

# Food Recognition System: A New Approach Based on Wavelet-LSTM

Ghulam Hussain<sup>1</sup>, Ali Raza Radhan<sup>2</sup>, Irfan Ali Tunio<sup>1</sup>, Mohsin Shaikh<sup>2</sup>, Umair Saeed Solangi<sup>1</sup>, Kamran Javed<sup>3\*</sup>

---

## Abstract:

An automated system for analyzing daily dietary intake is essential for human well-being and healthcare. This work presents a novel wearable necklace embedded with a piezoelectric sensor and a microcontroller to monitor food ingestion of users. To effectively represent the food ingestion patterns, the sensor signal is dynamically segmented using a bidirectional search technique. Each segmented food intake pattern consists of a chewing sequence and a swallow peak. We exploit wavelet transform to decompose the complex food ingestion patterns, collected by the sensor, into frequency sub-bands at discrete scales. The frequency sub-bands are used as sequences to train long short-term memory (LSTM) for the recognition of 5 food categories. Our proposed recognition model based on wavelet-LSTM recognizes 5 food classes with an accuracy of 98.1%.

**Keywords:** *Food recognition, Signal segmentation, Wearable sensors, Signal processing*

---

## Introduction

Obesity is the most common disease present in every third human being. It is defined as excessive fat deposition in a person's body, which is caused by the imbalance between energy intake and energy expenditure. People consuming high caloric food in large quantity suffer from obesity. The previous researches reported obesity as a source of other diseases, such as hepatitis, diabetes, cardiovascular and cancer [1]. A comprehensive review was previously performed to investigate the performance and usability of different food dietary monitoring systems, which incorporated various sensors

and applied numerous sophisticated signal processing techniques to determine eating behavior, classify food type, and estimate amount of food [2]. There were mainly two approaches discussed such as manual and automatic. Manual approaches are proven to be biased and inaccurate due to relying on subjective reports and recall. There is a need for an automated system to monitor the daily dietary intake of obese subjects. Moreover, the system should be non-invasive so that people can use it easily during routine tasks to regulate their energy intake. Improvement in the field of artificial neural networks provides an opportunity for

---

<sup>1</sup>Dept. Electronic Engineering, Quaid-e-Awam University, Larkana, Pakistan

<sup>2</sup>Dept. Computer Science, Quaid-e-Awam University, Larkana, Pakistan

<sup>3</sup>National Centre of Artificial Intelligence (NCAI), Saudi Data and Artificial Intelligence Authority (SDAIA), Riyadh, Saudi Arabia; Corresponding Author: [kamran@skku.edu](mailto:kamran@skku.edu), [enr.ghulam.hussain@quest.edu.pk](mailto:enr.ghulam.hussain@quest.edu.pk)

researchers to design an intelligent and non-invasive food intake system that can assist individuals to monitor their daily food intake without requiring a lot of effort. There are many studies to monitor the daily food intake using various sensors, such as a microphone, piezoelectric, accelerometer, gyroscope, camera, and strain gauge. The piezoelectric and microphone-based systems have attained higher recognition performance than the rest [3]. The authors in [3] attained an accuracy of 80.3% through the training of a conventional machine learning algorithm on manually extracted features. The manually extracted statistical features in different domains can improve food recognition as it is also validated in [4]. On the contrary to [3], Alshurafa et al. [4] improved the recognition by employing the handcrafted features in time- and frequency-domain (TFD), obtained using short-time Fourier transform. They have improved the food recognition performance, but their system lacks dynamic segmentation technique and efficient features extraction algorithm, which are essential for any recognition task.

The segmentation of the wearable sensor signals into essential parts assist in accurate analysis of events of interest. Here, the events are related to ingestion activity such as chewing and swallow. Segmentation of signal actually helps in separate out the ingestion activity related data from silent phases [5]. We were inspired by the previous methodology of signal segmentation [5], in which signals collected by the wearable devices [6-7] were segmented using detection and selection techniques. The important activity was detected first by applying some threshold maxima [5,8-9]. Later, the beginning and ending of the detected activity is determined in selection stage [5]. Unlike deterministic segmentation approaches, probabilistic Bayesian approach is also reported in literature for estimating the segments of signals [10].

Recently, a new approach based on deep learning (DL) algorithm automatically extracts the features in TFD to perform the

recognition of the ECG signals [5]. However, their approach uses traditional static segmentation (SS) to segment the signals. The SS cannot completely cover the events in the naturally occurring body-signals. The drawbacks of manual and static segmentation (SS) approaches [3-5] motivated us to design a novel bidirectional search (BS) algorithm that can segment the ingestion patterns (IPs) both automatically and dynamically. In this study, a piezoelectric sensor is employed and preferred over the microphone as the performance of microphone-based systems degrades owing to surrounding noise [4].

The contributions of the proposed study can be best described in three folds: 1) The novel wearable system is designed to collect ingestion patterns of different food categories, 2) An efficient signal segmentation approach known as bidirectional search algorithm is developed to dynamically segment ingestion events of varying size such as chewing and swallow, 3) Sophisticated wavelet-transform based LSTM model has outperformed current state-of-the-art studies by recognizing five food classes with an accuracy of 98.1%.

**Our Proposal:** The proposed food recognition system, shown in Figure 1, consists of a piezoelectric sensor, LilyPad microcontroller, and smartphone application. The sensor and the microcontroller are integrated into a stretchable necklace, which is worn around the neck by the subjects. The smartphone application communicates with the necklace via Bluetooth for data logging, and then it transfers the data to a cloud server for further data analytics. Moreover, the smartphone application provides an interface to users for interaction with the food intake system.

The sensor in the necklace generates distinct IPs for different food categories, as shown in Figure 2. Each intake pattern consists of chewing sequence and a swallow. Owing to different characteristics of food, chewing period while ingestion of each food varies. Therefore, dynamic segmentation is required

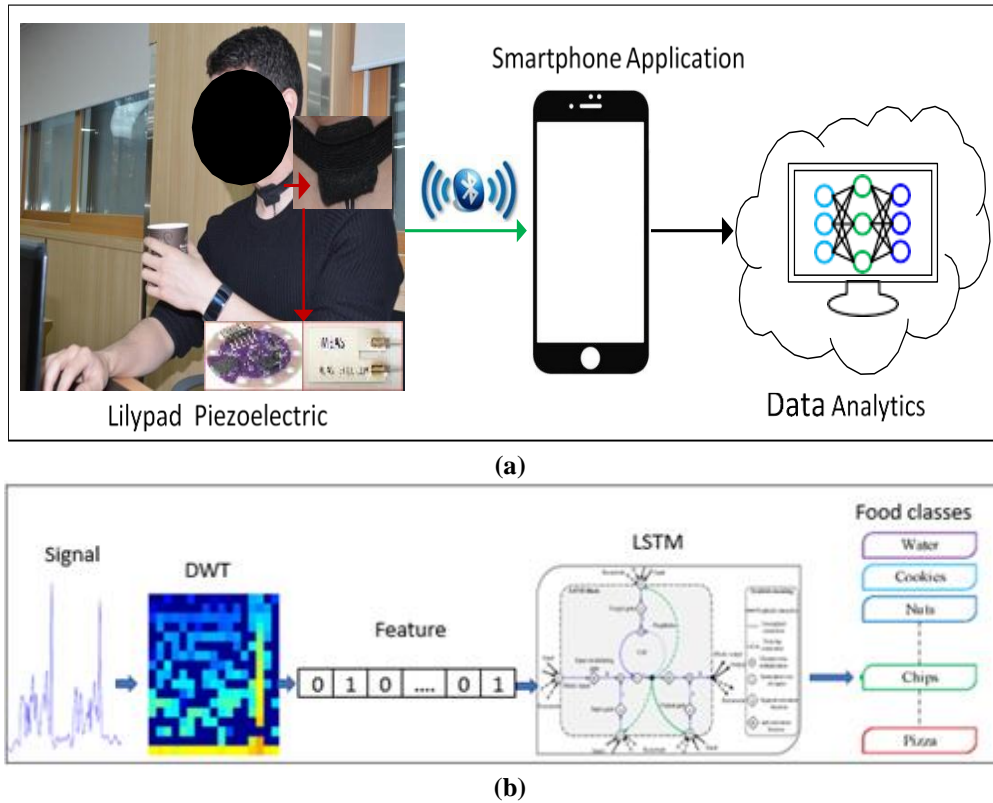


Figure 1: The proposed food recognition system (a) Wearable embedded module (b) Deep Learning based Food recognition model

to effectively represent the sensor data. The BS technique processes the input data and separates the IPs from unwanted data. In the proposed study, the word unwanted data refers to the data that is not related to the ingestion activity. BS technique, illustrated in Algorithm 1, performs two searches: first for swallow event in the forward direction; and second for supportive events or chewing sequence associated to the swallow event found during the first search. During the first search, the swallow event is spotted based on the change in the amplitude  $\Delta A_{qi}$  of neighboring samples.  $A_{qi}$  and  $A_{qr}$  denote  $i^{th}$  swallow event and related chewing sequence, respectively. Then, the related chewing sequence is searched in the backward direction. The start of chewing sequence is decided based on the change in the

amplitude  $A_{qr}$  of the samples. Swallow event and chewing sequence are determined by comparing the values of change in amplitudes ( $A_{qi}; A_{qr}$ ) to the thresholds ( $\theta_1; \theta_2$ ) as illustrated in the algorithm. Swallow event found in the forward search marks end of the IP. The chewing sequence found during backward search denotes the start of the IP. Thus, the BS technique helps to split the input data into the dynamic segments (DSs) automatically, shown in Figure 2(b). A fixed window length was chosen previously for the segmentation [3–14], but their accuracy degraded for the ingestion sequences of varying length as the duration of the IPs depends on the bite size and food type a person ingests. Therefore, the DSs are well suited to represent the IPs. The dimension of the DSs is

reduced further by the wavelet transform that forms the efficient sequences to be trained on the classifier. The computational complexity of the BS is  $O(N M)$ .  $N$  is length of each ingestion pattern and  $M$  is number of ingestion patterns.

**Wavelet Transform:** Wavelet transform decomposes the segments, containing complex food IPs, into frequency sub-bands at various discrete scales. Wavelet transform, unlike the Fourier transform, accurately analyses the data containing abrupt changes (swallows), by localizing spectral content of

changes as it carries the complex events of chewing and swallow together. Hence, there is a need for complex wavelet or basis function that can efficiently represent the complex IP. We implemented DWT using the filter-banks method. Different combinations of wavelets and levels of decomposition were tested using long short-term memory (LSTM). LSTM network was designed to overcome the vanishing gradient problem that can occur in traditional recurrent neural networks, allowing them to better handle long-term dependencies in sequential data [15].

---

**Algorithm 1** Bidirectional Search (BS) algorithm

---

**1: Input:**  $Y$  : Input data,  $\kappa=1$  : Search Limit,  $\phi_i$  : Segment index

**2: for**  $t \leftarrow 1, t_{max}$  **do**

**3: if** ( $y \leftarrow \text{FORWARDSEARCH}(Y[t:t+5])$ ) **is true then**

**4:** $\delta \leftarrow t + 5$

**5:** $v \leftarrow \text{BACKWARDSEARCH}(Y[\kappa:\delta])$

**6:** $\phi_i \leftarrow Y[v:\delta]$

**7:** $t \leftarrow \kappa = \delta + 1$

►  $\kappa$ : New Search Limit

**8:     else**

**9:         Search Again Go to Line2**

**10:     end if**

**11: end for**

**12: function** FORWARDSEARCH ( $Y[i:i+5]$ )

**13:     if** ( $\blacktriangle A_{qt} = \mu(Y[q=i+1:i+5]) - Y[i] < \theta_1$ ) **then**

**14:         return y as true**

**15:     else**

**16:         return y as false**

**17:     end if**

**18: end function**

**19: function** BACKWARDSEARCH( $Y[Y:i]$ )

**20:     for**  $r \leftarrow i, Y$  **do**

**21:         if** ( $\blacktriangle A_{qr} = \mu(Y[q=r-1:i-5]) - Y[r] < \theta_2$ ) **then**

**22:             return v  $\leftarrow r - 5$**

**23:         else**

**24:             Search Again Go to Line20**

**25:         end if**

**26:     end for**

**27: end function**

---

the signal in time. We employ a discrete wavelet transform (DWT) to characterize the oscillatory behavior of the IP segment. The IP signal consists of slow trends and abrupt

Recently, LSTM has emerged as the most widely applied model in analysis of sequential healthcare data [11]. Daubechi

wavelet with 4 levels of decomposition is selected as the best combination based on the recognition performance of the LSTM model. Therefore, we have chosen Daubechies

The chosen wavelet is scaled ( $s$ ) and shifted ( $\tau$ ) along the entire length ( $t$ ) of the segment to be multiplied later. The coefficient ( $\frac{1}{\sqrt{s}}$ )

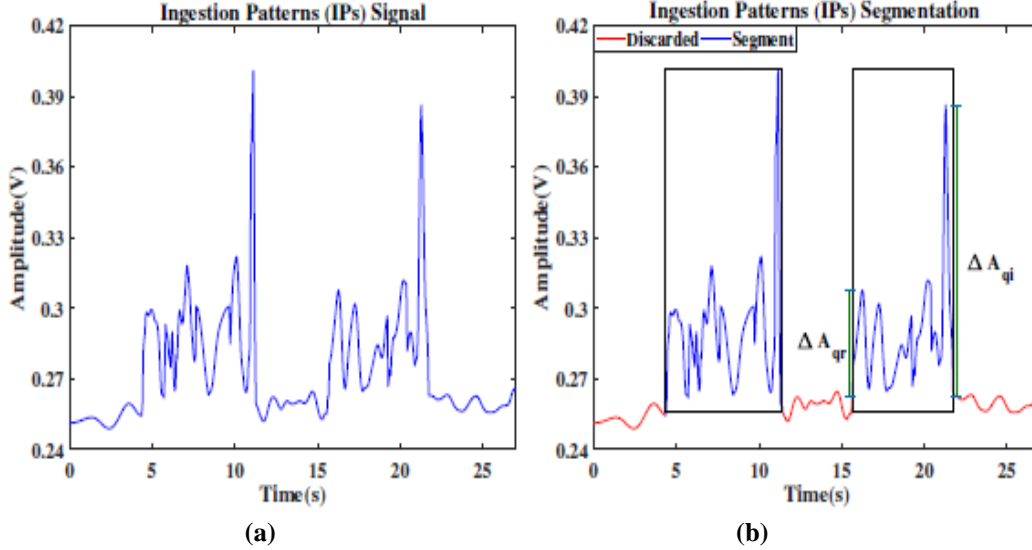


Figure 2: Wearable sensor signal (a) Ingestion patterns (IPs) signal, (b) IPS segmentation

wavelet as a mother wavelet ( $\Psi_{s,\tau}(t)$ ) given in Equation (1) to analyze the complex IP.

$$\Psi_{s,\tau}(t) = \frac{1}{\sqrt{s}} \Psi\left(\frac{t-\tau}{s}\right) \quad (1)$$

updated by substituting the parameters and given as.

$$\Psi_{j,k}(t) = 2^{\frac{1}{2}} \Psi(2^j t - k) \quad (2)$$

DWT analyze the input signal or segment ( $\phi_i$ ) by multiplying it with a wavelet function ( $\Psi$ ), which results in values of the coefficients  $c_i(j, k)$  as shown in Equation (3). The process is repeated for all segments to represent each  $\phi_i$  with fewer coefficients  $c_i(j, k)$ . The main aim of applying DWT on the segments is to represent the signal pattern with the efficient sequences, the coefficients, without acquiring redundancy.

normalizes energy of the wavelet. For the DWT representation, the  $s$  and  $\tau$  parameters are replaced with values of  $2^{-j}$  and  $k \cdot 2^{-j}$ , respectively. The  $j$  and  $k$  denote the scale and the shift parameters in the DWT. To define the wavelet basis ( $\Psi_{j,k}(t)$ ), the Equation (1) is

**Food recognition using wavelet-LSTM:** The coefficients, computed through the DWT, form efficient and accurate sequences for each

$$c_i(j, k) = \sum_t \phi_i(t) \Psi_{j,k}(t) \quad (3)$$

segment containing the IPs. The sequences for the IPs are fed into the LSTM. The LSTM contains input ( $\psi_t$ ), forget ( $f_t$ ), output ( $o_t$ ), and input modulating ( $\vartheta_t$ ) gates along with a memory cell ( $c_t$ ) and a hidden state ( $h_t$ ). The gates regulate the input information (sequences) so that it can be written to, read from, or stored in the memory during each time step as given in Equation (4). The LSTM, unlike other DL models, reduces the overall

number of parameters by sharing the same forward and backward processes temporally. The LSTM The forward and backward processes assist the trains on the input sequences by using the network to tune the parameters or weights

Table 1: The comparison of the previous studies and the proposed study

Study	Segmentation	Features (domain)	Accuracy
[3]	Static	Handcrafted features (time)	80.3%
[4]	Static	Handcrafted features (TFD)	90.0%
[11]	Static	Algorithm extracted features (TFD)	97.4%
<b>Proposed</b>	Dynamic	Algorithm extracted features (TFD)	98.1%

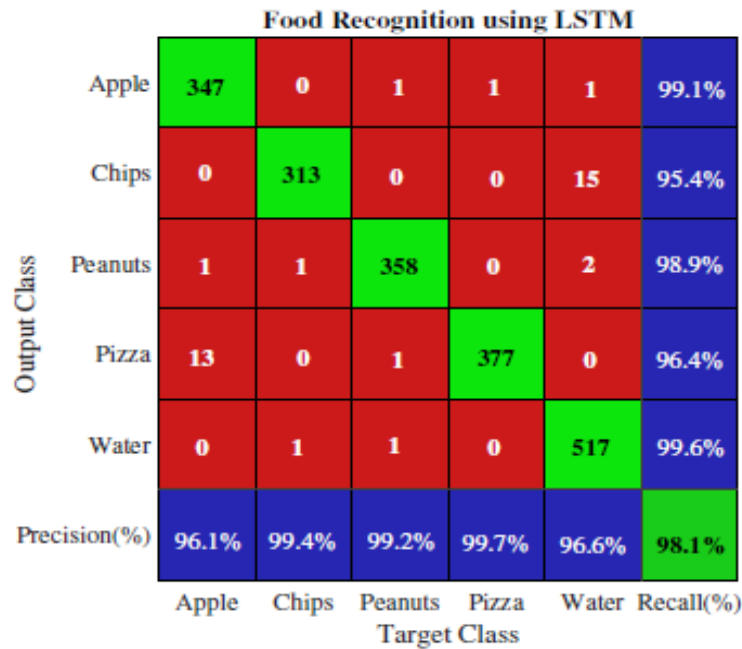


Figure 3: Food recognition using proposed system based on wavelet - LSTM

( $W, R, b, p$ ) using an optimization algorithm for minimizing the recognition error during training on the input sequences ( $x_t$ ) [4].

$$\vartheta_t = \tanh(W^\vartheta x_t + R^\vartheta h_{t-1} + b_\vartheta): \text{Input Modulating gate} \tag{4a}$$

$\psi_t = \text{sigm}(W^\psi x_t + R^\psi h_{t-1} + P^\psi \odot \theta_{t-1} + b_\psi)$ :  
Input gate (4b)

$f_t = \text{sigm}(W^f x_t + R^f h_{t-1} + p^f \odot \theta_{t-1} + b_f)$ :  
Forget gate (4c)

$o_t = \text{sigm}(W^o x_t + R^o h_{t-1} + p^o \odot \theta_{t-1} + b_o)$ :  
Output gate (4d)

$h_t = \tanh(\theta_t \odot o_t)$ : Hidden state or Block  
output (4e)

$\theta_t = \vartheta_t \odot \psi_t + \theta_{t-1} \odot f_t$  : Cell state (4f)

The LSTM network, a DL architecture, offers an end-to-end learning strategy, in which features extraction and classification are performed together automatically. Contrary to the conventional handcrafted features techniques, the DL model extracts the important and efficient features that enable the classifier to accurately predict the food class of the input sequences. We 4 bands discrete wavelet decomposition to generate components of detailed D1, D2, D3, D4 and approximate A4. These five detailed (D, D2, D3, D4) and approximate (A4) components are used as features for LSTM model. The data in the form of the sequences, carrying the tempo-spectral contents of IPs, are split into two sets: training (75%) and test (25%). A set of optimal hyperparameters is selected prior to training, as these parameters have a significant impact on the performance of the LSTM. Then, the LSTM model is trained on the training data, containing the IPs of 5 foods, to associate the wavelet sequences with the correct food class label. To assess the recognition-ability, the trained LSTM is evaluated on the test data (i.e., unseen sequences). The proposed model recognizes the test sequences of 5 food classes with an accuracy of 98.1% as shown in Figure 3.

**Results and Discussion:** A total of 15 subjects (10 males and 5 females, average age  $31.4 \pm 12.9$  years, average body mass index (BMI)  $27.1 \pm 6.09$  kg/m<sup>2</sup>) participated in the experiment. All subjects signed a consent form prior to the experiment and their information is protected following the declaration of Helsinki. There are five food categories, (such as chips, peanuts, pizza, apple and water), chosen for recording ingestive behavior of the

participants. Each subject took part in the experiment three times and ate food items from each of the categories. The subjects wore necklaces while they ingested the food items. The wearable sensor generates distinct IPs for each food category due to different characteristics such as hardness, stickiness, and crunchiness. These characteristics of foods alleviate the need for the complex classifier to recognize the IPs. Furthermore, the different characteristics present in each food category cause the different duration of the IPs, which cannot be covered with the conventional SS. For monitoring the IPs application, the SS approach is not suitable as the duration of the patterns varies depending on the bite size and food type. Electroglottograph (EGG) based system was designed to automatically monitor diet and food intake. They achieved reasonably accuracy for predicting eating episodes using static signal segmentation technique [16]. A multi-sensor fusion based system was developed to classify collected activity data between eating and non-eating activities [17]. Camera based objection detection method was used to determine normal eating and stressed eating [18]. The performance of method can be degraded if objects are not aligned to line-of-sight of camera. A wearable diet monitoring system was presented to classify liquid and solid food categories [19]. The system detected normal breathing and swallow containing breathing cycles, which were further associated to liquid and solid intakes. The proposed may falsely consider interfering activities, coughing and talking, as swallow breathing cycle that can reduce the recognition performance. An integrated wearable necklace was applied to collect ingestion patterns which were passed through LSTM network to detect and count the swallows [20]. The method was mainly focused on eating disorder and did not address the problem of food classification. These methods did not attain high accuracy in food classification due to traditional static signal segmentation. Comparing to the static signal segmentation approaches, the dynamic segmentation improves the representation of the data [5].

We have employed the BS to segment the IPs correctly around the two main events: chewing

and swallow. Moreover, we transformed the DSs into the wavelet sequences (i.e., data in TFD) using DWT, which are fed into the LSTM model. DWT converts the ingestion pattern containing DSs into the well-organized sequences. The food recognition model based on the LSTM as shown in Figure 1(b) is trained and evaluated using 15-fold cross-validation technique with leave-one-subject-out. We used leave-one-subject-out validation technique for enabling the model to gain generalization ability to recognize food categories based on ingestion data of participants. The model trained on data of fourteen subjects and left-one-subject-out for validation. The advantage of the leave-one-subject-out strategy is that it provides a rigorous evaluation of the model's performance on the data from previously unseen subjects or users.

Before training the recognition model, the optimal hyperparameters for the LSTM were selected through evaluating the model's performance with various combinations of parameters. The hyperparameters are computational components that can significantly impact the solution achieved by the learning algorithm [21]. The optimal set of hyperparameters was chosen. A stochastic gradient descent with momentum was used as the optimization algorithm since it is consistently faster than other gradient descent methods. The added momentum improved the convergence rate. The optimization algorithm helped the model minimize the loss function by iteratively optimizing the parameters. The initial learning rate was set to 0.01, and it started decreasing every 30 iterations using piece-wise learning rate scheduling [22]. A minimum batch size of 30 was set for each training iteration, and the training was limited to a maximum of 40 epochs. Batch normalization was conducted before training to avoid an internal covariate shift problem, which is a change in the network parameters transforming the distribution of the network.

The number of training and test instances are 5850 and 1950, respectively. The LSTM automatically extracts the efficient features and accurately recognizes the IPs of 5 food classes. Our proposed food recognition model

attains 98.1% accuracy and outperforms as compared to previous studies as detailed in Table. Our proposed approach achieves 18% and 7% higher accuracy, as compared to static segmentation handcrafted features approach [3] and [4], respectively.

**Conclusion and Future Work:** In this paper, we presented a new food recognition system based on wavelet-LSTM for recognizing the ingestion patterns of 5 food classes. DWT converts the dynamic segments into the sequences, which are fed into the LSTM model. We emphasized dynamic segmentation as it is essential to cover the chewing and swallow events of varying durations. The LSTM attains high recognition performance for 5 food classes by automatically extracting the efficient features from the wavelet sequences. The proposed system attains state-of-the-art performance by recognizing the 5 food classes with 98.1% accuracy. Our future work aims to expand the number of food classes and incorporate interfering activities such as coughing and talking to develop a more robust food recognition system. This will offer obese individuals more options to monitor their dietary intake effectively.

#### References:

- [1] Shaikh, Mohsin, et al. "Open-source electronic health record systems: A systematic review of most recent advances." *Health Informatics Journal* 28.2 (2022): 14604582221099828.
- [2] Vu, Tri, et al. "Wearable food intake monitoring technologies: A comprehensive review." *Computers* 6.1 (2017): 4.
- [3] Hussain, G., et al. 'Food intake detection and classification using a necklace-type piezoelectric wearable sensor system,' *IEICE TRANS. on Inform. and Syst.*, 2018, 101, (11), pp. 2795–2807.
- [4] Alshurafa, N., et al. 'Recognition of nutrition intake using time-frequency decomposition in a wearable necklace using a piezoelectric sensor,' *IEEE Sensors Journal*, 2015, 15, (7), pp. 3909–3916.
- [5] Alonso-Arévalo MA, Cruz-Gutiérrez A, Ibarra-Hernández RF, García-Canseco E,



- Conte-Galván R. Robust heart sound segmentation based on spectral change detection and genetic algorithms. *Biomedical Signal Processing and Control*. 2021 Jan 1;63:102208.
- [6] Hussain G, Javed K, Cho J, Yi J. Food intake detection and classification using a necklace-type piezoelectric wearable sensor system. *IEICE TRANSACTIONS on Information and Systems*. 2018 Nov 1;101(11):2795-807.
- [7] Cho HJ, Hussain G, Park JH, Kim JH, Cho JD. Visual fatigue measurement model based on multi-area variance in a stereoscopy. In *2016 18th International Conference on Advanced Communication Technology (ICACT) 2016 Jan 31* (pp. 814-817). IEEE.
- [8] West N, Roy T, O'Shea T. Wideband signal localization with spectral segmentation. *arXiv preprint arXiv:2110.00583*. 2021 Oct 1.
- [9] Al-Rawi M, Materka A. 'Voiced/Unvoiced Speech Signal Segmentation Using Wavelet Analysis.' *Computational Intelligence and Applications*. 23:85, (1999).
- [10] Fearnhead P. Exact Bayesian curve fitting and signal segmentation. *IEEE Transactions on Signal Processing*, 53(6):2160-6, (2005).
- [11] Saadatnejad, S., et al. 'LSTM-Based ECG Classification for Continuous Monitoring on Personal Wearable Devices,' *IEEE jour. of bio. and health informatics*, (2019).
- [12] Hussain, G., et al. 'Indoor Positioning System: A New Approach Based on LSTM and Two Stage Activity Classification,' *Electronics*, 8, (4), pp. 375, 2019.
- [13] Shaikh, M., Lee, K., and Lee, C. 'Assessing the Bug-Prediction with Re-Usability Based Package Organization for Object Oriented Software Systems.' *IEICE TRANSACTIONS on Information and Systems* 100.1, 107-117, (2017).
- [14] Shaikh, M. and Lee, C. 'Aspect Oriented Re-engineering of Legacy Software Using Cross-Cutting Concern Characterization and Significant Code Smells Detection.' *International Journal of Software Engineering and Knowledge Engineering* 26.03, 513-536, (2016).
- [15] Hochreiter, S. and Schmidhuber, J. 'Long short-term memory.' *Neural computation* 9.8, 1735-1780, (1997).
- [16] Farooq, Muhammad, Juan M. Fontana, and Edward Sazonov. "A novel approach for food intake detection using electroglottography." *Physiological measurement* 35.5 (2014): 739.
- [17] Farooq, Muhammad, and Edward Sazonov. "A novel wearable device for food intake and physical activity recognition." *Sensors* 16.7 (2016): 1067.
- [18] Rachakonda, Laavanya, Saraju P. Mohanty, and Elias Kougiianos. "iLog: An intelligent device for automatic food intake monitoring and stress detection in the IoMT." *IEEE transactions on consumer electronics* 66.2 (2020): 115-124.
- [19] B. Dong and S. Biswas, "Wearable diet monitoring through breathing signal analysis," *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Osaka, Japan, 2013*, pp. 1186-1189, doi: 10.1109/EMBC.2013.6609718.
- [20] Nguyen, Dzung Tri, et al. "SwallowNet: Recurrent neural network detects and characterizes eating patterns." *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. IEEE, 2017.
- [21] Greff, K.; Srivastava, R.K.; Koutnik, J.; Steunebrink, B.R.; Schmidhuber, J. LSTM: A search space odyssey. *IEEE Trans. Neural Netw. Learn. Syst.* 2017, 28, 2222–2232.
- [22] Yang, Jun, and Fei Wang. "Auto-ensemble: An adaptive learning rate scheduling based deep learning model ensembling.", *IEEE Access* 8 (2020): 217499-217509.