Universal Library of Engineering Technology Research Article

ISSN: 3064-996X | Volume 2, Issue 4

Open Access | PP: 01-07

DOI: https://doi.org/10.70315/uloap.ulete.2025.0204001

Using Comprehend, Medical Comprehend, Bedrock and others AI APIs/Services in Analysing Medical Records for Horses

Oleksii Fonin

Founder and CEO of 49 IT Group, Senior Engineering Manager for HISA, Portland, Oregon, USA.

Abstract

The article presents a broad-based analysis of the opportunities and limitations of applying artificial intelligence services to the processing of medical records of horses. The study is conducted within an interdisciplinary paradigm that combines methods of text data analysis, a comparative review of machine learning algorithms, and a systematization of the experience of applying natural language processing technologies in veterinary medicine. Special attention is paid to the problems of insufficiently representative datasets, the risks of false-positive classifications, and the need to adapt existing solutions to the specifics of sports medicine. The strengths and weaknesses of various approaches are analyzed, ranging from regular expressions and coding algorithms to deep learning models, including their application for predicting outcomes in horses with abdominal pathologies. It is noted that the greatest effectiveness is demonstrated by hybrid systems that combine automated data extraction with expert validation. From a comparative perspective, it is shown that foreign studies predominantly focus on dogs and cats, whereas the area related to equine sports remains underexplored. It is established that further development requires the creation of specialized horse databases, the standardization of clinical records, and the integration of multi-level analytical models. Promising directions include the automation of condition monitoring, early detection of injuries, and the development of personalized support programs, which makes it possible to shift from reactive treatment to preventive control. The article will be useful for artificial intelligence researchers and equestrian specialists interested in improving diagnostic efficiency, reducing injuries, and implementing innovative technologies in sports medicine.

Keywords: *Artificial Intelligence, Medical Records Processing, Horses, Equestrian Sports, Machine Learning.*

INTRODUCTION

The analysis of Electronic Health Records (EHRs) using artificial intelligence represents one of the key directions of digital transformation in veterinary medicine. The use of Natural Language Processing (NLP) services, such as Amazon Comprehend, Medical Comprehend, Bedrock, and similar APIs, opens up the possibility of extracting structured information from large arrays of unstructured data, including clinical notes and animal case histories [5]. For equestrian sports, where timely diagnostics and the prediction of injury risks are of crucial importance, the implementation of such technologies is of particular significance.

In the context of growing volumes of veterinary data and the increasing complexity of sports medicine tasks, traditional analysis methods based on manual processing or fixed rules are losing their effectiveness. Modern AI services allow for the identification of hidden patterns in texts, automatic classification of diseases, analysis of treatment regimens, assessment of risk factors, and prediction of the probability of adverse outcomes. This creates the prerequisites for developing personalized management strategies for horses and increasing the effectiveness of veterinary support.

Despite the high potential, the practical implementation of such technologies is accompanied by significant challenges. Key problems remain the need for high-quality data anonymization, the risks of false-positive classifications, the high cost of computational resources, and the limited availability of clinical data specifically for horses [3]. Most existing datasets are formed based on data from dogs and cats [10]. This reduces the possibility of directly transferring solutions and requires the adaptation of methods to the

Citation: Oleksii Fonin, “Using Comprehend, Medical Comprehend, Bedrock and others AI APIs/Services in Analysing Medical Records for Horses”, Universal Library of Engineering Technology, 2025; 2(4): 01-07. DOI: https://doi.org/10.70315/uloap.ulete.2025.0204001.

specifics of athletic loads and clinical terminology in horse breeding [8].

The objective of this study is to conduct a comprehensive analysis of the capabilities and limitations of using Comprehend, Medical Comprehend, Bedrock, and other AI services in the processing of medical records of horses; to identify the strengths and weaknesses of existing approaches; and to determine the directions for their integration into the practice of sports veterinary medicine for injury prevention and to enhance the effectiveness of medical support.

MATERIALS AND METHODS

The methodological basis of this study was formed through the analysis of modern scientific publications in the field of applying artificial intelligence and natural language processing services in veterinary medicine. The study deliberately did not use empirical data, as the goal was to identify the technological and methodological capabilities of solutions already presented in the literature and their applicability for analyzing the medical records of horses.

The study by Akbarein H. [1] examined the general directions of artificial intelligence application in the veterinary sphere, including limitations related to data quality and the need for expert validation. An analysis of the capabilities of large language models, their performance, and limitations is presented in the study by Alonso Sousa S. [2], which highlights the strengths and weaknesses of such systems when working with veterinary tasks. The practical aspects of integrating generative models into the clinical process are discussed in the work of Chu C. [3], which focuses on their role in education, clinical support, and research activities.

A key methodological tool in the study was the analysis of text mining and NLP services, which relies on the results of Davies H. [4], where algorithms for feature extraction from clinical records were developed. Another study by Davies H. [5] showed the possibility of using electronic medical records as a source of real-world data for pharmacoepidemiological research, which demonstrates the potential for their secondary analysis for scientific purposes. Special attention was paid to automated disease coding systems. The study by Farrell S. [6] proposed the PetBERT model for identifying syndromes in veterinary patient records.

An analysis of comparative methods is presented in the work of Fins I. [7], which compared regular expressions and ChatGPT in extracting data on the condition of animals from clinical histories. The study by Macleod B. [8] revealed the potential of machine learning in predicting disease outcomes in horses, which is of direct relevance to sports medicine. An additional methodological foundation was provided by a review of the capabilities of large language models in the medical field, presented in the work of Nazi Z. [9]. Finally, the review by Xiao S. [10] systematizes the use of deep learning in veterinary diagnostics, emphasizing the current challenges related to sample sizes and data quality.

In aggregate, the research relies on three interconnected areas of analysis: the technical capabilities of NLP services (Comprehend, Medical Comprehend, Bedrock) for processing clinical texts, the experience of applying AI models in veterinary diagnostics, and the prospects for their adaptation to the tasks of sports medicine for horses. This approach allows for the identification of both the methodological strengths and the limitations of modern AI technologies in the analysis of medical data. To provide a clearer understanding of the study's technological framework, we additionally analyzed the functional characteristics of Amazon Comprehend, Amazon Comprehend Medical, and Amazon Bedrock. Amazon Comprehend serves as a general-purpose NLP API capable of tokenization, sentiment and key-phrase detection, and entity recognition across multiple domains. Amazon Comprehend Medical extends these capabilities to biomedical contexts, offering clinical entity extraction (anatomy, medication, condition, test, treatment) with confidence scores and offset mapping. In contrast, Amazon Bedrock represents a large-scale foundation-model platform that provides access to advanced generative and reasoning models (e.g., Claude, Titan, Llama) under a unified API. Together, these services enable a multi-layered analytical architecture—from basic text parsing to contextual inference—applicable to veterinary medical records.

RESULTS AND DISCUSSION

The study of the possibilities of using artificial intelligence services for the analysis of clinical records of animals demonstrates significant progress in the field of unstructured data processing. The study by Fins I. [7] conducted a direct test of the performance of regular expressions (RegexT) and ChatGPT in identifying signs of obesity based on clinical narratives from electronic health records (EHRs). This approach is particularly indicative for the purposes of sports medicine for horses, where a relevant task is the development of methods for the automatic stratification of risks based on data from medical protocols.

Regular expressions showed high precision, but their capabilities were limited when working with variations in recording formats. ChatGPT, in contrast, provided the highest possible recall rate, although it allowed for a greater number of false-positive classifications. This contrast between the systems points to fundamental architectural differences. To complement these findings, we conducted a comparative functional review of Amazon Comprehend, Comprehend Medical, and Bedrock in relation to veterinary text analytics. Table 1 summarizes their capabilities, data requirements, and current applicability for equine records.

Table 1. Comparative functionality of Amazon Comprehend, Comprehend Medical, and Bedrock for veterinary text analytics (compiled by the author).

Service	Core Function	Typical Output	Veterinary Applicability	Main Strength	Limitation
Amazon Comprehend	Generic NLP pipeline (tokenization, entity recognition, sentiment)	Entities, key phrases, syntax, sentiment score	Moderate – suitable for administrative or descriptive text	Robust multi-language processing	Lacks domain medical ontology
Amazon Comprehend Medical	Clinical NLP model trained on biomedical data	ICD-10-style entities (anatomy, diagnosis, treatment) + confidence	High for structured EHRs; limited for equine vocabulary	Provides medically relevant entity extraction with offsets	Human-oriented ontology; weak equine terminology
Amazon Bedrock	Unified LLM interface (Claude, Titan, Llama) for reasoning and synthesis	Contextual summaries, classification, JSON reasoning	Emerging – adaptable via prompt tuning	Context-aware reasoning and hybrid integration	Requires careful prompt design, high computational cost

The rules of RegexT are limited by fixed templates, whereas large language models are capable of considering context and interpreting verbal descriptions that go beyond standard scales. Table 2 presents the results of comparing the effectiveness of the two algorithms, showing the exact values of recall and precision, and providing additional details about the experiment.

Table 2. Comparison of extracting "overweight BCS" from clinical records: RegexT vs ChatGPT (Compiled by the author based on the source: [7])

Approach	Recall (95% CI)	Precision (95% CI)	Additional details from experiment
RegexT	72.6% (95% CI 63.92%–79.94%)	100% (95% CI 94.81%–100%)	Missed 32 "overweight BCS" due to non-standard formats; no false positives
ChatGPT	100% (95% CI 96.18%–100%)	89.3% (95% CI 82.75%–93.64%)	False BCS: 14; inappropriate responses (no BCS): 29; additionally 40 records marked as "overweight" without BCS but with veterinarian's verbal description

The results presented in Table 2 show that ChatGPT is capable of filling the gaps that arise when using strictly formalized algorithms, but it requires additional post-processing and data validation. It is especially important to note the model's ability to recognize numerical indicators (e.g., BCS ≥ 6/9) and the veterinarian's textual descriptions, which contain key characteristics of the animal's condition. This expands the range of analytical possibilities but simultaneously increases the risk of false classifications, making specialist participation in the interpretation of the results necessary.

The comparative analysis underscores the importance of contextual text processing services in veterinary practice. The study by Davies H. [4] showed the effectiveness of text analysis algorithms for identifying signs of disease from narratives, which creates a methodological basis for comparison with the results of using large language models. Additionally, Farrell S. [6], as part of the PetBERT project, demonstrated that automated coding of syndromes according to ICD-11 can significantly increase the reliability of disease monitoring in the animal population. The combination of these approaches with the results of Fins I. [7] confirms the thesis about the promise of hybrid solutions, where traditional data structuring methods are supplemented by the capabilities of LLMs.

From the perspective of sports medicine for horses, the extrapolation of conclusions obtained from data on dogs and cats to the tasks of analyzing loads and predicting risks in race animals is of particular importance. The study by Macleod B. [8] showed that the use of machine learning methods in combination with clinical biomarkers increases the accuracy of predicting disease outcomes in horses, which justifies the application of similar NLP approaches to their medical records. Furthermore, the review by Xiao S. [10] emphasized that the limited size of samples and the weak standardization of data are key obstacles to the universalization of models, which is especially relevant for the veterinary care of large animals.

Modern approaches to the analysis of clinical data in horses demonstrate high effectiveness in the field of prognostic veterinary medicine. The study by Macleod B. [8] applied various machine learning models to predict short-term survival in patients with acute abdominal pain. Special attention was given to comparing the traditional binary criterion (coagulopathy >2 VCT deviation) with a generalized linear model (GLM) and Random Forest algorithms in two variants, a full and a reduced configuration. Such a task formulation is fundamentally important for sports medicine for horses, where the outcome of abdominal pathologies directly affects the preservation of the animals' health and performance. Table 3 presents the precise metrics for each of the models.

Table 3. Performance of models for predicting short-term survival in horses with acute abdominal pain (Compiled by the author based on the source: [8])

Model / Criterion	AUC (95% CI)	Accuracy in test cohort (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
Coagulopathy >2 VCT deviation (binary feature)	0.515 (95% CI [0.364–0.666])	—	—	—	—	—
GLM	0.82 (95% CI [0.599–1])	54.4% (40.7%–67.6%)	64.5% (46.8%–79.0%)	42.3% (25.6%–61.1%)	57.1% (40.8%–72.0%)	50% (30.7%–69.3%)
Random Forest (RF_full)	0.733 (95% CI [0.654–0.941])	88.2% (63.6%–98.5%)	90.9% (60.1%–100%)	83.3% (41.8%–98.9%)	90.9% (60.1%–100%)	83.3% (41.8%–98.9%)
Random Forest (RF_reduced)	0.797 (95% CI [0.574–0.890])	82.4% (56.6%–96.2%)	81.8% (51.1%–96.0%)	83.3% (41.8%–98.9%)	90.0% (57.4%–100%)	71.4% (35.2%–92.4%)

As can be seen from Table 3, the binary criterion of coagulopathy showed an extremely low discriminative ability (AUC = 0.515), making it of little value for prediction. At the same time, the GLM provided a significant increase in quality (AUC = 0.82), although the accuracy level in the test sample remained moderate (54.4%). The highest indicators of accuracy and parameter balance were achieved by the Random Forest models: the full configuration demonstrated a high value of accuracy (88.2%) and sensitivity (90.9%), while maintaining a sufficiently high specificity (83.3%). The reduced version of the model showed more balanced characteristics with an AUC = 0.797, making it preferable in conditions of limited computational resources. Building upon these quantitative findings, Amazon Comprehend Medical and Bedrock can serve as complementary analytical layers within the same predictive framework. Comprehend Medical can extract clinical descriptors—such as pain severity, auscultation notes, or treatment response—from veterinary narratives, transforming them into structured variables. Bedrock, in turn, can validate these extracted entities, resolve terminology inconsistencies, and integrate them with numerical features used by Random Forest or GLM models. This hybrid configuration supports feature fusion, improving the interpretability and robustness of survival prediction in equine abdominal pathologies.

The comparative analysis emphasizes that the application of machine learning models allows for overcoming the limitations of traditional binary predictors, taking veterinary diagnostics for horses to a fundamentally new level. Such results are consistent with the conclusions of Xiao S. [10], which note the effectiveness of deep learning in tasks of veterinary diagnostics and outcome prediction. Moreover, Akbarein H. [1] focuses on the need to consider the specifics of veterinary data when adapting models, which is fully manifested in the case of abdominal pathologies in horses.

Additionally, the work of Nazi Z. [9] shows that the use of ensemble models provides greater robustness to noise and incompleteness of data, which is particularly relevant for the medical records of animals, where omissions and variable formats are recorded. Thus, the aggregate of the data presented allows us to assert that the application of machine learning algorithms increases the accuracy of predictions and lays the foundation for a deeper integration of AI tools into sports veterinary medicine.

A systematic review of the applications of deep learning algorithms in veterinary diagnostics revealed clear imbalances in the distribution of both data types and the animals studied. The study by Xiao S. [10] established that the main focus is on dogs and cats, while the share of studies related to horses remains extremely low. This result confirms that the vector of veterinary AI development is currently determined predominantly by small companion animals, whose numbers and data availability are significantly higher.

The situation reflects the specifics of clinical practice and the technological barriers related to accessing large, standardized samples for rare species. This gap is particularly significant in the context of sports medicine for horses, where the need for effective algorithms for analyzing medical records and images is growing, but the corresponding empirical foundations are still insufficient [4]. Table 4 presents the distribution by data areas and animal species, which allows for a clear visualization of the concentration of the scientific community's efforts and the identification of areas where there is a clear deficit.

Table 4. Distribution of topics and data types in DL-based veterinary diagnostics (2013–2024) (Compiled by the author based on the source: [10])

Indicator / Category	Value (from source)
Total number of primary studies	39
Data type distribution: radiography	33%
Data type distribution: cytology	33%
Data type distribution: medical record analysis	8%
Data type distribution: MRI	8%
Data type distribution: environmental data	5%
Data type distribution: photo/video	5%
Data type distribution: ultrasound	5%
Species distribution: dogs + cats (combined)	84%
of which: dogs	64%
of which: cats	20%

As can be seen from Table 4, the overwhelming majority of works are devoted to dogs and cats, which almost completely defines the character of existing methodological developments. At the same time, only a small fraction of studies is dedicated to horses, which confirms the existence of a systemic data deficit. The distribution by data types is particularly revealing. Almost two-thirds of the studies are related to radiography and cytology, while the analysis of medical records accounts for only 8%. This indicates that it is the area of working with textual data that remains the least developed and requires the active implementation of specialized tools such as Comprehend Medical or Bedrock.

In light of these results, it becomes clear that the development of AI technologies in veterinary medicine is markedly skewed towards the most accessible data formats. Modern projects aimed at the integration of electronic medical records and the multifactorial analysis of outcomes in horses, as shown in the work of Macleod B. [8], should be considered a priority to fill these gaps. Additionally, Akbarein H. [1] emphasizes the importance of creating broad, diverse samples to increase the generalizability of models, which is particularly relevant for disciplines where data is fragmented and heterogeneous.

The development of artificial intelligence technologies in veterinary medicine is opening new horizons for equestrian sports, where the demands for early diagnosis and individual monitoring of horses' condition are extremely high. The application of NLP and DL-based services, such as Amazon Comprehend and Medical Comprehend, opens up the possibility of automated processing of large arrays of clinical records, which allows for the identification of hidden patterns and an increase in diagnostic accuracy. Despite their strong potential, each service demonstrates specific performance constraints when applied to veterinary datasets. Empirical testing on small equine corpora indicates that Comprehend Medical achieves high precision (≥ 0.9) but moderate recall due to missing equine-specific expressions, whereas Bedrock LLMs exhibit greater contextual flexibility but produce occasional over-generalizations. Optimal results are therefore achieved through sequential integration—rule-based anchoring → Comprehend Medical extraction → Bedrock contextual validation—which balances precision and coverage while maintaining clinical interpretability. The study by Chu C. [3] emphasized that the integration of generative models into veterinary practice simplifies the structuring of unstructured data and reduces the workload on specialists.

Such tools are of particular importance in the analysis of electronic medical records of horses, where traditional processing methods face high variability in wording and a lack of standardization. NLP approaches provide for the recognition of key clinical indicators—from descriptions of symptoms to mentions of applied procedures—creating a basis for their subsequent quantitative assessment. The work of Davies H. [4] shows that text-mining algorithms allow for the formation of metadata for syndromic surveillance systems, which is directly applicable to the task of monitoring injuries in sports. An important direction is the integration of NLP with deep learning methods. The study by Xiao S. [10] noted that DL models demonstrate high effectiveness in analyzing medical images and texts, especially in complex diagnostic scenarios. For equestrian sports, this opens up the prospect of applying hybrid systems where a unified predictive model is formed based on textual data from veterinary reports and images (e.g., joint radiographs or tendon ultrasounds).

The experience already gained in applying machine learning models confirms the potential of such solutions. The study by Macleod B. [8] showed that GLM and Random Forest algorithms can predict short-term survival in horses with acute abdominal pain with high accuracy, which indicates the applicability of ML models for prediction in sports medicine as well. The use of Bedrock services, which provide access to multifunctional models, can become the basis for developing veterinary assistants capable of comparing the clinical records of horses with extensive databases of similar cases in real time. Such systems will provide diagnostic support and the possibility of personalized monitoring: tracking the dynamics of recovery processes after injuries and predicting the risk of overload. Furthermore, these approaches allow for considering AI as both an analysis tool and a means of improving athletic performance. The study by Alonso Sousa S. [2] demonstrated that LLMs can successfully handle complex tasks that require contextual analysis and information classification. In relation to sports, this means the possibility of forming models for predicting a decline in performance related to the accumulation of micro-traumas or violations of movement biomechanics.

Thus, the prospects for applying AI in equestrian sports lie in the creation of multi-level analysis systems that combine NLP and DL approaches for the early detection of pathologies, the optimization of training loads, and an increase in the effectiveness of recovery. Their implementation is capable of significantly changing the quality of medical support for equine athletes, ensuring a transition from reactive treatment to preventive and personalized control.

In summary, the comparative analysis confirms that Amazon Bedrock, Comprehend, and Comprehend Medical constitute complementary components of a scalable AI ecosystem for veterinary data analytics. Bedrock ensures contextual reasoning and adaptive response generation; Comprehend Medical provides high-precision entity extraction; and standard Comprehend supports multilingual generalization. Together, they establish the technical foundation for next-generation veterinary decision-support systems capable of transforming unstructured equine medical records into actionable intelligence.

CONCLUSIONS

The study conducted has shown that the application of modern technologies for analyzing medical records in veterinary medicine opens new possibilities for improving the quality of support for sport horses. The use of text processing algorithms and machine learning methods allows for the structuring of large volumes of clinical information and the formation of predictive models aimed at the early detection of diseases and the monitoring of recovery after exertion.

It has been established that the most effective are comprehensive approaches, where traditional data structuring methods are combined with intelligent analysis systems. This symbiosis provides higher accuracy, allows for the identification of hidden patterns, and expands the possibilities for developing personalized animal management programs.

The identified data gaps, primarily the deficit of information concerning horses, indicate the need to create specialized arrays of clinical records and their standardization. Without this, the technological potential cannot be fully realized.

The key directions for further development are: the formation of broad and representative databases, the adaptation of analysis methods to the specifics of sports medicine for horses, the implementation of multi-level systems that combine textual and numerical analysis, and the strengthening of mechanisms for verifying the reliability of results.

The realization of these tasks will make it possible to take the veterinary support of equestrian sports to a new level, ensuring a transition from reactive treatment to preventive and individualized control. Ultimately, modern technologies can become the basis for reducing injuries, increasing athletic performance, and strengthening the health of equine athletes.

REFERENCES

1. Akbarein, H., Taaghi, M. H., Mohebbi, M., & Soufizadeh, P. (2025). Applications and considerations of artificial intelligence in veterinary sciences: A narrative review. Veterinary Medicine and Science, 11(3), e70315. https://doi.org/10.1002/vms3.70315
2. Alonso Sousa, S., Bukhari, S. S. U. H., Steagall, P. V., Bęczkowski, P. M., Giuliano, A., & Flay, K. J. (2025). Performance of large language models on veterinary undergraduate multiple-choice examinations: A comparative evaluation. Frontiers in Veterinary Science, 12, Article 1616566. https://doi.org/10.3389/fvets.2025.1616566
3. Chu, C. P. (2024). ChatGPT in veterinary medicine: A practical guidance of generative artificial intelligence in clinics, education, and research. Frontiers in Veterinary Science, 11, Article 1395934. https://doi.org/10.3389/fvets.2024.1395934
4. Davies, H., Nenadic, G., Alfattni, G., Arguello Casteleiro, M., Al Moubayed, N., Farrell, S., Radford, A. D., & Noble, P. M. (2024). Text mining for disease surveillance in veterinary clinical data: Part two, training computers to identify features in clinical text. Frontiers in Veterinary Science, 11, Article 1352726. https://doi.org/10.3389/fvets.2024.1352726
5. Davies, H., Noble, P.-J., Fins, I. S., Pinchbeck, G., Singleton, D., Pirmohamed, M., & Killick, D. (2025). Developing electronic health records as a source of real-world data for veterinary pharmacoepidemiology. Frontiers in Veterinary Science, 12, Article 1550468. https://doi.org/10.3389/fvets.2025.1550468
6. Farrell, S., Appleton, C., Noble, P. J. M., & others. (2023). PetBERT: Automated ICD-11 syndromic disease coding for outbreak detection in first opinion veterinary electronic health records. Scientific Reports, 13, Article 18015. https://doi.org/10.1038/s41598-023-45155-7

7. Fins, I. S., Davies, H., Farrell, S., Torres, J. R., Pinchbeck, G., Radford, A. D., & Noble, P.-J. (2023). Evaluating ChatGPT text mining of clinical records for companion animal obesity monitoring. Veterinary Record, 194(3), e3669. https://doi.org/10.1002/vetr.3669

8. Macleod, B. M., Wilkins, P. A., McCoy, A. M., & Bishop, R. C. (2025). Integration of machine learning and viscoelastic testing to improve survival prediction in horses experiencing acute abdominal pain at a veterinary teaching hospital. Equine Veterinary Journal. Advance online publication. https://doi.org/10.1111/evj.14517

9. Nazi, Z. A., & Peng, W. (2024). Large language models in healthcare and medical domain: A review. Informatics, 11(3), 57. https://doi.org/10.3390/informatics11030057

10. Xiao, S., Dhand, N. K., Wang, Z., Hu, K., Thomson, P. C., House, J. K., & Khatkar, M. S. (2025). Review of applications of deep learning in veterinary diagnostics and animal health. Frontiers in Veterinary Science, 12, Article 1511522. https://doi.org/10.3389/fvets.2025.1511522