Contactless Privacy-Preserving Head Movement Recognition Using Deep Learning for Driver Fatigue Detection

Hira Hameed
James Watt School of Engineering
Glasgow, UK
2683961H@student.gla.ac.uk

Lubna
Telecommunication Engineering Dept. UET
Peshawar, Pakistan
Lubnaxafi@gmail.com

Muhammad Usman
James Watt School of Engineering
Glasgow, UK
Muhammad.Usman@glasgow.ac.uk

Hasan Abbas
James Watt School of Engineering
Glasgow, UK
hasan.abbas@glasgow.ac.uk

Ahsen Tahir
James Watt School of Engineering
Glasgow, UK
ahsen.tahir@glasgow.ac.uk

Kamran Arshad
Faculty of Engineering and IT, Ajman
United Arab Emirates
k.arshad@ajman.ac.ae

Khaled Assaleh
Faculty of Engineering and IT, Ajman
United Arab Emirates
k.assaleh@ajman.ac.ae

Ahmed Alkhayyat
Department of computer technical engineering
Najaf, Iraq
ahmedalkhayyat85@iunaja.edu.iq

Muhammad Ali Imran
James Watt School of Engineering
Glasgow, UK
Muhammad.Imran@glasgow.ac.uk

Qammer H. Abbasi
James Watt School of Engineering
Glasgow, UK
qammer.abbasi@glasgow.ac.uk

***Abstract*—Head movement holds significant importance in conveying body language, expressing specific gestures, and reflecting emotional and character aspects. The detection of head movement in smart or assistive driving applications can play an important role in preventing major accidents and potentially saving lives. Additionally, it aids in identifying driver fatigue, a significant contributor to deadly road accidents worldwide. However, most existing head movement detection systems rely on cameras, which raise privacy concerns, face challenges with lighting conditions, and require complex training with long video sequences. This novel privacy-preserving system utilizes UWB-radar technology and leverages Deep Learning (DL) techniques to address the mentioned issues. The system focuses on classifying the five most common head gestures: Head 45L (HL45), Head 45R (HR45), Head 90L (HL90), Head 90R (HR90), and Head Down (HD). By processing the recorded data as spectrograms and leveraging the advanced DL model VGG16, the proposed system accurately detects these head gestures, achieving a maximum classification accuracy of 84.00% across all classes. This study presents a proof of concept for an effective and privacy-conscious approach to head position classification.**

***Index Terms*—Deep learning, RF sensing, Head Movement, UWB radar, Assistive Driving**

I. Introduction

Head movement-detection is used in a growing number of applications, such as assistive technology, video conferencing, and virtual reality. This has led to more research into making real-time head movement detection and tracking systems that are robust and accurate. Since the past decade, smart driving assistance systems have been a significant area of research. Such technologies depend significantly on driver attentiveness detection. It detects the driver's state to minimize distractions while driving. The position of the driver's head may indicate his level of concentration. Tracking the movement of a driver's head is a commonly used method to study driver attentiveness, replacing the monitoring of eye movements that can be affected by lighting conditions. In the context of head position estimation, various techniques have been explored, and a survey of these methods can be found in the work by [1]. [2] proposes a method to detect a cyclist's intention to cross the road using posture estimation. The system analyzes the cyclist's body position and movements to determine if they plan to cross. Experimental results demonstrate the effectiveness of the approach, providing valuable insights for autonomous driving. [3] presents a driver fatigue detection system that uses facial features and a gated recurrent unit (GRU) model. By analyzing facial expressions, the system accurately detects driver fatigue. The GRU model classifies fatigue levels based on extracted facial features, improving road safety by preventing accidents caused by drowsy driving. In [4], a computational framework was proposed for robust face identification and posture estimation. The researchers employed a multi-primitive closed-loop face analysis approach using video-arrays, which yielded precise and reliable outcomes. This approach allowed for enhanced face recognition and the estimation of head posture in various conditions. In [5], facial symmetry and anthropometric measurements were employed to determine head orientation. By considering factors such as eye distances and camera focal length, the researchers calculated the Y-Z coordinates of the head. Additionally, head X-axis orientation was estimated using face anthropometry. The effectiveness of this method was evaluated using real photographs as test samples. For accurately calculating the 3D posture of a user's head in real-time, a novel tracking system based on a stochastic filtering framework was suggested in [6]. This system calculates a user's head posture for each picture frame using a 3D model that was automatically created after initialization. The estimation approach is specifically made to define a motion model's diffusion factor in an adaptive manner. This approach contributes to concurrently enhancing the performance of robust tracking against sudden head movements and accurate pose estimation while the user is gazing at a spot in a scene. [7] proposed an estimation of head-pose independent method using a Biased-Manifold-Embedding framework, however, the projection-matrix lacking the ability to compute new data-points is a major shortcoming of this method. [8] proposed a view-based head pose estimation using eigenspaces. A study [9], showed that drivers may involve in various activities while driving which might get more sophisticated than driving normally. The most reported driving activities from UK-drivers are listed in Table. I. Computer vision based systems are one of the most frequent

Non-Driving Activities	Occurrence
Read Activities	7.6%
Sleep Activities	7.2%
Call or Texting Activities	5.5%
Other work engagements	4.9%
Watch Tv or Movies	4.2%
Play Games	1.9%

TABLE I: Non-Driving Activities and its Occurrence Rate.

research approaches for detecting head movements [10]–[12] but unless numerous cameras are utilised, a frequent restriction of camera-based head tracking systems is that adequate facial characteristics must be visible, which restricts the detectable angle of head movement and, consequently, the ability to quantify driving activities. Such techniques investigated need powerful computing hardware and are highly susceptible to privacy aspects of users. In [13], radar-based head movement recognition is presented for driving scenario where four cases of driver's head movements were recorded namely (a) Front, (b) Shaking head up and down, (c) Shaking head side to side, and (d) Lowering the head. [14] developed head position monitoring and classification system using thin strain-sensing threads on the neck. A bluetooth module and impedance reading circuits communicate the strain data to a computer which is then classified using machine learning algorithm. This study introduces a novel approach for recognizing head movement gestures using micro-Doppler signatures captured by a radar sensor. The proposed system considers five different gestures: HL45, HR45, HL90, HR90, and HD. Experimental data is collected using an ultra-wideband (UWB) radar, specifically the XeThru X4M03 model. The collected data is transformed into spectrograms, and spatiotemporal features are extracted using the VGG16 model. The achieved classification accuracy is 84%. Detailed information about the experimental setup, data collection process, deep learning algorithms, and experimental results can be found in the subsequent sections. The rest of the paper is organized as: Section II provides an outline of the utilized approaches concerning the experimental configuration, hardware details, data acquisition, and deep learning algorithms. In Section III, evaluation criteria and result and discussion. Our work is concluded in Section IV with some future findings.

II. Methodology

Table 1 presents a block diagram illustrating the methodology employed in this investigation, which consist of two primary stages. Firstly, diverse datasets of head movements were collected, constructed, and annotated. Then, the VGG16 deep learning model was used to classify the head movements. The subsequent sections offer detailed explanations of each step in the proposed methodology.

A. Experimental setup and data collection

This study proposes a head movement recognition system based on a UWB radar sensor (Xethru X4M03) developed by Novelda. The radar system integrates antennas and a transmitter, enabling highly accurate distance measurements and capturing detailed movement information. The radar was positioned at a distance of 0.79 meters from the subject/target during data collection, as shown in Fig. 2. Each head movement activity demonstrated in Fig. 4 was executed for a duration of 4 seconds, and the RF signals were transmitted and received within the designated range for each activity. The collected data was then stored as spectrograms, where the x-axis represents time and the y-axis represents Doppler frequency (Hz), as shown in Fig. 3. Spectrograms provide valuable information about the dynamic movement of the head. To ensure an adequate number of data samples, participants were instructed to perform each head gesture multiple times. The data collection process involved one male and one female participant to enhance the realism and diversity of the dataset. In total, 250 data samples were gathered for the five distinct head gesture classes: HL45, HR45, HL90, HR90, and HD.

The distribution of the head gesture dataset is presented in Table. III. In each experiment, 125 data samples were collected from each participant, with 25 samples per class. Among the collected spectrograms, 100 were used for training and 25 for testing. The spectrograms of all activities are visualized in Fig. 1. Further details about the experimental setup and system parameters can be found in Table. II. The specifications of the experimental setup and system parameters are listed in Table II.

Parameter	Value
Equipment	XetruRadar (X4MO3)
Radar-Range	9 m
Distance	0.79 m
Radar-Frequency	7 GHz
Gain (Receiver) (dB)	ETSI (14.1)
Power (Transmitter)	6.3 dBm
Activity Duration	4 seconds
Sample collected per class	25

TABLE II: System parameters of the experimental setup.

Classes	Experimental Dataset			
	Subject (S1)	Subject (S2)	Combined	Total
HL45	25	25	50	100
HR45	25	25	50	100
HL90	25	25	50	100
HR90	25	25	50	100
HD	25	25	50	100
Total	125	125	250	**500**

TABLE III: An overview of the collected data, number of subjects, and performed activities.

B. Deep learning Algorithm

The architecture utilized in this model is VGG16, as mentioned in [15], which comprises a total of 16 layers. This architecture incorporates 3x3 filters with a stride of 1 and consists of approximately 138 million parameters. The padding and maximum pooling layers within this architecture are implemented using a 2x2 filter size with a stride of 2. The hierarchy of layers in this approach includes ReLU layers, convolutional layers, and max pool layers. ReLU combines fast learning with effective computing. Finally, it has three completely connected layers and an output softmax. The parameters we have used for evaluation are Learning Rate (L-Rate), Batch Size (B-Size), Learning Algorithm (L-Algo), Loss Function (L-Fntn), Maximum Epochs (Max-Epochs), Iteration per Epoch (IPE), and Elapsed time (E-time). The parameter settings of VGG16 are shown in Table. IV.

III. System Evaluation and Results

In this section, we provide a comprehensive overview of the system evaluation and discuss the classification results achieved using the pre-trained model.

DL Model	Parameters	Settings
Subject (S1)	L-Rate	0.0001
	B-Size	32
	L-Algo	Adam
	L-Fntn	Cross-entropy
	Max-Epochs	15
	IPE	30
	E-time for 0.79 M	00:40:03
Subject (S2)	L-Rate	0.0001
	B-Size	32
	L-Algo	Adam
	L-Fntn	Cross-entropy
	Max-Epochs	15
	IPE	30
	E-time for 0.79 M	00:32:14
Combined	L-Rate	0.0001
	B-Size	32
	L-Algo	Adam
	L-Fntn	Cross-entropy
	Max-Epochs	15
	IPE	60
	E-time for 0.79 M	01:22:08

TABLE IV: Parameter settings for the selected models

A. Performance Metrics

In this context, the outcomes refer to the results obtained from evaluating the performance of the VGG16 deep learning model in classifying five different head movements. The Equation (4), was used for calculated average test accuracy which measures how accurately the model predicts the correct head movement class. Additionally, the F1 Score is calculated to further evaluate the model's performance. The F1 Score is a metric that combines both precision and recall, providing a balanced measure of the model's accuracy. It is calculated using specific equations, referred to as Equation (3) and Equation (1), which consider the number of true positives, false positives, and false negatives.

$$Precision = \frac{\sum TP}{\sum TP + \sum FN} \tag{1}$$

$$Recall = \frac{\sum TP}{\sum TP + \sum FN} \tag{2}$$

$$F1 - Score = 2\frac{Precision \times Recall}{Precision + Recall} \tag{3}$$

$$Accuracy = \frac{\sum(TP + TN)}{\sum(TP + FP + TN + FN)} \tag{4}$$

B. Result and Discussion

Experiments were conducted utilising test and train-split methodology, with 80% of the data serving as training-data and the remaining 20% serving as testing-data. The VGG16 pre trained models have 15 epochs, Adamax as the optimizer, having a learning rate of 0.001.

The outcomes of the experiments are given in Fig. 5a, 5b, and 5c. In this case, Fig. 5a shows the VGG16 model of a female subject's confusion matrix for the classification of the considered classes. The figure shows that the majority of the classes are accurately identified, with the lowest classification

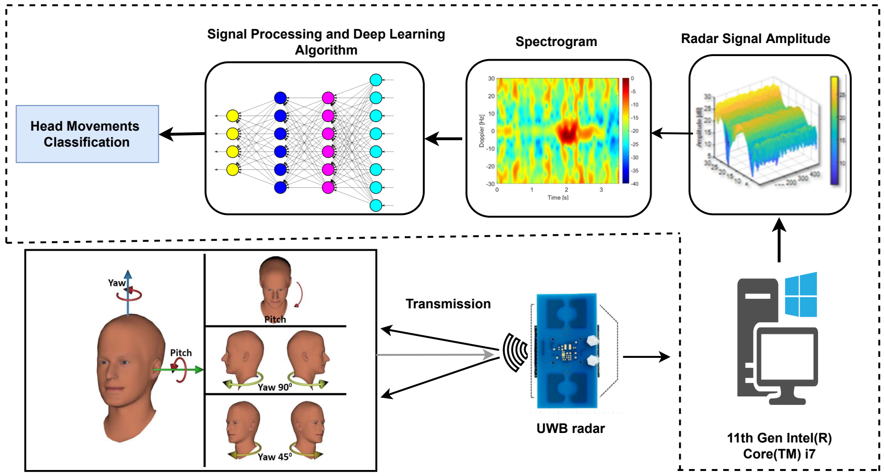


Fig. 1: Block diagram of the proposed architecture showcasing the UWB radar-based system, data gathering, and DL classification model for Head Movements.

Fig. 2: Experimental setup of the proposed Head movement recognition system.

Subject	VGG16 Model			
	Accuracy (%)	Precision	Recall	F1-Score
Subject (S1)	80.0	0.80	0.85	0.81
Subject (S2)	84.0	0.84	0.82	0.84
Combined	70.0	0.70	0.78	0.72

TABLE V: Metrics, including accuracy, recall, precision, and F1-score, were compared between subjects using VGG16.

accuracy of 60% for HL45 and HL90. This shows a 40% similarity between HR45 and HL45 and a 40% similarity between HL45 and HL90.

Furthermore, the Male (S2) confusion matrix is shown in Fig. 5b, in which the accuracy of the classification is around 100% for all classes excluding the HL45 and HR45, which seem to be similar to HR90 and HL45. Similarly, the confusion matrix of classifying the given head movement using combined dataset is presented in Table. 5c. With the exception of HL45, most of the classes are correctly classified, which exhibits resemblance with 30% HL90, 20% HR90, 10% HR45 and HR90 show similarities with 50% of HL90 and 10% HD. Table. V presents the total accuracy, precision, recall, and F1-score of the DL models under consideration. It is clear from the table that Subject (S2) performs better than Subject (S1) and combined dataset showing an overall test accuracy of 84.000

IV. Conclusion and Future Directions

Using the Xethru UWB Radar sensor and the DL algorithms, this research proposed a privacy-preserving head movement recognition system. The study focuses on five common head movements: HL45, HR45, HL90, HR90, and HD. Spectrograms capturing micro-Doppler features were recorded for each movement class and used to train VGG16 deep-learning models. The classification accuracy for most classes approached 100%, with the Male (S2) participant showing the highest performance. Overall, the system achieved an accuracy of 84.00% across all five classes. This preliminary work has resulted in a five-class head movement dataset, which will be further expanded and improved in future studies. Also we will take into consideration the key challenges in head-pose estimation systems which are real-time tracking as well as detection of non-driving activities, a few of these activities are mentioned in Table. I.

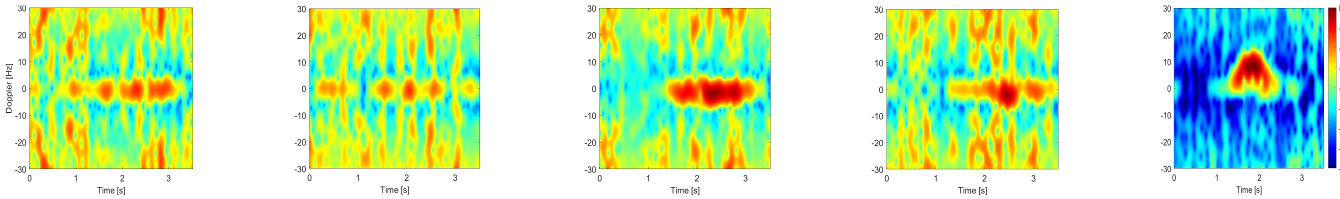
Acknowledgment

Partial support for this work was received from the Engineering and Physical Sciences Research Council (EPSRC) under the grant COG-MHEAR (ref. EP/T021063/1).

References

[1] A. Al-Rahayfeh and M. Faezipour, "Eye tracking and head movement detection: A state-of-art survey," *IEEE journal of translational engineering in health and medicine*, vol. 1, pp. 2100212–2100212, 2013.

[2] A. D. Abadi, Y. Gu, I. Goncharenko, and S. Kamijo, "Detection of cyclist's crossing intention based on posture estimation for autonomous driving," *IEEE Sensors Journal*, 2023.



(a) HL45 (b) HR45 (c) HL90 (d) HR90 (e) HD

Fig. 3: obtained a sample of the spectrum of (a) HL45, (b) HR45, (c) HL90, (d) HR90, (e) HD

(a) HL45 (b) HR45 (c) HL90 (d) HR90 (e) HD

Fig. 4: Gestures of the five considered Head Movement.

(a) Female(S1) (b) Male(S2) (c) Combined

Fig. 5: The Confusion Matrix of VGG16

[3] D. Li, X. Zhang, X. Liu, Z. Ma, and B. Zhang, "Driver fatigue detection based on comprehensive facial features and gated recurrent unit," *Journal of Real-Time Image Processing*, vol. 20, 02 2023.

[4] K. S. Huang and M. M. Trivedi, "Robust real-time detection, tracking, and pose estimation of faces in video streams," in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, vol. 3, pp. 965–968, IEEE, 2004.

[5] T. Horprasert, Y. Yacoob, and L. S. Davis, "An anthropometric shape model for estimating head orientation," in *3rd International Workshop on Visual Form, Capri, Italy*, 1997.

[6] K. Oka, Y. Sato, Y. Nakanishi, and H. Koike, "Head pose estimation system based on particle filtering with adaptive diffusion control.," in *MVA*, pp. 586–589, 2005.

[7] V. N. Balasubramanian, J. Ye, and S. Panchanathan, "Biased manifold embedding: A framework for person-independent head pose estimation," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–7, IEEE, 2007.

[8] S. Srinivasan and K. L. Boyer, "Head pose estimation using view based eigenspaces," in *2002 International Conference on Pattern Recognition*, vol. 4, pp. 302–305, IEEE, 2002.

[9] M. Sivak and B. Schoettle, "Motion sickness in self-driving vehicles," tech. rep., University of Michigan, Ann Arbor, Transportation Research Institute, 2015.

[10] J. Sherrah, S. Gong, and E.-J. Ong, "Face distributions in similarity space under varying head pose," *Image and Vision Computing*, vol. 19, no. 12, pp. 807–819, 2001.

[11] S. Niyogi and W. T. Freeman, "Example-based head tracking," in *Proceedings of the second international conference on automatic face and gesture recognition*, pp. 374–378, IEEE, 1996.

[12] E. Murphy-Chutorian and M. M. Trivedi, "Head pose estimation in computer vision: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 4, pp. 607–626, 2008.

[13] H. N. Nguyen, S. Lee, T.-T. Nguyen, and Y.-H. Kim, "One-shot learning-based driver's head movement identification using a millimetre-wave radar sensor," *IET Radar, Sonar & Navigation*, vol. 16, no. 5, pp. 825–836, 2022.

[14] Y. Jiang, A. Sadeqi, E. L. Miller, and S. Sonkusale, "Head motion classification using thread-based sensor and machine learning algorithm," *Scientific reports*, vol. 11, no. 1, pp. 1–11, 2021.

[15] A. Krishnaswamy Rangarajan and R. Purushothaman, "Disease classification in eggplant using pre-trained vgg16 and msvm," *Scientific reports*, vol. 10, no. 1, pp. 1–11, 2020.