C-Y (2025) Mediating effect of social support on the relationship between social activity and depressive symptoms among older widows in taiwan. J Women Aging 37(1):1–14
40. Jiang D, Warner LM, Chong AM-L, Li T, Wolff JK, Chou K-L (2021) Benefits of volunteering on psychological well-being in older adulthood: evidence from a randomized controlled trial. Aging Mental Health 25(4):641–649
41. Lee S (2022) Volunteering and loneliness in older adults: a parallel mediation model. Aging Mental Health 26(6):1234–1241
42. Lorenti A, De Rose A, Racioppi F (2025) Volunteering during early retirement reduces depression. Soc Sci Med, 117790
43. Qi P, Huang M, Zhu H (2024) Association between alcohol drinking frequency and depression among adults in the united states: a cross-sectional study. BMC Psychiatry 24(1):836
44. Zhang X, Lu Q, Luo Y, Wang L, Tian Y, Luo X (2024) The causal relationship between major depression disorder and thyroid diseases: a mendelian randomization study and mediation analysis. J Affect Disord 359:287–299
45. Wu Z, Yue Q, Zhao Z, Wen J, Tang L, Zhong Z, Yang J, Yuan Y, Zhang X (2023) A cross-sectional study of smoking and depression among us adults: Nhanes (2005–2018). Front Public Health 11:1081706
46. Cui Y, Duan Y, Du J, Yang L, Tian X, Liu H (2025) Relationship between leisure activity and depression in Chinese older adults: chain mediating effect of diet and cognition. BMC Geriatr 25(1):14
47. Bartrés-Faz D, Demnitz-King H, Cabello-Toscano M, Vaqué-Alcázar L, Saunders R, Touron E, Cattaneo G, Gonneaud J, Klimecki O, Bargalló N, et al. (2025) Psychological profiles associated with mental, cognitive and brain health in middle-aged and older adults. Nat Mental Health, 1–12
48. Petrova N, Khvostikova D (2021) Prevalence, structure, and risk factors for mental disorders in older people. Adv Gerontol 11:409–415
49. Tapia-Munoz T, Ajnakina O, Fancourt D, Steptoe A (2024) Personality traits and loneliness among older people in the UK: cross-sectional and longitudinal analysis from the english longitudinal study of ageing. Eur J Pers 38(4):599–614