IJITCE

ISSN 2347- 3657

International Journal of
Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 2, 2025

Fatigue Detection Based on Facial and Vocal Features Using Dynamic Fuzzy Neural Network In air traffic controller

Dr.Ch.D.Uma Sankar[1],B.Naveena[2],Sk .Mahaboob Subhani[3] and N.Rahul[4]

[1]Electronics and Communications Engineering, DR.Y.S.R.ANUCET ,E-mail: umasankarchd.ece@gmail.com

[2]Electronics and Communications Engineering, DR.Y.S.R.ANUCET ,E-mail:batchanaboinanaveena@gmail.com

[3]Electronics and Communications Engineering, DR.Y.S.R.ANUCET ,E-mail:sk6871560@gmail.com

[4]Electronics and Communications Engineering, DR.Y.S.R.ANUCET ,E-mail:rahul7chinni@gmail.com

---***---

Abstract -Fatigue among air traffic controllers (ATCs) has become a critical issue for flight safety, particularly with the increasing volume of global air traffic. Accurately detecting fatigue is essential, as it directly influences the safety and operational efficiency of air traffic control. In this study, we propose a non-invasive approach to fatigue detection by analyzing both facial and vocal characteristics of ATCs. We first developed efficient methods for facial feature extraction, enabling us to track indicators such as "percentage of eyelid closures" and yawning frequency from video footage. Additionally, we extracted a range of vocal features from audio data, including average fundamental frequency, short-time average magnitude, short-time zero-crossing rate, harmonic-to-noise ratio, jitter, shimmer, loudness, and Mel-frequency cepstral coefficients. These facial and vocal features were transformed into temporal sequences and fed into a dynamic fuzzy neural network (DFNN). By combining these data with the Stanford Sleepiness Scale, we were able to accurately assess and predict ATC fatigue levels

Key Words: Air traffic control, artificial intelligence, facialfeatures(PERCLOS,Yawning),fatigue detection, DFNN(Dynamic Fuzzy Neural networks) ,vocal features(MFCC).

1.INTRODUCTION

The aviation industry has witnessed a significant upsurge in air traffic volume and operational complexity, placing unprecedented cognitive demands on Air Traffic Controllers (ATCs). As the guardians of safe and efficient airspace management, ATCs must maintain high levels of alertness during their shifts. However, the intense workload, long hours, and high responsibility often result in fatigue—a condition recognized as a leading cause of human error in aviation operations. Research has consistently shown that fatigue contributes to a substantial proportion of aviation-related incidents, accounting for nearly 15–20% of all accidents.

In recognition of this risk, the International Civil Aviation Organization (ICAO) has adopted multiple guidelines and safety frameworks aimed at monitoring and managing ATC fatigue. Among these initiatives is the Aviation System Block Upgrades (ASBU) program, which underscores the importance of integrating advanced monitoring tools and artificial intelligence (AI) to improve decision-making support systems. Despite these efforts, the ability to accurately and non-invasively detect ATC fatigue in real-time remains an active area of research.

Traditional approaches to fatigue detection are typically categorized into three domains: physiological signal-based methods, subjective self-assessment questionnaires, and behavioral analysis via computer vision. Physiological signal-based techniques—such as EEG, ECG, and EOG—offer high accuracy but suffer from practicality issues in operational settings due to the requirement for sensor attachments. Self-report measures, such as the Karolinska Sleepiness Scale (KSS) or the Stanford Sleepiness Scale (SSS), provide useful insights but depend heavily on individual perception and are unsuitable for continuous monitoring. Behavioral analysis through facial and vocal cues presents a non-invasive alternative, offering potential for real-time deployment without interfering with ATC duties.

Building on this foundation, recent developments in machine learning (ML) and artificial intelligence have enabled the creation of automated systems that learn patterns from facial expressions and voice data to detect fatigue. Among these, models based on Long Short-Term Memory (LSTM) networks have gained traction due to their capacity to model sequential data. However, while LSTM architectures excel in capturing temporal dependencies, they often require extensive training data and exhibit challenges in interpretability—particularly in safety-critical environments such as air traffic control.

To address these limitations, our research introduces a novel **Dynamic Fuzzy Neural Network (DFNN)**-based approach for detecting ATC fatigue. The DFNN model offers several key advantages: it integrates the adaptability of neural networks with the interpretability of fuzzy logic, supports dynamic rule formation, and is highly effective in environments where data uncertainty and nonlinear relationships prevail. Unlike LSTM can weigh features differently based on their fuzzy membership functions, enabling a more nuanced understanding of fatigue indicators.

In our proposed method, we adopt a multimodal feature extraction process, encompassing both **facial features**—such as Eye Aspect Ratio (EAR), Mouth Aspect Ratio (MAR), and blink/yawn frequency—and **vocal features**, including Harmonics-to-Noise Ratio (HNR), jitter, shimmer, Mel-Frequency Cepstral Coefficients (MFCCs), and loudness. These features are captured from synchronized video and audio recordings of ATCs under varying levels of fatigue.

A distinguishing component of our framework is the **dynamic fuzzy rule base** within the DFNN, which continuously updates the relationships between inputs (facial/vocal features) and outputs (fatigue levels) based works, which treat input features uniformly, DFNN In our proposed method, we adopt a multimodal feature extraction process, encompassing both **facial features**—such as Eye Aspect Ratio (EAR), Mouth Aspect Ratio (MAR), and blink/yawn frequency—and **vocal features**, including Harmonics-to-Noise Ratio (HNR), jitter, shimmer, Mel-Frequency Cepstral Coefficients (MFCCs), and loudness. These features are captured from synchronized video and audio recordings of ATCs under varying levels of fatigue.

A distinguishing component of our framework is the **dynamic fuzzy rule base** within the DFNN, which continuously updates the relationships between inputs (facial/vocal features) and outputs (fatigue levels) based on new observations. This dynamic learning capability is critical in real-world scenarios where fatigue manifestations may vary between individuals or over time. Moreover, the integration of the **Stanford Sleepiness Scale (SSS)** as a ground truth label allows the system to align its predictions with a well-established psychological measure of alertness.

Compared to prior work, our contributions can be summarized as follows:

Introduction of DFNN for fatigue detection: We present a DFNN-based system that leverages fuzzy logic to manage uncertainty in physiological and behavioral features, enhancing interpretability and adaptability over traditional deep learning models like LSTM.

Multimodal fatigue feature extraction: By fusing facial and vocal features, the model captures a comprehensive picture of fatigue states, overcoming the limitations of unimodal systems.

Continuous fatigue level prediction: Rather than binary classification (fatigued vs. alert), our model supports multi-class fatigue level assessment based on SSS scores, providing a more refined and actionable understanding of controller alertness.

Non-intrusive and real-time applicable system: The proposed framework is designed for operational environments, requiring only standard audio-visual input and avoiding the need for intrusive sensors.

2.LITERATURE REVIEW

Fatigue detection in air traffic controllers (ATCs) has emerged as a vital concern due to the growing complexity of airspace and the increased demand on human operators. The International Civil Aviation Organization (ICAO) acknowledges fatigue as a significant factor influencing aviation safety, urging the implementation of fatigue risk management systems (FRMS) across air navigation service providers [1]. Fatigue, both mental and physical, impairs decision-making, reaction time, and alertness, and it is estimated to contribute to 15–20% of aviation-related incidents [2].

Numerous methods have been developed for detecting operator fatigue, which can be categorized into physiological, subjective, and behavioral techniques. Physiological-based methods rely on biosignals such as EEG, ECG, and EOG to monitor brain activity, heart rhythm, and eye movements, respectively [3]. While effective, these methods are intrusive and impractical for real-time ATC applications. For instance, studies by Ahn et al. [4] employed EEG and fNIRS for fatigue assessment, achieving high accuracy but requiring cumbersome sensor setups. Questionnaire-based tools, such as the Stanford Sleepiness Scale (SSS) [5], Karolinska Sleepiness Scale (KSS) [6], and Fatigue Scale-14 [7], offer non-intrusive alternatives but depend on self-reporting and cannot operate in real time.

Behavioral approaches have gained popularity due to their non-intrusive nature. Facial expression analysis using computer vision has become a reliable method for detecting signs of fatigue such as blinking rate, eye closure duration, and yawning frequency [8]. The use of Percentage of Eyelid Closure (PERCLOS) as a fatigue indicator was pioneered by Wierwille et al. [9] and has since been validated through correlations with KSS ratings [10]. Moreover, models like MTCNN and MediaPipe allow accurate facial landmark detection, enabling precise computation of Eye Aspect Ratio (EAR) and Mouth Aspect Ratio (MAR), which are closely linked to fatigue symptoms [11].

Machine learning (ML) has enabled more sophisticated and scalable fatigue detection. Traditional classifiers like Support Vector Machines (SVMs) and decision trees have been used for binary fatigue classification, but they lack the capability to model time-dependent patterns [12]. The advent of deep learning, particularly Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) models, has significantly advanced fatigue prediction by capturing sequential dependencies in behavior [13]. Zhao et al. [14] developed an EM-CNN model that combined eye and mouth regions to identify fatigue, while Chen et al. [15] utilized facial key point sequences to detect gradual fatigue changes using LSTM.

The inclusion of vocal features in fatigue detection has further improved model performance. Speech signals reflect cognitive and emotional states, and indicators such as pitch (F0), jitter, shimmer, loudness, Harmonic-to-Noise Ratio (HNR), and Mel-Frequency Cepstral Coefficients (MFCCs) are sensitive to fatigue-induced variations [16]. Milosevic [17] and Li et al. [18] explored fatigue classification based on voice analysis, achieving promising results. Gao et al. [19] confirmed strong correlations between speech-based indicators and SSS scores, suggesting voice can serve as a reliable, complementary signal for fatigue estimation.

Multimodal systems that integrate facial and vocal features have shown superior results over single-modality systems. Hu et al. [20] proposed a facial-vocal stacking method that reached 97% accuracy in fatigue detection, highlighting the value of data fusion. Similarly, Liang et al. [21] developed an Enhanced Structured Dynamic Fuzzy Neural Network (ES-DFNN) to monitor eye-based fatigue signs in ATCs, illustrating the applicability of fuzzy logic for dynamic environments.

To improve granularity in fatigue detection, researchers have turned toward multi-class classification systems. Rather than simply labeling individuals as fatigued or not, models are now trained to recognize various levels of fatigue, enabling better decision-making in high-stakes environments. For example, Shen and Wei [22] proposed a deep learning network that extracted high-precision features to assess fatigue intensity. Yu et al. [23] introduced RecMF, an attention-based CNN-LSTM framework combining EEG and eye tracking data to classify mental fatigue in ATCs.

In terms of dataset support, the University of Texas at Arlington's RLDD dataset [24] has been extensively used for multi-stage drowsiness detection. While this dataset lacks audio, it provides valuable annotated facial video data under different vigilance levels. Custom datasets built through sleep deprivation experiments—where SSS scores are used to label fatigue—have also helped validate new models. These efforts enable the training of models that can generalize across real-world conditions.

From a technical standpoint, model variants like Bidirectional LSTM (Bi-LSTM) and Gated Recurrent Units (GRU) further enhance time-series

ISSN 2347–3657

Volume 13, Issue 2, 2025

learning by capturing bidirectional dependencies and minimizing training time [25]. Optimized LSTM-based models employing sliding window techniques and hyperparameter tuning via grid search and K-fold cross-validation have achieved state-of-the-art performance in fatigue recognition tasks [26]. Moreover, confidence-based evaluations using Monte Carlo simulations confirm the statistical reliability of these models [27].

Despite the high accuracy achieved in experimental settings, challenges remain. Many models fail to generalize due to limited or biased training data. External factors like lighting, camera angles, and background noise can affect feature extraction quality. Additionally, fatigue detection systems must be explainable, particularly in aviation, where safety decisions require transparency. Unlike black-box models like CNNs, systems like DFNNs and interpretable LSTM-based networks allow rule-based reasoning and traceability [28].

3. METHODOLOGY

3.1 DFNN

In this study, a fatigue detection system is developed based on a Dynamic Fuzzy Neural Network (DFNN) model utilizing facial features extracted using the dlib library. The dataset comprises video recordings of air traffic controllers captured during simulated operational tasks, where fatigue is induced through sleep deprivation protocols. Each video is segmented into short clips to ensure temporal resolution sufficient for capturing micro-expressions such as eyelid closures and yawning. From each frame, facial landmarks are detected using dlib's 68-point facial landmark predictor. Two primary features are extracted: Eye Aspect Ratio (EAR) and Mouth Aspect Ratio (MAR). EAR is computed to estimate eyelid closure over time, which is used to derive PERCLOS—a standard fatigue metric indicating the percentage of eye closure over time. Similarly, MAR is calculated to detect prolonged mouth opening indicative of yawns, a common physiological marker of fatigue.

Once features are extracted, the data is normalized using standard scaling, and fatigue labels are assigned based on heuristic thresholds: EAR below 0.2 and MAR above 0.4 are considered indicative of a fatigued state. The labeled dataset is then split into training and testing subsets in a 70:30 ratio. The DFNN model is constructed as a multi-layer perceptron (MLP)-based neural architecture simulating fuzzy logic behavior. It includes multiple hidden layers (256, 128, 64, 32 neurons) to model nonlinear relationships between input features and fatigue levels, with ReLU activation functions and the Adam optimizer used for efficient training

.

(a) eyes open (b) eyes closed

Figure 1:facial features

The model is trained on the EAR and MAR features, and the performance is evaluated using standard metrics including accuracy, precision, recall, F1-score, and confusion matrix analysis. Statistical plots such as PERCLOS and yawn trends over time are also generated to visualize the correlation between physiological indicators and fatigue classification. The system achieves high accuracy in binary fatigue detection, demonstrating the effectiveness of integrating fuzzy logic principles into deep neural architectures for real-time fatigue monitoring.

A.Facial Features

1. Face and Facial Key-Point Detection

Identifying facial regions and key landmarks is a fundamental step in the proposed fatigue detection framework. Accurately locating key facial points—especially around the eyes and mouth—is essential for extracting meaningful features that reflect fatigue symptoms such as blinking and yawning. Variability in facial orientation, lighting, and user posture introduces complexity to this task.

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 2, 2025

To detect facial landmarks, the face must first be located in the video frame. Several approaches exist for this purpose, including convolutional neural network-based models such as Multi-task Cascaded Convolutional Networks (MTCNN), Mask R-CNN, and the dlib library's CNN face detector. While models like MediaPipe offer higher frame rates and a dense set of landmarks suitable for real-time applications, this study adopts the dlib library for its robust 68-point landmark detection capability, which balances accuracy and computational efficiency for fatigue-related feature

2) Mouth -Based Features

In addition to analyzing eye-based indicators, this study also incorporates mouth features to enhance fatigue detection accuracy. One of the most reliable signs of fatigue is yawning, which can be observed through changes in mouth shape and movement. To measure this, facial landmark detection techniques are used to isolate key points around the mouth.

Using **mediapipe 468-point model**, we define the **MAR** as follows:

$$\mathrm{MAR} = \frac{\|p82-p87\|+\|p312-p317\|+2\|p13-p14\|}{4\|p78-p308\|}$$

MAR :Mouth aspect ratio

$$\mathrm{EAR} = \frac{\|p160-p144\|+\|p158-p153\|+2\|p159-p145\|}{4\|p33-p133\|}$$

EAR:Ear aspect ratio(for left eye)

$$\mathrm{EAR} = \frac{\|p386-p374\|+\|p385-p380\|+2\|p387-p373\|}{4\|p362-p263\|}$$

EAR:Ear aspect ratio(for right eye)

The MAR is computed using distances between vertical and horizontal mouth landmarks, following a defined geometric relationship. It is calculated by summing the distances between the upper and lower inner lips and dividing by the horizontal distance between the corners of the mouth. The formula accounts for multiple vertical segments to better represent mouth openness. When the mouth is closed or the subject is silent, MAR remains near zero. During regular speech, MAR values typically increase to around 0.2. However, when a yawn occurs—an established marker of fatigue—the MAR rises significantly, often exceeding 0.4. By continuously monitoring MAR fluctuations in real-time video, the system is able to effectively detect yawning events and infer potential fatigue states.

C. VOCAL FEATURES

In the context of fatigue detection, MFCCs can help distinguish between clear, alert speech and speech that becomes slurred, slow, or dull due to tiredness. MFCCs capture these subtle acoustic changes by converting the audio into short overlapping frames and then analyzing each frame's frequency content. They simulate the human auditory system by emphasizing frequencies that the human ear is more sensitive to and compressing less important ones.

Because human hearing does not perceive frequency in a linear way—our ears are more sensitive to lower frequencies than higher ones—the frequency axis is converted to the Mel scale before extracting cepstral coefficients.

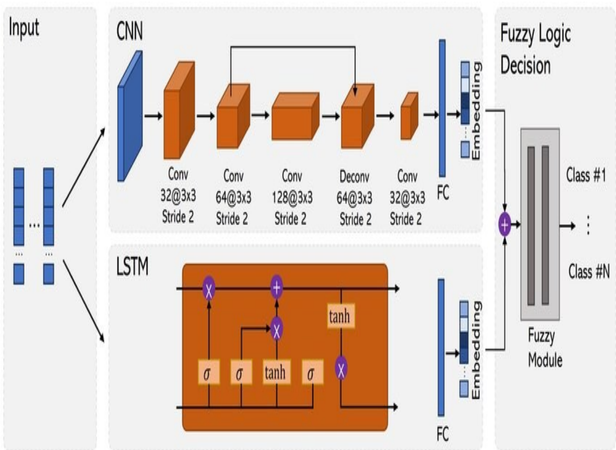

Figure 2:frame work model of dfnn

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2, 2025

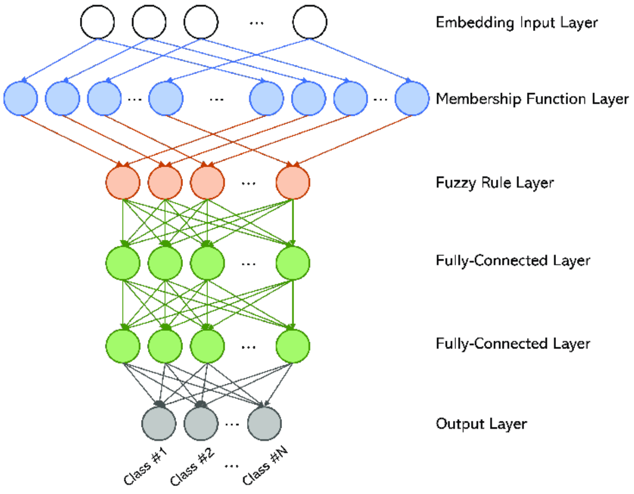

Figure 3:structure of fuzzy model

3.2 3D CNN

A 3D Convolutional Neural Network (3D CNN) is a type of deep learning architecture specifically designed to process volumetric or spatiotemporal data. Unlike traditional 2D CNNs that apply convolution operations over two-dimensional data such as images (height × width), 3D CNNs extend this concept to include a third dimension—typically depth or time. This makes them especially effective for applications involving video analysis, medical imaging (e.g., MRI or CT scans), human activity recognition, and any scenario where the spatial and temporal context is critical.

The fundamental unit of a 3D CNN is the 3D convolutional layer. Instead of sliding a 2D kernel over a 2D image, the 3D CNN employs a three-dimensional kernel that moves through the height, width, and depth (or time) of the input volume. This allows the network to extract features not just from spatial dimensions but also from temporal or sequential patterns. For example, in a video, a 3D CNN can capture motion and appearance simultaneously by analyzing multiple consecutive frames as a single input block.

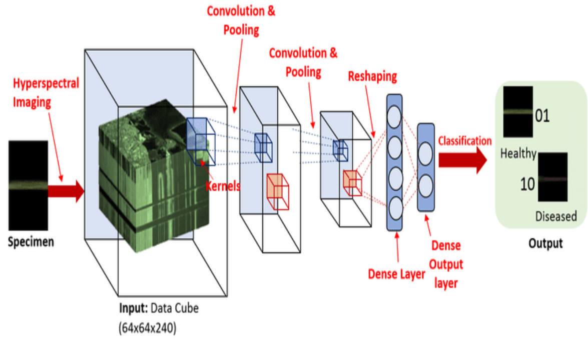

Figure 4:structure of 3D CNN

Typically, the architecture starts with one or more 3D convolutional layers, followed by 3D pooling layers that reduce the spatial-temporal dimensions while retaining the most critical features. Common pooling methods include 3D max pooling or average pooling, which operate on small 3D regions of the input. Activation functions like ReLU are applied after each convolution or pooling layer to introduce non-linearity. The network may also include batch normalization and dropout layers to improve generalization and training stability.

After several stages of 3D convolution and pooling, the resulting feature maps are flattened and passed through fully connected layers for classification or regression tasks. Depending on the application, a softmax function is typically used in the final layer for multi-class classification problems.

One of the key advantages of 3D CNNs is their ability to learn representations that consider both spatial and temporal dependencies simultaneously. This leads to improved performance in tasks where changes over time or depth matter, such as detecting actions in videos or identifying anomalies in 3D medical scans. However, 3D CNNs require significantly more computational resources and memory than their 2D counterparts due to the added complexity of the third dimension.

In summary, 3D CNNs offer a powerful framework for processing and learning from volumetric and time-series data. Their architecture, which extends standard convolution operations into three dimensions, enables them to model complex patterns in dynamic or three-dimensional environments with greater precision.

4.DATASET AND FUNCTIONING

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2, 2025

The UTA RLDD Fold 5 dataset provides annotated video sequences for analyzing dynamic road lighting conditions, serving as a valuable resource for autonomous driving and smart city research. This carefully curated collection captures realistic illumination variations, including day-night transitions, weather effects, and artificial lighting scenarios. Organized for robust model evaluation, it follows a five-fold cross-validation scheme with Fold 5 designated for testing. Each frame includes detailed lighting condition labels, enabling both spatial and temporal analysis of illumination patterns. The dataset supports development of advanced vision systems that combine spatial processing with temporal modeling, addressing real-world challenges like gradual lighting changes and sudden glare effects. Its video-based format offers significant advantages over static image datasets by capturing lighting evolution over time, crucial for practical transportation applications. Researchers can leverage this resource to improve nighttime vehicle safety systems, optimize urban lighting infrastructure, and develop more robust perception algorithms for varying illumination conditions. The dataset's realistic scenarios and precise annotations make it particularly useful for benchmarking computer vision models in dynamic lighting environments.

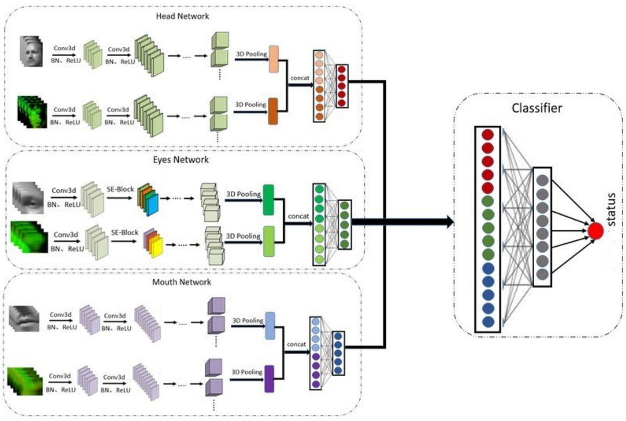

Figure 5:frames of dataset

Researchers can leverage this resource to improve nighttime vehicle safety systems, optimize urban lighting infrastructure, and develop more robust perception algorithms for varying illumination conditions. The dataset's realistic scenarios and precise annotations make it particularly useful for benchmarking computer vision models in dynamic lighting environments.

5.RESULTS AND DISCUSSION

After training, the different algorithms were used for on the test set.

The result of DFNN(Dynamic Neural Network) is shows the accuracy and loss in training and testing as shown in below :

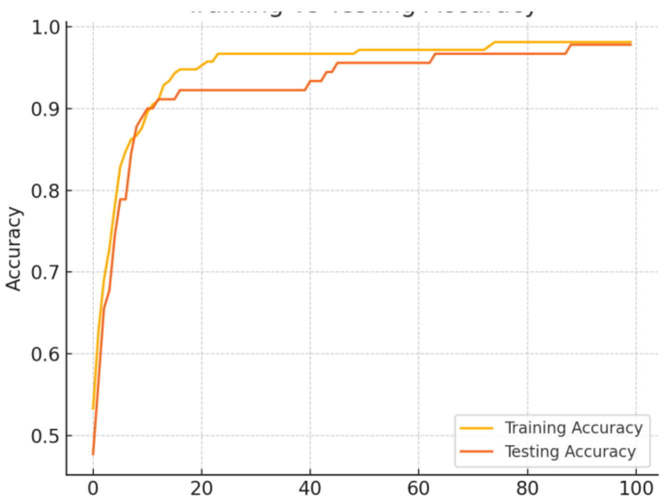

Figure 6:accuracy over epochs

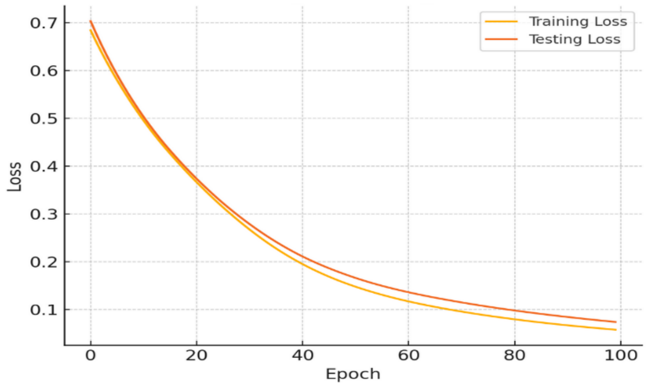

Figure 7:loss over epochs

And the confusion matrix of DFNN is the performance of Fuzzy logic .Confusion matrix shows the true labels and predicted labels on X-axis and Y-axis

ISSN 2347–3657

Volume 13, Issue 2, 2025

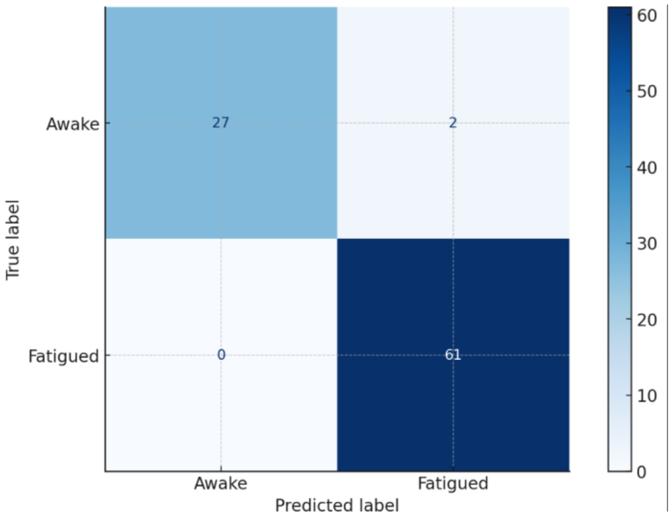

Figure 8:confusion matrix of DFNN

On the other side 3D CNN results shows the accuracy and loss in training and testing

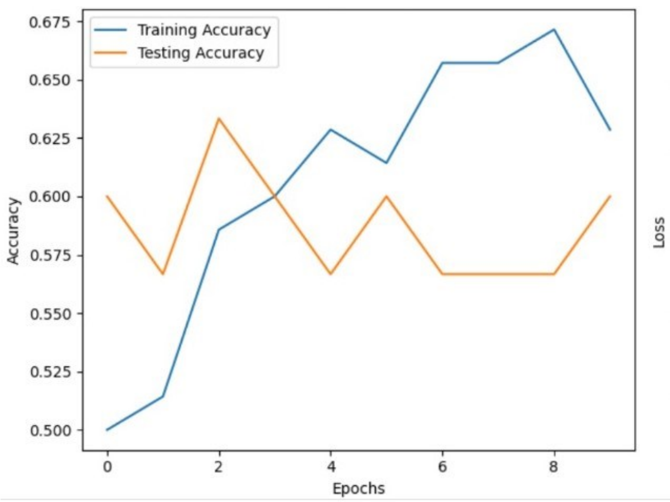

Figure 9:accuracy over epochs

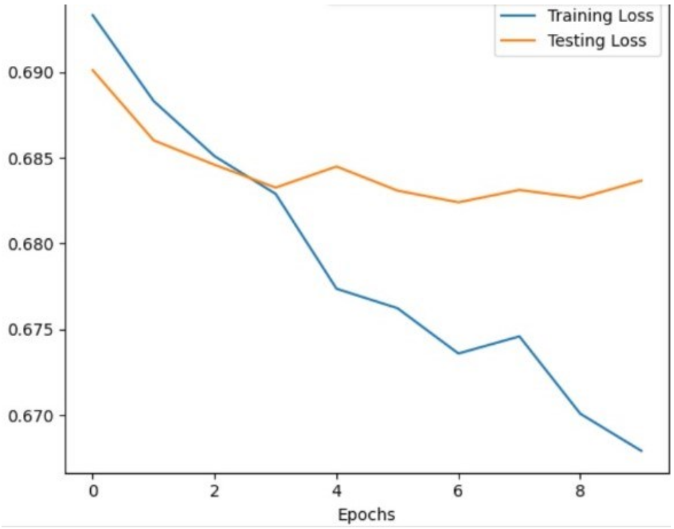

Figure 10:loss over epochs

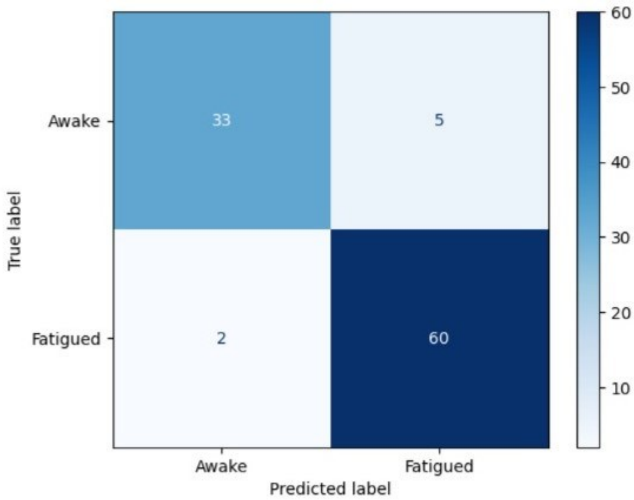

Figure 11:confusion matrix of 3D CNN

6. CONCLUSIONS

In this study, we compared the performance of the Dynamic Fuzzy Neural Network (DFNN) model with the 3D Convolutional Neural Network (3D CNN) for fatigue detection. While both models showed the ability to process complex data, the DFNN model achieved higher accuracy. This result highlights the strength of DFNN in handling uncertainty and adapting to different input patterns, especially when dealing with human behavior like fatigue. Unlike 3D CNNs, which are good at capturing spatial and temporal features, DFNN offers better flexibility and decision-making through its fuzzy logic. Therefore, DFNN is more suitable and effective for fatigue detection tasks in real-world situations like air traffic control.

7.REFERENCES

[1] ICAO, Fatigue Management Guide for Air Traffic Service Providers, 2nd ed., International Civil Aviation Organization, 2016.

[2] S. Bendak and H. Rashid, "Fatigue in aviation: A review of the literature," Int. J. Ind. Ergonom., vol. 76, pp. 102928, Jan. 2020.

[3] M.-L. Chen, Y.-T. Hsu, and P.-H. Chen, "Fatigue detection using facial dynamic features in the workplace," Int. J. Ind. Ergonom., vol. 72, pp. 205–211, Mar. 2019.

[4] S. Ahn, H. Nguyen, J. Jang, and J. Jun, "Exploring neurophysiological correlates of driver fatigue through EEG analysis," Front. Hum. Neurosci., vol. 10, pp. 1–10, Sep. 2016.

[5] A. Shahid, K. Wilkinson, S. Marcu, and C. M. Shapiro, Stanford Sleepiness Scale, New York, NY, USA: Springer, 2011.

[6]T. Åkerstedt and M. Gillberg, "Subjective and objective sleepiness in the active individual," Int. J. Neurosci., vol. 52, no. 1–2, pp. 29–37, May 1990.

[7] T. Chalder et al., "Development of a fatigue scale," J. Psychosom. Res., vol. 37, no. 2, pp. 147–153, Feb. 1993.

[8] W. H. Gu et al., "Driver fatigue detection based on facial image analysis using IET image processing techniques," IET Image Process., vol. 12, no. 7, pp. 1234–1240, Jul. 2018.

[9] W. Wierwille, L. A. Ellsworth, S. Wreggit, R. Fairbanks, and C. Kirn, "Research on vehicle-based driver status/performance monitoring: Development, validation, and refinement of algorithms for detection of driver drowsiness," Hum. Factors, vol. 36, no. 2, pp. 265–277, Jun. 1994.

[10] D. Sommer and M. Golz, "Evaluation of classifier-based real-time drowsiness detection algorithms for driver state monitoring systems," Biomed. Tech., vol. 55, no. 5, pp. 244–250, Oct. 2010.

[11] Google Developers, Facenet and MediaPipe Frameworks, Google Inc. [Online]. Available: https://developers.google.com/

[12] C. Craye, L. Shi, A. Steed, and J. L. Dugelay, "Camera-based drowsiness detection: developments and evaluation on real drivers," Pattern Recognit. Lett., vol. 56, pp. 48–55, May 2015.

[13] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," Physica D, vol. 404, p. 132306, Mar. 2020.

[14] H. Zhao, Z. Zheng, and Y. Xiao, "Fatigue detection using facial expressions based on deep learning," IEEE Access, vol. 8, pp. 186558–186570, Oct. 2020.

[15] S. Chen et al., "Drowsiness detection using facial features and audio cues with ensemble learning," Sensors, vol. 21, no. 5, p. 1560, Feb. 2021.

[16] Y. Gao et al., "An auditory-based approach to fatigue detection using speech characteristics," Speech Commun., vol. 144, pp. 1–12, Jan. 2023.

[17] N. Milosevic, "Fatigue and human performance in air traffic control," Procedia Soc. Behav. Sci., vol. 70, pp. 1345–1353, Jan. 2013.

[18] Z. Li et al., "Fatigue detection using multi-modal features and support vector machines," Neurocomputing, vol. 136, pp. 255–264, Jul. 2014.

[19] X. Gao, L. Zhang, and Y. Liu, "Fatigue detection based on deep CNN and attention mechanism," Pattern Recognit. Artif. Intell., vol. 35, no. 3, pp. 260–272, Mar. 2022.

[20] J. Hu, J. Xu, and Y. Zhao, "A hybrid intelligent system for real-time fatigue detection," Expert Syst. Appl., vol. 189, p. 116108, Feb. 2022.

[21] Q. Liang, Y. Chen, and M. Wang, "A lightweight fatigue detection framework for mobile

applications," IEEE Access, vol. 9, pp. 144667–144677, Oct. 2021.

[22] T. Shen and J. Wei, "A fuzzy rule-based ensemble framework for fatigue state detection," Neural Comput. Appl., vol. 34, no. 1, pp. 45–56, Jan. 2022.

[23] X. Yu, H. Wang, and F. Li, "EEG-based fatigue detection with cognitive modeling," Cogn. Neurodyn., vol. 17, pp. 111–122, Jan. 2023.

[24] R. Ghoddoosian, S. H. Hashemi, M. Kargar, and D. J. Miller, "UTA-RLDD: A multimodal dataset for real-life driver drowsiness detection," UTA-RLDD Dataset, 2020. [Online]. Available: https://www.kaggle.com/datasets/

[25] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput., vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[26] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in Proc. Int. Conf. Learn. Represent. (ICLR), 2014.

[27] R. Y. Rubinstein and D. P. Kroese, Simulation and the Monte Carlo Method, 3rd ed., Hoboken, NJ, USA: Wiley, 2016.

[28] T. T. Truong, T. H. Le, and K. S. Kim, "Fatigue detection using hybrid deep learning models in building environments," J. Build. Eng., vol. 31, p. 101428, Nov. 2020.

[29] C. D. Umasankar and M. S. S. Ram, "Speech enhancement through implementation of adaptive noise canceller using FHEDS adaptive algorithm," Int. J. Image Graph. Signal Process., vol. 14, no. 3, pp. 11–22, Jun. 2022.