|| Volume 9 || Issue 4 || April 2025 || ISSN (Online) 2456-0774
INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH
AND ENGINEERING TRENDS

Generic Artificial Intelligent Agent Using Iot And Deep Learning

Vaishnavi Satish Mahajan[1] , Dhawal Pravin Choudhary[2] ,Mayur Pravin Patil[3] , Sivaram Ponnusamy[4]
School of Computer Sciences and Engineering Sandip University, Nashik, Maharashtra, India.[1,2,3,4]

Abstract: the rapid integration of the Internet of Things (IoT) with artificial intelligence has unlocked new opportunities to develop adaptable, multi-domain artificial intelligence (AI) agents. However, the design of many AI agents for specific tasks limits their ability to generalize across different applications and environments. This paper introduces a generic artificial intelligent agent that utilizes IoT and deep learning, enabling autonomous adaptation to diverse domains such as smart homes, healthcare, industrial automation, and agriculture. The agent leverages IoT sensors for real-time data collection, while deep learning models process and analyze this data to make intelligent, context-aware decisions. Using a combination of initial training and domain adaptation techniques, the agent can learn to recognize patterns and perform tasks across diverse environments. This research proposes an intelligent facial identification and recognition system powered by deep learning. It automatically updates the identification records of individual personnel based on the results generated by the recognition process. Utilizing deep learning models, the system performs both facial recognition and object detection, ensuring high accuracy and reliable performance.

Keywords: *natural language processing (NLP),deep learning, IoT, Face Detection and convolutional neural networks (CNN)*

I.INTRODUCTION:

GAIA's core functionality revolves around its ability to authenticate individuals upon their entry into the office. By employing advanced facial recognition algorithms, GAIA swiftly checks whether an individual is authorized to be on the premises. If the system identifies an unauthorized person, it immediately captures their image and securely stores it on a server for subsequent review and action. This proactive security measure not only safeguards the office environment but also maintains a comprehensive log of unauthorized entries, which can be valuable for future analysis and decision-making. In addition to its security features, GAIA excels in enhancing user experience through its interactive capabilities. Leveraging sophisticated natural language processing (NLP) and deep learning algorithms, GAIA can hold meaningful conversations with visitors, addressing their queries, providing necessary information, and guiding them to appropriate resources or personnel within the office. This interaction mimics human-like assistance, making visitors feel welcome and well-informed, while simultaneously reducing the burden on human staff.

II.LITERATURE SURVEY

In [1], we detail the method for assessing learners' cognitive states during an online learning session. Confusion, dissatisfaction, contentment, and frustration are four intricate emotions identified as amalgamations of primary emotions. We have shown that more information is required to assess emotion from a single image capture. To assess the learner's cognitive condition, we analyzed a sequence of six images captured by the camera. The consideration of six image frames is also shaped by the notion that human emotions are continuous rather than discrete; they do not shift abruptly but need some time, although minimal, to transform.

According to [2], the machine learning approach employs distinct, salient extracted features to model the face. Consequently, it can only accurately discern emotions if the characteristics are meticulously crafted and based on prior information. Scholars from other disciplines have undertaken extensive research on the computational modeling of human emotion. Nonetheless, the human visual system needs more advancement. This research first recognized the eye and lip areas of a face using the Viola-Jones method, followed by a neural network. Furthermore, neural network techniques, deep learning frameworks, and machine learning methodologies are used for emotion recognition.

[3] presents a review of the current status of human emotional stress indicators and identifies the major prospective biomarkers for future wearable affective system sensors. Emotional stress has been recognized as a critical contributor to social challenges, including quality of life, crime, health, and the economy. Traditional methods such as electroencephalography, monitoring physiological parameters, and testing blood cortisol levels are widely regarded as the benchmarks for stress assessment. However, these techniques can be intrusive or uncomfortable, making them impractical for continuous monitoring through wearable devices. Alternatively, cortisol present in bodily fluids and volatile organic compounds released from the skin provide dependable biomarkers that sensors can use to identify episodes of emotional stress.

Graph-based methodologies for emotion recognition use face landmarks, as stated in [4]. We implemented a series of pre-processing phases in line with the plan. Facial key points need pre-processing prior to the extraction of facial emotion features. To recognize emotions on faces that are partially hidden, the main steps involve using the Haar-Cascade method to find faces, applying a MediaPipe model to map facial features, and teaching the model to recognize seven different emotions. The emotion recognition model was trained using the FER-2013 dataset, specifically targeting faces without coverings. We focused on the upper facial region by placing markers to identify key features. Once facial landmarks were detected, we recorded the coordinates corresponding to various emotional expressions and saved this data in a CSV file. Model weights were then assigned to represent

the dynamic emotional categories. To evaluate performance, we implemented a webcam-based application that tested the landmark-driven emotion recognition model on both static images and live video feed, focusing on the upper face.

According to [5], the single-network approach using the FER2013 dataset achieves the highest classification accuracy. "Leveraging the VGGNet framework, we meticulously adjusted its hyperparameters and experimented with diverse optimization strategies. Facial emotion recognition involves detecting expressions that reflect fundamental emotions like joy, fear, and contempt. This technology significantly improves human-computer interaction and is utilized across various fields, including analyzing customer feedback, interactive online gaming, digital advertising, and healthcare services.

In [6] a facial appearance recognition method that leverages convolutional neural networks (CNNs) in conjunction with edge detection techniques, effectively eliminating the need for manual feature extraction commonly associated with traditional approaches. After normalizing facial expression images, the convolutional layers are used to extract the boundary information of each image layer. The resulting edge features are applied to the corresponding feature maps, ensuring the retention of critical structural details from the texture. To classify facial expressions, a Softmax classifier is utilized, and test sample images are analyzed accordingly. Simulation experiments are carried out by integrating the FER-2013 dataset with the LFW dataset to evaluate the effectiveness and accuracy of the approach in recognizing facial expressions under challenging and diverse conditions.

As stated in [7], conversational skills in language learners can be evaluated using multimodal tasks that integrate spoken content, speech intonation, and visual signals. While extensive research has been conducted on linguistic and auditory aspects, visual elements like facial expressions and eye gaze remain comparatively underexplored. To address this, we compiled a dataset consisting of 210 online video interviews featuring Japanese learners of English, each paired with evaluations of their speaking skills. This dataset served as the foundation for developing an automated system that assesses speaking proficiency using a combination of multimodal features.

According to [8], The self-management interview software employs a multi-block deep learning framework to detect user emotions, offering an alternative to traditional approaches that analyze entire facial images. Instead of relying on holistic facial analysis, this method segments the face into multiple blocks and applies AdaBoost learning across these segments. It also incorporates similarity assessment techniques to efficiently screen and validate image blocks. To evaluate its effectiveness, the proposed model is compared against AlexNet, a widely recognized face recognition architecture. Key performance indicators include emotion identification accuracy and the time taken to extract features from the designated facial regions.

As outlined in [9], the central concept of the automated attendance system is to utilize facial recognition technology in a more accessible and user-friendly manner than conventional biometric approaches. Our study evaluates the face detection and recognition capabilities of three techniques: Haar Cascade Classifiers (Viola-Jones), Histogram of Oriented Gradients (HOG), and Convolutional Neural Networks (CNNs). Among these, the Viola-Jones method demonstrated the highest accuracy. In practical applications, real-time attendance tracking leverages face detection via Viola-Jones and identification through CNN. A key benefit of the proposed approach lies in its ability to handle challenges such as facial misalignment, distinct facial features, and suboptimal lighting conditions.

As demonstrated in [10], an attendance system based on video input utilizes real-time facial recognition technology to monitor and verify user presence. This system supports simultaneous participation by multiple users and includes a face-liveness detection feature. It automatically captures facial data and stores it in a database alongside attendance records. For liveness detection, a module built using the TensorFlow framework incorporates the Ensemble of Regression Trees (ERT) method, which can effectively detect blinking to verify user presence. The overall attendance system is developed in Python, with the graphical user interface designed using the Qt framework.

III.METHODOLOGY

The system design incorporates optimization methods for efficient face detection management, focusing on public-facing applications. Initially, the system captures an employee's face using a camera and saves the image to a storage device. The CNN framework powers the training module, which utilizes feature extraction and selection algorithms to store individual features in a training database. This module then identifies faces within input images and records the corresponding frames for object recognition. The test feature maps the entire training dataset and computes similarity weights for each entry, which are used to accurately identify the staff based on the derived weight system. The system automatically adjusts its behavior based on the identified staff's information, as specified by the ID system. CNN is utilized for both training and testing tasks within the system.

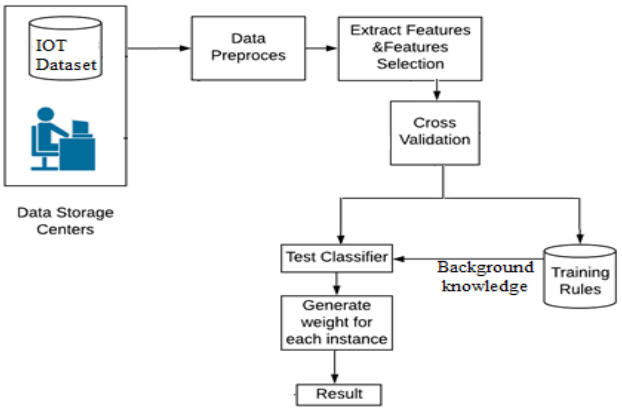

Figure 1: System Architecture

List of Modules

Module 1: Data Collection: We captured several staff images using a camera and saved them on a hard disk**.**

Module 2: Data Training: We gather both synthetic and authentic data using staff images and conduct training over time with comprehensive categorization**.**

Module 3: Testing with deep learning: We established a weight system with a deep learning classifier. The current staff facial and behavioral recognition system autonomously refreshes the database for the corresponding personnel.

Module 4: Analysis: We evaluate the accuracy of the proposed system and compare its performance with other existing systems.

Algorithm Details

Input:

Test dataset: A set of test instances, represented as TestDB-Lits[].

Training dataset: A set of data constructed during the training phase, represented as TrainDB-Lits[].

Threshold (Th): A predefined threshold value for similarity scoring.

Output: class label and weight

Step 1: For each testing record, the equation is provided below.

$$testFeature(k) = \sum_{m=1}^{n} (. \ featureSet[A[i] \ldots\ldots A[n] \leftarrow \text{TestDBLits})$$

Step 2: Generate a feature vector via the function provided below.

Extracted_FeatureSetx [t……n] = $?_{x=1}^{n}(t) \leftarrow testFeature$ (k)

Extracted_FeatureSetx[t] stores the features extracted for each instance in the testing dataset.

Step 3: For each training instance, use the following function.

$$trainFeature(l) = \sum_{m=1}^{n} (. \ featureSet[A[i] \ldots\ldots A[n] \leftarrow \text{TrainDBList})$$

Step 4: Create a new feature vector by training the features using the following function.

Extract_FeatureSet_Y[t……n] = $?_{x=1}^{n}(t) \leftarrow TrainFeature$ (l)

Extract_FeatureSet_Y[t] holds the features extracted for each instance in the training dataset.

Step 5: Subsequently, assess each test record against the whole training dataset.

$$weight = calcSim \ (\text{FeatureSetx} \ || \ \overset{n}{?}_{i=1} \ \text{FeatureSety[y]})$$

Step 6: Return label and Weight

IV.RESULT

The effectiveness of this system can only be evaluated by comparing it with other systems designed to address similar end-user challenges.

We have used a 1.2 GHz 64-bit quad-core ARM Cortex-A53 CPU with 32GB random access memory for execution. We have experimentally investigated proposed systems, including 5G networks, using the RESENT (32, 50, 101, and 152) version. Key factors taken into account for assessing the efficiency of the proposed systems include execution time (encompassing data processing, uploading, and downloading), memory usage, network overhead, and energy consumption. The microSD card slot (used for OS and data storage) typically runs Raspberry Pi OS (formerly Raspbian) or other Linux-based operating systems.

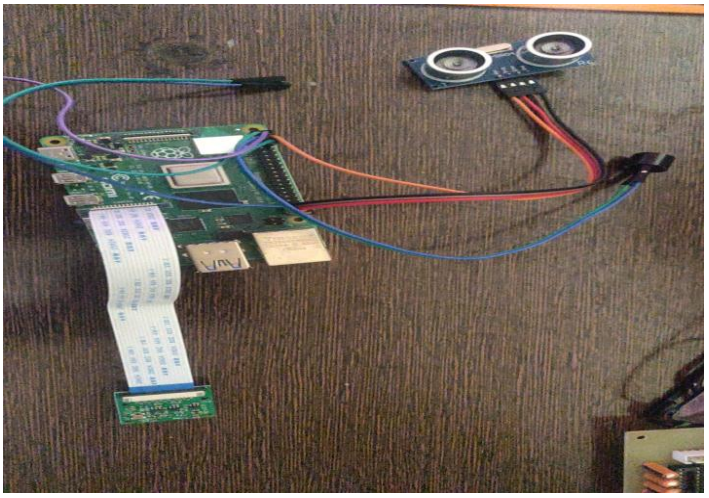

Figure 2: Hardware Connection

IV.CONCLUSION

The creation of a generic artificial intelligence agent using IoT and deep learning signifies significant progress in the development of versatile, cross-domain intelligent systems. This agent provides a flexible, scalable solution by merging IoT for data collection with deep learning for advanced analytics and decision-making, applicable across several domains, including smart cities, agriculture, healthcare, and industrial automation.

V.REFERENCES

[1] Mukhopadhyay, Moutan, et al. "Facial emotion detection to assess Learner's State of mind in an online learning system." Proceedings of the 2020 5th international conference on intelligent information technology. 2020.

[2] Ali, Md Forhad, Mehenag Khatun, and Nakib Aman Turzo. "Facial emotion detection using neural network." the international journal of scientific and engineering research (2020).

[3] Zamkah, Abdulaziz, et al. "Identification of suitable biomarkers for stress and emotion detection for future

personal affective wearable sensors." Biosensors 10.4 (2020): 40.

[4] Farkhod, Akhmedov, et al. "Development of Real-Time Landmark-Based Emotion Recognition CNN for Masked Faces." Sensors 22.22 (2022): 8704.

[5] Khaireddin, Yousif, and Zhuofa Chen. "Facial emotion recognition: State of the art performance on FER2013." arXiv preprint arXiv:2105.03588 (2021).

[6] Zhang, Hongli, Alireza Jolfaei, and Mamoun Alazab. "A face emotion recognition method using convolutional neural network and image edge computing." IEEE Access 7 (2019): 159081-159089.

[7] Saeki, Mao, et al. ”Analysis of Multimodal Features for Speaking Proficiency Scoring in an Interview Dialogue.” 2021 IEEE Spoken Language Technology Workshop (SLT). IEEE, 2021.

[8] Shin, Dong Hoon, Kyungyong Chung, and Roy C. Park. ”Detection of emotion using multi-block deep learning in a self-management interview app.” Applied Sciences 9.22 (2019): 4830.

[9] Patil, Payal, and S. Shinde. "Comparative analysis of facial recognition models using video for real time attendance monitoring system." 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA). IEEE, 2020.

[10] Huang, Shizhen, and Haonan Luo. "Attendance System Based on Dynamic Face Recognition." 2020 International Conference on Communications, Information System and Computer Engineering (CISCE). IEEE, 2020.