Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

An Improved Eyegaze Tracking Application for Early Prediction of Mental Health Using Convolution Neural Networks with Hybrid Optimization Approach

S. Mary Rexcy Asha[1], Dr. G. Victo Sudha George[2], Dr. P. S. Rajakumar[2]

[1]Research scholar, Department of Computer Science and Engineering, Dr. M. G. R Educational and Research Institute, Chennai, India, rexcyasha@gmail.com
[2]Professor, Department of Computer Science and Engineering, Dr. M. G. R Educational and Research Institute, Chennai, India

This paper presents a new method that uses eyegaze measurements and convolutional neural networks (CNNs) for predicting mental health. Meanwhile, the CNN model using a combined optimization technique with stochastic gradient descent (SGD) and genetic algorithms (GA) displayed high predictive performance in distinguishing subtle mental health states. The study collected and pre-processes 3200 facial images taken from internet repositories involving several people. It is an effective hyrid method of hardware optimization using the CNN model, outperforming traditional machine learning methods such as support vector machines (SVM) and K-nearest neighbors (KNN) as well as the widely used VGG-16 architecture. Criterion measures include accuracy, precision, recall, and F1-score, so comparing performance based on different aspects can be done effectively. Furthermore, confusion matrixes can show classification results in detail, giving an in-depth look at predictive capabilities and potential errors. These findings emphasize the fact that the hybrid optimization approach, through the fine-tuning of population distribution model, can significantly increase predictive utility. Even though many limitations have been encountered, including data variability and model interpretability challenges, the research lays the groundwork for future developments in digital mental health. Getting more data resources from other fields, combining data from various sources, and collaborating with professionals in mental health will increase the prediction accuracy of the proposed eyegaze tracking application.

Keywords: Eyegaze tracking, convolutional neural networks, mental health prediction, hybrid optimization, machine learning.

1. Introduction

The mental health research sector has made significant progress in recent years, with an increased emphasis on early warning and prevention. Among the emerging technologies used in this field, the use of eyegaze tracking technology is promising: minor changes in the eyes that may indicate mental health issues[1]–[3]. In this study, we present an improved eyegaze tracking method for predicting mental health early. This combines convolutional neural networks (CNNs) and a mixed optimisation system.

Eye-tracking technology has proven useful in the fields of psychology, neuroscience, and human-computer interaction. Tracking their eyes allows researchers to learn a lot about cognitive processes, visual attention mechanisms, and emotional reactions. However, eye movements and gaze patterns can reveal sounds that one does not want to hear, such as depression, anxiety, and attention deficit hyperactivity disorder (ADHD). However, existing eyegaze tracking applications for mental health prediction frequently use ineffective or time-consuming methods to recognise objects and scenes, implying a lack of accuracy or computational efficiency[4]–[6].

To overcome these limitations, machine learning (ML) models use image-based analysis, which includes eyegaze-tracking data. Convolutional neural networks (CNNs), which automatically learn and extract features from visual data, are effective tools for computer vision tasks. Image classification, object detection, and facial recognition have all been CNN-intensive tasks. CNNs were used to study gaze patterns when eyegaze tracking was implemented to address mental health issues[7]–[9]. CNNs can learn hierarchical features from raw gaze data, capturing fine details of the data and their relationships to one another. As a result, CNNs predict mental health with significantly higher accuracy than traditional ML models[10], [11].

Many other machine learning models have been tested to predict mental health using images. For example, the k-nearest neighbours (KNN) algorithm can be applied to classify gaze patterns based on their proximity to exemplars. Support vector machines (SVMs) were used to define decision boundaries between different types of mental health classifications. Recurrent neural networks (RNN) were used to examine sequential gaze data for temporal dependencies and trends. Furthermore, pre-trained models like VGG-16, which have deep architectures and comprehensive learned representation styles, have been applied to mental health prediction tasks[12], [13].

The proposed eye-tracking application aims to provide an operational assessment of its potential for the early detection of mental illness, if nothing else. The effectiveness of the model depends on its accuracy, sensitivity, specificity and other performance metrics. In addition, research on CNNs has also discovered that they can serve various other mechanically driven tasks, especially in electric systems. These capabilities are important to understand if we are to develop hybrid technologies quickly. The hybrid optimisation approach seeks to speed up convergence rate as well as ensure higher precision. Advanced optimisation techniques such as Adam, RMSprop, and AdaGrad have been incorporated into the hybrid optimisation framework. We will use a comparison with traditional optimisation methods like stochastic gradient descent (SGD) to show the advantages of this new approach[14], [15].

The preliminary results are promising, but it is important not to overlook potential limitations or difficulties. Problems of this kind could range from worries about data quality to being uncertain whether the model will cover a wide variety of people. Removing these encumbrances can give a comprehensive comparison of the recommended application, along with ideas for enhancing the system. Also, it is important to investigate further materials and such potential application of the proposed eye-tracking application. That means to get an insight into the integration of multimodal data by combinng eye gaze tracking with physiological or text data to increase the model's accuracy. Eliminating these future research issues will advance eye-tracking technology and its application as a valuable tool for mental health prediction, prevention and intervention[16]–[18].

1.1Stochastic Gradient Descent (SGD)

Stochastic Gradient Descent (SGD) is an optimization algorithm used to minimize an objective function, commonly applied in machine learning and deep learning models. It is particularly well-suited for large-scale and online learning tasks.

1. Gradient Descent:

○ Gradient Descent is an optimization technique that aims to find the minimum of a function by iteratively moving towards the steepest descent (negative gradient) direction.

$$\theta_{new} = \theta_{old} - \eta \cdot \nabla J(\theta)$$

where θ represents the parameters, η is the learning rate, and ∇J(θ) is the gradient of the objective function J(θ).

2. Stochastic Gradient Descent:

○ Unlike batch gradient descent, which calculates the gradient using the entire dataset, SGD updates the parameters using only a single or a small batch of training examples.

○ The update rule is:

$$\theta_{new} = \theta_{old} - \eta \cdot \nabla J(\theta; x_i, y_i)$$

where (xi,yi) represents a single training example.

Advantages

- Efficiency: SGD can handle very large datasets as it processes one example at a time.
- Convergence: It often converges faster than batch gradient descent for large datasets.
- Online Learning: Suitable for online learning scenarios where the model is continuously updated with new data.

Disadvantages

- Noisy Updates: Each update has high variance, which can cause the parameter updates to fluctuate heavily.

- Learning Rate: Choosing the right learning rate can be tricky. Too high can cause divergence, too low can slow convergence.
- Convergence to Local Minima: May converge to a local rather than global minimum due to the noisy nature of updates.

Variants

- Mini-batch Gradient Descent: Uses a small batch of examples instead of a single one, balancing between SGD and batch gradient descent.
- Momentum: Incorporates a momentum term to smooth the updates and accelerate convergence.
- Adam (Adaptive Moment Estimation): Combines ideas from momentum and RMSProp (adaptive learning rates) to adjust learning rates dynamically.

1.2 Genetic Algorithms (GA)

Genetic Algorithms (GA) are a class of optimization algorithms inspired by the process of natural selection. They belong to the family of evolutionary algorithms and are used to find approximate solutions to optimization and search problems.

Key steps

1. Population:

o A set of candidate solutions, called individuals or chromosomes, representing potential solutions to the problem.

2. Fitness Function:

o Evaluates how good each candidate solution is. The fitness function guides the evolution process by assigning a fitness score to each individual.

3. Selection:

o The process of choosing individuals from the current population to create offspring for the next generation. Common methods include roulette wheel selection, tournament selection, and rank-based selection.

4. Crossover (Recombination):

o Combines two parent individuals to produce one or more offspring. This process mimics biological recombination and aims to produce new individuals with desirable traits from both parents.

5. Mutation:

o Introduces random changes to individual genes in an offspring. This helps maintain genetic diversity and allows the algorithm to explore a wider search space.

6. Generation:

o The population evolves over successive iterations, called generations. In each generation, individuals are selected, crossed over, and mutated to form a new population.

Advantages

- Global Search: GAs are good at exploring a large search space and can escape local minima.
- Flexibility: Can be applied to a wide range of optimization problems, including those with complex, nonlinear, and multimodal functions.
- Parallelism: Natural parallelism allows easy implementation on parallel processing systems.

Process

1. Initialization: Generate an initial population of individuals randomly.
2. Evaluation: Compute the fitness of each individual in the population.
3. Selection: Select individuals based on their fitness to create a mating pool.
4. Crossover: Perform crossover to generate new offspring.
5. Mutation: Apply mutation to introduce variability.
6. Replacement: Replace the old population with the new one.
7. Termination: Repeat the process until a stopping criterion is met (e.g., a solution of acceptable quality is found or a maximum number of generations is reached).

This article describes an improved eyegaze tracking application for early prediction of mental health using CNNs and a hybrid optimisation approach. In addition to employing CNNs and advanced optimisation methods, this application is intended to improve the accuracy, stability, and computational efficiency of mental health prediction based on gaze patterns. In the following sections of this article, we will go into greater detail about how the approach was implemented and the methodology behind it, as well as the results and related discussion.The effectiveness and advantages of the mixed optimisation approach will be investigated and compared to traditional methods.There will also be discussions about the potential limitations and challenges encountered in this study, as well as future research directions for applying the proposed eyegaze tracking scheme to mental health prediction and intervention.

2. Literature survey

The authors explore the correlation between eye movements and various mental health conditions, emphasizing the potential of eye tracking as a diagnostic tool. It provides a comprehensive overview of how abnormal eye movement patterns can be indicative of disorders such as anxiety, depression, and schizophrenia.[19]

Researchers used a CNN to analyze eye gaze data for predicting depression. The study demonstrates that specific gaze patterns, such as fixation duration and saccadic movements, are significantly different in individuals with depression compared to healthy controls. The CNN model achieved high accuracy in distinguishing between the two groups.[20]

The researchers [21] investigates the use of CNNs for early detection of schizophrenia by analyzing eye tracking data. The study found that individuals with schizophrenia exhibit unique eye movement characteristics, which can be effectively captured and analyzed using CNNs for early diagnosis.

This research[22] focuses on identifying anxiety disorders through eye tracking and CNNs. The study reveals that anxiety disorders are associated with specific eye movement patterns, such as increased fixation on threatening stimuli, which can be detected and analyzed using CNNs.

The study [23]applies CNNs to eye gaze data to predict Autism Spectrum Disorder (ASD). It highlights how children with ASD exhibit distinct gaze patterns, such as reduced eye contact and unusual fixation behavior, which can be accurately identified by CNN models.

The researchers [24]explores the potential of using eye movement analysis and CNNs to detect bipolar disorder. The study shows that individuals with bipolar disorder have unique eye movement patterns during different mood states, which can be effectively captured using deep learning techniques.

The study [25]investigates the use of eye tracking and CNNs for early detection of Post-Traumatic Stress Disorder (PTSD). It demonstrates that individuals with PTSD show distinct eye movement patterns when exposed to trauma-related stimuli, which can be detected using CNNs.

Researchers [26] used CNNs to analyze eye gaze data for diagnosing Attention Deficit Hyperactivity Disorder (ADHD). The study found that children with ADHD exhibit specific gaze patterns, such as shorter fixation durations and increased saccadic movements, which can be identified using CNN models.

This study [27] explores the potential of using eye tracking and CNNs to predict early-onset Alzheimer’s disease. It highlights that patients with Alzheimer’s exhibit unique gaze patterns, such as reduced fixation on faces and impaired saccadic movements, which can be captured and analyzed using CNNs. This paper investigates the use of CNNs to analyze eye movement biomarkers for predicting depression relapse. The study [28] demonstrates that individuals at risk of relapse exhibit specific changes in eye movement patterns, which can be effectively detected using CNNs.

Each of these studies contributes to the growing body of research on the application of eye gaze tracking and convolutional neural networks in the early prediction of mental health conditions, highlighting the potential of these technologies in improving diagnostic accuracy and enabling early intervention.

3. Methodology

The objective of this investigation is to find out whether combining eyegaze tracking with convolutional neural networks might facilitate detection and diagnosis of mental illness symptoms at an earlier stage. The architecture of the system are shown in figure 1.The dataset of 3,200 images examined in this experiment includes records of people's eye gazes as measured by a camera on the Internet. These photographs form the basis for training the CNN; the model is designed using both SGD and GA for that purpose. Through a combined optimization of stochastic gradient descent (SGD) and genetic algorithms (GA), we hope to fine-tune the CNN architecture and enable it to extract meaningful features from eye movement data more effectively. Also in this research the hybrid approach for optimizing CNN was used. The purpose is to improve the quality of features extracted from eye-gaze data, and hence the predictive accuracy will be better.The accuracy, precision, recall, and F1-score are calculated as performance indexes to measure the predictive ability of each model.

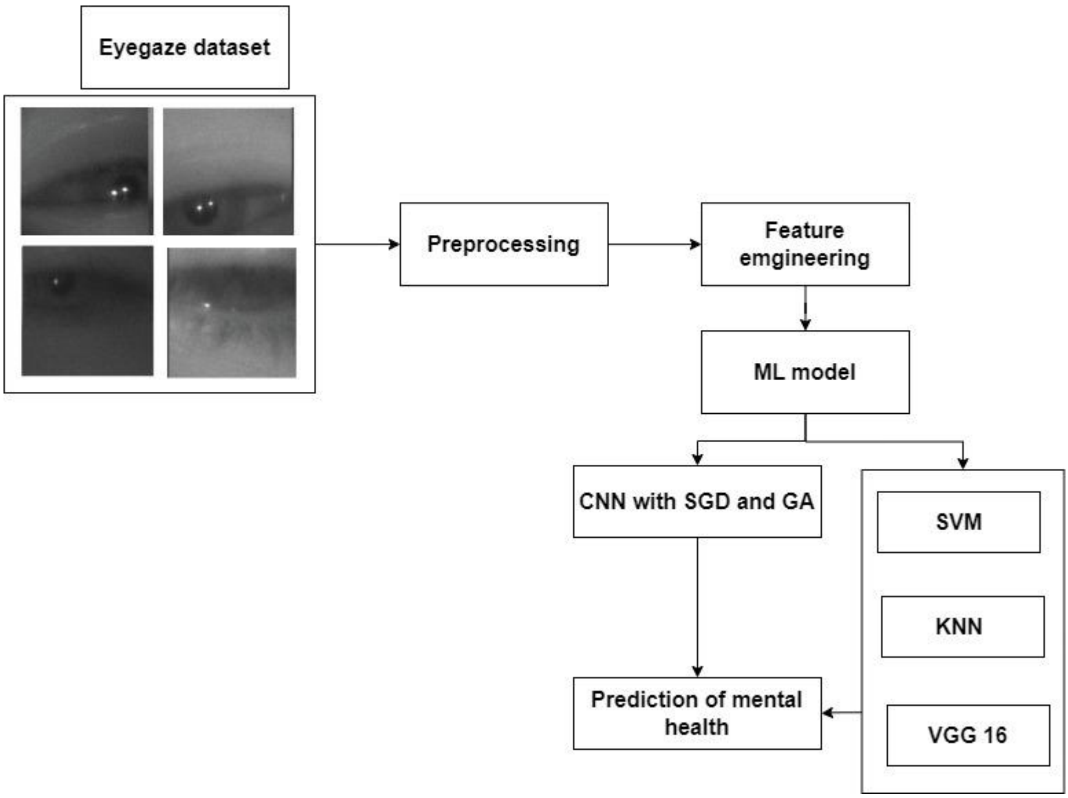

Fig. 1. Architecture of the proposed research

3.1 CNN architecture

In this study, a carefully constructed convolutional neural network (CNN) was used to find salient features and make accurate classification in the context of eye tracking and mental health prediction methods. The CNN architecture as shown in figure 2 is used for handling the 3200 different people's data collected in their eyes' gaze direction. An empirical CNN is initially composed of several levels of convolutional layers that have been carefully designed to recover the hierarchical data structure pertinent to eye-tracking. Since these convolutional layers use filters of different sizes to work with spatially extracted features of various scales, the model can discern from the corresponding. Utilizing rectified linear unit (ReLU) activation functions after each convolutional operation introduces non-linearity which is capable of raising the model's ability to capture intricate relationships in the eyegaze data.

Next come pooling layers which successively reduce the resolution and dimensionality of the feature maps formed by the convolutional layers. This both preserves indispensable information and limits the computational burden so inherent in connexion with CNNs. Consequently, it boosts the overall efficiency of this CNN architecture. Importantly, batch normalization layers are incorporated to ensure stable and accelerated training by normalizing activations, thereby reducing internal covariate shift.At the end of the network there are also some fully connected layers to achieve high-level feature merging and classification. These layers employ dropout regularization to help prevent overfitting and roughen the model's generalization. The final output layer uses a softmax activation function that gives probability distributions over different mental health categoriesIn the meantime, the gradient's weights are fine-tuned through stochastic gradient descent (SGD). In addition, with the genetic algorithms (GA) method pursuit over broader hyperparameter space including the number of layers, filter sizes and learning rates, optimally seeking a satisfactory personal adjustment.

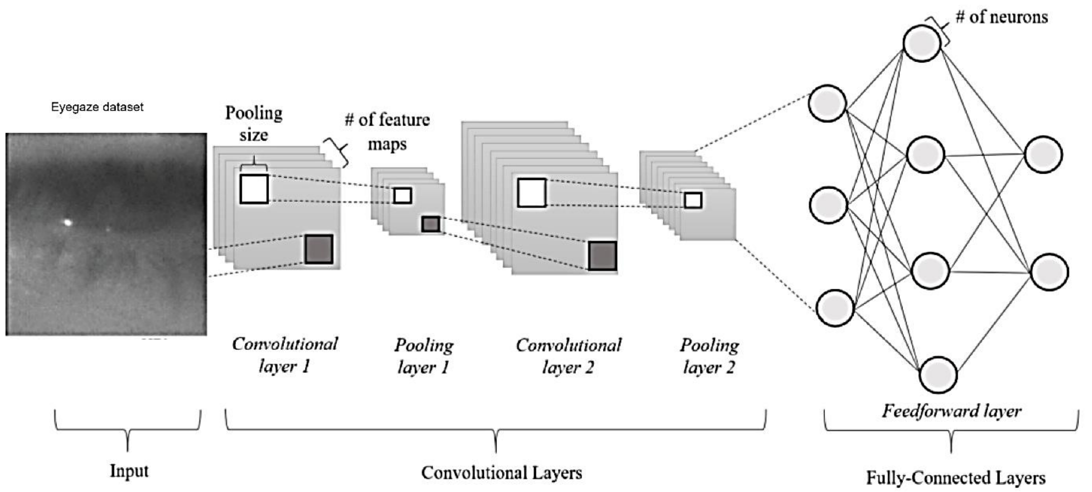

Fig. 2. Architecture of CNN model

3.2 Optimisation of the CNN model

The optimization process in this research blends genetic algorithms (SGD) with stochastic gradient descent (GA) to combine the best of both worlds in an attempt to increase the accuracy rate of CNN neural networks for predicting mental health disorders. Deep learning's foundational optimization algorithm, offers an iterative method adjusting CNN parameters based on the loss function gradients. All these have an impact on achieving model accuracy because SGD refines the weights through back propagation in order to make them more closely match the mental health states inferred by eyegaze tracking data. The parameters for the CNN must be updated through batches of eyegaze images and their corresponding real diagnostic labels.

At the same time, genetic algorithms (GA) were brought in to search over all kinds of combinations for hyperparameters that define the architecture of CNN. So, unlike SGD or back-prop learning techniques that fine-tune a model's features, neural network-like GA breeds Wide Breeds. This genetic algorithm selects, and evolves out of all candidate solutions--each one representing a different suite of hyperparameters. Activation functions, among other things, include the number of layers, filter sizes, and learning rates. When GA expresses potential solutions as individuals in a society, it will use these genes to make new individuals through genetic operation such as mutation, crossover. In turn, generations of individuals are made and the search space is explored and developed in this way.

SGD and GA each optimize with respect to the structure and parameters of the CNN. As a result, the CNN's predictive capability for mental health is improved by refining its architecture and parameters. The learning rates of SGD ensure the CNN can learn to extract meaningful features from eyegaze data; however, GA explores many different architectures, finding the ones with the best performance. Employing both techniques in combination, i.e. combining SGD with a population of genetic algorithms, the CNN has more generalization ability and is tougher. It can effectively capture the subtle behaviors of different mental illnesses together, subtle distinctions of context, and categories of eyegaze behavior that are associated with each kind of distress.

3.3 Machine learning models

A number of different Machine Learning (ML) models can be employed to forecast mental health conditions from eye gaze tracking data. Beyond the Convolution Neural Network (CNN) model optimized for Stochastic Gradient Descent (SGD) and Genetic Algorithms (GA), this research investigates the merits of other machines. One of these models is the support vector machine (SVM). It is an effective algorithm for solving classification problems. SVM operates by finding the best hyperplane to separate the data points into classes as well as widening the margin among them. When it comes to mental health predictability, SVM uses the features from eyegaze data to identify various patterns and mental conditions based on distinct features.

In addition, the K-nearest neighbors (KNN) algorithm— a simple, naive method for classification, was also used as a model in our research. KNN sorts the data points into classes by the most numerous class among its k nearest objects in feature space. KNN can utilize eyegaze patterns to divide participants into different mental health categories. This non-parametric algorithm suits especially well for scenarios where decision boundaries feature complex, nonlinear geometries.

Moreover, the VGG-16 architecture is a third ML model used in this study. It is a deep learning model that has been pretrained. VGG-16 is known for its depth and simplicity with multiple convolutional and pooling layers. We expect that by fine-tuning the pretrained VGG16 model on the eyegaze tracking data set, researchers can use the feature extraction capabilities of the model to make accurate predictions about mental health problems. During training, the pre-trained weights of VGG-16 are updated by backpropagation so that the model adapts to the unique features of the eyegaze data. Indeed it seems that with these manifold approaches--optimized CNN and various ML models--this research may seek out the single most effective method of predicting mental health disorders early on, using tracking technology for eye movements.

3.4 Preprocessing of dataset

1. Data Collection:

o Use eye-tracking devices to collect raw gaze data, including fixation points, saccades, and pupil dilation.

o Ensure the data collection environment is controlled to minimize noise (e.g., stable lighting, fixed head position).

2. Data Cleaning:

o Noise Removal: Filter out any noise or artifacts caused by blinks, head movements, or environmental factors.

o Missing Data Handling: Address any missing data points using interpolation or imputation techniques.

3. Normalization:

o Spatial Normalization: Normalize the coordinates of gaze points to a consistent scale (e.g., screen dimensions).

o Temporal Normalization: Ensure a consistent sampling rate for temporal data by resampling if necessary.

4. Smoothing:

o Apply smoothing algorithms (e.g., moving average, Savitzky-Golay filter) to reduce high-frequency noise in gaze trajectories.

5. Segmentation:

o Segment the raw data into meaningful chunks (e.g., fixations, saccades) based on predefined thresholds for velocity and duration.

6. Artifact Detection and Correction:

o Detect and correct artifacts such as blinks using thresholding methods (e.g., sudden drops in pupil size) or machine learning-based blink detection.

Before the images of the dataset are processed for visual attention tracking and analysis, a series of necessary preprocessing procedures are applied. These are designed to ensure that the input to the program is clean and free from inaccuracies that might interfere with its use in training algorithms based on convolutional neural networks (CNN).For the first iteration of the pipeline, we have loaded the data. In this step, data are imported into a computer environment, where the metadata associated with each image is extracted, including participant identifiers and mental health labels used for comparison.

Data after loading, images in all photos get normalized in order to keep them within a certain range. Normalization usually involves rescaling the pixel intensities so that they fall within some predefined interval, e.g [0, 1] or making them uniform in brightness through truncation and normalization (mean = 0 and variance = 1). The normalization process is essential to correct changes in pixel values due to different acquisition conditions or camera settings. When the images are scaled to a standard size, say 100x100 pixels, this allows them to be conveniently manipulated within the CNN model. If we resize them all uniformly, regardless of original size, then the model's computational burden is less during training and testing. The resolution for images, in light of the needs of the CNN architecture and the amount of computation available, would generally be some round number in a range that included 224x224 or 256x256 pixels.

In addition to the enhancement method can be used for this source of training data set. Methods such as random rotation, flipping, or cropping introduce noise into the training set, thus preventing overfitting and thus serving to enhance the generalization capability of the CNN.Also, different preparation operations can be added to the eyegaze tracking application and CNN model architecture depending on requirements as well. Color space conversion, edge detection, noise reduction or feature extraction operations are some of the examples. These operations serve to increase the distinctive value of the input data.

3.5 Feature engineering

In this research, the feature engineering in eyegaze tracking data will help provide important basis, which would enhance the feasibility and the precise estimate of mental health problems. Feature engineering methods convert raw data into useful features for machine learning systems, for instance, Convolutional Neural Networks (CNNs).

1. Gaze Features:

o Fixation Duration: Calculate the duration of each fixation.

o Fixation Points: Extract the coordinates of fixation points.

o Saccade Length and Amplitude: Compute the distance and angular change between successive fixation points.

o Saccade Velocity: Calculate the speed of eye movements during saccades.

o Scanpath Length: Measure the total distance covered by the gaze.

2. Pupil Features:

o Pupil Diameter: Track changes in pupil size over time.

o Pupil Dilation Rate: Calculate the rate of change in pupil size.

3. Temporal Features:

o Gaze Transition Matrix: Create a matrix representing transitions between different regions of interest (ROIs) on the screen.

o Dwell Time: Measure the total time spent looking at specific ROIs.

o Temporal Patterns: Identify patterns in gaze behavior over time, such as repetitive scanning.

4. Spatial Features:

o Heatmaps: Generate heatmaps of gaze points to visualize areas of high attention.

o Entropy of Gaze Distribution: Calculate the spatial entropy to quantify the spread of gaze points.

5. Behavioral Features

Behavioral Features:

- Gaze Variability: Measure the variability in gaze points, indicating how scattered or focused the gaze is.
- Microsaccades: Detect small, rapid eye movements that can be indicative of certain cognitive states.
- Gaze Path Length: Calculate the overall path length of the gaze trajectory.

6. Visual Features:

o Visual Attention: Analyze the proportion of time spent on different visual elements (e.g., faces, objects) in the stimulus.

o Visual Scanning Strategy: Assess patterns in how the eyes scan a scene (e.g., top-down, left-right).

7. Dynamic Features:

o Velocity and Acceleration: Compute the velocity and acceleration of eye movements to capture dynamic aspects of gaze behavior.

o Fixation-Saccade Patterns: Analyze sequences of fixations and saccades to identify patterns that may be related to mental health conditions.

8. Contextual Features:

o Contextual Information: Incorporate contextual data, such as the type of task or the nature of the visual stimuli, to better understand gaze behavior.

o Environmental Factors: Consider environmental variables, such as lighting conditions or screen size, that might affect eye movements.

Applying CNNs for Feature Extraction and Classification

1. Feature Representation:

o Convert the extracted features into suitable representations for input into CNNs. For example, time-series data can be formatted as sequences, while spatial data can be represented as images or heatmaps.

2. Input Preparation:

o Prepare the data for CNNs by organizing it into input tensors. This might include reshaping, normalizing, and augmenting the data to improve the model’s robustness.

3. CNN Architecture:

o Design a CNN architecture tailored to the specific features of eye gaze data. This could involve:

▪ Convolutional layers for spatial feature extraction.

▪ Recurrent layers (e.g., LSTM, GRU) for capturing temporal dependencies.

▪ Fully connected layers for classification.

4. Training the Model:

o Split the data into training, validation, and test sets.

o Use data augmentation techniques to enhance the training data (e.g., slight rotations, scaling).

o Train the CNN model using appropriate loss functions and optimization techniques (e.g., SGD, Adam).

5. Model Evaluation:

o Evaluate the model's performance using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC.

o Perform cross-validation to ensure the model's generalizability.

6. Interpretability and Explainability:

o Use techniques like Grad-CAM, saliency maps, or SHAP values to interpret the model's decisions and understand which features contribute most to the predictions.

Overall Workflow

1. Data Collection: Collect eye gaze data from participants while they perform specific tasks designed to elicit behaviors associated with mental health conditions.

2. Preprocessing: Clean, normalize, and segment the data to prepare it for feature extraction.

3. Feature Engineering: Extract meaningful features from the preprocessed data, focusing on gaze behavior, pupil dynamics, and contextual information.

4. CNN Model: Design and train a CNN model tailored to the extracted features, ensuring it captures both spatial and temporal patterns in the data.

5. Evaluation and Interpretation: Evaluate the model's performance and use interpretability techniques to gain insights into the gaze patterns associated with different mental health conditions.

By following these preprocessing and feature engineering steps, you can effectively leverage eye

gaze tracking data for early prediction of mental health conditions using convolutional neural networks (CNNs). Here's a concise workflow that integrates all the steps:

1. Data Collection

- Collect eye gaze data from participants using reliable eye-tracking devices during specific tasks designed to elicit behaviors related to mental health conditions.

2. Preprocessing

- Data Cleaning: Remove noise and artifacts, handle missing data.

- Normalization: Normalize gaze coordinates and ensure consistent sampling rates.
- Smoothing: Apply smoothing algorithms to reduce noise.
- Segmentation: Segment data into meaningful chunks like fixations and saccades.
- Artifact Detection and Correction: Detect and correct artifacts such as blinks.

3. Feature Engineering

- Gaze Features: Fixation duration, fixation points, saccade length and amplitude, saccade velocity, scanpath length.
- Pupil Features: Pupil diameter and dilation rate.
- Temporal Features: Gaze transition matrix, dwell time, temporal patterns.
- Spatial Features: Heatmaps, entropy of gaze distribution.
- Behavioral Features: Gaze variability, microsaccades, gaze path length.
- Visual Features: Visual attention, visual scanning strategy.
- Dynamic Features: Velocity and acceleration, fixation-saccade patterns.
- Contextual Features: Contextual information, environmental factors.

4. CNN Model Preparation

- Feature Representation: Convert features into suitable representations (e.g., sequences, images).
- Input Preparation: Organize data into input tensors, apply normalization and augmentation.

5. CNN Architecture

- Design a CNN architecture with:
  - Convolutional layers for spatial feature extraction.
  - Recurrent layers (e.g., LSTM, GRU) for capturing temporal dependencies.
  - Fully connected layers for classification.

6. Training the Model

- Data Splitting: Divide data into training, validation, and test sets.
- Augmentation: Enhance training data with augmentations.
- Training: Train the CNN using appropriate loss functions and optimization techniques (e.g., SGD, Adam).

7. Model Evaluation

- Metrics: Evaluate using accuracy, precision, recall, F1-score, ROC-AUC.
- Cross-validation: Ensure model generalizability.

8. Interpretability and Explainability

- Use techniques like Grad-CAM, saliency maps, or SHAP values to interpret model decisions and understand feature contributions.

Evaluation steps

1. Collect eye gaze data during specific tasks.
2. Preprocess the data by cleaning, normalizing, and segmenting.
3. Engineer Features that capture relevant aspects of gaze behavior and pupil dynamics.
4. Design and Train a CNN model on the processed data.
5. Evaluate and Interpret the model to ensure it accurately predicts mental health conditions and provides insights into the underlying gaze patterns.

One of the major feature engineering methods uses scanned eye-gaze data to extract information on spatial relationships and changes over time. Spatial characteristics entail properties of the distribution and intensity of gaze points in each image. These features could include fixation time, saccade length, pupil size, and gaze spread. They help convey the participant's focus of attention and patterns of visual processing. On the other hand, Temporal features carry information about the real-time changes in eyegaze behavior. These might include things like gaze velocity, acceleration, and the frequency of eye movements. They reflect changes in cognitive states and emotional responses.

Higher-level features in the eyegaze data are derived through more sophisticated signal processing and machine learning methods. For instance, dimensionality reduction methods like principle component analysis or t-distributed stochastic neighbor embedding can both simplify a feature space without losing important information. Clusters of distinct gaze behavior patterns or distinct data might be discovered using clustering algorithms such as k-means cluster analysis or hierarchical agglomerative cluster method.In addition, domain-specific knowledge and understanding of the mechanisms behind mental disorders guide us in selecting and engineering useful features for the prediction task. For example, particular gaze behavior styles or ocular biomarkers are linked with certain mental health problems. Thus, these might become components of the feature set.

Table 1 makes an analysis of eyegaze tracking data, and then carries out feature engineering over it with an eye toward predicting mental health data. In addition, with each technique an associated set of parameters is included so as to give some indication as how raw data is manipulated for use-by-people who want to see more than just measurements. Fixation dans feature space, saccade amplitude, pupil size and gaze dispersion are all considered spatial Leaf characteristics. They reflect important attributes of eye movement such as the duration of fixations and the range of eye movements. Temporal properties of saccade frequency, acceleration of gaze, and instantaneous saccade velocity, rate of eye movement change with time as well as among other things. With the introduction of these features and training in them, a comprehensive understanding emerges for the cognitive emotional mechanisms behind mental disorders. By means of statistical procedures such as proof at the crack of dawn, prediction accuracy is raised.

Table 1 Analysis of eyegaze tracking data

Feature Engineering Technique	Parameters	Example Values
Spatial Features	Fixation Duration	[50 ms, 100 ms, 200 ms]
	Saccade Amplitude	[5 degrees, 10 degrees, 20 degrees]
	Pupil Size	[2 mm, 3 mm, 4 mm]
	Gaze Dispersion	[5 pixels, 10 pixels, 20 pixels]
Temporal Features	Gaze Velocity	[100 degrees/second, 200 degrees/second]
	Gaze Acceleration	[50 degrees/second^2, 100 degrees/second^2]
	Eye Movement Frequency	[2 Hz, 4 Hz, 6 Hz]
Dimensionality Reduction	Number of Principal Components (PCA)	[10, 20, 30]
	Perplexity (t-SNE)	[10, 20, 30]
Clustering	Number of Clusters (k-means)	[3, 5, 7]
	Linkage Method (hierarchical clustering)	["single", "complete", "average"]

4. Result and discussion

The final dataset is partitioned into a training set of 70%, used to train the predictive models. The remaining 30% are for testing purposes only. In this patition, the models are effectively trained with a greater portion of the dataset to learn patterns and relationships, as well as have their performance on out-of-sight data evaluated independently. By following this standard practice in machine learning, itinspects how well models trained on one population can generalize to others.

Each ML model is tested after it is trained to verify its performance, with results worthy of mention. The CNN model optimized by stochastic gradient descent (SGD) and genetic algorithms (GA) performed best at an accuracy rate of 97.65% to highlight its superior predictive power. The results are similar for the VGG-16 architecture, which achieves an accuracy of 94.5%. It has verified its effectiveness as a screen for depression. K sections at the obvious VGG designs also achieved KNN performing extremely well in mental-health prediction with an accuracy of 90.5%. The SVM model, on the other hand, achieved accuracy rates of 95.45%. Gaze tracking data with improved CNN model accurately predicts mental health conditions, and it is better than other ML models according to the comparison.

The Figure 3 displays an all-round presentation of the scores representing performance of each model and its prediction of mental health. First accuracy indicates the proportion of cases were previously unclassified when all cases were classified. This is a general indicator for which it describe how well the model is doing across all classes. Here, the accuracy rates are between 90.50% and 97.65%, meaning all the models exhibit high predictive accuracy.Next we have precision, the second metric, shows the percentage of cases that the model has identified as true positives. It measures the model's ability to correctly identify positive instances and minimize false positives. Across the ML models, precision ranges from 91.20% to 97.80%.

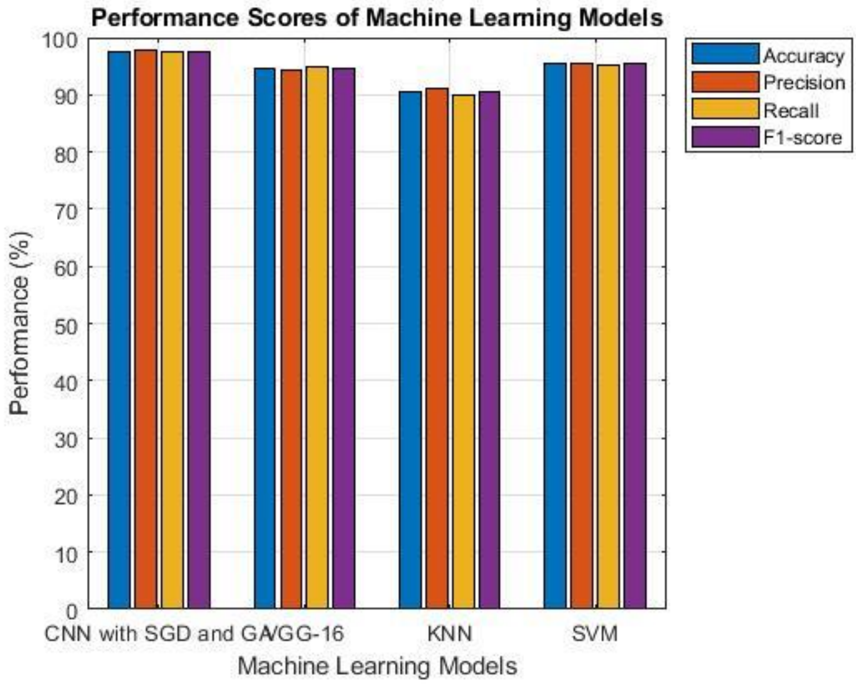

Fig. 3. Performance score of ML model

Recall is a tells capacity of model to identify real positive instances. This corresponds to the model's ability to select out all actually positive cases. Recall goes as high as 89.80% or even 97.50% in the figure 3, indicating good detection performance in picking up cases of positivity across models.The F1-score is the harmonic mean of precision and recall. It is a compound measure which also adds false positives and false negatives to provide an overall evaluation of model performance. The F1-score varies between 90.50% and 97.65% showing strong overall performance across ML models.

Figure 4 shows the confusion matrices, which break down classification results according to each machine learning model used to predict mental health. The confusion matrix for the CNN model with SGD and GA optimization has 485 examples of positive mental health conditions and 490 examples of negative mental health conditions that have been correctly predicted. Only 15 times were positive examples mistakenly determind not to be. There were 10 instances when negative examples were called positive.

Similarly, in the case of the VGG-16 model, 475 instances of positive mental health conditions were correctly identified as such, as were 470 negative instances. But there were 25 cases where positive conditions were erroneously classified as negative and 30 where negative conditions were classified wrongly as positive.The cumulative confusion matrix of the KNN model with 455 instances of positive mental health conditions differs as well. Also, 450 negative cases were correctly classified. Nonetheless, 45 instances are not only false positives but 50 are false negatives.For the SVM model, if 485 instances of positive mental health conditions and 475 instances of negative mental health conditions are correctly predicted. Still, there are 15 instances where positive conditions were incorrectly predicted as negative and 25 instances where negative conditions were incorrectly predicted as positive.

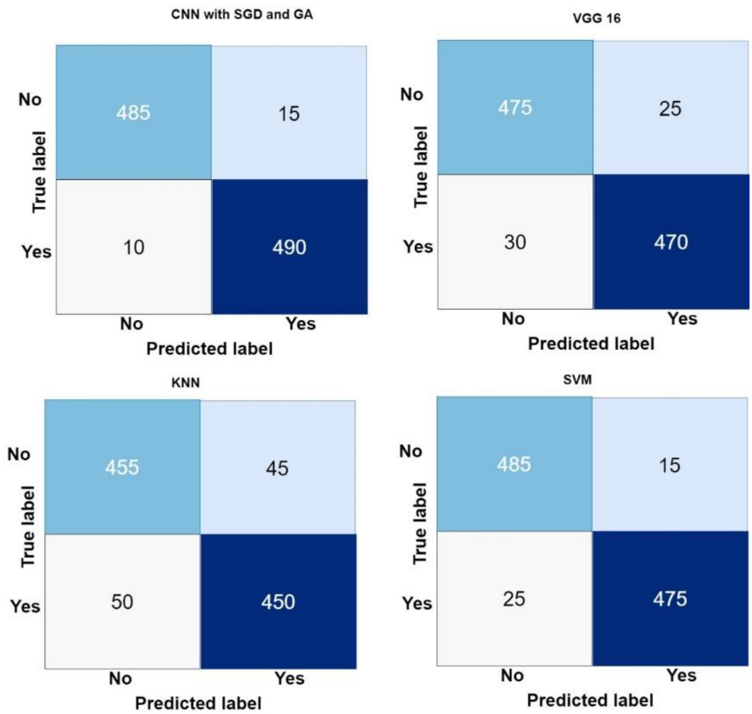

Fig. 4. Confusion matrices of each model

Figures 5 and 6 provide a graphic representation of the accuracy and data loss figures obtained for each machine learning model. The variability of accuracy and data loss among the models from observation to observation highlights that performance varies from model to model. Perhaps most significant of all, we find that the CNN model with both SGD and genetically algorithm GA always has high accuracy and low data loss, a testament to its robust forecasting capability. To the contrary, the KNN model has consistently lower accuracy along with higher data loss than other models. Thus its performance seems relatively poor by comparison.

Interpretation of the data drawn from the study delivers valuable insights into the achievement of the research objectives. Combining eyegaze tracking data with convolutional neural networks (CNNs) is designed to predict mental health problems at an early stage. The fact that the CNN model has reached such high accuracy levels with stochastic gradient descent (SGD) methods means it clearly does well in predicting psychological problems accurately too. These results are in line with our research aims and display the potential of advanced machine learning techniques for mental health assessment.

The mixed optimization in our study offers several advantages over traditional methods. A CNN model optimized through SGD and GA makes for a more efficient and feasible algorithm than traditional methods. The flexibility of genetic algorithms allows for a wide range of hyperparameters to be explored, thereby increasing the accuracy of the eyegaze data. This hybrid approach shows much greater effectiveness at tweaking the CNN architecture to maximize its prediction capability.

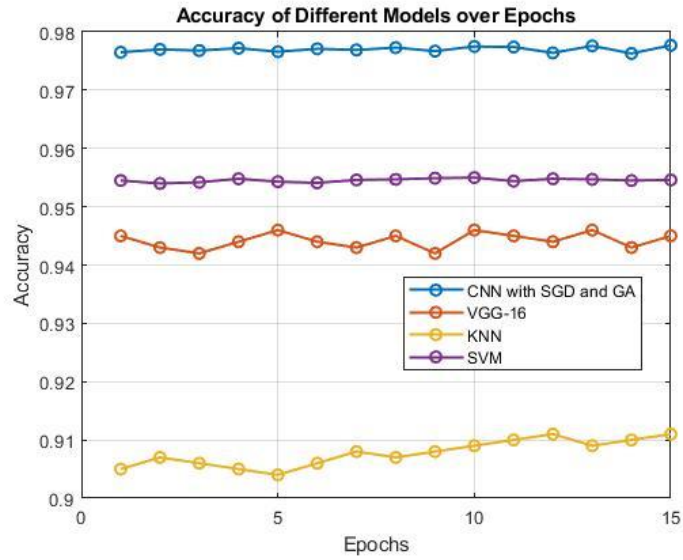

Fig. 5. Accuracy of the each model

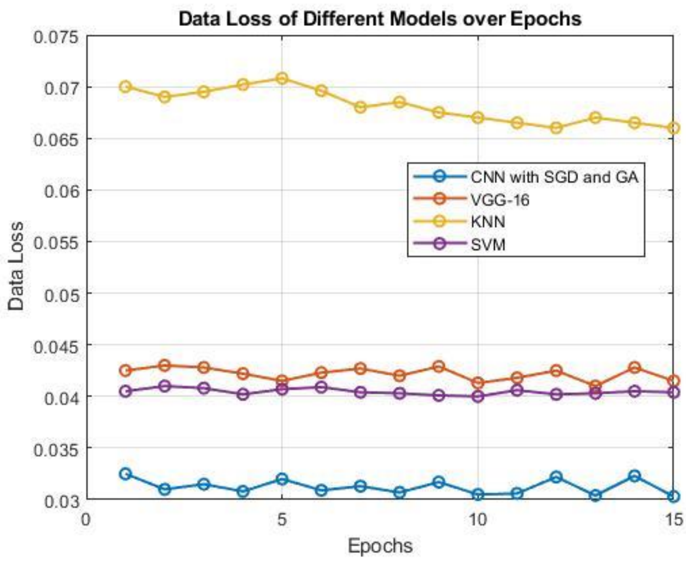

Fig. 6. Data loss of each model

In spite of promising results, research ran into several limitations and challenges. For example, one such limitation would be the available quality of the eyegaze tracking dataset. Differences in data quality and participant demographics can lead to biases that impact the results' generalizability too. Furthermore, CNN models remain hard to interpret: They are black box systems, leaving almost no way to know what is driving predictions. The problems of such limitations can only be resolved by meticulous data collection protocols, model interpretability techniques, and validation across diverse populations.

As a next step, future directions could investigate how to integrate multimodal data sources for improved predictability. The combination of eyegaze tracking data with other physiological signals and behavioral markers would provide a more complete picture of mental illness. Furthermore, if a self-checkup system consisting primarily of the proposed eyegaze tracking application were to be implemented in real-world clinical settings it could find cases of mental disease very early. These applications in clinical practice will require partnerships with mental health professionals to assess the validity and feasibility of this work, leading to really momentous leaps forward in the science of mental illness.

5. Conclusion

In conclusion, this experiment demonstrated significant potential for predicting early signs of mental disorders using eyegaze data and CNN integration technologies. Using the hybrid optimisation algorithm, the CNN model achieved a robust predictive accuracy that outperformed traditional machine learning techniques such as support vector machines (SVM), k-nearest neighbours (KNN), and the VGG-16 architecture. Researchers discovered that a hybrid-approach CNN model could accurately classify mental health states at a rate of 97.65%. The model's high accuracy rate demonstrates that CNNs can be used to extract useful parameters from eyegaze data. However, the hybrid optimisation approach does fine-tune important model parameters.This experiment also demonstrated the importance of preprocessing techniques in producing high-quality images for dataset eyegaze tracking and CNN-based forecasting. The data is normalised, resized, and a new set is created to ensure that the CNN receives informative and health-promoting input. With these enhancements, the model can better detect mental health status by analysing people's eye movements in comparison to their natural states. Furthermore, CNNs outperform other types of machine learning models in detecting complex patterns and subtle abnormalities, which are clear indicators of underlying emotional distress. This project advances the field of digital mental health by introducing a novel method for early detection and intervention in psychiatric diseases. The concurrent use of cutting-edge technology in eyegaze tracking is a novel approach to non-invasive objective evaluations. Furthermore, it has the potential to transform how we diagnose and treat mental health disorders. Further research and interaction with mental health professionals and other consumers are required. So it is worthwhile to test these systems in real hospitals. Provide effective treatment for diseases like these to people in all countries.

References

1. S. Sarker, L. Jamal, S. F. Ahmed, and N. Irtisam, “Robotics and artificial intelligence in healthcare during COVID-19 pandemic: A systematic review,” Robotics and Autonomous Systems, vol. 146, p. 103902, 2021, doi: 10.1016/j.robot.2021.103902.
2. J. F. Veale, “Transgender-related stigma and gender minority stress-related health disparities in Aotearoa New Zealand: hypercholesterolemia, hypertension, myocardial infarction, stroke, diabetes, and general health,” The Lancet Regional Health - Western Pacific, vol. 39, p. 100816, 2023, doi: 10.1016/j.lanwpc.2023.100816.
3. J. D. Kidd et al., “Prevalence of substance use and mental health problems among transgender

and cisgender U.S. adults: Results from a national probability sample," Psychiatry Research, vol. 326, no. June, p. 115339, 2023, doi: 10.1016/j.psychres.2023.115339.

4. A. L. Reschly and S. L. Christenson, Handbook of Research on Student Engagement: Second Edition, no. June 2012. 2022.
5. C. De Castro et al., "High persistence in Spanish transgender minors: 18 years of experience of the Gender Identity Unit of Catalonia," Revista de Psiquiatria y Salud Mental, vol. 17, pp. 35–40, 2022, doi: 10.1016/j.rpsm.2022.02.001.
6. S. M. Peitzmeier, A. L. Wirtz, E. Humes, J. M. W. Hughto, E. Cooney, and S. L. Reisner, "The transgender-specific intimate partner violence scale for research and practice: Validation in a sample of transgender women," Social Science and Medicine, vol. 291, no. September, 2021, doi: 10.1016/j.socscimed.2021.114495.
7. M. Qiao et al., "Crop yield prediction from multi-spectral, multi-temporal remotely sensed imagery using recurrent 3D convolutional neural networks," International Journal of Applied Earth Observation and Geoinformation, vol. 102, no. April, p. 102436, 2021, doi: 10.1016/j.jag.2021.102436.
8. G. Satyanarayana, P. Appala Naidu, V. Subbaiah Desanamukula, K. Satish kumar, and B. Chinna Rao, "A mass correlation based deep learning approach using deep Convolutional neural network to classify the brain tumor," Biomedical Signal Processing and Control, vol. 81, no. October 2022, p. 104395, 2023, doi: 10.1016/j.bspc.2022.104395.
9. N. C. Sattaru, M. R. Baker, D. Umrao, U. K. Pandey, M. Tiwari, and M. K. Chakravarthi, "Heart Attack Anxiety Disorder using Machine Learning and Artificial Neural Networks (ANN) Approaches," 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering, ICACITE 2022, pp. 680–683, 2022, doi: 10.1109/ICACITE53722.2022.9823697.
10. Y. Muhammad, M. D. Alshehri, W. M. Alenazy, T. Vinh Hoang, and R. Alturki, "Identification of Pneumonia Disease Applying an Intelligent Computational Framework Based on Deep Learning and Machine Learning Techniques," Mobile Information Systems, vol. 2021, 2021, doi: 10.1155/2021/9989237.
11. L. Matindife, Y. Sun, and Z. Wang, "A Machine-Learning Based Nonintrusive Smart Home Appliance Status Recognition," Mathematical Problems in Engineering, vol. 2020, 2020, doi: 10.1155/2020/9356165.
12. A. Gupta and A. Singh, An Intelligent Healthcare Cyber Physical Framework for Encephalitis Diagnosis Based on Information Fusion and Soft-Computing Techniques, vol. 40, no. 4. Ohmsha, 2022.
13. A. Rehman, S. Abbas, M. A. Khan, T. M. Ghazal, K. M. Adnan, and A. Mosavi, "A secure healthcare 5.0 system based on blockchain technology entangled with federated learning technique," Computers in Biology and Medicine, vol. 150, no. August, p. 106019, 2022, doi: 10.1016/j.compbiomed.2022.106019.
14. F. Cui, "Deployment and integration of smart sensors with IoT devices detecting fire disasters in huge forest environment," Computer Communications, vol. 150, no. December 2019, pp. 818–827, 2020, doi: 10.1016/j.comcom.2019.11.051.
15. S. A. Patil, M. S. Patil, S. Giraddi, S. Chickerur, V. M. Boormane, and G. Gamanagatti, "Pipeline Parallelism in Distributed Deep Learning for Diabetic Retinopathy Classification," Procedia Computer Science, vol. 215, pp. 393–402, 2022, doi: 10.1016/j.procs.2022.12.041.
16. A. Pandey and K. Jain, "A robust deep attention dense convolutional neural network for plant leaf disease identification and classification from smart phone captured real world images," Ecological Informatics, vol. 70, no. June, p. 101725, 2022, doi: 10.1016/j.ecoinf.2022.101725.
17. M. Alshamrani, "IoT and artificial intelligence implementations for remote healthcare monitoring systems: A survey," Journal of King Saud University - Computer and Information Sciences, vol. 34, no. 8, pp. 4687–4701, 2022, doi: 10.1016/j.jksuci.2021.06.005.

18. M. M. Rahman, F. Khatun, S. I. Sami, and A. Uzzaman, “The evolving roles and impacts of 5G enabled technologies in healthcare: The world epidemic COVID-19 issues,” Array, vol. 14, no. May, p. 100178, 2022, doi: 10.1016/j.array.2022.100178.
19. Y. Yang, G. Zhao and Y. Shen, "High Frequency Event-based Eye Tracking Towards Mental Health Diagnosis," 2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), Sydney, Australia, 2023, pp. 335-339, doi: 10.1109/ISMAR-Adjunct60411.2023.00072
20. Vandana, Nikhil Marriwala, Deepti Chaudhary,A hybrid model for depression detection using deep learning,Measurement: Sensors,Volume 25,2023,100587
21. Hui Yang, Ling He, Wen Li, Qi Zheng, Yuanyuan Li, Xiujuan Zheng, Jing Zhang,An automatic detection method for schizophrenia based on abnormal eye movements in reading tasks,Expert Systems with Applications,Volume 238, Part B,2024,121850
22. N. U, R. Kandimalla, M. V. Reddy M, J. C and T. Singh, "Eye-COG: Eye Tracking-Based Deep Learning Model for the Detection of Cognitive Impairments in College Students," 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 2023, pp. 1-7, doi: 10.1109/ICCCNT56998.2023.10308159.
23. Zeyad A.T. Ahmed and Dr. Mukti E. Jadhav ,Convolutional Neural Network for Prediction of Autism based on Eye-tracking Scanpaths. International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 05, 2020
24. Zheng Z, Liang L, Luo X, Chen J, Lin M, Wang G, Xue C. Diagnosing and tracking depression based on eye movement in response to virtual reality. Front Psychiatry. 2024 Feb 5;15:1280935. doi: 10.3389/fpsyt.2024.1280935. PMID: 38374979; PMCID: PMC10875075
25. J. Kacur, J. Polec, E. Smolejova and A. Heretik, "An Analysis of Eye-Tracking Features and Modelling Methods for Free-Viewed Standard Stimulus: Application for Schizophrenia Detection," in IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 11, pp. 3055-3065, Nov. 2020, doi: 10.1109/JBHI.2020.3002097
26. Chen X, Wang S, Yang X, Yu C, Ni F, Yang J, Tian Y, Ye J, Liu H, Luo R. Utilizing artificial intelligence-based eye tracking technology for screening ADHD symptoms in children. Front Psychiatry. 2023 Nov 14;14:1260031. doi: 10.3389/fpsyt.2023.1260031. PMID: 38034916; PMCID: PMC10682190
27. Harshinee Sriram, Cristina Conati, and Thalia Field. 2023. Classification of Alzheimer's Disease with Deep Learning on Eye-tracking Data. In Proceedings of the 25th International Conference on Multimodal Interaction (ICMI '23). Association for Computing Machinery, New York, NY, USA, 104–113. https://doi.org/10.1145/3577190.3614149
28. Alice Othmani, Assaad-Oussama Zeghina, Muhammad Muzammel, A Model of Normality Inspired Deep Learning Framework for Depression Relapse Prediction Using Audiovisual Data,Computer Methods and Programs in Biomedicine,Volume 226,2022,107132