IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Detection And Prediction Of Mental Health Illness Using Machine Learning And Deep Learning Techniques: A Survey

Prof Jayashree M Kudari
Dept of BCA, Associate Professor
School CS&IT
Jain (Deemed to Be University)

Dr Srikanth V
Associate Professor, Dept of MCA
School CS&IT
Jain (Deemed to Be University)

Abstract:

In modern era psychological illnesses have grown quite common and depression remains one of the most prevalent forms of mental illness. Based on WHO statistics, depression is the second greatest cause of illness burden worldwide. The reality for those with mental diseases is significantly worse, especially in emerging and undeveloped nations because healthcare assets are brief. Depression is a form of psychological condition where an individual experiences constant despondency, demotivation, mood fluctuations and lack of interest in everyday mental, physical and social endeavors resulting in emotional harm and bodily modifications in the patient's physical condition. It has a particular impact on a person's learning ability, produces mood changes and frequently impairs job productivity. This paper deals with the various, techniques used by various researchers to predicts the different kind of depression.

Keywords: CNN, ANN, SVM, KN, Mental health, Depression

I. Introduction:

As a result of the need to develop efficient approaches to depression issues, deep learning methods have been integrated into medical facilities for the identification and likely estimation of treatment results for mental health diseases. Data-driven algorithms like CNN and LSTM live on data since there are so many free datasets accessible, they perform better when identifying sad and non-depressed comments. RNNs are favored because they may be provided already trained embedding’s from outside sources. It may be used for processing text because it can be built quickly. The main objective behind selecting a hybrid deep learning approach is its ability to improve the performance or accuracy when compared to a single ML or DL algorithm. It can handle large scale complex data with high dimensional features and also hybrid DL algorithms can handle computational complexities.

II. Literature Review

1. Prediction and Detection of Depression

A. Predicting Depression using Machine Learning Techniques

1. Data Acquisition

A survey of Bangladeshi individuals of various ages was performed to gather data. A total of 604 people's replies makes up the dataset. To gauge each participant's real state of depression during the survey, the study used a revised version of the BDC with 25 questions. David Burns created the BDC, one of the most used rating scales for assessing depression. If a person's BDC score as a whole is more than 10, that individual is deemed sad; if not, they are not. [1]

2. Data description

One target variable and thirty predictor variables make up the dataset that was acquired from the survey. By administering the Burns Depression Checklist (BDC) to each participant, the target variable was created. [1]

3. Data analysis

397 out of the 604 participants in the total sample were judged to be depressive, while 34.27 percent were not. The survey's respondents have a 65.73% depression prevalence. [1]

4. Feature Selection Techniques

The selection of characteristics for a machine learning model should only include those that are essential. The performance of the model may be harmed by choosing irrelevant features. Feature selection aids in the elimination of redundant and pointless features that don't improve the performance of the model. [1]

5. Select K-Best Features

The study demonstrates that participants who are working or running their own businesses are more likely to experience depression than participants who are not. Participants who live in villages had a lower risk of developing depression. [1]

$$x^2 = \sum_{i=1}^{n} \frac{(OF_i - EF_i)}{EF_i}$$

Where EF*i* is the frequency that is anticipated for the i-th value of the feature F, and OF*i* is the frequency that is observed for the i-th value of the frequency F in this case.

Other univariate statistical tests, such as the analysis of variance (ANOVA) p-value and manual information approaches, can be employed in addition to the x^2 test to choose the K-best features.

6. Minimum Redundancy and Maximum Relevance(mRMR)

A feature selection approach with the names Minimum Redundancy and Maximum Relevance was proposed by the
author (mRMR). When ranking features, this algorithm takes into account both relevance and redundancy. [1]

7. Synthetic minority oversampling technique (SMOTE)

Synthetic samples of the minority class are created by working in feature space in order to improve the minority class's predictive accuracy. The K-nearest minority class sample neighbors as well as the line immediately adjacent to each minority class sample are where the synthetic samples are first introduced. [1]

8. Machine Learning Algorithms

A. KNN

KNN is a non-parametric distance-based method that can be used for classification and regression problems. The Minkowski Distance function can be used to explain the K value, which is the major classifier factor. [1]

$$distance(E,F) = \left(\sum_{i=1}^{n}(|e_i - f_i|)^q\right)^{1/q}$$

Here, q is the Minkowski distance. [1]

B. Adaptive Boosting classifier

An ensemble learning technique called AdaBoost classifier makes use of a number of weak classifiers. A poor classifier picks up on the mistakes made by its stronger counterpart. [1]

$$a = \frac{1}{2} ln\left(\frac{1-\varepsilon}{\varepsilon}\right)$$

Here, a is used to alter the weights of the dataset's samples, which produces a new dataset. Later, the next weak classifier is built using this fresh dataset. Finally, the vast majority of weak classifiers choose the sample's class label.

C. Gradient Boosting

Gradient Boosting (GB) classifier ensembles weak models to sequentially create new models, using the gradient descent method to compute the loss function. [1]

Using the linear combination of the weak models, g: Rn R, as the following equation, GB creates a function, f: Rn R:

$$f(x) = \sum_{j=1}^{M} \omega_i g_i(x{:}\,\theta_i)$$

In this instance, xERn represents the input vector, while M wiER shows the weight of the weak models. the function produced by minimizing the loss function iteratively by selecting a weak learner's weight wi and parameter Oi. [1]

The study considerably fills gaps in earlier research by predicting depression in persons of various ages, professions, and socioeconomic backgrounds. This study's accuracy score of 92.56% is respectable and encouraging. [1]

Six different machine learning classifiers were utilized in this study to predict depression in 604 participants. The accuracy of the AdaBoost classifier using the Select K-Best feature selection method was 92.56%. Although no biological markers were included in the dataset for depression prediction, this study used BDC as the gold standard for diagnosing depression. This research can be expanded in the future to determine a person's level of depression. [1]

B. A Review on Detection of Depression of Existing Literature

The authors of [3] use the Bayes Net (BN) classifier, logistic regression (LR), multilayer perceptron (MLP), sequential minimal optimization support vector machine (SMO), and decision table to address the prediction of depression in older individuals (DT). On a dataset provided by the Urban Health and Training Center (UHTC) Bagbazar Kolkata, they apply the WEKA Machine Learning Model. According to BN, accuracy is 86.67%, precision is 0.8, ROC Area is 0.96, and RMSE is 0.32. RMSE is 0.46, Precision is 0.79, ROC Area is 0.85, and Accuracy is 78.33% according to LR. MLP provides 85% accuracy, 0.85% precision, 0.94 ROC area, and 0.33 RMSE. SMO provides 88.33% accuracy, 0.88 precision, 0.88 ROC area, and 0.34 RMSE. 80% accuracy, 0.80 precision, 0.86 ROC area, and 0.38 RMSE are provided by DT. As a result, the Bayes Net classifier, which they investigated, provides the best result, with a maximum accuracy of 95% and a maximum precision of 0.95. [2]

The authors of [4] use ML approaches and ten classifiers—the BN, NB, Log, MLP, KS, RS, J48, and RF—to address the prediction of anxiety and depression in older individuals. A dataset of 510 elderly patients is used to train the classifiers. The KAR Medical College and Hospital in Kolkata is the source of the dataset. Additionally, they employ ten-fold cross-validation. And RF, which has a maximum accuracy of 89%, provides the best outcome, according to the authors. [2]

The NANA toolbox, a tool that makes use of ML approaches, is highlighted by the authors in [5]. The NANA toolbox gathers the required data from elderly individuals in their homes without the practitioner's knowledge. In [3], the author focuses on the application of ML approaches to the data gathered by the NANA toolkit from the homes of older people in order to produce an algorithm to predict the symptoms of depression in older persons. The authors employ the LASSO method in conjunction with the logistic regression classifier to conduct regularization. [2]

The application of ML approaches, integrating clinical and imaging characteristics, and predicting depression in older adults are highlighted by the authors in [6]. The goal is to assess the precise model (prediction models) for depression and the appropriate care, medication supplied to patients using machine learning techniques with model inputs (multi-modal imaging, non- imaging the brain of human). By using medicated post-recruitment, 33 patients and 35 older, non-depressed individuals were each recruited separately. Using multi-modal magnetic resonance imaging pretreatment, their brain features as well as demographic and cognitive scores were obtained. The authors report a maximum accuracy of 87.27% and a treatment response of 89.47% when evaluating the prediction models for depression in older persons using the alternating decision tree (ADT) machine learning method [4]. [2]

Before using ML to predict sadness, the authors of [7] use the ILIOU data preprocessing approach. Using the tenfold cross validation method, the effectiveness of the ILIOU data preprocessing method and the principal component analysis preprocessing method was assessed. The seven machine learning classification algorithms were the nearest-neighbor classifier (IB1), random forest, multilayer perceptron (MLP), support vector machine (SMO), JRIP, and fuzzy logic (FURIA). ILIOU preprocessing can be used to predict various forms of depression. Using SMO classifier from their original data, the maximum precision and maximum recall were attained at 91.2%. J48 classifier was used to obtain maximum precision and maximum recall of 96.2% on PCA preprocessed data. Using the MLP classifier from the ILIOU Method, 100% recall and 100% precision were attained. ILIOU thus provides the greatest results with the highest recall and highest precision. [2]

The authors of [8] discuss the use of ML approaches to identify major depressive disorder (MMD) using EEG-based functional connectivity. The authors of the study experiment with the SVM, LR, and NB classifiers. The Hospital University Sains Malaysia has made the dataset available (HUSM). These three classifiers were chosen because it is simple to train them on tiny datasets. Different accuracy, sensitivity, specificity, and f-M metrics are provided by these classifiers. LR provides accuracy of 91.7%, sensitivity of 86.66%, specificity of 96.6%, and f-m of 0.90, while SVM provides accuracy of 98%, sensitivity of 99.9%, and specificity of 95%. 93.6% accuracy, 100% sensitivity, 87.9% specificity, and 0.95 f-m are provided by NB. The authors draw the conclusion that SVM provides the best outcome with the greatest accuracy based on the results. [2]

C. Prediction of Depression with Machine Learning using Clinical Support System.

1. Approach, Data Selection and Preprocessing

The CP-CD, based on the ML algorithm, is the author's suggested method for identifying and predicting depression in patients. It is based on an examination of patient data and uses the supervised classification algorithm. [9]

Artificial intelligence continues to play a significant role in health informatics because the volume and diversity of data generated in the field of mental health pose a significant barrier for diagnosis. [9]

2. Feature Selection

To maintain a high enough level of precision in the classification of the psychological state in the context of the defined depression, the authors employed the Gini index to assess the impurity of a data partition or set of training tuples. [9]

$$Gini(v_i) = 1 - \sum_{j=1}^{k} \varphi_j^2$$

By calculating the information gain from developing the tree at a specific position matching to each characteristic, the best split is determined. [9]

$$Gain(p,t) = f(p) - \sum_{j=1}^{n} P_j$$

In this equation, p stands for the position in the tree, t for the test at branch n, Pj for the percentage of elements at position and going to position Pj, and f(p) for either Gini or (p). The node is split using the characteristic that offers the highest gain. [9]

3. Random Forest Algorithm

The suggested system can assist physicians in reaching pertinent conclusions when diagnosing a patient's mental state by automatically and intelligently classifying the patient's depressed state in accordance with each patient's symptoms. The random-forest technique was used to develop a classification model for the patient's depressive condition, and it produced findings that were adequate and of the highest precision. [9]

Input	Type	Description
Age	Numerical	[18. 70]
Antecedent	Numerical	[0, 4]
Symptom 1	Numerical	[0, 4]
Symptom 2	Numerical	[0, 1]
Symptom 3	Numerical	[0, 2]
Symptom 4	Numerical	[0, 4]
Symptom 5	Numerical	[0, 4]
Symptom 6	Numerical	[0, 4]
Symptom 7	Numerical	[0, 3]
Symptom 8	Numerical	[0, 4]
Period	Numerical	[0, 1] > 2 week
Decision	Nominal	Absence depression, mild depression, moderate depression, and major depression.

Table 1. Attributes used in the Determination of Depression [9]

4. Experimental Results and Discussion

The classification of patient data and the prediction of the patient's level of depression are the two primary steps in the authors' detailed explanation of the suggested CP-DDC technique. [9]

Precision					
Class	Naïve - Bayes	Decision Stamp	Random Forest	ZeroR	OneR
Absence Depression	0.901	0.320	1.000	0.300	0.752
Mild Depression	0.910	0.330		0.302	0.715
Moderate Depression	0.930	0.310		0.209	0.800
Depression major	0.909	0.350		0.310	0.790
Weight Average	0.912	0.327	1.00	0.300	0.764

Table 2. Comparison of the random-forest algorithm's results on the dataset used in the suggested system to those of other methods. [9]

5. Results

Table 2 compares the performance of the four methods (Naive-Bayes, DecisioStump, ZeroR, and OneR) with that of the random-forest approach. The method with the highest accuracy and weight average was random-forest. [9]

2. Machine Learning Approaches for Detecting and Attacking Heterogeneity in Schizophrenia

A. Machine Learning Techniques for Diagnosis or Prognosis of Psychiatric Diseases.

1. Treating heterogeneity (in)directly

The decision border between the classes would be more complicated, but it was still possible to attempt to separate patients and controls in a single operation. [10]

2. Transforming the data

Quadratic transforms can be used to build straightforward nonlinear relationships, although this method has the drawback of rapidly growing the feature set. Using a kernel function is a more versatile technique to implement nonlinear connections. [10]

3. Artificial neural networks (ANNs)

Patients with various substrates can be thought of as lying in distinct feature spaces that are delimited by distinct hyperplanes. An example of such a design using linear classifiers and AND/OR operators is a multi-layer perceptron (MLP), and artificial neural networks (ANNs) are its generalization. For modelling various abstraction levels of learned characteristics, deep learning networks have been built. They can be utilized to treat psychiatric illnesses and are excellent for image recognition. [10]

4. Combining linear SVMs

To improve classification accuracy and categorize the pathology, the authors of [11] suggest combining linear classifiers (SVMs). HYDRA is a non-linear semi-supervised machine learning approach that develops a convex polytope that divides the heterogeneous group of patients from the healthy controls. [10]

5. Hierarchical methods such as decision trees, random forests

Another type of AND/OR decision method is a decision tree, in which the classification of a subject depends on where in the (OR) branch's (series of ANDs) terminus the subject is located. Although it comes at the cost of having a very huge tree, they can simulate any difficult decision boundary. [10]

The methods mentioned above target heterogeneity to enhance classification accuracy, but they don't offer much insight into the disease's heterogeneous nature. [10]

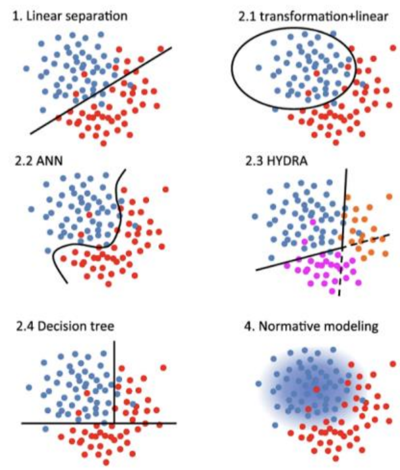

Figure 2. It illustrates various strategies for combating heterogeneity. Red circles stand in for patients, whereas blue circles represent controls. The borders of the decisions are shown by black lines. For an explanation of the techniques, read Section 3's text. The biological space is displayed in the left column and the clinical space is displayed in the right column for the approaches 3.2.5–3.2.7. Two successive steps are denoted by the numbers 1 and 2. A patient subtype or cluster is represented by an orange or pink circle.

6. Subdivision of patients according to explicit, clinical, criteria

The PDS was used by the authors in [12] to divide their sample of schizophrenia patients into subtypes with and without deficits. Between the deficiency group and the other groups, they discovered variations in the anatomy of the brain. According to the researchers in [13], individuals with cognitive deficits and estimated premorbid crystallized IQ (ePMC-IQ) had a more extensive pattern of grey matter decreases than patients with intact ePMC-IQ. Additionally, using structural MRI scans, the authors of [4] Gould et al. (2014) developed a classification strategy to distinguish between schizophrenia patients and control people. They discovered that the two cognitive categories' discriminative sMRI patterns mainly coincide. [10]

There are presumably more subtypes that are based on combinations of patient features, in addition to the well-defined subtypes by theoretical hypotheses or clinical evidence. [10]

7. Clustering based on clinical/demographic variables

This method uses an algorithm to look for patterns in the data that identify different patient subtypes. When there is no obvious way to divide the population into subgroups, it is utilized. [10]

Based on two PANSS composite scores, the authors in [15] segmented a sizable sample of schizophrenia patients into "deficit," "distress," and "low symptom" subgroups. In comparison to the other two groups, the "deficient" subgroup displayed a different activation profile. [10]

A large sample of patients with psychosis, their relatives, and unrelated control subjects were divided into seven categories, or "classes," by the authors in [16] using latent class analysis and factor analysis of clinical data. There were correlations between class membership and cognition and functioning. [10]

8. Clustering based on biological features

A sizable sample of individuals with various forms of psychosis were categorized by the authors in [17] into three classes of distinct "biotypes," each of which represented a distinctive composite pattern. The findings demonstrated that two of the biotypes differed considerably from healthy controls in terms of cognitive control. [10]

145 schizophrenia patients and healthy control participants were subjected to a method of brain subtyping, and the classification accuracy increased from 68.5% without subtyping to 73 and 79% for the two subtypes, respectively. [10]

The benefit of these clustering/subtyping methods is that they make classification models simpler, but the number of clusters selected can have a significant impact on the subtypes that are produced. The separation of patients and controls is a key component of most strategies mentioned thus far. Using a completely new methodology known as generative modelling, both groups are individually modelled as two independent distributions, and a detailed description of the two classes is provided. [10]

9. Normative modeling

This method models healthy people as a multidimensional Gaussian distribution, with the majority of persons concentrated in the center and getting more diluted (with lower probability) as one moves away from the center. Outliers might be found and classified as patients or not-normal. The scientists mapped the association between behavior (trait impulsivity) and biology (reward-related brain activity derived from fMRI) using a healthy cohort and linked extreme values to ADHD symptoms. Large samples are required, according to the scientists, to lessen the impact of these inclusions. [10]

Although normative modelling may work wonders for high-risk populations, it might not be helpful in clinical settings where it is unclear which condition a patient actually has. Discriminative modelling may be more beneficial in these circumstances. [10]

B. Prediction Of Schizophrenia Using Deep Learning

1. Participants

The Maryland Psychiatric Research Center, Johns Hopkins University, the Institute of Psychiatry (IOP), London, UK, and the Western Psychiatric Research Institute and Clinic at the University of Pittsburgh were the four locations from which the sMRI data was gathered, as stated in [8]. It has 191 controls (97 males; age = 40.2615.02; range 16-79) and 198 schizophrenia patients (121 males; age = 39.6812.12; range 17-81). [18]

2. MRI settings

A 1.5 T Signa GE scanner was used to obtain the MRI, and the following settings were used: TR = 35 ms, TE = 5 s, flip angle = 45 degrees, 1 excitation, slice thickness = 1.5 mm, field of view = 24 cm, and matrix size = 256256. [18]

3. Pre-processing

The T1-weighted images were smoothed, interpolated to 1.5 mm3 voxel size, and standardized to MNI standard space. 318 grey matter concentration photos remain after 71 images were removed for quality assurance. [18]

4. Data-driven simulator

Independent component analysis is used by the suggested data-driven simulator to minimize dimensionality and a random variable sampling technique to produce artificial samples. [18]

5. RV generator: Rejection sampling

Given the minimum and maximum observed samples, a random variable generator function takes a sample from the distribution after sampling two RVs from a probability density function. [18]

6. RV generator: Multivariate Normal

We create multivariate random normal samples using the spectral decomposition method, but we lose the generality of the marginal distributions. [18]

7. Structural MRI generator

Based on these presumptions, the generator divides the observed dataset X into submatrices AHC and ASZ before factoring the observed dataset X into A and S. Each submatrix AHC and ASZ are then fed to an RV generator separately. The data are first factorized using the synthetic structural MRI generation algorithm, followed by the creation of a loading matrix and the input of the synthetic data to the rejection sampling RV generator. [18]

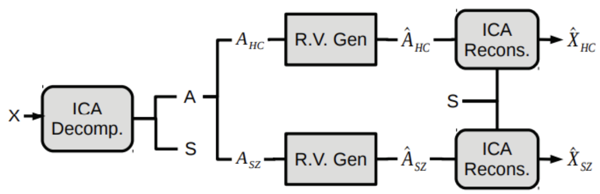

Figure 3. Block diagram for the development of synthetic structural MRI data, where X stands for data, A and S for the factorization of X, A for the created loading matrix, and X for the synthetic data.

The mean and covariance matrices are estimated, and M samples are created using the multivariate normal sampling technique. M pictures should be rebuilt, one for each diagnosis category. [18]

$$[\hat{X}_{\mathrm{HC}}]_{M\times m} = [\hat{A}_{\mathrm{HC}}]_{M\times c}[S]_{c\times m} + \bar{X}$$
$$[\hat{X}_{\mathrm{SZ}}]_{M\times m} = [\hat{A}_{\mathrm{SZ}}]_{M\times c}[S]_{c\times m} + \bar{X}, \quad [18]$$

Where X is the voxel mean computed at the beginning of the method and X is the resulting simulated image. The method RV generator method depicted in Figure 3. [18]

8. Multilayer Perceptron

To maximize the binary cross entropy, we employed a multilayer perceptron (MLP) with 10 layers, 1000 hidden nodes for each layer, sigmoid activation functions, 50% dropout for each layer, and L1 regularization. [18]

3. Application of Machine Learning to the Detection of Cognitive Decline

A. Utilizing a Cognitive Ability Test to Predict Cognitive Decline

1. Survey

Mild cognitive impairment (MCI) is a precursor to Alzheimer's and other types of dementia later in life, which can have a significant negative influence on one's physical health as well as their psychological, social, economic, and social well-being. [19]

Understanding the association between disorders like diabetes, hypertension, depression, and cognitive impairment is provided by the summary. The data is analysed using a variety of statistical methods, and a machine learning algorithm is then used to forecast the likelihood of developing Alzheimer's in later life. [19].

2. Proposed Work

The authors' suggested approach seeks to predict cognitive decline in individuals using their historical cognitive, clinical, and physical data. Early detection of those who are at high risk for dementia is crucial. [19]

The existing approach necessitates the collection of imaging data or fluid, which delays early detection. VAD and AD can be predicted using machine learning. Globally, the prevalence of chronic illnesses like diabetes, high blood pressure, heart disease, and kidney infections is rising. Patients with type 2 diabetes should routinely have their cognitive abilities evaluated because diabetes is seen as a major risk factor for cognitive impairment. [19]

Until the patient or the patient's family consents, doctors will not examine the patient's psychological capacity. A quick neuropsychological screening test called the Cognitive Ability Test (CAT) gives an overview of cognitive function. [19]

1. Data and Methodology

The most frequent risk factors for cognitive impairment in middle-aged persons, according to this study, are hypertension and diabetes. The best method for determining a causal relationship is a credible cross-sectional analysis. The study looks at general health data from the "Data globe" and discovers a link between elevated blood pressure in middle age and the onset of dementia and AD in their later years. [19]

2. Feature Selection

There are 19 features in the data set that describe age, cholesterol, glucose, systolic and diastolic blood pressure, and cholesterol. HbA1c was taken into account in the early detection of AD, and the correlation factor connected to each set of features aided in the extraction of pertinent characteristics for research. [19]

3. Process flow

Figure 3 depicts the proposed approach to forecast the degree of cognitive deterioration. To more precisely assess cognitive deterioration, it applies a 2-stage classification model to a set of diabetes and blood pressure data from "Data World". [19]

III. Algorithm

Step 1: Patients with Blood Pressure and Diabetes dataset input in the first step. Sort statistics for people aged 40 to 65. Deal with erroneous and incomplete data Result: processed data
Step 2: Employ multivariate analysis to find the features of the dataset that are correlated.
Step 3: Perform a stage 1 classifier's initial classification for Alzheimer's disease
Step 4: If AD possible, have those patients undergo CAT testing.
Step 5: Use second-level categorization to look for AD/HD, Uncer [19]

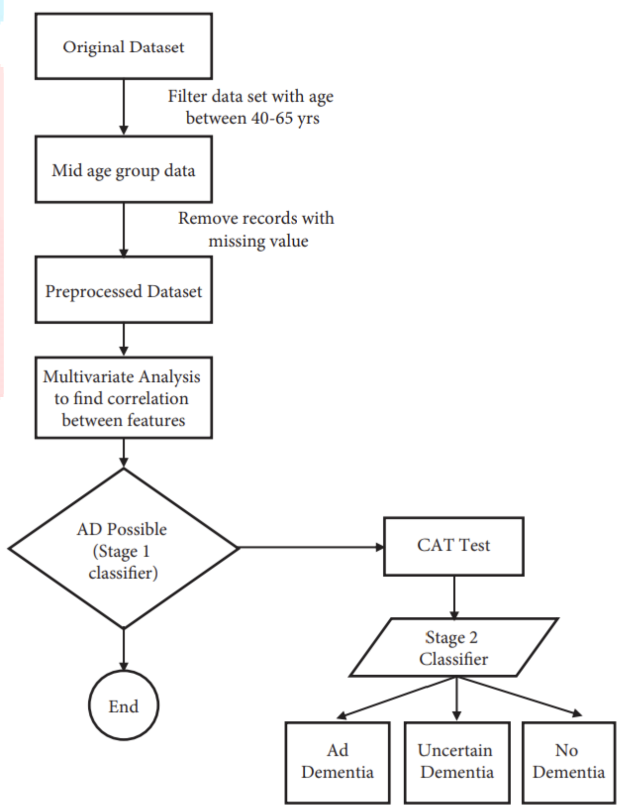

Figure 5. Flowchart

IV. *Flowchart's process*

Figure 2 shows a flowchart of the suggested approach to predict the degree of cognitive deterioration. [19]

V. *Support Vector Machine*

A popular supervised classifier is the Support Vector Machine (SVM), which uses a hyperplane to categorise data in N-dimensional space. It has been used extensively in the healthcare industry to forecast diseases using structural data. [19]

VI. *Random Forest*

The Random Forest approach mixes the output of n distinct decision trees that have been trained using various data subsets and tuning parameters. [19]

VII. *Mental Ability Test*

To evaluate a person's cognitive impairment, a cognitive ability test is administered. The test contains questions about orientation, memory, attention, recall, naming things, reacting to verbal and written orders, composing a phrase, and copying a figure. The CAT score is a gauge of cognitive deterioration. Each year, it drops by an average of two to four points. [19]

On the basis of the dependent variables, the probabilities of the N categories are estimated.

$$pr\,(Yi = kXi{:}\ \ \beta 1,\ \beta 2\ldots,\beta n)$$
$$= \frac{\exp\left(\beta 0k + Xi\,\beta' k\right)}{\sum_{j=1}^{n}\exp\left(\beta 0k + Xi\,\beta' k\right)},$$

Figure 6. k is the regression coefficient for the kth category of Y, where Y is the dependent variable and X is the set of explanatory variables.

Based on the association between the dependent variables, multinomial logistic regression is used to forecast the illness severity in CAT patients. [19]

IV. Results

The suggested classifier model's performance is evaluated using a confusion matrix. Using the provided formula, the model's overall accuracy, sensitivity, and specificity are determined, and its true positive rate, sensitivity, specificity, and true negative rate are assessed. With an AUC value of 0.90, the suggested SVM classifier surpasses the Random Forest algorithm. The probabilistic classifier illustrates the tradeoff between sensitivity and specificity. The Random Forest algorithm can be improved by changing the parameters, but in our case study we were unable to precisely adjust the parameters. If there is enough training data, a multinomial logistic regression model can predict with improved accuracy. [19]

B. Prediction of Cognitive Decline using Deep Learning from Brain Metabolism and Amyloid Imaging

The Alzheimer's Disease Neuroimaging Initiative-II (ADNI-II), a public-private collaboration headed by Principal Investigator Michael W. Weiner, MD, VA Medical Center, and University of California San Francisco, provided the subjects for this study's data collection. Using baseline FDG and AV-45 PET scans, we examined 182 healthy controls, 139 Alzheimer's disease patients, and 171 patients with mild cognitive impairment (MCI). The findings demonstrated a relationship between baseline PET biomarkers and long-term cognitive deterioration. [20]

1. FDG PET and AV-45 PET

The preprocessed PET pictures were retrieved from the ADNI database by the authors, who then uniformized the voxel size, co-registered the images, and applied smoothing specific to each scanner. We tested and trained deep CNN using these photos. [20]

2. Study design

The study created a sophisticated CNN-based approach for identifying patients who would eventually acquire AD and for predicting cognitive impairment. The approach produced a quantitative MCI converter score without any training data. [20]

3. Prediction of cognitive decline in MCI subjects

MatConvNet was used to create 8 CNNs, and cross-validation was used to determine the sensitivity, specificity, and accuracy of the classification between AD and NC. In order to determine whether or not MCI participants would eventually develop AD, we trained a CNN to predict cognitive deterioration in MCI subjects. Using ConvScore, the ROC analysis and accuracy, sensitivity, specificity, and predictive value of the CNN were assessed. [20]

4. Voxel wise Feature Extraction, Support Vector Machine and Statistics

The whole cortical mean uptake value expressed relative to uptake in the entire cerebellum was used to compute the AV-45 PET cortical uptake. Additionally, the suggested CNN-based method was contrasted with a traditional machine learning method employing voxelwise feature extraction. The outcomes demonstrated that for AD classification and MCI conversion prediction, the CNN-based strategy outperformed the traditional feature-based approach. [20]

5. Results

The accuracy of the CNN-based technique was much higher than that of the SVM classifier and the VOI-based analysis in this investigation, which included 492 participants. Additionally, the deep CNN's accuracy was much higher than that of traditional techniques. To forecast MCI conversion and AD using ConvScore, we used deep CNN. ConvScore's AUC was much higher than feature-VOI based analysis's AUC. ConvScore had a negative correlation with the MMSE and was strongly connected with the longitudinal change in cognitive assessments in MCI patients at 1 and 3 years. It could be employed as a quantitative biomarker to foretell the progression of AD in MCI patients as well as the reduction in their longitudinal cognitive assessments. [20]

IV. Conclusion

The objectives of this paper is to explore researchers work on prediction of mental health illness like depression and anxiety and schizophrenia. Compare to machine learning algorithms, deep learning algorithms provides higher accuracy. And conclude that CNN give more accuracy rate than the SVM.

References

[1] Zulfiker, Md & Ety, Nasrin & Biswas, Al Amin & Nazneen, Tahmina & Uddin, Mohammad. (2021). An In-Depth Analysis of Machine Learning Approaches to Predict Depression. Current Research in Behavioral Sciences. 2. 100044. 10.1016/j.crbeha.2021.100044.

[2] Usman, Muhammad & haris, syed & Fong, A.M.. (2020). Prediction of Depression using Machine Learning Techniques: A Review of Existing Literature. 10.1109/SBBS50483.2020.9314940.

[3] I. Bhakta and A. Sau "Prediction of depression among senior citizens using machine learning classifiers", International Journal of Computer Applications, vol. 144, no. 7, pp. 11-16, June 2016. DOI: 10.5120/ijca2016910429

[4] A. Sau and I. Bhakta, "Predicting anxiety and depression in elderly patients using machine learning technology," Healthcare Technology Letters, vol. 4, no. 6, pp. 238-243, 12 2017, doi: 10.1049/htl.2016.0096.

[5] J. A. Andrews, R. F. Harrison, L. Brown, L.M. MacLean, F. Hwang, T. Smith, E.A. Williams, C. Timon, T. Adlam, H. Khadra, and A.J. Astell, "Using the NANA toolkit at home to predict older adults' future depression", Journal of Affective Disorders, vol. 213, pp. 187–190, 2017. https://doi.org/10.1016/j.jad.2017.02.019.

[6] M. J. Patel, C. Andreescu, J.C. Price, K.L. Edelman, C.F. Reynolds III, and H. J. Aizenstein, (2015). Machine learning approaches for integrating clinical and imaging features in late Θ life depression classification and response prediction. International journal of geriatric psychiatry, 30(10), 1056-1067

[7] T. Iliou, G. Konstantopoulou, M. Ntekouli, C. Lymperopoulou, K. Assimakopoulos, D. Galiatsatos, and G. Anastassopoulos, "ILIOU machine learning preprocessing method for depression type prediction", Evolving Systems, vol. 10, pp. 29–39, 2019. https://doi.org/10.1007/s12530-017-9205-9

[8] W. Mumtaz, S. S. A. Ali, M.A.M. Yasin, A.S. Malik, A machine learning framework involving EEG-based functional connectivity to diagnose major depressive disorder (MDD), Medical & Biological Engineering & Computing, vol. 56, 2017. 10.1007/s11517-017-1685-z

[9] Benfares, Chaymae & Akhrif, Ouidad & Idrissi, Younès & Hamid, Karim. (2020). A clinical support system for classification and prediction of depression using machine learning methods. Computational Intelligence. 37. 10.1111/coin.12377.

[10] Schnack, Hugo. (2017). Improving individual predictions: Machine learning approaches for detecting and attacking heterogeneity in schizophrenia (and other psychiatric diseases). Schizophrenia research. 214. 10.1016/j.schres.2017.10.023.

[11] Varol, E., Sotiras, A., Davatzikos, C., Alzheimer's Disease Neuroimaging Initiative, 2017. HYDRA: revealing heterogeneity of imaging and genetic patterns through a multiple max-margin discriminative analysis framework. NeuroImage 145 (Pt B), 346–364.

[12] Takahashi, T., Takayanagi, Y., Nishikawa, Y., Nakamura, M., Komori, Y., Furuichi, A., Kido, M., Sasabayashi, D., Noguchi, K., Suzuki, M., 2017. Brain neurodevelopmental markers related to the deficit subtype of schizophrenia. Psychiatry Res. 266, 10–18.

[13] Czepielewski, L.S., Wang, L., Gama, C.S., Barch, D.M., 2017. The relationship of intellectual functioning and cognitive performance to brain structure in schizophrenia. Schizophr. Bull. 43 (2), 355–364.

[14] Gould, I.C., Shepherd, A.M., Laurens, K.R., Cairns, M.J., Carr, V.J., Green, M.J., 2014. Multivariate neuroanatomical classification of cognitive subtypes in schizophrenia: a support vector machine learning approach. Neuroimage Clin. 6, 229–236.

[15] Dickinson, D., Pratt, D.N., Giangrande, E.J., Grunnagle, M., Orel, J., Weinberger, D.R., Callicott, J.H., Berman, K.F., 2017. Attacking heterogeneity in schizophrenia by deriving clinical subgroups from widely available symptom data. Schizophr. Bull. 2017 Mar 20.

[16] Derks, E.M., Allardyce, J., Boks, M.P., Vermunt, J.K., Hijman, R., Ophoff, R.A., GROUP, 2012. Kraepelin was right: a latent class analysis of symptom dimensions in patients and controls. Schizophr. Bull. 38 (3), 495–505.

[17] Clementz, B.A., Sweeney, J.A., Hamm, J.P., Ivleva, E.I., Ethridge, L.E., Pearlson, G.D., Keshavan, M.S., Tamminga, C.A., 2016. Identification of distinct psychosis biotypes using brainbased biomarkers. Am. J. Psychiatry 173 (4), 373–384.

[18] Ulloa, Alvaro & Plis, Sergey & Erhardt, Erik & Calhoun, Vince. (2015). Synthetic structural magnetic resonance image generator improves deep learning prediction of schizophrenia. 10.1109/MLSP.2015.7324379.

[19] A., Revathi & Ramar, Kaladevi & Kadiyala, Ramana & Jhaveri, Rutvij & Kumar, Dr & Kumar, M.. (2022). Early Detection of Cognitive Decline Using Machine Learning Algorithm and Cognitive Ability Test. Security and Communication Networks. 2022. 1-13. 10.1155/2022/4190023.

[20] Choi, Hongyoon & Jin, Kyong. (2017). Predicting Cognitive Decline with Deep Learning of Brain Metabolism and Amyloid Imaging. Behavioural Brain Research. 344. 10.1016/j.bbr.2018.02.017.