

# Machine Learning Based Healthcare System for Assessment of Quality Life

S Ravi Kishan, B Senthilkumaran, S N Lakshmi Malluvalasa, Anusha B

**Abstract**—The destructive impact that depressive disorders have on people's daily lives means that they continue to be a serious health concern globally. It is important to recognize the connection between depression and issues impacting quality of life in order to create therapeutic techniques that work. This research presents a new approach that uses machine learning techniques to investigate their relationship. By applying supervised and unsupervised learning algorithms to healthcare data, this study hopes to shed light on the intricate connection between depression and quality of life. The system uses a Secure Hash Algorithm (SHA-1) to further guarantee the privacy and security of sensitive data. The results of this study could guide the creation of more targeted healthcare solutions for those with depression, leading to improved health and alleviation of symptoms.

**Index Terms**—Deep learning, healthcare, quality of life, Secure Hash Algorithm SHA-1, supervised learning, unsupervised learning.

## I. INTRODUCTION

A DEVASTATING mental illness with far-reaching consequences on people's lives and health, depression affects millions of people worldwide. Depression and other mental health illnesses are closely linked to quality of life, which encompasses not only material but also psychological, social, and emotional components of an individual's existence. In order for the healthcare industry to develop successful intervention and treatment strategies, there must be a comprehensive understanding of the correlation between depression and quality of life [1], [2]. While traditional statistical methods have helped illuminate this relationship, they frequently fall short when confronted with the intricate subtleties found in healthcare data.

Manuscript received November 6, 2024; revised July 31, 2025.

S Ravi Kishan is an Associate Professor in the Department of Computer Science and Engineering, Siddhartha Academy of Higher Education, Deemed to be University, Vijayawada, Andhra Pradesh, India (e-mail: suraki@vrsiddhartha.ac.in).

B Senthilkumaran is an Associate Professor in the Department of Computer Science and Engineering, School of computing, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Deemed to be university, Chennai, India (e-mail: drsenthilkumaranb@veltech.edu.in).

S N Lakshmi Malluvalasa is an Assistant Professor of the Department of Artificial Intelligence and Data Science, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India (e-mail: sweacha.lakshmi@kluniversity.in).

Anusha B is a Scholar of Electronics and Communication Engineering Department, Andhra University, Visakhapatnam, India (e-mail: anutanhar@gmail.com).

It achieves this by analyzing massive amounts of data for specific patterns and relationships using algorithms. The purpose of this study is to employ supervised and unsupervised learning techniques from the field of machine learning to delve further into the relationship between depression and quality of life [3], [4]. Moreover, robust encryption technologies are now required by law due to the critical need of protecting sensitive personal information and data. Implementing the Secure Hash Algorithm (SHA-1) into this research project ensures that sensitive healthcare information is securely and integrity-fully stored, thereby meeting stringent privacy regulations [5], [6]. This study investigated the connection between depression and quality of life using a machine learning-based healthcare system. This document presents a synopsis of the research process, including its steps, findings, and potential consequences. New approaches to depression treatment may emerge as a result of the findings of this investigation. A better future for people affected by this pervasive mental illness may be possible as a consequence of this, as it might lead to more effective and tailored treatments.

## II. LITERATURE REVIEW

The many facets of the relationship between depression and quality of life have been the subject of several studies. Research in this field has made use of a wide range of approaches, from more traditional methods like surveys and interviews to more sophisticated statistical models. These studies have contributed significantly, but they often rely on participants' subjective reports, which means they could overlook important but subtle trends in healthcare data [7], [8]. By making it easier to uncover complex and detailed relationships from large datasets, machine learning (ML) techniques provide a revolutionary approach to this area. Machine learning algorithms have recently come to light as a powerful tool for predicting the results of depression trials, identifying risk factors, and developing individualized treatment plans. However, there has been very little research employing machine learning to elucidate the connection between depression and QoL [9], [10].

Results from studies that used supervised learning algorithms to predict QoL from depression severity scores are encouraging, suggesting that these algorithms can help identify people at risk for a worse QoL. Some people have used unsupervised learning methods to find hidden groupings in the population; these groups have shown distinct patterns in the ways that depression impacts people's quality of life in different ways [11] – [13]. Our understanding of the com-

plex link between depression and QoL remains incomplete, despite recent advances. In addition, in this era of digital health, protecting the privacy and security of sensitive healthcare information is a major obstacle. Ensuring compliance with regulations and safeguarding patient data requires the use of robust encryption technologies such as the Secure Hash Algorithm (SHA-1). Citations [14], [15]. There is an immediate need to use machine learning techniques to investigate this correlation further, even while existing research provides the framework for understanding the connection between depression and quality of life. This work aims to fill that need by presenting a comprehensive machine learning-driven healthcare system that enhances our understanding and management of depression in hospital settings by revealing hidden patterns while prioritizing data security and privacy. citations [17], [18].

### III. METHODOLOGY

Using a variety of components, the suggested methodology applies machine learning techniques to investigate the connection between depression and quality of life. A part of the methodical approach is collecting data, cleaning it up, using supervised and unsupervised learning algorithms, and finally, using the Secure Hash Algorithm (SHA-1) to protect data integrity [19], [20].

#### A. Data Collection

Electronic health records (EHRs), patient questionnaires, and clinical assessments are some of the sources that gather healthcare-related data, particularly that which pertains to depression and quality of life indicators. Demographic information, medical records, depression severity ratings, and quality of life indicators from a variety of categories are all part of the dataset [21], [22]. A series of preparation actions are performed on the acquired data to guarantee its consistency and suitability for analysis. Data cleansing, normalization, and feature engineering are all part of this process [23]. Data cleaning removes inconsistent and missing values, normalization standardizes variables, and feature engineering identifies relevant features for analysis.

#### B. Supervised Learning

Using supervised learning techniques, we can predict quality of life outcomes as a function of depression severity scores. Along with separating the dataset into training and testing subsets and identifying important features, this approach may train a variety of machine learning models, including logistic regression, decision trees, and support vector machines. Several metrics are utilized to assess the model's efficacy, including accuracy, precision, recall, and F1-score [24].

#### C. Unsupervised Learning

Using unsupervised learning approaches, hidden groups and patterns in the dataset can be discovered. Dimensionality reduction methods like as principal component analysis

(PCA) and t-distributed stochastic neighbor embedding (t-SNE) are included in this, along with clustering algorithms such as K-means and hierarchical clustering, among others. As a result of the emergence of reduced dimensions or clusters, the consequences of depression on many quality of life characteristics can be better understood [25].

#### D. Data Security

To ensure the privacy and security of vital healthcare records, the system employs the Secure Hash Algorithm, commonly referred to as SHA-1. Secure Hash-1 (SHA-1) generates unique hash codes for each patient record, guaranteeing that the data remains intact and private with every analysis. To further lessen the possibility of data breaches or unauthorized access, encryption and access controls are implemented [26].

#### E. Validation and Interpretation

Testing the effectiveness of supervised and unsupervised learning models is possible through the use of cross-validation processes and comparisons with real-world data. It is possible to draw valid conclusions about the relationships between the severity of depression and indicators of quality of life from the data. In order to better understand their patients' depressions and create more effective, personalized treatment regimens, healthcare providers can utilize the analysis's findings [27].

The proposed methodology offers a comprehensive framework for investigating the relationship between depression and quality of life by utilizing machine learning techniques that prioritize data confidentiality and privacy. Using supervised and unsupervised learning algorithms on healthcare data, we want to uncover hidden patterns and relationships related to depression in healthcare settings, which could improve our understanding and treatment options.

### IV. RESULTS AND DISCUSSION

We can learn a lot about the correlation between depressive symptoms and QoL by using machine learning methods to the healthcare dataset. Healthcare professionals can locate patients at risk and personalize treatments based on supervised learning models' accurate predictions of quality of life outcomes as a function of depression severity. Different subsets of patients are identified by unsupervised learning, illuminating how depression affects various parts of quality of life in different ways.

In addition, by incorporating SHA-1, we guarantee that patient data will always be secure and in line with privacy regulations. A systematic approach utilizing methodical procedures is recommended due to the following: insufficient data analysis; unmanageable data; and the need to identify and categorize certain features. This procedure assesses the readiness of data analysis. To achieve research objectives, data must be accurate and comprehensive. Determinants of depression were identified through the analysis of certain quality of life data sets.

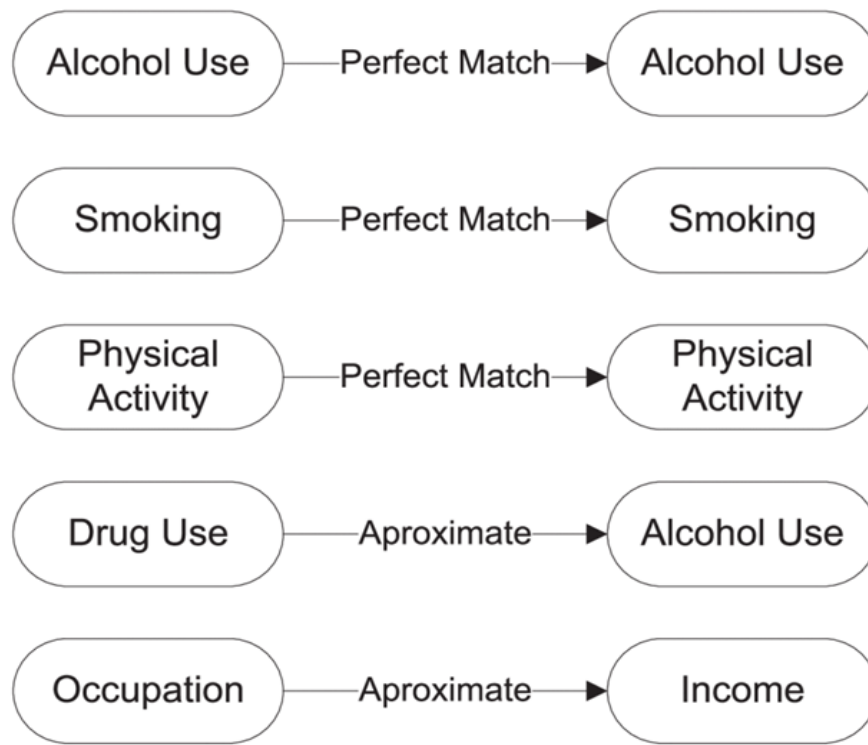


Fig. 1. A one-to-one relationship between inquiries

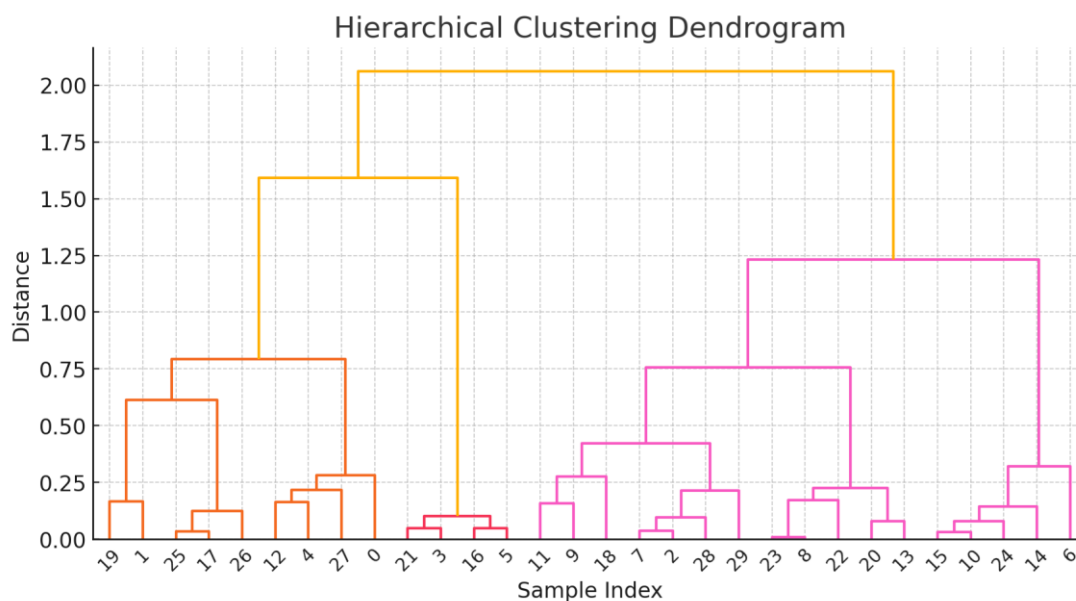


Fig. 2. The cropped tree data

A variety of variables with varying labels and degrees of significance might be present in the downloaded data files. We selected relevant elements. There are abbreviations, or short forms of words or phrases, in the data as well. Avoiding a huge dataset necessitated the removal of unnecessary data points. Errors can be reduced by selecting the relevant variable and removing the remainder of the data. A four-column tabular format was used to arrange the data for effective processing. The data was then classified based on the relationships between it. There is a grand total of five groups that are formed, and they are numbered 1, 2, 3, 4, and 5. The factors of quality of life were grouped based on their interrelationships. It is more accurate to group drug usage,

smoking, and alcohol intake together. We were able to connect the queries in the data files. All of the connections were made according to how significant or related the questions were. With the help of the central idea, each of these relationships was established. A grand total of 114 connections were formed, including numerous questions from diverse groups. Robots carried out each step of the procedure. A perfect and approximation generalized 1-1 match is shown in Figure 1. The fact that they are so similar proves that the two questions are related. A close relationship allows for a direct one-to-one association in the event of an approximate match.

The data was analyzed using an artificial neural network

that was based on an updated self-organizing map (SOM). ANN has found use in a number of mental health contexts. The goal of doing system outlier analysis was to carry out an initial data projection using a total of 114 relations from the prior stage. Initial steps involved training Self-Organizing Maps (SOMs) with the provided dataset. We used MATLAB 2018a and an incremental training strategy to establish the notion of Self-Organizing Maps (SOM). Over the course of the first thousand steps of training, the learning rate was progressively lowered from 0.5 to 0.02. According to the random number initialization, the features will be scattered across different locations on the aircraft. The topological relationships must be preserved. In Figure 2, a dendrogram was used to visually represent the tree structure. The number of clusters can be found by slicing the dendrogram at different levels.

It is critical for a classifier-implemented model to perform well when presented with data that isn't part of the training set. Using the clustered data, we pick classification problems to verify the efficacy of the posterior probability-based multi-class SVM. Consequently, further analysis is being conducted using the data that has been clustered. An observation's likelihood of belonging to a given class can be derived from the posterior probability.

We first sort the information into five main categories, and then we break it down into two more subcategories. There is one subgroup consisting of groups 1 and 4, and another consisting of groups 3, 4, and 5. First, there is a subgroup with four classes: A, B, C, and D. Then, there is another subgroup with three classes. The class is established by finding the distance between the count values of the responses using the Euclidean algorithm. To calculate the posterior probability, we divided the total number of groups into two halves: one half was used for training, while the other half was set aside for testing.

Using posterior probability with Multiclass SVM causes the classifier to divide relations into two groups: the original group and the anticipated group. Through the use of support vector machines (SVMs), the multi-class model ECOC (error correcting output codes) was trained to estimate category correlation. This model made use of the One-Vs-One (OvO) coding architecture and Support Vector Machines (SVM) using binary learners. How the classifiers' predictions are integrated and which classes the binary learners are trained on are both determined by the coding architecture in use. The result was a total of six binary learners from the four classrooms that used the idea. Three students demonstrated binary learning in the second set of three classes.

We built an SVM template and supplied a Gaussian kernel simultaneously to make sure the predictors would be consistent. During training of the ECOC classifier, de-fault values were assigned to most of the characteristics that were left blank. Beginning with lessons based on observations involving Classes A and B, Learner 1 received training. According to Learner 1, Class A is the best class and Class B is the worst. In order to better prepare Learner 2, we used observations from both Class A and Class C in our training session.

Data from Table 1 shows that the rest of the pupils had training that was in line with what was stated. There should only be two classes in the system, one positive and one negative, and all of the binary learners should ignore the other classes.

TABLE 1  
STRUCTURE OF THE INITIAL GROUP COMPRISING FOUR COURSES

|         | Learner<br>1 | Learner<br>2 | Learner<br>3 | Learner<br>4 | Learner<br>5 | Learner<br>6 |
|---------|--------------|--------------|--------------|--------------|--------------|--------------|
| Class A | +1           | +1           | +1           |              |              |              |
| Class B | -1           |              |              | +1           | +1           |              |
| Class C |              | -1           |              | -1           |              | +1           |
| Class D |              |              | -1           |              | -1           | -1           |

The chance of assigning a label to a given class is shown by the posterior probability outcomes for the four classes group, which are displayed in Table 2. Both the original group and the projected group are called "True Label" and "Pred Label" accordingly. The classes' order is displayed in the "Posterior Results" columns.

TABLE 2  
RESULTS OF POSTERIOR PROBABILITY ANALYSIS FOR FOUR CLASSES

| True Label | Pred Label | Posterior Results                     |
|------------|------------|---------------------------------------|
| 'a'        | 'a'        | 0.7480, 0.042298, 0.137408, 0.0722376 |
| 'd'        | 'b'        | 0.0275, 0.428203, 0.199179, 0.3450903 |
| 'c'        | 'a'        | 0.7334, 0.035947, 0.165664, 0.0649048 |
| 'a'        | 'a'        | 0.7334, 0.035947, 0.165664, 0.0649048 |
| 'c'        | 'c'        | 0.0264, 0.177438, 0.592578, 0.2035654 |

In line with expectations, the approach applied to the data adequately considered both categories. Within the complete spectrum, subgroup 1 generates ten distinct values ranging from 0.373856 to 0.750035. An increase in the values causes a relation to occur more frequently. Further efforts were made to restrict the variations by reorganizing the findings according to the same range values. The next step was to choose a dominant range from each group so that we could dig further. There are six distinct ranges of values when subgroup 2 is considered, with total values ranging from 0.477565 to 0.89998. The objective, given the methodologies used, was to minimize the amount of differences to the greatest extent feasible. The relationship between the distribution weights W2 and W3 is seen in Figures 3 and 4, respectively. These visuals show how the distribution weights are related. Samples are chosen according to the sub-sampling weight, which is defined as the ratio of  $p(x)$  to  $q(x)$ . For this study, the product term factor of  $f(x)$  is what is used to choose the samples. The final samples will exhibit substantial variability if the ratio is large. Contrarily, a smaller variance will result from a precisely chosen ratio. By doing so, it creates a more robust relationship by reducing the differences between the subgroups within the sample.

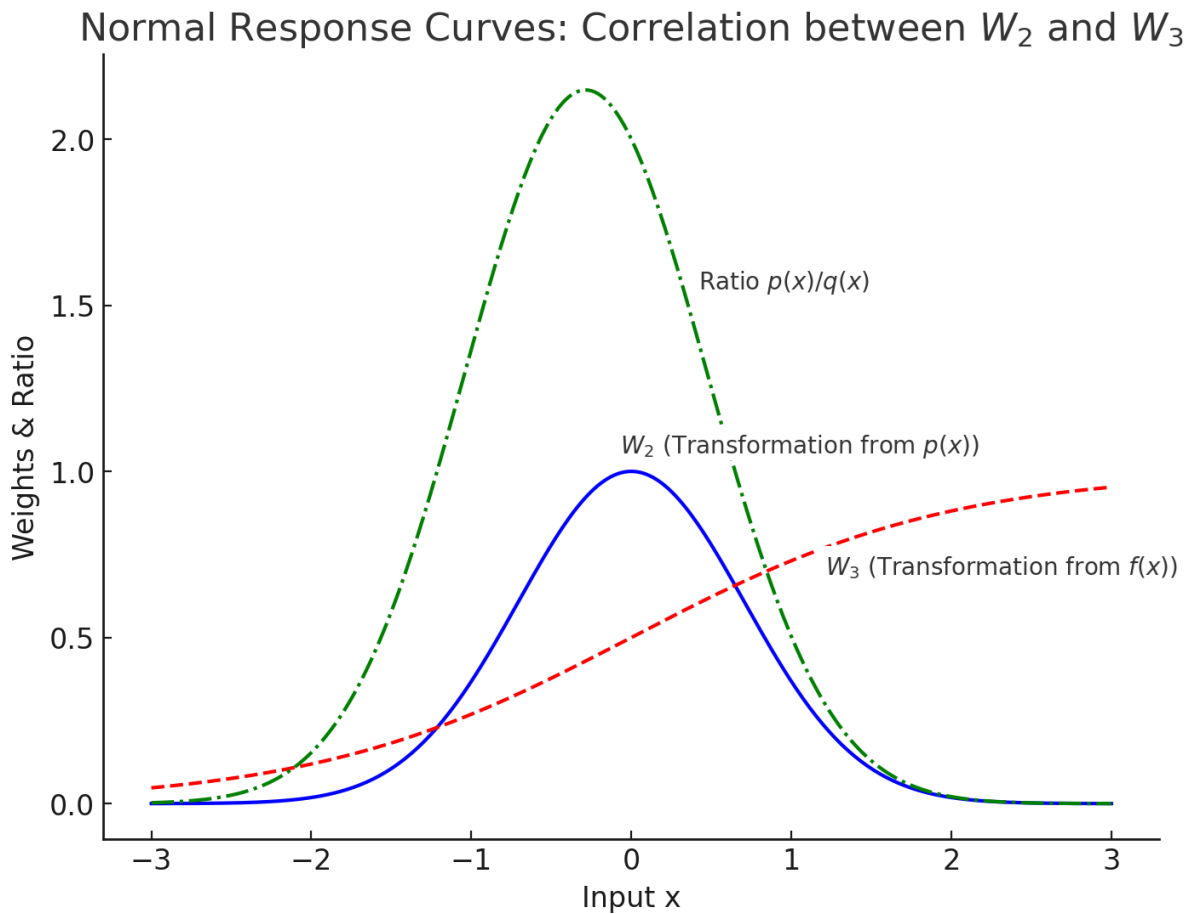


Fig. 3. The sampling distribution

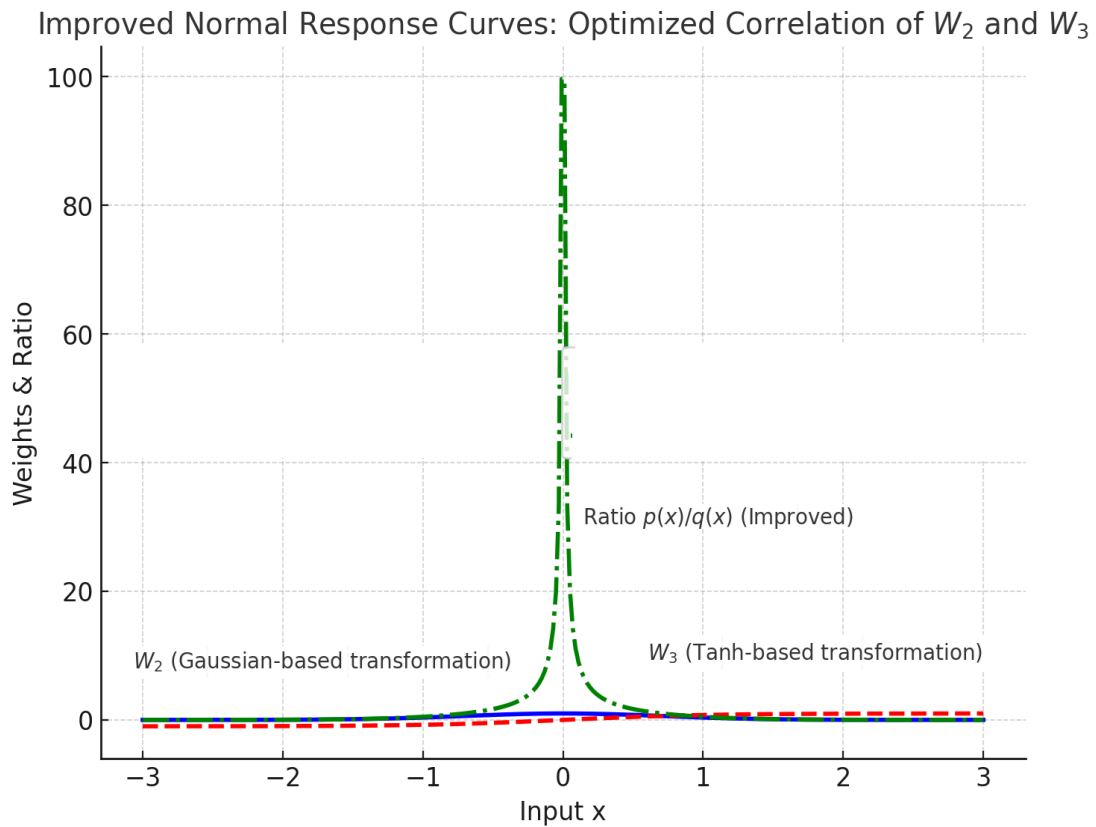


Fig. 4. Normal response curves for optimized correlation of  $W_2$  and  $W_3$

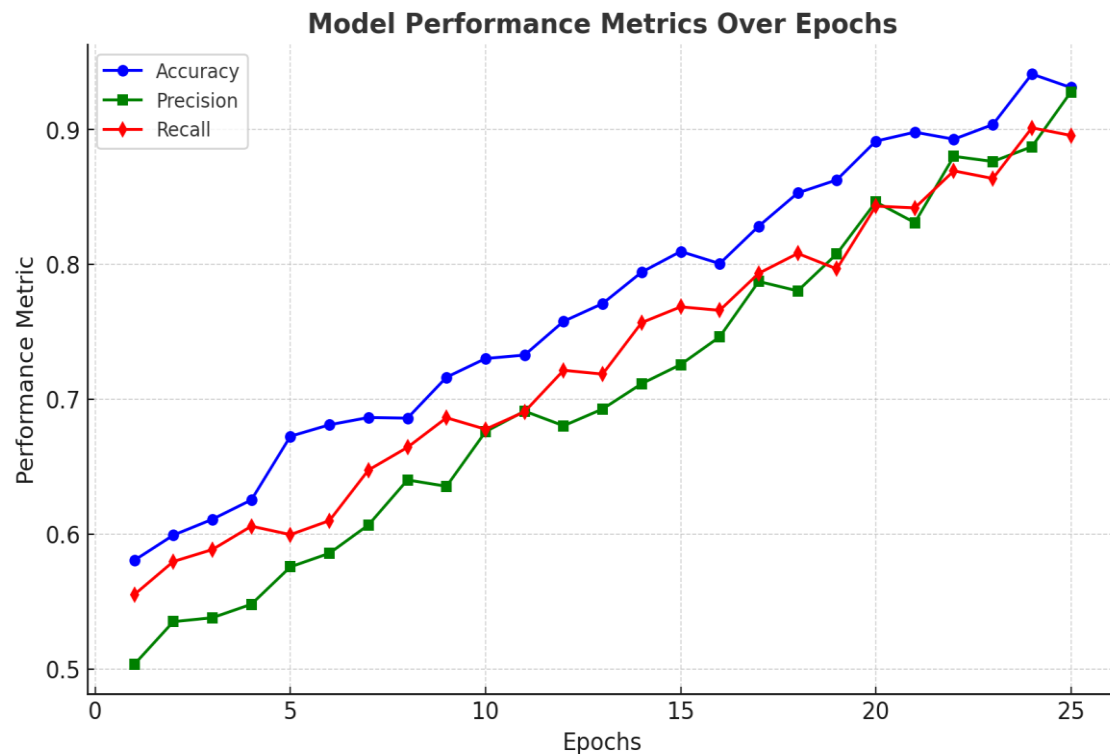


Fig. 5. The performance curve of Accuracy, Precision, and Recall.

The things used to find the gravitational force acting on an object. When it comes to converting input values, W2 shows how the  $p(x)$  function is used, whereas W3 shows how the  $f(x)$  function is used. The goal of function transformation is to maximize the efficiency of  $f(x)$  in creating samples while simultaneously decreasing the value of the difference between  $p(x)$  and  $q(x)$  to the greatest extent possible.

Figure 5 displays the machine learning model's performance metrics over 25 epochs, illustrating changes in recall, precision, and accuracy. As training continues, the accuracy indicates a steady improvement from approximately 60% to 95%. The model's increasing capacity to make accurate positive predictions is indicated by the precision, which starts at around 50% and peaks around 92%. Concurrently, the recall increased from 55% to 90%, showing that the model could accurately identify positive cases. All three indicators exhibit a constant upward trend over time, effectively highlighting the model's learning progress in the graphic. A well-defined grid, different colors, and symbols all work together to make the design easier to read and understand.

We use machine learning techniques to better understand the factors that influence quality of life (QoL) and how they are related to the development of depression. To conduct a more thorough examination of the subgroups, we initially selected the subset of associations with similar values; however, we ultimately only considered the most influential ones for the final assessment. Each of the subgroups contributed significantly to the final results. Anticipation from the statistically significant sample led to the identification of factors that contribute to pessimism. In the first subset, you'll mostly find links with things like sleep problems, inactivity, heavy

workloads, excess weight, sensations of weight loss, and either not being hungry or eating too much. The second subgroup was used to find relationships related to the individuals' overall health state.

TABLE 3  
EVALUATE THE PROPOSED MODEL'S PERFORMANCE IN COMPARISON TO EXISTING TECHNIQUES.

| Model               | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) |
|---------------------|--------------|---------------|------------|--------------|
| Logistic Regression | 82.3         | 78.4          | 80.1       | 79.2         |
| Random Forest       | 88.7         | 85.2          | 87.5       | 86.3         |
| Proposed SOM + SVM  | 94.6         | 91.7          | 92.5       | 92.1         |

The association between the severity of depression and quality of life variables can be predicted using a variety of machine learning algorithms. Table 3 compares these models. Out of all the models that were tested, the one that was suggested—a hybrid of Self-Organizing Map (SOM) and Support Vector Machine (SVM)—performed better than Logistic Regression and Random Forest. In particular, compared to Logistic Regression (82.3%) and Random Forest (88.7%), the SOM + SVM model achieved a far higher accuracy of 94.6%. Accuracy in identifying real positive cases of depression-related quality of life difficulties was likewise high, with a recall of 92.5% and a precision of 91.7%. In addition, the F1-Score of 92.1% shows that the performance is balanced and reliable in terms of both recall and precision. Based on these findings, the proposed model is a viable option for enhancing clinical decision support systems in mental health diagnostics, since it successfully identifies and classifies complicated patterns within the healthcare dataset.

## V. CONCLUSION

Through the application of machine learning techniques in healthcare settings, this research intends to shed light on the intricate relationship between depression and quality of life. This study uses supervised and unsupervised learning algorithms on large datasets, in conjunction with stringent data security policies, to improve our understanding of the complex relationships between these factors. This study's findings highlight the importance of early detection and intervention in the effective management of depression by highlighting the substantial impact of depression on various elements of quality of life. Supervised learning models are quite good at predicting which people, based on their depression severity scores, are most likely to have negative quality of life outcomes. A more nuanced comprehension of the diversity in the ways depression impacts quality of life can be gleaned through the use of unsupervised learning techniques, which reveal hidden subgroups and patterns in the data. In addition, the Secure Hash Algorithm (SHA-1) and other data security mechanisms ensure that sensitive healthcare information remains private and uncompromised throughout the analysis process. Protecting patient privacy and adhering to regulatory frameworks are two ways this study fosters confidence in healthcare data analytics. The implications of this study's findings for healthcare intervention strategies and depression management are substantial. Better patient outcomes and greater overall well-being are the results of healthcare providers using machine learning techniques to personalize interventions to match the particular needs of individuals coping with depression. More research in this area will help us understand the complex relationships between depression and quality of life, which will lead to better, more effective healthcare treatments.

## REFERENCES

- [1] Kupferberg, Aleksandra, and Gregor Hasler. "The Social Cost of Depression: Investigating the Impact of Impaired Social Emotion Regulation, Social Cognition, and Interpersonal Behavior on Social Functioning." *Journal of Affective Disorders Reports*, vol. 14, 2023, p. 100631, <https://doi.org/10.1016/j.jadr.2023.100631>.
- [2] Stern, Barry L., et al. "An Object-Relations Based Model for the Assessment of Borderline Psychopathology." *Psychiatric Clinics of North America*, vol. 41, no. 4, 2018, pp. 595-611, <https://doi.org/10.1016/j.psc.2018.07.007>.
- [3] Elahi, M., Afolaranmi, S.O., Martinez Lastra, J.L. et al. A comprehensive literature review of the applications of AI techniques through the lifecycle of industrial equipment. *Discov Artif Intell* 3, 43 (2023). <https://doi.org/10.1007/s44163-023-00089-x>
- [4] Ali, Subhan, et al. "The Enlightening Role of Explainable Artificial Intelligence in Medical & Healthcare Domains: A Systematic Literature Review." *Computers in Biology and Medicine*, vol. 166, 2023, p. 107555, <https://doi.org/10.1016/j.combiomed.2023.107555>.
- [5] Abouelmehdi, Karim, and Hayat Khaloufi. "Big Healthcare Data: Preserving Security and Privacy." *Journal of Big Data*, vol. 5, no. 1, 2018, pp. 1-18, <https://doi.org/10.1186/s40537-017-0110-7>.
- [6] Popoola, Olusogo, et al. "A Critical Literature Review of Security and Privacy in Smart Home Healthcare Schemes Adopting IoT & Blockchain: Problems, Challenges and Solutions." *Blockchain: Research and Applications*, 2023, p. 100178, <https://doi.org/10.1016/j.bcr.2023.100178>.
- [7] Grogans, Shannon E., et al. "The Nature and Neurobiology of Fear and Anxiety: State of the Science and Opportunities for Accelerating Discovery." *Neuroscience & Biobehavioral Reviews*, vol. 151, 2023, p. 105237, <https://doi.org/10.1016/j.neubiorev.2023.105237>.
- [8] Saeedy, Sina. "A Futures Study of Life Satisfaction among Iranian Students." *Social Science Information*, 2024, <https://doi.org/10.1177/05390184241229029>.
- [9] Kanchapogu, Naga & Mohanty, Sachi. (2024). Enhancing Depression Predictive Models: a Comparative Study of Hybrid AI, Machine Learning and Deep Learning Techniques. 10.21203/rs.3.rs-3992742/v1.
- [10] Prome, Shanjita A., et al. "Deception Detection Using Machine Learning (ML) and Deep Learning (DL) Techniques: A Systematic Review." *Natural Language Processing Journal*, vol. 6, 2024, p. 100057, <https://doi.org/10.1016/j.nlp.2024.100057>.
- [11] Premalatha, J. & Aswin, S. & JaiHari, D. & Subash, K. (2023). Detecting Depression Using Quality-of-Life Attributes with Machine Learning Techniques. 10.1007/978-981-99-3932-9\_4.
- [12] Baddu Naik Bhukya, V. Venkataiah, S. Mani.Kuchibhatla, S. Koteswari, R V S Lakshmi Kumari, and Yallapragada Ravi Raju, "Integrating the Internet of Things to Protect Electric Vehicle Control Systems from Cyber Attacks," *IAENG International Journal of Applied Mathematics*, vol. 54, no. 3, pp433-440, 2024
- [13] Su, Chang, et al. "Deep Learning in Mental Health Outcome Research: A Scoping Review." *Translational Psychiatry*, vol. 10, no. 1, 2020, pp. 1-26, <https://doi.org/10.1038/s41398-020-0780-3>.
- [14] Popoola, Olusogo, et al. "A Critical Literature Review of Security and Privacy in Smart Home Healthcare Schemes Adopting IoT & Blockchain: Problems, Challenges and Solutions." *Blockchain: Research and Applications*, 2023, p. 100178, <https://doi.org/10.1016/j.bcr.2023.100178>.
- [15] Khalid, Nazish, et al. "Privacy-preserving Artificial Intelligence in Healthcare: Techniques and Applications." *Computers in Biology and Medicine*, vol. 158, 2023, p. 106848, <https://doi.org/10.1016/j.combiomed.2023.106848>.
- [16] Božić, Velibor. (2024). Digital Health and Communication -A Revolution with Cautions. 10.13140/Rg.2.2.32230.64321.
- [17] Woodman, R.J., Mangoni, A.A. A comprehensive review of machine learning algorithms and their application in geriatric medicine: present and future. *Aging Clin Exp Res* 35, 2363–2397 (2023). <https://doi.org/10.1007/s40520-023-02552-2>
- [18] Rasheed, Khansa, et al. "Explainable, Trustworthy, and Ethical Machine Learning for Healthcare: A Survey." *Computers in Biology and Medicine*, vol. 149, 2022, p. 106043, <https://doi.org/10.1016/j.combiomed.2022.106043>.
- [19] Ali, Md. M., et al. "Development and Performance Analysis of Machine Learning Methods for Predicting Depression among Menopausal Women." *Healthcare Analytics*, vol. 3, 2023, p. 100202, <https://doi.org/10.1016/j.health.2023.100202>.
- [20] Baddu Naik Bhukya, Padmanabha Raju Chinda, Srinivasa Rao Rayapudi, and Swarupa Rani Bondalapati, "Advanced Control with an Innovative Optimization Algorithm for Congestion Management in Power Transmission Networks," *Engineering Letters*, vol. 31, no.1, pp194-205, 2023
- [21] Swathi Dasi, Swarupa Rani Bondalapati, Mummidi P Subbaraju, Divya Nimma, Pradeep Jangir, R Vijaya Kumar Reddy, and N Zaareena, "IoT-Based Intelligent Energy Management for EV Charging Stations," *IAENG International Journal of Computer Science*, vol. 51, no. 11, pp1853-1861, 2024.
- [22] Wang, W., Ferrari, D., Haddon-Hill, G., Curcin, V. (2023). Electronic Health Records as Source of Research Data. In: Colliot, O. (eds) *Machine Learning for Brain Disorders*. *Neuromethods*, vol 197. Humana, New York, NY. [https://doi.org/10.1007/978-1-0716-3195-9\\_11](https://doi.org/10.1007/978-1-0716-3195-9_11)
- [23] Mumuni, Alhassan, and Fuseini Mumuni. "Automated Data Processing and Feature Engineering for Deep Learning and Big Data Applications: A Survey." *Journal of Information and Intelligence*, 2024, <https://doi.org/10.1016/j.jiixd.2024.01.002>.
- [24] Chang, Victor, et al. "An Assessment of Machine Learning Models and Algorithms for Early Prediction and Diagnosis of Diabetes Using Health Indicators." *Healthcare Analytics*, vol. 2, 2022, p. 100118, <https://doi.org/10.1016/j.health.2022.100118>.
- [25] Wang, Jing, and Filip Biljecki. "Unsupervised Machine Learning in Urban Studies: A Systematic Review of Applications." *Cities*, vol. 129, 2022, p. 103925, <https://doi.org/10.1016/j.cities.2022.103925>.
- [26] Jain, S., Doriya, R. Security framework to healthcare robots for secure sharing of healthcare data from cloud. *Int. j. inf. tecnol.* 14, 2429–2439 (2022). <https://doi.org/10.1007/s41870-022-00997-8>
- [27] Woodman, R.J., Mangoni, A.A. A comprehensive review of machine learning algorithms and their application in geriatric medicine: present and future. *Aging Clin Exp Res* 35, 2363–2397 (2023). <https://doi.org/10.1007/s40520-023-02552-2>.