INTERNATIONAL JOURNAL ON HUMAN COMPUTING STUDIES
https://journals.researchparks.org/index.php/IJHCS e-ISSN: 2615-8159 | p-ISSN: 2615-1898
Volume: 03 Issue: 6 |Aug 2021

Hybridization of Machine Learning Techniques in Predicting Mental Disorder

[1]Emmanuel Oluwadunsin Alabi, [2]Oluwashola David Adeniji
[3]Tolulope Moyo Awoyelu, [4]Oluwakemi Dunsin Fasae
[1,2,4]Department of Computer Science, University of Ibadan, Ibadan, Nigeria
[3]Department of Computer Sciences, Ajayi Crowther University, Oyo, Nigeria

----------***----------

Abstract: This study applied a hybrid Random Forest and Artificial Neural Network (RF-ANN) model in predicting the chances of IT employees developing mental disorder. To measure the performance of the model, Random Forest and Artificial Neural Networks algorithms were separately developed, their results were recorded and compared with the results of the hybridized model. The result obtained from the hybridized model showed a significant improvement in its performance over the individual performances of the Random Forest model and Artificial Neural Networks models. Hybridizing Random Forest and Artificial Neural Networks using "Bagging Ensemble" produced a model that was able to correctly predict the chances of IT employees developing Mental Disorder with 94% recall and 80% F1-score compared to 65% and 60% respectively in the Random Forest Model. With these results, applying the RF-ANN model on improved dataset could be investigated and compared with the results found in this study.

Keywords: Hybridized, IT, Machine Learning, Mental Disorder, Mental health, RF-ANN

Introduction

Mental health is the psychological, social and emotional state of a person who is functioning at an acceptable level of behavioral and emotional adjustment. Mental health can be seen as a measure to which an individual can handle stress and make decisions in every facet of their life, as it severely impacts how such individual acts, thinks and feels. At any stage in life, be it adulthood or childhood, mental health is a very important factor [1].

According to the World Health Organization, depression - the leading cause of Mental Health Disorder worldwide, affects people individually and as a community. It is estimated that over 350 million people worldwide lives with depression [2]. The impact of mental health problems in the workplace has serious consequences not only for the individual, but also for the productivity of the organization. Unfortunately, people generally find it difficult to discuss mental health issues publicly, and sufficient awareness is not being made in the society.

In recent years, machine learning techniques have been adopted in numerous medical researches, especially in biomedicine and neuroscience to gain further insight into mental health disorders [3]. Machine learning, being an area of artificial intelligence involves the process of computers learning from data through the use of heuristic algorithms [4]. It can be generally classified into supervised and unsupervised machine learning. Supervised machine learning usually involves the process of assigning a set of attributes to a target class, it implies classification and regression. While unsupervised machine learning involves the process of describing the relationship or characteristics of a set of attribute data. Unsupervised machine learning mostly requires a process of Feature Selection, Clustering and Association rule mining.

Studies show that employees in IT industries are at high risk of developing mental disorder due to increased stress and pressure to meet targets and deadlines. In many cases, these disorders are not

https://journals.researchparks.org/index.php/IJHCS e-ISSN: 2615-8159 | p-ISSN: 2615-1898
Volume: 03 Issue: 6 |Aug 2021

obvious, known or diagnosed until they become life-threatening. Existing studies in different fields have implemented various machine techniques to predict mental disorder. However, there is the need to address the issue of late detection of mental disorder in IT employees.

Developing automated models that can predict, diagnose and classify mental health disorder is now possible with the help of computer-aided systems. By using these developed models, they help in saving manpower, time and other resources, while also removing the possibilities of human bias. Large amount of data is readily available thanks to the advancement in the usage, power and capacity of latest computer technologies. This has resulted in increase in the ability to collect, store and manipulate data. Subsequently, knowledge can be extracted from the data by bringing out patterns and relationships through the development of a methodology. Such methodology can be developed from a database of existing tools and methods available for the discovery of knowledge and data mining [5].

A hybridized model of neural network with random forest structure can produce a result with improved generalization ability and accuracy. The ability of this hybridized architecture to reduce the back-propagation algorithm to a more powerful and generalized decision tree structure makes it more effective than random forests. In addition, this model is more efficient to train as the number of training examples usually requires only a small constant factor [6]. By hybridizing two of the best performing models in previous studies, this study aims to develop a model that can better predict the chances of IT employees developing mental disorder by:

- Employing a Random Forest algorithm to predict mental health disorder among IT industry employees.
- Employing an Artificial Neural Network algorithm to predict mental health disorder among IT industry employees.
- Developing a hybridized RF-ANN model using Random Forest as the base model.
- Evaluating the performances of the hybridized model using standard evaluation metrics such as: precision, recall, f1-score and accuracy.

Literature Review

Mental health disorder, also called mental illness relates to a wide range of mental health conditions that affect a person's thinking, mood, and behavior. Examples of mental disorder include anxiety disorders, depression, addictive behaviors, schizophrenia, eating disorders and so on. From time to time, many people have various mental health concerns. However, these mental health concerns only become a mental disorder when the on-going signs and symptoms result in frequent stress and affect the ability for such person to function effectively.

Problems such as loss of pleasure or interest, poor concentration, lack of appetite, disturbed sleep, feelings of guilt and low energy as constituents that come with mental health disorder. These problems have the tendency to become chronic and recurrent, and thus impair a person's ability to take care of their daily responsibilities [7].

According to [8], more than 30% of people suffering from major mental disorder do not seek treatment, while more than 80% of people battling with some forms of mental disorder do not seek to be treated at all. Variations of mental illnesses include: depression, bipolar disorder, schizophrenia and other psychoses, dementia developmental disorders.

A. Machine Learning and Healthcare

Precision medicine is a way in which healthcare professionals can move to a more personalized care by adopting machine learning in finding patterns and reasons about data [9]. With the

https://journals.researchparks.org/index.php/IJHCS e-ISSN: 2615-8159 | p-ISSN: 2615-1898
Volume: 03 Issue: 6 |Aug 2021

large volume of data being collected about patients in the healthcare sector, it is near-impossible for humans to analyze. With sufficient data and permission to use, there are numerous ways in which machine learning can be applied in healthcare. In times past, hard-coded software has been developed based on external studies to provide recommendations and alerts for different medical practices. The limitation to this however is the problem with accuracy of data due to other factors such as location, environment, population, and so on.

With machine learning, data can be refined to a particular environment, for instance, refining a data from a hospital and the surrounding environment in a way that the patient's information is anonymized. Examples of ways in machine learning can help healthcare providers include: prediction of possible outbreak of disease, predicting the possibility of hospital re-admission for critically ill patients, prediction of cancer risks in patients, and so on [5]. According to [10], being able to identify patients that are most liable to the risk of hospital readmission helps healthcare providers to offer better support after discharge. The lives of those at risk are improved when the rate of readmission is lowered, and this can be made possible with the intervention of machine learning.

Implementation of artificial intelligence in healthcare organizations as a response to the needs of doctors to aid them in their daily decision-making activities is now on the increase. This hopes to improve decision making and reduce errors.

B. Related Works

Prediction of mental health problems among children using eight machine learning techniques of which the three classifiers multilayer perceptron, multiclass classifier and logical analysis of data (lad) tree produced more accurate results and there was only a slight difference between their performances over full attribute set and selected attribute set. The study found that by developing a high-performing model, early diagnosis of mental health problems in children will help healthcare professionals to treat it at an earlier stage and subsequently improve the quality of patients' life. Therefore, there comes an urgent need to treat basic mental health problems that persist among children which may lead to complicated problems, if not treated at an early stage [11].

By introducing genetic algorithm in developing a system for intelligent data mining and machine learning for mental health diagnosis, [12] was able to extract keywords from user's symptoms. The approach markedly increased the success of information retrieval and relevancy between keywords-matching and relevant user's symptom. In this research, a new study introduced a semiautomated system that helps in preliminary diagnosis of the psychological disorder patient. This was achieved by matching description of a patient's mental health status with the mental illnesses. The study constructed a semi-automated system based on an integration of the technology of genetic algorithm, classification data mining and machine learning. The goal was to help psychological analysts make informed, appropriate and intelligent assessment leading to accurate prognoses by ensuring that they are aware of all possible mental health illnesses that could match the patient's symptoms.

In a predictive research for mental health disease, [13] proposed a prototype that uses random forests classification to determine the mental state of a person based on attributes such as life style, age, education, gender, vision, occupation, sleep, personal income, mobility, diabetes and hypertension. With the amount of data produced by humans daily and with most of these data stored in semi-structured way, these researchers believed that by using this machine learning technique, hidden patterns can be found between the different attributes of data. With these, the system will be able to predict whether a patient is suffering from mental illness or not.

https://journals.researchparks.org/index.php/IJHCS e-ISSN: 2615-8159 | p-ISSN: 2615-1898
Volume: 03 Issue: 6 |Aug 2021

A critical review using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease was conducted by [14]. The study provided an overview of the method and reviewed studies that applied SVM in the investigation of schizophrenia, Alzheimer's disease, Parkinson's disease, bipolar disorder, pre-symptomatic Huntington's disease, major depression, and autistic spectrum disorder. Standard univariate analysis of neuroimaging data revealed a host of neuroanatomical and functional differences between healthy individuals and patients suffering a wide range of psychiatric and neurological disorders.

Mental health evaluation model based on fuzzy neural network was carried out by [15]. By selecting the important factors such as input vector, they model was used to evaluate the psychological health of college students in China. The introduction of neural network and fuzzy mathematics improved the accuracy of the mathematical model compared to other traditional models and made it easy to analyze the overall mental health trend of students.

Recurrent and linear models to detect depression early were developed by [16]. The work focused on early, automatic detection of depression from users' posts on the social media site – Reddit. For prediction, both sequential (recurrent neural network) and non-sequential (support vector machine) models were used. The results showed the superiority of sequential models over non-sequential models. The work did not sufficiently explore the broad range of possible features.

Different machine learning techniques such as K-nearest neighbor classifier, support vector machines, naïve bayes classifier, decision trees, and logistic regression to identify the state of mental health in a target group were proposed by [17]. The responses gotten from the target group for the designed questionnaire were initially subjected to unsupervised learning techniques. The labels obtained from clustering were validated by computing the Mean Opinion Score. The cluster labels were then used to build classifiers to predict the mental health of an individual. Population from wide range of groups such as college students, high school students, and working professionals were considered as target groups.

A survey on the analysis of the mental state of social media users to predict depression was conducted by [18]. The survey was done to detect depression and mental illness by the use of social media are surveyed. They found out that there was a very high rate at which depression and mental illness were being diagnosed in recent times. They observed that some symptoms linked to mental illness were detectable on Facebook, Twitter, and web forums. They suggested that using automatic methods would help in locating inactivity and other mental disease. Various automated detection methods could help to detect depressed people using social media. Mentally ill users were pointed out through the use of screening surveys, their twitter analysis based on community distribution, or by their membership in online forums, and they were detectable through the patterns in their language and online activities. Additionally, they observed that a number of authors experienced that numerous activities on social networking sites could be linked to low self-confidence, especially in young people and adolescents.

A predictive model for the determination of risk of depression among university students was also developed by [5]. The study extracted knowledge on the factors causing depression among university students. In the study, a predictive model for depression risk with a view to determining the risk of depression among university students was formulated, simulated and validated. The result of the study identified variables that have strong relevance to developing depression among university students. The simulation results showed that the model without feature selection gave a total of 465 correct classification out of 507 records with an accuracy of 91.7% while with feature

INTERNATIONAL JOURNAL ON HUMAN COMPUTING STUDIES
https://journals.researchparks.org/index.php/IJHCS e-ISSN: 2615-8159 | p-ISSN: 2615-1898
Volume: 03 Issue: 6 |Aug 2021

selection, gave a total of 475 correct classification with an accuracy 93.7%.

Data Collection/ Implementation

The input data for this model is the dataset provided by OSMI Ltd. The dataset is derived from a survey aimed at measuring the attitudes of people towards mental health in IT workplace and examining the rate of mental disorders among IT employees. The dataset contains 63 variables or columns (including the target variable) and 1,433 responses/ observations. The performances of the models built were evaluated on the basis of their precision, recall, f1-score and accuracy. To further understand and measure the performance of the hybrid model, different set of algorithms were implemented. These models include: random forest with default parameters, random forest with tuned parameters and artificial neural networks.

Methodology

The developed model consists of three different phases namely: the dataset pre-processing, features extraction (training and testing), and the results evaluation. At the pre-processing phase, values of missing data were replaced, even distribution of data was ensured with features scaling. Features of the pre-processed data were split into two with 67% of the data set aside for training and the remaining 33% set aside for testing. Using bagging ensemble method, the training set was passed through hybridized Random Forest and Artificial Neural Networks algorithms, using Random Forest as the base. The performance of the result was then evaluated using the standard evaluation metric namely: precision, recall, f1-score and accuracy. Figure 1 shows the developed methodology framework.

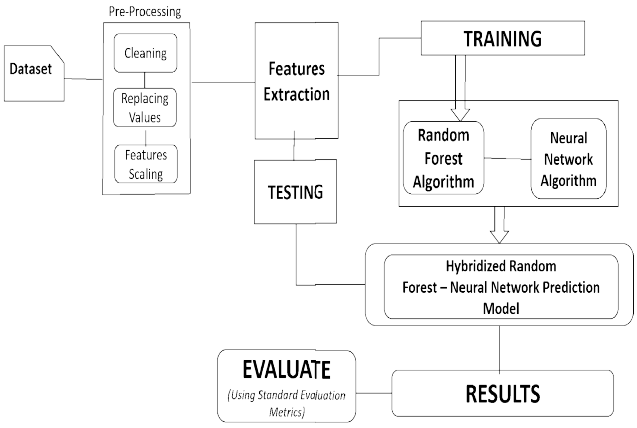

Fig. 1 Developed methodology framework

The performances of these models were observed and compared with the hybridized model. The results obtained from the trained models are discussed below:

A. Random Forest Model

The result obtained from the model as shown in figure 2 shows that the model performed very poorly in correctly classifying IT employees "not sure of their mental disorder status" with 23%, 17%, 19% and 50% precision, recall, f1-score and accuracy respectively. It was also very poor in correctly classifying IT employee "not having mental disorder", with barely a random guess of 53% each for precision, recall, f1-score, and 50% accuracy. Although it performed a little better in correctly classifying IT employees "having mental disorder" with performance metrics of 56%, 65%, 60% and 50% precision, recall, f1-score and accuracy respectively.

The default-parameterized random forest-based model performed at a weighted average of 48% precision, 50% recall, 49% f1-score and barely 50% accuracy. This poor performance called for the need to tune the parameter in order to obtain an improved performance.

INTERNATIONAL JOURNAL ON HUMAN COMPUTING STUDIES
https://journals.researchparks.org/index.php/IJHCS e-ISSN: 2615-8159 | p-ISSN: 2615-1898
Volume: 03 Issue: 6 |Aug 2021

	precision	recall	f1-score	support
0	0.23	0.17	0.19	103
1	0.53	0.53	0.53	183
2	0.56	0.65	0.60	187
accuracy			0.50	473
macro avg	0.44	0.45	0.44	473
weighted avg	0.48	0.50	0.49	473

Fig. 2 Result of the Random Forest Model

B. Artificial Neural Networks (ANN) Model

The ANN model as shown in figure 3 performed poorly in correctly classifying "Employees not sure of their mental health status". It however did better with 90%, 70%, 79% and 71% precision, recall, f1-score and accuracy respectively for correct classification of "Employees not having mental disorder"; and 65% precision, 95% recall, 77% f1-score and 71% accuracy for IT employee "actually suffering from mental disorder".

Overall, the model performed at a weighted average of 72% precision, 71% recall, 69% f1-score and 71% accuracy.

	precision	recall	f1-score	support
0	0.52	0.28	0.36	103
1	0.90	0.70	0.79	183
2	0.65	0.95	0.77	187
accuracy			0.71	473
macro avg	0.69	0.64	0.64	473
weighted avg	0.72	0.71	0.69	473

Fig. 3 Result of the ANN Model

C. RF-ANN Model

Using Bagging ensemble method, Random Forest and ANN algorithms were hybridized, using the Random Forest as the base model. The hybrid model was trained on the training set and its performance was evaluated on the testing set.

The result of the hybridized model in figure 4 showed a significant improvement in the performance of the hybrid model over the performances of the Random Forest model with default parameters and ANN model and a slight improvement over the performance of the parameter-tuned Random Forest model. The ANN model's weighted average performance was improved by the hybrid model from (72% 71%, 69% and 71%) to (74%, 74%, 80% and 74%) precision, recall, f1-score and accuracy respectively.

	precision	recall	f1-score	support
0	0.64	0.38	0.48	103
1	0.84	0.74	0.79	183
2	0.69	0.94	0.80	187
accuracy			0.74	473
macro avg	0.73	0.68	0.69	473
weighted avg	0.74	0.74	0.72	473

Fig. 4 Result of the Hybrid Model

Findings

From all the results obtained shown in table 1 reveals that hybridizing Random Forest and Artificial Neural Networks using "Bagging Ensemble" gave the best performance with weighted average performance of 76% precision, 75% recall, 73% f1-score and 75% accuracy. It can also be observed that the model is able to correctly predict IT employees suffering from Mental Health Disorder with 97% recall. Additional insight gotten from the results in this experiment reveal that there was only a marginal improvement in the performance of the hybridized model when compared with the result of the parameter-tuned Random Forest. This shows that Random Forest is one of the best classifiers in predictive algorithm which is consistent with the work of [13].

https://journals.researchparks.org/index.php/IJHCS e-ISSN: 2615-8159 | p-ISSN: 2615-1898
Volume: 03 Issue: 6 |Aug 2021

Table I Result Comparison of the 3 Models

Model/Evaluation Metric	Precision (%)	Recall (%)	F1 – Score (%)	Accuracy (%)
Random Forest	48	50	49	50
Artificial Neural Network	72	71	69	71
Hybrid RF-ANN Model	74	74	72	74

Summary

The study described various approaches to predicting mental disorder. It focused on the development of a hybridized predictive model for determining the risk level of mental disorder among employees in IT industry. Most of the existing models focused on predicting mental disorder using a single machine learning technique.

This study identified the variables measured in IT employees which are relevant to prediction of mental disorder in the dataset collected. The results obtained showed that developing a hybridized RF-ANN model had the best overall performance, hence, our developed model.

Conclusion

The study described various approaches to predicting mental disorder. It focused on the development of a hybridized predictive model for determining the risk level of mental disorder among employees in IT industry. Most of the existing models focused on predicting mental disorder using a single machine learning technique. However, there is also the need to address the issue of late detection of mental disorder in IT employees. This study identified the variables measured in IT employees which are relevant to prediction of mental disorder in the dataset collected.

In conclusion, the best performing model to predict mental disorder in employees of IT industry have been identified, the work has been able to develop a predictive model based on the most relevant factors causing mental disorder. From all the results obtained, hybridizing Random Forest and Artificial Neural Networks using "Bagging Ensemble" gave the best performance in predicting IT employees suffering from Mental Health Disorder with 94% recall. In addition, insights gotten from the results in this study reveal that there was only a marginal improvement in the performance of the hybridized model when compared with the result of the parameter-tuned Random Forest.

Recommendations

Based on the outcomes of this study, the following recommendations are suggested:

- This model can be adopted as an assistant tool by mental health professionals to help them in making early and more consistent diagnosis of mental disorder.
- The model can be integrated into an existing employees' Health Information System (HIS) which have clinical data about the employees. With the right information, data can be fed to the mental disorder risk prediction model, thereby helping healthcare professionals in detecting the risks early.
- It is recommended that variables monitored in IT employees be continually reviewed to increase the number of information relevant to developing an improved prediction model for mental health disorder.

In the future, hybridized algorithm can be used more in predicting mental disorder in other fields. Equally, applying random forest with parameter tuning on improved dataset could be investigated and compared with the results found in this study.

https://journals.researchparks.org/index.php/IJHCS e-ISSN: 2615-8159 | p-ISSN: 2615-1898
Volume: 03 Issue: 6 |Aug 2021

References

[1] P. Sandhya, and K. Mahek, "Prediction of Mental Disorder for employees in IT Industry." IJITEE (International Journal of Innovative Technology and Exploring Engineering, 2019), 8(6s): 2278-3075.

[2] S. S. Lim, T. Vos, A. D. Flaxman, G. Danaei, K. Shibuya, H. Adair-Rohani, M. A. AlMazroa, M. Amann, H. R. Anderson, and K. G. Andrews. "A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990 – 2010: a systematic analysis for the Global Burden of Disease Study 2010." The Lancet, 2012, 380(9859):2224-2260.

[3] J. F. Dipnall, J. A. Pasco, M. Berk, L. J. Williams, S. Dodd, F. N. Jacka, and D. Meyer. "Fusing data mining, machine learning and traditional statistics to detect biomarkers associated with depression." PloS one, (2016) 11(2):148-195.

[4] J. Han, J. Pei, and M. Kamber. "Data Mining: Concepts and Techniques." Elsevier, New York, United States, (2011).

[5] T. M. Awoyelu A. R. Iyanda and S. K. Mosaku. "Formulation of a Predictive Model for the Determination of Depression among University Students." Proceeding of the 14th International Conference on Smart Nations, Digital Economies and Meaningful Lives Nigerian Computer Society, (NCS, 2019), 098- 109, Nigeria.

[6] S. Wang, C. Aggarwal, and H. Liu. Random-Forest-Inspired Neural Networks. ACM Trans. Intell. Syst. Technol., 2018. 9, 6(69), 25 pages. https://doi.org/10.1145/3232230.

[7] P. Gilbert. Depression: The Evolution of Powerlessness. Routledge, London, United Kingdom, 2016.

[8] L. A. Cooper, J. J. Gonzales, J. J Gallo, K. M Rost, L. S. Meredith, L. V. Rubenstein, N. Y. Wang, and D. E. Ford. The Acceptability of Treatment for Depression among African-American, Hispanic, and White Primary Care Patients. Medical Care, (2003). 41(4):479-489.

[9] M. Kantardzic. Data Mining: Concepts, Models, Methods, and Algorithms. John Wiley and Sons, New Jersey, United States, 2011.

[10] P. Groves, B. Kayyali, D. Knott, and S. Van Kuiken, S. The Big Data Revolution in Healthcare. McKinsey Quarterly, 2:3-5, 2013.

[11] M. R. Sumathi, and B. Poorna. "Prediction of Mental Health Problems Among Children Using Machine Learning Techniques." IJACSA (International Journal of Advanced Computer Science and Applications, 2016). 7(1).

[12] G. Azar, C. Gloster, N. El-Bathy, S. Yu, R. H. Neela, and I. Alothman. "Intelligent Data Mining and Machine Learning for Mental Health Diagnosis using Genetic Algorithm." (2015)

[13] S. Chauhan, and A. Garg. Predictive Research for Mental Health Disease. Blue Eyes Intelligence Engineering & Sciences Publication, 2019. https://doi.org/10.35940/ijitee.I1038.0789S219.

[14] G. Orrù, W. Pettersson-Yeo, A. F. Marquand, G. Sartori, and A. Mechelli. "Using Support Vector Machine to Identify Imaging Biomarkers of Neurological and Psychiatric Disease: A Critical Review." Neuroscience and Biobehavioral Reviews 36 (2012) 1140–1152.

[15] L. Jing. "Mental Health Evaluation Model based on Fuzzy Neural Network." 2016 International Conference on Smart Grid and Electrical Automation (2016).

https://journals.researchparks.org/index.php/IJHCS e-ISSN: 2615-8159 | p-ISSN: 2615-1898
Volume: 03 Issue: 6 |Aug 2021

[16] F. Sadeque, D. Xu, and S. Bethard. "UArizona at the CLEF eRisk 2017 Pilot Task: Linear and Recurrent Models for Early Depression Detection." In CEUR workshop proceedings, volume 1866. NIH Public Access (2017).

[17] M. Srividya, S. Mohanavalli, and N. Bhalaji. "Behavioral Modelling for Mental Health using Machine Learning Algorithms." Journal of Medical Systems (2018). https://doi.org/10.1007/s10916-018-0934-5.

[18] A. Khan, M. H. Husain, and A. Khan. "Analysis of Mental State of Users using Social Media to Predict Depression! A Survey." Conference Paper at the International Journal of Advanced Research in Computer Science (2018). 9(2):0976-5697.