Research Article

Husam Ali Abdulmohsin*

A speech-based convolutional neural network for human body posture classification

https://doi.org/10.1515/jisys-2022-0326
received December 30, 2022; accepted February 22, 2024

Abstract
Problem – In recent years, computing and sensing advances have helped to develop efficient human posture classification systems, which assist in creating health systems that contribute to enhancing elder's and disable's life quality in context-aware and ambient assistive living applications. Other applications of body posture classification include the simulation of human bodies into virtual characters, security applications such as kidnapping and the body position of the kidnapper or victim, which can provide useful information to the negotiator or the rescue team, and other sports applications.
Aim – This work aims to propose a body posture classification system based on speech using deep learning techniques.
Methods – All samples pass through the preprocessing phase. The Mel-frequency cepstral coefficient (MFCC) with 12 coefficients was used as features, and the best features were selected by the correlation-based feature selection method. Two supervised learning techniques called artificial hydrocarbon networks (AHN) and convolutional neural networks (CNN) were deployed to efficiently classify body postures and movements. This research aims to detect six human positions through speech, which include walking downstairs, sitting, walking upstairs, running, laying, and walking. The dataset used in this work was prepared by the authors, named the human activity speech dataset.
Results – The best performance gained was with the AHN classifier of 68.9% accuracy and 67.1% accuracy with the CNN.
Conclusion – It was concluded that the MFCC features are the most powerful features in body position classification, and the deep learning methods are powerful in classifying body positions.

Keywords: classification of body movements, health system, body position detection, Mel-frequency cepstral coefficient, correlation-based feature selection, artificial hydrocarbon networks, and convolutional neural network

1 Introduction

Body posture analysis is useful for many applications, the most important of which being at-home health monitoring and assisted living. Tracking body postures in computer vision is a fundamental problem, taking into account the complexity of human postures, where there are thousands of human postures [1], and the lack of a publicly accessible dataset.

* **Corresponding author: Husam Ali Abdulmohsin,** Department of Computer Science, College of Science, University of Baghdad, Al-Jaderya, Baghdad, 10081, Iraq, e-mail: husam.a@sc.uobaghdad.edu.iq

Through non-verbal communication, everyone possesses a wealth of knowledge, beliefs, and experience [2–5]. When you engage with people on a daily basis, three different types of non-verbal communication are used. The first deals with non-verbal communication; the second, with receiving them; and the third, with the intricate interaction between the first two. One, first communicates with people through sending (or encoding) non-verbal cues, sometimes consciously and sometimes not. In the former, your objective is for the other person to comprehend a certain message that you have sent to him through one or more non-verbal cue channels, such as your tone of voice, posture, and facial expression. Additionally, you could provide others with non-verbal cues without meaning to sent such a false message [6]; this kind of communication is not the focus of the research. However, there are times in which important information about your emotional state [7], attitudes [8], and intentions [9] leaks out of you non-verbally.

Researchers have used non-verbal actions to translate non-verbal communication in many applications in life [10], such as in health [11–13], where non-verbal communication was used in human physical therapy and health-care systems, in human–robot communication [14–22], to increase the intelligence of robots as understanding humans reactions and questions.

Virtual persons, also known as virtual characters with human-like bodily structures, have attracted a lot of researcher's attention in recent years. We can employ non-verbal cues, which are widely used in face-to-face or human–robotics communication to clarify one's intentions or the context of the words uttered, by incorporating these virtual individuals into a system [23–25].

Now that we have covered the non-verbal posture classification systems and their applications, it is time to talk about the verbal posture classification systems, which are human body posture classification systems (BPCS) that are based on speech as input information to classify body postures. The applications of such systems are not different from non-verbal systems since both are aiming at posture classification, but the difference between them is the input information. In non-verbal systems, the input is either image, video, or signals coming from sensors installed on the human body. Rather than verbal systems, where the input is speech and nothing else than speech.

Now, regardless of all the research on non-verbal human posture classification, there is a big gap in verbal human posture classification, which is the target of this work. Verbal human posture classification means translating speech to human body postures, as this article will define in detail; in other words, this work aims to identify the human body postures from the human speech at that same moment.

The limitation of the literature is in translating human verbal to human body postures [26], mentioning that some of the researchers have detected robot head actions and translated them to emotions, some translated body actions to emotions [27,28], but none have translated speech to body actions, which is the target of this work.

The main contributions of this article are:

Presenting a novel BPCS based on speech using artificial hydrocarbon networks (AHN), which in turn will classify six different body postures of the human being using their speech. These six postures are walking, walking upstairs, walking downstairs, sitting, laying, and running.

- Such systems are essential when the data source is limited to speech only.
- Extracting Mel-frequency cepstral coefficient (MFCC) features from human speech, and the correlation-based feature selection (CFS) algorithm was used to remove the most correlated features and improve the system time-consuming.

The first problem for the proposed work is noise; if the speech source contains high noise, results will be affected. The second problem is the implementation of such systems in environments where speech is absent, which can be for many reasons, such as speechless human beings.

The remaining sections of the article are as follows: Section 2 reviews the methodology of the proposed system, Section 3 discusses the results, Section 4 discusses the limitations of the proposed work, Section 5 explains the limitations of the proposed work, and Section 6 discusses the conclusion and future work.

2 Related works

There are many approaches to classifying human postures depending on the type of input to those systems, such as image or video based, sensor based, and speech based as the target of this system.

Many researchers have worked on image-based posture classification, such as in 2021 [29], built a model using ResNet. The model was trained using the hierarchically labeled dataset Yoga-82. A novice will be able to understand several classification levels connected to a specific yoga pose with the aid of the designed model. The proposed model significantly outperformed previous models produced for tracking people's fitness levels, providing a major boost to yoga applications.

Gaikwad et al. [30] implemented an LSTM-based self-data procurement system that was followed by real-time body posture recognition. The accuracy of the posture classification is 93.75%. If the present frames are not correct, the results will be shown right away. Because they can see the accuracy of their performance in real time, users can rapidly fix their posture even if they perform their exercises incorrectly.

Madake et al. [31] proposed a viable method for identifying golf swings with computer vision techniques, which may have an impact on improving sports performance monitoring. To identify different swing types and postures, including the backswing, downswing, and follow through, videographic data of golf swings were used. Dense and sparse optical flow were employed by the system to determine the trajectory of the golf club. The histogram of oriented gradients is then used to extract the characteristics from these trajectories. With the help of morphological image frames and the scale-invariant feature transform, the golfer's posture feature during a particular stroke is extracted. Several classification approaches are applied to the retrieved features once they have been normalized and concatenated. In random forest classification, the system obtains 84.4% accuracy when tested on the standard GolfDB dataset.

Zhang et al. [32] developed a novel method to identify depression using the skeletal data while assessing the scale. Human skeletal data were captured using Kinect V2, and participants were divided into two groups: one for depression and the other for control. Participants were evaluated using the Hamilton Rating Scale for Depression. They developed a Temporal Spatial Attention (Dep-TSA) for depression identification in order to improve the classification ability of temporal and spatial elements within skeletal data. A total of 202 participants met the criteria of the scale; these comprised the depression group, which consisted of 89 patients with depression, and the control group, which consisted of 113 socially recruited healthy persons. In comparison to other models, the proposed Dep-TSA model produced exceptional results, obtaining an accuracy of 72.13%.

Other researchers have proposed posture classification systems depending on sensor data input. In 2024, they created a rule-based model to categorize squatting and kneeling by including data from three AX3 accelerometers – which measure the acceleration of body segments – into the samples of fifteen healthy people. The sensors were affixed below the hip's iliac crest, on the anterior and distal regions of the femur, and immediately below the head of the fibula. The postures were captured with a GoPro camera for validation, and special software was utilized to synchronize data from the accelerometers with the annotated video recordings. Totally 98% of the time, accuracy was attained when kneeling and 57% when squatting. Additionally, they discovered that when sensors were placed on just one side of the body, their model functioned a little bit less accurately.

Qiu et al. [33] provided two categories of automated fall detection methods that make use of camera visual data: (1) A self-supervised autoencoder that recognizes falls as an anomaly and separates them from normal behavior and (2) A supervised human posture-based fall activity recognition system. Two publicly accessible video datasets that include everyday living activities and simulated falls in an office setting are used to train and assess five models. We created two new datasets, one with more complicated backgrounds and scenarios and the other with recordings of actual falls in senior citizens, in order to test the models for fall detection in real-world scenarios. According to the experimental findings, autoencoder detectors may accurately anticipate falls based only on photos in which the background has been previously learnt. The stance-based method, on the other hand, better targets complicated scenes and backgrounds by using foreground body poses exclusively for AI learning.

Zeng et al. [34] introduced PyroSense, a passive infrared sensor (PIR sensor), that is widely available off-the-shelf (COTS) that allows for fine-grained 3D posture reconstruction. PyroSense reconstructs the appropriate human postures in real time by sensing airflow caused by movement and heat signals produced by the human body. By enhancing the COTS PIR sensor's sensitivity to body movement, boosting spatial resolution without adding to the deployment overhead, and creating clever algorithms that can adjust to a variety of environmental conditions, PyroSense significantly improves upon the previous PIR-based sensing architecture. Using readily available hardware, they constructed a low-cost PyroSense prototype. The results of the trial show that PyroSense achieves 99.46% classification accuracy in 15 classes and less than 16 cm mean joint distance error for 14 body joints in posture reconstruction.

With all the state-of-the-art works related to body posture classification, none were based on speech in classifying body postures. Therefore, the work proposed in this article is considered novel and a new idea of body posture classification.

3 Materials and methods

The proposed work block diagram is shown in Figure 1, where all the main steps followed are shown clearly, and through this section, each step will be explained in detail.

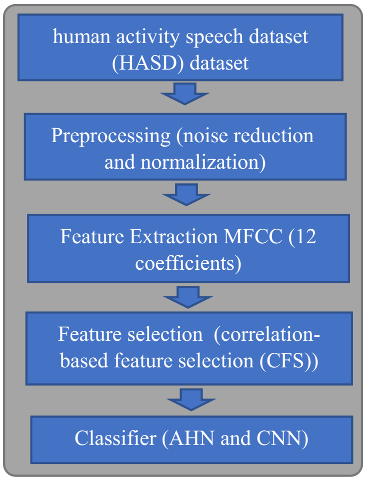

human activity speech dataset (HASD) dataset
Preprocessing (noise reduction and normalization)
Feature Extraction MFCC (12 coefficients)
Feature selection (correlation-based feature selection (CFS))
Classifier (AHN and CNN)

Figure 1: The block diagram of the proposed BPCS.

3.1 Dataset description

The author recorded a dataset called the human activity speech dataset (HASD) specifically for this study. Version VR01 of the EVIDA recording equipment was used to record the dataset. The gadget uses a professional recording chip and an enhanced noise-cancelling microphone. High-quality audio can be recorded at 128, 192, or 1,536 kbps using WAV format. Voice decibels can be adjusted at six different levels with this tiny recorder gadget. Regardless of the volume level, this recorder will only capture spoken voice when the voice decibel level is set correctly, reducing blank and whispery snippets. MP3 Audio and WAV are the two recording formats supported by this device. The device's noise reduction feature was enabled by setting it to level 6 decibels and WAV format.

Thirty-four volunteers, 17 of whom were female and 17 of whom were male, between the ages of 25 and 72, participated in the trials. Wearing the EVIDA recording device on their chest, each subject completed six different activities: walking downstairs, sitting, walking upstairs, running, laying, and walking.

Through every action, two sentences were recorded for every object. "Today is a beautiful day for doing sports, I'm reading a history book in my backyard" was one of the two sentences that were utilized. A total of

408 recordings, lasting 3 s each, were acquired, with 68 recordings for each action. It should be mentioned that the recordings were made on separate days and at various periods of the day. There was silence in the room where the recording was made. The objects were instructed to put their phones and other electronics in flight mode, leave them outside the room, and refrain from wearing any jewelry or watches.

The captured audio was manually labeled. After the dataset was received, it was randomly divided into three sets: 70% were chosen for neural network training, 15% were chosen for validation, and 15% were chosen for testing.

3.2 Preprocessing

The preprocessing phase contained two important functions, the noise reduction Wiener filter and the normalization function to limit the boundaries of the values generated from the noise reduction phase.

3.3 Feature extraction

Twelve coefficients of the MFCC, which is frequently utilized in voice recognition as well as earlier speech-gesture investigations [35,36], were included as features in the recognition system. In order to represent change over time, we examined a feature set that only included the pitch (F0) and its first and second derivatives.

3.4 Feature selection

The discriminative and pertinent features for model construction are determined by this process. With simple models, feature selection approaches are used to reduce training time and increase generalization potential by lowering overfitting. Its main goal is to get rid of the unnecessary and minor aspects.

This study uses a CFS methodology [37]. The discriminative features that have a strong link with a class instance are the only ones that are chosen by the CFS method after evaluating the subset of attributes. The qualities are ranked by CFS using a correlation-based heuristic evaluation algorithm. It is employed to gauge how similar different traits are. CFS ignores qualities that do not correlate well with the class designation. The following is the CFS criterion shown in equation (1):

$$\mathrm{CFS} = \mathrm{Max}\left[\frac{r_{cf1} + r_{cf2} + \cdots + r_{cfn}}{\sqrt{k + 2(r_{f1f2} + r_{fifj} + \cdots + r_{fnfn-1})}}\right], \tag{1}$$

where r_{cfi} is a feature classification correlation, n is the number of features, and r_{fifj} represents the correlation between features. The chosen features are provided to the classifiers for speech recognition.

3.5 Classification

In comparison to other well-known machine learning models, AHN, a supervised learning technique inspired by organic chemical structures and mechanisms, have demonstrated gains in prediction power and interpretability. However, AHN are time-consuming and have not yet been able to handle enormous amount of data [38,39]. The convolutional neural network (CNN) was used as a second classifier for comparison purposes between the two classifiers.

4 Results and discussion

The confusion matrices and the ROC line charts were used to measure the numerical and visual outcomes of the proposed system when it came to the CNN and AHN classifier. Only the AHN results will be presented and contrasted with the CNN results.

The ROC curve is a performance measurement diagnostic tool, which is translated by the area under the curve of each body position; the greater the area is, the higher the discriminative ability measure will be for that specific body position related specifically to that curve. Figure 2(a) displays the ROC curves of the training phase, Figure 2(b) displays the ROC curves of the validation phase (also known as the training record error values), and Figure 2(c) displays the ROC curves of the testing phase. Figure 2 displays the ROC curves for the classification processes on the HASD dataset. The power of the system’s discriminative ability is demonstrated by the fact that every curve in the ROC line chart displayed in Figure 2(a) and (c) is above the average gray line, which runs from the left down corner to the right up corner. The system only deviated from the average line during the validation phase when it was trying to distinguish between the walking position and other positions. This can be explained by the system’s inability to distinguish between the two positions. Anticipating the ROC curves in the testing phase, we find that the suggested BPSD system achieves 68.9% accuracy in recognizing all positions represented in the HASD dataset.

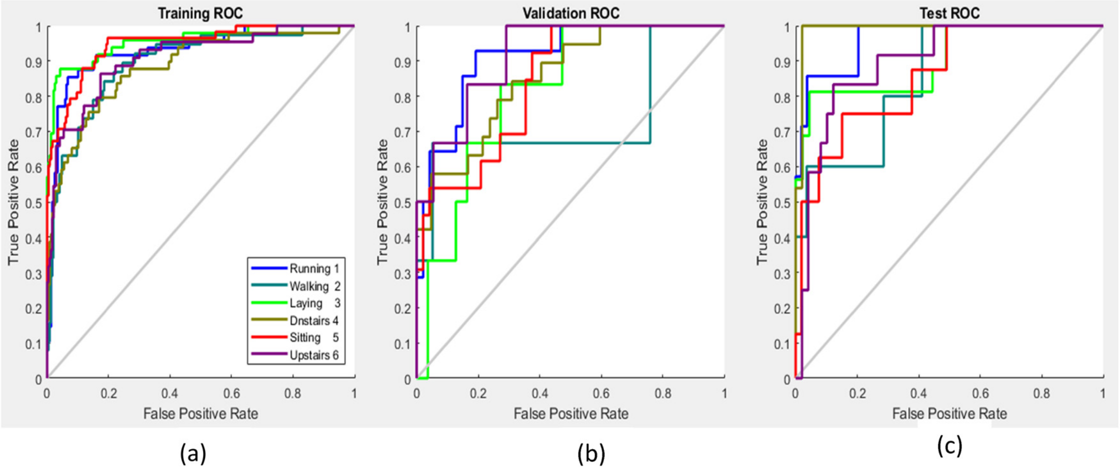
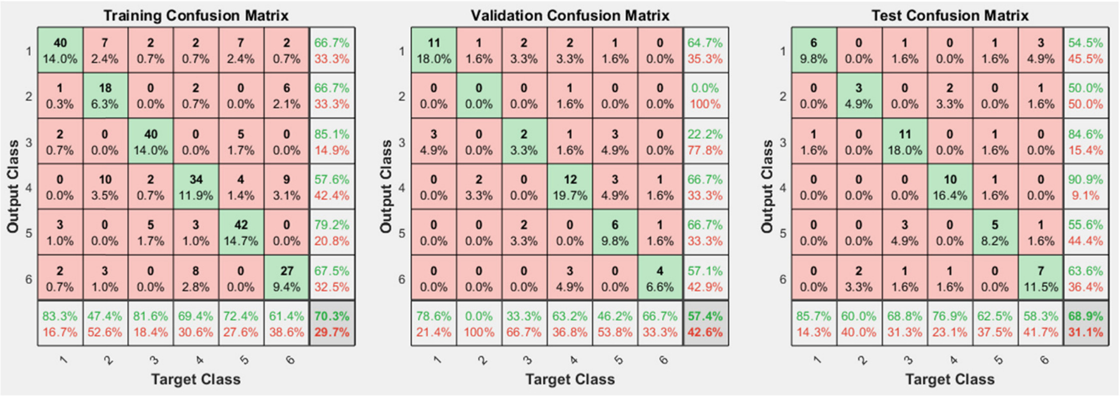


Figure 2: ROC curves performance measurement for the proposed BPCS: (a) for the training phase, (b) for the validation phase, and (c) for the testing phase.

The walking downstairs body position in Figure 2(c) is the closest ROC curve to the upper left corner, indicating that the BPSD system was more successful in differentiating this body position than the other positions. This can be explained by the bumps that result from a person falling down stairs and the voice-generated reflections that follow. Because of this, all of the testing phase’s ROC curves, as seen in Figure 2(c), are above the average line, demonstrating the BPSD system’s success in differentiating everybody's position with an accuracy level above 50%.

Now, we will discuss numerically the results gained through the confusion matrices shown in Figure (3). The numbers of the columns from 1 to 6 represent the positions (walking downstairs, sitting, walking upstairs, running, laying, and walking), respectively. The testing results of the proposed BPSD system proposed according to discriminating the six body positions, where (85.7, 60, 68.8, 76.9, 62.5, and 58.3%) for the (walking downstairs, sitting, walking upstairs, running, laying, and walking), respectively, shown in Figure 3(c). The average of discriminating all six body positions was 68.9% during the testing phase. As we can see, the highest discriminating accuracy was with the walking downstairs body position, and this was justified previously. The lowest accuracy was gained with the walking body position, with 58.3% accuracy. Walking downstairs



Training Confusion Matrix

Output Class

	1	2	3	4	5	6	
1	40 14.0%	7 2.4%	2 0.7%	2 0.7%	7 2.4%	2 0.7%	66.7% 33.3%
2	1 0.3%	18 6.3%	0 0.0%	2 0.7%	0 0.0%	6 2.1%	66.7% 33.3%
3	2 0.7%	0 0.0%	40 14.0%	0 0.0%	5 1.7%	0 0.0%	85.1% 14.9%
4	0 0.0%	10 3.5%	2 0.7%	34 11.9%	4 1.4%	9 3.1%	57.6% 42.4%
5	3 1.0%	0 0.0%	5 1.7%	3 1.0%	42 14.7%	0 0.0%	79.2% 20.8%
6	2 0.7%	3 1.0%	0 0.0%	8 2.8%	0 0.0%	27 9.4%	67.5% 32.5%
	83.3% 16.7%	47.4% 52.6%	81.6% 18.4%	69.4% 30.6%	72.4% 27.6%	61.4% 38.6%	70.3% 29.7%

Target Class

(a)

Validation Confusion Matrix

Output Class

	1	2	3	4	5	6	
1	11 18.0%	1 1.6%	2 3.3%	2 3.3%	1 1.6%	0 0.0%	64.7% 35.3%
2	0 0.0%	0 0.0%	0 0.0%	1 1.6%	0 0.0%	0 0.0%	0.0% 100%
3	3 4.9%	0 0.0%	2 3.3%	1 1.6%	3 4.9%	0 0.0%	22.2% 77.8%
4	0 0.0%	2 3.3%	0 0.0%	12 19.7%	3 4.9%	1 1.6%	66.7% 33.3%
5	0 0.0%	0 0.0%	2 3.3%	0 0.0%	6 9.8%	1 1.6%	66.7% 33.3%
6	0 0.0%	0 0.0%	0 0.0%	3 4.9%	0 0.0%	4 6.6%	57.1% 42.9%
	78.6% 21.4%	0.0% 100%	33.3% 66.7%	63.2% 36.8%	46.2% 53.8%	66.7% 33.3%	57.4% 42.6%

Target Class

(b)

Test Confusion Matrix

Output Class

	1	2	3	4	5	6	
1	6 9.8%	0 0.0%	1 1.6%	0 0.0%	1 1.6%	3 4.9%	54.5% 45.5%
2	0 0.0%	3 4.9%	0 0.0%	2 3.3%	0 0.0%	1 1.6%	50.0% 50.0%
3	1 1.6%	0 0.0%	11 18.0%	0 0.0%	1 1.6%	0 0.0%	84.6% 15.4%
4	0 0.0%	0 0.0%	0 0.0%	10 16.4%	1 1.6%	0 0.0%	90.9% 9.1%
5	0 0.0%	0 0.0%	3 4.9%	0 0.0%	5 8.2%	1 1.6%	55.6% 44.4%
6	0 0.0%	2 3.3%	1 1.6%	1 1.6%	0 0.0%	7 11.5%	63.6% 36.4%
	85.7% 14.3%	60.0% 40.0%	68.8% 31.3%	76.9% 23.1%	62.5% 37.5%	58.3% 41.7%	68.9% 31.1%

Target Class

(c)

Figure 3: Confusion matrices of the proposed BPCS: (a) for the training phase, (b) for the validation phase, and (c) for the testing phase.

achieved the highest accuracy, and this can be justified by the unique attitude of the human body, because when the body hits the ground, getting downstairs affects the voice from the reactions of the human body hitting the ground, not like running, generates less reaction in the human body. Three of the walking samples were classified as walking downstairs, and this depends on the nature of the human being walking, some human beings walk softly, and some hit the ground strong when they walk. Some of the results have no reasonable reasons, but only can be clarified to the bad performance of the subjects deployed in recording this dataset or to the failure of the classifiers used in this work, such as three samples of the walking upstairs were classified as laying down, two samples of the sitting records were classified as walking, two samples of the running were classified as sitting, and three samples of the walking were classified as walking downstairs.

Through the confusion matrices, it is noticed that the random samples selected for the testing phase contained 7 walking downstairs recording samples, 5 sitting recording samples, 16 walking upstairs recording samples, 13 running recording samples, 8 laying recording samples, and 12 walking recording samples, which shows 61 recording samples used for the testing phase, which is 15% of the samples of the HASD dataset, and also shows that there was no balancing in choosing the samples according to the body positions, but this is caused from the random selection of the samples in the all three phase. In order to achieve balance in choosing the samples according to body position, the random selection has to be neglected, and this will affect the judgment of the proposed system.

Figure 4 shows the best validation performance of the proposed BPCS, during the training phase, which shows that the proposed system needed 17 epochs to learn how to discriminate the samples in the HASD dataset, which is considered a reasonable time.

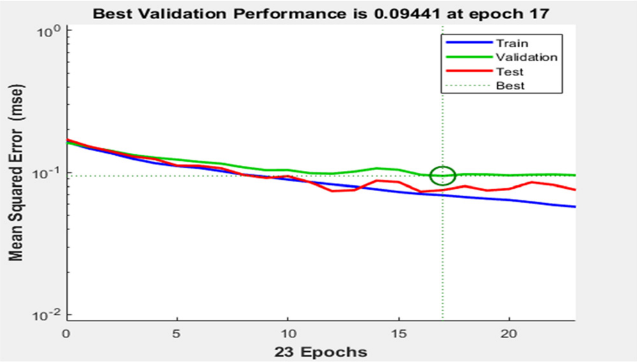


Figure 4: The best validation performance of the proposed BPSD system.

In seconds, the system using the AHN needed 103 s to learn about the 34 human speeches, and needed less than one second to recognize the human positions, which is considered a real-time system. To be mentioned, the CNN needed less time, which was 56.66 s, but achieved less accuracy because the CNN operates faster on huge data.

The results are reasonable because of the following considerations: first, the final MSE is small, the test and validation lines are characteristically similar in wave behavior, and no significant overfitting has occurred between the validation and testing lines.

A comparison study was implemented to evaluate the performance of the proposed approach. Table 1 shows the results of verbal and non-verbal studies compared to the results of the proposed study. The results of the literature listed in Table 1 show that all literature research studies were implemented on robotics, and none was implemented on human beings, but not including [28], who worked on elder people only, rather than our project, which worked on different ages, and it is important to mention that dealing with the flexible body of the human being with different dimensions is much harder than dealing with robotics and its sharp body parts. The results show that the proposed system competes with the literature but needs to be enhanced, as mentioned in the limitation in the introduction. Table 1 shows that the highest accuracy achieved was 83.16% with ANOVA, but still, they only detected the head position only, rather than the whole body as implemented in this work. The highest accuracy that can be compared to our results is with the RBF network (RBFN) implanted by McColl and Nejat [28], which achieved 77.9% accuracy, but with elder people only, as mentioned before.

Table 1: Comparison study with literature

Ref.	Classification method	Accuracy %	Part of the body included
[28]	*K*-nearest neighbor	72.1	All body, for body language purpose
	Naïve bayes	70.7	
	Logistic regression	70.0	
	Random forest (RF)	72.9	
	Adaptive boosting (Adaboost) with Naïve Bayes (ANB)	70.0	
	RBFN	77.9	
	Support vector machines	66.4	
[27]	ANOVA	83.16	Six positions for the robotic head only for emotional recognition
[26]	ANOVA	60	Six parts of the body (arm, hand, neck, mouth, eyes, and chest) for detecting robot behavior
Proposed work	CNN	67.1	Six positions for the human body position detection
	AHN	68.9	

5 Limitation of the proposed work

The main limitation of the proposed work is the number of human body postures that can be classified using speech, which is very limited compared to the hundreds of postures that the human body can generate because many human postures affect speech in a similar way that no system can distinguish. The second limitation of this work, it does not propose a system to the real world, never the less, not using speech through phone calls, really does not reflect the real purpose of this work, but because of the limitation of resources and the legal authorization needed to use recorded phone calls, this work depended on a recorded dataset.

6 Conclusion

The highest accuracy obtained from the implemented experiments was 68.9%, indicating that the system was powerful in differentiating some body positions but partially failed in differentiating others. This can be demonstrated by using additional features or alternative classifiers.

Through experimentation, it was discovered that wave behavior of the (walking and walking downstairs recording samples), (walking upstairs and running samples), and (sitting and laying samples) shared some similarities as pairs. Because of these similarities in wave behavior, the proposed system was not good enough to distinguish some body positions. It is crucial to note that the reason for this resemblance in wave behavior is the same influence that the body's posture has on the lungs and other parts of the body. Therefore, it might be very difficult to distinguish some body positions from speech, especially in a noisy environment.

The primary motivation for this work was to assess human activities in order to gauge the individual's health and, second, for security reasons. Therefore, creating a realistic dataset will be more usable in real-world applications. The dataset created through this work included many samples; however, only the best-recorded samples were chosen.

It is recommended for future work to study other positions of the human body and use datasets recorded through languages other than English. The most important is to build a dataset that is gathered from a realistic environment that will bring this work to real life.

This work did not deal with any noise reduction, because the dataset deployed in this work was limited to noise, so adding a noise reduction method during the preprocessing phase of this work will bring this work more to reality.

Acknowledgements: The authors want to thank all researcher's works cited in this work in the field of body position detection.

Funding information: This project is funded by the author only.

Author contribution: The author confirms the sole responsibility for the conception of the study, presented results and manuscript preparation.

Conflict of interest: The author declares that there is no conflict of interest regarding the publication of this article.

Data availability statement: Data sharing is not applicable to this article, as no datasets were generated or analyzed during the current study.

References

[1] Hewes GWJSA. The anthropology of posture. Sci Am. 1957;196(2):122–33.

[2] Martin AM, O'Connor FM, Rosemary L. Non-verbal communication between nurses and people with an intellectual disability: a review of the literature. J Intellect Disabil. 2010;14(4):303–14.

[3] Sime D. What do learners make of teachers' gestures in the language classroom?. Int Rev Appl Linguist Lang Teach. 2006;44(2):211–30.

[4] Ting-Toomey S, Chung LC. Understanding intercultural communication. New York: Oxford University Press; 2005.

[5] Attard A, Coulson NS. A thematic analysis of patient communication in Parkinson's disease online support group discussion forums. Comput Hum Behav. 2012;28(2):500–6.

[6] Knapp ML, Hall JA, Horgan TG. Nonverbal communication in human interaction. Holt, Rinehart and Winston: Cengage Learning; 2013.

[7] Wojciechowski J, Stolarski M, Matthews GJPO. Emotional intelligence and mismatching expressive and verbal messages: A contribution to detection of deception. PloS one. 2014;9(3):e92570.

[8] Jagoe C, Wharton T. Meaning non-verbally: the neglected corners of the bi-dimensional continuum communication in people with aphasia. J Pragmat. 2021;178:21–30.

[9] Roseano P, González M, Borràs-Comes J, Prieto P. Communicating epistemic stance: How speech and gesture patterns reflect epistemicity and evidentiality. Discourse Process. 2016;53(3):135–74.

[10] Riggio RE, Feldman RS. Applications of nonverbal communication. New York: Psychology Press; 2005.

[11] Ambady N, Koo J, Rosenthal R, Winograd CH. Physical therapists' nonverbal communication predicts geriatric patients' health outcomes. Psychol Aging. 2002;17(3):443–52.

[12] Friedman HS. Nonverbal communication between patients and medical practitioners. J Soc Issues. 1979;35(1):82–99.

[13] Singh NN, McKay JD, Singh AN. Culture and mental health: Nonverbal communication. J Child Family Stud. 1998;7(4):403–9.

[14] Saunderson S, Nejat G. How robots influence humans: A survey of nonverbal communication in social human–robot interaction. Int J Soc Robot. 2019;11(4):575–608.

[15] Breazeal C, Kidd CD, Thomaz AL, Hoffman G, Berlin M. Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In 2005 IEEE/RSJ International Conference On Intelligent Robots And Systems. IEEE; 2005.

[16] Shen Z, Elibol A, Chong NY. Understanding nonverbal communication cues of human personality traits in human-robot interaction. IEEE/CAA J Autom Sin. 2020;7(6):1465–77.

[17] Brooks AG, Arkin RC. Behavioral overlays for non-verbal communication expression on a humanoid robot. Auton Robot. 2007;22(1):55–74.

[18] Bicho E, Louro L, Erlhagen W. Integrating verbal and nonverbal communication in a dynamic neural field architecture for human-robot interaction. Front Neurorobot. 2010;4:5.

[19] Vogeley K, Bente G. "Artificial humans": Psychology and neuroscience perspectives on embodiment and nonverbal communication. Neural Network. 2010;23(8–9):1077–90.

[20] Admoni H. Nonverbal communication in socially assistive human-robot interaction. AI Matters. 2016;2(4):9–10.

[21] Brock H, Sabanovic S, Nakamura K, Gomez R. Robust real-time hand gestural recognition for non-verbal communication with tabletop robot haru. In 2020 29th IEEE International Conference On Robot And Human Interactive Communication (RO-MAN). IEEE; 2020.

[22] Mavridis NJR, Systems A. A review of verbal and non-verbal human–robot interactive communication. Robot Auton Syst. 2015;63:22–35.

[23] Llargues Asensio JM, Peralta J, Arrabales R, Bedia MG, Cortez P, Peña AL. Artificial intelligence approaches for the generation and assessment of believable human-like behaviour in virtual characters. Expert Syst Appl. 2014;41(16):7281–90.

[24] Lee J, Marsella S. Nonverbal behavior generator for embodied conversational agents. International Workshop on Intelligent Virtual Agents. Springer, Berlin, Heidelberg: Springer; 2006.

[25] Jiang X. Perceptual attributes of human-like animal stickers as nonverbal cues encoding social expressions in virtual communication. Types of nonverbal communication. London, United Kingdom: IntechOpen Limited; 2021. p. 83.

[26] Wang E, Lignos C, Vatsal A, Scassellati B. Effects of head movement on perceptions of humanoid robot behavior. Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-robot Interaction. 2006.

[27] Beck A, Cañamero L, Bard KA. Towards an affect space for robots to display emotional body language. In 19th International symposium in robot and human interactive communication. IEEE; 2010.

[28] McColl D, Nejat G. Determining the affective body language of older adults during socially assistive HRI. In 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE; 2014.

[29] NL JV, Bagaria CK. Yoga pose classification using resnet of deep learning models. i-Manager's J Comput Sci. 2021;9(2):29.

[30] Gaikwad S, Bhatlawande S, Dusane A, Bobby D, Durole K, Shilaskar S. Vision-based posture detection for rehabilitation program. Second International Conference on Emerging Trends in Engineering (ICETE 2023). Atlantis Press; 2023.

[31] Madake J, Sirshikar S, Kulkarni S, Bhatlawande S. Golf shot swing recognition using dense optical flow. In 2023 IEEE International Conference on Blockchain and Distributed Systems Security (ICBDS). IEEE; 2023.

[32] Zhang K, Wan Y, Ning C, Fan Y, Wang H, Xing L, et al. A novel approach for depression detection through extracted skeletal data during scale assessment. Durham, USA: Research square; 2024.

[33] Qiu Y, Meng J, Li BJJI. Automated falls detection using visual anomaly detection and pose-based approaches: experimental review and evaluation. J ISSN. 2024;2766:2276.

[34] Zeng H, Li G, Li T. PyroSense: 3D posture reconstruction using pyroelectric infrared sensing. Proc ACM Interact Mob Wearable Ubiquitous Technol. 2024;7(4):1–32.

[35] Ferstl Y, Neff M, McDonnell R. Adversarial gesture generation with realistic gesture phasing. Comput Graph. 2020;89:117–30.

[36] Kucherenko T, Hasegawa D, Henter GE, Kaneko N, Kjellström H. Analyzing input and output representations for speech-driven gesture generation. Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents. 2019.

[37] Li B, Wang Q, Hu J. Feature subset selection: a correlation-based SVM filter approach. IEEJ Trans Electr Electron Eng. 2011;6(2):173–9.

[38] Ponce H, de Campos Souza PV, Guimarães AJ, Gonzalez-Mora G. Stochastic parallel extreme artificial hydrocarbon networks: An implementation for fast and robust supervised machine learning in high-dimensional data. Eng Appl Artif Intell. 2020;89:103427.

[39] Ponce-Espinosa H, Ponce-Cruz P, Molina AJAON. Artificial organic networks. Switzerland: Springer Cham; 2014. p. 53–72.