Establish an Effective Framework for Accurate Prediction of Employee Mental Health for Public Health

Omprakash Dewangan[1], Vasani Vaibhav Prakash[2]
[1]*Assistant Professor, Department of CS & IT, Kalinga University, Raipur, India*
[2]*Research Scholar, Department of CS & IT, Kalinga University, Raipur, India*

KEYWORDS

Mental health, dynamic lion optimized VGG16 (DLO-VGG16), psychological suffering, Sequential Floating Forward Selection (SFFS), work environment, public health.

ABSTRACT

Employee mental health is a primary feature of workplace efficiency and overall public health. With rising mental health problems impaired by variables contains stress, burnout, and isolation, there is an urgent need for effective prediction models that can identify at-risk employees. This study proposes a robust predictive model utilizing dynamic lion optimized VGG16 (DLO-VGG16), for employee mental health prediction represents a significant advancement in public healthcare. The use of these sophisticated prediction models can improve medical results for groups at risk and is essential in the battle against heart disease. Employee's mental health data collected from Kaggle website. The data preprocessing phase employed Z-score normalization to standardize input values, thereby addressing data inconsistencies and errors. The Sequential Forward Floating Search Algorithm (SFFS) was utilized for extracting the employee's mental health. This enhanced heuristic search method begins with an empty feature set, incrementally adding features until the optimal set is identified. The proposed DLO-VGG16 framework to balance accuracy, ROC curve, and f1-score ensuring reliable detection and forecasting of employee mental health issues, thereby contributing to the broader public health landscape.

1. Introduction

A system focused on mental health prediction has emerged as the value of mental well-being, is growing more generally recognized in the rapidly evolving place of employment. The application [1] combines the use of machine learning (ML) abilities and data from a 2014 Tech sector assessment to proactively detect and address mental health concerns. The mental and emotional health of employees is given priority in a supportive work environment [2], which promotes open communication, managerial assistance, and access to mental health treatments for working efficiently. Employee turnover has an enormous effect on financial performance, which affects the industrial sector [3]. Mental health problems are on the rise among working people, throwing workers, organizations, and public health system [5] such as greater disability pensions for mental disorders at risk [4]. These include workplace mental health and compensation authorities. It is positive or negative, with positive attrition causing less productivity and low attrition stagnating ideas [6]. High attrition costs corporations and hinders intellectual growth. Negative attrition results in high-performing employees leaving for better opportunities [8]. In fields including image identification, natural language processing, healthcare, finance, and public health, ML transforms complicated problems and yields insightful information from digital data [9]. It also revolutionizes innovation and automation [10].

2. Literature Review

In [11] found that stress issues are prevalent among IT workers due to changing work cultures and lifestyles. When the boosting models were used on data from the 2017 Open-Source Mental Illness (OSMI) mental health survey, they proved to be the most accurate. Decision Trees (DT) showed that advantages related to occupational health, family history, and gender were significant factors influencing stress [12]. To predict increased anxiety, sadness, post-traumatic stress disorder (PTSD), and suicidal thoughts and actions, an early mental health disorder prediction approach, In [13], utilized objective stressor exposure, subjective self-report, multimodal stimulation, neuro-physiological predictor features, and statistical and ML techniques. In [14] examined how COVID-19 pandemic significantly impacted psychological suffering, with young, female, and non-binary individuals experiencing higher rates of anxiety, sadness, and post-traumatic stress disorder. In [16] utilized ML algorithms to forecast stress, anxiety, and depression in the contemporary environment [7]. Random

Forest (RF) classifier accurately predicted psychological stress among unemployed working individuals, emphasizing the importance of precise prediction methodologies for effective management [15]. In addition to addressing the rising incidence and need for efficient healthcare solutions, [17] examined ML techniques for predicting mental health issues and addressed challenges, constraints, and potential solutions.

3. Methodology

The DLO-VGGG16 approaches recommend generating an accurate prediction of employee mental health in public. The collection of employee mental health data, and preprocessed the gathered data, feature extraction from the preprocessed data, and the final conversion of the data to the suggested structure. Figure 1 shows the structure of proposed DLO-VGGG16 methodology.

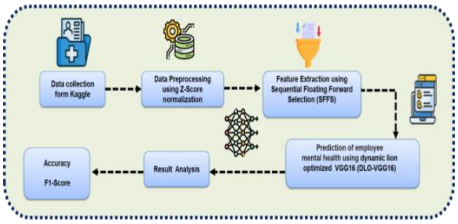

Figure 1: Structure of proposed DLO-VGGG16 methodology

Data collection

To collect the 756 employees mental health public data form Kaggle website [19]. In addition to examining the prevalence of mental health illnesses among IT professionals, the 2017 poll attempts to gauge attitudes about employee mental health for public health. The mental health is seen in the high tech and IT sectors, as well as common various public mental health conditions are there. This information is used by the volunteer team behind open sourcing mental illness to guide our efforts in increasing awareness and enhancing the working environment for people with mental health issues in the IT industry.

Data preprocessing using Z-score normalization

After collecting the dataset preprocessing is normalizing and fixing mistakes constitute certain of ways that preprocessing assists in clearing up the data. This method is the most frequently employed normalization technique that standardizes all input values to a single scale with 0 as average and 1 as standard deviation (SD). SD and mean are computed for every sensor feature. It has been z-score normalized the calculated SD and mean. The air quality transformation is provided as in the Eq (1).

$$X = \frac{(y - mean(Y))}{std(Y)} \quad (1)$$

This method's benefit originates from the reality that it minimizes the effect of outliers on mental health for public data. The preprocessed data is transformed into the feature extraction phase. The Z-score normalization method offers high effectiveness in preprocessing data for enhanced accuracy and reliability in predictions.

Feature Extraction using Sequential Floating Forward Selection (SFFS)

Following the preprocessing, Sequential Forward Floating Search Algorithm (SFFS) is used to extract the features. An enhanced SFFS is a heuristic method of searching. When the interruption requirement is met, a feature is added to the starting feature set, which is initially configured to be zero. A reliable technique that is used in public health to improve the precision of forecasting employee mental health outcomes is SFFS. It optimizes forecasting method by iteratively including and eliminating estimators in a forward selection strategy as it navigates across subsets. By iteratively selecting the most pertinent variables, the model's predictive ability for mental health issues in employers is enhanced. SFFS ensures that public health creativities could be fixated effectively by focusing on crucial indicators, which might lead to better workplace mental health approaches and findings. The feature extracted information is changed into the suggested DLO-VGG16 strategy phase. Greater effectiveness in eliminating patient features enhances the overall performance of the predictive accuracy of employee mental health for public health.

Prediction of employee mental health using dynamic lion optimized VGG16 (DLO-VGG16)

The proposed DLO-VGG16 method predicts employee mental health for public health by leveraging advanced feature extraction and dynamic optimization techniques, enhancing accuracy in identifying well-being indicators to enable timely interventions for a healthier workplace.

Visual Geometry Group 16 (VGG 16)

The VGG16 network uses five convolution layers, a max pooling layer, ReLU activation function, flattened layer, dropout layer, and densely linked output layer for efficient data conversion.

Input layer: This is the first layer in which we load the data into the input layer where image data is read.

Convolution layer: The result is computed by convolution rather than matrix multiplication. This sheet provides function maps. In case, they employ the approximation time convolution technique, use equation (2).

$$t(s) = (w * x)(s) = \sum_{\alpha=-\infty}^{\infty} w(b)x(s-b) \quad (2)$$

$X: kernel, w: input, s: times, t: result$ Equation (3) is used for two-dimensional data inputs like images, referencing the new matrix's position after convolution, typically positioned with the source at the filter core.

Activation function: Deep learning (DL) values are translated into non-linear terms using activation functions. There are numerous activation features accessible. $Tanh$, Sigmoid, and $Relu$ are the most regularly utilized. $Relu$ was used in the procedure that was developed and is shown in equation (3).

$$Relu: e(w) = \begin{matrix} 0, \mathrm{w} < 0 \\ 1, \mathrm{w} \geq 0 \end{matrix} \quad (3)$$

Max polling layer: The bundling layer in CNN reduces function, while pooling methods share output neurons. The max-pooling layer's performance is determined using equation (4).

$$\left[\frac{\mathrm{P} + 2\mathrm{ob} - 2}{\mathrm{ts}}\right] + 1 \quad (4)$$

Dropout layer: A model stops being able to generalize once it has learned a particular set of results. To prevent overtraining, the dropout layer would remove nodes and links that were over learned during training. This approach regulates the data's weighting. Dropout should only be utilized in class to prevent overfitting.

Fully connected layer: More epochs improve classification accuracy in deep learning (DL). The activation method distinguishes two-layer node performance levels, with complete correlations between subsequent layers are depicted in equation (5).

Softmax: Softmax must be provided at the classification layer. It computes a probabilistic sequence to generate values for every class. Utilize this layer to determine the percentages for every class.

$$O(z = i \backslash w; X, c) = \frac{exp^{w^{S}w_i}}{\sum_{i=1}^{m} exp^{w^{S}w_i}} \quad (5)$$

Classification layer: This convolutional layer is the final one in the final location. This layer is given the same number of credits as the entire production quantity. For those of who are aware, this stands for both non-tumorous and healthy tissue. Figure 2 depicts the architecture of the VGG16 network [18].

Figure 2: Architecture of the VGG16 network

Dynamic lion optimization(DLO)

This DLO method is advanced hyperparameter tuning technique that enhances model performance, inspired by natural principles like dynamic lion behavior. It involves dynamic adjustments, evolutionary-inspired algorithms, or strategies mimicking adaptive behaviors in biological systems.

Let them currently investigate the goal function in equation (6).

$$B^{optimal} = arg \min_{b_j \in (b_j^{min}, b_j^{max})} e(b_1, b_2, \dots, b_m); m \geq 1 \qquad (6)$$

In equation (7), $e\,(.)$ represents a continual uni-modal or multi-modal functioning, with the space of a single solution being the same size as another solution's$\mathfrak{R}^m$. Here, $\mathfrak{R}$ symbolizes the real integers, and $bj : j = 1, 2, \dots, n$ is the j^{th} vectors size of the 9th variable solution, correspondingly. The j^{th} variable's minimal and maximal bounds are represented by b_j^{min} and b_j^{max} respectively. Equation (8) represents the optimization algorithm's target solution, denoted as$B^{optimal}$. The dimension of the solution area $e(.)$ can be calculated in the following manner:

$$\mathfrak{R}^m = \prod_{j=1}^{m} (b_j^{max} - b_j^{min}) \qquad (7)$$

$$w^{Optional} = w : e(w) < e\left(b' \middle| b' \neq b; b_j' \in \left(b_j^{max} - b_j^{min}\right)\right) \qquad (8)$$

Here B represents the vector of the depicted arrangement $B = [b_1, b_2, \dots . b_m]$. Equation (9) depicts the goal function that could be obtained by a minimization method. This can occasionally be a retraining process. Thus, a rational computation necessitates the application of an adequate determination procedure.

Pride generation: Pride is demonstrated with the localized lion B^{male} and its lioness B^{female}, while B^{normal}provides the wrong example of a lion, based on the feeling of pride as well as condition. This is stated that whenever pride is developed, a migratory lion isn't a person. The arrangement's vector illustration and the lion's depiction are identical. Whether the vector components of B^{male}, B^{female}, and B^{normal}, that is, $b_\int^{male}$, $b_\int^{female}$, and $b_\int^{normale}$ have been self-confirmed numbers at their base, the greatest limitation points were $m > 1$ (such as searching using the original coding). The lion's length, denoted by $k = 1, 2, \dots, K$, may be determined in the following manner:

$$K = r\begin{cases} o; & m > 1 (genralcase) \\ n; & otherwise\ (SpecialCase) \end{cases} \qquad (9)$$

In this instance $o\ and\ r$ are numbers used to calculate the lion's lengths. Within the case of$m = 1$, the computation needs to be carried out in parallel. In this approach, the vector elements are created as either $1\ or\ 0$, depending on the limitations specified in equations (10) and (11).

$$g(b_\int) \in b_\int^{min}, b_\int^{max} \qquad (10)$$

Where $g(b_{\int}) = \sum_{\int}^{k} b_{\int} \, 2^{\left(\frac{K}{2}-1\right)}$ (11)

A binary lion, formed by terms (10) exists within the control region and equation (11). Complexity of double coded lion is not addressed due to various attempts. The exorcism generated B^{normal}, which impacts one of the migratory lions' locations. Two traveling lions are scheduled to approach the region, introducing additional migrating lions during localized conservation efforts. As a result, the second sediments B^{normal}will have B^{normal}1 with a zero location. The DLO-VGG16 effectively improved the prediction of employee mental health by leveraging advanced feature extraction and hyperparameter tuning, resulting in improved accuracy and robustness in identifying mental health for public health.

4. Results and discussion

In this investigation, we evaluate the DLO-VGG16 approaches under investigation's ability to forecast the employee mental health, using several existing methods including Decision Tree (DT) [20], Navies Bayes (NB) [20], and Deep Neural Network (DNN) [20] in terms of accuracy and f1 score. Table 1 shows the numerical outcomes of the following accuracy and F1-Score.

Table 1: Numerical outcome of Accuracy and F1-Score

Methods	**Values**	
	Accuracy (%)	**F1-Score (%)**
DT [20]	54.2	50.06
NB[20]	68.44	63.7
DNN [20]	88.4	87.28
DLO-VGG16 [Proposed]	90.86	89.55

Accuracy

One of the crucial factors in calculating classification models is accuracy. In general, the proportion of forecasts is used to assess accuracy. Equation (12) defines accuracy and comparison and outcomes of accuracy, as shown in Figure 3. The proposed DLO-VGG16 achieved 90.86% accuracy when compared to other existing models, DT (54.2%), NB (68.44%), and DNN (88.4%).

$$Accuracy(\%) = \frac{No.of.Correct\ Prediction}{Total.no.of.Prediction} \times 100 (12)$$

F1-Score

Given the uneven distribution of classes, the F1 Score becomes crucial for evaluating the quality of predictions. This metric is essential to balance false positives and false negatives, which is determined in Equation (13), ensuring reliable and meaningful forecasting in the context of public health. The comparison of the f1-score and results are displayed in Figure 3, the proposed DLO-VGG16 achieved 89.55% of F1-score when compared to other traditional techniques, DT (50.06%), NB (63.7%), and DNN (87.28%).

$$F1 - score = \frac{2 \times recall \times Precision}{recall + Precision} \quad (13)$$

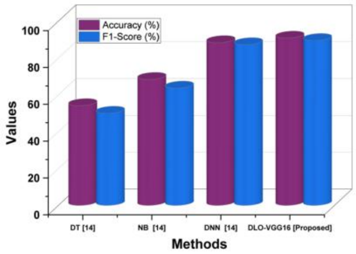

Figure 3: Comparison of Accuracy and F1-Score

ROC Curve

These DLO-CNN methods are taken into special consideration for this comparison since they were suggested in other methods that were taken into consideration during study. The specificity assessment for each technique in Figure 4 displayed the ROC curve.

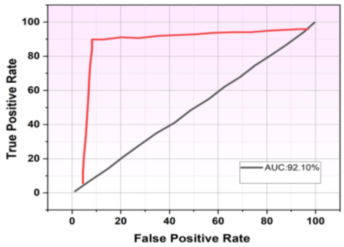

Figure 4: ROC Curve for proposed DLO-CNN method

5. Conclusion and future scope

This study developed a robust predictive model for employee mental health using data from various sectors. The DLO-VGG16 framework, adapted from the VGG16 with DL optimization architecture, ensures accurate detection and forecasting of mental health issues. The DLO-VGG16 framework, derived from a Kaggle dataset, was meticulously preprocessed using Z-score normalization and the SFFS to address inconsistencies and errors. This robust framework ensures reliable detection and forecasting of employee mental health issues, contributing to the broader public health landscape and providing early intervention and support for employee mental well-being. The proposed model (DLO-VGG16) achieved accuracy (90.86%) and F1-score (89.55%), making it a valuable tool for early intervention and support for employee mental well-being, contributing to public health.

Reference

[1] N.T.Renukadevi, R.Sridhar, S.A. Kumar, and S.Vignesh, "MENTAL HEALTH PREDICTION FOR EMPLOYEES USING MACHINE LEARNING," MENTAL HEALTH, 19(1), 2024.

[2] A.R.Skelton, D. Nattress, and R.J.Dwyer, "Predicting manufacturing employee turnover intentions," Journal of Economics, Finance and Administrative Science, 25(49), pp.101-117, 2020. https://doi.org/10.1108/JEFAS-07-2018-0069

[3] B.H.Sujal, K. Neelima, C. Deepanjali, P.Bhuvanashree, K.Duraipandian, S. Rajan, and M.Sathiyanarayanan, "Mental health analysis of employees using machine learning techniques," In 2022 14th International Conference on Communications Systems &NETworkS (COMSNETS) (pp. 1-6). IEEE, January2022,https://doi.org/10.1109/ICICCS51141. 2021.9432259

[4] S. Neelima, Manoj Govindaraj, Dr.K. Subramani, Ahmed ALkhayyat, & Dr. Chippy Mohan. (2024). Factors Influencing Data Utilization and Performance of Health Management Information Systems: A Case Study. Indian Journal of Information Sources and Services, 14(2), 146–152. https://doi.org/10.51983/ijiss-2024.14.2.21

[5] K.Pugazharasi, P. Kalaivani, J. Jayapriya, S. Kadhambari, and S.Mailvizhi, "Machine learning algorithms for predicting depression, anxiety and stress in modern life," In AIP Conference Proceedings, 2587(1),November2023, AIP Publishing. https://doi.org/10.1063/5.0150604

[6] T. E. Ramya, "An Effective Approach for Mental Health Prediction Using Machine Learning algorithm". 10(13), 81–84,

2022. https://doi.org/10.17577/IJERTCONV10IS13016

[7] Sotnikova, O., Zhidko, E., Prokshits, E., & Zolotukhina, I. (2022). Administration of Sustainable Development of Territories as One of the Approaches for Creating A Biosphere-Compatible and Comfortable Urban Environment. Archives for Technical Sciences, 1(26), 79–90.

[8] G.Mazzetti, E. Robledo, M. Vignoli, G. Topa, D. Guglielmi, and W.B.Schaufeli, "Work engagement: A meta-analysis using the job demands-resources model," Psychological reports, 126(3), pp.1069-1107, 2023.

[9] Sonya, A., & Kavitha, G. (2022). A Data Integrity and Security Approach for Health Care Data in Cloud Environment. Journal of Internet Services and Information Security, 12(4), 246-256

[10] S.Timsina, "Employee Turnover and Engagement Programs for Retention,"2024.

[11] G.K. Awal, and K.Rao, "Can Machine Learning Predict an Employee's Mental Health?, "In Information, Communication and Computing Technology: 6th International Conference, ICICCT 2021, New Delhi, India, May 8, 2021, Revised Selected Papers 6 (pp. 235-247), 2021. Springer International Publishing.

[12] Bobir, A.O., Askariy, M., Otabek, Y.Y., Nodir, R.K., Rakhima, A., Zukhra, Z.Y., Sherzod, A.A. (2024). Utilizing Deep Learning and the Internet of Things to Monitor the Health of Aquatic Ecosystems to Conserve Biodiversity. Natural and Engineering Sciences, 9(1), 72-83.

[13] K. Ćosić, S. Popović, M. Šarlija, I. Kesedžić, and T. Jovanovic, "Artificial intelligence in prediction of mental health disorders induced by the COVID-19 pandemic among health care workers," Croatian medical journal, 61(3), p.279, 2020.

[14] T.A. Prout, S. Zilcha-Mano, K.Aafjes-van Doorn, V. Békés, I. Christman-Cohen, K. Whistler, T. Kui, and M.Di Giuseppe, "Identifying predictors of psychological distress during COVID-19: a machine learning approach, "Frontiers in psychology, 11, p.586202, 2020.

[15] Lomotey, R.K., & Deters, R. (2013). Facilitating Multi-Device Usage in mHealth. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 4(2), 77-96

[16] A.Priya, S. Garg, and N.P. Tigga, "Predicting anxiety, depression and stress in modern life using machine learning algorithms," Procedia Computer Science, 167, pp.1258-1267, 2020.

[17] J.Chung, and J.Teo, "Mental health prediction using machine learning: taxonomy, applications, and challenges," Applied Computational Intelligence and Soft Computing, 2022(1), p.9970363, 2022.

[18] Pittala, C.S., Sravana, J., Ajitha, G., Lakshamanachari, S., Vijay, V., & Venkateswarlu, S.C. (2021). Novel architecture for logic test using single cycle access structure. Journal of VLSI Circuits and Systems, 3(1), 1-6.

[19] Data collected from kaggle website (https://www.kaggle.com/ datasets/ osmihelp/osmi-mental-hea lth-in-tech-survey-2017)

[20] N. Patel, S. Trivedi, andN. Faruqui, "An innovative deep neural network for stress classification in workplace," In 2023 International Conference on Smart Computing and Application (ICSCA) (pp. 1-5), February2023, IEEE.