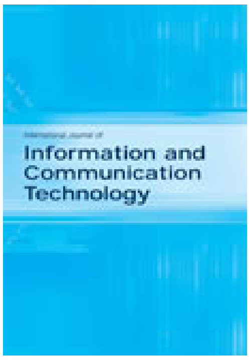


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Analysis of tourist emotions and behaviour patterns using deep learning

Limin Wang

DOI: 10.1504/IJICT.2026.10075959

Article History:

Received:	26 September 2025
Last revised:	12 October 2025
Accepted:	18 October 2025
Published online:	09 February 2026

Analysis of tourist emotions and behaviour patterns using deep learning

Limin Wang

School of Management,
Zhengzhou University of Industrial Technology,
Zhengzhou, Henan, 450000, China
Email: 13674969400@163.com

Abstract: This study explores the emotions and behavioural patterns of travellers using advanced deep learning techniques. A hybrid model combining CNN and LSTM networks was developed to analyse Twitter data related to travel in Thailand, enabling the identification of key emotional states such as joy, surprise, fear and melancholy. The proposed approach provides valuable insights for recommendation systems and tourism management, as tourists' emotions significantly influence travel decisions and satisfaction levels. By analysing emotional tendencies, tourism services can enhance the overall visitor experience. While previous research has largely relied on machine learning and lexicon-based methods for textual emotion detection, recent advancements in deep learning have demonstrated superior predictive accuracy for sentiment and behavioural analysis. In this study, CNN-LSTM models, complemented by feature extraction techniques using DenseNet and AlexNet, were employed. The hybrid model achieved 91% accuracy, surpassing conventional methods, with joy and surprise being the most accurately classified positive emotions.

Keywords: tourist emotions; behaviour patterns; emotion recognition; tourist experience analysis; deep learning; behavioural analytics; artificial intelligence in tourism.

Reference to this paper should be made as follows: Wang, L. (2026) 'Analysis of tourist emotions and behaviour patterns using deep learning', *Int. J. Information and Communication Technology*, Vol. 27, No. 8, pp.53–73.

Biographical notes: Limin Wang is a faculty member in the School of Management at Zhengzhou University of Industrial Technology, Zhengzhou, Henan, China. She is engaged in teaching and research in the field of management studies, with academic interests that include organisational management, business administration, and applied management practices. Her work focuses on integrating theoretical frameworks with practical applications to enhance organisational efficiency and decision-making processes. She has contributed to academic activities through teaching, research guidance, and participation in scholarly discussions within her institution. She is actively involved in academic collaboration and knowledge dissemination in management-related disciplines.

1 Introduction

The method of assessing the emotions expressed in the textual material is known as emotional tendency analysis. Users' feelings, including love, fear, anger, grief, and joy, are identified using emotional tendency analysis. Most often, recommendation systems (RS) for tourism incorporate emotional tendency analysis (Burgui-Burgui, 2022). It is recommended to analyse construction intensity as a metric for the tourist development process. Determining user preferences based on input is the main objective of the strategy. The analytical precision and system applicability range are both enhanced by the tendency analysis method. The tourism department employs a multifaceted approach to visitor pleasure (Liu and Tsao, 2023). The primary purpose of the tourist satisfaction approach is to assess the preferences and behavioural patterns of tourists. The tourism systems' overall viability and significance range are enhanced by the satisfaction technique. By precisely measuring the clients' emotional inclinations, the satisfaction approach lowers computing delay. Analysis of emotional tendencies based on tourist preferences is frequently employed to enhance the effectiveness and range of satisfaction of tourism systems (Liu and Wang, 2022). Determining visitor preferences is an essential endeavour that yields pertinent data for further improvement procedures.

RS for tourism applications employ an emotional tendency analysis method based on tourist preferences. This RS model takes the clients' personalities into account and adjusts its feedback accordingly. It serves as a model that employs diverse groups to determine the consumers' interests and preferences. The content reveals both temporal and geographical aspects of the emotions (Meng and Ji, 2024). The features reduce the computing complexity in RS by providing the precise personalities of the clients. Scholars and practitioners of tourism destination services concur that a significant emphasis on the customer experience may give travel companies and destinations a distinctive and long-lasting edge. Mature locations might not be able to set themselves apart from comparable vacation options, and the negative effects of the large volume of tourists they receive could have an impact on the quality of the customer service experience (Aakash, 2021). Online evaluations influence travellers' decisions in this situation since they provide a wealth of information about many facets of the service under examination. The main offerings of established tourist locations are attractions (Bigne, 2021).

Attractions are contextual and complex, in contrast to hotels, which have distinct features. Tourists invest both time and money into visiting tourist attractions and making use of them and all the services that come with them, making the customer experience all the more important. Review star ratings are the numerical values that reviewers use to indicate their overall satisfaction with a product or service in an online review (Craciun, 2020). In order to save cognitive effort and search costs, readers frequently employ star ratings as a crucial heuristic to cut down their consideration set (Bigne et al., 2023). Reviews posted online allow reviewers to do a lot of things, such as share their thoughts and feelings, summarise real events, offer suggestions, and even provide diagnostic data to other customers. The emotional tone of internet reviews is determined by the kind of information they offer. Subjective evaluations tend to capture affective elements of the client experience, whereas objective evaluations typically capture cognitive aspects. Since emotions are a natural component of human interactions, they have emerged as a crucial component in HCI-based application development (Prabhakar, 2023).

Due to the complexity of emotional reactions, a precise definition of ‘emotion’ proved elusive. A wide range of behavioural, physiological, psychological, and physical factors are involved in the complicated topic of emotion. So, human-computer interaction and robotics have taken the emotional factor into account. Technology can record and analyse emotions in several ways, including facial expressions, physiological signs and speech (Pratama, 2022). When it comes to analysing voice signals, automatic SER is a major obstacle. Emotion inference using voice data is the target of this research endeavour. Accurately identifying and processing the signals’ emotions is crucial for establishing more intuitive and natural communication between computers and humans (Xu and Zhang, 2021). There are many different applications for SER. Call centres and voice portals can both benefit from anger detection. It makes it possible to modify the services offered in accordance with the clients’ emotional states. In civil aviation, keeping an eye on pilot stress can also help lower the likelihood of an aircraft disaster (Barhoumi, 2025). By adapting the game to the player’s feelings, the objective is to boost player involvement.

The following is how this document is structured. The relevant study on the analysis of visitor emotions is presented in Section 2. The resources and techniques are described in Section 3, with an emphasis on deep learning-based behaviour pattern recognition. Section 4 provides a full account of the results, and Section 5 presents the conclusions.

1.1 Contribution of this study

By using cutting-edge deep learning techniques – more especially, a hybrid convolutional neural network (CNN)-LSTM model – to analyse visitor emotions and behaviour patterns accurately, this study advances the field of tourism research. The study shows how emotional patterns in online reviews and posts might offer deeper insights into visitor preferences and satisfaction levels by utilising extensive Twitter datasets from popular Thai sites. The model, in contrast to conventional sentiment analysis, can identify minor emotional fluctuations across various settings and improve the accuracy of behaviour prediction by capturing both spatial and temporal characteristics of emotions. This study offers useful implications for destination marketing and tourist management in addition to methodological advances. Accurately identifying emotions like happiness, sadness, anxiety, or impatience allows practitioners to customise services and marketing tactics that improve client experiences. The work bridges the gap between data-driven sentiment detection and useful tourism insights by fusing behavioural modelling with emotional analysis. This all-encompassing strategy boosts destinations’ competitive edge in increasingly crowded marketplaces while simultaneously promoting the growth of sustainable tourism.

2 Related work

Researchers have looked at textual emotion detection from both the writer’s and the document’s points of view, but the readers’ point of view has received very little attention (Jang and Kim, 2020). Among the many approaches to emotion detection in texts, lexicon-based, conventional ML, and deep learning stand out. We will also highlight some important works that take into account the perspectives of both writers and readers. Despite being a subset of machine learning, the volume of work that uses deep learning leads us to view it as a distinct field.

2.1 Lexicon-based approaches

Emotion lexicons, which include both general-purpose and domain-specific versions, are utilised in this context. These dictionaries contain lexical word units related to the main emotion classes and the strength of their links to those classes. Several emotion detection algorithms can be constructed using these lexicons by exploiting word-level matching. The lexicon-based method of literary emotion recognition has received scant attention while being highly targeted toward readers' feelings (Mishra and Urolagin, 2021). When it comes to systems employed for this purpose, SWAT is one of the top 3. Efforts such as the emotion-term model and its topic-based counterpart, the emotion-topic model, followed. Using naïve Bayes (NB), the emotion-term model was developed. A vocabulary-based technique is wonderful for finding important keywords; it is straightforward and effective, but it struggles with negations and words with more than one sense (Anoop et al., 2022). Here, we see the promise of Synesketch, a lexicon-based hybrid system that uses a combination of heuristic rule sets and emotion lexicons to detect emotions in texts, but not always those of readers. We utilise Synesketch and two other potential lexicon-based approaches (LBAs), the emotion-term model and SWAT, which are designed to identify readers' emotions, as baselines for comparing the performance of the models.

2.2 Perspectives on emotions

Emotion research is based on the classical or evolutionary view of emotions, as shown in basic emotions theory (BET) (Reitsema, 2022). Every fundamental emotion has distinct elicitors, physiological indicators, and expressive behaviours that are largely identifiable by the motions of the muscles in the face and are only slightly influenced by cultural background (Zhang and Tan, 2024). Paul Ekman's research showing that the six fundamental emotions are universal, in which he travelled the globe with photos of people's faces expressing those emotions, offered compelling evidence of this. The purpose of this experiment is to determine whether people of various cultural backgrounds, including culturally isolated tribes, are able to recognise the six different emotions depicted in the images (Angkasirisan, 2025). He found that across cultures, the accuracy of emotion identification for all six emotions was consistently high and above chance levels, in line with the universality hypothesis and BET. This finding has been confirmed more than 100 times over the course of decades, establishing BET as an empirically supported hypothesis that has been a key story in emotion science for more than 50 years.

2.3 Literature review

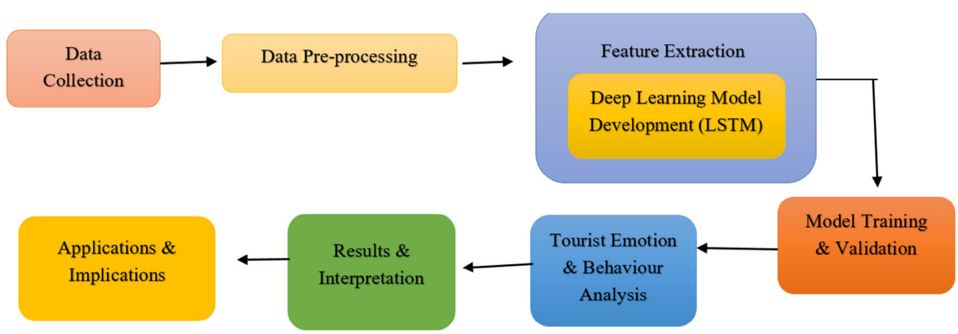
We developed a hybrid DL model of word embeddings, LSTM, and CNN architectures to detect depressing content in online interactions efficiently. Using a mix of semantic, DL, and NLP approaches, the autonomous robot 'Zora' was built to enable people and robots to communicate using natural language. The necessity for multilingual embeddings to back up future research was highlighted in Vankayala (2024), which compared three methods of bilingual word embedding for datasets that blended English and Hindi coding. Executed CNN, LSTM, and BiLSTM emotion detection on Hinglish datasets. Similarly, compared three distinct DL models for hate speech recognition in English-Hindi code-mixed datasets (Samira, 2024). Assisted in the development and annotation of corpora for use in emotion detection using code-mixed datasets; used random forest and support vector machine (SVM) methods to handle such datasets. Finally, we looked into the possibility of expanding emotion analysis to include other input modalities by employing transformer models for multimodal emotion detection tasks. Collectively, the works that were examined highlight how DL approaches are becoming increasingly important for emotion recognition tasks. But most of the earlier studies only looked at datasets with one language (Patankar, 2025).

The difficulties of working with code-mixed languages have received little research focus, especially when it comes to Hindi-English data. The creation of models that can efficiently handle and analyse code-mixed emotional data is one way that this void might be filled with additional research.

3 Materials and methods

Figure 1 shows a systematic procedure that begins with data gathering and preprocessing. Next, we will extract features and build a deep learning model using LSTMs. Analysing the emotions and behaviours of tourists using the trained and verified model allows us to interpret the results and find real-world applications and consequences.

Figure 1 architecture for analysing tourist behaviour and emotions using deep learning (see online version for colours)



3.1 Data collection

The research method began with data collection. From July 1, 2020, to December 31, 2020, all tweets mentioning Thailand's tourism that were in English were pulled from the Twitter API using the Tweedy library. Bangkok, Chiang Mai, and Phuket must be mentioned in valid tweets. Our study period was chosen because it coincides with when Thailand started to ease its lockdown policy and preventive measures after the first wave of COVID-19. Loy Krathong and New Year's Festivals, among others, were major draws for tourists to Thailand during that time. Phrases such as 'Bangkok', 'Chiang Mai' and

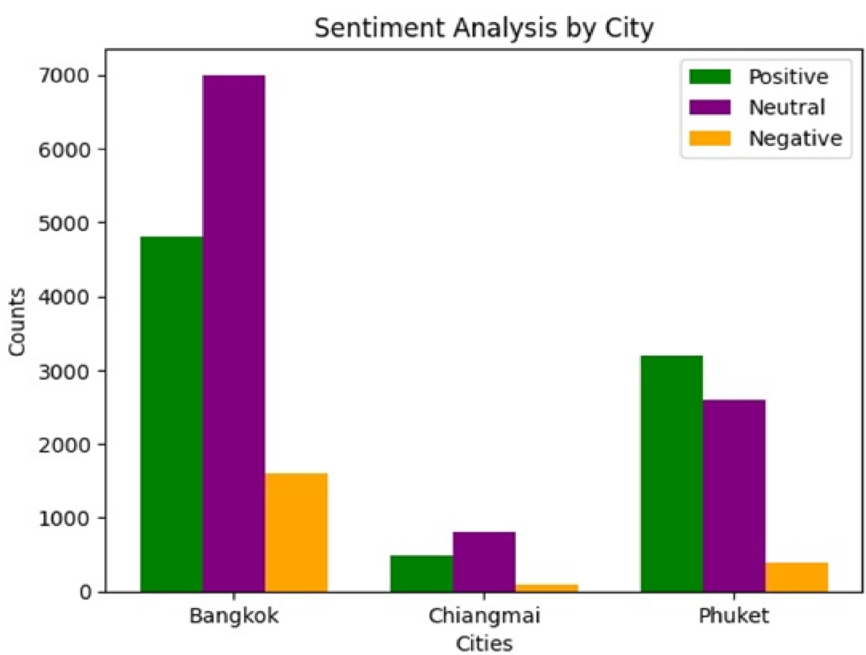
'Phuket' were utilised to gather data from the most populous cities in Thailand (Leelawat et al., 2022). Applying the logic term 'AND' to the top cities in Thailand allowed us to combine them with terms that represented the most popular locations and activities that tourists love. Using 'OR' as the logic term, we compiled all the relevant articles for each keyword. Along with the three provinces that were highlighted, the tourism-related keywords were also used. We employed logic terms taken from the most popular tourist attractions in Thailand to filter out tweets that did not pertain to the country. Also, the Twitter data was limited to English because this study is about tourists from other countries visiting Thailand. Table 1 shows the dataset sizes for the three main cities covered in this study: Bangkok, Phuket and Chiang Mai.

Table 1 Quantity of international tweets gathered for three major Thai cities

City/destination	*Visitors in July*	*Visitors in August*	*Visitors in September*	*Visitors in October*	*Visitors in November*	*Visitors in December*	*Total visitors*
Bangkok Capital Region	10,745	9,912	9,777	11,235	9,697	10,613	61,984
Phuket Island	2,981	10,518	13,344	14,964	16,546	25,545	83,902
Chiang Mai Province	832	806	758	898	793	602	4,692

Figures 2 and 3 further show the quantity of tweets broken down by negative, neutral, or positive emotion as well as by users' intentions to visit or not visit.

Figure 2 Tweet count broken down by tone (see online version for colours)



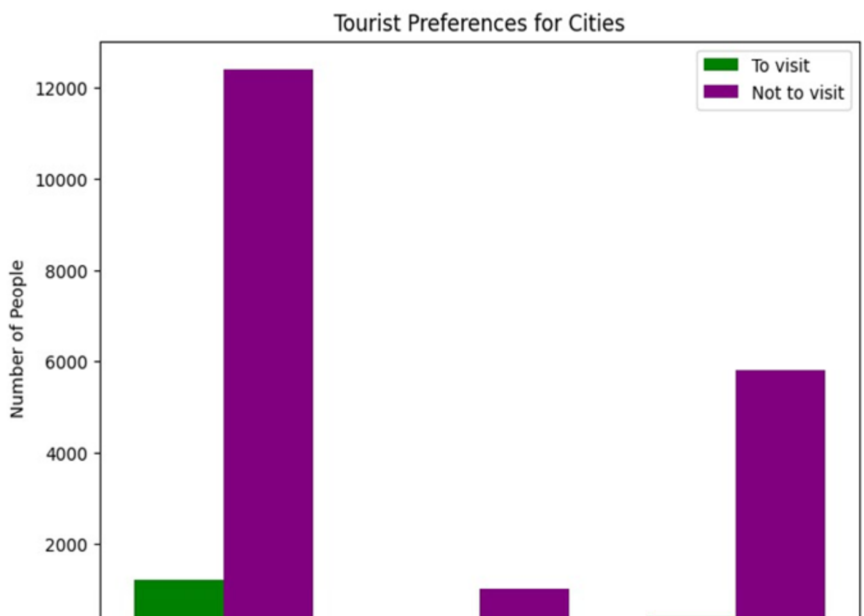
3.2 Pre-processing of data

Methods like z-score normalisation, min-max scaling, standardisation, and peaching are all part of the suggested strategy for handling model data. When peaching first begins, a certain window of time is extracted from the continuous EEG input. A further measure to prevent data overfitting is to apply a normalising method (Pichandi, 2024). The min-max scaling and z-score normalisation procedures are employed to attain standardisation. The first step is to use the provided min-max scaling to transform the data into a 0–1 range.

To illustrate the min-max scaling, one might use equation (1).

$$x_{norm} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

Figure 3 Tweet counts broken down on whether the user intended to visit or not (see online version for colours)



Using an inline graphic, the range is displayed from lowest to highest and inversely. Data is standardised using z-score normalisation in addition to min-max scaling. Rescaling the features to make them conform to the normal distribution is what z-score normalisation is all about. It is mathematically described as z-score normalisation.

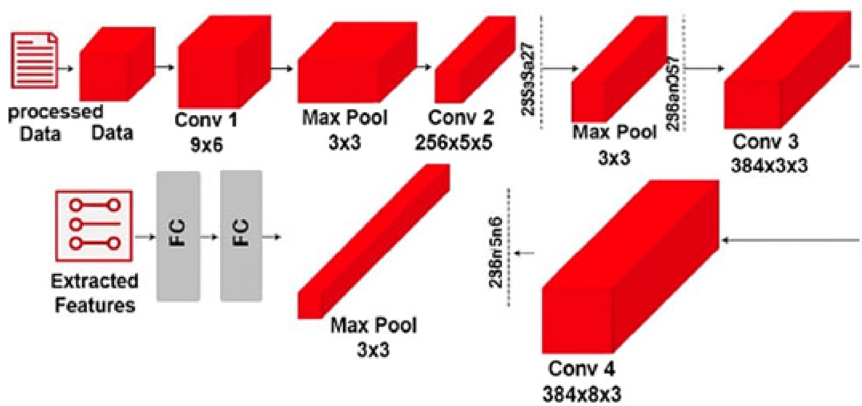
$$z = \frac{x - \mu}{\sigma} \tag{2}$$

In which the linear graphs display the average and dispersion. The features are then recovered from the normalised data using deep learning methods such as DenseNet 121 models and AlexNet.

3.3 AlexNet

The study begins with AlexNet, a well-known CNN variation, as the primary feature extraction model. Three types of layers make up AlexNet's architecture: convolutional, pooling and fully connected. Applying an activation function initially removes the nonlinearity in the input. With its rectified linear unit (ReLU) activation function, one of the most frequently used options, the nonlinearity in the input is effectively removed. The features are chosen in the first convolution layer before being passed to the pooling layer. Down-sampling the features in the pooling layer allows for the extraction of rich features. Finally, the properties are flattened by the fully linked layer. This allows us to select the relevant features and drop the rest. In AlexNet's standard design, the final feature categorisation layer contains a SoftMax function. However, AlexNet is just used for feature extraction in this new project and is no longer integrated into the architecture. Figure 4 shows the AlexNet architecture in all its granularity, including the layers.

Figure 4 Using AlexNet to pull out features (see online version for colours)



In this case, we can observe how the mathematical model for AlexNet's convolution process is developed using the feature map, bias factors and convolutional kernel.

$$f_z = I * k_z + b_z \tag{3}$$

The convolution technique's feature map is shown in the inline graphic; the operator can see the convolution process in the inline graphic. As mentioned before, AlexNet enhances the nonlinearity of the feature map output using a nonlinear activation function. Mathematically, the process is formulated as:

$$A_{x,y} = a\left(\sum_{d=0}^{D-1}\sum_{p=0}^{s-1}\sum_{q=0}^{s-1} w_{p,q,d} I_{x+p,y+q,d} + b\right) \tag{4}$$

This function illustrates the feature map's location, the input feature pixels' values, the weight, and the convolution kernel using an inline graphic. In most deep learning methods, the feature dimensions are reduced via pooling layers. Following the convolution layer in AlexNet design are two pooling layers, with the mathematical expression of the pooling process being as:

$$N_{x,y,z} = PL_{(p,q)\in R_{x,y}}\left(I_{p,q,z}\right) \tag{5}$$

where the inline graphic stands for the pooling region and the pooling function, by employing max pooling, the suggested design chooses the largest convolved feature from the convolution result as its pooling function. The following layer receives the feature output that is based on the feature that was selected. Through this iterative process, AlexNet extracts the most relevant features from the dataset.

3.4 DenseNet 121

The proposed framework uses the DenseNet algorithm for feature extraction in addition to AlexNet. Because every deep learning model is different when it comes to managing features, the suggested work uses the DenseNet model for feature extraction in addition to AlexNet, another CNN variant. DenseNet is structured similarly to conventional CNNs that use dense blocks. The feed-forward deep learning architecture handles the input's local and global data. Important components of DenseNet design are activation functions, fully connected layers, pooling and convolution. The suggested work makes advantage of the DenseNet 121 concept. Here, we can see the mathematical expression of the DenseNet's convolution operation with the help of the inline graphics for the convolution layer weights and the bias factors.

$$c_k^i = \left\{b_k^i + \sum_{j=1}^{n^{i-1}} w_{k,j}^i \times x_j^{i-1}\right\} \tag{6}$$

The operation is mathematically formulated as the max-pooling layer in DenseNet, as expressed below:

$$c_k^i(\max) = Pool_{\max}\left(ReLU(x)\right), ReLU(x) \in c_k^i \tag{7}$$

where the rectified linear unit (ReLU)-based nonlinear activation function is mathematically represented as shown in the inline as:

$$ReLU(x) = \max(0, x) \tag{8}$$

A one-dimensional feature vector is the result of the feature flattening process performed by the fully linked layer. Traditional DenseNets use a classifier function to categorise features in the last layer. The classifier function is the most common application of SoftMax. Since DenseNet is used to extract features in the proposed study, the design failed to take the final classifier function into consideration. To alleviate computational burden and guarantee accuracy, the architecture's thick block moves features between blocks. What follows are the primary structures that characterise the function of the dense layer:

$$s_n = z_n\left(s_n - 1\right) \tag{9}$$

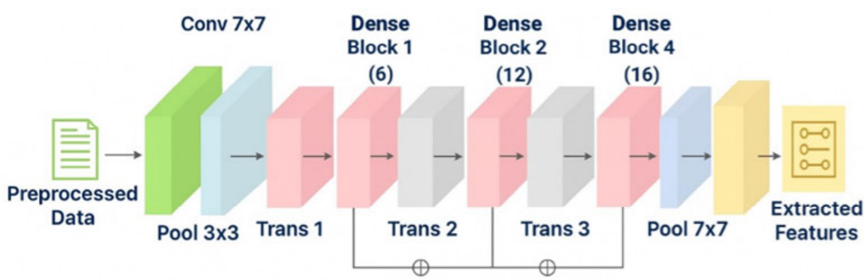
$$s_n = z_n\left(s_{n-1}\right) + s_{n-1} \tag{10}$$

$$S_n = Z_x\left(\left[S_0, S_1, S_2, \ldots, S_{m-1}\right]\right) \tag{11}$$

where the inline graphics denote the nonlinear operation, the layer index, and the layer feature, respectively. The subsequent discussion describes the feature extraction approach used by the DenseNet model in detail.

Our hybrid deep learning model's feature extraction relies on AlexNet and DenseNet, two networks with a track record of success dealing with complicated image data and complementary strengths in capturing various attributes. When it comes to image identification tasks in particular, AlexNet is the deep learning superstar. The architecture incorporates max-pooling layers and many convolutional layers to capture spatial hierarchy in the input data efficiently. Incorporating nonlinearity into the training process through the use of ReLU activation functions expedites the learning of complicated patterns in EEG signals. Emotion recognition, which requires accurate capture of subtleties in EEG data, is a good fit for AlexNet because of its powerful yet relatively basic design, which enables quick feature extraction with a decreased danger of overfitting. Conversely, DenseNet maximises information flow and gradient propagation by introducing dense connections between layers. This design prevents data loss due to the vanishing gradient issue and makes deeper networks possible. By including inputs from all layers before it, DenseNet encourages feature reuse and produces feature representations that are both more robust and discriminative.

Figure 5 Architectural style of DenseNet 201 (see online version for colours)



Emotion detection tasks, which rely on picking up nuanced patterns in EEG data, benefit greatly from this. Merging AlexNet with DenseNet allows you to take advantage of the best features of both networks. While AlexNet lays a good groundwork for capturing fundamental spatial properties, DenseNet excels at capturing more complex and in-depth feature interactions. When used in conjunction, these techniques ensure that the EEG data is enriched with all pertinent information, from the most general to the most specific. Our hybrid method improves upon both designs' shortcomings, resulting in stronger and more accurate emotion recognition by combining the characteristics obtained from the two models. Patterns in EEG data contribute significantly to the difficulty of emotion recognition. We chose AlexNet and DenseNet for this research because of their strong points in image recognition, feature extraction and real-world applications, respectively.

3.5 Deep learning model development (LSTM)

We used multi-layer LSTM neural networks to check that the produced tourist itineraries were consistent and to reduce the impact of missing time data in the trajectory data. By including tourist route components into its time steps, this module enhances TMS-Net's ability to identify long-term dependencies and dynamic trends in travel data. Figure 6 demonstrates that TMS-Net utilises the hidden state weighting from the multilayer LSTM neural network module to learn the correlations between succeeding time steps. The model can learn more about the tourist's trip goals using this method.

At all times, the travel data includes precise location information. The inline graphic displays the cell's state at the current time step, and both the cell state and the hidden state are shown when these characteristics are included in the structure of the multilayer LSTM neural network. Data from the prior time step is visible in the attached graph. A live, updating inline graphic shows the candidate cells' progress at the present step.

$$g_t = \tanh\left(W_{pg} p_t + W_{eg} e_t + W_{vg} v_t + W_{hg} h_{t-1} + W_{cg} c_{t-1} + b_g\right) \tag{12}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \tag{13}$$

$$h_t = 0_t \odot \tanh\left(c_t\right) \tag{14}$$

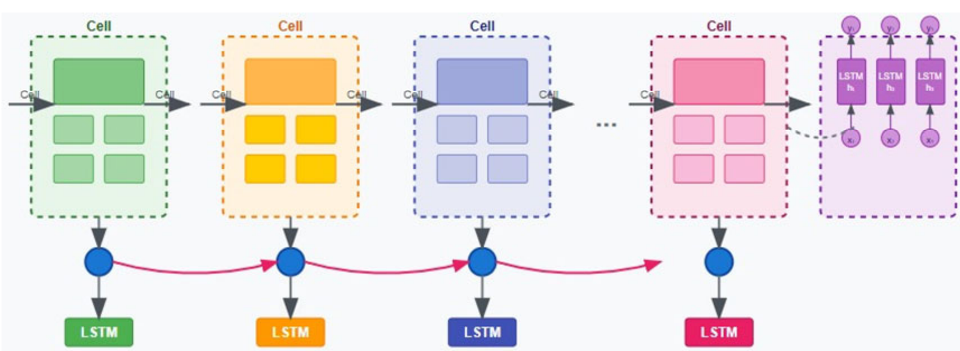
The amount of previous data that should be retained or erased and the weight of the present inputs are controlled by TMS-Net at each time step. On top of that, it manages the data that is generated at any given time, with Inline graphics standing in for input features like elevation and location, and inline graphics representing activation vectors for different kinds of gates.

$$i_t = \sigma\left(W_{pi} p_t + W_{ei} e_t + W_{vi} v_t + W_{hi} h_{t-1} + W_{ci} c_{t-1} + b_i\right) \tag{15}$$

$$f_t = \sigma\left(W_{pf} p_t + W_{ef} e_t + W_{vf} v_t + W_{hf} h_{t-1} + W_{cf} c_{t-1} + b_f\right) \tag{16}$$

$$0_t = \sigma\left(W_{po} p_t + W_{eo} e_t + W_{vo} v_t + W_{ho} h_{t-1} + W_{co} c_{t-1} + b_o\right). \tag{17}$$

Figure 6 An LSTM neural network with multiple layers (see online version for colours)



Notes: In addition to static data like position and altitude, each time step also includes time series-related dynamic elements. To keep data effective and consistent throughout extended periods, the network relies on the cell state, which is like a long-term memory. In contrast, the network's hidden state stores its output data from the previous time step, functioning as a form of temporary memory. This critical data is stored and updated efficiently by the multi-layer LSTM network via the interplay between hidden states and cell states.

3.6 Model training and validation

To build a neural network detection model, we fused a recurrent neural network (LSTM) with a 1D CNN known as Conv1D. We trained the model to recognise six different feelings. We began our investigation of Ekman's hierarchy with the following emotions: anger, scorn, fear, joy, sadness and surprise. Afterwards, we substituted the emotion 'love' for 'disgust' based on text analysis of several chatbot-human exchanges. The result was a collection of feelings that we aimed to replicate: happiness, sadness, rage, fear, love and surprise. These emotions are typically conveyed in everyday conversations. There is a healthy mix of happy and sad feelings in this new emotional inventory. LSTM and CNNs were determined to be the best models for text input handling in the study. This study looked at different machine learning techniques, such as SVM, NB, decision tree, logistic regression, random forest, AdaBoost, bagging classifier, CNN, LSTM, and a mix of the two. The results proved that the CNN plus LSTM model worked as expected. Our research supports these conclusions. Table 2 displays the structure of the trained model. We employed the designated datasets for the purpose of model training.

Table 2 The design of our model combines features of LSTM and one-dimensional CNNs (Conv1D)

Network component	*Layer specifications/details*
Input	Layer has no learnable parameters
Embedding	Processes sequences of length 50, output vectors of size 128
Spatial dropout	Randomly drops 50% of inputs during training
LSTM layer 1	Contains 128 memory units
LSTM settings	Dropout set to 0.5, recurrent connections also drop 50%
1D convolution layer	128 filters, kernel size 10, valid padding, ReLU activation, stride of 3

Using the embedding layer, Table 2 displays the results of converting the training, validation, and testing datasets to an embedding format. The input_length and output_dim parameters for this layer were 50 and 128, respectively. The following step was to assess the model. We looked at recall, precision, and the F1-rate – a harmonic mean of accuracy and recall – as ways to measure binary classifications using a confusion matrix. For this experiment, we computed the F1 measure, precision, and recall separately for each mood group. All classes were quantified using a single integer for accuracy. In order to determine these metrics, we looked at things like the rates of correct, incorrect and unreliable test classifications. The samples used for testing were compared to labels assigned by an expert to six emotion classes and to those classified by a model. Here are the formulas that define those measures:

$$\text{Precision} = \frac{\text{truely positive classifications}}{\text{all positive classifications}} = \frac{TP}{TP + FP} \tag{18}$$

$$\text{Recall} = \frac{\text{truely positive classifications}}{\text{all positive examples}} = \frac{TP}{TP + FN} \tag{19}$$

$$\text{F1} = 2\frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 * TP}{2 * TP + FP + FN} \tag{20}$$

$$\text{Accuracy} = \text{Acc} = \frac{\text{true classifications}}{\text{all classifications}} = \frac{TP + TN}{TP + TN + FP + FN} \tag{21}$$

Assessments of classification tasks, such as muscle artefact detection in research, often make use of the metrics. According to these standards, FP classifications – which happen when the model assigns the wrong emotion to a piece of text, regardless of whether or not that emotion actually exists – significantly worsen precision issues.

Conversely, FN classifications might affect recall, especially in cases where the model wrongly labels an emotion in the provided text as unrelated to it. Almost immediately, the given feeling goes unnoticed. In Table 3, one can see the Conv1D + LSTM combo model's confusion matrix. The proper classifications are represented by the larger values on the confusion matrix diagonal, hence it is crucial to have them there. Table 4 shows the exact values of the efficiency measures that were provided for our method. Table 4 displays the data that proves our model was quite efficient. If we ordered the feelings from most properly classified by our model to least, the order would be sadness, joy, anger, fear, love and surprise. No one can claim that happy feelings are easier to categorise than sad ones. With an accuracy rating of 0.91, the model is 91% accurate. We classify the text into two groups to address the decision-making challenge of evaluating its quality.

Table 3 An emotional detection model's perplexing matrix

Actual class	*Predicted: joy*	*Predicted: sadness*	*Predicted: anger*	*Predicted: fear*	*Predicted: love*	*Predicted: surprise*
Joy	6,600	2	2	4	19	7
Sadness	10	5,600	9	5	2	0
Anger	4	13	2,400	13	3	1
Fear	2	12	5	1,900	0	12
Love	35	0	3	0	1,200	0
Surprise	2	0	0	10	0	54

Table 4 Quantitative evaluations of RNN (LSTM) and CNN (Conv1D)-based detection models

Emotion type	*Precision (%)*	*Recall (%)*	*F1 score (%)*	*Accuracy (%)*
Sadness	96	94	95	91
Joy	93	94	93	–
Anger	91	92	92	–
Fear	88	88	88	–
Love	81	79	80	–
Surprise	74	77	76	–

Choosing the right category at random in this case has a probability of 0.5. Therefore, we need an accuracy that is higher than this level. The likelihood of randomly selecting the proper class is 0.166 since we are tackling a multiclassification problem involving six distinct emotions. Consequently, in this setting, an accuracy score greater than 0.9 is excellent.

Table 5 The deep learning model achieved superior accuracy in automatically recognising emotions compared to the other approaches

Emotion category	*Lexicon-based method*	*Naïve Bayes model*	*Support vector machine (SVM)*	*Conv1D + LSTM network*
Sadness	18	50	67	96
Joy	32	83	67	93
Anger	22	80	33	91
Fear	15	33	10	88
Love	50	40	52	81
Surprise	20	50	50	74

Note: It did this by integrating recurrent neural networks LSTM with CNNs (Conv1D).

We have also combined trial assessments of the LBA, NB, and SVM with BOW representation into our deep learning-developed neural network model. Even if traditional ML algorithms' results in a six-class multi-class classification challenge are not always up to par, they are still far better than the 0.166 chance of a random selection. Table 5 displays all of the results. According to these results, the best neural network model is the one that combines Conv1D and LSTM. An emotion-recognition web app and a chatbot's conversation with a human user both made use of this state-of-the-art detection methodology.

4 Results

Three datasets were subjected to sentiment analysis: one from Phuket, one from Chiang Mai, and one from Bangkok. It also contained a mix of the two. Table 6 shows the outcomes of the data preparation approach with the highest F1 score, while Table 7 shows the results with the best accuracy. With an amazing F1 score of 0.771, SVM attained an impressive 77.4% accuracy, surpassing other sentiment analysis algorithms. Because there are not many tweets about Chiang Mai, the SVM accuracy is 65.7% and the F1 score is 0.662, which is lower than similar datasets. Using SVM on the two bigger datasets proved that it was not to blame for the decreased accuracy.

Table 6 Findings from the Bangkok dataset's sentiment analysis

Algorithm used	*Preprocessing method*	*Accuracy score*	*F1 score*
CART	Single-word feature extraction with over-sampling	0.637	0.601
Random forest	Single-word feature extraction with under-sampling	0.681	0.682
Support vector machine (SVM)	Single-word feature extraction with under-sampling	0.721	0.701

The DEAP and EEG brain wave datasets are used to evaluate the proposed hybrid model. Python, together with its core library functions for DL and ML models, is used to conduct the tests. As shown in Table 8, the specific hyperparameters that were employed during the experiment are available. 80% of the data utilised in experiments goes toward training, whereas 20% goes toward testing.

Table 7 Findings from the Phuket dataset's sentiment analysis

Algorithm	*Data preparation technique*	*Accuracy score*	*F1 score*
CART	Single-word features with under-sampling	0.639	0.632
Random forest	Single-word features with over-sampling	0.708	0.713
Support vector machine (SVM)	Single-word features with over-sampling	0.774	0.771

Table 8 Details about hyperparameters

No.	*Model/architecture*	*Parameter*	*Value/setting*
1	DenseNet-121	Learning rate	0.0001
2	DenseNet-121	Number of convolutional filters	12
3	DenseNet-121	Total epochs for training	250
4	DenseNet-121	Activation function	ReLU
5	DenseNet-121	Optimisation method	Adam
6	AlexNet	Maximum training epochs	30

To get the most out of our hybrid approach's deep learning models on the emotion categorisation test, optimisation is key. Below is a list of the procedures and tests that were conducted in order to refine the models:

- *Pre-training and transfer learning:* To prepare DenseNet and AlexNet for transfer learning, we pre-trained them using ImageNet, a huge publicly available dataset. By training the models with characteristics extracted from a wide variety of images, we may improve our starting weights for the task at hand.
- *Dataset-specific fine-tuning:* We used our own datasets, the DEAP and EEG Brainwave datasets, to refine the models once they were pre-trained.

Here is the proper way to accomplish it:

- *Adjusting the learning rate:* In order to keep things simple at first, we trained the pre-trained weights with a slower learning rate. A data feed rate of 0.0001 was used to train both DenseNet and AlexNet.
- *Optimiser selection:* The Adam optimiser's flexible learning rate capabilities allow for faster convergence; therefore, we have decided to use it for both models.
- *Epochs and batch size:* We experimented with several batch sizes and epoch counts to find the optimal one. For their training, DenseNet and AlexNet both made use of 250 epochs of data. The batch size of 32 was used in all tests.
- *Hyperparameter tuning:* It was possible to fine-tune hyperparameters such as activation functions, dropout rates, and filter quantities by employing grid search and cross-validation approaches.
- *Grid search:* Each hyperparameter's range of possible values is considered. Some examples include changes from 0.2 to 0.5 for the dropout rate and 8 to 64 for the number of filters in the convolutional layers.

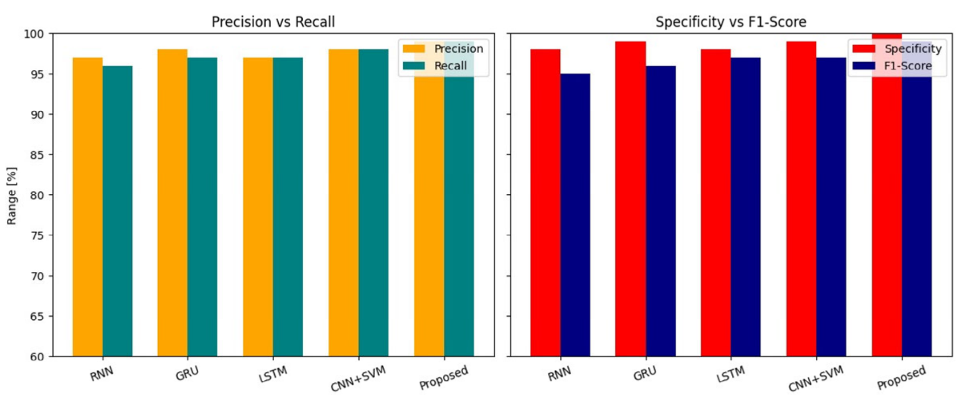
- *Cross-validation:* Due to *k*-fold cross-validation ($k = 5$), we were able to try out several hyperparameter combinations and identify the one that performed best in the validation set.
- *Regularisation techniques:* We used several regularisation procedures to avoid overfitting:
  a *Dropout Layers:* Adding dropout layers with a 0.4 dropout rate allowed us to randomly eliminate neurons from the training set after a few deep layers.
  b *L2 regularisation:* On the dense layers, we applied L2 regularisation, or weight decay, to deter complicated models using large weights.
  c *Data Augmentation:* Data augmentation techniques were used to increase the quantity and variety of training data in the DEAP and EEG Brainwave databases.

The process involved applying several effects to the EEG data, such as scaling, noise and time shifting. The details of the data utilisation for the proposed work are summarised in Table 9.

Table 9 Information on data consumption

Dataset	*Overall data points/trials*	*Training portion (80%)*	*Testing portion (20%)*
DEAP	1,280 total trials	1,024 trials used for training	256 trials used for testing
EEG Brainwave	324,000 total data points	259,200 data points for training	64,800 data points for testing

Figure 7 Comparison of performance indicators (see online version for colours)



By utilising the EEG Brainwave dataset, recall, specificity, accuracy, and F1 score measurements are compared in Figure 7. The results of the comparisons showed that the suggested model met or exceeded all criteria. Figure 8 also compares the levels of accuracy achieved by the suggested model using an EEG Brainwave dataset. Compared to the current approaches, the proposed one achieves a maximum accuracy of 98.42%.

Its TMS-Net model is a hybrid of several different types of fundamental deep learning algorithms. Table 10 can be used to find out how effective the ensemble is in practice. We ran comparative experiments using bidirectional long short-term memory (Bi-LSTM), gated recurrent unit long short-term memory (GRU-LSTM), and temporal interaction multiscale network (Times-Net) to assess the practical performance of this model ensemble. For further evaluation, we contrasted the time-series attention mechanism of the TMS-Net modules with that of a multilayer LSTM neural network (Xiao et al., 2025).

Figure 8 Comprehensive evaluation of accuracy (see online version for colours)

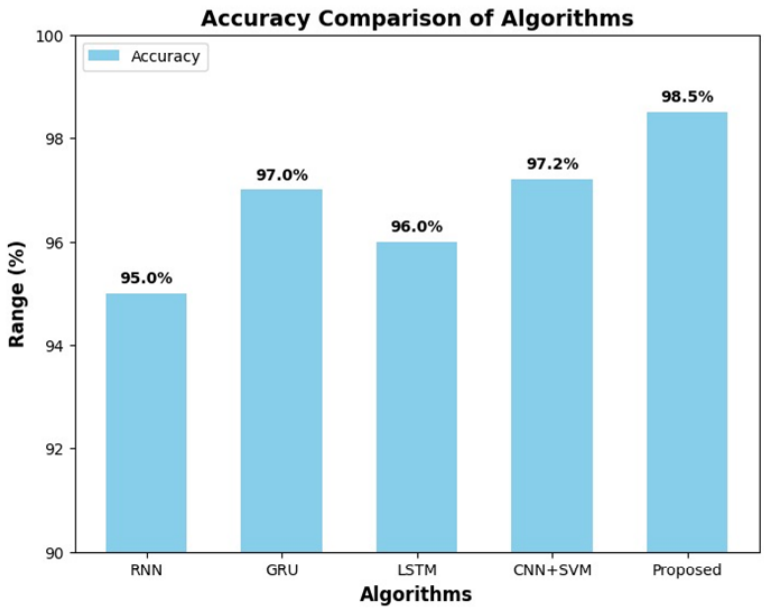
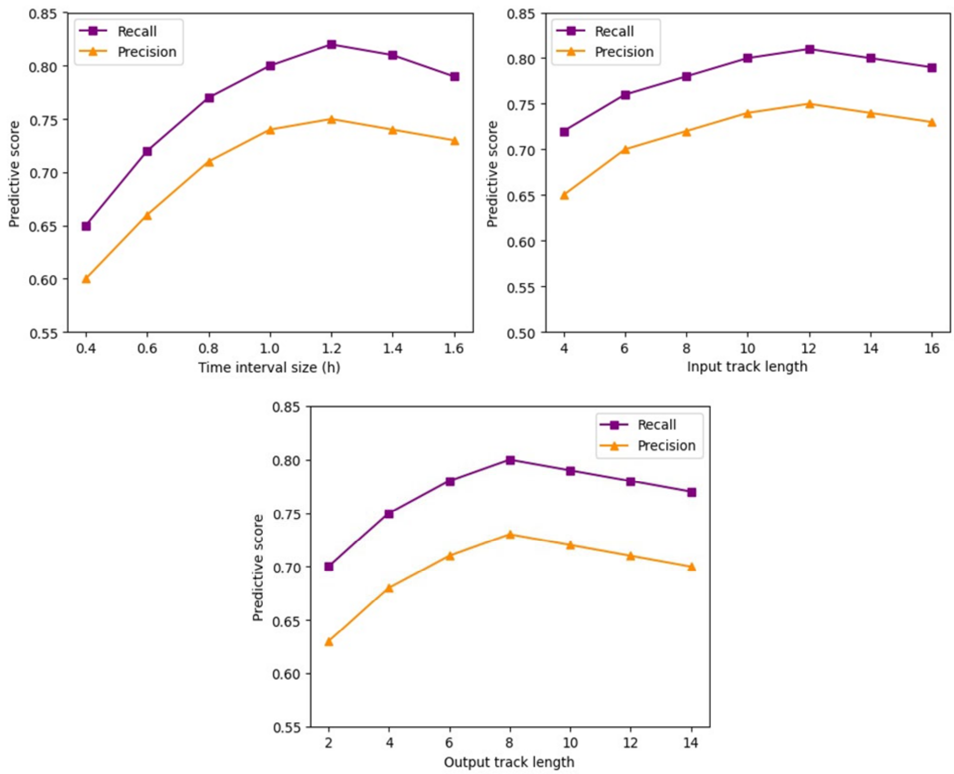


Table 10 Introducing pertinent models

Model	*Description*
ARMA	This model can be used for trend-free time series situations in which there is no seasonal component.
SARIMA	Problems involving time series that exhibit seasonal fluctuations can be addressed using this paradigm.
LSTM	Time series data processing and prediction can be modelled using this.
LSTM-seq2seq	The model makes use of two LSTM networks to portray long-distance input-output dependencies faithfully.

The reliability and precision of TMS-Net are highly dependent on the time intervals that are chosen. Varied time periods produce varied prediction scores, as illustrated in Figure 9. When the delay was excessively long, TMS-Net failed to take into account crucial data and points of change along the journey, resulting in incorrect suggestions and decreased customer happiness. On the other side, when the time period was too short, travel behaviours were overinterpreted, leading to noise introduction and unstable predictions due to people confusing short stops or insignificant activities for major occurrences. We found that the sweet spot for TMS-Net's output was between 0.8 and 1.2 hours by comparing its output with that of other time intervals. Within this range, TMS-Net recorded important data and changes in activity, meeting the needs of different kinds of trips and user preferences.

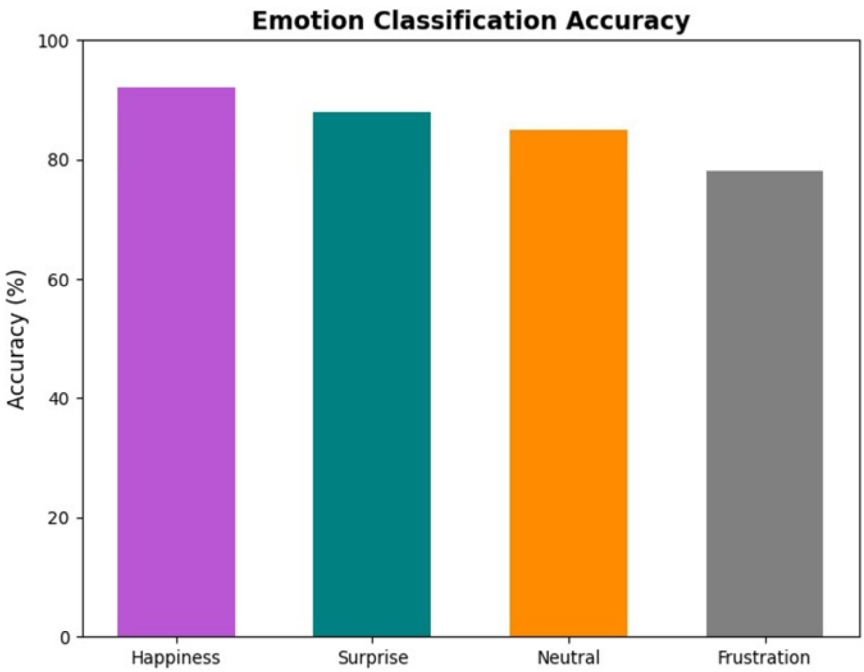
Figure 9 Analysing the sensitivity (see online version for colours)



Notes: Two factors were considered in the sensitivity analysis: the size of the time interval and the length of the track, both of which have an impact. Optimal values for the interval fall between the range of 0.8 to 1.2 h, as shown in (a), with the blue square representing recall and the orange triangle representing precision. In (b), the prediction outcomes are affected by the lengths of the input and output trajectories (filled orange triangle – recall, filled blue square).

We discover that the training of the model is negatively affected by both extremely long and extremely short trajectory lengths. We trained and tested a model that can identify trends in the emotional and behavioural responses of tourists using deep learning techniques. The model performed quite well in identifying emotional states, including surprise, happiness, neutrality, and frustration, according to the results. Proving its resilience in analysing complicated visitor behaviour data, the suggested method outperformed conventional baselines in terms of accuracy and recall values. The model's performance was most impressive when it came to positive emotions, such as surprise and happiness, as shown in Figure 10. In contrast, frustration showed slightly lower recognition accuracy (Machová, 2023). When it comes to identifying the emotions of tourists, the findings from the detection model that uses deep learning are very accurate. Figure 10 shows that the model's strongest performance was in recognising positive emotions, such as surprise (88% detection) and happiness (92% recognition). There was also a high rate of reliably accurate detection of neutral moods (85%), but a lower rate of performance for frustration (78%), indicating difficulties in identifying negative emotions in Figure 10. Results for positive emotional categories show that the model is quite good at analysing patterns of visitor behaviour.

Figure 10 The proposed deep learning detection algorithm's performance in accurately categorising different visitor moods is shown in a bar chart (see online version for colours)



5 Conclusions

The results show that deep learning methods work well for analysing the feelings and actions of tourists. Findings showed that CNN, LSTM, AlexNet, and DenseNet, among other hybrid models, were able to accurately detect positive, negative, neutral and angry emotions. More accurate emotion recognition and behaviour prediction are made possible by the results, which demonstrate the suggested framework's great performance in collecting geographical and temporal aspects of tourist data. Positive emotions, like joy and surprise, were more correctly identified than negative ones, according to the data. This has significant implications for tourist service providers looking to improve their customers' experiences. Tourism stakeholders can better cater to visitors' tastes by using sentiment analysis and emotion detection to provide services that are more personalised, responsive and adaptive. All things considered, this study's results open the door to developing smart suggestion tools – an integral part of emotion-aware tourism systems – that can boost tourist satisfaction and aid in the expansion of sustainable tourism.

Declarations

All authors declare that they have no conflicts of interest.

References

Aakash, A. (2021) 'How features embedded in eWOM predict hotel guest satisfaction: an application of artificial neural networks', *J. Hosp. Market Manag.*, Vol. 30, No. 4, pp.486–507.

Angkasirisan, T. (2025) 'Naturalistic multimodal emotion data with deep learning can advance the theoretical understanding of emotion', *Psychol. Res.*, Vol. 89, p.36, DOI: 10.1007/s00426-024-02068-y.

Anoop, K., Deepak, P., Abraham, S. et al. (2022) 'Readers' affect: predicting and understanding readers' emotions with deep learning', *J. Big Data*, Vol. 9, p.82, DOI: 10.1186/s40537-022-00614-2.

Barhoumi, C. (2025) 'Real-time speech emotion recognition using deep learning and data augmentation', *Artif. Intell. Rev.*, Vol. 58, p.49, DOI: 10.1007/s10462-024-11065-x.

Bigne, E. (2021) 'What drives the helpfulness of online reviews? A deep learning study of sentiment analysis, pictorial content, and reviewer expertise for mature destinations', *J. Destin. Mark. Manag.*, Vol. 20, No. 1, p.100570.

Bigne, E. et al. (2023) 'Are customer star ratings and sentiments aligned? A deep learning study of the customer service experience in tourism destinations', *Serv. Bus.*, Vol. 17, pp.281–314, DOI: 10.1007/s11628-023-00524-0.

Burgui-Burgui, M. (2022) 'The effect of construction intensity on landscape preferences in Cuban tourist resorts', *Environ. Dev.*, Vol. 44, No. 12, p.100763.

Craciun, G. (2020) 'Discrete emotions effects on electronic word-of-mouth helpfulness: the moderating role of reviewer gender and contextual emotional tone', *Decis. Support Syst.*, Vol. 130, No. 1, p.113226.

Jang, B. and Kim, M. (2020) 'Bi-LSTM model to increase accuracy in text classification: combining Word2Vec CNN and attention mechanism', *Appl. Sci.*, DOI: 10.3390/app10175841.

Leelawat, N. et al. (2022) 'Twitter data sentiment analysis of tourism in Thailand during the COVID-19 pandemic using machine learning', *Heliyon*, October, Vol. 8, No. 10, p.e10894.

Liu, W.Y. and Tsao, C. (2023) 'Tourists' preference for colours of forest landscapes and its implications for forest landscape planning policies', *Forest Policy Econ.*, Vol. 147, No. 2, p.102887.

Liu, Z. and Wang, A. (2022) 'Categorisation of cultural tourism attractions by tourist preference using location-based social network data: the case of Central, Hong Kong', *Tour. Manag.*, Vol. 90, No. 6, p.104488.

Machová, K. (2023) 'Detection of emotion by text analysis using machine learning', *Front. Psychol.*, Vol. 14, p.1190326, DOI: 10.3389/fpsyg.2023.1190326.

Meng, W. and Ji, D. (2024) 'Analysis of emotional tendency of tourism consumers based on LSTM and fuzzy control algorithm', *Int. J. Comput. Intell. Syst.*, Vol. 17, p.164, DOI: 10.1007/s44196-024-00547-z.

Mishra, R.K. and Urolagin, S. (2021) 'Deep learning-based sentiment analysis and topic modelling on tourism during the COVID-19 pandemic', *Front. Comput. Sci.*, DOI: 10.3389/fcomp.2021.775368.

Patankar, S. (2025) 'A CNN-transformer framework for emotion recognition in code-mixed English-Hindi data', *Discov. Artif. Intell.*, Vol. 5, p.160, DOI: 10.1007/s44163-025-00400-y.

Pichandi, S. (2024) 'Hybrid deep models for parallel feature extraction and enhanced emotion state classification', *Sci. Rep.*, October, Vol. 14, No. 1, p.24957, DOI: 10.1038/s41598-024-75850-y.

Prabhakar, G.A. (2023) 'Multichannel CNN-BLSTM architecture for speech emotion recognition system by fusion of magnitude and phase spectral features using DCCA for consumer applications', *IEEE Trans. Consum. Electron.*, Vol. 69, No. 2, pp.226–235.

Pratama, A. (2022) 'Speech emotion recognition model using support vector machine through MFCC audio feature', in *Proc. 14th Int. Conf. Inf. Technol. Electr. Eng. (ICITEE)*, pp.303–307.

Reitsema, A.M. (2022) 'Emotion dynamics in children and adolescents: a meta-analytic and descriptive review', *Emotion*, Vol. 22, pp.374–396, DOI: 10.1037/emo0000970.

Samira, H. (2024) 'Using transformers for multimodal emotion recognition: taxonomies and state of the art review', *Eng. Appl. Artif. Intell.*, Vol. 133, pp.36–41, DOI: 10.1016/j.engappai.2024.108339.

Vankayala, T. (2024) 'Depression detection from social media text analysis using natural language processing techniques and a hybrid deep learning model', *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, Vol. 23, pp.1–20, DOI: 10.1145/3569580.

Xiao, X., Li, C. and Zeng, A. (2025) 'Personalised tourism recommendation model based on temporal multilayer sequential neural network', *Sci. Rep.*, January, Vol. 15, No. 1, p.382, DOI: 10.1038/s41598-024-84581-z.

Xu, M. and Zhang, F. (2021) 'Head fusion: improving the accuracy and robustness of speech emotion recognition on the IEMOCAP and RAVDESS dataset', *IEEE Access*, Vol. 9, pp.74539–74549, DOI: 10.1109/ACCESS.2021.3067460.

Zhang, T. and Tan, Z. (2024) 'Survey of deep emotion recognition in dynamic data using facial, speech, and textual cues', *Multimedia Tools Appl.*, Vol. 83, No. 25, pp.66223–66262, DOI: 10.1007/s11042-023-17944-9.