Human-Intelligent Systems Integration
https://doi.org/10.1007/s42454-026-00096-0

ORIGINAL PAPER

AI-driven motion-based emotion recognition in interactive virtual reality of equine-assisted services

Roya Haratian[1] · Ann Hemingway[1] · Fred Charles[1] · Xun He[1] · Joenam Coutinho[2] · Liucheng Guo[2]

Received: 22 December 2025 / Accepted: 8 March 2026
© The Author(s) 2026, modified publication 2026

Abstract
This paper explores how technology can replicate equine-assisted services without requiring physical interaction with horses. Human-horse interaction is utilised in equine-assisted therapy to improve mental wellbeing; however, access to such interventions is limited due to the high horses' cost and maintenance. Horses possess advanced emotional intelligence, interpreting human gestures and non-verbal cues to respond adaptively, often mirroring human emotions. By modelling these capabilities in a virtual environment using Artificial Intelligence (AI) and Machine Learning techniques, users can access the replicated services while interacting with virtual horses. A system is designed to detect participants' emotional states reflected in their motion via motion analysis and integrate these cues into an interactive virtual reality (VR) experience to foster calmness and improve communication skills. An experiment was conducted to train an emotion-recognition model for replicating horse-human interaction in VR. Results demonstrate the system's accuracy in detecting emotional states and the findings suggest the potential of VR-based equine-assisted services to replicate the real ones to enhance accessibility and engagement with such services.

Keywords Motion capture · Machine learning · Virtual reality · Emotion recognition · Mental wellbeing · Artificial intelligence

1 Introduction

Equine-assisted therapy services leverage human-horse interaction in improving mental health and wellbeing. The bond between humans and horses is utilised in equine-assisted services (EAS) interventions to reduce incidents of domestic violence (Hemingway and Sullivan 2022), reduce anxiety and increase levels of calmness (Hemingway 2019). However, not everyone can have access to such services, as horses are expensive and high maintenance and for their own wellbeing need to live outside in green spaces with other equines. In this paper, it is explored how technology can be utilised to roll out the mental health benefits of such interactive equine-assisted services by using technology and interacting with virtual horses.

In terms of the mechanism of action of EAS (Hemingway et al. 2019), horses have a sophisticated sense of awareness and an ability to perceive human emotions and intentions. Their emotional intelligence makes them adept at interpreting non-verbal cues from humans, such as gestures and postures, enabling them to adjust their behaviour in response. From such cues, horses can sense human emotions such as fear, anxiety, happiness, or calmness and will respond to requests if the human is calm, assertive and consistent. This project considered if this body language-based emotional communication process could be reproduced in the virtual world where the virtual horses would respond to requests when the human playing the game becomes calm assertive and consistent in their body language. Furthermore, during EAS, participants are taught to understand and respect

✉ Roya Haratian
rharatian@bournemouth.ac.uk

[1] Science and Technology Faculty, Bournemouth University, Poole, UK

[2] TG0, London, UK

Published online: 17 April 2026

the horse's perspective as a prey animal and that the horse requires clear, non-aggressive communication and body language control from the human to build a method of communication. The intervention is structured around participants' learning to communicate with the horses using body language and involves a series of horse-human interaction games, which aim to foster calmness and improve empathy, assertiveness and communication in participants. The project reported here started to explore whether this process can be replicated by use of virtual technology.

Development of technology in which emotional state is detected from user's motion will provide input to an interactive VR adapting and responding to the user's emotional cues. The interactive VR will adjust the intensity and content of the experience based on the user's emotional reactions, creating a more effective and engaging mental health intervention. The face-to-face equine-assisted intervention has proved to be highly effective for individuals and families socially excluded due to poor psychosocial skills (Hemingway and Sullivan 2022). This innovation will enable access to an effective intervention at mass scale through affordable technology. The impact will be demonstrated through development and uptake of this innovative approach to mental health intervention delivery.

Using technology to address social and mental health issues benefits from the aspect that not only the technology is scalable and can provide valuable feedback on users' cognitive and emotional reactions to various situations, but also it enjoys enthusiastic uptake by young people, who are particularly vulnerable and need social services support. Particularly, virtual reality (VR) is a technology that is still in growth path with interest from customers depending on available and affordable options. It is estimated that over 171 million people use VR technology globally as of 2025. VR users mainly use this technology for gaming, education, entertainment, and healthcare purposes (Kumar 2025).

Use of VR in healthcare is increasing for instance Food and Drug Administration (FDA) in the USA authorised the usage of VR treatment solution RelieVRx to help patients manage their chronic lower back pain (Tolochkov 2025). While it is mostly seen as an alternative to current treatment measures, VR technology has the potential to transform mental health care with its capabilities to create and provide immersive experiences (Rodrick 2025; Kuleli et al. 2024). The global market for VR therapy in mental health was valued at USD1.83 billion in 2024. This market is expected to grow significantly, reaching USD10.71 billion by 2032, with an average annual growth rate of 24.7% (Data Bridge 2025).

In this paper, after a background review about equine-assisted therapy services and VR technology in addressing mental wellbeing, it is explored how emotions can be modelled informing development of a model in recognising user's emotions. In Sect. 3, an experiment is designed and conducted for a pilot study to train a model for replicating human-horse interaction in VR; and the process of model development is discussed. Results are presented in Sect. 4 demonstrating how accurate the developed system can recognise participants' emotion while interacting with virtual horses. The findings are concluded in Sect. 5 on how the equine-assisted therapy can be replicated using VR technology to make it accessible for more users without the need for real horses and to benefit from leading to wider utilisation of such an approach.

2 Background

Equine-assisted therapy services leverage human-horse interaction in improving mental health and wellbeing. The bond between humans and horses is utilised in equine-assisted services (EAS) interventions. Quantitative research indicates that equine assisted learning specifically (which is the type of EAS intervention that this VR pilot was modelled on (TheHorseCourse 2025) can significantly enhance emotional regulation across a variety of participant groups. Reported benefits include reductions in domestic violence in families with multiple problems, (Hemingway and Sullivan 2022) improved social competence in children, (Pendry and Roeter 2013) and improved behavioural outcomes (Pendry et al. 2014). Hopson et al. (2023) also found notable gains in self-efficacy among young people with disabilities participating in equine assisted learning programmes.

Qualitative studies further support these findings. Evidence suggests that equine assisted learning can strengthen self-confidence and self-esteem in young people (Burgon 2011), and enhance social, cognitive, and psychomotor skills in school-aged children (Stoppard 2024). These improvements may be linked to the sense of emotional safety fostered within equine assisted learning environments, with themes such as self-esteem, connection, and respect highlighted by Cagle-Holtcamp et al. (2019).

Taken together, this body of research positions equine assisted learning as a promising approach for supporting social development and emotional regulation across diverse populations. Future studies could explore additional client needs that may benefit from this method, as well as assess the effectiveness of alternative delivery formats (e.g., virtual reality-based programmes).

VR therapy is set to significantly change mental healthcare by providing an engaging and immersive method for

treating various psychological conditions, helping patients achieve substantial progress. VR therapy represents a cutting-edge tool with the potential to transform a patient's mental well-being journey (Mihajlovic 2025). The rising number of mental health disorders like anxiety, post-traumatic stress disorder (PTSD), and depression, along with a demand for new treatments without medicine, is boosting the market for VR therapy in mental health. This growth is also helped by better VR technology making therapies more engaging and effective, and by more people accepting digital health options like telehealth (Data Bridge 2025).

Using VR in mental health treatment and assessment makes the process more engaging experiences. As VR is interactive, patients are usually less resistant to VR than traditional tests and assessments. This approach is especially effective for younger patients or anyone who usually resist mental health assessments and treatments by turning the experience into an enjoyable one (Roddick 2025). Virtual Reality Exposure Therapy (VRET) is one of the therapeutic options that is used to treat patients with anxiety orders and psychiatric conditions in an affordable and accessible way (Mihajlovic 2025). According to Mihajlovic (2025), these are some examples of VRET usage customised for the needs of patients including simulating crowded social environments to help patients gradually become comfortable with interactions to treat social anxiety disorders, creating virtual scenarios that expose patients to their specific fears and phobias in a controlled manner, allowing clinicians to adjust parameters like intensity and duration or even personalise environments to target specific patient triggers, and providing peaceful virtual natural settings for relaxation and stress reduction.

There is evidence of an increasing demand for VR-based solutions for mental health treatment in academic literature, industry reports and user feedback. The increasing interest in VR therapy is evident from the over 7 thousand research papers published since 2015, with 1,412 of those coming out in just 2024 (Silber 2025). Successful examples are Psious, BehaVR, OxfordVR and DeepVR with increased interest for VR applications to treat anxiety, depression, PTSD, and stress management. Psious merged with XRHealth uses VR environments to create controlled scenarios for therapy, such as simulated heights or public speaking situations; BehaVR and OxfordVR merged companies as RealizedCare accrued by XRHealth are providing experience of social situation in VR (XR Health 2025); and DeepVR provides experience of underwater world controlled by deep breathing in rising and falling movement (Deep VR 2025).

Challenges of using VR games for therapies are linked to the need for further personalisation in VR therapies instead of generic scenarios (May 2024), affordability of VR therapy due to cost of VR headsets and scalability of solutions (Cushnan et al. 2024), and lack of integration with traditional therapies in which therapists monitor, adjust and interact with patients in real time (May 2024). Therefore, informing this research by the need for integration of the VR to the existing therapies with personalised features which are addressed in this paper.

2.1 System modelling

Addressing mental health through the VR technology, there is a need to explore how emotions are represented, recognised, and modelled using the technology to address personalisation of the VR experience. From a psychological perspective, emotions are represented through three emotion models: discrete (categorical), dimensional and appraisal. Continuous prediction of emotions is achievable through utilising algorithms to find out the correlation between a user's physiological signal changes and the emotion in a real-time manner.

In the categorical model, emotions are represented by discrete labels such as happiness, anger, disgust, sadness, anxiety, and surprise which are the basic emotions (Ekman, 1971). The mathematical modelling of logistic regression for discrete classifications of the six states is to take an input vector $\mathbf{x}$ and to assign it to one of k discrete classes C_k where $k=1, \ldots, K$ (Murphy 2012). Therefore, it is necessary to know the probabilities of the classes given the input data, which are given by $p(C_k|\mathbf{x})$. Using Bayes' theorem, these probabilities can be expressed in the form of (1).

$$p(C_k|\mathbf{x}) = \frac{p(\mathbf{x}|C_k)\ p(C_k)}{p(\mathbf{x})} \tag{1}$$

Whereas any of the quantities appearing in Bayes's theorem can be obtained from the joint distribution $p(C_k, \mathbf{x})$. In this equation, $p(C_k)$ is the prior probability for the class C_k and $p(C_k|\mathbf{x})$ is the corresponding posterior probability. The conditional probabilities can be modelled by representing them as parametric models and then optimising the parameters using a training set. A generative approach can be adopted in which models the class-conditional densities given by $p(\mathbf{x}|C_k)$, together with the prior probabilities $p(C_k)$for the classes and then computing the required posterior probabilities using Bayes' theorem. Minimum probability of making a mistake is obtained if each value of $\mathbf{x}$ is assigned to the class for which the posterior probability $p(C_k|\mathbf{x})$ is largest. This is equivalent to the minimum misclassification rate decision rule, which assigns each value

of $\mathbf{x}$ to the class having the higher posterior probability $p(C_k|\mathbf{x})$. Therefore, solving the optimization problem, the discrete classification can be achieved.

In the dimensional model (Fig. 1), emotions are labelled in a two- or three-dimensional space where three independent and bipolar dimensions, valance (pleasure-displeasure), degree of arousal and dominance-submissiveness are necessary and sufficient to adequately define emotional states (Russell 1980).

The mathematical modelling for this purpose can be seen from Linear Regression viewpoint that is predicting the value of one or more continuous target variables $\mathbf{t}$ given the value of a dimensional vector $\mathbf{x}$ of input variables. Therefore, making predictions for the target variable $\mathbf{t}$ given some new value of the input variable $\mathbf{x}$ on the basis of a set of training data. The input values are represented as $\mathbf{x} = (x_1, \ldots, x_N)^T$ and the corresponding target values are represented as $\mathbf{t} = (t, \ldots, t_N)^T$. Specifically, there is need to fit the data using a polynomial function of the form (2).

$$y(x, \mathbf{w}) = \omega_0 + \omega_1 x + \ldots + \omega_M x^M = \sum_{j=0}^{M} \omega_j x^j \tag{2}$$

Whereas $y(x, \mathbf{w})$ is the polynomial function, which is a nonlinear function of $\mathbf{x}$, and a linear function of the coefficients $\mathbf{w}$; M is the order of polynomial and $\omega_0, \ldots, \omega_M$ are polynomial coefficients. The values of the coefficients will be determined by fitting the polynomial to the training data. This can be done by minimizing an error (cost) function which is the sum of the squares of the errors as (3).

$$\frac{1}{2} \sum_{n=1}^{N} \{y(x_n, \mathbf{w}) - t_n\}^2 \tag{3}$$

Gradient descent algorithm which is a first-order optimization algorithm can be utilised to find a local minimum of the error function that takes steps proportional to the negative of the gradient (or of the approximate gradient) of the function at the current point. Gradient descent is also known as steepest descent in the literature. The classification can be considered from Bayesian curve fitting perspective whereas a Bayesian treatment simply corresponds to a consistent application of the sum and product rules of probability. It allows the predictive distribution to be written in the form fully Bayesian requires integrating over all values of $\mathbf{w}$ as (4). By maximising the posterior distribution over the parameters, it is possible to determine $\mathbf{w}$ by finding the most probable value of $\mathbf{w}$ given the data; therefore, achieving the classification goal.

$$p(t|x, \mathbf{x}, \mathbf{t}) = \int p(t|x, \mathbf{w})\, p(\mathbf{w}|\mathbf{x}, \mathbf{t})\, d\mathbf{w} \tag{4}$$

In the appraisal model, emotions are defined as processes where appraisal of significance of an event is shown by emotion quality, intensity and duration while considering dynamics of emotion as discussed in Scherer et al. (2013; Fig. 2). To mathematically represent the appraisal model, dynamic system theory could be utilised that provides the language in which embodied and situated stances can be developed into a scientific theory. Dynamic field theory (DFT) is a mathematical and conceptual framework which was developed to model embodied human cognition (Haratian, 2018).

The basic computational element of DFT is Dynamic Neural Field (DNF). The response of a neural population is modelled through a mathematical representation of DNFs. The dynamics of DNF are mathematically formulized as follows. In (5), $u(x,t)$ is the activation of DNF over dimension x, to which the underlying neural population is

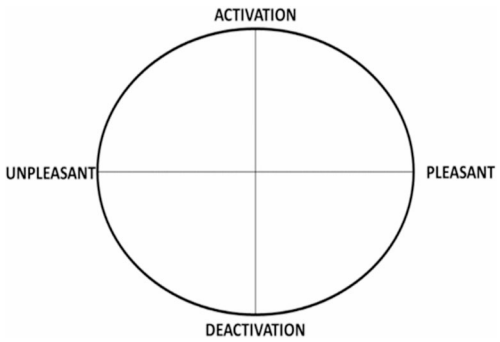


Fig. 1 Dimensional model of emotions (Russell 1980)

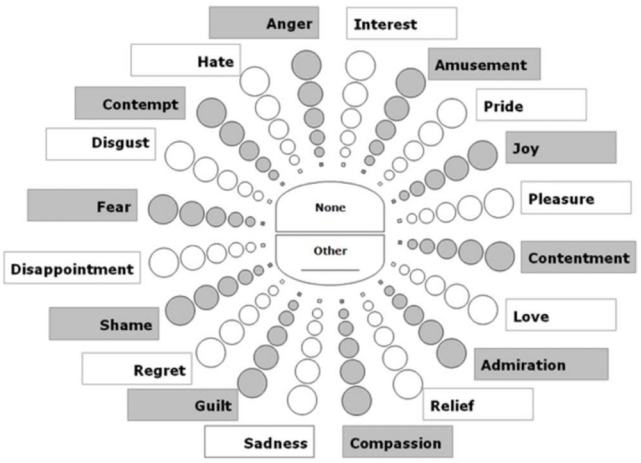


Fig. 2 Geneva emotion wheel (Scherer et al. 2013)

responsive. h is a negative resting level and $S(x,t)$ is an external input. The lateral interactions in DFT are modelled through a short-range excitation and a long-range inhibition where σ_{exc}, σ_{inh}, c_{exc} and c_{inh} are width and amplitude of the excitatory and inhibitory parts of interaction kernel as (6). The $f(u(x,t))$ in (7) shapes the output of DNF through a sigmoid function where β is the slope of sigmoid (Haratian 2018).

$$\tau \dot{u}(x,t) = -u(x,t) + h + \int f(u(x,t'))\omega(x-x')dx' + S(x,t) \tag{5}$$

$$\omega(x-x') = c_{exc}\exp\left[-\frac{(x-x')^2}{2\sigma_{exc}^2}\right] - c_{inh}\exp\left[-\frac{(x-x')^2}{2\sigma_{inh}^2}\right] \tag{6}$$

$$f(u(x,t)) = \frac{1}{1+exp[-\beta\, u(x,t)]} \tag{7}$$

The model parameters are τ as the time constant of each neuron, β as the slope parameter of output function for each neuron, as well as the coupling coefficients and kernel excitation and inhabitation parameters. After initialisation of model parameters, a cost function will be calculated using the training data. The model simulator estimates the quality and intensity regarding the input training data of multiple trials as (8) and (9).

$$Cost_{quality} = \frac{1}{T*L}\sum_{l}\sum_{t}\left|\widehat{Q}_l(t) - Q_l(t)\right|^2 \tag{8}$$

$$Cost_{intensity} = \frac{1}{T*L}\sum_{l}\sum_{t}\left|\widehat{I}_l(t) - I_l(t)\right|^2 \tag{9}$$

To identify model parameter, the cost functions need to be minimised. The cost function is calculated for training trials and the average of all Mean Square Errors, MSEs, of multiple trials is used as the function which should be minimized to derive the model parameters. In (8) and (9), the model quality and intensity estimates are represented by $\widehat{Q}_l(t)$ and $\widehat{I}(t)$ while the targeted ones are represented by $Q_l(t)$ and $I_l(t)$.

3 Methodology

To develop an artificial intelligence (AI)-AI-driven human-horse interaction in VR to replicate equine-assisted services, a pilot study is conducted to design a motion-based emotion recognition system learning from participants' motion while interacting with virtual horses. For this purpose, a VR environment is developed in which horses can behave differently based on the controls applied to the system (Fig. 3).

3.1 Experiment setup and design

The VR headset of Meta Quest 3 which is a standalone device developed by Reality Labs, is utilised for the VR experience with rendering via Meta Quest App AirLink (Meta and quest 2025). Environment is designed and implemented in Unity3D (Unity 2025) which is a cross-platform game engine developed by Unity Technologies and simulation was run from Lenovo laptop with AMD Ryzen 7 CPU, 16GB RAM and RTX 3070 laptop GPU. The full-body motion capture framework used in this study is powered by TG0's etee trackers (etee Tracker 2025), which provide low-latency, high-fidelity motion data through a proprietary BLE streaming architecture for seamless interaction with VR. Participants were equipped with 12 etee trackers strategically placed on the wrists, shoulders, hips, knees and ankles to monitor full-body motion (Fig. 4). Each tracker measures $20\times50\times37.5$ mm, weighs 60 g, and contains a 700-mAh LiPo battery with USB-C fast charging. The trackers communicate via BLE using a USB-C dongle compatible with SteamV Base Station 1.0 and 2.0, enabling precise and stable motion capture throughout the VR sessions.

The etee ecosystem supplied the Unity3D integration modules and drivers that enabled real-time skeletal reconstruction (Fig. 5) and seamless connectivity with the Meta Quest VR environment. These hardware–software components ensured accurate body-language capture and stable long-duration data collection across all participants. As shown in Fig. 5, the motion-capture system generates a

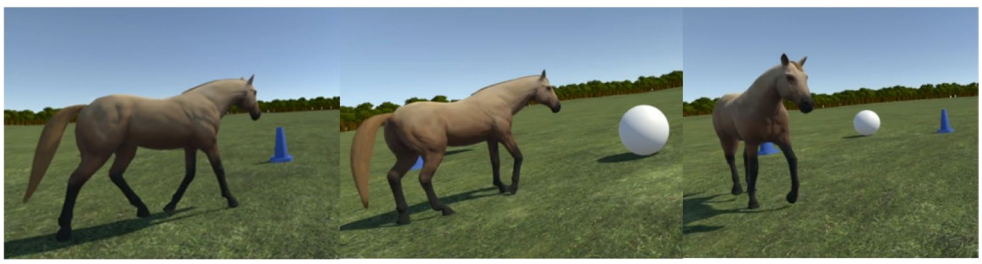


Fig. 3 Simulated environment with a virtual horse in low arousal state by lowering the head to more arousal state by raising head and rearing from left to right, respectively



Fig. 4 Motion capture setup with on-body markers and the VR headset

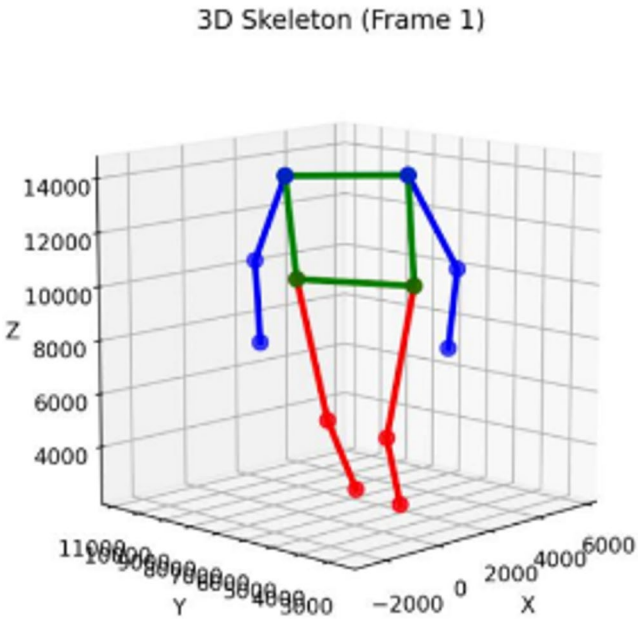


Fig. 5 The stick figure generated as the result of motion capture

full-body stick-figure model based on the positional data from the etee trackers. The input signals to the Machine Learning (ML) model consist of the 3D positions of all trackers expressed relative to a reference point, defined as the trackers attached to the participant's body.

An experiment is designed for a pilot study in which participants interact with the virtual horses, rooted in the idea that horses are highly perceptive to human emotion. To build trust and alignment, humans are encouraged to approach the horses in a calm state of emotion. Based on the model developed by TheHorseCourse (2025), while interacting with the virtual horses, the participants will be either in the Calm state which shows a relaxed posture and minimal movement, signalling alignment with the horse, or in the Neutral (Ask) state which is active but composed gestures to navigate the horse in specific direction, or Tense which is sharp or anxious movements, causing tense movement such rearing from the horse. These states are mapped to the dimensional model on the vertical axes showing the level of arousal from low arousal or deactivation to high arousal and activation as indicated by Calm, Neutral(ask) and Tense, respectively in TheHorseCourse model. The virtual horse in low arousal (deactivation/calm) state would lower the head to more arousal state by raising head and rearing (activation/tense).

Ethical approvals are in place for this research, and participants provided consent in attending the experiment. Six participants, three male and three female, attended the experiment and participants were briefed on how to identify and express their emotional state during the session. As they announced their emotion, the observer would label the experienced emotion on a scale from low (deactivation) to high (activation) arousal. In reciprocation, the virtual horse would demonstrate postures such as head lowering and grass eating when relax for low arousal and towards more tense state, they would be rising their head and rearing for high arousal.

3.2 Signal processing and machine learning

The system processes body motion signals collected from full-body trackers, applying pre-processing, feature extraction, and data augmentation before feeding it into ML models to predict the participants' emotion. The primary model, a Multi-Layer Perceptron (MLP), captures temporal patterns to classify emotional states. For comparison, three baseline models are used including a Long Short-Term Memory (LSTM) network, a Two-Stream MLP and a Random Forest classifier.

Motion data from the etee full-body tracker were synchronised and down-sampled from 200 Hz to 60 Hz. Joint positions were transformed into a hip-centred coordinate system to normalise inter-participant differences, and noise artefacts caused by occlusions or signal spikes were removed via threshold-based filtering and manual inspection. Segments with ambiguous emotion labels were excluded.

The cleaned dataset consisted of 2,137 overlapping 4-second windows (Calm: 569, Neutral: 1,242, Tense: 326). The model evaluation used Leave-One-Subject-Out (LOSO) cross-validation to ensure subject-independent testing. To address the class imbalance without data leakage, augmentation was applied only to the training data within each LOSO fold. Minority classes were up sampled using geometric coordinate shifting, whereby a random constant 3D shift was applied to joint positions within each window.

Additional strategies, including applying Gaussian noise, random rotation of the skeleton about the hip joint and hip-joint centring, were implemented but not activated during training. The post-augmentation step included each training fold balanced to the majority class, yielding an average of approximately 1,035 windows per class, while test sets remained unaltered.

For each 4-second window, a set of motion features was extracted to capture the dynamics of body posture and movement. These included joint velocity, joint acceleration, and relative angles between connected body segments. Statistical descriptors such as mean, standard deviation, and temporal variance were computed for each feature dimension. To simplify learning and reduce ambiguity, the continuous emotion ratings collected during the experiment were grouped into three representative categories—Calm, Neutral, and Tense—mapped to the lower, middle, and upper levels of the arousal axis. The resulting feature vectors were used as inputs to the subsequent machine-learning models. Multiple machine-learning models were evaluated to identify the most effective approach for classifying emotional states from motion data. A Naive Bayes classifier was first implemented to establish a baseline and assess the separability of the dataset. More advanced models were then explored, including a Multi-Layer Perceptron (MLP) with three hidden layers (256, 128, 64 units), a Two-Stream MLP that separately processes kinematic and postural features (each stream uses a small MLP to extract feature representations), a Long Short-Term Memory (LSTM) network designed to capture temporal dependencies (two hidden layers, 128 units), and a Random Forest classifier with 100 estimators.

All models were trained using class-weighted loss functions to account for imbalance across emotional categories. To prevent over-fitting, early stopping and learning-rate scheduling were employed throughout the training process. All models were trained using only statistics from the training split. Loss functions were weighted to handle class imbalance. Early stopping and learning rate scheduling were used to prevent over-fitting. For each held-out user, the performance measures for the machine learning techniques are recorded accuracy, F1-score, precision, recall, and confusion matrices. To evaluate generalizability across individuals, and Leave-One-Subject-Out (LOSO) was employed for cross-validation.

4 Results

The evaluation metrics in machine learning, especially for classification tasks are accuracy, F1-score, precision, recall, and confusion matrices. Accuracy is defined as the proportion of correctly predicted instances out of all instances as

Table 1 Confusion matrix

		Predicted	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN

Table 2 Binary classification performance (Mean ± Standard Deviation)

Model	Accuracy (Mean ± SD)	Precision (Mean ± SD)	Recall (Mean ± SD)	F1 (Mean ± SD)
MLP	71.94% ± 4.81%	63.64% ± 5.56%	67.58% ± 5.61%	63.20% ± 4.97%
Two-Stream MLP	71.99% ± 4.12%	61.69% ± 8.83%	68.15% ± 4.57%	62.60% ± 7.17%
Random Forest	71.44% ± 4.49%	54.56% ± 13.24%	58.32% ± 17.44%	58.84% ± 9.95%
LSTM	71.40% ± 6.39%	60.77% ± 10.59%	68.35% ± 6.25%	62.89% ± 9.23%

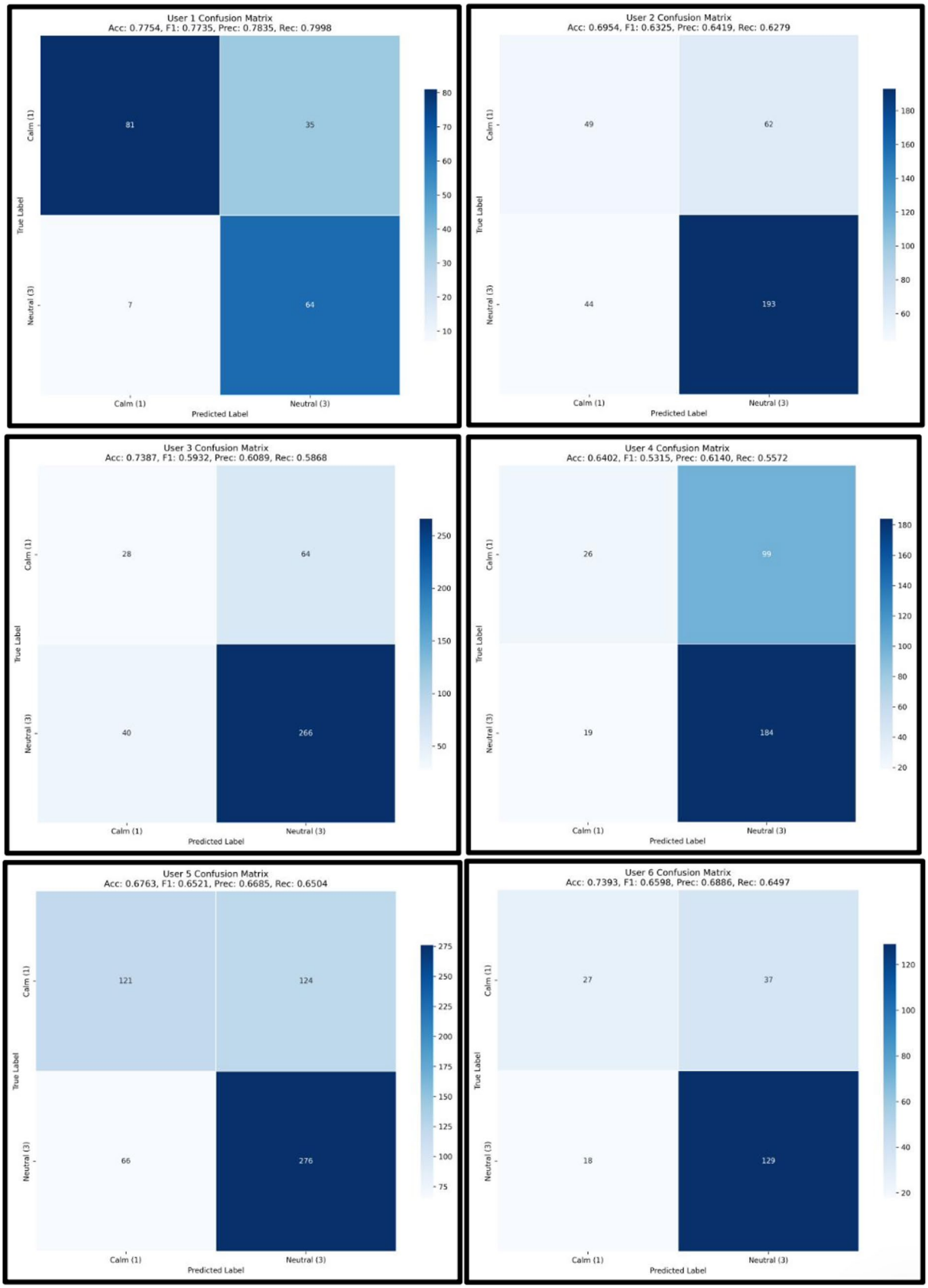


Fig. 6 Confusion matrix for each LOSO (binary classification)

(10) in which TP is True Positives, TN is True Negatives, FP is False Positives and FN is False Negatives. Such metric is suitable when classes are balanced but misleading for imbalanced datasets.

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \quad (10)$$

Precision, which is about of all predicted positives, how many are actually positive is defined as (11) and important when the cost of false positives is high.

$$\text{Precision} = \text{TP} / (\text{TP} + \text{TF}) \quad (11)$$

Recall (Sensitivity) is about of all actual positives, how many is predicted correctly and defined as (12) and is important when missing a positive is costly.

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) \quad (12)$$

F1-Score is Harmonic mean of Precision and Recall and is defined as (13) and is used when balances precision and recall, useful for imbalanced datasets.

$$\text{F1} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (13)$$

Confusion Matrix is a table showing counts of predictions vs. actual classes and has structure as Table 1 and provides a detailed view of classification performance beyond a single metric.

The Naive Bayes classifier was used as an initial baseline due to its simplicity. As expected, it achieved relatively low performance, indicating that emotional-state differences in whole-body movement cannot be captured effectively through simple statistical assumptions. This baseline confirmed the need for more expressive models (MLP, Two stream MLP, LSTM and Random Forest) capable of handling temporal and spatial dependencies in motion data. Further manual cleaning was also conducted to improve label consistency—removing samples where ambiguous movements, such as trembling, introduced confusion for the model. This helped boosting the overall reliability and clarity of the classification results.

A binary classification experiment was conducted to evaluate how well each model distinguishes between Calm and Neutral states. Table 2 summarises the results using LOSO cross-validation. The Two-Stream MLP, which processes kinematic (Acceleration, velocity) and postural (joint position) features in parallel before feature fusion, achieved the highest overall accuracy and strong recall performance, indicating its effectiveness in capturing complementary movement characteristics. The confusion matrix for the Two-Stream model (Fig. 6) across all the 6 users provides a detailed breakdown of prediction results.

Multi-Class Classification is used for the more challenging three-class emotion recognition task, encompassing Calm, Neutral, and Tense states, the performance of all models is summarised in Table 3. The MLP achieved the highest overall accuracy among the models. The confusion matrix for the MLP model (Fig. 7), aggregated across all six users, provides a detailed view of the predictions for each emotional state, highlighting which classes were most frequently misclassified.

In the three-class emotion recognition task (Calm, Neutral, Tense), the MLP achieved the highest overall accuracy, with performance patterns consistent across models (Table 3). Per-class evaluation and the confusion matrices (Fig. 7) reveal that the Neutral state functions as a statistical attractor, receiving 62% of Calm misclassifications and 35% of Tense predictions. The class-level analysis confirms that Tense is the most challenging class, reflected in lower recall scores and its tendency to be absorbed into Neutral. These class-specific dynamics contribute to a conservative macro-F1 (~52%), indicating that while models accurately identify Neutral behaviour, the boundary between Calm and Tense remains more subtle.

With the current dataset, the model performs as reported. However, there are considerations for future work based on the findings in this study. Improved data collection protocols in briefing participants more thoroughly to familiarise them on how better to control the horses. Expanded dataset would also help by including more participants and balanced coverage across all emotional states, particularly tense. Adaptive learning also helps exploring models that personalise predictions over time, adapting to individual movement patterns. Furthermore, multi-modal sensor fusion for integration of data from additional sources (e.g., heart rate, or audio) will complement body movement and improve emotional state detection.

Table 3 Multi-class classification performance (Mean ± Standard Deviation)

Model	Accuracy (Mean ± SD)	Precision (± SD)	Recall (± SD)	F1 (± SD)
MLP	66.60% ± 7.44%	63.60% ± 14.55%	55.09% ± 13.12%	52.45% ± 14.52%
Two-Stream MLP	65.27% ± 6.90%	63.04% ± 8.12%	58.51% ± 11.60%	56.31% ± 12.87%
Random Forest	64.82% ± 5.55%	58.74% ± 18.40%	53.66% ± 9.40%	48.43 ± 8.51%
LSTM	65.53% ± 8.20%	60.78% ± 13.26	57.28% ± 15.54%	54.17% ± 17.22%

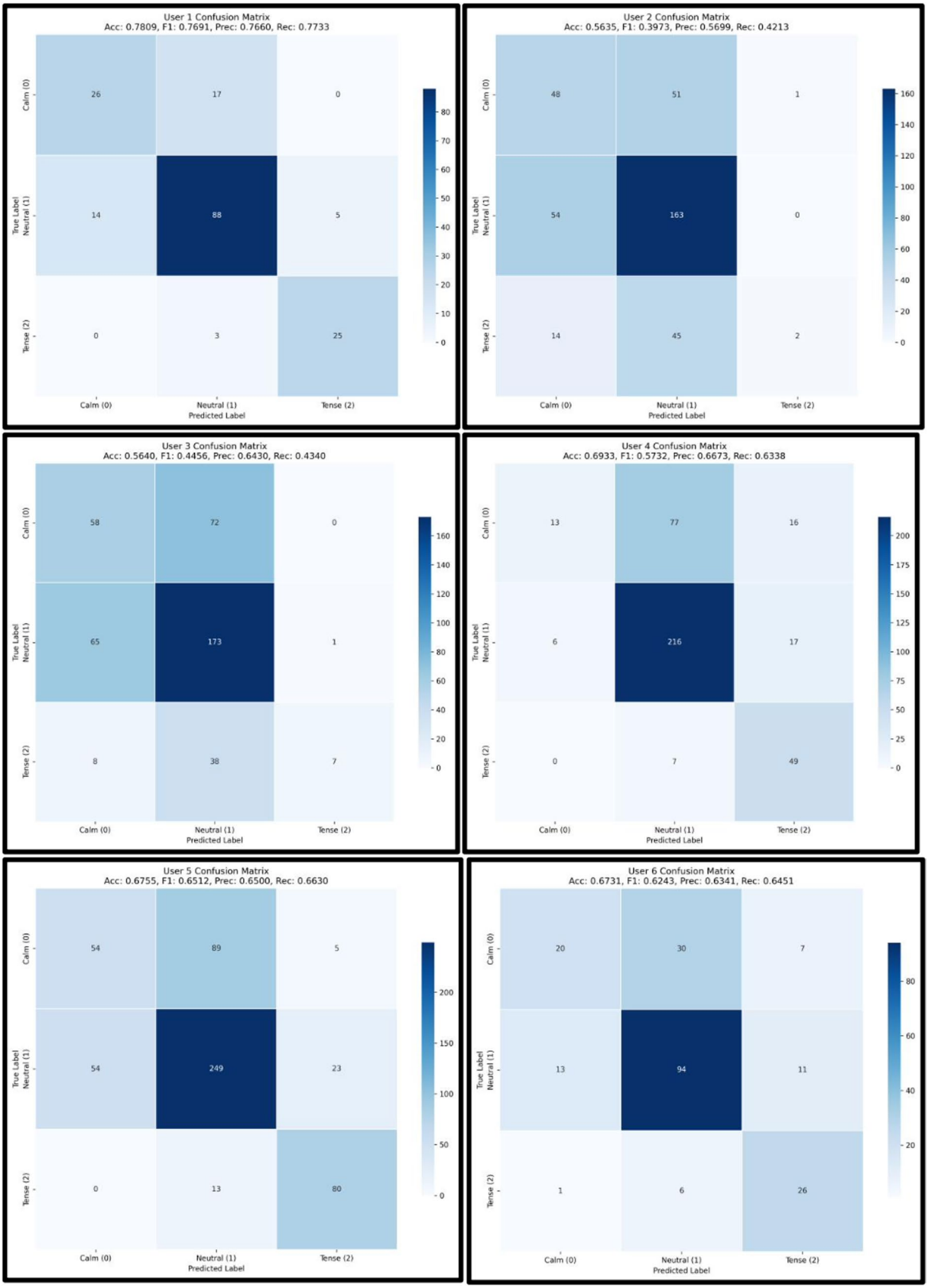


Fig. 7 Confusion matrix for each LOSO (multi-class classification)

5 Conclusion

This study explored the feasibility of replicating aspects of equine-assisted therapy through an AI-driven, motion-based emotion recognition system integrated into an interactive VR environment. In a pilot study, by combining full-body motion capture information with machine-learning models, the system was able to classify users' emotional states based on a technical foundation for emotion-state modelling from whole-body motion. Building on these results, the proposed study will expand the dataset to achieve more balanced representation across the emotional states for enabling a richer modelling of calm and tense patterns. Data collection protocols will be enhanced to ensure participants are thoroughly familiarised with tasks. In addition, the integration of multimodal sensor fusion from additional sources (e.g., heart rate, or audio cues) will increase sensitivity to capture subtle information providing richer signals than motion alone. Adaptive learning will further enable models to personalize the predictions over time capturing individuals' movement signation and improving long-term accuracy.

Acknowledgements The authors would like to acknowledge the contributions of Mel Hughes, professor of social work at Bournemouth University in informing the study from public health perspective, Ediz Akcay, senior academic in digital marketing at Bournemouth University in informing the market analysis, Paul Brown alumni of Bournemouth University in supporting the VR environment setup, and Harriet Lury, MBE and CEO of TheHorseCourse, in informing the VR horse interaction scenario based on their practice in TheHorseCourse charity.

Author contributions Roya Haratian: Project lead, system modelling, and experiment design. Ann Hemingway: Public health and experiment design. Fred Charles: Creative Technology and VR experiment setup. Xun He: Psychology Advisor. Joenam Coutinho: Motion capture and Machine Learning modelling. Liucheng Gue: Motion capture and Machine Learning.

Funding The project is funded by UK The Higher Education Innovation Funding (HEIF).

Data availability The supporting data would be available upon request.

Declarations

Disclosure of potential conflicts of interest Authors has no conflict of interest.

Research involving human participants and informed consent The participants were informed about the experiment procedure through participant information sheet, and they provided consent to take part in the experiment. Bournemouth University (BU) Ethical approval is in place and accepted by BU Ethics Panel.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Burgon HL (2011) Queen of the world': Experiences of 'at-risk' young people participating in equine-assisted learning/therapy. J social work Pract 25(02):165–183

Cushnan J, McCafferty P, Best P (2024) Clinicians' perspectives of immersive tools in clinical mental health settings: a systematic scoping review. BMC Health Serv Res 24(1):1091

Data Bridge (2025) Global Virtual Reality (VR) therapy for mental health market size, share, and trends analysis report – industry overview and forecast to 2032. DataBridge. Available from https://www.databridgemarketresearch.com/reports/global-virtual-reality-vr-therapy-for-mental-health. Accessed 10 June 2025

Ekman P, Friesen WV (1971) Constants across cultures in the face and emotion. J Personal Soc Psychol 17(2):124–129. https://doi.org/10.1037/h0030377

Etee Tracker (2025) Available from https://eteexr.com/products/new-eteetrackers. Accessed 10 June 2025

Haratian R, Timotijevic T (2018) On-body sensing and signal analysis for user experience recognition in human machine interaction. In: 4th International conference on frontiers of signal processing

Hemingway A (2019) A Study Exploring the Implementation of an Equine Assisted Intervention for Young People with Mental Health and Behavioural Issues. J 2:236–246. https://doi.org/10.3390/j2020017

Hemingway A, Carter S, Callaway A, Kavanagh E, Ellis S (2019) An exploration of the mechanism of action of an equine-assisted intervention. Animals (Basel). 9(6):303. https://doi.org/10.3390/ani9060303

Hemingway A, Sullivan K (2022) Reducing the incidence of domestic violence: An observational study of an equine-assisted intervention. Fam Process 61:549–570. https://doi.org/10.1111/famp.12768

Hopson L, Bailey JONR, Dugard J, McGraw K, Hanenberg C, Anderson D et al (2023) Feasibility Study of an Equine-Assisted Learning Intervention for Children with Disabilities. People Animals: Int J Res Pract 6(1):12

Kuleli D, Charles F, Guo L, Vuillier L, Liu CH, Gregory N, He X (2024) Exploring influence of social anxiety on embodied face perception during affective social interactions in VR. In: Proceedings of the 24th ACM international conference on intelligent virtual agents (IVA '24). Association for Computing Machinery, New York, NY, USA, Article 32, 1–5. https://doi.org/10.1145/3652988.3673952

Kumar N (2025) Virtual reality statistics 2025: users & trends. DemandSage. Available from https://www.demandsage.com/virtual-reality-statistics. Accessed 10 June 2025

May M (2024) How virtual reality therapy is shaping mental health. Nature Medicine. Available from https://www.nature.com/immersive/d41591-024-00032-2/index.html. Accessed 10 June 2025

Meta, quest (2025) Available from https://www.meta.com/gb/quest/. Accessed 10 June

Mihajlovic A (2025) Could virtual reality be a game changer for mental health? Sussex Publishers, LLC. Available from https://www.psychologytoday.com/us/blog/psychiatrys-think-tank/202503/could-virtual-reality-be-a-game-changer-for-mental-health. Accessed 10 June 2025

Murphy K (2012) Machine learning: a probabilistic perspective, MIT Press

Pendry P, Carr AM, Smith AN, Roeter SM (2014) Improving adolescent social competence and behavior: A randomized trial of an 11-week equine facilitated learning prevention program. J Prim Prev 35(4):281-293

Pendry P, Roeter S (2013) Experimental trial demonstrates positive effects of equine facilitated learning on child social competence. Human-Animal Interaction Bulletin 2013. https://doi.org/10.1079/hai.2013.0004

Roddick DG (2025) Understanding VR for Cognitive Behavioral Therapy (CBT): techniques and benefits

Rodrick G (2025) Revolutionizing mental health with VR therapy. Psytechvr Inc. Available from https://psytechvr.com/mental-health-and-vr-therapy. Accessed 10 June 2025

Russell JA (1980) A circumplex model of affect. J Personal Soc Psychol 39:1161–1178. https://doi.org/10.1037/h0077714

Scherer KR, Shuman V, Fontaine JRJ, Soriano C (2013) The GRID meets the wheel: assessing emotional feeling via self-report. In: JJR Fontaine, KR Scherer, C Soriano (eds) Components of emotional meaning: a sourcebook. Oxford University Press, Oxford. https://doi.org/10.1093/acprof:oso/9780199592746.003.0019

Silber D (2025) What Is the future of therapeutic VR and why does it matter? VR for health. Available from https://vrforhealth.com/2025/05/20/what-is-the-future-of-therapeutic-vr-and-why-does-it-matter/. Accessed 10 June 2025

Stoppard L (2024) Evaluation of equine assisted learning in education for primary school children: a qualitative study of the perspectives of teachers. Front Educ 9. Frontiers Media SA, p 1275280

TheHorseCourse (2025) Available from https://www.thehorsecourse.org/method. Accessed 10 June 2025

Tolochkov P (2025) The future of mental health: how AI, VR and AR are changing treatment. Forbes Media LLC. Available from https://www.forbes.com/councils/forbesbusinesscouncil/2025/05/07/the-future-of-mental-health-how-ai-vr-and-ar-are-changing-treatment/. Accessed 10 June 2025

Unity (2025) Available from https://unity.com/ Accessed 10 June

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.