International Journal for Housing Science and Its Applications
Publish August 4, 2025. Volume 46, Issue 3 Pages 2558-2567

https://doi.org/10.70517/ijhsa463214

Random Forest Algorithm-Based User Emotion Recognition Model Construction for Housing Information Systems

Gaofeng Huang[1,*] and Xiangjun Xu[1]
[1] School of Electronic and Information Engineering, Wuhan Donghu University, Wuhan, Hubei, 430212, China
Corresponding authors: (e-mail: hgf001x@163.com).

Abstract The housing information system involves personal sensitive data, and the user's emotional state is crucial to the system interaction experience. This study constructs a housing information system user emotion recognition model based on the random forest algorithm, and explores the problems of EEG signal feature extraction and emotion classification. The study adopts the DEAP dataset and performs feature dimensionality reduction by two types of feature extraction methods, namely power spectral density and differential entropy, combined with Savitzky-Golay feature smoothing and minimum redundancy maximum correlation algorithms. The experiments set different emotion label thresholds on two emotion dimensions, arousal and validity, and compared the emotion recognition effects of decision tree and random forest algorithms. The results show that the classification accuracy of the random forest algorithm in the arousal and valence dimensions reaches 92.2% and 91.0%, respectively, which is much higher than that of the multilayer perceptual machine and close to the discrete emotion model. Compared with the three classifiers, LSTM, KNN and LR, the Random Forest classifier has an average training time of only 522 ms and a test prediction time of 29.6 ms, which is the best overall performance. The study confirms that the random forest algorithm is computationally efficient and resistant to overfitting when processing high-dimensional EEG signal data, and provides feasible technical support for emotion recognition for the optimization of user experience in housing information systems.

Index Terms random forest algorithm, emotion recognition, EEG signal, power spectral density, feature extraction, housing information system

I. Introduction

Accompanied by the development of the economy and the improvement of salary level, the purchase of housing has become the demand of more and more people, and the housing information system as a modern real estate enterprise's basic management information system, is the basic basis of the real estate enterprise's management of the property and the owner [1]-[4]. The application of housing information system is an important part of perfecting enterprise information management [5], [6]. With the gradual increase in the number of housing users, relying only on manual has been unable to meet the complex and cumbersome management requirements, in the face of a huge amount of information, housing information system greatly reduces the workload of the staff, improves the work efficiency, so that the original complex and boring work has become simple and easy, and to a certain extent reduces the management costs [7]-[10].

In the housing information system, if we want to better provide better service for users, it is indispensable to recognize users' emotions [11], [12]. Emotion recognition refers to judging the user's current emotional state by recognizing data such as voice, facial expression, and heart rate [13], [14]. In housing information systems, emotion recognition can provide the system with more accurate information about users' needs, thus optimizing the user experience [15]. In addition, emotion recognition can also help the system handle some emergencies, such as users experiencing low mood, and provide timely help and support [16], [17]. However, to realize the application of emotion recognition in housing information systems, technology is a challenge that must be overcome [18], [19]. As an integrated learning method, the application of random forest algorithm in emotion recognition has attracted much attention [20]. In the housing information system, the advantage of constructing a user emotion recognition model based on the random forest algorithm is that it can integrate multiple weak classifiers and get the results of user emotion recognition by voting, which is of great significance for improving the quality of user service [21]-[24].

As an important research direction in the field of human-computer interaction, emotion recognition technology is of great significance for improving the user experience of information systems. In application scenarios involving personal sensitive data, such as housing information systems, accurate recognition of the user's emotional state can help the system better understand the user's needs, provide personalized services, and at the same time reduce the negative emotions generated during the user's operation. Currently, emotion recognition technology is mainly based on multimodal data such as facial expressions, speech, body movements and physiological signals for analysis. Among them, physiological signals represented by electroencephalogram (EEG) have received extensive attention in emotion recognition research due to their strong objectivity and difficulty in disguising. However, EEG signals are characterized by high dimensionality, nonlinearity, and nonsmoothness, and traditional emotion recognition methods often face problems such as low feature extraction efficiency and low classification accuracy when dealing with such data. How to extract features highly related to emotional states from complex EEG signals and construct efficient emotion classification models has become the focus and difficulty of current research. Machine learning algorithms have shown better performance in EEG signal emotion recognition, especially decision trees, support vector machines, neural networks and other methods are widely used. In recent years, integrated learning methods have gradually become a research hotspot in the field of emotion recognition due to their strong generalization ability in dealing with high-dimensional data. In particular, the random forest algorithm, as an integrated learning method based on multiple decision trees, can effectively avoid the overfitting problem of a single decision tree and improve the classification accuracy by constructing multiple classifiers and integrating their results, while it has the advantages of high computational efficiency and insensitivity to outliers, which provides a new technological path for EEG signal emotion recognition. In this study, we propose an emotion recognition model construction method based on random forest algorithm for the problem of emotion recognition of users in housing information system. Firstly, the DEAP dataset is preprocessed, and the EEG signals are segmented using the Hamming window without overlapping windows; secondly, two types of features, power spectral density and differential entropy, are extracted from the EEG signals, and feature smoothing is carried out by the Savitzky-Golay method, and feature dimensionality reduction is carried out by combining with the minimum redundancy maximum correlation algorithm in order to increase the representativeness of the features and reduce the redundancy of the data; and then the arousal and validity two emotion dimensions, different emotion label thresholds are set to compare the emotion recognition effect of decision tree and random forest algorithms; finally, the effectiveness and superiority of the proposed method is verified through comparison experiments with commonly used classification algorithms, such as LSTM, KNN, and LR. This study aims to provide theoretical support and technical guidance for user emotion recognition in housing information system through algorithm optimization and model construction, and then promote the intelligent and humanized development of housing information system.

II. Housing information system

II. A.Needs analysis

Individual housing information, housing fund information and housing security information are all important data and information related to the national economy and people's livelihood, and their main business requirements for data have six aspects.1) Security. Individual housing information, housing provident fund information and housing security information all involve sensitive information of public interest and have certain significance to social stability, therefore, it is the core demand to ensure the safe survival of related data.2) Real integrity. The data must completely cover the main business content of business production, including business object status data, business object historical data, business rule management data and business process record data to ensure the authenticity and integrity of the data.3) Dynamic real-time. The data must be synchronized with the production system's data update record segment generation process in real time to ensure the freshness of data resources.4) Breakpoint continuity. Network and other resources interruption recovery, the data must be from the breakpoint smooth succession, to ensure that the data is not due to the network and other resources interruption data crippling phenomenon. 5) Independence. Housing information system should be relatively independent of the production system, the housing information system must not affect the normal operation of the production system. 6) Application scope. The data should support the business application of the housing information system with the main content of housing information query and analysis, housing fund risk supervision, and housing security resource allocation supervision. The system should support the local housing management department, provident fund management department, housing security department and local finance, taxation, finance, public security and other departments to query the national housing ownership information (including mutual housing and security housing information) business needs, to support the provident fund regulatory department to record the provident fund centers and the bank on-line transaction information, real-time control of housing provident fund funds arbitrarily transferred to the deposits and embezzlement of the case occurs.

II. B.System Programming

II. B. 1) Data acquisition program design

According to the requirements of real integrity and dynamic real-time data, the data collected by the housing information system must be obtained directly from the production systems of local housing management departments, provident fund management departments and housing security departments in real time. The current mainstream feasible technical solutions are: data reporting technology, data mirroring technology and data pumping technology.

1) Data reporting technology. It is in the production business system in the business logic module or data access module to increase the business data change trigger module, when the business data changes, the changes in business data will be pushed to report. The design of the data collection program is shown in Figure 1.

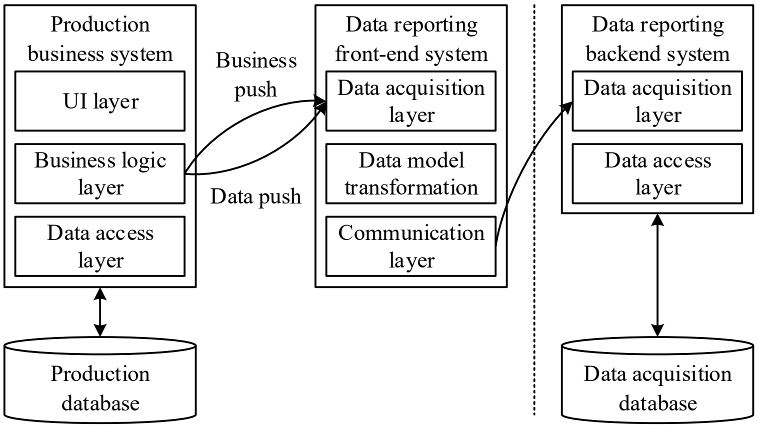

Figure 1: Schematic diagram of the principle of data reporting technology

2) Data mirroring technology. It is the preferred technology for remote data mirroring disaster recovery by using the log analysis results of the source database to remotely realize data replication.

3) Data pump technology. It is a technology that realizes data query and extraction by using the data pump software distributed in the production system.

II. B. 2) Database program design

Overall database solution.1) Adopting cheap X86 computer and local storage as the carrier to carry the database system and its data.2) Based on the standard relational database system (DBMS), the data of each city is stored and managed independently to realize the simple correspondence and traceability of the data of the production system.3) Absorbing the storage management concepts of Hadoop, with the technology of multi-living body data storage to Ensure the reliability of data and provide high-performance data query and analysis performance. 4) Referring to the basic concept of database management of distributed database, independently develop the phase query and analysis agent component to realize the metadata management and query analysis function.

II. B. 3) Query analysis program design

Considering that personal housing information, housing fund information and housing security information data involves sensitive personal information, the main design premise of data querying is: 1) Respecting information sovereignty. Any query request for independent datasets requires specific authorization from the information sovereignty authority.2) Support unified query. Establish a unified navigation meta-database to manage the node distribution of data slices and query routing, shield the organizational structure and semantic representation differences between independent data sets, and establish query request distribution routing through query navigation.3)Support local query. It is necessary to support the completion of nationwide query of relevant data on each city node.4) Support data calibration. Each independent dataset is synchronized with the local production system in real time, due to the lack of data cleansing, there may be many data missing error phenomenon.

III. Random forest algorithm-based emotion recognition methodology

III. A. Random Forest Algorithm

Random Forest is an integrated learning method based on decision trees and is a very powerful machine learning technique with high prediction accuracy and robustness. The core idea of the random forest algorithm is to perform classification and regression through multiple decision trees. In the random forest algorithm, each decision tree is constructed on a randomly selected subset of features, which avoids the overfitting problem of the decision tree algorithm. The predictions of each decision tree are voted and the final prediction is the average of the predictions of each decision tree. Compared with other machine learning classification algorithms, the Random Forest algorithm utilizes the classification results of multiple decision trees, which can deal with high-dimensional data and unbalanced data, eliminate the influence of noise in the data, and in this way, improve the accuracy of classification and ensure the accuracy of classification [25].

The algorithmic flow of Random Forest is shown below:

First Bootstrap performs resampling to produce T training sets $S_1, S_2, ..., S_T$. Let there be n different samples in the set S $\{x_1, x_2, ..., x_n\}$, if a sample is drawn from the set S each time there is a putback, a total of n times, to form a new set S^*, the probability that the set S^* does not contain a certain sample $x_i (i = 1, 2, \ldots, n)$ within $p = \left(1 - \frac{1}{n}\right)^n$, when $n \to \infty$, there is $\lim\limits_{n\to\infty} p = \lim\limits_{n\to\infty}\left(1 - \frac{1}{n}\right)^n = e^{-1} \approx 0.368$. At this point the new set S^* contains about 63-2% of the samples in the original set. Second, for each training set in T^*, generate the corresponding decision trees $C_1, C_2, ..., C_T$. On this basis, each decision tree is tested using a test set of samples X to obtain the corresponding categories $C_1(X), C_2(X), ..., C_T(X)$. Finally, the output categories in the T decision trees are counted, and the category with the most number of categories is used as the category to which the test set sample X belongs.

III. B. Emotion Recognition Process

The flow of the algorithm based on multi-feature deep forest is shown in Fig. 2. Firstly, the original data is preprocessed, including sentiment labeling processing and band segmentation. Second, the sentiment calculation is performed on the data to obtain two types of features: power spectrum (PSD) and differential entropy (DE). Third, data smoothing and dimensionality reduction are performed on the features. Finally, the raw data are converted into feature vectors and the feature vectors are input into the deep forest for sentiment recognition and classification.

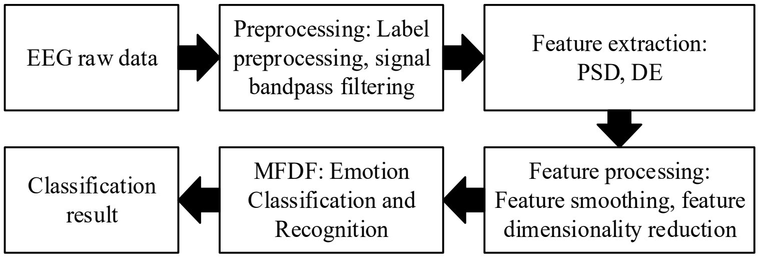
EEG raw data
Preprocessing: Label preprocessing, signal bandpass filtering
Feature extraction: PSD, DE
Feature processing: Feature smoothing, feature dimensionality reduction
MFDF: Emotion Classification and Recognition
Classification result

Figure 2: Flowchart of the algorithm based on multi-feature deep forest

III. B. 1) DEAP data set

The DEAP dataset is a publicly available multimodal dataset that includes EEG signals and is used to detect human emotional states.The DEAP dataset consists of two parts.The first part contains online self-assessments of arousal, potency, and dominance based on 120 one-minute videos from 14 to 16 subjects. The second part contains ratings of arousal, potency, and dominance by an experimental subject, in which a portion of the video of the subject's facial expression was also captured. Each of the 32 subjects watched 40 music videos, and for each experiment 8 other physiological signals and 32 channels of EEG signals were recorded synchronously, and the videos were rated according to the procedure described above.

III. B. 2) Data pre-processing

For multi-feature extraction the hyperparameter is the size of the window, because the DEAP dataset represents 1s for every 128 samples, we set the window size to 0.5-3s, after experimentation we have to the most suitable window size is 1s. For the 32×8064 size samples, the Hamming window with no overlap of the window of 1s is used for scanning the process.

III. B. 3) Data characterization

In the feature extraction stage, this paper extracts features such as PSD and DE from EEG signals for analysis. According to previous studies, it has been found that PSD estimation using Welch's method provides very strong features and it is also a good feature representation of EEG signals. It has been demonstrated that DE features are suitable for processing EEG signals and are the most accurate and stable EEG features reflecting changes in human alertness.

(1) Power spectral density (PSD) feature extraction

It is known from previous studies that the power spectral density of Welch's method is a method that can provide very strong data features, and it can also be used as a stable feature of EEG emotional signals. In this paper, the average value of power spectral density is used as a feature of EEG emotion data [26]. The power spectral density is defined in equation (1):

$$P_i(f) = \frac{1}{L}\left|\sum_{n=1}^{L-1} x_i[n]e^{j2\pi fn}\right|^2 \tag{1}$$

where L is representing the length of the EEG emotional signal and $P_i(f)$ is the fast Fourier transform of the EEG emotional signal $x_i[n]$. Then, the PSD feature $P(x_i)$ of the EEG emotional signal $x_i[n]$ is defined as in equation (2):

$$P(x_i) = \frac{1}{K}\sum_{i=1}^{K} P_i(f) \tag{2}$$

where the value of K denotes the number of frequency points of the EEG signal when doing the discrete Fourier transform. In calculating the power spectral density of the EEG emotional signal, this paper uses the PSD features of the signal extracted using a Hanning window with a non-overlapping window size of 1s. Where the peak, mean, variance and other features of the PSD can be used as the features of the original data, this paper calculates the PSD averaged over the signals in each window, thus obtaining the PSD features of the EEG emotional signal in that time period.

(2) Differential entropy (DE) feature extraction

It is known from previous studies that differential entropy is a very stable and highly accurate and stable EEG signal feature that can EEG emotional signal, where differential entropy is defined as in equation (3):

$$h(x_i[n]) = \frac{1}{2}\log 2\pi e\sigma^2 \tag{3}$$

where the distribution of the signal $x_i[n]$ conforms to a Gaussian distribution as $N(\mu,\sigma^2)$. In this paper, the Hanning window with no overlapping window size of 1s is also used for the extraction of DE features of EEG emotional signals. By calculating the variance σ of the EEG emotion signals within each window, the DE features of the EEG emotion signals within that time period can be calculated [27].

III. B. 4) Feature dimensionality reduction

When extracting features from a signal, the extracted features may not be relevant to the emotional state, which may lead to unsatisfactory performance of the classifier. In order to reduce the feature noise, avoid the occurrence of “dimensionality disaster”, and improve the performance of the classifier. In this paper, we firstly perform feature smoothing based on the Savitzky-Golay method, with the parameters set to span 5 and order 3. In addition, we ensure low redundancy and high correlation of the data through the data dimensionality reduction method. The basic principle of Minimum Redundancy Maximum Relevance (MRMR) algorithm is to use the mutual information between the data as a correlation measure, while the maximum dependency Huai rule ensures the relevance of the data and the class labels, while the redundancy between the data is measured by the minimum redundancy criterion [28]. The rule of maximum correlation criterion is to calculate the average value of mutual information between a single feature vector x_i and a class c in the data, and the maximum correlation criterion is defined in equation (4):

$$\max D(S,c), D = \frac{1}{|S|^2}\sum_{x_i\in S} I(x_i;c) \tag{4}$$

When two feature vectors in the data are equally dependent on the same class, the correlation between the two features and the class is considered to be strong, and deleting any one of the features will not affect the class differentiation ability of the total feature vector pair. Minimum redundancy criterion is used to reduce the redundancy between the data and the minimum redundancy criterion is defined as in equation (5):

$$\min R(S), R = \frac{1}{|S|^2}\sum_{x_i,x_j\in S} I(x_i;x_j) \tag{5}$$

The two constraint criteria in Eq. (4) and Eq. (5) are known as the “Minimum Redundancy Maximum Relevance” (MRMR) criterion. If the two criteria are put together, the operator $\varphi(D,R)$ can be defined to combine D and R, and the simplest definition of the MRMR can be expressed as Eq. (6):

$$\max \varphi(D,R), \varphi = D - R \tag{6}$$

IV. Experimental results and analysis

IV. A. Indicators for model evaluation

The commonly used model evaluation metrics for binary classification problems are classification accuracy (Acc), precision (P) recall (R) and Fl values and confusion matrix, with classification accuracy defined as:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

The precision rate is defined as:

$$P = \frac{TP}{TP + FP} \tag{8}$$

Recall is defined as:

$$R = \frac{TP}{TP + FN} \tag{9}$$

$F1$ value is defined as:

$$F1 = \frac{2 \times P \times R}{P + R} \tag{10}$$

where: TP denotes the number of positive classes predicted as positive classes, i.e., true classes; FN denotes the number of positive classes predicted as negative classes, i.e., false negative classes; FP denotes the number of negative classes predicted as positive classes, i.e., false positive classes; TN denotes the number of negative classes predicted as negative classes, i.e., true negative classes. In this paper, samples labeled with high arousal and high valence of emotions are recorded as positive class samples, and samples with low arousal and low valence are recorded as negative class samples.

IV. B. Analysis of results

IV. B. 1) Effect of Sentiment Label Threshold on Classification Results

In this paper, the emotion labeling thresholds are divided into two major categories on the two emotion dimensions of arousal and validity. By selecting different emotion label thresholds to study its impact on the emotion recognition results, the results of the model evaluation indexes are shown in Table 1.The results in the table show that in the arousal dimension, the classification accuracy and F1 value of utilizing the Decision Tree algorithm with the Class II threshold setting are 89.4% and 0.911 respectively, while the results of Class I threshold are 81.2% and 0.859.The classification accuracy and F1 value of utilizing the Random Forest algorithm with the Class II threshold setting the classification accuracy and F1 value reached 92.2% and 0.947, respectively, while the results for the Type I threshold were 87.6% and 0.907.The same is true for the validity dimension. It can be seen that when using decision tree or random forest algorithms to classify emotions in terms of arousal and validity, the classification results of emotion recognition with the class II threshold setting are better than those with the class I threshold in terms of classification accuracy, ACC, precision rate, P, recall rate, R and F1 value. The reason mainly lies in the fact that the number of samples for model training for the class II threshold setting, both in arousal and validity dimensions, is more than that for the class I threshold setting, which does not result in underfitting, and can fully learn the emotionally rich features, thus facilitating the recognition of emotions. Therefore, when performing emotion labeling threshold division, not only need to consider the intensity of emotion expression in arousal and validity, but also need to take into account the impact of the number of samples on emotion classification. This suggests that choosing the appropriate emotion threshold and the number of training samples is conducive to improving the emotion recognition accuracy.

Table 1: The classification of different emotional tag thresholds

Affective dimension	Algorithm	Categories	Model evaluation index			
			ACC/%	P	R	F1
Arousal	DT	Ⅰ	81.2	0.844	0.863	0.859
		Ⅱ	89.4	0.922	0.889	0.911
	RF	Ⅰ	87.6	0.877	0.946	0.907
		Ⅱ	92.2	0.918	0.964	0.947
Valence	DT	Ⅰ	78.4	0.802	0.846	0.826
		Ⅱ	86.9	0.897	0.897	0.894
	RF	Ⅰ	86.2	0.863	0.916	0.893
		Ⅱ	91.1	0.917	0.945	0.928

IV. B. 2) Effect of classification algorithms on emotion recognition results

In this paper, the decision tree and random forest algorithms were used for emotion recognition on arousal and potency respectively, and the results in the above table show that the results of emotion classification using the random forest algorithm on arousal and potency are better than the decision tree algorithm, both in terms of classification accuracy, ACC, and precision rate, P, recall rate, R and F1 value. It is thus shown that in emotion recognition of skin electrical signals, Random Forest's combined classifier based on integrated learning can improve the classification accuracy compared to a single decision tree classifier, and effectively reduce the overfitting phenomenon and improve the generalization ability of the model, and at the same time the algorithm has the characteristics of fast speed, low time overhead, and good robustness.
The time domain statistical features, power spectral density maximum and wavelet packet entropy features of the skin electrical signals are extracted from this paper, and the decision tree and random forest algorithms are used to classify the emotions in terms of arousal and validity, and compared with other machine learning methods, better classification results are achieved, and the results are shown in Table 2. For the dimensional sentiment model, the classification results using the decision tree and random forest algorithm are better than the classification accuracy achieved using the multilayer perceptual machine in the two sentiment dimensions of arousal and validity. Among them, the classification accuracy of emotion using Random Forest algorithm is higher than that of emotion recognition using probabilistic neural network in terms of arousal and validity. Comparing the dimensional emotion model used in this paper with the emotion recognition based on the discrete emotion model, the classification accuracy of the Random Forest algorithm in the two emotion dimensions of arousal and validity reaches 92.2% and 91.0%, respectively, which is close to the recognition accuracy based on the discrete emotion model, and is much higher than the recognition accuracy of the dimensional emotion model. It can be seen that better emotion recognition results and performance can be obtained by extracting the time-domain, frequency-domain, and time-frequency nonlinear features of the skin electrical signal in the dimensional model for emotion recognition, and using the tree model algorithm of random forest integrated learning.

Table 2: Classification accuracy of different classification methods

		DT%	Our Method RF%	MLP%	PNN%	SVM%
Dimensional arousal	Arousal	88.9	92.2	80.4	88.9	-
Emotional model	Valence	87.2	91.0	70.1	87.3	-
Discrete emotion model	-	-	-	-	-	92.5

IV. B. 3) Comparison of model performance

In order to further validate the performance of the RF method, in terms of sample division, 1090 samples were randomly sampled according to the sentiment category dimensions stratified by 70% of the samples as the training set and 30% of the samples as the unknown test set. In this way, the model performance comparison of the four classifiers, RF, LSTM, KNN and LR, is carried out. The comparison results are shown in Table 3.

Among them, LSTM is an improved model of recurrent neural network, which is able to apply previous information in the process of processing the current information, therefore, compared with the general neural network, it is able to solve the problem of gradient vanishing and gradient explosion in the process of training long sequences well, and it is widely used in the processing of sequence information.

KNN is a supervised classification algorithm and a theoretically mature method, which can classify new unknown categories of data into a specific class. The idea of the method is:If most of the K most similar (i.e., closest neighbors in the feature space) samples in the feature space of a sample belong to a certain class, then the sample also belongs to this class.

LR is a well-known classification model in the field of machine learning, and its commonly used to solve binary classification problems. The principle of logistic regression for multiple classifications is to partition multiple classifications of the dependent variable into multiple binary logistic regressions sequentially, thus realizing the classification of data with unknown categories.

Table 3: The parameters of the four classifiers are constructed

Classifier	Build parameter details
RF	The decision tree in the forest is :20
	Random number seed: 50
	LTSM layer: output: 128. activation function: relu"
	Dropout bed
LSTM	Lstm_1 layer: output: 68, activation function: sigmoid"
	Dropout bed
	Layer of space: 3. Activation function: sigmoid"
LR	Optimization algorithm selection: saga'
	The maximum iteration number: :1200
KNN	K = 6

In the process of model performance analysis, we adjusted each model parameter as much as possible to make all four classifiers achieve the best performance in EEG emotion recognition, which ensures that the experiments are more comparable. The results are shown in Figure 3, which shows the detailed information of the parameters when building the four classifiers.

In the process of debugging the RF parameters, it is found that no matter for all classes or a separate class of emotions, when the model complexity is insufficient, the machine learning is insufficient, and the underfitting phenomenon occurs, and the generalization error becomes larger; when the complexity is gradually increased to the optimal model complexity, the generalization error reaches the minimum (i.e., the highest accuracy); if the model complexity is still being increased, the generalization error starts from the minimum value and gradually increases, and over overfitting phenomenon. Finally, to ensure the best performance of the model, the number of decision trees in the random forest is set to 20.

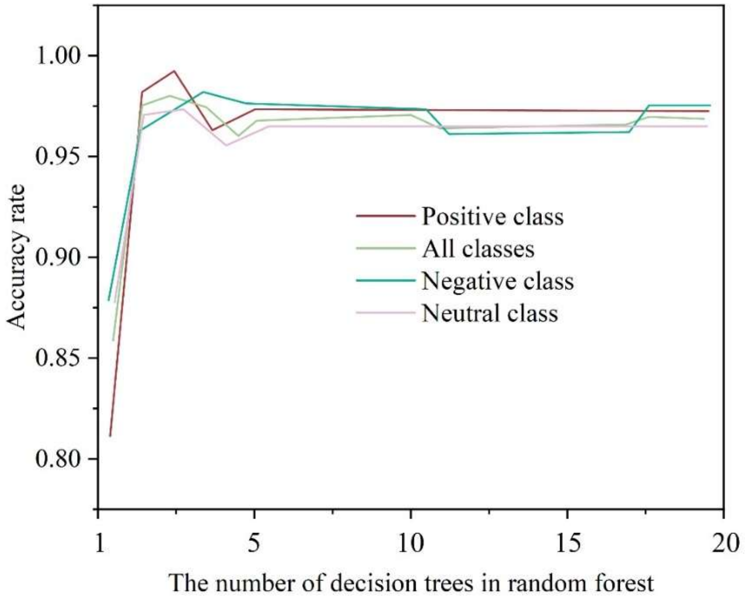

Figure 3: The emotional classification accuracy varies with the decision tree

Taken together the Random Forest classifier has a short training time and faster prediction in the test set. In particular, the overall classification accuracy of the LTSM model is very close to that of the random forest, but its training and prediction time consuming are much higher than that of the random forest method, which is based on the ability of the random forest feature selection, which does not need to normalize the dataset, and is better able to deal with multichannel, high-dimensional data such as EEG. The efficiency of emotion recognition is improved, and this property facilitates the widespread placement of the model in commercial EEG headbands. Therefore, the comprehensive performance of the random forest classifier is better than the other three classifiers, and it is a better-fitting model with certain value for engineering applications. The experimental results are shown in Fig. 4, indicating that the accuracy of emotion recognition using random forest is higher than the other three models.

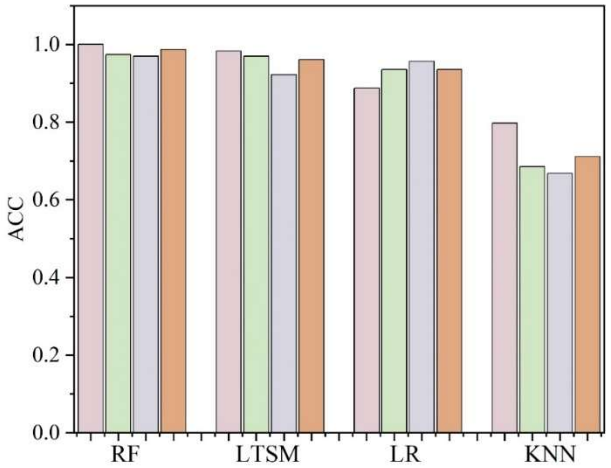

Figure 4: The four categories of classifiers are accurately compared

The average time for training and prediction for the four classifications is shown in Table 4, and taken together the Random Forest classifier has a short training time and faster prediction in the test set. In particular, the overall classification accuracy of the LTSM model is very close to that of the random forest, but its training and prediction time consuming are much higher than that of the random forest method, which is based on the ability of the random forest feature selection, which does not need to normalize the dataset, and is better able to deal with multichannel, high-dimensional data such as EEG. The efficiency of emotion recognition is improved, and this property facilitates the widespread placement of the model in commercial EEG headbands. Therefore, the Random Forest classifier has better overall performance than the other three classifiers and is a better-fitting model with some engineering applications.

Table 4: Four kinds of classification training and forecasting average time

Classifier	Training takes a while (ms)	It takes a time to predict (ms)
RF	522±10.4	29.6±1.52
LTSM	60000±136	83.7±1.22
LR	7342±105	2.0±0.09
KNN	160±8.34	684±10.3

V. Conclusion

Random forest algorithm-based user emotion recognition model for housing information system shows significant advantages in EEG signal processing and emotion classification. The comparative analysis of different emotion label thresholds reveals that the classification accuracy and F1 value of the Random Forest algorithm in arousal and validity dimensions under the Class II threshold setting are 92.2%, 0.947 and 91.1%, 0.928, respectively, which are superior to the results of the Class I threshold setting, indicating that a reasonable selection of emotion label thresholds has a significant impact on the accuracy of emotion recognition. Algorithm comparison experiments confirm that the Random Forest algorithm as a combined classifier with integrated learning has a significant improvement in emotion recognition accuracy over a single decision tree classifier, and effectively reduces the overfitting phenomenon. Compared with the three classifiers of LSTM, KNN and LR, the random forest classifier has obvious advantages in terms of training time consuming (522 ms on average) and prediction speed (29.6 ms on average), especially when dealing with multi-channel and high-dimensional EEG data without the need of dataset normalization, which greatly improves the efficiency of emotion recognition. The power spectral density and differential entropy features combined with Savitzky-Golay feature smoothing and minimum redundancy maximum correlation algorithms form a feature extraction and dimensionality reduction framework, which is important for extracting the features in EEG signals that are highly correlated with emotional states. The housing information system can realize real-time monitoring and feedback of users' emotional state with the help of this emotion recognition model, which provides data support for the optimization of system interaction experience and personalized service.

Funding

This work was supported by Science and Technology Research Project of Hubei Provincial Department of Education (No.B2023292).

References

[1] Hassan, M. M., Ahmad, N., & Hashim, A. H. (2021). The conceptual framework of housing purchase decision-making process. International Journal of Academic Research in Business and Social Sciences, 11(11), 1673-1690.

[2] Muczyński, A., Dawidowicz, A., & Źróbek, R. (2019). The information system for social housing management as a part of the land administration system–A case study of Poland. Land use policy, 86, 165-176.

[3] Abualloush, S., Bataineh, K., & Aladwan, A. S. (2017). Impact of information systems on innovation (product innovation, process innovation)-field study on the housing bank in Jordon. International Journal of Business Administration, 8(1), 95-105.

[4] Pletnev, D., Fink, O., & Dyachenko, O. (2020, February). State Information System of Housing. In Eurasian Business Perspectives: Proceedings of the 25th Eurasia Business and Economics Society Conference (Vol. 12, p. 127). Springer Nature.

[5] Fedorova, I. Y., Urunov, A. A., Rodina, I. B., & Ostapenko, V. A. (2020). Financing and quality of housing construction: introduction of information systems as a regulatory tool. Revista Inclusiones, 7(S2-1), 328-339.

[6] Ermilova, M., & Laptev, S. (2021). Information technology as a tool to improve the efficiency of the housing market. In E3S Web of Conferences (Vol. 234, p. 00047). EDP Sciences.

[7] Shamsuddin, S., & Srinivasan, S. (2021). Just smart or just and smart cities? Assessing the literature on housing and information and communication technology. Housing Policy Debate, 31(1), 127-150.

[8] Yulia, K. (2015). The development of information technologies in the sphere of housing service and utilities as a factor of national life quality increase. Procedia-Social and Behavioral Sciences, 166, 557-561.

[9] Chernenko, Y., Danchenko, O., Mysnyk, B., Bielova, O., & Adamov, O. (2024, May). Optimizing Housing and Communal Services Management Through Digital Transformation and Integrated Information Systems. In International Scientific-Practical Conference" Information Technology for Education, Science and Technics" (pp. 33-49). Cham: Springer Nature Switzerland.

[10] Bitter, N. A., Roeg, D. P., van Nieuwenhuizen, C., & van Weeghel, J. (2016). Identifying profiles of service users in housing services and exploring their quality of life and care needs. BMC psychiatry, 16, 1-11.

[11] Mahadzir, N. H., Omar, M. F., & Nawi, M. N. M. (2018). A sentiment analysis visualization system for the property industry. Int. J. Technol., 9.

[12] Bardhan, R., Sunikka-Blank, M., & Haque, A. N. (2019). Sentiment analysis as tool for gender mainstreaming in slum rehabilitation housing management in Mumbai, India. Habitat International, 92, 102040.

[13] Mehta, D., Siddiqui, M. F. H., & Javaid, A. Y. (2018). Facial emotion recognition: A survey and real-world user experiences in mixed reality. Sensors, 18(2), 416.

[14] Dzedzickis, A., Kaklauskas, A., & Bucinskas, V. (2020). Human emotion recognition: Review of sensors and methods. Sensors, 20(3), 592.

[15] Lee, C. (2025). Textual data analysis for enhancing housing management. International Journal of Housing Markets and Analysis.

[16] Hossain, M. S., & Muhammad, G. (2017). An emotion recognition system for mobile applications. IEEE Access, 5, 2281-2287.

[17] Renigier-Biłozor, M., Janowski, A., Walacik, M., & Chmielewska, A. (2022). Human emotion recognition in the significance assessment of property attributes. Journal of Housing and the Built Environment, 37(1), 23-56.

[18] Syahputra, R. A., Arifin, R., & Iqbal, M. (2024). Sentiment Analysis on Tabungan Perumahan Rakyat (TAPERA) Program by using Support Vector Machine (SVM). Journal of Applied Informatics and Computing, 8(2), 531-541.

[19] Boughareb, D., Boughareb, R., Guerri, N., & Seridi, H. (2023, September). FindLoc: A Sentiment Analysis-Based Recommender System. In 2023 3rd International Conference on Computing and Information Technology (ICCIT) (pp. 101-105). IEEE.

[20] Guo, R., Guo, H., Wang, L., Chen, M., Yang, D., & Li, B. (2024). Development and application of emotion recognition technology—a systematic literature review. BMC psychology, 12(1), 95.

[21] Noroozi, F., Sapiński, T., Kamińska, D., & Anbarjafari, G. (2017). Vocal-based emotion recognition using random forests and decision tree. International Journal of Speech Technology, 20(2), 239-246.

[22] Chen, L., Su, W., Feng, Y., Wu, M., She, J., & Hirota, K. (2020). Two-layer fuzzy multiple random forest for speech emotion recognition in human-robot interaction. Information Sciences, 509, 150-163.

[23] Pu, X., Fan, K., Chen, X., Ji, L., & Zhou, Z. (2015). Facial expression recognition from image sequences using twofold random forest classifier. Neurocomputing, 168, 1173-1180.

[24] Wang, Y., Li, Y., Song, Y., & Rong, X. (2019). Facial expression recognition based on random forest and convolutional neural network. Information, 10(12), 375.

[25] Bao Tran Trong Gia & Hao Vu. (2025). Positive and Negative Emotions Recognition using Multichannel EEG Analysis with Spectral Entropy and Machine Learning approaches. Journal of Physics: Conference Series,2949(1),012008-012008.

[26] Usman Goni Redwan,Tanha Zaman & Hazzaz Bin Mizan. (2025). Spatio-temporal CNN-BiLSTM dynamic approach to emotion recognition based on EEG signal. Computers in biology and medicine,192(Pt A),110277.

[27] Md Raihan Khan,Airin Akter Tania & Mohiuddin Ahmad. (2025). A comparative study of time–frequency features based spatio-temporal analysis with varying multiscale kernels for emotion recognition from EEG. Biomedical Signal Processing and Control,107,107826-107826.

[28] Zhangfang Hu,Yi Wang & Yuan Yuan. (2025). Fca-ProRes2Net Speaker Recognition Based on Fusion Feature Dimensionality Reduction. IAENG International Journal of Computer Science,52(4).