International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 11 Issue 3 ǁ March 2023 ǁ PP. 528-533

Machine learning based healthcare system for investigating the association between depression and quality of life

Y. Sowjanya[1],K .Madhavi Reddy[2], G.Radhika[3], Y.Ragasree[4]
Department of Information Technology
Anurag Group of Institutions

ABSTRACT: *New technological advancements are reshaping the healthcare system's future. Depression is becoming a major public health problem across the globe. The identification of elements that cause depression may lead to novel research and therapies. The machine learning techniques are used to predict the reason for depression. This paper proposes a comprehensive methodological framework to find relationship between quality of life and depression. The proposed framework has two phases namely data consolidation and prediction. The data connection is constructed, and the Secure Hash Algorithm idea is used to uniquely identify each data relation. Hashing is used to find and index the data's real objects. The second section offered a model that used both unsupervised and supervised machine learning approaches. The consolidation technique is aided in the creation and confirmation of the research hypothesis. The self organizing map generated 08 cluster solutions, and the classification problems were drawn from the clustered data to verify the performance of the posterior probability multi-class Support Vector Machine. The suggested model was used to enhance classification performance, yielding a classification accuracy of 91.16%.*

Keywords *–Depression, healthcare, quality of life, secure hash algorithm SHA-1, supervised learning, unsupervised learning.*

Date of Submission: 14-03-2023 Date of acceptance: 29-03-2023

I. INTRODUCTION

Healthcare is one of the most pressing issues confronting the whole globe, regardless of whether the nation is developing or developed. Smart, efficient, and secure healthcare systems are being created as a major focus globally in order to improve people's quality of life. Early studies of human behavior drew academics from several backgrounds to engage in the disciplines of psychology and neuroscience. The same is true for the expanding area of computer science and machine learningresearch[1].Identifying a patient's mental health difficulties is a persistent difficulty for physicians and healthcare organizations, particularly among younger individuals[2] .Depression causes insecure feelings and bad thinking. It has an impact on the depressed person's behaviour and physical health more frequently, depression is identified as one of the main factors contributing to suicide and drug usage[12].Recent breakthroughs in the area of machine learning and deep learning have shown their ability to spot psychiatric illnesses in people as well as the effect of such disorders on their lifestyle. The most significant growth-related shift among individuals all across the globe is a change in mental health. As a result, sadness and anxiety are regarded as the two most serious age-related diseases[4]. Both have a negative impact on patients' quality of life (QoL) and impair the decision-making system, resulting in a high degree of suffering and, in the end, a suicide attempt. "Depression" is regarded as one of the most complicated and dangerous psychological issues, with a negative effect and being the leading source of illness burden among all disorders [5]. As a result, many scholars and medical personnel have directed their research efforts on the study of depression. Depression, according to the World Health Organization, will be one of the top causes of mortality and disease by 2030[8]. After effective therapy, the influence of depression remains and continues its fight in reducing performance and overall impairing an individual's QoL.

The word "Quality of Existence" refers to several aspects of a person's life, such as emotional, physical, and psychological well-being. These characteristics describe an individual's life experience and are being studied by many researchers and health specialists. Similarly,focuses on two points: (1) gaining information of the linkages between quality of life, nutritional status, and depression in patients; and (2) discovering the potential to improve patients' quality of life and psychosocial outcomes [10]. Some people never discuss their personal troubles, such as financial difficulties, sex lives, and personal relationships, until they are questioned. As a result, assessing QoL traits reveals underlying difficulties in these patients and leads to better treatment. Numerous researches have been conducted in the past 30 years to examine QoL in a variety of medical fields [11]. Determine

the importance of QoL in connection to mental issues, particularly in psychology and psychiatry. Machine learning has the potential to explore for QoL factors in a wider method to uncover the relationship between depression and QoL qualities [6].

II. LITERATURE REVIEW

Big data and machine learning are increasingly being used in health sciences research. Big data includes genomics and other "omic" categories, as well as administrative, biological, medicinal, environmental, sociodemographic, and social media information. Machine learning is a technique for evaluating vast volumes of data by spotting interactions between properties in patterns.Machine learning algorithms enable predictions and clinical result segmentation at the individual subject level, in contrast to traditional statistical approaches [1]. The purpose of this study is to synthesize the literature on machine learning (ML) for mental health, emphasizing current research and practical implementations.

There were 300 papers on the application of machine learning to mental health, with four key application fields being identified: I detection and diagnosis; (ii) prognosis, treatment, and support; (iii) public health; and (iv) research and clinical administration. The most often treated mental health conditions were schizophrenia, depression, and Alzheimer's disease. ML techniques that were used include support vector machines, decision trees, neural networks, and clustering. The application of ML in mental health has demonstrated a number of benefits in diagnosis, therapy and support, research, and clinical administration. There is also much of room to apply ML to other branches of psychology and mental health [2]. The worrying trend of rising mental health disorders, along with the worldwide failure to discover effective solutions, is impeding both individual and society success [3].

The study used the Hospital Anxiety and Depression Scale (HADS) to assess the presence of anxiety and depression in cancer inpatients receiving palliative care at an oncology department, and to determine whether anxiety and depression contribute to a lower quality of life when pain and illness severity were controlled for. The participants in this c[3oss-sectional research were 225 advanced cancer patients (a mean age of 65.1 years). The HADS, EORTC QLQ-C30, and Karnofsky Performance Status scale were used to gather data. Anxiety was discovered in 33.9% of patients, while depression (HADS-d 8) was observed in 47.6%. Anxiety, depression, and all quality of life characteristics were shown to have correlations ($r = 0.31$-0.63).Anxiety and sadness, according to multiple regression analysis, correlate to worse physical and emotional functioning. The findings show that when a cut-off of 8 is employed, there is considerable agreement between the MADRS and the HADS in identifying depressed individuals in 151 patients with diverse cancer pathologies. Managing anxiety and depression may help to enhance some aspects of quality of life[4][5] From the above study, we can infer that there is some relationship between depression and QoL.Older age, lower level of education, lower income, unemployment, a worse subjective sense of health, obesity, and problems with mental health were linked to QoL impairments in depressive people[6].Quality of life is a multidimensional concept that includes psychological, social, environmental and physical health, and is influenced by social support [7].Therefore it is important to understand the relationship between depression and QoL helps patients with depression achieve better long-term results and experience less disability[6].

The ML-based depression detection algorithms are categorized into three classes, classification, deep learning, and ensemblemodels. To find the association between QoL and Depression Classification algorithms like Random Forest, SVM, DecisionTrees ,Naïve Bayes, PPMCSVM are been used to in order to test the model .

III. METHODOLOGY

Support Vector Machine (SVM) is a fast and reliable classification technique that is utilized in a variety of real-world applications. SVM's performance and efficiency are mostly determined by its parameters. When compared to other classification algorithms, SVM has a greater recognition rate. Vapnik presented the SVM theory at Bell labs. Though SVMs were initially designed to be binary classifiers, there have been various suggested adaptations to support multiclass classification. Additional parameters and restrictions are introduced to the optimization problem in these additions to manage the separation of the distinct classes. The most frequent approaches for applying SVMs to multi-class classification issues divide multi-class problems into two-class problems. Because of its ideal learning efficiency, SVM has been a research hotspot in machine learning in recent years.

Disadvantages:

1. The model's inefficiency.
2. The identification of elements that cause depression may lead to novel research and therapies.
3. Because depression is becoming a major public health problem across the globe.

Proposed System

Using machine learning methods, this paper proposes a comprehensive methodological framework presented in figure1- for processing and exploring heterogeneous data in order to better understand the relationship between quality of life and depression. As a result, the experimental investigation is separated into two sections. The first section describes a data consolidation procedure. The data connection is constructed, and the Secure Hash Algorithm idea is used to uniquely identify each data relation. Hashing is used to find and index the data's real objects. The second section offered a model that used both unsupervised and supervised machine learning approaches. The consolidation technique aided in the creation and confirmation of the research hypothesis. The self organizing map generated 08 cluster solutions, and the classification problems were drawn from the clustered data to verify the performance of the posterior probability multi-class Support Vector Machine.

Advantages:

1. The suggested model was used to enhance classification performance.
2. The results demonstrated that our suggested model has a good level of classification accuracy.

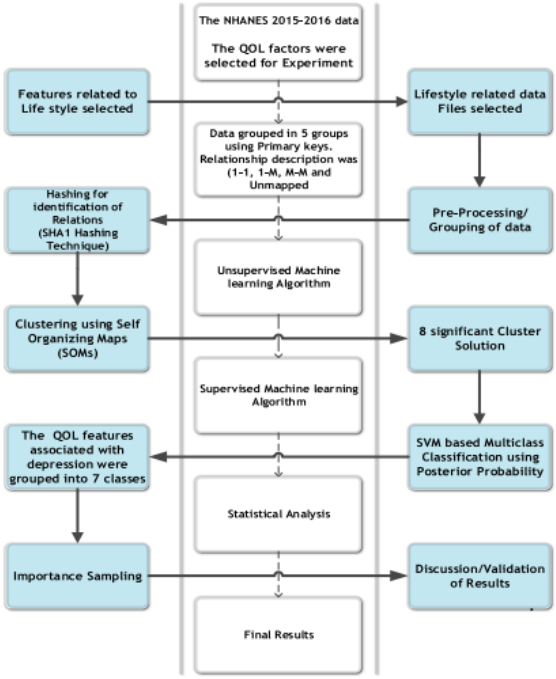

fig.1: system architecture

The methodology entails a number of steps, including as data extraction, pre-processing of the extracted data, feature extraction methods for choosing the necessary set of characteristics for identifying depressive symptoms, and ML classifiers for categorizing the input data [9].

MODULES:

To carry out the aforementioned process, we created the modules listed below.

- **Data collection**: Data is collected based on the QoL factors [12]. The NHANES 2015–2016 data was chosen for this study because it comprises QoL variables that are appropriate for our analysis and is a large dataset with 5400 data point .This Dataset contains QoL variables like alcoholic, smoking, age etc.
- **Data exploration**: This module reads the data and helps to find the missing values, outliers, colinearity and patterns in the dataset.
- **Data split**: In this phase, data split into training and testing dataset.
- **Models used**: Various machine learning algorithms such as Random Forest, Decision Tree, Naive Bayes, Support Vector Machine, PPMCSVM used to predict the depression in a person based on QoL factors.
- **Prediction:**The final projected value presented by using PPMCSVM

IV. IMPLEMENTATION

ALGORITHMS:

Random Forest:A Supervised Machine Learning Algorithm that is commonly utilized in Classification and Regression applications. An ensemble of decision trees that have not been trimmed is known as a random forest.

When we have very big training datasets and a lot of input variables, random forests are frequently used. A classifier known as a random forest model is made up of numerous decision trees and produces classes that are the mode of the classes produced by individual trees [13].
1. Assume that there are M variables in the classifier and N training cases.
2. We are informed of the m input variables that will be used to decide a node's choice; m should be substantially smaller than M.
3. Pick N times with replacement as the training set for this treefrom all N training instances available (i.e. take a bootstrap sample). By anticipating their classes, use the remaining cases to estimate the tree's error.
4. Select m random variables for each node in the tree to use as the basis for that node's decision.Based on these m training-set variables, determine the optimal split.
5. No trees have been clipped and are completely grown (as may be done in constructing a normal tree classifier).
Decision Tree: A decision tree is an algorithm for classifying features with numerous layers of observations by creating a tree-like structure. The "leaves" of the substructures denote the class of the objects, whilst the "branches" denote the features [14].Decision trees use numerous methods to determine whether or not to divide a node into two or more sub-nodes. The development of sub-nodes promotes the homogeneity of the sub-nodes that arise. In other words, the purity of the node rises in relation to the target variable.
NaiveBayes: This classifier operates on the assumption that features are analytically independent and supervises machine learning processes using the Bayes theorem. The nave assumption that the input variables are independent of one another [15]. Naive and Bayes, which together make up the Nave Bayes algorithm, are defined as follows: Naive: It is so called because it makes the assumption that the existence of one feature is unrelated to the occurrence of other features. It is known as Bayes because it utilizes the Bayes' Theorem .Bayes' Theorem: Also referred to as Bayes' Rule or Bayes' Law, Bayes' Theorem is a formula that can be used to estimate a hypothesis' likelihood given previous information.The posterior probability, P(A|B), measures the likelihood that a given hypothesis (A) will actually occur.The probability of likelihood is P(B|A): the likelihood that the evidence provided supports a certain theory.
SVM: Support-vector machines (also known as support-vector networks) are supervised learning models with corresponding learning algorithms that examine data for regression and classification. Support vector machines (SVMs) are a group of supervised learning techniques for classifying data, performing regression analysis, and identifying outliers. The benefits of support vector machines are as follows: efficient in high-dimensional environments. Still useful in situations where there are more dimensions than samples. A supervised machine learning technique called the Support Vector Machine (SVM) may carry out classification, regression, and even outlier detection. The linear SVM classifier functions by connecting two classes using a straight line[16].

```
# Building a Support Vector Machine on train data
Svm1_model =SVM(kernel='rbf')
Svm1_model.fit(X_train, y_train)
```

Performance metrics Used:
Accuracy: One parameter for assessing classification models is accuracy. The percentage of predictions that our model correctly predicted is known as accuracy. The following is the official definition of accuracy:

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

Where,
TP=True Positives, TN=True Negatives, FP=False Positives, FN=False Negatives

Precision: The proportion of accurately categorised positive samples (True Positive) to the total number of positively classified samples is known as precision (either correctly or incorrectly).

$$Precision = \frac{TP}{TP + FP}$$

Where,
TP=True Positives,FP=False Positives
Recall: The recall is determined as the proportion of Positive samples that were correctly identified as Positive to all Positive samples**.**

$$Recall = \frac{TP}{TP + FN}$$

Where,
TP=True Positives, FN=False Positives

V. EXPERIMENTAL RESULTS

ALGORTHIM	ACCURACY	PRECISION	RECALL
RANDOM FOREST	100.0	1.0	1.0
DECISION TREE	100.0	1.0	1.0
NAÏVE BAYES	93.35	0.96	0.91
SVM	100.0	1.0	1.0
PPMCSVM	95.54	0.96	0.95

fig2: performance metrics table

From fig2, we can see that PPMCSVM is the best model suited for the classification as it has the best accuracy, recall, and precision metrics, while others are falling into overfitting conditions.

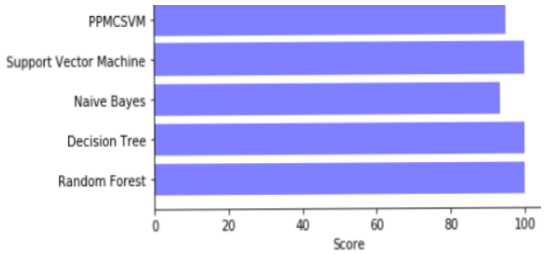

Fig3: Classification Performance Graph

VI. CONCLUSION

Identifying the mental health issues in a person is difficult to physicians and healthcare organizations. Using NHANES data, our results reveal that there is a relationship between depression symptoms and specific quality of life indicators. Clustering is the initial stage in grouping factors linked to quality of life in order to better investigate the varied data connected with mental health concerns. SOM restructures the clusters, and the data in the classes is readjusted to achieve a balanced collection of groups. PPMCSVM is an extension of conventional SVM that is used to take classification issues and estimate the correlation between data points. The multi-class model ECOC was trained using SVM. Based on the results, the suggested model may successfully predict elements that cause depression and given accuracy of 91.6%. It is important to characterize the correlation between the two distributions, and sampling is used to choose the final samples from the aforesaid groupings. The samples are chosen based on the sampling ratio, which serves as a correction factor to compensate for the probability sampling from the distribution. Finally, this research discovered a link between the expected quality of life indicators and depression. Comparisons with the most renowned state-of-the-art approaches and previously published work show that the suggested strategy produces much more trustworthy findings. However, further research is needed to enrich the dataset, which contains numerous characteristics, and determine the severity levels of depression based on many parameters. For example, which component has a low, medium, or high degree of depression? This will aid other researchers and doctors in recognizing risk factors for depression and other psychiatric illnesses.

REFERENCES

[1]. Passos, Ives Cavalcante, Pedro Ballester, Jairo Vinicius Pinto, Benson Mwangi, and Flávio Kapczinski. "Big data and machine learning meet the health sciences." Personalized psychiatry: big data analytics in mental health (2019): 1-13.

[2]. Shatte, A. B. "hutchinson dM, Teague SJ." Machine learning in mental health: a scoping review of methods and applications. Psychol Med 49 (2019):1426-48. doi:10.1017 /S003329171 9000151.

[3]. Kolenik, Tine, and Matjaž Gams. "Persuasive technology for mental health: One step closer to (Mental health care) equality?." IEEE Technology and Society Magazine 40, no. 1 (2021): 80-86. doi: 10.1109/MTS.2021.3056288.

[4]. Ahmad, Roslee, Azizan Mohd Isa, Mohd Radzniwan A. Rashid, Nasir Yusoff, Lokman Mohd Tahir, Miftachul Huda, and Rahmiwati Marsinun. "Prevalence of Anxiety and Depression among Caregivers of Cancer Patients A Case Study in a Public Hospital in Malaysia." IIUM Medical Journal Malaysia 22, no. 1 (2023). Doi:10.31436/imjm.v22i1.2146.

[5]. Yohannes, Abebaw M., Martino Belvederi Murri, Nicola A. Hanania, Elizabeth A. Regan, Anand Iyer, Surya P. Bhatt, Victor Kim et al. "Depressive and anxiety symptoms in patients with COPD: A network analysis." Respiratory medicine 198(2022):106865.doi:10.1016/j.rmed.2022.106865.

[6]. Cho, Yunji, Joo Kyung Lee, Do-Hoon Kim, Joo-Hyun Park, Moonyoung Choi, Hyun-Jin Kim, Myung-Ji Nam, Kang-Uk Lee, Kyungdo Han, and Yong-Gyu Park. "Factors associated with quality of life in patients with depression: A nationwide population-based study." PloS one 14, no. 7 (2019):e0219455. doi:10.1371/journal.pone.0219455

[7]. M. M. Alsubaie, H. J. Stain, L. A. D. Webster & R. Wadman (2019) The role of sources of social support on depression and quality of life for university student International Journal of Adolescence and Youth, 24:4, 484-496, DOI: 10.1080/02673843.2019.1568887

[8]. Mathers, C. D., and D. Loncar, "Projections of global mortality and burden of disease from 2002 to 2030," PLoS Med., vol. 3, no. 11, pp. 2011–2030, 2006, doi: 10.1371/journal.pmed.0030442.

[9]. Aleem, Shumaila, Noor ul Huda, Rashid Amin, Samina Khalid, Sultan S. Alshamrani, and Abdullah Alshehri. "Machine learning algorithms for depression: diagnosis, insights, and research directions." Electronics 11, no. 7 (2022): 1111. doi: org/10.3390/electronics1107111

[10]. Daniel, Shawona C., Andres Azuero, Orlando M. Gutierrez, and Karen Heaton. "Examining the relationship between nutrition, quality of life, and depression in hemodialysis patients." Quality of Life Research 30 (2021): 759-768. doi: 10.1007/s11136-020-02684-2.

[11]. Narayanrao, Purude Vaishali, and P. Lalitha Surya Kumari. "Analysis of machine learning algorithms for predicting depression." In 2020 international conference on computer science, engineering and applications (iccsea), pp. 1-4. IEEE, 2020. doi: 10.1109/ICCSEA49143.2020.9132963.

[12]. Kumar, Prince, Shruti Garg, and Ashwani Garg. "Assessment of anxiety, depression and stress using machine learning models." Procedia Computer Science 171 (2020): 1989-1998. doi:10.1016/j.procs.2020.04.213

[13]. Gao D, Zhang YX, Zhao YH. Random forest algorithm for classification of multiwavelength data. Research in Astronomy and Astrophysics. 2009 Feb 1; 9(2):220. Doi: 10.1088/1674-4527/9/2/011

[14]. Qasrawi, Radwan, Stephanny Vicuna Polo, Diala Abu Al-Halawa, Sameh Hallaq, and Ziad Abdeen. "School children' depression and anxiety risk factors assessment and prediction: Machine learning techniques performance analysis." (2022). doi: 10.2196/32736

[15]. Priya, Anu, Shruti Garg, and Neha Prerna Tigga. "Predicting anxiety, depression and stress in modern life using machine learning algorithms." Procedia Computer Science 167 (2020):1258-1267. doi:10.1016/j.procs.2020.03.442.

[16]. Prof. P.M. Kamde, Shantanu Ramteke ,Ruturaj Kadam , GunjHundiwala, Suryansh "SVM Classification Technique to Analyze Mental Health and Stress Levels", Volume 2, Issue 7, May 2022, DOI: *10.48175/IJARSCT-4368*