Survey on Machine Learning Approach to Predict The Depression Level

Gopinath C B
Research Scholar
Department of Information Science and Engineering
Adichunchanagiri Institute of Technology, Chikkamagaluru
Visvesvaraya Technological University, Belagavi

Dr. Sunitha M R
Professor & Head
Department of Artificial Intelligence and Machine Learning
Adichunchanagiri Institute of Technology, Chikkamagaluru
Visvesvaraya Technological University, Belagavi

Abstract - **Collaborative Machine Learning (CML) stands as a novel paradigm that revolutionizes traditional machine learning methodologies by amalgamating the collective intelligence of multiple entities. This paper investigates the increasing landscape of Collaborative Machine Learning, exploring its underlying principles, methodologies, and applications across various domains. It delves into the multifaceted nature of CML, highlighting its significance in enhancing model performance, scalability, and robustness by harnessing the collective knowledge embedded in distributed datasets. The abstract concept of collaboration within machine learning extends beyond mere aggregation of data; it encompasses federated learning, ensemble methods, transfer learning, and multi-party computation. These techniques enable disparate entities to collaborate without compromising data privacy, fostering advancements in domains like healthcare, finance, and IoT. Collaborative Machine Learning facilitates the development of personalized mental health interventions. By leveraging the collective insights gleaned from diverse sources, these models can identify subtle patterns and markers indicative of mental health conditions, enabling early detection and intervention. For instance, predictive models can identify high-risk individuals or foresee potential crises, allowing for timely interventions or proactive support systems**

Keywords - **Machine Learning, Prediction, collaborative approach.**

I. INTRODUCTION

Collaborative Machine Learning (CML) has emerged as a promising avenue in the domain of mental health prediction, revolutionizing the way we approach the diagnosis, understanding, and treatment of various mental health conditions. This innovative approach transcends the constraints of individual datasets and models, harnessing the collective knowledge embedded in distributed sources to enhance predictive accuracy, personalized treatment, and early intervention strategies [1].

Mental health prediction through Collaborative Machine Learning involves amalgamating insights from diverse data sources, including electronic health records (EHRs), wearable devices, social media activity, and more. By leveraging federated learning techniques, where models are trained locally on individual data sources and then collaboratively aggregated, CML ensures privacy preservation while deriving collective intelligence.
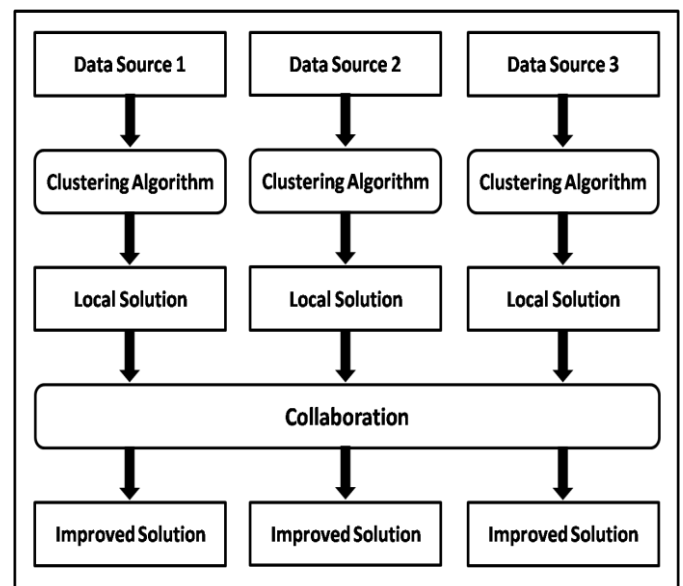

Figure 1: Collaborative machine learning approach for classification

One of the significant advantages of CML in mental health prediction lies in its ability to incorporate multifaceted data types. For instance, wearable sensors can capture physiological signals, while EHRs contain clinical notes and diagnostic information. Integrating these disparate data modalities enables a holistic understanding of an individual's mental health state, allowing for more accurate predictions and tailored interventions [2, 3]. The collaborative approach fosters robust predictive models by mitigating issues related to data scarcity and variability often encountered in mental health datasets. Federated learning techniques enable model training across multiple institutions or individuals without centralizing sensitive data, thereby addressing concerns of privacy and data ownership. This collaborative framework also allows for continual model refinement as more data becomes available, ensuring adaptive and up-to-date predictive capabilities [4, 5].

Moreover, Collaborative Machine Learning facilitates the development of personalized mental health interventions. By leveraging the collective insights gleaned from diverse sources, these models can identify subtle patterns and markers indicative of mental health conditions, enabling early detection and intervention. For instance, predictive models can identify

high-risk individuals or foresee potential crises, allowing for timely interventions or proactive support systems.

However, the application of Collaborative Machine Learning in mental health prediction poses several challenges. Ensuring data security and privacy while aggregating information from disparate sources remains a paramount concern. Techniques like federated learning, secure multi-party computation, and differential privacy play a pivotal role in addressing these challenges by allowing collaborative model training while preserving data confidentiality.

Additionally, ethical considerations surrounding the use of predictive models in mental health care demand careful attention. Issues related to model transparency, fairness, and potential biases need to be meticulously addressed to ensure equitable and just outcomes for all individuals.

II. BRIEF REVIEW OF MACHINE LEARNING TECHNIQUES

Machine learning techniques encompass a diverse set of algorithms and methodologies that enable computers to learn from data and make predictions or decisions without explicit programming [6]. These techniques can be broadly categorized into supervised learning, unsupervised learning, reinforcement learning, ensemble learning, deep learning, semi-supervised learning, and transfer learning, each with its unique applications and principles.

Supervised learning involves training a model on labeled data, where the algorithm learns the mapping between input features and corresponding output labels. Regression techniques predict continuous outcomes, such as linear or polynomial regression, while classification algorithms like logistic regression or decision trees predict discrete class labels. This approach is widely used in various fields, including finance, healthcare, and natural language processing.

In contrast, unsupervised learning works with unlabeled data to discover hidden patterns or structures within the dataset. Clustering algorithms, such as k-means or hierarchical clustering, group similar data points together, while dimensionality reduction techniques like PCA or t-SNE help in reducing the number of features while retaining essential information. Reinforcement learning operates on the concept of an agent interacting with an environment, learning through trial and error to maximize cumulative rewards. Algorithms like Q-learning or Deep Q Networks (DQN) are used in various applications, including robotics, game playing, and autonomous systems [7, 8, 9].

Ensemble learning combines multiple models to improve predictive performance. Bagging methods train multiple models independently and aggregate their predictions, while boosting techniques sequentially train models, giving more weight to misclassified instances. These approaches, seen in Random Forest or AdaBoost, often enhance accuracy and robustness. Deep learning, a subset of machine learning, utilizes neural networks with multiple layers to learn complex patterns directly from raw data. Convolutional Neural Networks (CNNs) excel in image-related tasks, Recurrent Neural Networks (RNNs) are effective in sequential data

analysis, and Transformer models dominate in natural language processing [10].

Semi-supervised learning techniques combine elements of both supervised and unsupervised learning by utilizing a small amount of labeled data along with a more extensive set of unlabeled data. Transfer learning leverages knowledge from pre-trained models to efficiently solve related tasks by fine-tuning existing models on specific datasets or tasks. Each machine learning technique possesses distinct strengths, limitations, and ideal use cases, dependent on data characteristics, problem complexity, and desired outcomes. The selection of a technique often involves understanding the nature of the data and the problem domain, with ongoing advancements continually expanding the applications of these methodologies across diverse fields and industries.

III. FUNDAMENTAL BACKGROUND OF DIFFERENT SIGNIFICANT DATA ATTRIBUTES

Predicting depression using machine learning methods relies on the selection of important attributes or features that are indicative of an individual's risk or presence of depression [11]. These attributes can be derived from various data sources, including self-reported surveys, clinical assessments, behavioral data, and more. Here are some important attributes commonly used in machine learning-based depression prediction:

Demographic Information:
Age: Age can be a significant factor, as the risk of depression may vary across different age groups.
Gender: Some studies have shown gender-related differences in depression prevalence.
Socioeconomic Status: Factors like income, education, and employment status can influence depression risk.

Personal History:
Previous Episodes of Depression: A history of depression is a strong predictor of future episodes.
Family History: A family history of depression may indicate a genetic predisposition.
Medical History: Chronic illnesses and comorbidities can contribute to depression risk.

Psychosocial Factors:
Stress Levels: Chronic stressors or recent life events can increase depression risk.
Social Support: The presence or absence of a strong support network can impact mental health.
Trauma and Adverse Childhood Experiences (ACEs): Experiencing childhood trauma or ACEs can increase vulnerability to depression.

Behavioral Patterns:
Sleep Patterns: Irregular sleep patterns, insomnia, or excessive sleep can be indicative of depression.
Physical Activity: Low physical activity levels or sedentary behavior can be associated with depression.
Substance Use: The use of alcohol, drugs, or tobacco can contribute to depression risk.

Cognitive and Emotional Factors:
Cognitive Distortions: Negative thought patterns and cognitive distortions can be signs of depression.
Emotional Regulation: Difficulty in regulating emotions and experiencing prolonged sadness or hopelessness.
Rumination: Frequent rumination on negative thoughts or events may be a predictive factor.

Questionnaire and Survey Responses:
PHQ-9 and Other Depression Screening Tools: Scores on validated depression screening questionnaires like the Patient Health Questionnaire-9 (PHQ-9) can directly indicate depressive symptoms.
Anxiety Levels: Co-occurring anxiety symptoms may provide additional predictive information.
Quality of Life: Self-reported quality of life assessments can reflect depressive symptoms and overall mental well-being.

Biological Markers:
Genetic Markers: Specific genetic variations associated with depression susceptibility.
Hormonal Levels: Elevated cortisol levels, a stress hormone, can be indicative of increased depression risk.
Neuroimaging Data: Brain imaging can reveal structural and functional differences associated with depression.

Social Media and Digital Footprints:
Sentiment Analysis: Analyzing social media posts and online activity for emotional distress or changes in sentiment.
Linguistic Cues: Identifying linguistic patterns in digital communication that may signal depressive symptoms.

Wearable Device Data:
Heart Rate Variability: Irregular heart rate patterns may indicate stress and potential depressive symptoms.
Activity Levels: Decreased physical activity and mobility can be associated with depression.

Environmental Factors:
Geographic Location: Regional factors, such as climate and access to healthcare, can influence depression risk.
Seasonal Variation: Seasonal affective disorder (SAD) is a subtype of depression related to seasonal changes.

IV. METHODOLOGY

The methodology for mental health prediction using collaborative machine learning approach is as represented in figure 2.
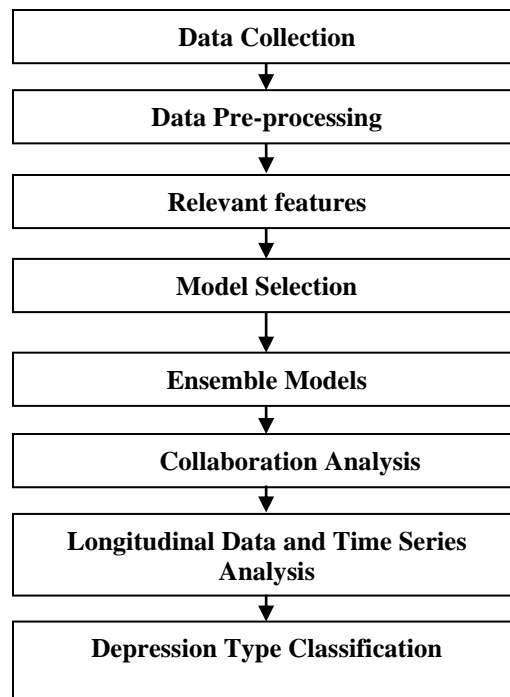

Figure 2: Proposed prediction methodology

The selection of attributes depends on the available data and the specific machine learning algorithms being used. Feature selection techniques and domain expertise are often employed to identify the most relevant attributes for accurate depression prediction. Combining multiple types of attributes and data sources can enhance the predictive power of machine learning models for depression. Depression prediction is a complex task that can benefit from a multifaceted approach that considers various data sources and techniques.

Data Collection:
Gather a diverse range of data sources, including demographic information, medical history, psychosocial factors, behavioural patterns, and biomarkers.
Collect self-reported data through validated depression screening questionnaires and surveys.
Consider incorporating data from wearable devices, social media activity, and electronic health records, if available and ethically appropriate.

Data Pre-processing:
Clean and pre-process the data, handling missing values and outliers appropriately.
Normalize or standardize numerical features to ensure they have a consistent scale.
Encode categorical variables into numerical format using techniques like one-hot encoding.

Feature Engineering:
Identify and create relevant features that may be indicative of depression risk.
Consider domain knowledge and expert input to select meaningful features.
Feature selection techniques like recursive feature elimination or feature importance from tree-based models can help identify the most relevant features.

Model Selection:
Choose appropriate machine learning algorithms based on the nature of the problem (classification or regression) and the data characteristics.
Consider trying multiple algorithms to compare their performance, such as logistic regression, decision trees, random forests, support vector machines, and gradient boosting algorithms.

Ensemble Models:
Experiment with ensemble methods, such as random forests or gradient boosting, which can often enhance prediction accuracy and robustness.

Collaboration Analysis:
Collaborative Machine Learning presents a transformative approach to mental health prediction, offering a pathway towards more accurate, personalized, and ethical interventions. By synergistically leveraging distributed data sources while preserving privacy, CML holds the potential to revolutionize mental health care, fostering early intervention strategies and tailored treatments for individuals worldwide
Consider the class imbalance issue in depression prediction and use metrics like area under the precision-recall curve (AUC-PR) in addition to ROC AUC.

Longitudinal Data and Time Series Analysis:
Consider the longitudinal aspect of depression by analysing data over time to detect trends or patterns that may indicate risk or recovery.
Depression prediction is a challenging task that requires a multidisciplinary approach, ethical considerations, and ongoing validation to create a reliable and effective model for identifying individuals at risk of depression.

V. CONCLUSION

In conclusion, Collaborative Machine Learning presents a transformative approach to mental health prediction, offering a pathway towards more accurate, personalized, and ethical

interventions. By synergistically leveraging distributed data sources while preserving privacy, CML holds the potential to revolutionize mental health care, fostering early intervention strategies and tailored treatments for individuals worldwide. However, continued research, ethical guidelines, and technological advancements are imperative to harness the full potential of CML in transforming mental health prediction and care. Collaborative machine learning paves the way for personalized interventions, early detection, and tailored solutions. By amalgamating insights from diverse sources, it enables the identification of subtle patterns and markers, facilitating proactive measures and targeted support systems, especially in domains like mental health prediction or disease diagnosis.

REFERENCES

[1] J. Forough and S. Momtazi, "Ensemble of deep sequential models for credit card fraud detection", Appl. Soft Comput., vol. 99, Feb. 2021.

[2] V. Arora, R. S. Leekha, K. Lee and A. Kataria, "Facilitating user authorization from imbalanced data logs of credit cards using artificial intelligence", Mobile Inf. Syst., vol. 2020, pp. 1-13, Oct. 2020.

[3] O. Balogun, S. Basri, S. J. Abdulkadir and A. S. Hashim, "Performance analysis of feature selection methods in software defect prediction: A search method approach", Appl. Sci., vol. 9, no. 13, pp. 2764, Jul. 2019.

[4] Madhu Belur Gopala Gowda1, Naveen Kumar Boraiah, Varun Eshappa, GopalaKrishna Chandrashekara, "Classification of Epileptic EEG Signals Using Improved Atomic Search Optimization Algorithm", International Journal of Intelligent Engineering and Systems, Vol.16, No.6, 2023, DOI: 10.22266/ijies2023.1231.12.

[5] Y. Fang, Y. Zhang and C. Huang, "Credit card fraud detection based on machine learning", Comput. Mater. Continua, vol. 61, no. 1, pp. 185-195, 2019.

[6] J. Kim, H.-J. Kim and H. Kim, "Fraud detection for job placement using hierarchical clusters-based deep neural networks", Int. J. Speech Technol., vol. 49, no. 8, pp. 2842-2861, Aug. 2019.

[7] M. Günay and T. Ensari, "EEG signal analysis of patients with epilepsy disorder using machine learning techniques", Proc. IEEE Comput. Sci. Biomed. Engineerings' Meeting Electric Electron., pp. 1-4, 2018.

[8] M.-P. Hosseini, D. Pompili, K. Elisevich and H. Soltanian-Zadeh, "Random ensemble learning for EEG classification", Artif. Intell. Medicine, vol. 84, pp. 146-158, 2018.

[9] J. Hu and J. Min, "Automated detection of driver fatigue based on EEG signals using gradient boosting decision tree model", Cogn. Neurodyn., vol. 12, no. 4, pp. 431-440, Aug. 2018.

[10] E. Hramov et al., "Classifying the perceptual interpretations of a bistable image using EEG and artificial neural networks", Frontiers Neurosci., vol. 11, pp. 674, 2017.

[11] Varun E and Dr. PushpaRavikumar, "Community Mining In Multi-Relational and Heterogeneous Telecom Network", IEEE 6th International Conference on Advanced Computing (IACC-2016), DOI:10.1109/IACC.2016.15.

[12] W.-Y. Hsu, "Assembling a multi-feature EEG classifier for left–right motor imagery data using wavelet-based fuzzy approximate entropy for improved accuracy", Int. J. Neural Syst., vol. 25, no. 08, 2015.