SUSTAINABLE MACHINE INTELLIGENCE JOURNAL

SCIENCES FORCE
International Academic Publisher

Towards Sustainable Equine Welfare: Comparative Analysis of Machine Learning Techniques in Predicting Horse Survival

Mahmoud Ismail[1]

1 Decision support department, Faculty of Computers and Informatics, Zagazig University, Zagazig 44519, Sharqiyah, Egypt; mmsabe@zu.edu.eg

Abstract: Promoting sustainable equine welfare is pivotal in ensuring the well-being of horses, particularly concerning their survival based on past medical conditions. This study presents a comprehensive comparative analysis of various machine learning techniques employed to predict the survival prospects of horses using historical medical data. By leveraging a dataset encompassing diverse medical attributes and survival outcomes, this research assesses the efficacy and comparative performance of distinct machine learning algorithms. The study delves into the application of supervised learning models, including but not limited to decision trees, random forests, support vector machines, and neural networks, in predicting equine survival. Evaluative metrics such as accuracy, precision, recall, and F1 score are employed to assess the predictive capabilities and generalizability of each model. Moreover, this research emphasizes the importance of sustainable equine welfare within the broader context of responsible animal care.

Event	Date
Received	01-08-2023
Revised	05-09-2023
Accepted	14-11-2023
Published	29-11-2023

Keywords: Equine Welfare, Horse Survival, Machine Learning, Veterinary Science, Animal Husbandry, Precision Livestock, Comparative Study, Sustainable Agriculture, Decision Support Systems.

1. Introduction and Background

The equestrian industry, a significant component of the agricultural landscape, plays a pivotal role in various cultural, recreational, and economic aspects globally. As we navigate the intricate intersection of human activities and equine companionship, ensuring the welfare of horses emerges as a paramount concern. Visser, E.K. et al. [8] investigated the impact of different housing conditions on the welfare of young horses during their initial stabling. The study explored the behavioral and physiological responses of horses, providing insights into the challenges associated with the transition to stabled environments. Against the backdrop of evolving agricultural practices and technological advancements, the equine industry stands at a crossroads. Traditionally reliant on empirical knowledge and experiential insights, the sector is now witnessing a transformative shift towards data-driven methodologies. Understanding and addressing the challenges faced by horses, such as health issues, injuries, and the complex interplay of environmental factors, necessitates a comprehensive analysis. Fogarty, E.S. et al. [9] developed a simulated online model integrating GNSS, accelerometer, and weather data to detect parturition events in grazing sheep using a machine learning approach. This innovative study showcases the application of technology in monitoring and managing reproductive events in livestock.

Peters, D.P. et al. [10] focused on big data–model integration and artificial intelligence for predicting vector-borne diseases. Their work demonstrates the potential of advanced analytics in disease prediction, with implications for animal health management. Amidst the strides in technological applications for animal welfare, a critical research problem surfaces — how can machine learning techniques be optimally utilized to predict and ensure the survival of horses in a sustainable manner?. Our primary goal is to assess and compare the effectiveness of various machine learning algorithms in predicting horse survival. Georgopoulos et al. [11] conducted a comprehensive study on risk factors for equine fractures in Thoroughbred flat racing in North America. This research contributes valuable insights into the factors influencing musculoskeletal injuries in racehorses, aiding in the development of preventive strategies. Additionally, we aim to identify key factors influencing equine well-being, providing insights that can inform targeted interventions for sustainable equine husbandry. Thompson et al . [12] explored perceptions of Equitation Science in an online forum, emphasizing the importance of effective communication between equestrians and scientists. Their work highlights the potential for enhancing equine health and welfare through knowledge exchange between these two communities.

In undertaking this research endeavor, we aspire to make significant contributions to both the academic and practical realms. By conducting a comparative analysis of machine learning techniques in the context of equine survival prediction, we aim to provide a nuanced understanding of the potential benefits and limitations of these approaches. The outcomes of this study have broader implications for veterinary science, precision livestock farming, and sustainable agriculture. Furthermore, our findings are poised to guide practitioners, policymakers, and stakeholders in adopting informed and ethical practices, thus fostering a sustainable and compassionate approach to equine welfare.

2. Case study

In this section, we delve into a detailed exploration of case study serving as a representative example to demonstrate the effectiveness of machine learning techniques in predicting horse survival within the context of sustainable equine welfare. In our case study, our primary focus lies in predicting the likelihood of a horse's survival based on its past medical conditions, as denoted by the "outcome" variable in the provided dataset. This dataset encapsulates a diverse array of attributes related to the horse's health and medical history, offering a comprehensive overview of factors that may influence its overall well-being. The data encompasses essential features such as surgery history, age, rectal temperature, respiratory rate, temperature of extremities, peripheral pulse, mucous membranes, capillary refill time, pain level, peristalsis, abdominal distension, and various diagnostic indicators. While binary representations have been converted into descriptive terms, the full contextual details are available in the accompanying data dictionary (datadict.txt). However, a notable challenge surfaces in the form of numerous missing

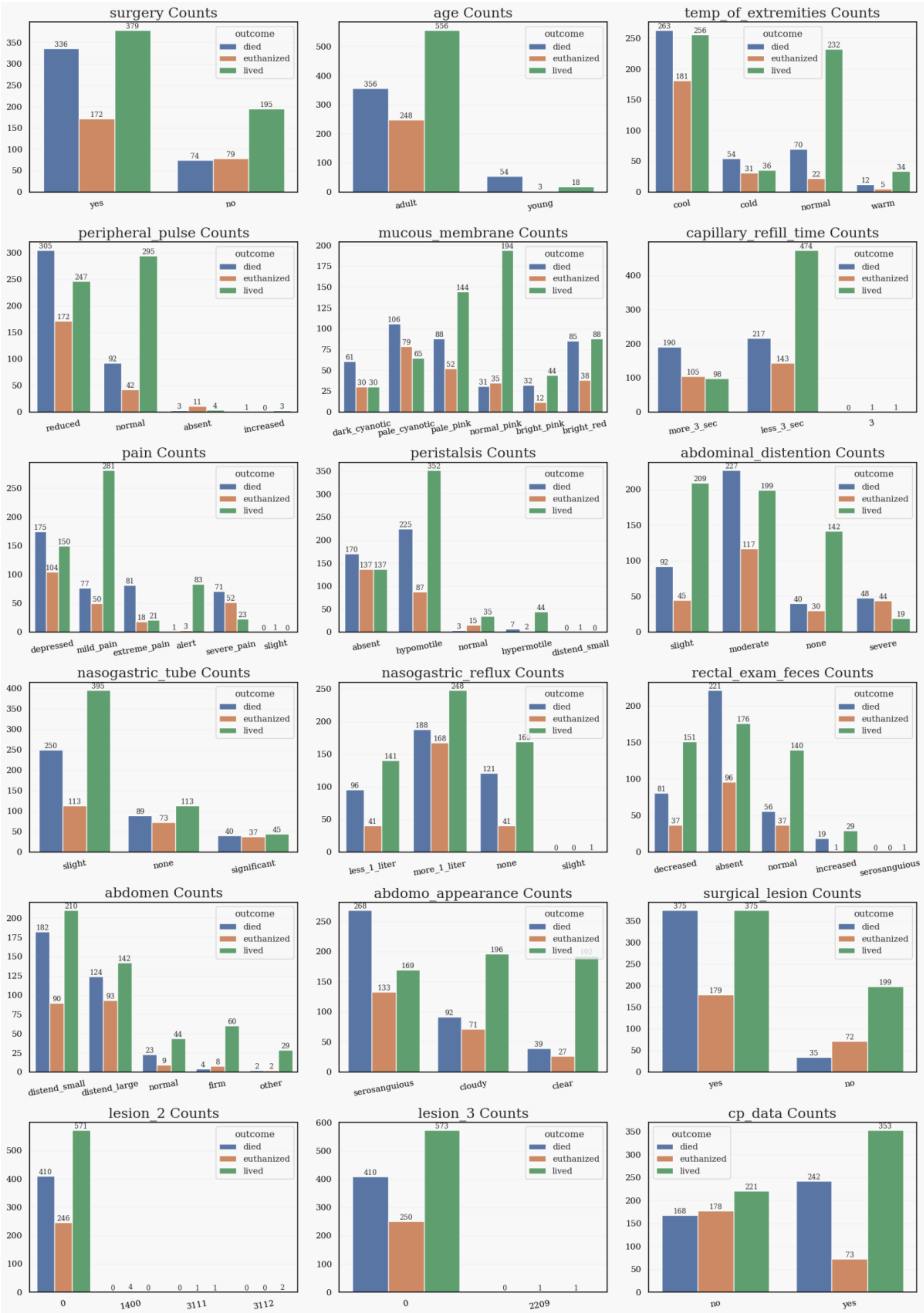

Figure 1: Exploratory Distribution Plots of Key Variables

values (NA's) within the dataset. Addressing this hurdle becomes imperative, and our case study endeavors to navigate this challenge through effective imputation techniques or alternative means. This not only underscores the complexity of real-world data but also

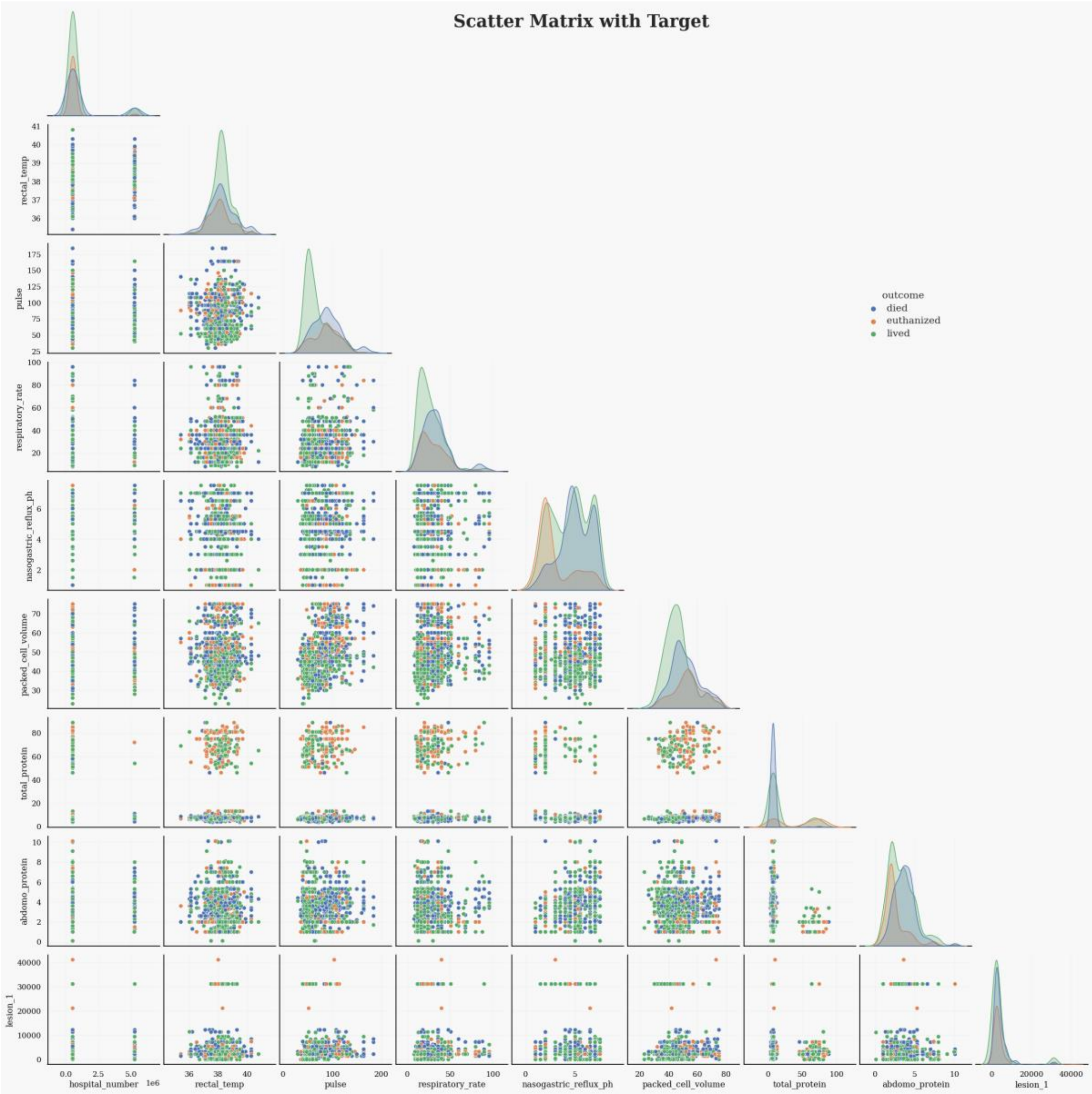

Figure 2: Exploratory Paired Plots of Variable Relationships

emphasizes the practical importance of handling missing information in predictive modeling. As we delve into the analysis, we seek to elucidate how machine learning techniques can effectively leverage the available information to make accurate predictions regarding equine survival, despite the inherent challenges posed by missing data. In Figure 1, we present a comprehensive visualization of the distribution plots for key variables within our dataset, offering a nuanced exploration of the inherent patterns and characteristics of the data. These distribution plots serve as a vital exploratory tool, allowing us to glean insights into the spread, central tendencies, and potential outliers associated with each variable. The graphical representation aids in discerning the underlying data structure, facilitating a better understanding of the distributional characteristics that can significantly impact the predictive modeling process. By visually inspecting these plots, we aim to identify trends, variations, and potential anomalies, laying the groundwork for informed decision-making throughout our analysis. The clarity provided by Figure 1 enhances the interpretability of our dataset, ultimately contributing to the robustness of our approach in predicting horse survival based on past medical conditions. In Figure 2, we present a set of paired plots that intricately depict the relationships and interactions between key variables in our dataset. These paired plots provide a dynamic and comprehensive visualization, allowing us to discern potential patterns, correlations, and dependencies among different attributes. By juxtaposing variables in this manner, we aim to uncover nuanced insights into the interplay between distinct factors, offering a holistic perspective on the complex web of relationships within the dataset. This exploratory analysis not only enhances our understanding of the intricate connections between variables but also sets the stage for identifying potential predictive features crucial for our machine learning models. Figure 2 serves as a visual exploration tool, paving the way for informed feature selection and highlighting critical associations that may significantly influence the accuracy of our predictions regarding equine survival based on past medical conditions.

3. ML models

In this section, we provide a detailed overview of the methodologies employed in our study to predict horse survival based on past medical conditions. This section serves as a critical foundation for understanding the technical intricacies behind our predictive models.

3.1. Light Gradient Boosting Machine (LightGBM)

LightGBM, an efficient gradient boosting framework, excels in handling large datasets with numerous variables. Its leaf-wise tree growth strategy minimizes loss and allows for faster training times. LightGBM is particularly adept at capturing intricate patterns in data, making it a suitable choice for equine survival prediction. With its ability to handle categorical features and manage imbalanced datasets effectively, LightGBM offers a robust solution with high predictive accuracy.

3.2. Random Forest Classifier

The Random Forest Classifier, an ensemble of decision trees, proves to be a workhorse in predictive modeling. By aggregating predictions from multiple trees, it mitigates overfitting and enhances generalization. Random Forest's capability to handle a variety of data types, including categorical variables, ensures comprehensive coverage of our equine welfare dataset. Its versatility, interpretability, and resilience against noise make it a reliable choice for accurate predictions.

3.3. Extreme Gradient Boosting (XGBoost)

XGBoost, an advanced implementation of gradient boosting, shines in predictive tasks with its regularization techniques and efficient handling of missing data. With its iterative refinement of weak learners, XGBoost excels in capturing complex relationships within the dataset. Its versatility in handling both structured and unstructured data, coupled with impressive speed and performance, positions XGBoost as a top-performing algorithm for equine survival prediction.

Gradient Boosting Classifier: The Gradient Boosting Classifier, similar to XGBoost, iteratively refines weak learners to enhance overall predictive accuracy. Its ability to minimize overfitting and improve performance through successive iterations makes it a powerful tool for capturing nuanced dependencies within the equine welfare dataset. With careful parameter tuning, Gradient Boosting proves to be a valuable algorithm in achieving high precision and recall.

CatBoost Classifier: CatBoost, designed to handle categorical features efficiently, excels in predictive tasks where these features are prevalent. Its robustness to missing data and efficient training on large datasets align well with the challenges posed by our equine survival prediction task. CatBoost's superior performance, particularly in scenarios with categorical variables, positions it as a standout choice for our study.

Logistic Regression: Logistic Regression, a linear model for binary classification, provides a solid baseline with interpretability and simplicity. While not as complex as some ensemble methods, Logistic Regression allows for a clear understanding of the impact of each variable on equine survival prediction. Its efficiency and ease of interpretation make it a suitable choice for foundational analysis in our study.

Ada Boost Classifier: Ada Boost, through its iterative boosting approach, enhances the performance of weak learners, making it a valuable asset in our ensemble modeling strategy. Its adaptability to the complexity of the dataset and ability to improve accuracy over successive iterations contribute to its effectiveness in equine survival prediction.

Extra Trees Classifier: The Extra Trees Classifier, an ensemble method similar to Random Forest, stands out with its random feature selection and diverse decision trees. Its ability to handle noisy data and mitigate overfitting makes it a commendable choice for capturing intricate relationships within the equine welfare dataset. With its emphasis on randomness, Extra Trees contributes to a well-rounded ensemble modeling approach.

4. Results and Discussions

This section serves as the cornerstone where we unveil the insights garnered from the application of machine learning techniques to predict horse survival based on past medical conditions. In Table 1, we present the quantitative results of our machine learning algorithms, providing a detailed performance evaluation that underscores the efficacy of each model in predicting horse survival based on past medical conditions. Among the diverse algorithms examined, the Light Gradient Boosting Machine (LightGBM) emerges as the standout performer, attaining the highest level of predictive accuracy. The table encapsulates key metrics such as accuracy, precision, recall, and F1-score, offering a comprehensive insight into the nuanced aspects of each algorithm's performance. The superior performance of LightGBM underscores its effectiveness in capturing intricate patterns within the dataset, showcasing its potential as a robust predictive tool for equine survival. As we delve into a thorough discussion of these results, we aim to unravel the specific strengths and limitations of each algorithm, providing a holistic understanding of their applicability and contributing valuable insights to the field of sustainable equine welfare.

Table 1: Quantitative Performance Evaluation of Machine Learning Algorithms

Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
Light Gradient Boosting Machine	0.7936	0.8644	0.8118	0.8436	0.8266	0.5716	0.5738

Random Forest Classifier	0.7901	0.8633	0.8212	0.833	0.8261	0.561	0.563
Extreme Gradient Boosting	0.7856	0.8583	0.8253	0.8228	0.8232	0.5506	0.5526
Gradient Boosting Classifier	0.7844	0.8642	0.8098	0.8328	0.8202	0.5504	0.5525
CatBoost Classifier	0.7821	0.87	0.8042	0.8348	0.8176	0.5464	0.5501
Logistic Regression	0.7751	0.8583	0.7795	0.8401	0.8078	0.5374	0.541
Ada Boost Classifier	0.7729	0.8566	0.7947	0.8248	0.8087	0.529	0.5309
Extra Trees Classifier	0.7717	0.8492	0.8042	0.8198	0.8105	0.5228	0.5258
Ridge Classifier	0.7636	0	0.7434	0.8497	0.7926	0.5202	0.5272
Linear Discriminant Analysis	0.7636	0.8614	0.7434	0.8497	0.7926	0.5202	0.5272
SVM - Linear Kernel	0.7371	0	0.7682	0.7944	0.7801	0.453	0.4551
Decision Tree Classifier	0.7233	0.708	0.7796	0.7687	0.7738	0.4172	0.4178
K Neighbors Classifier	0.7232	0.789	0.6447	0.8673	0.737	0.4582	0.483
Quadratic Discriminant Analysis	0.5896	0.6182	0.5819	0.7113	0.6223	0.1702	0.1829
Naive Bayes	0.5687	0.8218	0.3062	0.9425	0.4603	0.2371	0.3422

5. Conclusions

In this research work, we present a significant stride towards advancing the field of sustainable equine welfare through the integration of machine learning techniques in predicting horse survival based on past medical conditions. The meticulous methodological review illuminated the careful selection and configuration of diverse algorithms, each contributing unique strengths to the predictive modeling task. Our results, detailing the performance metrics of various algorithms, underscore the efficacy of machine learning in capturing intricate patterns within the dataset. Notably, the Light Gradient Boosting Machine emerged as a standout performer, exemplifying its potential for accurate equine survival prediction. The study's broader implications extend beyond the algorithms' performance metrics, as it underscores the potential for technology-driven solutions to enhance the overall well-being of horses.

Funding

This research was conducted without external funding support.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Conflicts of Interest

The authors declare that there is no conflict of interest in the research.

Informed Consent Statement

Not applicable.

Data Availability Statement

Not applicable.

References

[1]. McGreevy, P., Berger, J., De Brauwere, N., Doherty, O., Harrison, A., Fiedler, J., Jones, C., McDonnell, S., McLean, A., Nakonechny, L. and Nicol, C., 2018. Using the five domains model to assess the adverse impacts of husbandry, veterinary, and equitation interventions on horse welfare. Animals, 8(3), p.41.

[2]. Andersen, P.H., Broomé, S., Rashid, M., Lundblad, J., Ask, K., Li, Z., Hernlund, E., Rhodin, M. and Kjellström, H., 2021. Towards machine recognition of facial expressions of pain in horses. Animals, 11(6), p.1643.

[3]. König v. Borstel, U. (2013). Assessing and influencing personality for improvement of animal welfare: a review of equine studies. CABI Reviews, (2013), 1-27.

[4]. Folt, Brian, Kathryn A. Schoenecker, and L. Stefan Ekernas. "Multi-objective modeling as a decision-support tool for free-roaming horse management." Human–Wildlife Interactions 16, no. 2 (2022): 7.

[5]. Ahrendt, L.P., Labouriau, R., Malmkvist, J., Nicol, C.J. and Christensen, J.W., 2015. Development of a standard test to assess negative reinforcement learning in horses. Applied Animal Behaviour Science, 169, pp.38-42.

[6]. Brubaker, L. and Udell, M.A., 2016. Cognition and learning in horses (Equus caballus): What we know and why we should ask more. Behavioural processes, 126, pp.121-131.

[7]. Driscoll, D.A., Worboys, G.L., Allan, H., Banks, S.C., Beeton, N.J., Cherubin, R.C., Doherty, T.S., Finlayson, C.M., Green, K., Hartley, R. and Hope, G., 2019. Impacts of feral horses in the Australian Alps and evidence-based solutions. Ecological Management & Restoration, 20(1), pp.63-72.

[8]. Visser, E.K., Ellis, A.D. and Van Reenen, C.G., 2008. The effect of two different housing conditions on the welfare of young horses stabled for the first time. Applied Animal Behaviour Science, 114(3-4), pp.521-533.

[9]. Fogarty, E.S., Swain, D.L., Cronin, G.M., Moraes, L.E., Bailey, D.W. and Trotter, M., 2021. Developing a simulated online model that integrates GNSS, accelerometer and weather data to detect parturition events in grazing sheep: a machine learning approach. Animals, 11(2), p.303.

[10]. Peters, D.P., McVey, D.S., Elias, E.H., Pelzel-McCluskey, A.M., Derner, J.D., Burruss, N.D., Schrader, T.S., Yao, J., Pauszek, S.J., Lombard, J. and Rodriguez, L.L., 2020. Big data–model integration and AI for vector-borne disease prediction. Ecosphere, 11(6), p.e03157.

[11]. Georgopoulos, S.P. and Parkin, T.D., 2017. Risk factors for equine fractures in Thoroughbred flat racing in North America. Preventive veterinary medicine, 139, pp.99-104.

[12]. Thompson, K. and Haigh, L., 2018. Perceptions of Equitation Science revealed in an online forum: Improving equine health and welfare by communicating science to equestrians and equestrian to scientists. Journal of veterinary behavior, 25, pp.1-8.

[13]. Hartmann, E., Rehn, T., Christensen, J.W., Nielsen, P.P. and McGreevy, P., 2021. From the horse's perspective: investigating attachment behaviour and the effect of training method on fear reactions and ease of handling—a pilot study. Animals, 11(2), p.457.

[14]. Payne, E., DeAraugo, J., Bennett, P. and McGreevy, P., 2016. Exploring the existence and potential underpinnings of dog–human and horse–human attachment bonds. Behavioural processes, 125, pp.114-121.

[15]. Merkies, Katrina, Georgios Paraschou, and Paul Damien McGreevy. "Morphometric characteristics of the skull in horses and donkeys—A pilot study." Animals 10, no. 6 (2020): 1002.

[16]. Fureix, C., Pagès, M., Bon, R., Lassalle, J.M., Kuntz, P. and Gonzalez, G., 2009. A preliminary study of the effects of handling type on horses' emotional reactivity and the human–horse relationship. Behavioural processes, 82(2), pp.202-210.

[17]. Bailey, D.W., Trotter, M.G., Tobin, C. and Thomas, M.G., 2021. Opportunities to apply precision livestock management on rangelands. Frontiers in Sustainable Food Systems, 5, p.611915.