FMDs: Forecasting Global Trends in Mental Health Disorders

Diaa Salama AbdElminaam[1],
Moshir Ashraf [2], Khaled Yehia [3], Adam Loay [4],Dalia Raafat[5]
Faculty of Computer Science
Misr International University, Cairo, Egypt
diaa.salama[1],
moshir2107190[2], khaled2103399[3], adam2100199[4], dalia2110446 [5]{@miuegypt.edu.eg}

***Abstract*—In a world grappling with deteriorating mental health, this paper aims to forecast the prevalence and frequency of mental disorders across three distinct datasets obtained from Kaggle. The disorders examined include schizophrenia, anxiety disorders, bipolar disorder, eating disorders, drug addiction, alcoholism, and depression. Statistical models such as ARIMA, AR, and MA, as well as deep-learning models like LSTM and RNN, were comprehensively assessed and compared based on their performance across four key metrics: MSE, MAE, RMSE, and SMAPE. The results underscore the relevance of trend-based patterns in mental disorders globally. However, the study strongly favors Exponential Smoothing for its superior predictive capabilities and robust trend capture. This research contributes to advancing mental health forecasting methodologies, providing valuable insights into effective predictive modeling approaches for global trends in mental health disorders.**

Keywords: Forecasting; Time series Analysis; Mental health; Deep Learning; ARIMA; CNN; RNN; LSTM; FNN; Moving Average; Exponential Smoothing

I. Introduction

Mental health disorders haunt an ever increasing segment of the human population all around the globe, many of which go on to miserably live their entire lives as social outcasts without ever getting properly treated or diagnosed. As a result, they lead needlessly burdened lives severely limiting their quality of life and overall happiness and satisfaction. However, in recent years, people have become aware of the importance of their mental health and the impact of ignoring the drawbacks of untreated mental disorders. Ultimately, as is the case with any health issue, mental disorders must be tracked and monitored to disseminate whether global efforts to combat them are actually effective.

In recent years, more and more health organizations have leveraged Artificial Intelligence to keep track and counteract various global health issues through a plethora of different applications most prominent of which is forecasting.Forecasting is the practice of analysing previous historical data to recognize trend patterns allowing one to produce an educated assumption about how future data may actually look like.

So, in an attempt to gain a tangible metric for the efficacy of global efforts being undergone to treat mental disorders and illnesses, Deep-learning and Statistical models have been deployed to attempt to forecast variable disorder percentages. This paper utilizes Moving Average, Exponential Smoothing and ARIMA statistical models while also implementing CNN, RNN, LSTM and FNN to accurately forecast many probable disorder percentages among the populace including schizophrenia, depression and anxiety disorders in the upcoming years.

First off, the next section of the paper shall showcase a selection of related works concerned with mental health forecasting. Moving on, a faithful detailed description of the collected datasets will be illustrated along with the various data pre-processing methods applied in the following section. Furthermore, in the proceeding section, the implemented models shall be explained with due care and attention to the finer details of each accompanied by a suite of performance metrics used to arbitrate each model's efficiency. Finally, the paper wraps up with the forecasting results and analysis in the penultimate section followed by the terminal section giving a definitive conclusion of this research discussing the best forecasting model along with a summary of it's predicted projections for the upcoming years. [1]

Fig. 1. Illustration of forecasting

II. Related Work

In the field of mental health forecasting, many researchers have carried out an intensive study on and developed different forecasting methods. The present research paper puts emphasis on key researches that have been rigorously examined so as to enlighten and support our study and we will be referring them in the references section.

In the paper "Planning the Future in a Longer Perspective: "A one Week Forecast of Mental Health" [2] - Naoki Tateyama et al. proposed mental health predictor with a unique point of planning long-term. A research having 105 participants within 4 weeks analyzed that as much as using one-week forecast was considered. The results indicate the conventional mood indicators stimulate the traditional scheduling habit with no innovative approaches, but our mood indicator significantly helps with regard to shaping new work schedules. The research supplies a blueprint to ideal presentation of forecasts when dealing with mental issues.

The article "Forecasting Mental Stress Utilizing Machine Learning Algorithms" [3] authored by Elias Hossain, Tanvir Rahman, and Saiful Islam examines the effectiveness of machine learning techniques to forecast mental stress. They applied many models such as decision trees, support vector machines, and neural networks to the real data of physiological parameters and survey responses. This study established that the algorithms had a high level of prediction, indicating that timely interventions and good stress management were possible when stress levels were accurately forecasted.

The paper "A Novel Approach to Forecasting the Mental Well-Being Using Machine Learning" [4] by Rayan Alanazi and Saad describes the concept of a machine-based learning system that predicts people's mental well-being. They employed different algorithms including Random Forest, Gradient Boost, Support Vector Machine, etc. , to analyze physiological data and behavioral signals. It was figured out from this study that machine learning algorithms could do mental health predictions successfully. This would facilitate the implementation of such tools in monitoring and treatment of psychological problems.

In the paper "The Future of Mental Health Care: "Trends and Future Directions"" [5] - James Lakes addresses emerging trends and, the new directions in mental health care. The paper will tackle digital health technologies that have made leaps and bounds and the integration of mental health services into primary care as well as personalized treatment approaches. It highlights the fact that early and preventive methods of treating mental health are more efficient and reliable than the conventional methods which suggests that in the future a significant boost in the accessibility and efficiency of mental health care systems is very possible.

In their "DeepMood" [6] article, Yoshihiko Suhara et al. introduce a mood forecasting framework through the implemenation of Deep Learning techniques. The system utilizes sensors in the mobile apps and wearable devices to enable user's moods to be predicted accurately. The research presented is the application of LSTM (long short term memory) networks in conjunction with a potential real-time mood monitoring and prediction. obtained data in this study have a huge implication of mental health interventions for improving user well being through continuous mood monitoring.

In the paper "Anxiety and Affective Forecasting" [7] the author Van Dang describes how people with anxiety mispredict their future emotions. The study investigates if feeling in control impacts prediction of negative feelings in highly apprehensive subject. The finding suggests that people who believe that events are out of their control show stronger negative affect, which damages the accuracy of their predictions. In this researcher stresses cognitive bias in anxious people which are responsible for maladaptive behaviors and deterioration of functioning.

In the Journal "Time Series Forecasting in Anxiety Disorders of Outpatient Visits using Data Mining", [8], Vatinee Sukmak, Jaree Thongkam as well as Jintana Leejongpermpoon study forecasting anxiety disorder outpatient visits with Data Mining. The study contrasts between two neural models, Radial Basis Function (RBF), and Multi-Layer Perceptron(MLP), and as a result RBF turns out to be a more accurate one. Systematic forecasting requires allocation and planning of resources in health care institutions.

The paper "Forecasting of The Number of Schizophrenia Disorder by using The Box-Jenkins of Time Series Analysis" [9],by Syifa Putri Humaira, Indah Nursuprianah, and Darwan makes use of the ARIMA Box-Jenkins method to forecast the cases of schizophrenia disorderData from 2014 to 2017 which will be studied in order to identify ARIMA (0,1,1) as the perfect model. In healthcare, one of the greatest applications of models like the one developed in this report is forecasting total number of patients who will attend healthcare facilities in one calendar month. Therefore, the generated results aid in the planning of healthcare resources and interventions needed, thus demonstrating the model's accuracy in forecasting.

In the paper "Time Series Analysis for Psychological Research: "Examining and Forecasting Change" [10], Andrew T. Jebb, Louis Tay, Wei Wang, and Qiming Huang emphasize the fact that time series analysis is of utmost importance in psychological one of the key method for psychological research, (Jebb et al. 2015). They underline the use of such procedures to realize these goals- temporal adaptability, seasonal modeling, and forecasting future psychological states. A job search behavior of online job example also demonstrates that the Autoregressive Integrated Moving Average (ARIMA) modeling is a very useful way of forecasting. The paper focuses on the necessity of linking collective analyses with swift predictions in the psychological theory and practice.

In the paper "Forecasting the Future, Remembering the Past: Misrepresentations of Daily Emotional Experience in Generalized Anxiety Disorder and Major Depressive Disorder" [11], Danielle C. Mathersul and Ayelet Meron Ruscio undertake the research to determine if patients with GAD and MDD provide the impression of a negative emotional state associated with their illness. The clinical group stands out demonstrating the fact that while forecasting, experiencing, and remembering more negative and a less positive are more frequent in such group than in controls. Biases in emotional forecasting and memory are the two targets for interventions that are recommended in the effectiveness of these standards of care.

This paper[12] raises concerns on how COVID-19 could increase risks of mental health issues. They aimed to propose an approach to forecast mental health in Argentina using social media data. They used a dataset containing 150 million tweets. The data was used with 3 time series forecasting strategies (Univariate Forecasting, Multivariate Forecasting, and Deep learning forecasting). They concluded that the forecasting strategies gave different capabilities in forecasting, but the neural network strategy was effective in forecasting even with limited data. The focus on twitter is a limitation as it doesn't represent the entire populatio

This paper's[13] goal is to forecast depression in elderly Chinese people by comparing the results of both regression based models and Machine learning model. They used 3 regression models (logistic regression, lasso, ridge), and used a machine learning model (random forest). The models were assessed using repeated nested 10-fold cross-validation, and the area under the receiver operating characteristic curve was the main measure of performance. They concluded that the regression based and machine learning models performed equally well, and that model choice should be based on ease of use in the specific situation. This paper's limitation is that it used self-reported data which could introduce recall bias.

The research paper [14] investigates the increasing number of suicide rates in India over the past five decades by projecting future numbers and formulating preventive strategies. The purpose of the study is to study historical data from National Crime Records Bureau reports using the ARIMA model of time series analysis to forecast suicide rates for the forthcoming ten years. Among the different ARIMA models tested, it was found that ARIMA(4,1,0) was the most appropriate model, which had the best performance with the lowest AIC, BIC, and AICc values. The model predicts that the suicide rate will be 10. 15 per 100,000 within a 95% confidence interval of 6. 92 to 13. 38 for the coming decade. The research confirmation the growing concern about suicidal deaths in India and validates the need for focused interventions and policies. Yet the study noted the possibility of under-reporting and bias in suicide reporting in India which warned the generalizing results and recommended more studies to improve data accuracy and analysis.

The research [15] was targeted at assessing the incidence of schizophrenia in Hamadan province, Iran, between 2008 and 2016, using the Holt-Winters exponential smoothing (HWES) method as the analytical tool. Researchers used medical records from a psychiatric hospital and the HWES model to calculate the frequency of schizophrenia. The results demonstrated a highly significant increase in the prevalence of schizophrenia from 2016 onwards, the HWES method was found to be a very effective tool for analyzing and predicting such data. The study revealed that the Epidemiologic investigations, especially time series analysis like HWES, can be very useful in public health efforts to reduce schizophrenia. Nonetheless, the reliance on hospital records and the absence of data before 2008 are two main shortcomings of this study, indicating the necessity of a more comprehensive population-based research with validated instruments to fill in the gaps in the literature and optimize the epidemiological boundaries for the detection of disorders within populations.

This research textcolorblue[16] sought to use Twitter data to forecast depression and PTSD in the initial stages, thereby meeting the call for the availability of low-cost and accessible screening methods for mental health. The goal was to design a computational model that would use the content of tweets from a user to determine whether they were more likely to experience these mental illnesses. A technique designed for the supervised learning classifiers training with the Random Forests model displaying the best accuracy. The outcome demonstrated that the models were efficient with recall values of 0. 852, specificity of 0. 866, and precision of 0. While the accuracy and recall metrics attained were 0. 858 and the F1 score was 0. Support Vector Machines are known to reach an accuracy of 0. 672, while PTSD classifiers have a recall of 0. 683, specificity of 0. 988, precision of 0. 0. 882, and F1 score of 0. 769. After analyzing Twitter data, the study revealed that mental health issues can be detected among the population and intervention can be administered early. While there were some limitations, such as participants' unwillingness to reveal Twitter data, the need for regular model calibration, and the term "depression" used, non-specifically in the participant's survey, the study succeeded.

The study entitled "Predicting Depression via Social Media" [17] the researchers focus on the use of social media, specifically Twitter, to identify and diagnose major depressive disorder in patients. The authors argue that many people suffer from depression but remain untreated or even diagnosed, which leads to the consideration of using social media as a possible means of early detection. Their goal is to establish a model that would predict the probability of one having depression before its actual occurrence by analyzing different behavioral parameters on social media posts. The approach includes crowdsourcing for ground truth data on depression, building feature vectors from time series data of behavioral measures, and training SVM classifiers. The results presented here demonstrate good performance, with the best-performing model achieving an average accuracy of approximately 70% and high accuracy for predicting depression. The study ends on a positive note that social media analysis has the potential to quantify depression and identify the behavioral markers of the condition. However, some limitations include the following: The study used self-reported data; there could be some biases in the data collected from social media; there is a need to conduct more studies to include the models for better predictive accuracy and generalizability.

In the paper with the title of “Forecasting Suicide Rates in India Using ARIMA Models”, [18] the authors are concerned with the task of explaining and predicting the suicide rate in India using data from the period of 1967-2019 to assist the government in preventing the issue. Their goal is to analyze suicide rate patterns and predict them in the future with the help of the right models. For the secondary data collection, they went to the National Crime Record Bureau, and for the predictions, they used the ARIMA model i.e., Autoregressive Integrated Moving Average. These results show that the optimized model is ARIMA (1,2,2) with a minimum AIC of 921. 13 and a log-likelihood of -456. 56 for total deaths, which shows a better capability in predicting future suicide rates. They also stated that the next 15-year suicide rates are expected to rise. Therefore, the study has a few incongruities including the analysis of historical data that may not reflect changes in the future social economic status affecting suicide rates.

In the paper titled "Analyzing Seasonal Variations in Suicide With Fourier Poisson Time-Series Regression: A Registry-Based Study From Norway, 1969–2007," [19] the authors are concerned with the problem of determining whether the seasonality in suicide rates has decreased over time in Norway. They fitted additive Fourier Poisson time-series regression models with monthly suicide data spanning 39 years, assessing both linear and nonlinear temporal trends. Performance measures include the Akaike Information Criterion (AIC) and analysis of variance (ANOVA). They found out that models incorporating a seasonal pattern performed better than those without, with significant evidence for a reduction in seasonality over the years included in the study. The best-performing model, which had an AIC of 3,050.5, indicated a change point in seasonality, demonstrating that the seasonality of suicides has diminished. They concluded that the Fourier Poisson time-series regression model is proficient for studying seasonal phenomena. Some limitations of the paper include possible bias because of data aggregation

In the paper titled “ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks” [20] the authors’ aim to compare the performance of ARIMA models and machine learning in time series forecasting in the context of data driven networks. The Authors also reviewed the literature to assess the performance of the ARIMA models against the machine learning ones. They reviewed several works that compare ARIMA with other methods based on machine learning and deep learning for forecasting. The results suggest that the deep learning model is more accurate than the ARIMA model and has a 21% lower MSE compared to the ARIMA model. They stated that the choice of the forecasting method should depend on the complexity of the system represented by the data, underlining the need to include the traffic features on the network scale and the characteristics of the specific location for better results. However, the paper had some limitations including the sensitivity of the results to the quality of input data and the need to conduct more studies to establish the applicability of these findings in other forecasting conditions.

In the paper titled “A Machine Learning-Based Approach to Forecasting Alcoholic Relapses,” [21] the researchers sought to create machine learning models that will be used in predicting relapses in patients receiving alcoholism treatment. They aimed to contribute to the discussion on the lack of appropriate tools to predict the effectiveness of addiction treatment. The goal was to evaluate the predictive accuracy of decision trees, random forests, support vector machines, and K-nearest neighbor regressors relative to baseline predictors for relapses. They used patient data to train the models, and assessed the models’ accuracy using RMSE. The findings also depicted that the random forest model was more accurate than the other algorithms with the test RMSE of 12.124 and a test variance of 0. 841. The authors stated that the results of the study showed that machine learning models could predict the relapse of alcohol abuse but the results are not generalizable to other settings because of the data set used. Therefore, future research should be conducted in this area to improve the generalization of these models across various contexts.

III. Proposed Methodology

Numerous algorithms were used, and a research was done on each algorithm before training the models. The following diagram represents the steps the datasets went through to get the results.

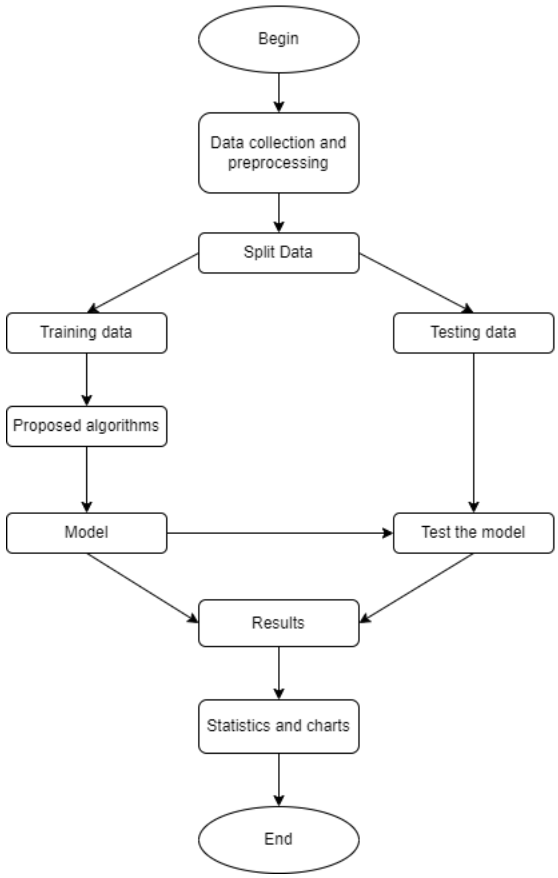


Fig. 2. Prediction and Forecasting Process

A. *Datasets Descriptions*

The datasets consist of 10 features and include multiple records spanning several years. A detailed description of the features can be found below.

- **Entity**: A categorical feature representing the name of the country.
- **Code**: A categorical feature representing the country code.
- **Year**: This is a numerical feature representing the year of the data.
- **Schizophrenia disorders** : A numerical feature representing the prevalence of schizophrenia disorders in the population, expressed as a percentage.
- **Depressive disorders** : A numerical feature representing the prevalence of depressive disorders in the population, expressed as a percentage.
- **Anxiety disorders** : A numerical feature representing the prevalence of anxiety disorders in the population, expressed as a percentage.
- **Bipolar disorders** : A numerical feature representing the prevalence of bipolar disorders in the population, expressed as a percentage.
- **Eating disorders** : A numerical feature representing the prevalence of eating disorders in the population, expressed as a percentage.
- **Drug use disorders** : A numerical feature representing the prevalence of drug use disorders in the population, expressed as a percentage.
- **Alcohol use disorders** : A numerical feature representing the prevalence of alcohol use disorders in the population, expressed as a percentage.

TABLE I
FEATURES OF THE DATASET

Feature	Type	Values
Entity	Categorical	The name of the country
Code	Categorical	The country code
Year	Numerical	The year of the data
Schizophrenia disorders	Numerical	The prevalence of schizophrenia disorders
Depressive disorders	Numerical	The prevalence of depressive disorders
Anxiety disorders	Numerical	The prevalence of anxiety disorders
Bipolar disorders	Numerical	The prevalence of bipolar disorders
Eating disorders	Numerical	The prevalence of eating disorders
Drug use disorders	Numerical	The prevalence of drug use disorders
Alcohol use disorders	Numerical	The prevalence of alcohol use disorders

B. *Pre-processing*

1) Data Loading and Cleaning: First things first, all datasets were sequentially fully loaded, except for the 'Mental health Depression disorder Data', which was loaded only up to the 6467th row due to incomplete data. Additionally, the 'Code' column was found to be redundant and was swiftly dropped. Moreover, the 'country' column was renamed to 'Entity' to maintain data conformity across all datasets.

2) Data Transformation: Since this paper focuses on forecasting mental disorders globally rather than on a country-by-country basis, yearly global averages were calculated for each disorder. This final form of the data represents the aggregation necessary for passing it on to the models for training.

3) Importance of Identifying Patterns: Identifying patterns in the data is of paramount importance in data analysis. Patterns can provide valuable insights into underlying trends, correlations, and anomalies. By systematically analyzing and visualizing the data, patterns emerge, revealing recurring relationships and behaviors. This process enables researchers to gain a deeper understanding of the phenomena under study, facilitating informed decision-making and predictive modeling. In our case, plotting all disorders over the years after preprocessing allows us to identify trends and patterns that may exist in the global prevalence of mental disorders.

C. *Used Algorithms*

The mentioned datasets were analyzed using 8 different Machine Learning and Deep Learning algorithms, including

Holt-Winters Exponential Smoothing, Moving Average, AutoRegressive Integrated Moving Average (ARIMA), Auto Regression, Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), Feedforward Neural Network (FNN), and Long Short-Term Memory (LSTM). The results, along with charts and detailed discussions, can be found later in the paper.

1) **Holt-Winters Exponential Smoothing:**

Holt-Winters' seasonal method extends Holt's method to capture seasonality. It consists of a forecast equation and three smoothing equations: one for the level (ℓ_t), one for the trend (b_t), and one for the seasonal component (s_t), with corresponding smoothing parameters α, β^*, and γ. The frequency of the seasonality is denoted by m, which represents the number of seasons in a year (e.g., $m = 4$ for quarterly data, $m = 12$ for monthly data).

- **Additive Method**: Used when the fluctuations in the seasonal patterns are relatively constant over the series. The analysis of the seasonal variation is done in the absolute values and the series is then deseasonalized by subtracting the seasonal part. That means, for each year, the sum of the seasonal part of the component will amount to about zero.

Holt-Winters' Additive Method

The component form for the additive method is:

$$\hat{y}_{t+h|t} = \ell_t + hb_t + s_{t+h-m(k+1)}$$

$$\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$$

$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$$

$$s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}$$

where k is the integer part of $(h-1)/m$, ensuring that the estimates of the seasonal indices used for forecasting come from the final year of the sample. The level equation shows a weighted average between the seasonally adjusted observation $(y_t - s_{t-m})$ and the non-seasonal forecast $(\ell_{t-1} + b_{t-1})$ for time t. The trend equation is identical to Holt's linear method. The seasonal equation shows a weighted average between the current seasonal index, $(y_t - \ell_{t-1} - b_{t-1})$, and the seasonal index of the same season last year (i.e., m time periods ago).

The equation for the seasonal component can also be expressed as:

$$s_t = \gamma^*(y_t - \ell_t) + (1-\gamma^*)s_{t-m}$$

Substituting ℓ_t from the smoothing equation for the level, we get:

$$s_t = \gamma^*(1-\alpha)(y_t - \ell_{t-1} - b_{t-1}) + [1 - \gamma^*(1-\alpha)]s_{t-m}$$

which is identical to the smoothing equation for the seasonal component, with $\gamma = \gamma^*(1-\alpha)$. The usual parameter restriction is $0 \leq \gamma^* \leq 1$, translating to $0 \leq \gamma \leq 1-\alpha$. [22]

2) **Moving Average:**

The simple moving average (SMA) is a basic technical indicator that is arrived at by adding up the figures in a given range and then dividing by the number of periods. SMA is utilized in signalling the traders on when to enter or exit in a certain market. An SMA is lag-based for the reason that it depends on the past price data for a specific time period. High, low, open, close: It can be computed for different types of prices.

In trading financial securities in the stock market, the SMA indicator is used by analysts and investors to decide when to buy or when to sell. Support and resistance prices are used in the SMA to provide a signal on when to open an entry or an exit position.

When computing the SMA, the trader has to arrive at it through the following formula: Add up prices over the specified period and divide the sum by the total of the periods. The data is then transferred on a graph called the pictogram.

The formula for Simple Moving Average is written as follows:

$$\text{SMA} = \frac{A_1 + A_2 + \ldots + A_n}{n}$$

Where:

- A is the average in period n
- n is the number of periods [23]

3) **Autoregressive Integrated Moving Average**

An autoregressive integrated moving-average (ARIMA) is a type of statistical model analysis in which values in a data series are described as a continuous function of past values measured at several intervals of time. If a statistical model attempts to predict future values with current as well as past value then its called autoregressive model. For instance, an ARIMA model could aim at trying to estimate future stock prices for a given firm based on its historical performance or future earnings of a company given other periods' results.

An ARIMA model can be understood by outlining each of its components as follows:

- **Autoregression (AR)**: Refers to a model that shows a changing variable that regresses on its own lagged, or prior, values.
- **Integrated (I)**: Represents the differencing of raw observations to allow the time series to become stationary (i.e., data values are replaced by the difference between the data values and the previous values).
- **Moving Average (MA)**: Incorporates the dependency between an observation and a residual error

from a moving average model applied to lagged observations.

Each component in ARIMA functions as a parameter with a standard notation. For ARIMA models, a standard notation would be ARIMA(p, d, q), where integer values substitute for the parameters to indicate the type of ARIMA model used. The parameters can be defined as:

- p: The number of lag observations in the model, also known as the lag order.
- d: The number of times the raw observations are differenced; also known as the degree of differencing.
- q: The size of the moving average window, also known as the order of the moving average. [24]

4) **Autoregressive Model**

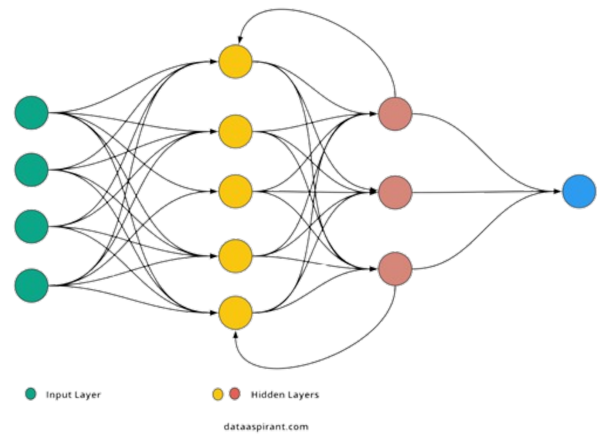
Autoregressive models are based on the idea that lagged variables have some impact on current values, which is why this statistical tool is widely used for the analysis of nature, economics, etc. , which are characterized by variable changes over time. Multiple regression models predict a variable based on a set of variables while autoregressive models predict a variable on the basis of a set of past values of the variable.

An AR(1) autoregressive process means that the current value is dependent on the previous value while an AR(2) means the value of the current variable depends on the two previous values. AR (0) process is employed for white noise and it does not have any interrelationship between the terms. and there are different ways of estimating the coefficients involved in such computations, for instance the least square method. [25]

5) **Recurrent Neural Networks (RNNs)**

Recurrent Neural Networks (RNNs) are one type of neural network employed for sequential data modeling. Deriving from the feedforward networks, RNNs work like the human brain and contain the capability to predict the occurrence of sequential data in ways that most other algorithms cannot.

RNNs are complex and powerful because of their ability to remember past inputs while working on current inputs. Hence, internal memory enables RNNs to have essence of the details in the input sequences, thus making their forecast of the future sequences extremely accurate. Used mostly in sequential data such as time series, text, speech, financial records, voices, sounds, videos and chronological climatic patterns, recurrent neural networks offer deeper analysis on the sequential nature of the data as well as context compared to other kinds of algorithms.[26]



[27]

Fig. 3. Illustration of Recurrent-Neural-Network[27]

6) **Convolutional Neural Networks (CNNs)**

Convolutional Neural Networks (CNNs) are deep neural networks primarily used for analyzing visual imagery. CNNs use convolution, a mathematical operation on two functions that produces a third function, showing how one shape is modified by the other. It typically has three main layers: convolutional, pooling, and fully connected layers.

Convolutional Layer

The convolutional layer performs a dot product between the input data and learnable parameters known as kernels. These kernels slide across the input data, creating activation maps that highlight specific features.

Pooling Layer

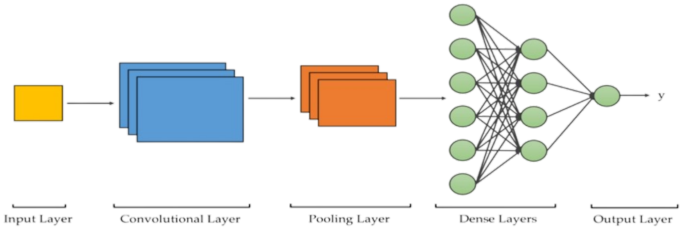
The pooling layer reduces the spatial size of the activation maps, decreasing computational load and helping to prevent overfitting. Max pooling, which selects the maximum value from a defined region, is the most common pooling operation.

Fully Connected Layer

In the fully connected layer, each neuron is connected to every neuron in the preceding layer. This layer maps the learned features to the final output.

Non-Linearity Layers

Non-linearity layers such as ReLU, sigmoid, and tanh are applied after the convolutional layer to introduce non-linearity into the model.
[28]



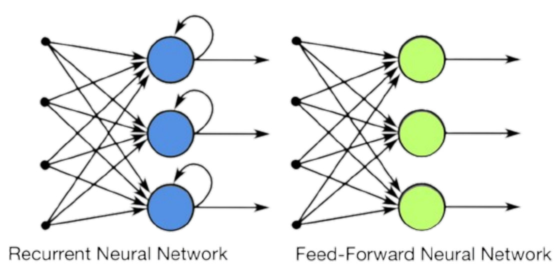
[29]

Fig. 4. Illustration of Convolutional-Neural-Network[29]

7) **Feedforward Neural Networks (FNNs)**
Feedforward Neural Networks (FNNs) are the relatively simple form of ANN where the structure is aimed to be only unidirectional one where the information flows from the input node via the hidden node planes to the output node plane. FNNs have no cycle or loops and are less complicated than RNNs and CNNs. It consist of three types of layers: To achieve the desired output, two hidden layers or phases for input and output. Each layer element is a neuron connected to all neurons in the next layer through weights.
FNNs operate in two phases: feedforward and backpropagation.

Feedforward Phase: Input data propagates forward through the network. Each hidden layer calculates the weighted sum of inputs, applies an activation function, and passes it on until the output layer produces a prediction.

Backpropagation Phase: The error between predicted and actual outputs is calculated and propagated back through the network. Weights are adjusted to minimize this error using gradient descent.[30] [28]



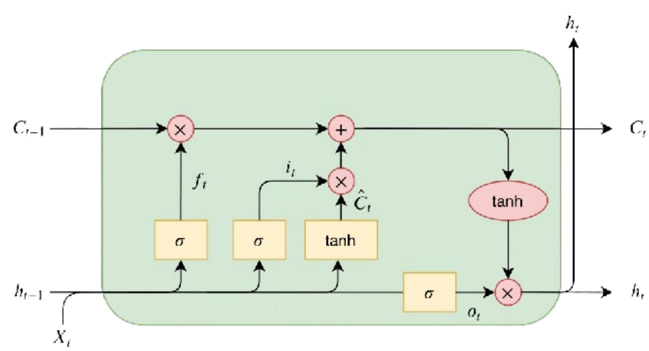
[29]

Fig. 5. Illustration of Recurrent-Neural-Network vs Feedforward-Neural-Network[29]

8) **Long Short-Term Memory (LSTM)**
Long Short-Term Memory (LSTM) is a specific kind of Recurrent Neural Network (RNN) that has a specific form of connection within the network to be able to capture the long-term dependencies of sequential data. One of the key benefits of LSTMs is that they are particularly useful when it comes to the data involving time series, texts, and speech. They incorporate memory cells and gates to regulate the incoming or out going information when needed so that it does not retentive or discard the information hence do not experience the vanishing gradient problem of recurrent RNNs. LSTM is highly beneficial in fields such as NLP, Speech to Text conversion, and Time series analysis.

An LSTM network is a combination of LSTM cells and within these cells, there are three main aspects: namely the input gate, the output gate, and the forget gate. These gates enable the LSTM to either give up or maintain some information from the previous time steps of the input data and provide long-term memory.

The last intermediate is LSTM cell and it also has incorporated one of the memory cell that contains information for previous time step in its computation to give the current output. The little square symbol on the right of the equation represents the output of each LSTM cell that is fed to the next cell in the network so as to analyze sequential data over time steps for multiple times.
[31]



[32]

Fig. 6. Illustration of Long Short-Term Memory (LSTM)[32]

Seeing as the LSTM is conceptually a more powerful iteration of RNNs that compensates for their shhortcomings, it was decided that it would be more efficient to implement LSTM only.

D. *Performance Metrics*

Mean Squared Error (MSE), Mean Absolute Deviation (MAD), Root Mean Squared Error (RMSE), and Symmetric Mean Absolute Percentage Error (SMAPE) are common metrics used to evaluate the performance of predictive models.

Mean Squared Error (MSE):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (1)$$

Mean Absolute Deviation (MAD):

$$MAD = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (2)$$

Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (3)$$

Symmetric Mean Absolute Percentage Error (SMAPE):

$$SMAPE = \frac{100\%}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{(|y_i| + |\hat{y}_i|)/2} \quad (4)$$

IV. Results and Analysis

The following results were collected from the output of one deep learning model [*LSTM*] and four statistical time-series forecasting models [*HWES*, *ARIMA*, *MA*, *AR*] after undergoing training on three different data-sets that capture a variety of mental diseases [*Schizophrenia*, *Depressive*, *Anxiety*, *Bipolar*, *Eating*, *Drug use*, *Alcohol use*] (%) Disorders. These were the MSE results upon forecasting the validation sets.

Shizophrenia: A chronic and severe mental disorder that affects how a person thinks, feels, and behaves.

TABLE II
Mean Squared Error of Global Schizophrenia Disorders

Model	Data-set 1	Data-set 2	Data-set 3
LSTM	0.00399603	0.12610709	0.00673738
MA	0.00000092	0.00000300	0.00000300
AR	0.00001354	0.00640246	0.00624262
HWES	0.00000004	0.00000002	0.00000002
ARIMA	0.00000001	0.00000007	0.00000007
Best Model	ARIMA	HWES	HWES

Depression: A mood disorder characterized by persistent sadness and feelings of worthlessness.

TABLE III
Mean Squared Error of Global Depressive Disorders

Model	Data-set 1	Data-set 2	Data-set 3
LSTM	0.00391188	0.07616258	0.00003644
MA	0.00109821	0.00103472	0.00103472
AR	0.00796262	0.00660729	0.00660737
HWES	0.00128554	0.00006987	0.00006987
ARIMA	0.00003738	0.000260353	0.00026028
Best Model	ARIMA	HWES	LSTM

Anxiety: A mental health condition causing excessive worry and fear.

TABLE IV
Mean Squared Error of Global Anxiety Disorders

Model	Data-set 1	Data-set 2	Data-set 3
LSTM	0.08017553	0.31049339	0.00056277
MA	0.00135751	0.00000502	0.00000502
AR	0.01772680	0.00000296	0.00000296
HWES	0.00005326	0.00014699	0.00014699
ARIMA	0.00021815	0.00000009	0.00000009
Best Model	HWES	ARIMA	ARIMA

Bipolar Disorder: A mental illness causing extreme mood swings from mania to depression.

TABLE V
Mean Squared Error of Global Bipolar Disorders

Model	Data-set 1	Data-set 2	Data-set 3
LSTM	0.00012783	0.00340332	0.00457757
MA	0.00000001	0.00000350	0.00000350
AR	0.00000553	0.00002443	0.00002372
HWES	0.00000036	0.000000003	0.000000003
ARIMA	0.000000043	0.000000009	0.000000009
Best Model	MA	HWES	HWES

Eating Disorder: An unhealthy relationship with food that can lead to obsession with weight or body image.

TABLE VI
MEAN SQUARED ERROR OF GLOBAL EATING DISORDERS

Model	Data-set 1	Data-set 2	Data-set 3
LSTM	0.01572257	0.01471523	0.01141222
MA	0.00008329	0.00015106	0.00015106
AR	0.00010693	0.00000337	0.00000336
HWES	0.00000095	0.00000239	0.00000239
ARIMA	0.00000397	0.00000043	0.00000042
Best Model	ARIMA	HWES	LSTM

Drug Use: The repeated use of a substance (other than prescribed medication) that alters mood or behavior.

TABLE VII
MEAN SQUARED ERROR OF GLOBAL DRUG USE DISORDERS

Model	Data-set 1	Data-set 2	Data-set 3
LSTM	NA	2.23787543	0.50067806
MA	NA	0.00041606	0.00041606
AR	NA	0.00001749	0.00001748
HWES	NA	0.00000226	0.00000226
ARIMA	NA	0.00000090	0.00000090
Best Model	NA	ARIMA	ARIMA

Alcoholism: A chronic and severe condition characterized by excessive and uncontrolled drinking.

TABLE VIII
MEAN SQUARED ERROR OF GLOBAL ALCOHOL USE DISORDERS

Model	Data-set 1	Data-set 2	Data-set 3
LSTM	NA	0.06036105	0.61716326
MA	NA	0.00031545	0.00031545
AR	NA	217.02851894	47419.75247827
HWES	NA	0.00028950	0.00028950
ARIMA	NA	0.00022096	0.00022098
Best Model	NA	ARIMA	ARIMA

V. CONCLUSION

To summarize, global mental health disorders and issues are a serious and rising challenge that affects people from all countries and demographics. The prevalence of illnesses such as schizophrenia, depressive disorders, anxiety disorders, bipolar disorders, eating disorders, drug use disorders, and alcohol use disorders emphasizes the critical need for comprehensive mental health policies and services. Anticipating the prevalence of these illnesses in the community is critical for a variety of reasons.

After conducting our in-depth training and evaluating, it's been noted that highest resulting models in our forecast were **HWES:** ***Holts Winter Exponential Smoothing*** and **ARIMA:** ***Auto-Regressive Integerated Moving Average*** among all out data-sets and disorders. Based on these models, we projected the prevalence of mental health issues over the following ten years, beginning a year after the last year in our dataset.

Shizophrenia:

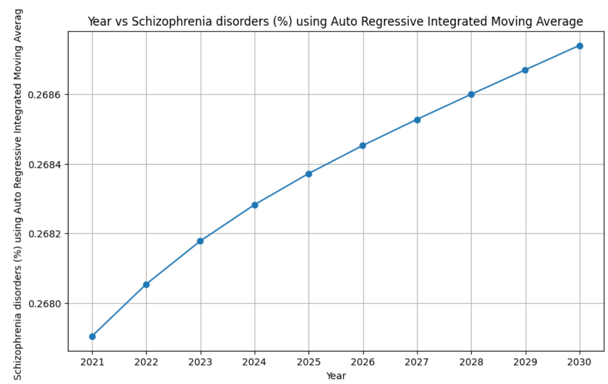


Fig. 7. Forecast using Dataset 1 & ARIMA

Anxiety Disorders:

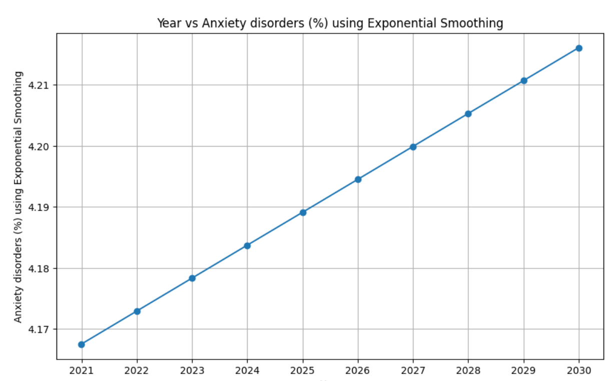


Fig. 8. Forecast using Dataset 1 & HWES

Bipolar Disorder:

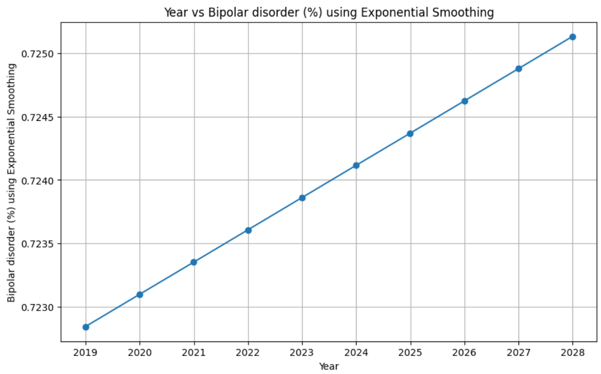


Fig. 9. Forecast using Dataset 2 & ARIMA

Eating Disorders:

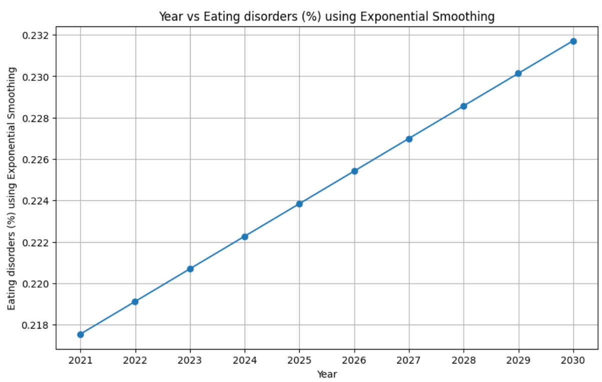


Fig. 10. Forecast using Dataset 1 & HWES

Drug-Use Addiction:

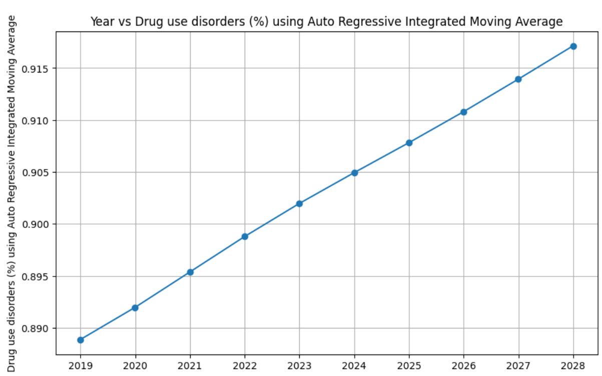


Fig. 11. Forecast using Dataset 2 & ARIMA

Depression:

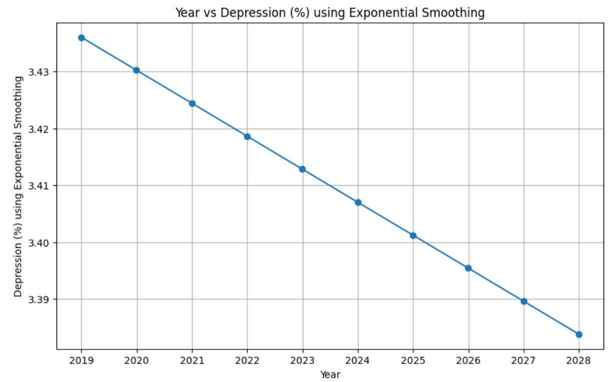


Fig. 12. Forecast using Dataset 2 & HWES

Alcoholism:

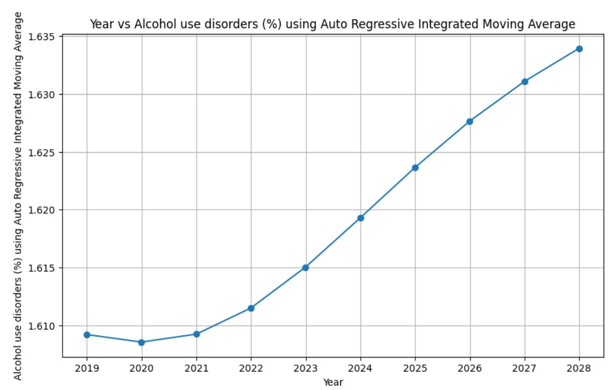


Fig. 13. Forecast using Dataset 3 & ARIMA

In our final analysis its evident that **HWES** is however, the best achieving model with the lowest errors and the most accurate trend projections.

TABLE IX
FORECAST EVALUATION METRICS FOR HWES MODEL

Metric	**Value**
Mean Squared Error (MSE)	0.000000003524272
Mean Absolute Deviation Error (MAE)	0.000053016918021
Root Mean Squared Error (RMSE)	0.000059365581812
Symmetric Mean Absolute Percentage Error (SMAPE)	0.007341066351758

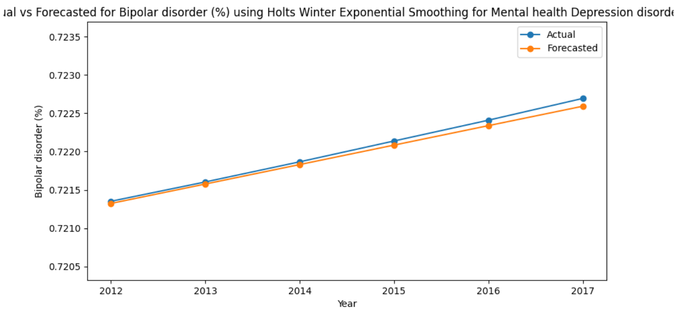


Fig. 14. Actual vs Forecasted Using HWES

VI. Acknowledgment

To begin with, we would like to establish our immeasurable gratitude for the many passionate educators and staff at Misr International University and the computer science faculty in particular for striving endlessly in the pursuit of spreading knowledge and success. We are immensely appreciative of Prof. Mohamed Shebl El Komy University President, Prof. Ayman Nabil dean faculty of Computer Science, Prof. Abdelnasser Zaied Vice Dean of Student Affairs and Professor of Computer Engineering for permitting us to have a fruitful and beneficial academic journey at this university. And last but definitely not the least, would like to commend with utmost sincerity Dr. Diaa AbdelMoneim associate professor in information systems and Eng. Donia Mostafa teaching assistant for their mentorship and unyielding and unrelenting efforts to provide assistance in our most treasured project yet.

References

[1] W. P. Castelli, R. D. Abbott, and P. M. McNamara, "Summary estimates of cholesterol used to predict coronary heart disease." *Circulation*, vol. 67, no. 4, pp. 730–734, 1983.

[2] N. Tateyama, R. Yokomura, Y. Ban, S. Warisawa, and R. Fukui, "Planning the future in a longer perspective: Effects of a one-week forecast of mental health," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 8, no. 1, pp. 1–20, 2024, accessed 18 May 2024. [Online]. Available: https://doi.org/10.1145/3643538

[3] E. Hossain, T. Rahman, and S. Islam, "Forecasting mental stress using machine learning algorithms," *ResearchGate*, 2022, accessed 18 May 2024. [Online]. Available: https://www.researchgate.net/profile/Elias-Hossain-2/publication/360087966_Forecasting_Mental_Stress_Using_Machine_Learning_Algorithms/links/626188408e6d637bd1f248da/Forecasting-Mental-Stress-Using-Machine-Learning-Algorithms.pdf

[4] R. Alanazi and S. Alanazi, "A novel approach to forecasting the mental well-being using machine learning," *Alexandria Engineering Journal*, vol. 84, pp. 175–183, 2023, accessed 18 May 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1110016823009717

[5] J. Lake, *The Future of Mental Health Care: Trends and Forecast*. Springer, 2019, pp. 235–241, accessed 18 May 2024. [Online]. Available: https://doi.org/10.1007/978-3-030-15285-7_10

[6] Y. Suhara and et al., "Deepmood," in *DSpace@MIT (Massachusetts Institute of Technology)*, 2017, accessed 2 Sept. 2023. [Online]. Available: https://doi.org/10.1145/3038912.3052676

[7] V. Dang, "Anxiety and affective forecasting," Master's thesis, Texas A&M University, 2023, accessed 18 May 2024. [Online]. Available: https://hdl.handle.net/1969.1/199898

[8] V. Sukmak, J. Thongkam, and J. Leejongpermpoon, "Time series forecasting in anxiety disorders of outpatient visits using data mining," *KKU Research Journal*, vol. 20, no. 2, pp. 241–253, 2015, accessed 18 May 2024. [Online]. Available: http://resjournal.kku.ac.th

[9] S. P. Humaira, I. Nursuprianah, and Darwan, "Forecasting of the number of schizophrenia disorder by using the box-jenkins of time series analysis," *Journal of Robotics and Control (JRC)*, vol. 1, no. 6, pp. 213–219, 2020, accessed 18 May 2024. [Online]. Available: http://journal.umy.ac.id/index.php/jrc

[10] A. T. Jebb, L. Tay, W. Wang, and Q. Huang, "Time series analysis for psychological research: Examining and forecasting change," *Frontiers in Psychology*, vol. 6, 2015, accessed 18 May 2024. [Online]. Available: https://doi.org/10.3389/fpsyg.2015.00727

[11] D. C. Mathersul and A. M. Ruscio, "Forecasting the future, remembering the past: Misrepresentations of daily emotional experience in generalized anxiety disorder and major depressive disorder," *Cognitive Therapy and Research*, vol. 43, pp. 1068–1082, 2019, accessed 18 May 2024. [Online]. Available: https://doi.org/10.1007/s10608-019-10048-5

[12] A. Tommasel, A. Diaz-Pace, J. M. Rodriguez, and D. Godoy, "Capturing social media expressions during the covid-19 pandemic in argentina and forecasting mental health and emotions," *arXiv preprint arXiv:2101.04540*, 2021.

[13] S. Lin, Y. Wu, and Y. Fang, "Comparison of regression and machine learning methods in depression forecasting among home-based elderly chinese: A community based study," *Frontiers in psychiatry*, vol. 12, p. 764806, 2022.

[14] P. K. Swain, M. R. Tripathy, S. Priyadarshini, and

S. K. Acharya, "Forecasting suicide rates in india: An empirical exposition," *PLoS one*, vol. 16, no. 7, p. e0255342, 2021.
[15] M. E. Ghaffari, A. Ghaleiha, Z. Taslimi, F. Sarvi, P. Amini, M. Sadeghifar, and S. Yazdi-Ravandi, "Forecasting schizophrenia incidence frequencies using time series approach," *International Clinical Neuroscience Journal*, vol. 4, no. 4, p. 152, 2017.
[16] A. G. Reece, A. J. Reagan, K. L. Lix, P. S. Dodds, C. M. Danforth, and E. J. Langer, "Forecasting the onset and course of mental illness with twitter data," *Scientific reports*, vol. 7, no. 1, p. 13006, 2017.
[17] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, "Predicting depression via social media," in *Proceedings of the international AAAI conference on web and social media*, vol. 7, no. 1, 2013, pp. 128–137.
[18] P. Roy, B. B. Boruah, K. Barhoi, and K. K. Mahanta, "Suicide in india: A time series analysis," *Int. J. Mech. Eng.*, p. 6, 2021.
[19] J. G. Bramness, F. A. Walby, G. Morken, and J. Røislien, "Analyzing seasonal variations in suicide with fourier poisson time-series regression: a registry-based study from norway, 1969–2007," *American journal of epidemiology*, vol. 182, no. 3, pp. 244–254, 2015.
[20] V. I. Kontopoulou, A. D. Panagopoulos, I. Kakkos, and G. K. Matsopoulos, "A review of arima vs. machine learning approaches for time series forecasting in data driven networks," *Future Internet*, vol. 15, no. 8, p. 255, 2023.
[21] N. Katardjiev, S. McKeever, and A. Hamfelt, "A machine learning-based approach to forecasting alcoholic relapses," in *Proceedings of the ITISE 2019 6th International Conference on Time Series and Forecasting*, 2019.
[22] R. J. Hyndman and G. Athanasopoulos, "7.3 holt-winters' seasonal method | forecasting: Principles and practice," 2016, accessed 21 May 2024. [Online]. Available: https://otexts.com/fpp2/holt-winters.html
[23] Corporate Finance Institute, "Moving average," 2024, accessed 21 May 2024. [Online]. Available: https://corporatefinanceinstitute.com/resources/equities/moving-average/
[24] Investopedia, "Autoregressive integrated moving average (arima)," 2024, accessed 21 May 2024. [Online]. Available: https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp
[25] ——, "Autoregressive model," 2024, accessed 21 May 2024. [Online]. Available: https://www.investopedia.com/terms/a/autoregressive.asp
[26] Built In, "Recurrent neural networks and lstm," 2024, accessed 21 May 2024. [Online]. Available: https://builtin.com/data-science/recurrent-neural-networks-and-lstm
[27] Data Aspirant, "How recurrent neural network (rnn) works," 2024, accessed 21 May 2024. [Online]. Available: https://dataaspirant.com/how-recurrent-neural-network-rnn-works/
[28] Towards Data Science, "Convolutional neural networks explained," 2024, accessed 21 May 2024. [Online]. Available: https://towardsdatascience.com/convolutional-neural-networks-explained-9cc5188c4939
[29] Medium, "An overview on convolutional neural networks," 2024, accessed 21 May 2024. [Online]. Available: https://medium.com/swlh/an-overview-on-convolutional-neural-networks-ea48e76fb186
[30] DeepAI, "Feed forward neural network," 2024, accessed 21 May 2024. [Online]. Available: https://deepai.org/machine-learning-glossary-and-terms/feed-forward-neural-network
[31] Simplilearn, "Understanding lstm networks," 2024, accessed 21 May 2024. [Online]. Available: https://www.simplilearn.com/tutorials/artificial-intelligence-tutorial/lstm
[32] M. Saba, "Long short-term memory (lstm)," 2024, accessed 21 May 2024. [Online]. Available: https://medium.com/@saba99/long-short-term-memory-lstm-fffc5eaebfdc