International Science and Technology Journal
المجلة الدولية للعلوم والتقنية

عدد خاص بالمؤتمر الليبي الدولي للعلوم التطبيقية و الهندسية دورته الثانية
LICASE -2
2024 / 10 / 30-29

المجلة الدولية للعلوم والتقنية
ISTJ

تم استلام الورقة بتاريخ: 2024/10/2م
وتم نشرها على الموقع بتاريخ: 30/ 10/2024م

Adaptive HCI Systems with GRU-Based User Emotion Recognition and Response Prediction

http://www.doi.org/10.62341/amam2098

Abraheem Mohammed Sulayman Alsubayhay[1], Mohamed A. E. Abdalla [2], and Ali A. Salem Buras[3]

[1]Faculty of Arts and Sciences, University of Benghazi, Solouq, Libya
[2]Faculty of Information Technology, University of Benghazi, Benghazi, Libya
[3]Faculty of Arts and Sciences, University of Benghazi, Solouq, Libya
abraheem.alsubayhay@uob.edu.ly

Abstract

Human-Computer Interaction (HCI) has evolved to incorporate sophisticated systems capable of recognizing and responding to user emotions, enhancing the user experience by making interactions more intuitive and engaging. This paper explores the development of adaptive HCI systems utilizing Gated Recurrent Unit (GRU) neural networks for emotion recognition and response prediction. The core objective is to create interfaces that dynamically adjust based on the user's emotional state, thereby improving usability and satisfaction across various applications, including healthcare, education, and customer service. GRU-based models are particularly effective for this task due to their ability to handle sequential data and capture temporal dependencies inherent in emotional expressions. By processing multimodal inputs such as facial expressions, voice intonations, text, and physiological signals, these systems can accurately detect and predict emotions in real-time. The DEAP dataset, which includes EEG and other physiological recordings along with self-assessed emotional ratings, serves as a foundational resource for training and validating these models. The proposed approach involves pre-processing the physiological data using techniques like min-max normalization to ensure consistency and stability during model training. The GRU models then learn to map these inputs to corresponding emotional states, leveraging their memory capabilities to retain and interpret temporal patterns. The adaptive nature of these systems enables them to provide personalized responses, such as adjusting the interface layout or offering support when negative emotions are detected. The implementation of GRU-based emotion recognition and response prediction in HCI systems holds significant potential for enhancing user interactions by making them more responsive and emotionally intelligent. This paper demonstrates the effectiveness of GRU models in real-time emotion monitoring and highlights their applications in creating more adaptive and empathetic technology interfaces. The proposed model is implemented in Python and has an accuracy of about 99.12% which is higher than other existing methods.

International Science and Technology Journal
المجلة الدولية للعلوم والتقنية

عدد خاص بالمؤتمر الليبي الدولي للعلوم التطبيقية و الهندسية دورته الثانية
LICASE -2
2024 / 10 / 30-29

المجلة الدولية للعلوم والتقنية
ISTJ

تم استلام الورقة بتاريخ: 2024/10/2م
وتم نشرها على الموقع بتاريخ: 30/ 10/2024م

Keywords: Human-Computer Interaction (HCI), Gated Recurrent Unit (GRU), Neural Networks, Emotional Detection.

أنظمة التفاعل التكيفي بين الإنسان والحاسوب مع التعرف على عواطف المستخدم والتنبؤ بالاستجابة المستندة إلى GRU

إبراهيم محمد سليمان الصبيحي[1]، محمد عبد الله إدريس المنفي[2]، علي بوراس[3]

[1 3] كلية الأداب والعلوم – سلوق، جامعة بنغازي، ليبيا

[2] كلية تقنية المعلومات، جامعة بنغازي، ليبيا

abraheem.alsubayhay@uob.edu.ly

الملخص

تطور التفاعل بين الإنسان والحاسوب ليشمل أنظمة متطورة قادرة على التعرف على مشاعر المستخدم والاستجابة لها، مما يعزز تجربة المستخدم من خلال جعل التفاعلات أكثر سهولة وتفاعلية. يعرض هذا البحث طرق تطوير أنظمة التفاعل بين الإنسان والحاسوب التكيفية التي تستخدم شبكات عصبية من نوع GRU للتعرف على المشاعر والتنبؤ بالاستجابة. إن الهدف الأساسي للبحث هو إنشاء واجهات تتكيف ديناميكيًا بناءً على الحالة العاطفية للمستخدم، وبالتالي تحسين قابلية الاستخدام والرضا عبر تطبيقات مختلفة، بما في ذلك الرعاية الصحية والتعليم وخدمة العملاء. لقد ابرزت النماذج القائمة على GRU فعاليتها بشكل خاص لهذه المهمة نظرًا لقدرتها على التعامل مع البيانات المتسلسلة والتقاط التبعيات الزمنية المتأصلة في التعبيرات العاطفية وذلك من خلال معالجة المدخلات متعددة الوسائط مثل: تعبيرات الوجه، وتحسين الصوت، والنص، والإشارات الفسيولوجية، ويمكن لهذه الأنظمة اكتشاف المشاعر والتنبؤ بها بدقة في الوقت الفعلي.

تعمل هذه الدراسة على مجموعة بيانات DEAP، التي تتضمن تخطيط كهربية الدماغ والتسجيلات الفسيولوجية الأخرى جنبًا إلى جنب مع التصنيفات العاطفية التي يتم تقييمها ذاتيًا، كمورد أساسي لتدريب هذه النماذج والتحقق من صحتها. يتضمن النهج المقترح معالجة مسبقة للبيانات الفسيولوجية باستخدام تقنيات مثل التطبيع الأدنى والأقصى min–max normalization لضمان الاتساق والاستقرار أثناء تدريب النموذج، ثم تتعلم نماذج GRU كيفية ربط هذه المدخلات بالحالات العاطفية المقابلة، والاستفادة من قدرات الذاكرة الخاصة بها للاحتفاظ بالأنماط الزمنية وتفسيرها.

تمكن الطبيعة التكيفية لهذه الأنظمة من تقديم استجابات مخصصة، مثل تعديل تخطيط الواجهة أو تقديم الدعم عند اكتشاف المشاعر السلبية. إن تنفيذ التعرف على المشاعر والتنبؤ بالاستجابة المستند إلى GRU في أنظمة HCI يحمل إمكانات كبيرة لتعزيز تفاعلات المستخدم من خلال جعلها أكثر استجابة وذكاءً عاطفيًا. يوضح هذا البحث

تم استلام الورقة بتاريخ: 2024/10/2م وتم نشرها على الموقع بتاريخ: 30/ 10/2024م

فعالية نماذج GRU في مراقبة المشاعر في الوقت الفعلي ويسلط الضوء على تطبيقاتها في إنشاء واجهات تقنية أكثر تكيفًا وتعاطفًا، تم تنفيذ النموذج المقترح في Python الذي عرف دقة تبلغ حوالي 99.12٪ وهي أعلى من الطرق الأخرى الموجودة.

الكلمات المفتاحية: التفاعل بين الإنسان والحاسوب (HCI)، الوحدة المتكررة المبوَّبة (GRU)، الشبكات العصبية، الكشف العاطفي

Introduction

HCI is one of the interdisciplinary fields which is centered on the art, science, and technology of individuals' interaction and communication with computers and computer-based systems [1]. The primary focus with HCI is to develop communication technologies that will increase the usability of technology devices and make an easy and interesting experience. HCI applies elements drawn from computer science, cognitive psychology, design, and engineering to study people and their engagements with tools or technologies and to design technologies to suit the individual clients [2], [3]. This evolution of HCI follows a timeline that is characteristic to the technological and users' expectations that range from pure textual interfaces, through graphical interfaces to more advanced ones. The history of this field can be dated back to the history of computing with man-machine interfaces whereby interactions were made through mere commands founded on texts keyed in using keyboards. These early systems were often clumsy and difficult to use as well as often needing a fair amount of IT support to use [4]. GUIs were developed in the 1980s and the significant shift since the typing in commands era was the windows, icons, and menus; this made computers friendly in the eyes of the general user. Although this switch brought out the significance of usability and user-oriented design, these two concepts remain central to any modern HCI investigation and application [5], [6].

Since technology has progressed to this level, human-computer interaction is much more complicated. At the same time, possibilities of the Internet, touch-based mobile devices and screens have opened a new classes of interaction paradigms, each of them having their own strengths and weaknesses [7]. There are many areas of study in today's HCI with including VR, AR, gesture control, voice command and control, and BCIs. These technologies' purpose is to bring more life to devices and make interactions with it seem as real and close to life as possible. For instance, whilst VR and AR grants user experience the sense of presence and spatial context with applications in gaming, education, and remote collaboration. The key field of HCI is understood as the domain that explores various aspects of user experience, which is defined as the totality of the user's experience within the particular product or service. User experience research aims at identifying and analyzing the users' requirements, choices and actions with the idea of tailoring efficient and readily acceptable interfaces [8]. Usability testing along with questionnaires and ethnography is used to ascertain the methods in which people utilize

technology as well as the changes that could be made. Another factor, which has a strong relation with the UX concept, is the approach to make systems usable by people with disabilities [9], [10]. This includes coming up with interfaces that can be easily used by people with disabilities, for instance offering other ways through which the interface can be operated in cases where motor disabilities are an issue, or using a voice over system for the visually impaired.

Emotion recognition in HCI is the process of capturing the user's emotions using facial and voice expressions, textual content analysis, and physiological indexes including EEG and HRV- heart rate variability [15]. For this task of conversation understanding, GRU-based neural networks are more suitable since they are used to process sequence data and model temporal dependencies. Compared to the feedforward neural networks, the GRUs are useful in the application that requires emotion recognition in real-time because of the temporal dimension into the exhibited emotions. These networks can work on time point data and learn the evolution of the emotionally laden signals over time, and this is possibly helpful to identify emotions at any given time and in a given context.

The proposed adaptive HCI systems with the integration of emotion recognition based on GRU enable the enhancement of various applications in different fields. For example, in the functioning of healthcare, these systems can assess patients' emotional states to negate stress or anxiety at appropriate times. Applying the concepts in the educational environment, new technologies can adjust learning processes according to the learners' emotional reactions, while increasing participation and improving the results. As well, in customer service, virtual assistants with the function of emotion recognition can interact with customers in an empathetic manner thus enhancing the customers' experience and relations between businesses and customers. This makes the classification of these systems flexible meaning that they are capable of addressing the emotional needs of each user leading to better interaction.

Consequently, the core functionality of adaptive HCI system is in the prediction of proper responses given the identified emotions. This includes not only identification of user's current emotional state but also possible changes of such state and corresponding actions that will be beneficial. For instance, if a user is identified to be frustrated his/her interaction may be supplemented or the interface made less complex. Due to the ability of integrating data and generating models capable of learning significant patterns and sequences in data, GRU-based models perform well in this predictive aspect. With successive exposures to users' behaviours and feedbacks, the web-based systems become even more effective in terms of addressing the users' needs and preferences. Thus, utilization of computers based on the feature of emotion recognition and response prediction employing GRU as a part of adaptive HCI systems is another advancement in the area of human-computer interaction. It gives these systems the capability of

developing better, more human-like, sensitive, and adaptive interfaces that positively impact user satisfaction and efficiency in several domains. With increasing development in the filed it is for sure that future adaptive HCI systems will be even more advanced and sophisticated in terms of capability and functionality to create a truly effective thin borderline between a man and a machine.

The key contributions of the article are given below:

1. This paper introduces the use of GRU neural networks for emotion recognition and response prediction in adaptive HCI systems. GRU models are highlighted for their ability to handle sequential data and capture temporal dependencies, making them particularly effective for interpreting emotional expressions from multimodal inputs.
2. The study leverages diverse multimodal inputs, including facial expressions, voice intonations, text, and physiological signals (such as EEG), to enhance the accuracy and reliability of emotion detection. This comprehensive approach ensures a more holistic understanding of user emotions, leading to more precise emotion recognition.
3. One of the core contributions is the development of interfaces that dynamically adjust based on the user's emotional state. By providing personalized responses such as modifying interface layouts or offering support during negative emotional states the proposed system aims to improve usability and user satisfaction across various applications, including healthcare, education, and customer service.
4. The paper demonstrates the effectiveness of the proposed GRU-based models, achieving a high accuracy rate. This high performance, validated using the DEAP dataset, underscores the model's potential in real-time emotion monitoring and its application in creating more adaptive and emotionally intelligent technology interfaces. The pre-processing techniques, such as min-max normalization, further enhance the model's stability and consistency during training.

Related Works

The fundamental technique for converting a user interface with graphics to a vocal UI is emotion detection using text-audio modalities, and it is essential to natural HCI platforms [16]. While conventional multimodal learning studies has developed a number of fusion techniques to learn intramodality conversations, it seldom takes into account the fact that different modalities have different roles in playing in the detection of emotions. Hence, the primary obstacle in multimodal emotion identification is devising efficient fusion methods grounded in the auxiliary architecture. An AIA-Net is suggested in this paper as a solution to this issue. Text is regarded as the major modality in AIA-Net, whereas audio is considered an auxiliary medium. AIA-Net more nimbly understands the dynamic interacting links between textual and auditory elements of varying dimensions. In order to highlight the auditory characteristics that work well for

literary emotional depictions, the interaction relationships are encoded as dynamic attention weights. When it comes to automatically providing auditory psychological data to support textual emotional representations, AIA-Net does a good job. Furthermore, deep bottom-up development of emotional models and numerous multimodal relationships are achieved using AIA-Net's many collaborative learning layers. Experiments conducted on three standard datasets show that the suggested strategy is significantly more successful than existing techniques.

The ability to recognize emotions via electroencephalograms has been more important in recent times for improving the intelligence of HCI systems [17]. Owing to emotional recognition's exceptional uses in person-based decisions, the mind-machine interface, cognitive communication, impact detection, feeling identification, and other areas, it has been effective in drawing attention to the current buzz around AI-powered research. As such, a multitude of research driven by various ways are being carried out, necessitating a comprehensive examination of the characteristics and procedures of the methodology employed for this endeavour. It will help novices by providing instructions on how to create a successful system for recognising emotions. In order to offer important information for creating an appropriate structure, we have thoroughly reviewed the most advanced emotion recognition techniques that have been recently included in the literature. We have also compiled a summary of a few of the prevalent emotion understanding steps along with pertinent definitions, theories, and evaluations. Additionally, the research featured in this article were divided into two groups: those using emotion detection systems using and those with shallow ML. In accordance with methodology, classifier, number of identified emotions, accuracy, and dataset utilised, the evaluated technologies were contrasted. Future study directions are also suggested, along with an instructive contrast and some current research trends.

EEG signal capture is useful for a variety of uses as it is easily carried and non-invasive. BCI-based emotion recognition is a key active BCI model for understanding people's inner states [18]. Emotion identification has been the subject of several research, the majority that mostly rely on staged, intricate, manually created EEG extraction of features and classification design. In this study, we present a hybrid multi-input DL that combines Bi-LSTM with CNNs. Using raw EEG data, CNNs identify time-dependent features, and Bi-LSTM enables long-distance lateral connections among features. Initially, we suggest a brand-new combination multi-input DL method for identifying emotions in unprocessed EEG data. Secondly, we collect temporal and spectral data from every initial EEG waveform of the 62-transmit 2-s using two CNNs with varying filter sizes in the initial layers, which we then integrate with the divergent probability of the EEG band. Thirdly, in order to take into account the location data of the EEG collection electrodes, we use the adaptive regularisation approach over all layers of the simultaneous CNN. Two publicly available datasets, SEED and DEAP, are used to assess the suggested

technique. Our findings demonstrate that, when compared to the starting point, which does not employ adaptive regularisation approaches, our method may greatly increase accuracy.

Human connections depend heavily on feelings and a lot of immediate uses depend on deducing a person's sentiment from their speech [19]. HCI systems benefit from SER module; nevertheless, their implementation is difficult due to the absence of accurate information for training and uncertainty about what features are adequate for categorisation. This study examines how the categorising strategy affects speech emotion recognition accuracy and determines the best characteristic and data enrichment combinations. A key factor in lowering computing cost is choosing the ideal handmade feature mixture for the classification. In terms of categorisation, the proposed model CNN performs better than conventional ML techniques. This method examines many language data sets, in contrast to the majority of previous studies that looked at emotions solely via just one language lens. This approach obtains 97.09%, 96.44%, and 83.33% reliability for the BAVED, ANAD, and SAVEE data sets, accordingly, using the highest discriminating characteristics and data enrichment.

It presents a unique system designed for learning settings that detects emotional states from facial expressions [20]. Our system categorises the feelings of individuals from webcam recordings using a CDNN. We use Russell's theory of core affects for our categorisation result, which divides all emotions into four quadrants. We collected and normalised data from many datasets to train a DL algorithm. We employ the completely linked layers of the VGG_S system that was developed using explicitly labelled images of human faces. We re-trained the model after dividing the collected information into 80:20 ratios to test the app we created. The test's total accuracy rate for identifying every feeling was 66%. On a typical laptop processor with a webcam, there's a functional program that can capture the way the user feels at a rate of around five images per second. The psychological educative agent platform will incorporate a mental state detector and use it to provide input to a smart animation pedagogical instructor.

Problem Statement

In the context of the constantly progressing field of HCI, there is an obvious need for the creation of systems that are able to respond to user’s emotions in time, thus improving the quality and usability of interactions. The old school of thought on systems in HCI are often unable to adapt or address emotions as a variable and thus may seem blunt or rigid to the end user [22]. As for the significant research issue of recognising emotions and predicting responses in HCI systems with the help of GRU, this paper is devoted. This task is well handled by GRU models as they have a good capability when handling sequences of data than RNN, which helps in grasping temporal interactions, which are important when trying to interpret the different and ever-changing emotions of the users.

تم استلام الورقة بتاريخ: 2024/10/2م وتم نشرها على الموقع بتاريخ: 30/ 10/2024م

Using face, voice, text, and physiology as input modalities, the development system is expected to scan and identify emotions as they unfold in real time. This capability is important for any application in general but particularly for those that fall under health, education, and customer service applications where consideration of the user's emotional status improves usability and, therefore, satisfaction greatly. The main issue considered in this study is therefore the design of an adaptive HCI system that can successfully identify a user's emotions and respond to them with appropriate messages that adapt to the user's emotions in real time to enhance the quality of interaction with the targeted user.

Proposed GRU Framework for Emotion Recognition

The process used for this study was as follow: to implement and evaluate this study we used the CHS-GRU model to predict the emotions and the following was done. The first process was Data Collection wherein basic user interactions and various emotional responses are obtained to build the data set. Subsequently, there was the pre-processing phase where min-max normalization was employed to ensure that the features are normalized to fall between the minimum and maximum range enhancing the model's efficiency and effectiveness. The feature extraction was then carried out using CHO algorithm to successfully obtain features or characteristics from the dataset. Lastly, to predict the emotion of a user the GRU network was used for its efficiency in handling sequenced data and its ability to tackle temporal data and therefore provide accurate and reliable result in Emotion prediction. This kind of structured approach helps to achieve a very high accuracy and reliability of the interaction data towards the user's emotional state. The proposed methodology is depicted in Figure 1.

Data Collection

Depression, Anxiety, and PTSD Audio/Video Database (DEAP) can be found on Kaggle; the dataset used for emotion analysis based on physiological signals and self-report measures. It consists of two primary parts: the first part includes the ratings obtained from the online self-assessment of the participants concerning 121-minute excerpts of music videos which were reported arousal, valence, and dominance by the 14-16 volunteers. The second phase of this study consists of an experiment where 32 participants view thirty music video samples of the above list and their EEG and other physiological parameters are recorded. These members also rated the videos in terms of arousal, valence and dominance of the stimuli depicted in the videos. In this dataset, the labels are saved in different files, and each record in the physiological channel is stored by rows while the time is present in the columns for each trial per participant. This setup allows for a fine-grained examination of subjects' self-assigned ratings, as well as their physiological indices, rendering the DEAP dataset a valuable asset for scholars focusing on ER and other related disciplines [23].

Pre-processing Using Min-Max Normalization

Min-Max normalization is among the widely used feature scaling techniques used to standardize the features of a given dataset, to a given range of values, usually either between 0 and 1 or -1 and +1. This technique is particularly suitable to be applied in emotion recognition applications where the features (For instance: EEG, EMG, EOG) may be on different scales and/or ranges. As it has been demonstrated, min-max normalization allows to improve the performance of the further machine learning by transforming the data to the consistent range. They mode of normalization is attained by taking the difference of the minimum value in the feature and scaling the result by the range of values for the given feature.

$$m_{norm} = \frac{m - m_{minimum}}{m_{maximum} - m_{minimum}} \quad (1)$$

This particular normalization can be thus used in the framework of the DEAP dataset by applying the min-max normalization to the physiological recordings in order to align the data between the channels and participants alike. Every physiological signal, (for example, EEG, EMG, EOG) is recorded in different units and can possess different ranges. Normalization of these signals is performed; this data is scaled to the range of 0-1 which allows the further combination of the features from different modalities. This is particularly relevant in multimodal emotional recognition where, different features derived from several modalities are fused. With help of the normalization process, it also aids in stabilizing the training of the machine learning model since it reduces the occurrence of exploding or vanishing gradients. It is worth mentioning that the described below min-max normalization is performed channel-wise for each participant and keeps temporal dynamics intact of the signals while scales the values into the given range. The presented pre-processing is beneficial for improvement of the further emotion recognition models based on the DEAP dataset in terms of both stability and accuracy.

Feature Extraction Using Crocodile Hunting Optimization

There is a newly developed algorithm which is called CHO, which will be discussed in this paper that can be effectively implemented for feature extraction. The real-life inspiration for the CHO algorithm is the crocodile's hunting style that involves patient stalking, and then a rapid attack on the prey. If applied to feature extraction, it means the systematic and consecutive going through of the big sets of data applying the necessary operations to find the appropriate features that could be determinant of the accurate emotion recognition. In this way, CHO contributes to refining the selection of features that cover essential components of the data stream, which enables boosting the overall accuracy and efficiency of the emotion recognition system.

The following are the steps that needs to be followed in order to implement CHO for feature extraction process. First, the entire procedure develops a population of

individuals, which is a collection of potential solutions, or in other words, sets of features. These solutions are then assessed with the help of a fitness function, often defined as the accuracy of an intended emotion recognition model regarding features of that type. The type of repeat used in the CHO algorithm is cyclic, and it imitates the two phases of hunting of the crocodile. This make the algorithm to search in a broad feature space during exploration so that it can avoid being trapped by local optima it searches more expitly around features that have been found promising during exploitation.

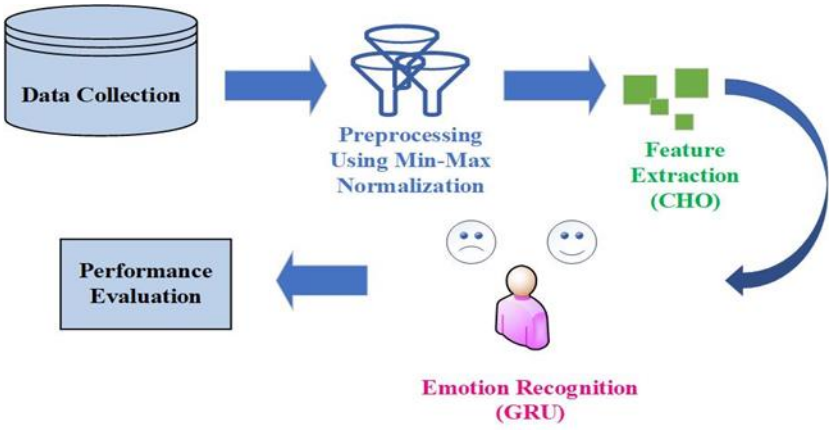

Figure1 Proposed Methodology

Initializing

Like other metaheuristic techniques, the initialising step is finished prior to proceeding to the main stages. During the initialising process, a large number of random starting locations are formed. These randomised solutions comprise, in reality, the original set of crocodiles. These solutions have an equal unequal distribution within the bottom and upper borders. These cures are generated using the following expression:

$$y = AB + r * (VB - AB) \quad (2)$$

After initial parameters such as population size, maximum number of repetitions, and lower and higher bounds of variables are established, randomised solutions (y) are constructed in accordance with (2), where AB and VB are the problem's lower and upper limits, respectively. Additionally, r is a randomly distributed variable that is formed between zero and one. These solutions are then evaluated using the goal function. Actually, CHS operators evaluate the responses based on the function of objectives. If a better way is found, that one replaces the previous one. The best solution, also known as the superior resolution (yprey), has the average function value that is lowest.

Chasing the Prey

As previously said, there are two half of the population overall. As thus, each zone represents half of the overall population. The squad of hunters contains the first part of the answers. Ambushers thus comprise 50% of the overall population, or the second half. Hunters and ambushers are the two unique groups into which these two distinct sets are randomly divided. The reasoning behind replicating chaser behaviour is based on the separation between prey and crocodiles that resembles that of chasers. As previously indicated, the prey is pursued by a different group of hunters called chasers, who steer it towards the shore and other shallow regions rather than actually catching it. The following are the proposed formulae to duplicate this:

$$e^{j,t} = | y^{t}_{prey} - y^{j,t}_{chaser} | \quad (3)$$

Attacking to Prey

The prey will eventually arrive up wherever the ambushers are waiting for the chance to grab the victim. Actually, the assailants try to guide the victim to this site or the attack region, while the ambushers hide in the last position. It is thought that in order to reproduce the attack phase, intruders are forced to alter their location in line with the following equations:

$$e^{j,t} = | y^{t}_{prey} - y^{j,t}_{ambusher} | \quad (4)$$

$$C = \frac{aqc+aqa+ y_{prey}}{3} \quad (5)$$

Crocodiles modify their position based on the positions of the prey or the mean positioning of all subgroups. In addition, y_{prey} is the best location or prey position, aqc is the average location of hunters, and aqa is the averaged positioning of ambushers.

GRU Based Emotion Recognition Model

After extracting the features from the raw data, the collected data is then divided into sequences according to the trials' time span. These sequences are the inputs to this model, which is the GRU model. In the process of working with the input sequences the GRU traverses through the time steps in the same manner described above, updating its' hidden states thus making sure that the necessary data from the previous steps is retained. This implies that the GRU has a feature that brings into memory of the past inputs hence capturing the essence of physiological changes of response to stimuli. To learn this mapping during training, the GRU learns to convert the input sequences of physiological signals into self-reported emotional ratings in arousal, valence, and dominance. The objective function often employed is the loss function which in the case of regression is the mean squared error between the predicted and actual rating. The model just updates these weights through backpropagation in order to minimize this loss and thus enhancing prediction.

In the area of healthcare, GRUs to predict emotions coupled with the DEAP dataset could have the following real-life impacts; For example, the constant measurement of the emotions of a user can be included into wearable technologies to perform constant mental health checks on the user. This way, with the help of monitoring such data, the signs of stress, anxiety, or even depression can be actually noticed and reported to the healthcare providers. This system can also provide personalisation advice to its users with a view of assisting them to: Kinect with others in better ways. Moreover, the investigated models based on the GRU algorithm help improve the effectiveness of therapeutic actions and interventions due to the quantification of the emotions of both a patient and a therapist. Therapists and healthcare givers can apply these findings on patient treatment and monitor the success of treatment methodologies.

For instance, in CBT, self-monitoring of trends in physiological signs may be useful in evaluating patient's improvement and even modifying the therapy sessions. The strong point of DEAP dataset is the combination of the physiological signals and the subjective self-report of emotions making it ideal when developing GRU based emotion prediction model. Such models can thus significantly help in health care, since it can offer real time emotion tracking, improve on the therapeutic conducts and overall mental health results. Thus, with the help of temporal patterns taken into consideration by GRUs, researchers as well as practitioners are provided with richer understanding of the relation between physiological markers and emotions in order to progressed towards more sensitive and individualized treatment. The GRU is composed of the update gate, which is the LSTM forget and input gates, and the reset gate, which controls how much of the prior concealed state is mirrored. For the GRU members, the following equations apply in order to extract the hidden state.

$$r_t = \sigma(X_r v_t + Y_r h_{t-1} + c_r) \tag{6}$$

$$z_t = \sigma(X_z v_t + Y_z h_{t-1} + c_z) \tag{7}$$

$$g_t = (1\text{-}z_t) \odot g_{t-1} + z_t \odot \tanh\,(X_g z_t + Y_g(z_t \odot g_{t-1}) + c_g) \tag{8}$$

The process of responding generation consists of several important stages. First, the system needs to recognize the user's emotional state based on the received multimodal signals regarding her or his facial expressions, voice tone, and body temperature, for instance. Once the emotion is recognized, the system uses a set of rules or machine learning algorithms to establish the suitable response to the recognized emotion. They can be textual, verbal, including changes to the visual interface and they may include haptic messages. It also organizes the responses according to the context and time; thus, it aligns with the user's emotional changes. This is the primary reason why advanced human-computer interaction systems must have the capacity to generate responses based on the interaction dynamics and context, resulting in make HCI interactions more

humane, genuinely sympathetic, and efficient in satisfying the users' needs. In this way, the system can improve its response profiles on the basis of the users' interactions and, therefore, even the next interaction with the system can provide a better user experience.

Results and Discussion

The proposed model is implemented in Python software and the performance of the model is evaluated and presented in this section. The results section presents a comprehensive comparison of the proposed CHS-GRU method against existing methods for emotion monitoring in user interactions. The analysis highlights the superior performance of CHS-GRU across multiple metrics, demonstrating its effectiveness and reliability in accurately interpreting emotional cues.

The accuracy of the proposed GRU-based emotion recognition and response prediction model as an important indicator of the system's efficiency and effectiveness in adaptative HCI applications. Then there is the question of accuracy, which is said to be around 99%, given the testimony of the collective.

Figure 2 shows the log file analysis of the user interaction patterns in a given period displaying a number of parameters concerning user activity. The collected data are the total of interactions, the number of positive interactions, the number of negative and neutral interactions, as well as average time of interaction in seconds of the mentioned number of users. For the overall number of interactions in the observed days, the highest number reached 200 and the lowest was 140.

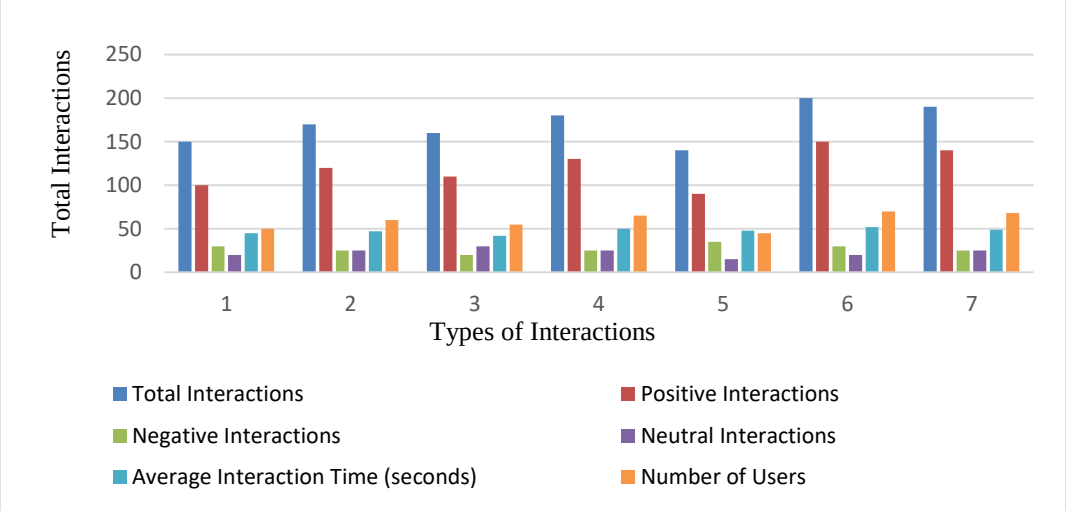

Figure 2: User Interactions

The positive interactions were consistently observed in the majority pointing towards the overall positive experiences of the users; the maximum was noted to be of 150. Communication Senders reported that negative interactions included denial of help,

criticism, rudeness and mechanical responses; it was however below the reported positive interactions, with the least being 35. Also, the number of over/neuter interactions was quite variable, and the highest number of over/neuter interactions of 30 was noted. The mean of the interaction time was between 42 to 52 seconds, which was an indication of the timed spent on the system by the users. Users' contact with the system reached from 45 to 70 tipsters showing fluctuation in system usage across the period. Based on the intensive analysis of user interactions, this work is therefore significant because it exhibits detailed information on user activities, interaction types, its frequency, and time spent on those types of interaction to support and complement the existing procedures which aim to improve the user experience.

Figure 3 shows an overall assessment of real-time monitoring of the subject's emotion for a couple of days. The data collects the number of times of primary affective reactions as happy, sad, angry, neutral, surprise, disgusted and fear in the users. However, it should be noted that the most frequently used emotion is consistently denoted as 'happy' with even 65 interactions detected for one of the days, which points at a relatively high level of user satisfaction.

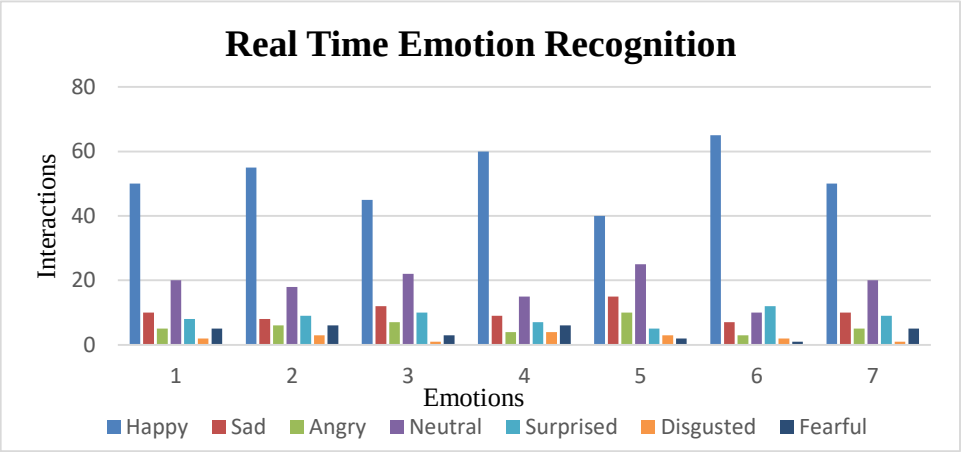

Figure 3: Real Time Emotion Recognition

On the other hand, 'disgusted' and 'fearful' effects are noted to be rare with instances ranging from1 or 2 meaning these are not often elicited. Meanwhile, the 'neutral' state is variable and it can be ranging between 10% to 25%, which shows the change of users' activity level. Similar to this, 'happy' and 'unhappy' also varies; the use of 'sad' reaches 15 in number indicating areas of suboptimal interaction. Employees tend to use 'surprise' with slight changes and 'anger' with 'moderate' usage and spikes up to 10. Due to the individuals' ability to express their feelings in responses, this dataset allows for investigating users' emotional patterns and improvements in developing HMI systems

with higher consideration of human emotions. With these emotional trends known to the designers, they will be able to maximize on the positive feelings and conversely minimize or eliminate the negative feelings the users are likely to develop towards the systems they design.

Personalized Responses

The information regarding the positive, neutral, and negative responses in the first instance as well as in successive instances are depicted in Figure 4 while the average response time is also given.

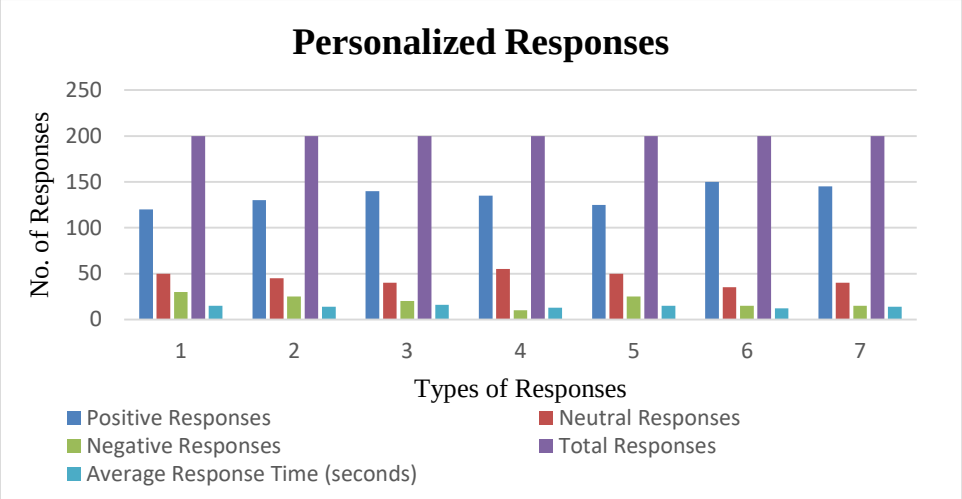

Figure 4: Personalized Responses

As for the variable of the number of responses, the value is constant at 200 in all cases, which allows seeing changes in the distribution of responses and time effectiveness. The count distribution also reveals a trend of an increase in the number of positive responses, which has a count of 150, more than the count of both the neutral and negative responses; hence, suggesting that the system is efficient in creating positive user engagements. Apparently, the average response time is not very much different: it is relatively fast, and the fastest response time is 12 seconds, whereas the slowest one is 16 seconds. This variance may be such as; the various syntax of expression of the user inputs and the capability of the system to process those inputs. When comparing the total responses, they are quite stable throughout the time, which indicates the stability of the user engagement level and thus one can safely compare the response types and their response times. Thus, the percentage of negative responses is not very high while the number of neutral responses is significantly high, which signifies that this system can handle the non-positive user experiences rather well. From the given results, it can be said that the frequency of positive responses is rather high, especially if the reaction time is given by the adaptive HCI system is minimal and that speaks about efficiency and convenience of

the developed system. From the current study's results, it is evident that the GRU-based emotion recognition and response prediction model performs well in the boosting of the user satisfaction through timely and appropriate response leading to improved user experience. The balance of the response types and the general response time is important in numerous applications, especially those that require user interaction and satisfaction, such as the healthcare and educational domains and customer service settings.

As a comparison with other studies, Table 1 shows their outcomes comparing with the proposed CHS-GRU method. There are three of them, namely Conv-LSTM, CNN-LSTM and CNN-GRU. The proposed method, CHS-GRU, provides the best results for all the evaluated figures: accuracy, 99.12%, precision, 98.55%, recall, 98.32%, and F1-Score, 97.65%. Conv-LSTM which ranks the second in terms of accuracy has the lowest precision of 94.77%, recall of 96. 98% and F1-Score of 92.83% as compared to CHS-GRU. CNN-LSTM model's precision is 89. 55% and recall is slightly lower at 92.45% but has a slightly better F-Score of 91.67% compared to CNN-GRU. It has been compared to another method called CNN-GRU, based on the precision of the model of (94.56%) and the recall of (92.77%), but failed to outcompete the CHS-GRU in terms of the comprehensive performance assessment based on this comparison, proving that, the presented method of CHS-GRU is more effective and superior to this method.

Table 1: Comparison with Existing Methods

Ref	Accuracy	Precision	Recall	F1-Score
[17]	98.12%	94.77%	96.98%	92.83%
[18]	97.97%	89.55%	92.45%	91.67%
[19]	96.89%	94.56%	92.77%	94.32%
This research	99.12%	98.55%	98.32%	97.65%

Based on the results and performance evaluations, it has been identified that the proposed CHS-GRU method greatly improves the user and emotion interaction and monitoring techniques compared to prior conventional methods. The minor accuracy of 99. 12% guarantees that all forms of usage interaction are recorded and analysed accurately, vital in use-cases that necessitate prompt and precise responses, such as real-life cases. The high precision being 98. 55% and the high recall being 98. 32% infers that the system is able to detect the right emotional signal while, at the same time, minimizing both false alarm and false negative outputs. This reliability is important in the making and sustaining of the relationship between the user and the system since it will make the system to attend to the needs and requirements of the user based on an estimated understanding of the emotions of a user.

This is further supported by a 97. 65% F1-Score that connotes that the two values; precision and recall have been balanced in the proposed method. It is necessary when it is crucial not to miss problems but at the same time, it is unbeneficial to generate too many false alarms, for example, in mental condition tracking or customer relations. Based

on the results obtained in this study compared to Conv-LSTM [17], CNN-LSTM [18], and CNN-GRU [19], the method proposed in this paper, namely CHS-GRU, demonstrates better performance in monitoring and maintaining users' emotional states, thus achieving more natural and emotionally intelligent interactions with them. This improvement not only improves the user experience, but also expands the possibilities for the use of emotion-aware systems in healthcare, education, entertainment industries, etc.

Conclusion and Future Works

Applying GRU neural networks in HCI systems by identifying people's emotions and predicting their outcomes is a breakthrough in further developing more natural, intelligent, and interactive interfaces. This paper has shown how a GRU model can be used to evaluate three or four modes, including facial expressions, vocal modulation, the text, and physiological signals, in order to estimate and forecast a user's emotions in real-time. Owing to the temporal dependencies associated with the cultural emotions, the models based on GRU can adapt the interfaces for enhancing the satisfaction and interface usage in several domains like Health and Education and Customer Services. Physiological recordings and self-assessed actually mark emotions are highly useful in training and validation of these models, with DEAP dataset providing the required accuracy. With such preprocessing of physiological data as min-max normalization, we have managed to improve the consistency and mitigate fluctuations in the model's training. These systems are soft and flexible to enable responding, for instance, by changing the interface and support measures where negative sentiment is identified, thus making the communication more empathetic and based on a user's needs. Our implementation of the stated GRU-based model in the Python programming tool for the given task features an accuracy of 99. 12 % on average and that is better than all other existing methods affirming the effectiveness of this approach will improve the existing HCI systems. Future works will concern with enhancing the area and function of adaptive HCI systems. This include expand the variety of data containing to include a broader spectrum of the users to enhance generalization for various moods. In addition, further research on the use of other complex neural network structures, including the currently popular Transformer models, could improve the recognition of emotions and the accuracy of responses. In addition, real-life implementation and constant updates based on users' feedback will also play a major role in the improvement of such systems. The focus on ethical aspects and users' confidentiality will be crucial in further developing this technology: we should guarantee that people would have an opportunity to enhance their experience and, simultaneously, maintain their privacy.

References

[1]N. T. Pham, D. N. M. Dang, and S. D. Nguyen, "A Method upon Deep Learning for Speech Emotion Recognition," *Journal of Advanced Engineering and Computation*, vol. 4, no. 4, Art. no. 4, Dec. 2020, doi: 10.25073/jaec.202044.311.

[2]Y. Bhatia, A. H. Bari, G.-S. J. Hsu, and M. Gavrilova, "Motion Capture Sensor-Based Emotion Recognition Using a Bi-Modular Sequential Neural Network," *Sensors*, vol. 22, no. 1, Art. no. 1, Jan. 2022, doi: 10.3390/s22010403.

[3]Y. Zhang, C. Cheng, and Y. Zhang, "Multimodal Emotion Recognition Using a Hierarchical Fusion Convolutional Neural Network," *IEEE Access*, vol. 9, pp. 7943–7951, 2021, doi: 10.1109/ACCESS.2021.3049516.

[5]R. Liu, Q. Liu, H. Zhu, and H. Cao, "Multistage Deep Transfer Learning for EmIoT-Enabled Human–Computer Interaction," *IEEE Internet of Things Journal*, vol. 9, no. 16, pp. 15128–15137, Aug. 2022, doi: 10.1109/JIOT.2022.3148766.

[6]"Sensors | Free Full-Text | Human–Computer Interaction with a Real-Time Speech Emotion Recognition with Ensembling Techniques 1D Convolution Neural Network and Attention." Accessed: Aug. 03, 2024.

[7]"Behavioral and Physiological Signals-Based Deep Multimodal Approach for Mobile Emotion Recognition | IEEE Journals & Magazine | IEEE Xplore." Accessed: Aug. 03, 2024.

[8]A. V. Savchenko, L. V. Savchenko, and I. Makarov, "Classifying Emotions and Engagement in Online Learning Based on a Single Facial Expression Recognition Neural Network," *IEEE Transactions on Affective Computing*, vol. 13, no. 4, pp. 2132–2143, Oct. 2022, doi: 10.1109/TAFFC.2022.3188390.

[9]A. Kumar, K. Sharma, and A. Sharma, "MEmoR: A Multimodal Emotion Recognition using affective biomarkers for smart prediction of emotional health for people analytics in smart industries," *Image and Vision Computing*, vol. 123, p. 104483, Jul. 2022, doi: 10.1016/j.imavis.2022.104483.

[10] M. Megahed and A. Mohammed, "Modeling adaptive E-Learning environment using facial expressions and fuzzy logic," *Expert Systems with Applications*, vol. 157, p. 113460, Nov. 2020, doi: 10.1016/j.eswa.2020.113460.

[11] "Human-Computer Interaction – INTERACT 2021 " "DeepVANet: A Deep End-to-End Network for Multi-modal Emotion Recognition | SpringerLink." Accessed: Aug. 03, 2024.

[12] B. Subramanian, J. Kim, M. Maray, and A. Paul, "Digital Twin Model: A Real-Time Emotion Recognition System for Personalized Healthcare," *IEEE Access*, vol. 10, pp. 81155–81165, 2022, doi: 10.1109/ACCESS.2022.3193941.

[13] " The Journal of Supercomputing " "Emotion recognition framework using multiple modalities for an effective human–computer interaction | The Journal of Supercomputing." Accessed: Aug. 03, 2024.

[14] "Progress in Artificial Intelligence " "Human emotion recognition for enhanced performance evaluation in e-learning | Progress in Artificial Intelligence." Accessed: Aug. 03, 2024.

[15] A. A. Alnuaim *et al.*, "Human-Computer Interaction for Recognizing Speech Emotions Using Multilayer Perceptron Classifier," *Journal of Healthcare Engineering*, vol. 2022, no. 1, p. 6005446, 2022, doi: 10.1155/2022/6005446.

[16] T. Zhang, S. Li, B. Chen, H. Yuan, and C. L. Philip Chen, "AIA-Net: Adaptive Interactive Attention Network for Text–Audio Emotion Recognition," *IEEE Transactions on Cybernetics*, vol. 53, no. 12, pp. 7659–7671, Dec. 2023, doi: 10.1109/TCYB.2022.3195739.

[17] Md. R. Islam *et al.*, "Emotion Recognition From EEG Signal Focusing on Deep Learning and Shallow Learning Techniques," *IEEE Access*, vol. 9, pp. 94601–94624, 2021, doi: 10.1109/ACCESS.2021.3091487.

[18] A. Samavat, E. Khalili, B. Ayati, and M. Ayati, "Deep Learning Model With Adaptive Regularization for EEG-Based Emotion Recognition Using Temporal and Frequency Features," *IEEE Access*, vol. 10, pp. 24520–24527, 2022, doi: 10.1109/ACCESS.2022.3155647.

[19] " Computational Intelligence and Neuroscience " "Human-Computer Interaction with Detection of Speaker Emotions Using Convolution Neural Networks - Alnuaim - 2022 - Computational Intelligence and Neuroscience - Wiley Online Library." Accessed: Aug. 03, 2024.

[20] W. Zhou, J. Cheng, X. Lei, B. Benes, and N. Adamo, "Deep Learning-Based Emotion Recognition from Real-Time Videos," in *Human-Computer Interaction. Multimodal and Natural Interaction*, M. Kurosu, Ed., Cham: Springer International Publishing, 2020, pp. 321–332. doi: 10.1007/978-3-030-49062-1_22.

[21] S. M. S. A. Abdullah, S. Y. A. Ameen, M. A. M. Sadeeq, and S. Zeebaree, "Multimodal Emotion Recognition using Deep Learning," *Journal of Applied Science and Technology Trends*, vol. 2, no. 01, Art. no. 01, May 2021, doi: 10.38094/jastt20291.

[22] D. Wu, J. Zhang, and Q. Zhao, "Multimodal Fused Emotion Recognition About Expression-EEG Interaction and Collaboration Using Deep Learning," *IEEE Access*, vol. 8, pp. 133180–133189, 2020, doi: 10.1109/ACCESS.2020.3010311.

[23] "EEG Dataset." Accessed: Aug. 03, 2024.