The project report is submitted for the fulfilment of IEEE Internship program

Depression Detection from Activity Patterns: A Machine Learning Study

Submitted by: Ananya Hegde, Ananya P Shetty, Aditi Khasnis

of

PES University

Under the guidance of :

Dr. Chetana Hegde

Lead Manager - Data Science

Fractal Analytics, Bangalore

Table of Contents

1.Abstract:

Background:

Depression is a prevalent mental health condition with significant societal impact. Detecting and classifying individuals as depressed or non-depressed based on behavioural data can provide valuable insights for early intervention and support. This study aims to leverage time series data analysis and machine learning techniques to classify individuals into these two categories.

Methods:

The dataset used in this research consists of minute-by-minute activity records collected from a diverse group of 55 subjects, including 23 depressed patients and 32 non-depressed individuals. To prepare the data for analysis, a rigorous preprocessing step was applied. This involved grouping the data by date and hour and calculating hourly average activity levels. Subsequently, a TimeSeriesForest classifier was selected for its suitability in handling time series data.Hyperparameter tuning was a crucial aspect of the methodology, and Randomized Search CV was employed to identify the optimal combination of hyperparameters. This process involved systematically exploring hyperparameter spaces to maximise classification accuracy.

Results:

After extensive experimentation, the final model, trained with the best-performing hyperparameters, achieved an impressive accuracy of 78.6% and AUC of 0.85 on a separate test dataset. This outcome underscores the potential of time series analysis and machine learning in effectively classifying individuals as depressed or non-depressed based on their activity patterns.

Discussion:

The successful classification of individuals into depressed and non-depressed categories using behavioural data highlights the promise of this approach for early detection and support in mental health. By leveraging time series data and machine learning, this study offers a valuable contribution to the field of mental health research and underscores the importance of data-driven approaches in addressing mental health challenges. Further research and validation are warranted to refine and generalise this classification method for broader applications in mental health monitoring and intervention.

2.Introduction to The Global Mental Health Challenge

The World Health Organization (WHO) defines health as “a state of complete physical, mental and social well-being, and not merely an absence of disease or infirmity”, and also includes the ability to lead a “socially and economically productive life” [10,11]. Thus, mental conditions have been recognised as one of the most important dimensions of health and well-being of every individual. Depression and anxiety are two important states in the wide spectrum of mental health disorders.They affect all age groups, from paediatrics to geriatrics, and both women and men. Effects of anxiety and depressive disorders on health and well-being are multidimensional. On one hand, they are responsible for multiple somatic symptoms such as acid reflux, gastritis, palpitation, tremor, insomnia or hypersomnia, significant weight loss or gain, and on the other hand, different psycho social manifestations such as social withdrawal, depressed mood, suicidal ideation or attempt suicide, decrease productivity in the workplace, and lack of concentration [12]. People suffering from anxiety and depression are often stigmatised by society and excluded by family. They may underperform in educational institutes and workplaces. As a consequence, they become increasingly deprived of economic and social opportunities, leading to a poor quality of life [13].

Mental health disorders have emerged as a significant global health challenge, encompassing a broad spectrum of conditions that affect individuals from all walks of life. Among these, depression looms as a pervasive concern, impacting over 264 million individuals worldwide, spanning diverse age groups and backgrounds. The World Health Organization (WHO)[1] ranks depression as the foremost cause of global disability, underscoring its far-reaching impact on society.

In the face of this mounting crisis, there is a disheartening reality: even though there are known, effective treatments for mental disorders, over 75% of individuals in low- and middle-income countries receive no treatment. This staggering treatment gap persists due to formidable barriers that limit access to mental health care. These barriers include a lack of substantial investment in mental health infrastructure, an insufficient number of trained health-care providers, and the enduring social stigma attached to mental disorders.

As we delve deeper into this complex landscape, it becomes evident that traditional approaches to mental health care, including depression, have limitations. A significant challenge lies in early detection and intervention. More than 70% of individuals experiencing mental health issues, particularly depression, do not seek medical assistance in the early stages. This hesitation often results in a deterioration of their conditions, further intensifying the challenge of treatment and recovery.

However, amid these challenges, a ray of hope emerges with the advent of machine learning (ML) and artificial intelligence (AI). These technological advancements offer promising avenues for understanding, diagnosing, and effectively treating mental health disorders, including depression. They bring to light new possibilities, offering fresh perspectives and innovative solutions to bridge the treatment gap and provide comprehensive support for individuals grappling with complex mental health challenges.

2.1 Understanding Mental Health Disorders: A Complex Landscape

Mental health disorders represent a diverse spectrum of conditions, each distinguished by unique symptoms and complexities. These disorders influence an individual's thoughts, emotions, and behaviour and can be shaped by a multitude of factors, including genetic predispositions, environmental influences, and life experiences. Within this intricate landscape, several prevalent mental health disorders stand out, each demanding specific approaches to diagnosis and treatment.

2.2 Depression - A Complex Mental Health Challenge

Depression, in particular, occupies a central position within the realm of mental health disorders. It manifests as a multifaceted condition, characterised by persistent feelings of sadness, hopelessness, and a profound loss of interest or pleasure in daily activities. Beyond these core symptoms, depression extends its reach into various aspects of an individual's life, contributing to social isolation, impaired work performance, and an increased risk of physical health problems.

For many individuals, depression is not an isolated episode but a recurrent battle, punctuated by periods of remission and relapse. Some individuals experience a form of chronic mild depression, which, although less intense, still diminishes their capacity for joy and engagement in life. Moreover, depression can manifest with suicidal thoughts, necessitating immediate intervention. It often co-occurs with other mental health conditions, such as anxiety disorders or substance use disorders, further complicating treatment and exacerbating symptoms. Understanding and addressing these multifaceted dimensions of depression is vital for comprehensive support and effective treatment.

2.3 The Role of Machine Learning and AI in Mental Health

Amid the complexities of mental health disorders, machine learning and artificial intelligence offer a glimmer of hope and a transformative force in the field. These technologies have begun to play a pivotal role in advancing our understanding of various mental health conditions, including depression, and in redefining how we approach diagnosis, treatment, and support.

These technological advancements have ushered in a new era in mental health care, with far-reaching implications:

- **Early Detection and Diagnosis**: Machine learning algorithms, fueled by extensive datasets, including electronic health records and social media posts, have the capacity to identify early signs of mental health issues, employing natural language processing (NLP) techniques to examine text and speech patterns for signs of these disorders.

- **Personalised Treatment Plans**: AI systems take into account a patient's unique genetic, environmental, and lifestyle factors to create personalised treatment plans, resulting in more effective interventions tailored to the specific needs and circumstances of each individual.
- **Predictive Analytics**: Machine learning models predict the likelihood of relapse or symptom exacerbation in patients grappling with these conditions, enabling timely interventions and adjustments to treatment strategies.
- **Enhanced Accessibility**: The digital age has given rise to AI-driven chatbots and virtual mental health assistants, expanding access to mental health services. These tools offer immediate support, dispense coping strategies, and facilitate connections with mental health professionals, transcending geographical barriers and wait times.
- **Advancements in Neuroimaging**: Machine learning analysis of neuroimaging data, such as functional magnetic resonance imaging (fMRI) and electroencephalography (EEG), aids in identifying brain patterns associated with these conditions, contributing to objective diagnosis and treatment progress tracking.
- **Drug Discovery**: AI accelerates the development of novel treatments by identifying potential compounds and predicting their efficacy, potentially revolutionising the pharmaceutical industry's approach to mental health.
- **Natural Language Processing in Therapy**: NLP algorithms assist therapists by analysing therapy session content and sentiment, enabling more targeted and effective interventions.

These advancements are instrumental in ushering in a new era of mental health understanding and care, enhancing our ability to diagnose, treat, and support individuals dealing with complex mental health challenges like depression. As we explore these transformative developments further, it becomes clear that harnessing the potential of machine learning and artificial intelligence is not just an option but a necessity in navigating this challenging terrain.

This report is outlined as follows: Section 2 reviews related works, connecting them to our dataset. Section 3 elaborates on our model's architecture and problem statement. Section 4 covers implementation, while Section 5 focuses on performance analysis, comparing our model with related works. Finally, in Section 6, we conclude and discuss future enhancements.

3. Related Works

In this section, we review and discuss related research works conducted by different researchers in the field of mental health, specifically focusing on depression and anxiety prediction using machine learning techniques.

Recent research by Radwan Qasrawi et al[2] addressed the prediction of maternal depression and anxiety during the COVID-19 pandemic, focusing on pregnant and postpartum women in five Arab countries (Jordan, Palestine, Lebanon, Saudi Arabia, and Bahrain). Their cross-sectional regional study aimed to develop machine learning models for this prediction task.Qasrawi et al. collected a dataset from July to December 2020, comprising 3,569 women, of whom 1,939 were pregnant and 1,630 were postpartum. The study evaluated the performance of seven machine learning algorithms in predicting depression and anxiety symptoms.The Gradient Boosting (GB) and Random Forest (RF) models exhibited the highest accuracy values of 83.3% and 83.2%, respectively, for predicting depression. The Mathew's Correlation Coefficient was evaluated for the ML models; the Naïve Bayes (NB) and GB models presented the highest performance measures (0.63 and 0.59) for depression The study identified several significant features for predicting anxiety and depression, including stress during pregnancy, family support, financial issues, income, and social support

Ishita Bhakta and Arkaprabha Sau conducted a study targeting depression prediction in senior citizens[3]. Their research emphasised the importance of socio-demographic factors, such as age, sex, earning status, living spouse, and family type, in influencing depression among elderly individuals. They employed machine learning classifiers, including Bayes Net, Logistic, Multi-Layer Perceptron, SMO, and Decision Table, and compared their performance. Their findings revealed that the Bayes Net classifier provided the best results.The Bayes Net classifier achieved the highest accuracy (91.67%) among all the classifiers, indicating that it correctly predicted depression in senior citizens in approximately 91.67% of cases. Additionally, it exhibited a high precision of 0.92, suggesting that when it predicted depression, it was accurate 92% of the time. The ROC Area of 0.98 signifies excellent discrimination ability, while the low RMSE value of 0.25 suggests that the model's predictions were generally close to the actual values.

In an another study conducted by Sau and Bhakta, titled "Screening of Anxiety and Depression Among Seafarers Using Machine Learning Technology,"[4] the focus was on the application of machine learning algorithms to assess the mental health of seafarers.Their research emphasised the vulnerability of seafarers to mental health disorders, particularly anxiety and depression, owing to the demanding and isolated nature of their profession.The dataset for this study comprised socio-demographic, occupational, and health-related information collected from 470 seafarers at Haldia Dock Complex, India. The dataset was meticulously structured to facilitate the assessment of anxiety and depression using machine learning algorithms.In their research, Sau and Bhakta employed many machine learning classifiers-Catboost,Random Forest,Logistic Regression,Naive Bayes and SVM,compared their performances and discovered that the CatBoost machine learning algorithm appeared to be the most effective measure for screening anxiety and depression among seafarers. It achieved an accuracy of 82.6% , a precision of 84.1% and a roc-auc score of 0.882. These findings underscore the potential of machine learning technology in identifying at-risk seafarers for early referrals to psychological counselling and treatment.

[5] This study focuses on detecting depression through Twitter data using a novel approach called "Multimodal Depressive Dictionary Learning" (MDL). The researchers collected data from depressed and non-depressed Twitter users, including user profiles, anchor tweets, and tweets posted within a month from the anchor tweet. They employed various models, such as Naive Bayesian (NB), Multiple Social Networking Learning (MSNL), Wasserstein Dictionary Learning (WDL), and MDL, to identify depressed users. The results revealed that WDL was more effective than NB by 10%, demonstrating the value of latent and sparse representation. Both MSNL and MDL outperformed WDL by 5% to 8%, highlighting the benefits of modelling relationships among data types. MDL, in particular, excelled, surpassing WDL by 3% and achieving the highest performance with an 85% F1-Measure. This underscores the effectiveness of combining a multimodal strategy with dictionary learning for depression detection.

In this research [6], a survey involving 503 employees, various aspects of their well-being, job-related factors, and workplace environment were assessed using 27 predictor variables to determine their risk of depression. Three machine learning models, sparse logistic regression, support vector machine, and random forest, were compared in terms of prediction accuracy, precision, sensitivity, specificity, and AUC. Surprisingly, all models yielded similar results, with random forest achieving the highest accuracy at 88.7%, while sparse logistic regression and support vector machine reached 86.8%. Key influencing factors identified encompassed gender, physical health, job-related aspects, psychosocial protective factors, and psychosocial risk factors in the workplace. These findings underscore the potential of machine learning in accurately predicting depression risk among employees and highlight the significance of addressing these factors in developing intelligent mental healthcare systems for early depression detection in the workplace.

These studies contributed to our understanding of depression prediction and provided valuable insights into the application of machine learning techniques

3.1 Relevance to Our Dataset: A Fresh Perspective on Depression Prediction

Our dataset emerges as a significant contribution to the field of depression prediction, offering a unique and comprehensive perspective that distinguishes it from related works. While prior research endeavours have largely concentrated on forecasting depression within specific demographic niches such as pregnant women, senior citizens, seafarers, and Twitter users, our dataset takes a more expansive and inclusive approach.

At its core, our dataset captures an extensive spectrum of daily activities, providing an exceptional window into behaviour patterns intricately linked with depression. Notably, it features key attributes like timestamp, date, and activity hour, enabling profound temporal analyses. These attributes empower researchers to delve into the temporal dynamics of depression-related behaviours, deciphering how these patterns evolve over time. This emphasis on daily activities, coupled with temporal insights, adds a layer of depth and granularity to the study of depression that sets our dataset apart.

Furthermore, our dataset transcends the boundaries of Twitter-centric data, which has been a focal point in some previous investigations. It accommodates a diverse range of data types and characteristics, endowing it with a remarkable versatility. Researchers can explore a myriad of machine learning methodologies, leveraging this diversity to gain novel insights and devise innovative approaches to depression prediction. This multifaceted nature of our dataset transforms it into a dynamic platform, facilitating a more comprehensive examination of depression prediction.

Crucially, our dataset doesn't merely stop at inclusivity; it embraces a wide array of demographic groups, from diverse age brackets to various socio-economic backgrounds. This inclusivity is paramount as it enriches the broader landscape of depression prediction research. It fosters a more exhaustive exploration of how depression manifests across varied populations, recognizing that depression is not constrained by demographic boundaries. Our dataset, by encapsulating this diversity, thus serves as an invaluable resource for understanding the intricate tapestry of depression across different contexts and among various population groups.In summation, our dataset paints a holistic portrait of daily activities intricately intertwined with features like timestamp, date, and activity hour, providing critical insights into behaviour patterns and their associations with depression. This multi-dimensional dataset is poised to make substantial contributions to the field of depression prediction, offering valuable perspectives across diverse contexts and population groups. Its emphasis on daily activities, temporal dynamics, versatility, and inclusivity positions it as a cornerstone for future research in this vital domain.

4. Architecture and Modeling

In this chapter, we will delve into the architecture and modelling aspects of time series forests, a powerful and innovative approach in time series forecasting. Time series forests are a cutting-edge ensemble machine learning technique designed specifically for handling time series data.

4.1 Problem statement:

The problem at hand is the detection of depression using time series data. This includes developing a predictive model to distinguish individuals with mental disorders such as depression, and schizophrenia from those without, utilising time series data of average human activity collected on an hourly basis. The objective is to enable accurate disorder detection based solely on the patterns observed in average human activity.The rationale behind using time series data is that individuals with mental disorders often exhibit distinct patterns and deviations in their daily activities compared to those without such conditions .For instance, disruptions in sleep patterns or a decrease in physical activity may serve as indicators of depressive episodes, while unusual social behaviour or activity fluctuations might be indicative of depression.

Approach : Our solution is to use a **Timeseries forest classifier.** Timeseries forest works on random forest algorithms such that data has been modified to be used on tabular data.

You will have multiple models to be aggregated and finally the majority among the predictions has been taken. It constructs multiple trees, the input is fed to the top most node and then based on threshold values and information gain it decides whether the node should be a left child or a right child.

Some other time series classification algorithms are Dynamic Time Warping which uses a similar concept as KNN where Euclidean matching is used to check how similar two time series input sequences are with respect to each other .

Why choose Times series forest over DTW?

DTW is suitable for only smaller datasets as its time complexity is very high and cannot be used on real time datasets.Instance based algorithms provide very limited insights into temporal characteristics useful for classification.

Time series forest is a Feature based method .

Here is detailed approach of how time series forest works:

1.Ensemble Construction for Time Series Classification:

TSF builds an ensemble of decision trees, akin to the concept of Random Forest.Each decision tree in the ensemble is trained on a different random subset of the time series dataset.

2.Feature Extraction for Classification:

TSF introduces a novel approach to feature extraction from time series data.Instead of utilising the raw time series data directly, TSF transforms each time series into a set of features suitable for classification purposes.These features capture crucial information about the underlying time series, rendering them suitable for decision tree-based classification.

3.Random Feature Subsetting:

Similar to the Random Forest methodology, TSF incorporates randomness during decision tree construction.At each node within each tree, a random subset of features is considered for splitting.This random feature selection mitigates overfitting and enhances the diversity of the decision trees within the ensemble.

4.Time Series Classification:

TSF is particularly well-suited for time series classification tasks.Each decision tree within the TSF ensemble yields its own prediction for a given time series input.To obtain the final classification result, the predictions from individual trees are aggregated.Common aggregation techniques include majority voting, where the class with the most votes is assigned as the final prediction.

Time series classifiers can use various distance-based metrics, including the Euclidean distance, to compare and classify time series data. The choice of distance metric depends on the specific requirements and characteristics of the problem at hand. Here's how the Euclidean distance can be used in time series classification:

5.Distance Calculation:

Once the time series data is represented as a set of features, the Euclidean distance can be computed between these feature vectors.The Euclidean distance is a common choice because it measures the straight-line distance between two points in a multidimensional space.In the context of time series classification, this means that time series with similar feature vectors (i.e., small Euclidean distances) are considered more similar, while those with larger distances are considered less similar.

6. Hyperparameter Tuning:

TSF has hyperparameters that can be tuned to optimise its performance.These hyperparameters include the number of trees in the ensemble, the number of features to consider at each split, and the depth of the decision trees.Proper hyperparameter tuning is essential to ensure that TSF delivers the best possible results for a given time series classification model.

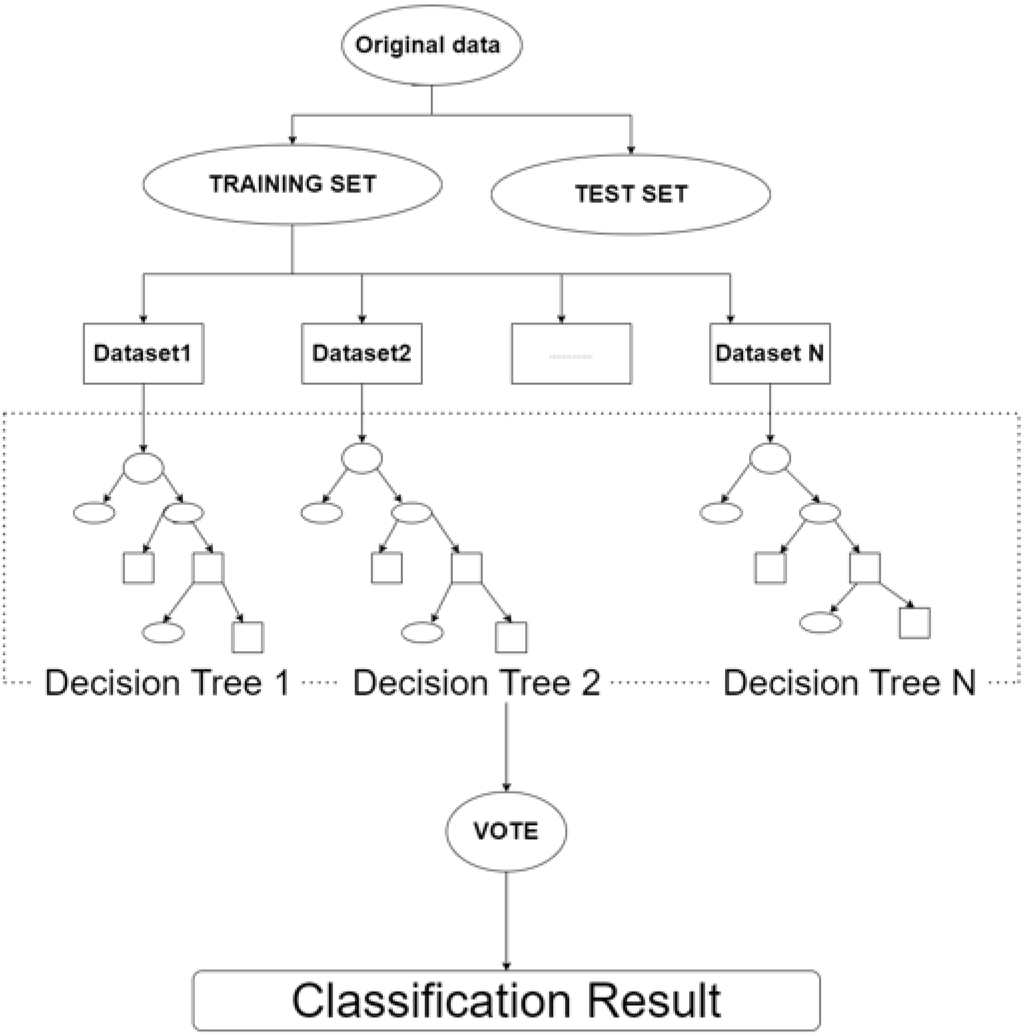

Fig. 1 Pictorial representation of Time series forest algorithm

In time series forest we define temporal features and train a classifier based on the temporal features of data.

To decide the best split in building a decision tree as shown in Fig 1 we need to build multiple treesThere are many methods such as calculating entropy and information gain. Higher the information gain the better the split is .Many factors such as purity, entropy comes into consideration .But we use class based measure to decide the split.

So, we start by setting a threshold and take our input features from k=1 to k=3 ,we have input features as date, hour and average activity. $\square_{\square}$ is a function where we calculate mean standard deviation of input features in a particular time interval t1 to t2.

$$\square_{\square}(t1, t2) \leq \tau$$

if the computed function output is less than threshold then it becomes a left child otherwise it is a right child of that particular node.

Splitting criterion is entropy gain which is again the common slitting criteria

$$E(S) = \sum_{i=1}^{c} - p_i \log_2 p_i$$

Where pi is proportion of a particular class (positive or negative)

where c is the corresponding to classes {1,2,….c} at a tree node.

Entropy gain is a concept used in decision tree algorithms, such as ID3 and C4.5, to find the optimal split at each node as seen in fig1 algorithm creates multiple decision trees. Entropy is a measure of disorder or randomness within a dataset. When you split a dataset into subsets based on a particular feature, you calculate the entropy of each subset. The entropy gain is then determined by measuring the reduction in entropy achieved by the split. A higher entropy gain indicates that the split has effectively reduced the disorder in the data, making it a good choice for decision tree construction.

On the other hand, purity and impurity measures are commonly used in various classification algorithms, including ensemble methods like Random Forest and gradient boosting. Purity is a measure of how homogeneous or pure a subset of data is with respect to a specific class label. Conversely, impurity is the opposite – it measures the degree of mixing of class labels within a subset. Common impurity measures include Gini impurity and cross-entropy. In the context of these measures, a split is considered good if it reduces impurity or increases purity.

Information gain is a concept used in decision tree algorithms, particularly in the context of feature selection and splitting nodes to build effective decision trees. It quantifies the reduction in uncertainty or entropy achieved by splitting a dataset based on a particular feature. Information gain helps in selecting the best feature to split on by evaluating how much it contributes to improving the overall purity or homogeneity of the subsets created after the split.

Δ=Entropy of the parent node – (Average weighted entropy of the child nodes)

$$\Delta = I(parent\ node) - \sum_{j=1}^{k} \frac{N(v_j)}{N} I(v_j)$$

Impurity of parent node

Average weighted entropy OR Average Information

N - total no. of instances at the parent node
k - total possible attribute values
N (□□) - total no.of instances associated with child node □□

In summary, when making decisions about splitting data, practitioners consider entropy gain as a measure of reducing disorder in decision tree algorithms, while purity and impurity measures guide the creation of pure and homogeneous subsets in classification tasks. Both concepts are essential in constructing accurate and interpretable models in data analysis and machine learning.

Time Series Forest emerges as a highly promising approach within the realm of time series analysis. Its ensemble-based methodology, combining the power of decision trees and randomness, offers an effective means of tackling the inherent complexity of time series data. Not only does it excel in classification and regression tasks, but it also provides a robust solution for handling diverse types of temporal data. Its scalability and competitive performance make it a valuable addition to the data scientist's arsenal, enabling the extraction of meaningful insights and informed decision-making from time series datasets.As the field of time series analysis continues to evolve, Time Series Forest stands as a versatile and dependable choice, ready to address the challenges of the future

4.2 Algorithm

Algorithm1 : sample() function: randomly samples a set of intervals < T1, T2 >, where T1 is the set of starting time points of intervals, and T2 is the set of ending points. The function RandSampNoRep(set, samplesize) randomly selects samplesize elements from the set without replacement [7].

```
T1 = ∅ , T2 = ∅

W = RandSampNoRep({1, ..., M}, √ M)

for w in set W do

  T1 = RandSampNoRep({1, ..., M − w + 1}, √ M − w + 1)

  for t1 in set T1 do

   T2 = T2 ∪ (t1 + w − 1)

  end for

end for

return < T1, T2 >
```

Algorithm 2 tree(data): Time series tree. For simplicity of the algorithm, we assume different types of features are on the same scale so that E can be compared [7].

< T1, T2 >=sample() calculated Threshold, the set of candidate thresholds for each feature type k

$E* = 0$, $\triangle Entropy* = 0$, $t*1 = 0$, $t2* = 0$, $\tau* = 0$, $f* = \emptyset$

for < t1, t2 > in set < T1, T2 > do

for k in 1:K do

for τ in Threshold do

calculate $\triangle$Entropy and E for $fk(t1, t2) \leq \tau$

if $E > E*$ then

$E* = E$, $\triangle Entropy* = \triangle Entropy$, $t*1 = t1$, $t*2 = t2$, $\tau* = \tau$, $f* = fk$

end if

end for

end for

end for

if $\triangle Entropy* = 0$

then label this node as a leaf and return

end if

data(left) ← time series with $f*(t*1, t*2) \leq \tau*$

data(right) ← time series with $f*(t*1, t*2) > \tau*$

tree(data(left))

tree(data(right))

Basic workflow:

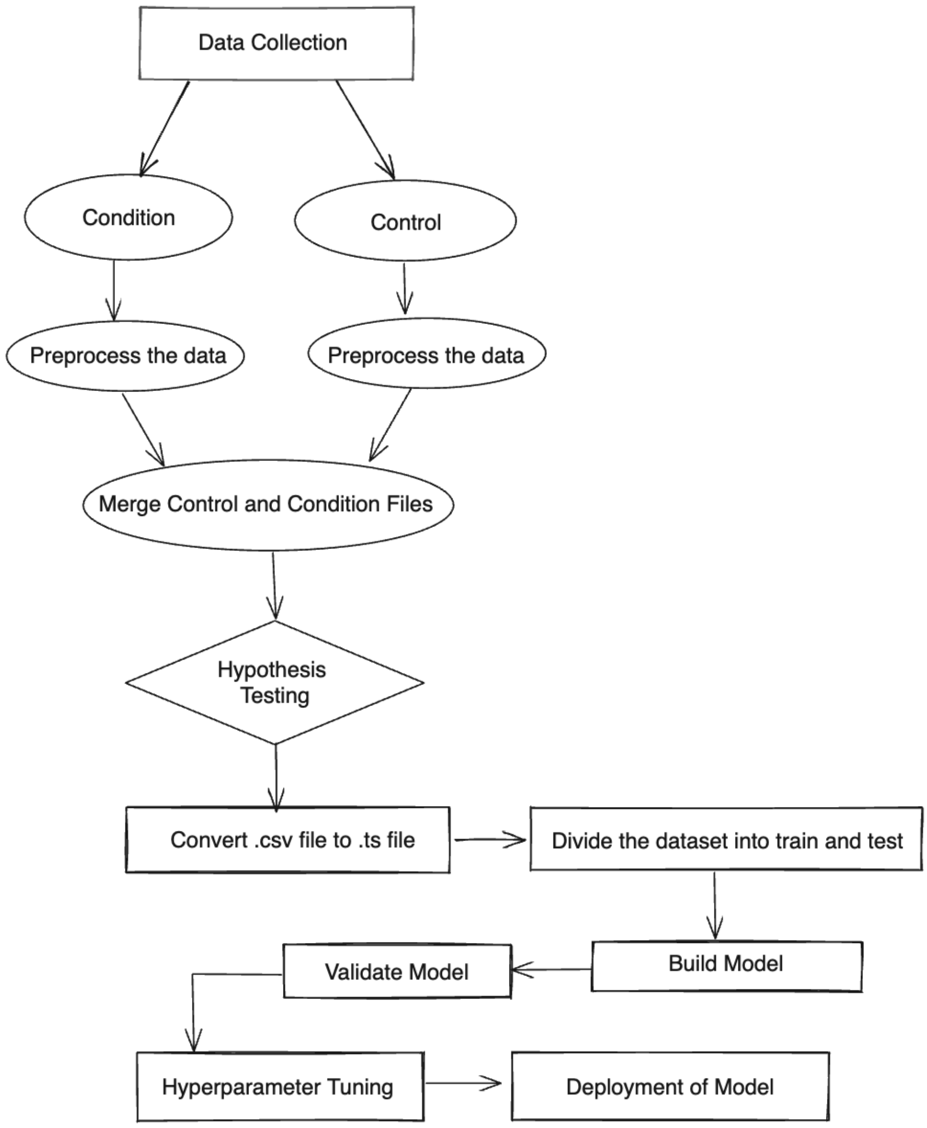

Fig. 2 Workflow of Depression Classification

Fig. 2 outlines our analytical workflow. We preprocess two datasets, "Control" and "Condition," merging them into a unified dataset with a binary "Status" column (1 for mental disorder patients, 0 for non-patients).

- **Data Preprocessing:**

1. Data cleaning ensures data integrity and consistency.
2. Missing values are handled to avoid data gaps.
3. Merging of "Control" and "Condition" datasets into a single dataset.
4. Introduction of a binary "Status" column (1 for depressed patients, 0 for non-depressed patients)

- **Hypothesis Testing:**

1. Objective: Determine if mean activity differs significantly between the two groups.
2. Statistical tests (e.g., t-tests or ANOVA) are conducted.
3. Null hypothesis (H0) There is no significant mean difference between them;
4. alternative hypothesis (Ha) There is a significant difference.

- **Model Training:**

1. Develop a predictive model to classify individuals as patients or non-patients.
2. Machine learning algorithms are used.
3. Model hyperparameters are tuned for optimal performance.
4. Descriptive Analysis for Model Deployment:

- **Model Evaluation**:

1. Assess performance metrics (accuracy, precision, recall, F1-score,AUC-ROC curve).
2. Feature Importance: Identify key variables contributing to predictions.

By adhering to this systematic workflow, we aim not only to statistically investigate the differences in activity between individuals with and without depression but also to construct a robust predictive model. This model, once validated, holds the potential to serve as a valuable tool for early detection and intervention in mental health disorders.

5. Implementation

This section provides a comprehensive overview of our project's implementation, covering key aspects such as data preprocessing, hypothesis testing, and the construction of our model.

5.1 Dataset Description

Fig. 3 represents the snapshot of the dataset. The dataset used for this project comprises three primary columns: "Time Stamp," "Date," and "Activity." It is a collection of records that capture the minute-by-minute activities of individuals [8].

	A	B	C
1	timestamp	date	activity
2	07-05-2003 12:00	07-05-2003	0
3	07-05-2003 12:01	07-05-2003	143
4	07-05-2003 12:02	07-05-2003	0
5	07-05-2003 12:03	07-05-2003	20
6	07-05-2003 12:04	07-05-2003	166

Fig. 3 Snapshot of the Original Dataset

Data Sources:

- Depressed Patients: The dataset includes 23 separate CSV files, each dedicated to a specific depressed patient. These files contain detailed activity records for each patient.
- Non-Depressed Individuals: Additionally, the dataset encompasses 32 distinct CSV files, each corresponding to a non-depressed individual. These files contain activity data for individuals who do not exhibit signs of depression.

Data Columns:

- **Time Stamp:** This column denotes the precise time at which each activity observation was recorded. It captures the minute-by-minute timeline of activities.
- **Date:** The "Date" column provides the date on which the activities were observed. It helps in organising and indexing the data temporally.
- **Activity:** The "Activity" column records the specific activity undertaken by the individuals at each time stamp. This could encompass a wide range of activities and behaviours.

Purpose:

The purpose of this dataset is to analyse and compare the activity patterns of depressed patients and non-depressed individuals. By examining minute-level activity data, this project aims to uncover potential correlations or differences in behaviour that may be associated with depression. Such insights can contribute to a better understanding of depression and inform future research and intervention strategies.

Data Organization:

- Each CSV file represents a unique individual, either a depressed patient or a non-depressed individual.
- The data within each file is organised chronologically, with time stamps indicating when each activity occurred.

5.2 Data Preprocessing

As part of our data preprocessing pipeline in Python, we utilised the "groupby" function to aggregate and summarise the raw activity data. The primary objective was to transform the data into a more manageable and meaningful format for subsequent analysis.

Grouping by Date and Hour:
The initial step involved grouping the data based on two key temporal dimensions: "Date" and "Hour." This grouping allowed us to organise the dataset into hourly segments, which is conducive to identifying patterns and trends in the activity data.

Calculating Hourly Averages:
Within each hourly group, we calculated the average activity level by summing all the data points collected on a minute-by-minute basis and then dividing this sum by 60 (representing the number of minutes in an hour). This computation yielded a representative hourly average activity value.

Consistency Across CSV Files:

This transformation process was applied uniformly to all CSV files, both for the depressed patients and the non-depressed individuals. This consistency ensured that our subsequent analysis would be based on standardised hourly averages, allowing for meaningful comparisons between the two groups.

Following the completion of the data transformation process, we arrive at the final representation of our dataset, as depicted in Fig.4.

	date	hour	Avg_activity
0	2003-03-18	15	156.48
1	2003-03-18	16	264.55
2	2003-03-18	17	1105.13
3	2003-03-18	18	338.63
4	2003-03-18	19	567.83

Fig. 4 Snapshot after data preprocessing

5.3 Hypothesis Testing

In Fig. 5, a line chart plots hourly data on the x-axis against average activity on the y-axis, visually illustrating differences between two groups: control (non-depressed) and condition (depressed).To confirm these distinctions statistically, we conducted a two-sample z-test [9]. This test assesses whether there is a significant difference between the means of two independent samples, commonly used to determine if groups exhibit statistically significant mean variations.

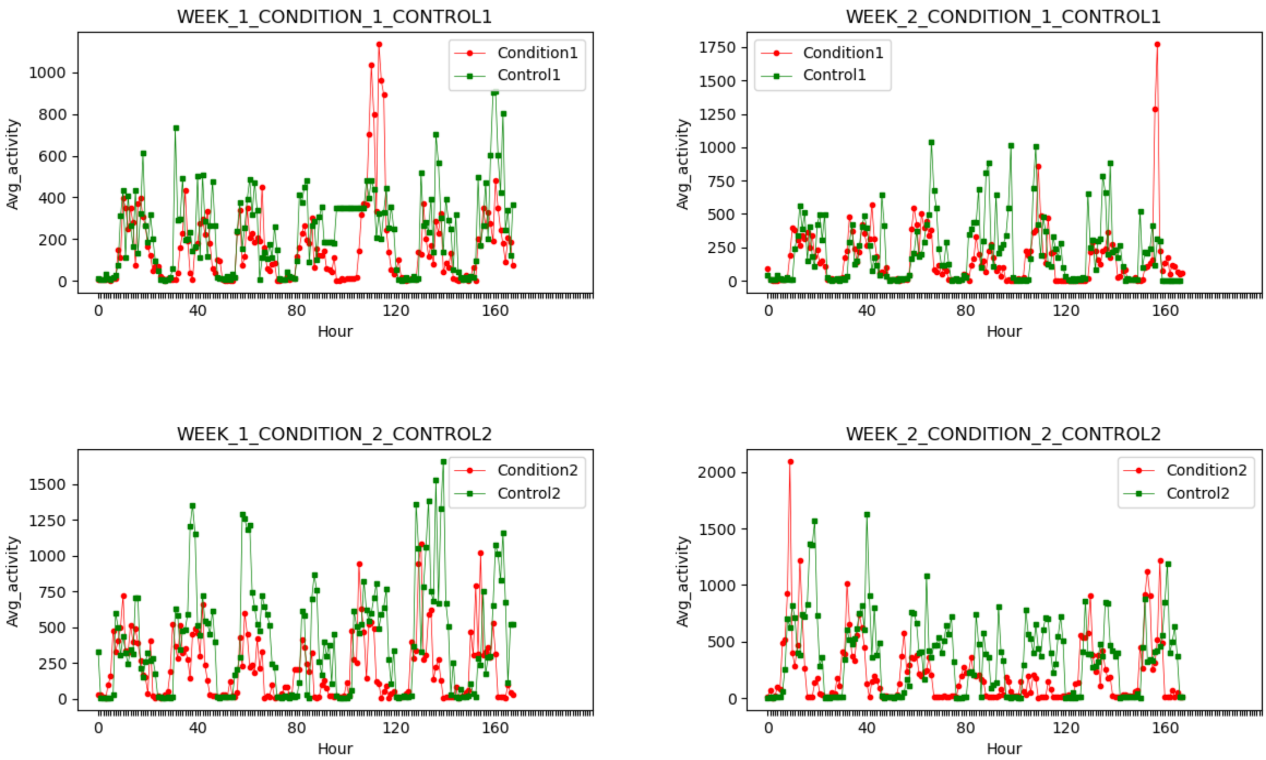

Fig.5 The line chart depicting an overview of average hourly activity across various weeks

● **Formulate the hypothesis**

The null hypothesis (H0): It serves as the baseline assumption, positing that there is no statistically significant distinction between the means of two populations: one comprising the average activity of depressed patients (with a mean denoted as $\mu 1$), and the other consisting of the average activity of non-depressed individuals (with a mean denoted as $\mu 2$). In mathematical terms, the null hypothesis can be expressed as $H0: \mu 1 - \mu 2 = 0$.

Alternate hypothesis(Ha): It proposes the presence of a statistically significant distinction between the means of two populations: one representing the average activity of depressed patients (with a mean denoted as $\mu 1$), and the other representing the average activity of non-depressed individuals (with a mean denoted as $\mu 2$). In mathematical terms, the alternative hypothesis can be expressed as $Ha: \mu 1 - \mu 2 \neq 0$.

- **Select Significance Level (α):**

We selected a significance level (α) of 0.05, which signifies the probability of committing a Type I error, where the null hypothesis is incorrectly rejected when it is actually true.

- **Calculate the test statistic**

The test statistic for a two-sample z-test is calculated using the formula:

$$z = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

where,

1. x1 and x2 denote the sample means of average activity for depressed patients and non-depressed individuals, respectively.
2. s1 and s2 represent the sample standard deviations of average activity for depressed patients and non-depressed individuals, respectively.
3. n1 and n2 indicate the sample sizes for the group of depressed patients and the group of non-depressed individuals, respectively.

To address data skewness, we applied a log transformation to the "Avg_activity" column. This transformation helps make the data more suitable for statistical analysis. We selected random samples of size 500 each from both the group of depressed patients and the group of non-depressed individuals. These samples served as representative subsets for our analysis.To calculate the z-statistics, we computed their means and standard deviations. We employed Python's NumPy module, utilising the functions numpy.std() for standard deviation and numpy.mean() for mean calculation. For hypothesis testing, we implemented a function called ztest_ind, which takes mean,standard deviation and sample size as the arguments to calculate the z-statistics and associated p-value. The z-statistics helps quantify the difference between the sample means.To determine the p-value, we utilised the SciPy library's stats.norm.cdf function. This function calculates the cumulative distribution function (CDF) of a standard normal distribution, enabling the derivation of the p-value.

Upon performing the z-test, we obtained a p-value that was less than the significance level (alpha) of 0.05. As a result, we rejected the null hypothesis and accepted the alternative hypothesis. This finding indicated that there were indeed significant differences in the average activity levels between depressed and non-depressed patients.
Our hypothesis testing provided valuable insights into the relationship between activity levels and depression, contributing to the overall understanding of the factors influencing mental health outcomes.

5.4 Proposed Model : Timeseries Forest Classifier

5.4.1 Preprocessing

The TimeSeries Forest classifier is a powerful machine learning algorithm specifically designed for time series classification tasks. It draws inspiration from ensemble learning techniques, combining multiple decision tree classifiers to make accurate predictions. What sets the TimeSeries Forest apart is its unique approach to handling time series data. Instead of using the entire time series, it randomly extracts subsequences from each series, allowing it to capture various temporal patterns and dependencies. This diversity in training data, coupled with the ensemble's majority voting mechanism, helps mitigate overfitting and enhances the model's ability to generalise. The TimeSeries Forest classifier has found applications in a wide range of domains, including finance, healthcare, and manufacturing, where it excels at classifying time series data into different categories, making it a valuable tool in the field of time series analysis.

To effectively utilise the TimeSeriesForest classifier for our task of classifying individuals as either depressed or non-depressed, we carefully structured our time series data. In our dataset, each individual's data served as a separate sample, and we collected observations over multiple time steps, defining the length of each time series. To meet the model's input requirements, we transformed our time series sequences into a 2D array format. This transformation resulted in a 2D array with dimensions (n_samples, n_timestamps), where 'n_samples' represented the number of individuals or examples in our dataset, and 'n_timestamps' denoted the count of time steps or observations recorded at each timestamp within the time series. This format ensured that the PyTS TimeSeriesForest model could analyse the temporal patterns within each individual's time series to classify them accurately as either depressed or non-depressed, based on the provided labels.

In the context of time series data, a '.ts' file format is a standardised way to store time-series datasets along with their associated metadata. Encoded in utf-8, these files can be easily opened and inspected using basic text editors like Notepad. Within a .ts file, string identifiers, denoted by strings beginning with '@', play a crucial role in specifying metadata and dataset information. The content of a .ts file is organised into three distinct blocks: a description block, a metadata block, and a dataset block. The description block contains lines beginning with '#' and can include any utf-8 text, commonly used for providing dataset descriptions, data dictionaries, or citations. The metadata block starts with '@' and is followed by various string identifiers, each with its associated value, describing the dataset's properties. There is no strict order for these identifiers, except that '@data' must appear at the end of this block. The dataset block consists of float values representing the dataset, either as a simple comma-separated list or in more complex cases with timestamps enclosed in round brackets. This structure enables the storage of diverse time-series data while preserving essential metadata and organisational clarity within .ts files.

In the "final.ts" file, depicted in Fig. 6, several key string identifiers are used to specify its characteristics and content for analysis. "@timestamps false" signifies that timestamps are not included in the file. The identifier "@univariate true" indicates that there is a single dimension for the time series, simplifying its structure. Furthermore, "@equalLength true" conveys that each instance within the dataset shares equal length, ensuring uniformity. Specifically, there are 24 timestamps in each instance, denoted by "seriesLength 24". The presence of class labels is denoted by "@ classLabel true 0 1", signifying that class labels are indeed included, with values "0" and "1" indicating the possible class assignments. The presence of actual data is marked by "@data," and the target value is mentioned after the colon(:) at the end of each row.

```
@problemName final.ts
@timestamps false
@univariate true
@equalLength true
@seriesLength 24
@classLabel true 0 1
@data
0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,346.55,284.57,279.18,218.78,238.55,209.85,129.12,60.47,99.95,41.53,32.45,148.73:1
8.0,4.08,7.07,4.03,12.22,0.97,10.0,13.7,146.73,111.45,395.65,351.95,247.33,348.03,279.88,75.92,368.17,394.63,307.8,263.57,164.32,120.05,46.52,67.87:1
69.7,22.25,11.02,2.08,12.65,4.52,6.43,5.67,36.55,161.82,228.83,432.48,192.4,38.35,7.73,160.68,181.9,276.12,298.47,223.47,333.12,178.12,58.2,36.7:1
100.62,98.63,7.72,3.28,3.7,0.58,1.97,17.78,231.55,339.67,72.85,116.67,347.07,207.83,227.57,184.88,206.7,192.22,447.58,157.02,60.08,47.67,79.97,85.48:1
2.77,0.92,4.08,6.52,20.37,7.95,13.73,10.87,117.97,158.27,228.07,267.48,196.82,178.4,299.97,62.7,152.28,124.48,123.87,144.5,57.2,52.15,43.97,111.23:1
2.77,2.67,12.45,5.25,9.82,9.68,9.68,9.87,16.9,144.57,317.8,368.25,367.7,702.27,1034.85,795.77,332.2,1137.3,959.95,892.68,244.2,140.2,56.37,31.8:1
52.0,100.48,15.73,9.53,0.3,7.3,12.53,16.45,6.45,136.48,127.62,368.3,202.28,119.78,170.33,80.25,284.63,230.13,325.3,42.45,86.9,66.05,132.22,9.9:1
3.78,2.92,14.02,5.63,25.75,2.5,17.32,65.73,1.32,199.37,169.95,349.18,264.87,330.02,274.02,192.35,481.03,349.68,246.48,179.87,91.55,206.88,187.47,74.75:1
89.08,9.9,0.9,3.6,3.57,6.45,9.57,11.58,22.47,194.1,394.03,382.55,305.28,262.12,336.23,313.35,364.55,249.0,342.5,183.78,228.3,132.62,149.73,105.78:1
8.42,19.27,13.98,7.92,6.18,17.12,26.25,177.35,221.75,481.32,371.12,243.78,138.08,218.65,423.95,353.75,269.15,316.22,569.87,317.67,180.38,57.62,70.68,61.73:1
98.4,13.47,8.02,7.07,17.83,0.92,7.98,8.35,8.88,44.3,386.42,549.1,424.75,195.88,505.1,399.88,448.15,337.48,379.08,87.7,68.45,120.53,51.02,97.48:1
72.88,11.73,13.6,6.45,6.97,10.77,9.45,47.07,42.78,3.5,117.75,163.27,328.85,198.23,115.35,147.03,65.78,225.22,276.07,177.97,75.45,99.75,32.65,96.93:1
3.95,9.03,4.92,4.22,6.92,17.55,11.8,25.32,223.67,190.92,225.68,362.5,381.48,857.32,485.22,191.55,137.02,472.08,211.33,212.0,0.0,0.0,0.0,0.0:1
0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.13,0.13,14.95,213.05,292.28,226.12,157.3,124.63,224.65,243.17,361.9,174.58,271.68,231.22,28.85,32.37,78.7:1
86.32,7.08,8.67,7.52,22.08,1.72,4.83,12.77,96.87,107.62,134.43,161.15,1289.32,1775.4,224.47,72.32,136.77,177.27,53.5,118.38,111.93,68.65,51.65,59.63:1
48.05,9.23,13.42,4.32,18.18,8.13,11.17,20.05,67.18,9.73,154.35,385.8,368.88,546.12,11.77,0.0,0.0,0.0,0.08,0.0,23.67,0.0,6.57,0.0:1
```

Fig.6 - Data after the conversion from .csv to .ts file

5.4.2 Data Conversion to '.ts' Format

We initialised an empty list grouped_data to store grouped time series data.It iterates through the sorted list of filenames, reads each CSV file into a pandas DataFrame (df), and converts the 'Date' column to datetime format.The data is then grouped by 'Date,' and the 'Avg_activity' values for each date are collected into a list using apply(list). The results are reset into a new DataFrame 'grouped' .Each grouped DataFrame is appended to the 'grouped_data' list.Then we concatenated all the grouped DataFrames into a single DataFrame named 'all_grouped_data'. The 'to_time_series_dataset' function is used to convert the 'Avg_activity' column of 'all_grouped_data' into a time series dataset.Finally, we created target labels y. In this case, you generate an array of '1' values with the same number of elements as in reshaped_array. These labels can be used for classification tasks. Output of the conversion is shown in Fig.7

```
          Date                                        Avg_activity
0   2003-05-07  [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...
1   2003-05-08  [8.0, 4.08, 7.07, 4.03, 12.22, 0.97, 10.0, 13....
2   2003-05-09  [69.7, 22.25, 11.02, 2.08, 12.65, 4.52, 6.43, ...
3   2003-05-10  [100.62, 98.63, 7.72, 3.28, 3.7, 0.58, 1.97, 1...
4   2003-05-11  [2.77, 0.92, 4.08, 6.52, 20.37, 7.95, 13.73, 1...
..         ...                                                 ...
400 2002-06-24  [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...
401 2002-06-25  [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...
402 2002-06-26  [0.57, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,...
403 2002-06-27  [0.0, 0.0, 0.28, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,...
404 2002-06-28  [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...

[405 rows x 2 columns]


Shape of reshaped array:  (405, 1, 24)
(405,)
```

Fig. 7 - The data after grouping by 'date'

The same data processing steps are executed for both the datasets containing information about individuals classified as depressed and non-depressed. Subsequently, the 'reshaped_array' and the target array 'y' for both groups are concatenated. To transform this combined array into the final.ts file format, we utilise the write_to_tsfile function available in the 'aeon.datasets' library. This process streamlines the conversion of our structured data into the standardised '.ts' file format for further analysis and assessment.

5.4.3 Model building

We've constructed a TimeSeriesForest model for classifying individuals into depressed patients and non-depressed individuals. The code depicted in the Fig.10 outlines the steps involved in building this model:

Data Loading: To start, we utilise the ‘load_from_tsfile’ function from the ‘sktime.datasets’ module to load our time series data. We specify ‘return_data_type="numpy2d"’ to receive the data in the form of a NumPy 2D array.

Data Splitting: The dataset is divided into training and testing subsets using the ‘train_test_split’ function from the ‘sklearn.model_selection’ module. In this process, 'X' represents the time series data features, while 'y' signifies the target labels. For reproducibility, we set a random seed with random_state=42. Furthermore, we allocate 20% of the data for testing, specifying test_size=0.2. Importantly, we use stratify=y to ensure that the class distribution is maintained during the split, a crucial step for balanced classification datasets

Model Initialization: We create an instance named ' clf ' for the TimeSeriesForest classifier. To ensure reproducibility, we set a random seed with random_state=43. Additionally, we specify n_estimators=20, determining the number of decision trees in the ensemble (forest).

Model Training: The fit method is employed to train the classifier on the training data, composed of 'X_train' and 'y_train'.

This establishes a streamlined pipeline for handling time series data. It encompasses data loading, splitting, classifier initialization, and model training. Ultimately, the trained model ('clf') can be utilised for making predictions and evaluating its performance on the test data ('X_test' and 'y_test').

5.4.4 Hyperparameter Tuning

Hyperparameter tuning is a crucial step in optimising machine learning models. It involves finding the best set of hyperparameters for a model to achieve the highest performance. Below are two common approaches for hyperparameter tuning:

Grid Search CV: It is a comprehensive hyperparameter tuning approach used in machine learning. In this method, you define a predetermined set of hyperparameter values to explore. It systematically generates a grid of all possible combinations of these hyperparameters and rigorously assesses the model's performance using cross-validation for each combination. While Grid Search is known for its thoroughness in searching for optimal hyperparameter settings, it can become computationally demanding, particularly when dealing with a vast hyperparameter space. Therefore, Grid Search is an ideal choice when you have a good understanding of the hyperparameter values likely to yield the best results.

Randomized Search CV: It offers an efficient alternative to Grid Search in hyperparameter tuning. Rather than exhaustively testing all potential combinations, it randomly selects a predefined number of hyperparameter configurations. This randomness significantly enhances computational efficiency, particularly in scenarios involving expansive hyperparameter spaces. Randomized Search is particularly valuable when you need to explore a broad range of hyperparameter values, but the exhaustive examination of all combinations is impractical. It strikes a balance between thoroughness and computational feasibility, making it a valuable tool in the hyperparameter tuning toolkit.

The Hyperparameter tuning using Randomized Search CV for a machine learning classifier is as follows:

'param_dist': This dictionary defines a range of hyperparameter values to be explored during the search. The hyperparameters being tuned include "max_depth" (tree depth), "criterion" (splitting criterion), "max_leaf_nodes" (maximum leaf nodes in a tree), "n_estimators" (number of decision trees in the forest), and "min_samples_leaf" (minimum samples per leaf node).

Classifier Initialization: An instance of the TimeSeriesForest classifier, clf, is created without specifying any hyperparameters.

Randomized Search CV: The RandomizedSearchCV function is used to perform randomized hyperparameter search. It takes several arguments:
clf: The classifier for which hyperparameters are being tuned (TimeSeriesForest in this case).
param_distributions: The dictionary of hyperparameter ranges defined earlier (param_dist).
n_iter: The number of random combinations of hyperparameters to try during the search (50 in this case).
cv: The number of cross-validation folds to use for evaluating each combination of hyperparameters (5-fold cross-validation).
scoring: The evaluation metric to optimise during the search (accuracy in this case).

Fitting the Randomized Search: The random_search object is fitted to the training data (X_train and y_train). It systematically explores different hyperparameter combinations, evaluating their performance using cross-validation.With the optimal hyperparameters identified to maximise accuracy, we can now proceed to train the model using these settings. Once trained, this tuned model can be employed for making predictions on new data.

6. Performance Analysis

Performance Metrics: They are quantitative measures used to assess the quality or effectiveness of a model, system, process, or any other entity or operation. In the context of machine learning and data analysis, performance metrics are crucial for evaluating the performance of models and algorithms.They provide objective and measurable insights into how well a model or system is performing in terms of its intended goals. They help in comparing different models, making informed decisions, and optimising processes.

6.1 Performance Metrics

1. **ROC-AUC (Receiver Operating Characteristic - Area Under the Curve):**
   This serves as a widely-used metric for evaluating the performance of binary classification models. It provides valuable insights into the model's ability to distinguish between the positive and negative classes. The ROC curve is a graphical representation that illustrates the trade-off between the true positive rate (sensitivity) and the false positive rate (1-specificity) at various threshold values. By quantifying the area under this curve, ROC-AUC summarises the model's discriminatory power. A higher ROC-AUC value, approaching 1, signifies superior class separation, with a perfect model achieving an ROC-AUC of 1, while a random guess corresponds to an ROC-AUC of 0.5.

2. **Accuracy:** It is a simple metric that measures the proportion of correctly classified instances out of the total. It provides an overall assessment of model correctness. However, it's less suitable for imbalanced datasets, where one class dominates, as high accuracy can result from merely predicting the majority class.

3. **Precision:** Precision, or positive predictive value, assesses the accuracy of a model's positive predictions. It's calculated as the ratio of true positives to the total positive predictions (true positives + false positives). Precision is vital when minimising false positives is important, revealing the proportion of correctly predicted positives.

4. **Confusion Matrix:** It is a fundamental tool in the evaluation of classification models, providing a comprehensive summary of the model's performance by breaking down predictions and actual outcomes into four categories. These categories are:

   True Positives (TP): These are instances where the model correctly predicted the positive class. In other words, the model identified a positive case, and it was indeed a positive case.
   True Negatives (TN): These are instances where the model correctly predicted the negative class. The model identified a negative case, and it was indeed a negative case.

False Positives (FP): These are instances where the model predicted the positive class incorrectly. The model identified a negative case as positive.
False Negatives (FN): These are instances where the model predicted the negative class incorrectly. The model identified a positive case as negative.

The confusion matrix is typically presented in a tabular format, making it easy to visualise model performance. It serves as the basis for computing various performance metrics such as accuracy, precision, recall, and F1-score.

Fig. 8 illustrates the confusion matrix for the training data. On the x-axis, you'll find the predicted class labels, while the y-axis corresponds to the actual labels. Inside the confusion matrix, each cell captures the combination of predicted and actual class labels. The diagonal cells, running from the top-left to the bottom-right, signify accurate predictions, while cells off the diagonal indicate incorrect predictions. This matrix is a crucial tool for evaluating the model's performance and assessing how well it aligns with the actual class labels. Based on the information within the confusion matrix, the training accuracy stands at 77.06%, while the training precision is recorded at 77.73%.

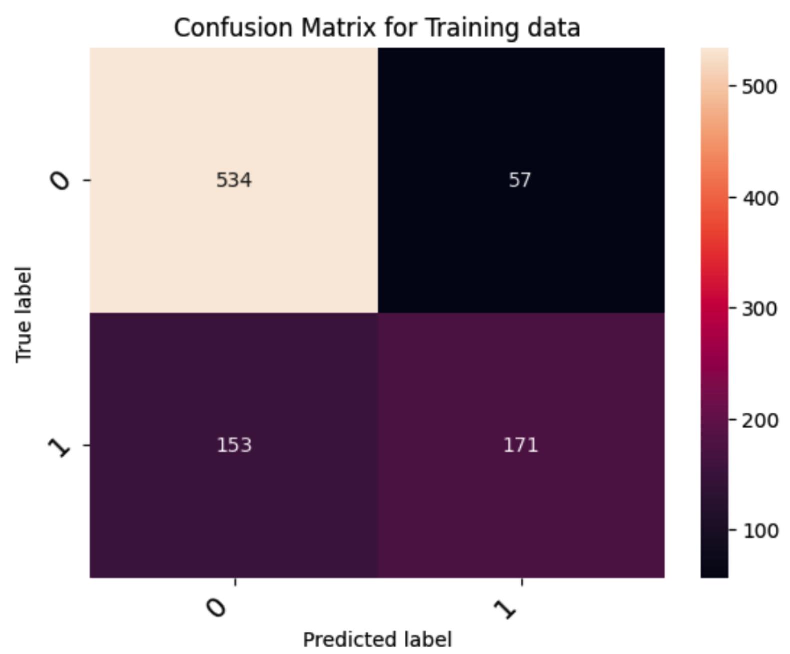

Fig.8 Confusion Matrix for Training data

Fig. 9 displays the confusion matrix for the testing data, enabling the calculation of testing accuracy at 78.6%. Additionally, the precision for the testing dataset is found to be 77.73%.Observations indicate that the model demonstrates generalisation rather than memorization, as there is no evidence of overfitting.

Fig.9 Confusion Matrix for Testing data

Fig. 10 illustrates the ROC curve, showing that both training and testing AUC (Area Under the Curve) values stand at a commendable 0.85.

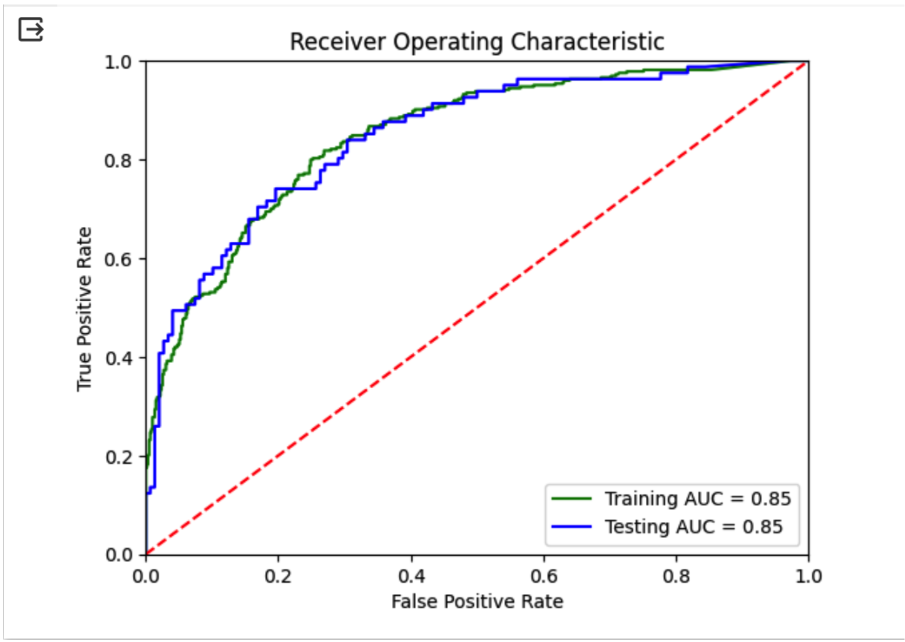

Fig. 10 ROC Curve

6.2 Comparative Performance Analysis: Our Model Versus Related Works

In this comparative performance analysis, we delve into the technical aspects of our depression prediction model, highlighting its training and testing accuracy along with its ROC AUC score in comparison to related works.

Overview of Related Works

Previous research has primarily focused on specific demographic groups, including pregnant women, senior citizens, seafarers, Twitter users, and employees.These studies have reported notable achievements in terms of prediction accuracy, with the highest reported accuracy reaching 91.67% in a study focused on senior citizens [2].

Training and Testing Accuracy of Our Model

Our model is designed to predict depression based on minute-by-minute activity data, offering a nuanced perspective on depression dynamics.Notably, our model achieved a training accuracy of approximately 77.06%, closely aligned with the testing accuracy of 78.6%.This near equivalence between training and testing values indicates that our model has not overfitted to the training data, demonstrating its robustness.

ROC AUC Score

Our model achieved an ROC AUC (Receiver Operating Characteristic - Area Under the Curve) score of 0.85.The ROC AUC score of 0.85 reflects the model's effectiveness in distinguishing between depressed and non-depressed individuals, further supporting its predictive capability.

Comparison with Related Works

Table. 1 represents the results of various studies, each of which used different machine learning models for their classification tasks. These models include Gradient Boosting (GB), Random Forest (RF), Bayes Net, CatBoost, and TimeSeriesForest.

Study	Model Used	Training Accuracy	Testing Accuracy	ROC AUC Score

Qasrawi et al. [2]	GB	—-	83.3%	—-
Qasrawi et al. [2]	RF	—-	83.2%	—-
Bhakta & Sau [3]	Bayes Net	—-	91.67%	0.98
Sau & Bhakta [4]	CatBoost	—-	82.6%	0.882
Study [5]	MDL	—-	85%	—-
Our Model	TimeSeriesForest	77.06%	78.6%	0.85
Study [6]	Random Forest	—-	88.7%	—-

Table. 1 Comparative Performance Analysis: Our Model Versus Related Works

Significance in Technical Terms

The near-equal training and testing values suggest that our model generalises well to new data and has not overfitted indicating its robustness and generalisation capability.Its performance, comparable to related works, underscores its potential as a valuable tool for depression prediction.The inclusion of a diverse demographic group within our dataset strengthens the model's utility in capturing depression dynamics across various contexts and population groups.

In conclusion, our depression prediction model demonstrates solid technical performance, with closely aligned training and testing accuracy values (approximately 77.06% and 78.6%, respectively) and an ROC AUC score of 0.85. While it may not outperform related works in every aspect, its near-comparable accuracy values ensure that we have a promising approach to depression prediction, particularly within diverse demographic contexts.

7. Conclusion

In this study, we embarked on a comprehensive exploration of depression classification utilising time-series data. We meticulously preprocessed and transformed the data, ensuring its suitability for analysis. Employing a TimeSeries Forest classifier, we achieved promising results in classifying individuals into depressed and non-depressed categories.

The absence of overfitting and the model's ability to generalise underscore its robustness. The confusion matrices provided insights into the model's predictive performance, with training and testing accuracies both exceeding 77%, and precision consistently at 77.73%. Moreover, the ROC curve, with an AUC of 0.85, further validated the model's discriminative power.

This project contributes to the growing body of research on depression classification, emphasising the potential of time-series data and machine learning techniques in identifying individuals at risk. The methods and insights presented here offer valuable guidance for future studies in this domain, with implications for early detection and intervention in mental health.

Key Findings and Contributions:

Through this research endeavour, we made noteworthy contributions:

Data Standardization: By standardising minute-level data into hourly averages, we enabled meaningful comparisons and analyses, a crucial step in revealing behavioural patterns.

Prediction Model: The TimeSeriesForest classifier exhibited substantial potential in forecasting depression based on activity patterns, offering a valuable tool for early detection and intervention.

7.1 Future Enhancements

While this research has made significant strides, we recognize several avenues for further exploration and development:

Feature Engineering: Investigating advanced features or transformations, potentially incorporating external data sources, could enhance the model's predictive capabilities.

Temporal Analysis: Incorporating sophisticated temporal analysis techniques, such as time series decomposition or recurrent neural networks (RNNs), may capture nuanced behavioural changes over time more effectively.

Multi-Modal Data Fusion: Combining various data modalities, such as sensor data, social interactions, or sentiment analysis of text data, could yield a more holistic and accurate model.

Explainability: The development of methods to elucidate model predictions is paramount, particularly in healthcare applications where interpretability is crucial.

Clinical Deployment: If viable, adapting the model for clinical deployment as a decision support tool for healthcare professionals in diagnosing or monitoring depression could be transformative.

Longitudinal Data: Acquiring longitudinal data to track individuals' activity and depression status over time can facilitate more precise predictions and risk assessments.

Ethical Considerations: Ensuring ethical standards, particularly regarding data privacy and informed consent, is of paramount importance when working with sensitive healthcare data.

In conclusion, this research has laid a solid foundation for depression prediction based on minute-level activity data. While the obtained results are promising, continuous research and enhancements are imperative to refine the model's accuracy and applicability in real-world contexts. Depression, a complex and multifaceted condition, deserves ongoing efforts that can contribute significantly to early diagnosis and intervention, ultimately enhancing the lives of individuals affected by depression.

8. References

[1] https://www.who.int/news-room/fact-sheets/detail/depression
[2] Qasrawi R, Amro M, VicunaPolo S, Abu Al-Halawa D, Agha H, Abu Seir R, Hoteit M, Hoteit R, Allehdan S, Behzad N, Bookari K, AlKhalaf M, Al-Sabbah H, Badran E, Tayyem R. Machine learning techniques for predicting depression and anxiety in pregnant and postpartum

women during the COVID-19 pandemic: a cross-sectional regional study. F1000Res. 2022 Apr 4;11:390. doi: 10.12688/f1000research.110090.1. PMID: 36111217; PMCID: PMC9445566.

[3] Bhakta, Ishita, and Arkaprabha Sau. "Prediction of depression among senior citizens using machine learning classifiers." International Journal of Computer Applications 144, no. 7 (2016): 11-16.

[4] Arkaprabha Sau, Ishita Bhakta ,Screening of anxiety and depression among seafarers using machine learning technology,Informatics in Medicine Unlocked,Volume 16,2019,100228,ISSN 2352-9148

[5] Shen, Guangyao, Jia Jia, Liqiang Nie, Fuli Feng, Cunjun Zhang, Tianrui Hu, Tat-Seng Chua, and Wenwu Zhu. "Depression detection via harvesting social media: A multimodal dictionary learning solution." In *IJCAI*, pp. 3838-3844. 2017.

[6] Kim SS, Gil M, Min EJ. Machine learning models for predicting depression in Korean young employees. Front Public Health. 2023 Jul 12;11:1201054. doi: 10.3389/fpubh.2023.1201054. PMID: 37501944; PMCID: PMC10371256.

[7] Deng, H., Runger, G., Tuv, E., & Vladimir, M. (2013). A time series forest for classification and feature extraction. Information Sciences, 239, 142–153.

[8] Enrique Garcia-Ceja, Michael Riegler, Petter Jakobsen, Jim Tørresen, Tine Nordgreen, Ketil J. Oedegaard, and Ole Bernt Fasmer. 2018. Depresjon: a motor activity database of depression episodes in unipolar and bipolar patients. Proceedings of the 9th ACM Multimedia Systems Conference (MMSys '18). Association for Computing Machinery, New York, NY, USA, 472–477.

[9] Sandra M.C. Pereira, Gavin Leslie,Hypothesis testing,Australian Critical Care,Volume 22, Issue 4,2009,Pages 187-191,ISSN 1036-7314

[10] World Health Organization. "Preamble to the Constitution of WHO as adopted by the International Health Conference." New York 19 (1946): 100.

[11] Park, Kumud. "Park's textbook of preventive and social medicine." Preventive Medicine in Obstet, Paediatrics and Geriatrics (2005).

[12] Michael Van Ameringen, M. D. "Comorbid anxiety and depression in adults: Epidemiology, clinical manifestations, and diagnosis."

[13] World Health Organization. "Depression in India, Lets talk." World Health Organization, http://www. searo. who. int/india/depression_in_india. pdf (2017).